



Tijdschrift voor onderwijsresearch 1994

<https://hdl.handle.net/1874/208872>

TIJDSCHRIFT VOOR ONDERWIJS RESEARCH

ISSN 0166-591 X

Jaargang 19
1994

Swets & Zeitlinger B.V.
Lisse

Redactie:

Bert P.M. Creemers (secretaris)
Christiaan Hamaker
Bernadette van Hout Wolters
Leo J.Th. van der Kamp
Joost Lowyck
G. Wim Meijnen
Johan M.M. van der Sanden
Jaap Scheerens
Marinus J.M. Voeten
Wynand H.F.W. Wijnen

Univ. of Utrecht
BIBLIOTHEEK CENTRUM UITHOF

SW

VERENIGING VOOR ONDERWIJSRESEARCH

RIJKSUNIVERSITEIT TE UTRECHT



2510 550 7

TIJDSCHRIFT VOOR ONDERWIJSRESEARCH

Een drie-maandelijks tijdschrift voor onderzoekers van onderwijs, gewijd aan theoretische, methodologische en professionele problemen in de onderwijsresearch.

Redactie:

Bert P.M. Creemers (secretaris)
Christiaan Hamaker
Bernadette van Hout Wolters
Leo J.Th. van der Kamp
Joost Lowyck

G. Wim Meijnen
Johan M.M. van der Sanden
Jaap Scheerens
Marinus J.M. Voeten

Redactiesecretariaat:

Annelies Weersing
GION
Postbus 1286
9701 BG Groningen
tel. 050-636673

Manuscripten zende men in drievoud, getypt en persklaar aan Bert P.M. Creemers, GION, Postbus 1286, 9701 BG Groningen.

Manuscripten dienen te worden voorbereid overeenkomstig de Aanwijzingen voor auteurs.

De abonnementsprijs bedraagt f. 132,50; voor studenten f. 109,00; voor instituten f. 204,00.

Men kan zich abonneren bij de uitgever: Swets & Zeitlinger, Heereweg 347, 2161 CA Lisse. Voor leden van de Vereniging voor Onderwijsresearch (VOR) is het abonnementsgeld reeds inbegrepen in de contributie. Leden van de VOR hoeven zich dan ook *niet* afzonderlijk te abonneren bij de uitgever.

Losse nummers zijn (à f. 35,00) verkrijgbaar bij de uitgever.

Advertenties dienen te worden gezonden aan het adres van de uitgever. Tarieven: 1/1 f. 450,00; 1/2 f. 260,00; 1/4 f. 190,00; exclusief zetkosten.

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotocopie, microfilm of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

Het *Tijdschrift voor Onderwijsresearch* wordt uitgegeven door de Vereniging voor Onderwijsresearch.

Vereniging voor Onderwijsresearch

Secretariaat:
Dr. P.J.J. Stijnen
Open Universiteit
Postbus 2960
6401 DL Heerlen
tel. 045-762293

Abonnementenadministratie:

Swets & Zeitlinger
Heereweg 347
2161 CA Lisse
tel. 02521-35111
giro 13984

Mededeling van de uitgever

Jaargang 19 (1994) van het *Tijdschrift voor onderwijsresearch* zal in 4 nummers verschijnen in plaats van in 6. Het aantal bladzijden voor de gehele jaargang blijft echter wel 384.

Van de redactie

De redactie kijkt met tevredenheid terug op het jaar 1993. Het is gelukt om de grote achterstand die in 1992 was ontstaan in de publikatie van de verschillende afleveringen van het tijdschrift enigszins in te halen. Alleen nummer 6 van jaargang 18 zal in 1994 verschijnen.

Het feit dat ongeveer alle nummers in het betreffende jaar konden verschijnen is te danken aan de auteurs, de redactieleden en niet in de laatste plaats aan de uitgever Swets & Zeitlinger die waar mogelijk productieprocessen heeft versneld. De auteurs werden in de voorgaande zin bedankt, omdat zij bereid waren om hun manuscripten aan het Tijdschrift voor Onderwijsresearch aan te bieden, maar ook om vervolgens op korte termijn opmerkingen en kritiek van de redactie te verwerken.

Graag zou de redactie nog meer mensen i.c. auteurs bedankt hebben. Alhoewel het aanbod van manuscripten op dit moment redelijk is, zou de redactie graag willen dat veel meer manuscripten werden aangeboden. De verhouding aantal lezers – aantal auteurs is ongunstig, in die zin dat slechts een betrekkelijk gering aantal lezers, gevormd door de professionele beroepsorganisatie de Vereniging voor Onderwijsresearch, ook als auteur in het tijdschrift blijkt op te treden.

Het Tijdschrift voor Onderwijsresearch bestaat bij de gratie van de Vereniging voor Onderwijsresearch. Dat blijkt uit het feit dat de vereniging optreedt als uitgever en al een groot aantal jaren het tijdschrift op alle mogelijke wijzen ondersteunt, maar het is ook het tijdschrift van de leden die het lezen en die ook als auteur optreden c.q. zouden moeten optreden.

Het is duidelijk dat dit een bijna agressieve oproep tot het aanbieden van manuscripten is. De redactie ontvangt ze graag en zal waar mogelijk haar bijdrage leveren om manuscripten te publiceren.

De regelmaat van uitgave die het afgelopen jaar bereikt is, is voornamelijk te danken aan de samenstelling van themanummers. De redactie hecht er aan de gastredacteuren op dit moment nog eens extra te bedanken voor hun bijdrage aan het samenstellen en het redigeren van deze themanummers.

De redactie heeft de indruk dat de themanummers door de lezers gewaardeerd worden. Ze verschaffen een overzicht van onderzoek en onderzoeksverslagen op een specifiek, maar interessant terrein van het onderwijs en de redactie zal dan ook het beleid van het publiceren van een aantal themanummers per jaar voortzetten. Echter, ze hecht eraan om daarnaast ook niet gethematiseerde nummers uit te brengen en het vrije aanbod van bijdragen een kans te geven. Deze nummers zullen een weerspiegeling zijn van het scala van onderwijsonderzoek dat plaatsvindt.

De verschijningsfrequentie van het tijdschrift zal worden teruggebracht van zes naar vier nummers per jaar. Deze nummers zullen echter omvangrijker zijn. Dit brengt met zich mee dat naast de grotere themanummers het ook mogelijk zal zijn om in een nummer, waarin een bepaald thema wordt behandeld, niet gethematiseerde bijdragen op te nemen.

Het moge duidelijk zijn uit het voorafgaande dat de redactie graag bijdragen en voorstellen voor themanummers blijft ontvangen. Wat themanummers betreft stelt de redactie een vroegtijdig contact met de samenstellers op prijs om afspraken te maken over het thema, het aantal bijdragen, de inhoud van de bijdragen, en tenslotte om een tijdschema af te spreken.

Wynand Wijnen, voorzitter van de redactie, heeft per 1 januari 1994 de redactie verlaten in verband met het groeiende aantal andere verplichtingen. Wynand Wijnen was al lid van de redactie in de periode 1975 tot en met 1982. In die tijd vertegenwoordigde hij de didacometrische (wie herinnert zich dit woord?) onderdelen van het Nederlandse Tijdschrift voor de Psychologie. Na een onderbreking van een aantal jaren kwam hij in 1987 weer terug in de redactie, nu als representant van het onderzoek in het wetenschappelijk en hoger onderwijs. Geheel in overeenstemming met zijn deskundigheden werd hij natuurlijk voorzitter van de redactie. De redactie bedankt hem voor de zakelijke en efficiënte wijze waarop hij in de afgelopen jaren de redactie heeft geleid.

Rinus Voeten zal vanaf 1 januari 1994 het voorzitterschap overnemen.

Effective Instruction as a Basis for Effective Education in Schools

Bert P.M. Creemers

*GION, Institute for Educational Research, University of Groningen**

ABSTRACT

Outcomes of instruction greatly depend on aptitudes, social background and motivation of students. But instruction also contributes to educational outcomes. Important as mediating variables in the instructional process are time available for learning and the learning experiences which are provided during this academic learning time and the opportunity to learn. Factors within the instructional process can influence academic learning time and opportunity to learn. Important components of the instructional process at the classroom level are the teachers, curricula and grouping procedures. Especially teachers are important because they carry out instruction in the classroom making use of curricular material and grouping procedures.

Based on reviews of research one can expect larger effects when the effective characteristics of the different components are consistent with each other.

1. INTRODUCTION

British (Mortimore et al., 1988) and Dutch studies (Brandsma and Knuver, 1989, p. 202) show that factors at the school level and the classroom level can explain between 9 and 27 per cent of attainment variance and about 20 per cent of the total variance of student progress (Brandsma, 1993). These factors at the school and classroom level can be influenced to increase the effectiveness of education. Initially, research was especially directed to factors at the school level. It was expected that the (educational) organization of the school greatly contributed to the effectiveness of it. Later on, the research was criticized for neglecting the classroom level and theories were developed including factors at different levels of education (Scheerens, 1992; Stringfield and Slavin, 1992; Creemers, 1992).

Furthermore, attention was given to the theoretical basis of these conceptual frameworks. For example, Scheerens advocated the contingency and public choice theory. Others took as a point of departure the learning process of the student, expecting that all levels, like the classroom, the school and the wider educational context, are expected to contribute to the students' learning and the outcomes of the learning process. This implies a search for factors at different levels that can be related to students' learning. In that respect, the classroom level is especially important, because it is the environment in which predominantly students' learning takes place. In this way educational effectiveness primarily depends on the instructional process at the classroom level. In this article we will further develop this idea and look for components at the classroom level and their characteristics that can 'create' instructional effectiveness (section 3). This search will be led, as indicated before, by Carroll's theory about learning in school (section 2).

* Address: GION, P.O. Box 1286, 9701 BG Groningen.

2. STUDENT LEARNING AS A POINT OF DEPARTURE FOR EDUCATIONAL EFFECTIVENESS

Carroll (1963) stated that the organization of education should be based on a theory that can explain school learning of students, and that is sufficiently empirically supported. To find out if teacher activities are effective, it is necessary to view education at the classroom level from the perspective of students' learning.

Carroll developed a theory that takes students' learning in schools (and classrooms) as a point of departure. Variables at the classroom level were explicitly involved in this theory. Later on, in the 1960's and 1970's, several other important models were introduced, like those of Bruner (1966), Bloom (1976), Harnischfeger and Wiley (1976), Glaser (1976) and Bennett (1978). All these models acknowledge students' aptitude, background and motivation and, in addition, instructional factors on the school's side as important constructs for (school) learning (Haertel, Walberg and Weinstein, 1983).

Some of these models of school learning got empirical support during the last decennia. Therefore, they are used as a point of departure for the evaluation of factors that contribute to learning. Furthermore, these models of school learning may be extended to educational theories contributing to effective education, by adding elements at the classroom level as well as the school level.

However, Walberg was not so positive in his review of learning theories for education (1986). He points out that, although all theories yield some good ideas, theories were not sufficiently supported empirically, and research and practice do not have enough to go by.

This does rather less apply to Carroll's model. In developing his model, Carroll explicitly aimed at designing a learning theory that would be of use for educational practice. Especially because of the elaboration Bloom provided within an instructional frame (although Carroll (1989) thinks this is a rather mechanical and technical elaboration of his original intentions), the influence of his learning theory on educational practice was substantial.

The Carroll model states that the degree of mastery is a function of the ratio of the amount of time students actually spend on learning tasks, to the total amount of time they need. Time actually spent on learning is defined as equal to the smallest value of three variables:

- a) opportunity (time allowed for learning);
- b) perseverance (the amount of time students are willing to engage actively in learning);
- c) aptitude (the amount of time needed to learn, under optimal instructional conditions).

This time might be increased by extra time needed, because of poor quality of instruction and lack of ability to understand less than optimal instruction (Carroll, 1963, p. 730).

The Carroll model can be criticized for being an instructional model more than a learning model. In fact, it does not provide information about how learning itself takes place, only that learning takes time and that this depends on multilevel interrelated factors. On the other hand, the general framework itself provides possibilities to elaborate on the different components.

Especially in recent years the idea of how to learn as an objective of education was given more emphasis. In addition to knowledge and skills, students should acquire learning strategies. A strategy concerns the knowledge and skills of how to proceed in a task and how to acquire knowledge. Strategies are also called meta-cognitive skills. A meta-cognitive skill consists of activities such as orientation, planning, testing, evaluation and reflection. From a review by Wang, Haertel and Walberg (1990), it can be concluded that meta-cognitive skills can predict learning outcomes. So far we expect that the same factors that are contained in Carroll's model contribute to achieving objectives in the area of learning strategies. Authors about meta-cognition state that other learning processes as well as other instructional strategies are required for meta-cognition than for (common) objectives like skills and knowledge (Palincsar and Brown, 1989; Prawat, 1989). On the other hand, Prawat emphasizes the fact that the acquisition of

“ordinary” cognitive knowledge and skills is required for this more active learning directed to meta-cognitive knowledge and skills.

3. EFFECTIVE COMPONENTS OF INSTRUCTION

The model of educational effectiveness was developed according to the factors of students' learning in Carroll's model. Like other models developed by Scheerens (1992) and Stringfield and Slavin (1992), the model contains more levels: student, classroom, school and educational context. The model is a combination of various components, that can contribute to achievement to a higher or lesser extent. The basic idea of the model is that student achievement is influenced by effective learning time and opportunity to learn and quality of educational processes at different levels next to several student background factors. Education, i.e. instruction at the classroom level is supposed to influence learning time and student achievement (directly or indirectly, via extension of learning time). We expect that variables at the school level and at the context level influence the quality of instruction, as well as the opportunity to learn and the time spent on learning.

Next to these indirect effects of the context and school level on education at the classroom level, there are also direct but weak influences of factors at these levels on student learning (see Brandsma, 1993; Knuver, 1993; R.H. Hofman, 1993; W.H.A. Hofman, 1993).

Essential at the classroom level are academic learning time and opportunity to learn, because they are mediating all the other variables at the school level, except for the student variables. At the student level these two variables appear again, now in the way students make use of the time for learning (academic learning time) and the opportunities to learn (opportunities used).

At the classroom level the model distinguishes three components: curricula, grouping procedures and teacher behaviour. It is expected that the effective characteristics of these components affect time and opportunity and contribute to student outcomes. In the remainder of this section the effective characteristics of the components will be described.

3.1. Curriculum materials

Curricula are important instruments for the planning and implementation of education at the classroom level and for the effectiveness of classroom instruction. What teachers will teach and what students will learn is laid down in curricula. Curricula create opportunity to learn; achievement is expected to be higher when curricula cover more subject matter. Research shows that opportunity to learn is an important variable at the school level, accounting for variations in student achievement. As is apparent from Alexander's research, the substance of the curriculum needs more attention instead of physical and organizational arrangements (Alexander, 1992, p. 57).

Comparative research on curricula and elements of curricula has yielded characteristics that are important for the effectiveness of instruction. These characteristics do not only concern the objectives curricula try to achieve, but also the structuring of objectives: the hierarchical organization of learning objectives and the ordering of subject matter. A related curricular characteristic is the degree of control of achieving objectives of lesson units and the extent to which individual students or groups of students are offered corrective procedures based on this control.

These curricular characteristics enhance effective learning time and, by this, student achievement. Certain effective characteristics, such as evaluation and corrective procedures, also influence student motivation and, by this, indirectly, effective learning time. However, the actual effectivity of these curricular characteristics is bound by the extent to which teachers actually make use of curricula and textbooks. When teachers are more or less free to act as they want to (because no support is offered to stimulate fidelity implementation of curricula, or because curricula do not incorporate guidelines for implementation), they think of curricula as a source of ideas, especially concerning the subjects to be covered and the order of subjects, but in

general they do not hold on very strictly to curricula when they are teaching. When curricula demand certain teacher activities, there are indications that teachers always want and need to have ample freedom to adapt education to their classrooms.

Therefore, capitalizing on the design of curricula and on fidelity implementation to enhance educational effectiveness, does not seem to be a very successful approach.

Nevertheless, a basis for effective instruction can be created by emphasizing educational elements that are inherent to curricula and are acknowledged by teachers as such in the design of curricula. These elements are the subjects to be taught in classroom instruction, the structuring and ordering of educational objectives and subject matter, and, related to these, evaluational procedures. Support in the implementation of curricula should also focus on these elements. The further design of education should be regarded as the professional responsibility of teachers, which they should acquire in their initial training and further schooling.

3.2. Grouping procedures

In reviewing the results of the analyses of within-class grouping, moderately positive conclusions can be drawn, on the basis of international literature, about the effects of grouping procedures. The positive results concern, in a descending degree of effectiveness, grouping on the basis of pacing (forms of mastery learning), grouping on the basis of ability, and cooperative learning. Especially the concept of mastery learning, in several forms (and in combination with other grouping procedures and educational strategies) was shown to be effective at all levels of education.

United States analyses often state that effectiveness of grouping is dependent on the way teachers implement grouping (Gamoran, 1986). Dutch research also points at the importance of teachers and curricula. The effects of grouping within classrooms may be shaped to a higher extent by teachers than by grouping procedures. This might be the most important explanation for the disappointing effects of grouping in Dutch education. Some analyses refer to curricula and the availability of tests; but the effects are largely determined by the way teachers use these procedures (Reezigt, 1993).

In general, adaptive instruction, which is instruction geared to learning characteristics and needs of individual students, is more effective than whole-class instruction with respect to cognitive as well as to affective measures. This conclusion is empirically affirmed, especially by meta-analytic studies (see, for example, the recapitulation of several meta-analyses by Waxman et al., 1985). They concluded that adaptive instruction improves student learning. The programs studied were quite different, but programs in the synthesized studies featured at least one of the following adaptive characteristics:

- Instruction is based on the assessed abilities of each student.
- Students work at their own pace.
- Students receive periodic reports on their mastery.
- Students plan and evaluate their own learning.
- Alternative materials and activities are provided.

To make it more complicated, according to these characteristics, adaptive instruction does not necessarily imply some form of grouping. Adaptive instruction is not restricted to the grouping of students in subgroups but can also be carried out in whole-group instruction.

In the end, grouping procedures are directed to increase the effectiveness of instruction by organizing instruction at the classroom level in such a way that instruction can be adapted to the individual characteristics of students. By grouping of students, time and opportunities for learning can be more adequately allocated. This means that time and opportunities are offered to individual (or groups of) students and that the learning environment is organized in such a way that the students' engagement in the learning task, and therefore the outcomes of the instructional process, are maximized. The effects of grouping are strongly determined by the "quality" of the effective characteristics of the instruction within a grouping procedure. To make the proce-

For effective grouping it is required that differentiated material is available. For curricular materials, the list of effective characteristics mentioned is applicable. Especially important with respect to grouping are the characteristics related to evaluation, feedback and corrective instruction, because in this way teachers can adapt instruction to students, the ultimate objective of grouping procedures. Applying this idea to the earlier mentioned list of features of adaptive instruction (Waxman et al., 1985), we expect especially positive effects of those features which can be clearly related to time, opportunity and quality of instruction. Grouping procedures do offer a frame and a structure, as curricula do; however, what actually happens in classrooms is decided by teachers. In the end, they are responsible for achieving effects for several groups of students.

3.3. Teacher behaviour

Teacher behaviour especially important in this respect encompasses the allocation and use of learning time and classroom management aimed at an orderly atmosphere to promote learning. It also encompasses teacher activities in several educational components, such as structuring the content, questioning, evaluation, feedback and corrective instruction.

These teacher behaviours have positive effects on student achievement. The effects can be expected to increase, as some experimental studies show, when teachers are trained in showing these behaviours (Bennett and Carré, 1993). On the basis of the effects of the model of direct instruction as well as of derivated or comparable educational models, a combination of the above-mentioned teacher behaviours can be expected to be even more effective.

For the quality of instruction the following characteristics of teacher behaviour are important.

- Management of the classroom in order to create a situation where learning can take place. This implies an orderly and quiet atmosphere in the classroom. Although, as mentioned before, learning itself requires more than a well-organized class. Moreover, effective teaching itself contributes to the management of the class.
- Provision of homework. If properly organized, homework contributes to effectiveness. This implies a clear structure of assignments, and supervision and evaluation of homework.
- Expectations teachers (and schools) have of their possibilities to influence student outcomes probably influence what teachers do. We expect that this becomes apparent in actual teacher behaviour.
- Clear goal setting. This includes a restricted set of goals, emphasis on basic skills and emphasis on cognitive learning and transfer. The content should be chosen in line with these goals.
- Structuring the content. This includes the ordering of the content according to the (hierarchically ordered) goals. The use of advance organizers can also structure the content for students. The use of prior knowledge can increase students' own contributions and responsiveness to learning.
- Clarity of presentation, which implies the above-mentioned elements, but also refers to the transfer process itself (avoiding vagueness and cripple sentences).
- Questioning (by means of low and higher order questions) keeps students at work and can be used to check their understanding.
- Immediate exercise after presentation. Like questioning it provides a check for understanding and can be used to clarify problems.
- Evaluating whether the goals are obtained by testing, providing feedback and corrective instruction.

The presented characteristics are based on research on teaching and school effectiveness research. It is not an enumeration of all (possible) correlates, but a choice of characteristics based on empirical evidence on the one hand and on our theoretical notions on the other hand.

The relation between teacher behaviour in classrooms and the other elements of the teaching-learning situation, i.e. curricula and grouping procedures, should be emphasized. Rosenshine (1983), and later also Rosenshine and Stevens (1986) state, in discussing active teaching, that it

is important that guidelines for teacher behaviour have been put together now; however, they should also be integrated in curricular teacher manuals and in teacher training.

Teacher behaviour is often partly elicited, planned or directed by curricula or grouping procedures, which can be more or less effective in their own way. We expect that better tuning of the components of instruction will induce a higher correlation with outcome measures. A next step is to check whether it is possible to relate these elements in some way or another, in order to enhance the effectiveness of instruction.

4. EMPIRICAL EVIDENCE FOR THE CHARACTERISTICS OF EFFECTIVE INSTRUCTION

In the previous sections we developed parts of a basic model of educational effectiveness. The basic model itself was based on Carroll's model of school learning. Essential in his model are time for learning and quality of instruction, next to the aptitude of students.

We added the concept of opportunity to learn to this model. Time and opportunity are discerned both at the classroom level and at the student level. In this way, we make a distinction between available and actually used time and opportunity. We distinguished three components within quality of instruction at the classroom level: curricular materials, grouping procedures and teacher behaviour. We reviewed the literature to find characteristics of these components that contribute to effectiveness. We included characteristics based on two arguments: theoretical notions and empirical evidence.

Theoretical notions are at the core of the basic model, especially, the expectation that the characteristics which constitute quality of instruction are related to time and opportunities for learning offered to the students.

The leading idea was not to include as many characteristics as possible, but to develop an economic and relevant as well as an insightful model. As a consequence we only included major characteristics, which can be considered as "collections" of several minor characteristics. An example of such a minor characteristic is wait time, which can explain variance in student outcomes, but is so detailed that it confuses or mixes up the theoretical framework. We regard wait time, more precisely "taking into account at least 3 seconds between phrasing the question and asking for answers" as a qualitative dimension of questioning. There are more guidelines for "good questioning" than wait time only. For the theoretical model, "questioning", including all these qualitative aspects, is sufficient.

We also left out characteristics which could be confusing between the components of effective instruction. An example is pacing, an effective characteristic of teacher behaviour. The way it is usually defined, it is related to grouping procedures, which take care of the pace of instruction, adapted to groups of students. So we decided to leave pacing out of the list of characteristics of teacher behaviour, but we assume that the effectiveness of grouping procedures is determined by the fact whether teachers are able to pace the instructional process.

Sometimes we included characteristics in the model even when the empirical evidence for them was quite small. For example, this is the case with the curriculum, which turned out not to be a very influential factor in itself. We assume that the contribution of the curriculum to effectiveness can be enlarged when it will be developed according to the guidelines described, and when the components are interrelated.

In summary, the model was developed departing from a theoretical model about learning in school. The addition of variables to the model was judged by theoretical notions and empirical evidence.

The three components as such got empirical support, but the distinction between curriculum and teaching behaviour is problematic in this respect. Teacher behaviour is possible without curricular materials, but to investigate the effectiveness of curricular materials we need teachers and

teacher behaviour. Research designs to find out the specific contribution of curricular materials are quite complicated to arrange in educational practice. The empirical evidence for effects of teacher behaviour and grouping procedures is strong, the empirical evidence for curricular materials is moderate.

The characteristics differ from each other with respect to the empirical evidence. For the curriculum research on mastery learning provides evidence for the importance of explicitness and the ordering of goals and content. For structure and clarity of content this is less obvious, but their importance can be assumed based on empirical evidence in the field of teacher behav-

Table 1. Overview of empirical evidence for the characteristics of effective instruction.

Characteristics	Strong empirical evidence	Moderate empirical evidence	Plausible
<i>Curriculum</i>		X	
Grouping procedures	X		
Teacher behaviour	X		
<i>Curriculum</i>			
Explicitness and ordering of goals and content	X		
Structure and clarity of content		X	
Advance organizers	X		
Evaluation	X		
Feedback	X		
Corrective instruction			X
<i>Grouping procedures</i>			
Mastery learning	X		
Ability grouping		X	
Cooperative learning			X
Differentiated material			X
Evaluation	X		
Feedback		X	
Corrective instruction		X	
<i>Teacher behaviour</i>			
Management/orderly and quiet atmosphere	X		
Homework	X		
High expectations		X	
Clear goal setting		X	
Restricted set of goals		X	
Emphasis on basic skills		X	
Emphasis on cognitive learning and transfer			X
Structuring the content		X	
Ordering of goals and content		X	
Advance organizers	X		
Prior knowledge		X	
Clarity of presentation		X	
Questioning	X		
Immediate exercise		X	
Evaluation	X		
Feedback		X	
Corrective instruction			X

ious. Moreover, the characteristics in the area of structuring need more research, probably after better definition. For advance organizers, evaluation and feedback the evidence is quite substantial. The importance of corrective instruction has to be assumed on the basis of more "circumstantial evidence": the fact that components of adaptive instruction, such as evaluation, feedback and individual working plans, turn out to be effective in grouping procedures like mastery learning.

Although there is a debate about the research evidence on grouping, some conclusions can be drawn. The empirical support for the grouping procedures is different. There is strong empirical evidence for mastery learning, for ability grouping the evidence is moderate, and for cooperative learning evidence can be expected, but we need further research. The characteristics of effectiveness of the grouping procedures got sufficient empirical support. This holds for the conclusions with respect to evaluation, and somewhat less for feedback and corrective instruction. The importance of these characteristics can also be deduced from the negative effects of grouping in educational practice. Empirical evidence is not so clear for differentiated material.

With respect to teacher behaviour, the empirical support for the characteristics mentioned is quite strong in general. Management and a quiet and orderly atmosphere were substantially supported. The same holds for homework, but the effects are highly dependent on the clearness of assignments, evaluation and feedback. High expectations are empirically supported. The same holds for goal setting. Some evidence for the effectiveness of a restricted set of goals comes from the research on mastery learning; with respect to basic skills, empirical support stems from school effectiveness studies, but it could not be replicated by other studies. Emphasis on cognitive learning and transfer can be expected to contribute to effectiveness, based on research on learning. With respect to structuring the content, the empirical evidence comes from quite different areas of research. The ordering of goals and content again comes from research in the field of mastery learning, the other aspects of structuring are supported by research on teaching. Clarity of presentation is found to be effective in teacher effectiveness studies. Immediate exercise after presentation was moderately supported in empirical studies. In teacher effectiveness research sufficient support can be found for questioning and evaluation, and moderate support for feedback. Corrective instruction can be expected to be effective.

Table 1 provides an overview of the empirical evidence for the distinguished characteristics.

5. INTEGRATION OF EFFECTIVE CHARACTERISTICS OF INSTRUCTIONAL COMPONENTS: THE CONSISTENCY PRINCIPLE

Research on education at the classroom level shows that isolated components or effective elements of individual components do not result in strong effects on student achievement. Good curricula need teachers who can make adequate use of them, and who show effective instructional behaviour. The same also holds for grouping procedures. Each component contributes to learning in classrooms, but the other components are necessary for substantial effects. Adequate management behaviour in classrooms will result in more time for learning, but good curricula, effective grouping procedures and adequate instructional behaviour are necessary to make the available learning time productive.

An integrated approach of education at the classroom level is necessary. In this integrated approach, the educational components curricula, grouping procedures and teaching behaviour are adapted to each other, they have the same characteristics of effectiveness. In this way these characteristics strengthen each other, which results in a synergetic effect. In case the characteristics are not in line with each other the effect can be minimized or even disappear (It can be assumed that effects of differentiation in the Netherlands do not occur because teachers do not behave according to the procedures).

Effective instruction should provide time and opportunities to learn. For that purpose, curricular materials like textbooks should be developed according to certain well-stated and clear goals. In using grouping procedures, it should be clear what kind of goals have to be achieved by

individual students or groups of students. Teacher behaviour should be guided by the goals teachers and students have to accomplish, although operationalizing goals is not meant to be a part of the instructional behaviour of teachers. When teachers make use of curricular materials, it should be clear that the goals they have in mind are also part of the curriculum they are using.

The same holds for structuring. Structuring should be an important characteristic of curricular materials and textbooks. Elements of structuring are the ordering of goals, the availability of advance organizers, links with prior knowledge, the ordering of subjects according to the goals, and immediate exercises after presentation of new content. The same characteristics show up in order to provide well-structured instruction. Teachers heavily depend on curricular materials and textbooks which make it possible to structure instruction. Actually, textbooks are, in a way, scaffolds for teachers. Another effective teacher characteristic, questioning, is also structured by curricular materials and textbooks. These materials provide teachers with all kinds of possibilities to raise questions and to adapt instruction to different groups of students. That is even more so with some other elements of curricular materials, like evaluation, provision of tests, possibilities for feedback and the inclusion of different kinds of corrective instruction. These elements can provide guidelines for effective teacher behaviour, and are of crucial importance to the effectiveness of grouping procedures.

As was said before, the same characteristics of effective teaching should be apparent in the different components. It is even more important that the actual goals, structuring, and evaluation in curricular materials, grouping procedures, and teacher behaviour are in the same line. This does not mean that they have to be the same, but they should support each other. They should be consistent with each other. In this way a synergetic effect can be achieved.

Consistency of the effective characteristics of the main components of instruction can cause a synergetic effect, an effect that is greater than the sum of the effects of the separate components. We can observe such an effect in the implementation of new, experimental curricula in which teachers in fact contribute to the effect of the curriculum (more than the curriculum itself). The opposite can be seen in the negative results of the introduction of material in situations in which teachers did not have the chance to adapt themselves to the curriculum, or to adapt the curriculum (Olson and Eaton, 1987; Kuiper, 1993). Figure 1 provides an overview of the components and characteristics of effective instruction at the classroom level.

The question could be raised as to whether it is possible to create instructional arrangements of the different components, including effective characteristics, that are in line with each other, mutually reinforce each other and have a synergetic effect that exceeds the effectiveness of separate components. Direct instruction and mastery learning show to some extent that this is possible.

Moreover, some successful school improvement projects were developed more or less in line with the principles explained in this study. The Milwaukee project combines the curriculum, grouping (whole-class instruction at the grade level and small group instruction at the performance level) and instruction at the classroom level. The results showed a significant progress of the schools with respect to the outcomes in mathematics and reading (McCormack-Larkin, 1985). Other examples are "Success for All" (Slavin, Karweit and Madden, 1989), the project of the Halton Schoolboard (Stoll and Fink, 1992) and the program in Sederoth (Bashi et al., 1990). In these programs, empirically effective factors of components in classroom instruction processes are integrated, and supported by effective factors at other educational levels (school level or national level).

School improvement projects should be evaluated very carefully, in order to know what contributes to effectiveness and what should be changed. Sometimes school improvement projects are more focused on improvement than on careful evaluation, but the programs in Sederoth and "Success for All" are carefully evaluated. The project in Sederoth shows impressive results, although the stimulating Israeli context and its high expectations must be taken into account in

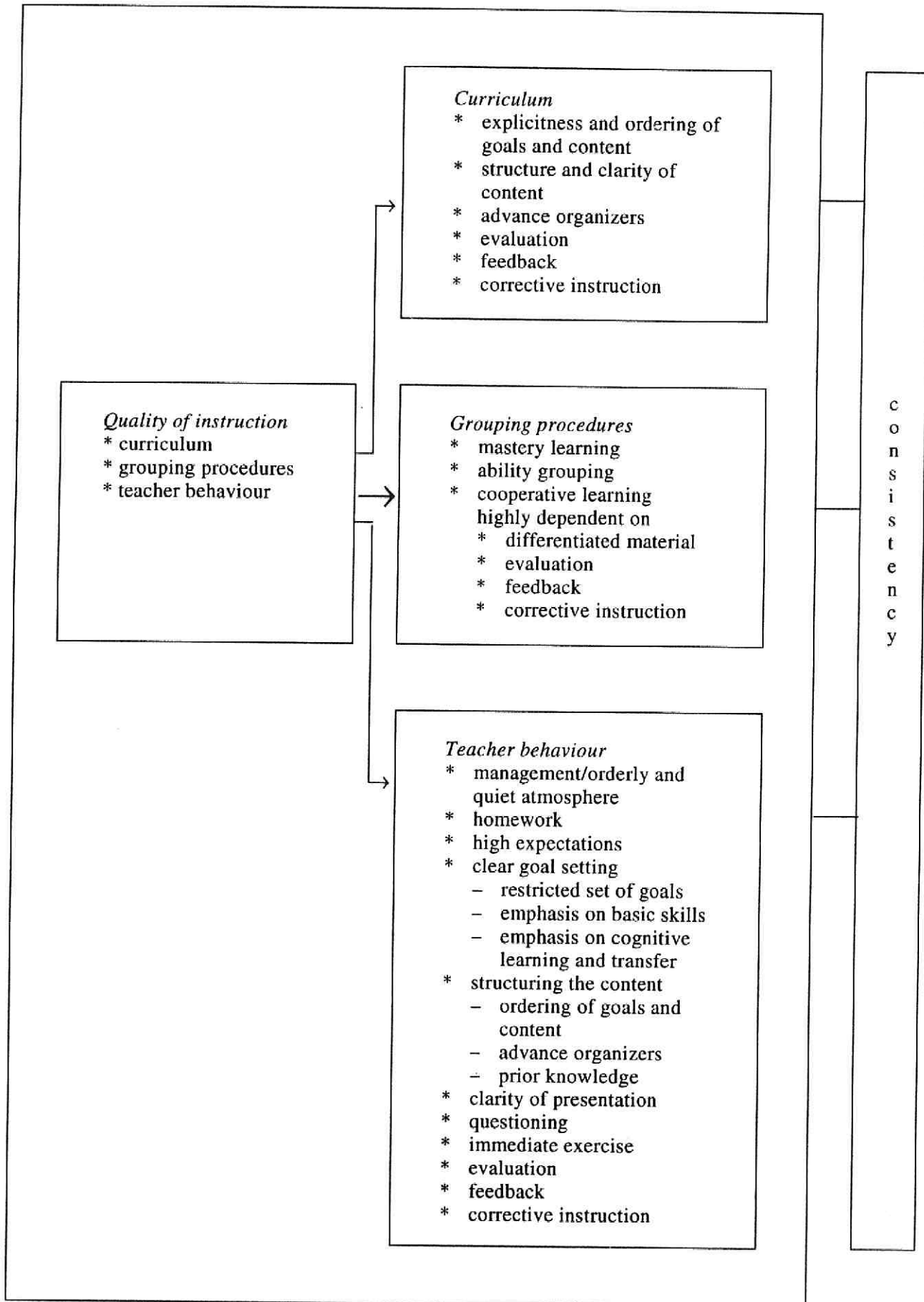


Fig. 1. Basic model of educational effectiveness: consistency of effective characteristics and components.

the interpretation of results. However, high expectations are one of the important factors for increasing effectiveness. Essential for the Sederoth program at the classroom level is the emphasis on achieving (intermediate) objectives. For that purpose, mastery learning is used with a possibility to provide extra time to master objectives.

The program developed by Slavin also shows favourable results (Slavin et al., 1990). This program too combines effective characteristics of curricular materials, grouping procedures and – again – the provision and effective use of time to master objectives. In this program, the emphasis on prevention of reading problems and immediate remediation of reading difficulties in a tutoring system is essential and impressively contributes to the results.

Unfortunately, research on effective instructional arrangements in school improvement projects is limited, so that support for synergetic effects cannot be illustrated. In these projects, the evaluation was directed to establish the fact that improvements were put into practice by teachers or established in curricular materials. But the ultimate test for effectiveness was not included: achievement test results of students, in connection with improvements and instructional processes. Some school improvement projects concentrated on factors at the school level, for example team building, the organization of the school, and the development of evaluation policy. They were not directed primarily at the classroom level, although some of them are supposed to link to increasing effectiveness at the classroom level. These projects hypothesize that teachers become more effective in their instructional behaviour based on the guidelines they receive from principals. The results of some of these projects are encouraging (Stringfield and Teddlie, 1989), although Freiberg et al. (1990) point to the fact that the results tend to decline in schools in severely disadvantaged areas when they are no longer supported.

In the Netherlands, improvement projects and research projects are being prepared or started already, based on the principles described in this study. Most of these projects take place in the context of the Dutch Educational Priority Policy for disadvantaged children. An example within this policy is the Amsterdam EGAA-project, which strives for a general improvement of education in 14 disadvantaged schools. This project makes use of the results of research on effective instruction and effective schools and concentrates on reading, on language instruction and on the implementation of a student monitoring system. The first research results are positive (Haanen et al., 1990). A program that focusses on the classroom level is directed to improve reading instruction (Osinga, Boersma and Houtveen, 1993).

The results of these kinds of projects are supported by the results of the evaluation of the Educational Priority Program. In general, outlier-analyses of high and low effective schools show that high effectiveness schools and low effectiveness schools can be distinguished based on characteristics of effective instruction, especially in the area of teacher behaviour and curricular materials (Van der Werf and Weide, 1991; Weide, 1993).

A similar conclusion of the importance of the combination of characteristics of effectiveness in classroom instruction can be drawn based on some of the programs described earlier, directed to improve education towards higher effectiveness. Although these programs want to change the whole school, the starting point is the classroom level. The Louisiana School Effectiveness Study (Teddlie and Stringfield, 1993) provides an empirical argument for starting at the classroom level and to address more than one component at the same time. Effective schools are schools that can achieve effective classrooms, showing a combination of the effective characteristics of the different components.

Mostly, this synergy between components is not developed carefully and is more or less arbitrary even in improvement programs, a result of different development activities which by incident turn out to be congruent or supplementary. Consistency between the components and the effective characteristics should be systematically established, and, once achieved, consistency can be expected to have a synergetic effect, that is to increase the effects of the separate components and characteristics.

6. CONCLUSIONS

In the previous section, a plea was made for instructional arrangements of all the components and the effective characteristics of components. Most of the time improving education at the classroom level starts with a teacher making a decision about grouping procedures and the choice and use of curricular materials. This is where effective instruction at the classroom level starts. Effective instruction cannot develop from scratch in an empty space. Effective instruction starts with teachers in classrooms.

This has two important implications:

1. Teachers as a central factor have to make a lot of decisions at the classroom level: about goals, the allocation of time to groups of or individual students about the use of material, their own instructional and management behaviour during the instructional process. The complexity of the instructional process requires serious planning, in which consistency can be systematically developed.
2. To guide the planning of the instructional process by teachers and the development of effective arrangements, central guiding ideas are of crucial importance. These can be found, in addition to the goals of education, in theories and research about learning and teaching, and in theories and research about quality and effectiveness of instruction. Time for learning and opportunity to learn can be used as mediating constructs. They can guide the development of effective instruction.

Ultimately, teachers are the crucial factor in education at the classroom level. They should implement curricular materials and grouping procedures adequately, and show effective instructional behaviours. Considering the actual context, in which education is provided by teachers to a large extent, it is important for teachers to train and practise these effective behaviours.

Teachers should let themselves be guided by good curricula and effective grouping procedures in preparing and implementing education at the classroom level. Moreover, teachers should have a repertoire of capacities that enables them to structure instruction, order content adequately, ask questions, use tests, and give feedback to their students.

We restricted ourselves to instruction at the classroom level. At the school and context level conditions can be formulated that can enhance effectiveness at the classroom level. These conditions should be related to what happens at the instructional level and especially to time and opportunity and quality of instruction. It is assumed that effective factors at other levels influence either the time, the opportunity or the quality of instruction at the classroom level.

REFERENCES

- Alexander, R. (1992). *Policy and practice in primary education*. London: Routledge.
- Bashi, J., Sass, Z., Katzit, R. and Margolin, J. (1990). *Effective schools from theory to practice: an intervention model and its outcomes*. Jerusalem: Nevo Publ. Ltd.
- Bennett, S.N. (1978). Recent research on teaching: a dream, a belief, and a model. *British Journal of Educational Psychology*, 48, 127-147.
- Bennett, N. and Carré, C. (eds) (1993). *Learning to teach*. London: Routledge.
- Bloom, B.S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.
- Brandsma, H.P. (1993). *Basisschoolkenmerken en de kwaliteit van het onderwijs* (Characteristics of primary schools and the quality of education). Groningen: RION.
- Brandsma, H.P. and Knuver, J.W.M. (1989). Organisational differences between Dutch primary schools and their effect on pupil achievement. In D. Reynolds, B.P.M. Creemers and T. Peters (eds), *School Effectiveness and Improvement. Proceedings of the First International Congress, London* (pp. 199-212). Groningen/Cardiff: RION/University of Wales.
- Bruner, J.S. (1966). *Toward a theory of instruction*. New York: Norton.
- Carroll, J.B. (1963). A model of school learning. *Teachers College Record*, 64, 723-733.

- Carroll, J.B. (1989). The Carroll Model: a 25-year retrospective and prospective view. *Educational Researcher*, 18, 26-31.
- Creemers, B.P.M. (1992). School effectiveness and effective instruction: the need for a further relationship. In J. Bashi and Z. Sass (eds), *School Effectiveness and Improvement. Proceedings of the Third International Congress, Jerusalem* (pp. 105-132). Jerusalem: The Magnes Press.
- Croll, P. (1988). Teaching methods and time on task in junior classrooms. *Educational Research*, 30, 2, 90-97.
- Freiberg, H.J., Prokosch, N., Treister, E.S. and Stein, T. (1990). Turning around five at-risk elementary schools. *School Effectiveness and School Improvement*, 1, 1, 5-25.
- Gamoran, A. (1986). Instructional and institutional effects of ability grouping. *Sociology of Education*, 59, 185-198.
- Glaser, R. (1976). Components of a psychology of instruction: toward a science of design. *Review of Educational Research*, 46, 1-24.
- Haanen, M., Lagendaal, P., Roders, R. and Wolf, K. van der (1990). *Kleurrijk en effectief. Eindrapportage van de eerste fase van het Amsterdamse EGAA-project* (Colourful and effective. Final report of the first phase of the Amsterdam EGAA project). Delft: Eburon.
- Haertel, G.D., Walberg, H.J. and Weinstein, T. (1983). Psychological models of educational performance: a theoretical synthesis of constructs. *Review of Educational Research*, 53, 1, 75-91.
- Harnischfeger, A. and Wiley, D.E. (1976). The teaching learning process in elementary schools: a synoptic view. *Curriculum Inquiry*, 6, 5-43.
- Hofman, R.H. (1993). *Effectief schoolbestuur* (Effective schoolboards). Groningen: RION.
- Hofman, W.H.A. (1993). *Effectief onderwijs aan allochtone leerlingen: een empirische studie naar de invloed van school- en klasfactoren op de loopbaan van allochtone en autochtone leerlingen in het voortgezet onderwijs* (Effective education for immigrant students). Proefschrift Rotterdam. Delft: Eburon.
- Knuver, J.W.M. (1993). *De relatie tussen klas- en schoolkenmerken en het affectief functioneren van leerlingen* (The relationship between class and school characteristics and the affective functioning of pupils). Groningen: RION.
- Kuiper, W.A.J.M. (1993). *Curriculumvernieuwing en lespraktijk* (Curriculum reform and teaching practice) PhD Thesis, University of Twente, Enschede.
- McCormack-Larkin, M. (1985). Ingredients of a successful school effectiveness project. *Educational Leadership*, 42, 6, 31-37.
- Mortimore, P., Sammons, P., Stoll, L., Lewis, D. and Ecob, R. (1988). *School Matters: The Junior Years*. Wells: Open Books.
- Olson, J.K. and Eaton, S. (1987). Curriculum change and the classroom order. In J. Calderhead (ed.), *Exploring teachers' thinking* (pp. 179-194). London: Cassell.
- Osinga, N., Boersma, R. and Houtveen, Th. (1993). *School improvement in the Netherlands*. Leeuwarden: GCO.
- Palincsar, A.S. and Brown, A.L. (1989). Instruction for self-regulated reading. In L.B. Resnick and L.E. Klopfer (eds), *Toward the thinking curriculum: current cognitive research* (pp. 19-39). Alexandria, VA: Association for Supervision and Curriculum Development.
- Prawat, R.S. (1989). Promoting access to knowledge, strategy, and disposition in students: a research synthesis. *Review of Educational Research*, 59, 1-41.
- Reezigt, G.J. (1993). *Effecten van differentiatie op de basisschool* (Effects of grouping in primary education). Groningen: RION.
- Rosenshine, B. (1983). Teaching functions in instructional programs. *The Elementary School Journal*, 83, 4, 335-351.
- Rosenshine, B. and Stevens, R. (1986). Teaching Functions. In M.C. Wittrock (ed.), *Handbook of Research on Teaching, Third Edition* (pp. 376-391). New York: Macmillan.
- Scheerens, J. (1992). *Effective Schooling: Research, Theory and Practice*. London: Cassell.
- Slavin, R.E., Karweit, N.L. and Madden, N.A. (1989). *Effective programs for students at risk*. Boston: Allyn & Bacon.
- Slavin, R.E., Madden, N.A., Karweit, N.L., Livermon, B.J. and Dolan, L. (1990). Success for All: first-year outcomes of a comprehensive plan for reforming urban education. *American Educational Research Journal*, 27, 2, 255-278.
- Stoll, L. and Fink, D. (1992). Reorganization for effectiveness: the Halton approach. In J. Bashi and Z. Sass (eds), *School Effectiveness and Improvement. Proceedings of the Third International Congress, Jerusalem* (pp. 370-380). Jerusalem: The Magnes Press.

- Stringfield, S.C. and Slavin, R.E. (1992). A hierarchical longitudinal model for elementary school effects. In B.P.M. Creemers and G.J. Reezigt (eds), *Evaluation of Educational Effectiveness* (pp. 35-69). Groningen: ICO.
- Stringfield, S. and Teddlie, C. (1989). The first three phases of the Louisiana school effectiveness study. In B.P.M. Creemers, T. Peters and D. Reynolds (eds), *School Effectiveness and School Improvement. Proceedings of the Second International Congress, Rotterdam* (pp. 281-294). Lisse: Swets & Zeitlinger.
- Teddlie, C. and Stringfield, S. (1993). *Schools make a difference: lessons learned from a 10-year study of school effects*. New York: Teachers College Press.
- Walberg, H.J. (1986). Synthesis of research on teaching. In M.C. Wittrock (ed.), *Handbook of Research on Teaching, Third Edition* (pp. 214-229). New York: Macmillan.
- Wang, M.C., Haertel, G.D. and Walberg, H.J. (1990). What influences learning? A content analysis of review literature. *Journal of Educational Research*, 84, 1, 30-43.
- Waxman, H.C., Wang, M.C., Anderson, K.A. and Walberg, H.J. (1985). Synthesis of research on the effects of adaptive education. *Educational Leadership*, 43, 1, 27-29.
- Weide, M.G. (1993). Effective provisions for ethnic minorities in primary schools. *Tijdschrift voor Onderwijsresearch*, 18, 2, 74-84.
- Werf, M.P.C. van der, and Weide, M.G. (1991). Effectief onderwijs voor allochtone leerlingen (Effective instruction for immigrant students). *Tijdschrift voor Onderwijsresearch*, 16, 4, 231-243.

Manuscript submitted: October 13, 1993

Accepted for publication: January 6, 1994

Teaching Effectiveness: An overview of studies

Joost Lowyck

Centre for Instructional Psychology and Technology, Leuven University

ABSTRACT

In this contribution, different approaches to teaching effectiveness studies are described. In a first section, effectiveness interpreted by 'process-product' studies is reviewed. It is assumed that teaching directly influences learning output. Isolated effective teaching behaviors are described as well as models in which effective teaching behaviors are grouped. The second section is about mediating variables. In this conception, learning outcomes are not directly influenced by teaching behaviors, but by variables mediating between teaching and learning, like 'academic learning time' and pupil cognitions. Finally, in the third section, teacher cognitions are reviewed in terms of their possible contribution to teaching effectiveness. These descriptive studies reveal the idiosyncratic nature of teaching rather than powerful effectiveness criteria.

INTRODUCTION

Since almost a century scholars in the field of education use many and divergent theoretical frameworks, models, methods and criteria to define teaching effectiveness without, however, any univocal or definite answer. Their stubborn endeavours bear a likeness to the Greek myth of Sisyphus, a king of Corinth condemned forever to roll a stone up a hill in Hades only to have it roll down again on nearing the top. Indeed, research on teaching passed through several phases, dependent on dominant ideas about education and human behavior (see Anderson, 1989; Creemers, 1991; De Corte & Lowyck, 1983; Dunkin & Biddle, 1974; Gage, 1978; OECD, 1990; Lowyck & Pieters, 1993; Scheerens, 1991; Shulman, 1986). Though teaching effectiveness research came at its very height within the process-product approach (see Wittrock, 1986, p. 1008), some predecessors are still worthy of mention.

In the days of test psychology and personality studies effective teaching is perceived as a by-product of the teacher's personality. Not teaching behavior in itself but the teacher's personality is the paramount indicator for effectiveness. The research design is a correlational one, since a strong correlation between teachers' characteristics and estimated teaching quality is hypothesized. The outcomes of ability tests and questionnaires are compared with expert's and students' opinions on the teacher's quality. Literature reviews, however, indicate few valuable results; there are some reasons to account for this shortage (see Getzels & Jackson, 1963). First, the results tell little about the 'teaching' profession since qualities like being pupil-centered, friendly, enthusiastic or psychologically stable apply to all jobs where interaction with people is predominant. Second, subjective and idiosyncratic perceptions by school people, mostly reflect traditional and superficial indicators of quality. Third, broad personality traits do not fit in with elementary, discrete teaching behaviors, since the unit of analysis highly differs. Fourth, research reviews and meta-analyses emphasize grouped results leaving out more fine-grained characteristics gathered from individual studies.

In later studies more attention is paid to the refinement of both design and criteria. A clear example is the search for a democratic teachers' style as a reaction to the European fascist scene during the thirties and forties. In this context teachers' democratic attitudes or beliefs ('leadership style') and concomitant democratic classroom behaviors became the paramount object of study (see Lippitt & White, 1943). The criterion of effectiveness is not only operationalized in

*Address: Vesaliusstraat 2, B-3000, Leuven.

terms of opinions of school people, but in measurements of learning results as well. Here, the criterion-of-effectiveness is no more the mere perception of quality, but the learning output itself. Anderson (1959) analyzed two types of effect in 32 studies on democratic leadership styles: productivity in terms of learning gains and positive classroom climate. For learning outcomes 11 studies report gains with a democratic (integrative) style; 8 studies indicate effects with the authoritarian ('dominative') style and 13 reports are not conclusive at all. Regarding the classroom climate the results are more in favour of the democratic style though the criteria used were very divergent. Following Anderson (1959) the main conclusion is clear: "We were not fortunate enough to find that one style is consistently different from the other; thus we are forced to explore new avenues" (p. 212).

This lack of consistency in research outcomes has been interpreted in different ways (see Anderson, 1959; Dunkin & Biddle, 1974; Flanders, 1965). First, the methodology of many studies was weak or inadequate. Second, the concepts 'authoritarian' and 'democratic' held value judgments which in a subtle way influenced the design and interpretation of the studies. Third, the concepts were differently operationalized due to the lack of a precise description of classroom behaviors which account for one or another style. Fourth, a binary dimension (authoritarian-democratic) seems inadequate for a valid interpretation of classroom complexity, since the classroom is multidimensional.

PROCESS-PRODUCT STUDIES: TEACHER'S CLASSROOM BEHAVIORS

The lack of clear research outcomes together with the rise of behaviorism as the dominant psychological paradigm forced researchers towards the study of effective classroom 'behaviors'.

The basic question is which specific teaching behaviors (process variables) are effective so as to bring about learning gains in pupils (product variables). In order to make teaching observable, the complex teaching activity is fragmented into elementary or discrete teaching behaviors in line with the analytical paradigm as advocated by Gage (1963). Within this view, scientific problems can be solved by being analyzed into small problems, whose variables are less complex. So-called 'micro-criteria' are the central issue leaving out the search for over-all effectiveness of teachers in the many, varied facets of their roles.

The research method is a correlational one. A lot of time and energy is invested in category systems for the registration of process variables. Indeed, almost each researcher worked out an idiosyncratic observation system or modified already existing ones, which resulted both in the proliferation of observation systems (see Simon & Boyer, 1970) and in a shortage of conceptually consistent models and frameworks (Lowyck, 1980). Moreover, the product variables are weakly operationalized which brought Heath & Nielson (1974) to qualify the operational definitions of student achievement as 'shallow'.

In their meta-analysis of fifty well selected studies Rosenshine & Furst (1973) report nine 'powerful' process variables, like clarity, variability, enthusiasm, task orientation, criticism, indirectness, student opportunity to learn, structuring comments, and use of different levels of questions and cognitive activities. Notwithstanding possible positive interpretations, the problems with this approach are recognized by many scholars in the field of teaching research. First, the analytical and formal nature of the process variables dismisses two important types of variables, namely subject-matter content and pupil characteristics (Heath & Nielson, 1974). It seems obvious that one single set of teaching behaviors cannot account for learning outcomes in all subject-matter and for pupils from different socio-economic background or from all grades, the research being 'content-proof' and 'pupil-proof'. Second, almost all observational systems register behaviors in a traditional 'classroom teaching' context with emphasis on lecturing and question-answer cycles. Individual work or groupwork are not studied and, hence, the research design is 'method-proof'. Third, teaching is conceived of as a unidirectional activity of teachers

neglecting any influence of the situation on the teacher, as recognized by Doyle's mediating paradigm (Doyle, 1978). Since, this approach is 'situation-proof'. Fourth, because emphasis is exclusively on observable classroom behaviors, no information is provided on cognitive processes of teachers and pupils nor on the teachers' complex information processing during the planning, interactive and post-interactive phase of teaching (Jackson, 1968). Consequently, the approach is 'cognition-proof'.

As a solution to the analytical deficiency, researchers try to 'chunck' isolated teaching behaviors into more encompassing models or patterns, like 'active teaching' (Brophy, 1986) or 'adaptive teaching' (Good, 1979; Wang, 1980). In these models, several single teaching behaviors are grouped and a clear interdependency of discrete behavioral elements within one teaching model is hypothesized. Active teaching, for example, claims that students learn best in classes where they spend most of the time being taught or supervised by their teacher rather than working on their own or not working at all. The teacher presents information, develops concepts through lecture and demonstration, offers feedback, prepares the student for follow-up assignments, and monitors progress. There is a great deal of academic teacher talk rather than procedural or managerial, and much of it involves asking questions and giving feedback rather than extended lecturing (Brophy, 1986). This active teaching model seems very promising, mainly for groups of students with a need for intensive instructional support, like in the first grades of elementary education or in special education settings, for well-structured subject-matter (mathematics, reading and grammar) and in teacher-centered approaches (see Brophy, 1991; Carnine, 1989).

A possible disadvantage of this kind of model-driven teaching research is the fact that effective teaching models are defined by means of research reviews, without strong empirical validation of the composite behaviors, or as Gage (1978) states "From hundreds of correlations, especially the significant ones, the investigators and reviewers then synthesize the style or pattern of teaching that seems to be associated with desirable kinds of pupils achievement and attitudes". Further experimental studies with these composite behaviors indicated that not all effective teachers used the 'desirable' behaviors and that not all skills trained showed stable correlations with learning gains. Consequently, not all composites seemed necessary for effective teaching (Shulman, 1986).

Despite efforts to remediate the deficiencies mentioned above, the direct correlation of teacher behavior and pupil achievement remains problematic. It is hypothesized, then, that the search for direct correlations should at least be complemented by studying the 'intervening' or 'mediating' variables which determine classroom effectiveness (see Doyle, 1986).

MEDIATING VARIABLES: LEARNING TIME AND TEACHER/PUPIL COGNITIONS

Since the direct link between instructional input (teacher behavior) and output (learning outcomes) caused problems for defining effectiveness, emphasis is laid on the throughput, the variables mediating between teaching activities and learning outcomes. The learner himself becomes the decisive factor in the definition of effectiveness "What a teacher does at any one moment while working in a circumscribed content area affects a student primarily at only that particular moment and in that particular content area. The link between teacher behavior and student achievement is, therefore, the ongoing student behavior in the classroom learning situation" (Berliner, 1979, p. 123).

Both studies on 'Aptitude-Treatment-Interaction' (Snow, 1986) and the 'Beginning Teacher Evaluation Study' (Fisher, Berliner, et al., 1978) revealed the importance of individual differences and their concomitant cognitive and affective processes for active learning. This especially means that information delivered in classrooms results in differential effects in pupils due to the complexity of their processing characteristics and motivation. This increased attention to individual differences was compatible with the focus on individualization in the sixties and

seventies. The democratization of education with its increasing number of students in secondary and higher education, brought about the need for differentiated learning routes. Moreover, it was expected that individualization could enhance both creativity and competence in 'basic' subject-matter.

In a first approach 'time-on-task' is considered to be the decisive predictive effectiveness variable, in line with Carroll's model (1963) and the mastery-learning strategy (Bloom, 1974). It is hypothesized that powerful indicators for teaching effectiveness are found in observable and measurable actual pupil behaviors. If researchers could find a 'direct' criterion of effectiveness, they could define effectiveness 'on-the-spot' without having to wait for outcomes from achievement tests at the end of the year. In this view, pupil competence is defined as the amount of learning time necessary for task fulfilment. Brophy and Good (1986), in their review of effects of teaching on pupil achievement, report time-on-task as the most general positive finding from research on teaching: "At least two common themes cut across the findings, despite the need for limitations and qualifications. One is that academic learning is influenced by the amount of time that students spend engaged in appropriate learning tasks..." (p. 366). However, the amount of time is not a definite measure for effectiveness, since learning processes activated during the allocated time strongly determine the learning outcomes.

The cognitive-mediational perspective supports conceptions of learning and instruction that are associated with meaningful and 'higher-order' student learning, teacher efficacy, and equity of educational opportunities. Teaching is interpreted by the pupils and their interpretation influences to a high degree the effects of any teaching activity. Here, the most important findings from cognitive research enter the field of teaching effectiveness. The criterion of effectiveness is the cognitive quality of instruction. Anderson (1989) describes the criterion as follows "Learning occurs when learners actively transform incoming information and construct meaning in terms of their prior knowledge. Teachers and others influence learning not through mere transmittal of information but through their influence on students' cognitive processing of that information" (p. 101). In this phase of research, most emphasis is laid on the direct translation of findings from psychological literature, like stimulation of metacognition and self-regulation, schema activation, linking prior knowledge with new information by means of advance organizers, supporting cognitive processes through scaffolding, modelling, cognitive apprenticeship (see Brown, Collins, & Duguid, 1989).

Concerning the social mediation, attention is paid to the analysis of meaning and purposes attributed by the participants to social settings. One example of this approach is the 'hidden curriculum' of schools and classrooms. It is not the classroom context in itself, but the complex attribution of meaning by the actors which is of paramount importance. Anderson (cit. in Shulman, 1986) analyzed ways in which primary-grade children cope with seatwork. "Whereas the process-product researcher would have correlated frequency of seatwork with student achievement, and the ALT scholar would have observed the proportion of seatwork time during which the student was apparently engaged along with the degree of difficulty of the tasks, Anderson proceeds to record student comments to themselves and to classmates during assignments and to interview the students" (p. 16). Her focus is on what students are thinking and feeling as they work on their tasks, and on what these phenomena reveal about the mediation of instruction by the students.

In line with the social mediation paradigm, emphasis is laid on support activities of the peer group. Examples are: reciprocal teaching (Palincsar & Brown, 1984), collaborative knowledge construction in the classroom (Scardamalia & Bereiter, 1992), and cooperative learning (Slavin, 1990). Common effectiveness criteria are explicit goals, individual responsibility, equal opportunities for peers, clear procedure and task relevance. Effectivity studies reveal interesting and lasting outcomes of the original training programs (see Palincsar & Brown, 1984). This effect, however, fades the more the experimental training procedure is implemented under usual classroom conditions (see Palincsar, Brown & Martin, 1987).

Interesting though this line of research may seem, some problematic issues remain as to the definition of effectiveness criteria. In most cases, a direct translation of descriptive findings into

prescriptions for instruction take place without, however, a strong control of the transition steps necessary to validate the effectiveness criteria. As Shulman (1986) remarks "We get too little sense of how variations among tasks, or among forms of teaching, relate to differences in how students mediate those instructional presentations. As with the paradigms discussed earlier, far too little attention is devoted to differences in the content of subject matter being taught. ... We thus learn important new things about teaching from this research, but also forgo parts of the portrayal available from the work in other research programmes" (p. 16).

A positive aspect of this recent line of effectiveness research, however, is the combination of the process-product paradigm with the cognitive and social approach on teaching and learning. It seems a challenging endeavour indeed to search for the most essential effectiveness variables within the complex teaching and learning situation, making use of a parsimonious model.

TEACHER COGNITIONS

The shift in the psychological scene from behaviorism to cognitivism, along with a refinement of qualitative research methods during the seventies paved the way for studies on teachers' cognitions. Teaching is no more perceived as a quality of a teachers' personality nor as a set of observable behaviors, but as a profession with a call for higher-order cognitive skills in order to cope with the complexity of an educational environment. It is a job resembling the one of physicians, engineers and architects with emphasis on judgment, problem-solving and decision-making. Although this more qualitative line of research started in the sixties (see Smith & Geoffrey, 1969; Jackson, 1968) it was first acknowledged as a research field by the National Institute for Education in its publication: "Teaching as Clinical Information Processing" (1975). In spite of the fact that most studies of teacher's thinking are explorative and descriptive in nature, they nevertheless offer some indications as to effective teaching. At the beginning, most attention is paid to the isolated description of the 'naturalistic' cognitive processes of teachers during the planning, interactive and post-interactive phase of teaching, defined in terms of information processing. The methods used are protocol-analysis, retrospection, stimulated recall, thinking aloud, rep-grid, logbook keeping, etc. Because these methods focus almost exclusively on isolated cognitive processes, no links between cognition and action are studied. Later on, the interaction between the phases becomes object of study, like the interaction between planning and classroom interaction, or the relationship between planning, interaction and post-interactive reflection. One of the most common observations is the psychological way in which teachers function which seems to counter a mere logical teaching model (Lowyck, 1980).

Whereas the description of cognitive processes tends to be content-free, the study of teacher's subjective or implicit theories (teacher beliefs, teacher's practical knowledge) throws light on the very content of their cognitions. It is shown that teachers do not possess a complete nor clear model of teaching, but an eclectic mix of rules of the thumb, generalizations, opinions, values and expectations. An important issue is how the teacher's cognitive understanding of subject-matter content can be elucidated. Shulman (1986) suggests to avoid any unnecessary isolation of types of knowledge and to focus on the intrinsic interaction between subject-matter knowledge, pedagogical knowledge and curricular knowledge.

One of the main problems with this kind of research is the lack of conceptual refinement (see Pajares, 1992) and the closed-circuit characteristics of the research object, namely its reduction to cognitive processes as internal representations without any relationship to teaching behaviors. Moreover, shortage of validity and methodological strength in many idiosyncratic research endeavours caused a lot of free-wheeling descriptions, lacking model building and hypothesis testing. The danger of an over-estimation of idiosyncratic narratives of teachers lays in its disconnection from any search for professional expertise or teaching effectiveness, the latter labeled as a 'positivist assumption' (see Elbaz, 1991, p. 5).

It is clear that insight in the subjective theories of teachers can elucidate their knowledge-base about teaching which allows clarification and classification of subjective teaching models.

However, this line of research is not clear at all about the effectiveness criteria of teaching used. Even if one tries to deduce criteria of effectiveness from expert's cognition and behavior, it is an enormously difficult task due to the idiosyncrasy of the expert's functioning, the vagueness of criteria that should warrant the experienced teacher under study is an expert, the influence of the complex context or task and the difficulty for experts to fully report their cognitive processes or practices.

The integration of teacher cognitions in process-product research was advocated by Berliner (1989), but empirical research in this tradition is scarce. Clark & Peterson (1986), for example, mention only three studies that bear on the relationship between teacher's decision making in classrooms and student achievement. And, the solution proposed by Peterson (1988) to include teacher's cognitional knowledge as a variable in effective teaching did not lead towards solid empirical investigations neither. It seems time, now, to reconsider the paradigms of research on teaching and to "invite scholars in the field to reflect carefully upon the possible effects of centrifugal tendencies in research on teaching so as to look for more integration and cooperation" (Lowyck & Clark, 1989). We do not need a dissension of the research domain on teaching, but as Berliner (1989) contends "There is no doubt that continued descriptive and taxonomic work to better understand the teacher's interactive thoughts and decisions is highly desirable...But I believe that it will be the establishment of relations between interactive thoughts and decision making on the one hand, and criteria of effectiveness on the other, that will lead to the greatest growth of knowledge in this field" (p. 12).

In short, the whole domain of teaching research suffers from a lack of a well functioning apparatus of concepts, models and theories. "The vagueness of the conceptual framework, the babel of unclear, idiosyncratic definitions, the proliferation of models, the isolation of paradigms, and the suggested incompatibility of qualitative and quantitative methodology all hinder the consolidation of thousands of research outcomes. How can we further the understanding of teaching in its complexity, if the concepts of teaching are apparently more complex than the teaching reality itself" (Lowyck, 1990, p. 92-93).

DISCUSSION

As has been illustrated in the previous sections, research on teaching is strongly influenced by the "Zeitgeist", the state-of-the-art in research, its available methodology and the expectations towards or the frustrations with research outcomes.

First, societal needs, like democratization of citizens ('authoritarian-democratic' paradigm), the competitive edge of society (back-to-the-basics), the adaptation of schools to new target groups (eg. special education), and the fit between school and industry (cognitive flexibility, transfer of knowledge) all are in a subtle way filtered through education policy and plugged into research endeavours. Needless to say that not only social aspects, but also economical-financial factors influence the way education and schooling are perceived.

Second, evolutions within research itself determine to a high degree the focus, expectations and frustrations with regard to paradigms, theories and models. Examples are the successive shift from 'trait psychology' toward behaviorism and cognitivism. These shifts are generated both by disappointing outcomes of the research paradigm at hand and the attractiveness of new avenues.

Third, methodological refinement and availability of new research techniques in different domains, like psychology, sociology or antropology all challenge researchers to find new technical solutions. Examples are ethnography, symbolic interactionism, verbal reports, protocol analysis, rep-grids, as well as multi-level analysis, etc.

Fourth, the conceptualization of teaching dominates the choice of research design. In the case teaching is perceived as the behavioral output of the teacher's personality, emphasis is laid on the study of teachers' characteristics. If teaching is conceptualized as the sum of overt but discrete teaching skills, process-product studies seem appropriate. When teaching is seen as the idiosyncratic model of any individual teacher, teacher cognitions are focused at.

The preceding examples of influences, however, cannot explain all variability in approaches of teaching research. Often more psychological factors determine the research at hand. Examples are the dominant research groups, or the brilliant ideas of individual researchers, like Carroll, Dewey, Gage, Jackson, Vygotsky, who unexpectedly launch influencing research contrary to the dominant line of thought.

Consequently, any review of teaching effectiveness struggles with conceptual, methodological and operational difficulties and affects opinions about the possibility to build synergy in this particular field of research. Some researchers emphasize incompatibility between research paradigms, epistemological backgrounds and suitable methodology. Examples are the opposition of qualitative and quantitative methodology (whatever this means), or the study of teaching as an idiosyncratic attribution of meaning by teachers versus the so-called positivist assumption with a search for more generalizable outcomes. Others claim openness and cooperation between researchers in order to broaden the scope as well as to challenge the integration of originally controversial paradigms. Examples are the recent studies on cognitive apprenticeship, who try to combine a process-product approach with recent insights in cognitive psychology, or the mediating paradigm which inserts new but important variables between teaching and learning outcomes.

This contribution reviewed some lines of thought and empirical work which could contribute to a more synergic approach of the study of teaching effectiveness. It has been revealed that many factors influence both the conceptualization and the design of effectiveness studies. Moreover, the knowledge accumulated through reflection and empirical research during more than half a century, enables at present an incremental view on teaching effectiveness.

REFERENCES

- Anderson, R.C. (1959). Learning in discussions: a resume of the authoritarian-democratic studies. *Harvard Educational Review*, 29, 20-215.
- Anderson, L.M. (1989). Classroom instruction. In M.C. Reynolds (Ed.). *Knowledge base for the beginning teacher* (pp. 11-115). Oxford/New York: Pergamon.
- Berliner, D.C. (1979). Tempus educare. In P.L. Peterson & H.J. Walberg (Eds). *Research on Teaching*. Berkeley: McCutchan.
- Berliner, D.C. (1989). The place of process-product research in developing the agenda for research on teacher thinking. In J. Lowyck & C.M. Clark (Eds). *Teacher Thinking and Professional Action* (pp. 3-21). Leuven: Leuven University Press.
- Bloom, B.S. (1974). Time and learning. *American Psychologist*, 29, 682-688.
- Brophy, J. (1986). Teacher influences on student achievement. *American Psychologist*, 41, 10, 1069-1077.
- Brophy, J. (1991). Effective schooling and disadvantaged students. In M.S. Knapp & P.M. Shields (Eds). *Better schooling for children of poverty: Alternatives to conventional wisdom* (pp. 211-234). Berkeley: McCutchan.
- Brophy, J. & Good, T.L. (1986). Teacher behavior and student achievement. In M. Wittrock (Ed.). *Handbook of Research on Teaching, Third Edition* (pp. 328-375). New York: Macmillan.
- Brown, J.S., Collins, A. & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18, 1, 32-42.
- Carnine, D. (1989). Direct instructional model: special education. In T. Husèn & T.N. Postlethwaite (Eds). *International Encyclopedia of Education, Research and Studies. Suppl. Vol. 1* (pp. 251-257). Oxford: Pergamon.
- Carroll, J.B. (1963). A model of school learning. *Teachers College Record*, 64, 724-733.
- Clark, C.M. & Peterson, P.L. (1986). Teachers' thought processes. In M.C. Wittrock (Ed.). *Handbook of Research on Teaching, Third Edition* (pp. 255-296). New York: Macmillan.
- Creemers, B.P.M. (1991). *Effectieve instructie: een empirische bijdrage aan de verbetering van het onderwijs in de klas*. 's Gravenhage: Instituut voor het Onderzoek van het Onderwijs.
- De Corte, E. & Lowyck, J. (1983). Heroriëntatie in het Onderzoek van het Onderwijzen. *Tijdschrift voor Onderwijsresearch*, 8, 242-260.
- Doyle, W. (1978). Paradigms for research on teacher effectiveness. In L.S. Shulman (Ed.). *Review of Research on Teaching*. Itasca: Peacock.

- Doyle, W. (1986). Classroom organization and management. In M.C. Wittrock (Ed.). *Handbook of Research on Teaching, Third Edition* (pp. 392-431). New York: Macmillan.
- Dunkin M.J. & Biddle, B.J. (1974). *The Study of Teaching*. New York: Holt, Rinehart & Winston.
- Elbaz, F. (1991). Research on teacher's knowledge: the evolution of a discourse. *Journal of Curriculum Studies*, 23, 1, 1-19.
- Fisher, C.W., Berliner, D.C., et al. (1978). *Teaching and learning in the elementary school: a summary of the Beginning Teacher Evaluation Study*. San Francisco: Far West Laboratory for Educational Research and Development.
- Flanders, N.A. (1965). *Teachers influence, pupil attitudes and achievement*. Washington: U.S. Department of Health, Education and Welfare, Office of Education.
- Flanders, N.A. (1970). *Analyzing teaching behavior*. Reading, Mass: Addison-Wesley.
- Gage, N.L. (1963). Paradigms for research on teaching. In N.L. Gage (Ed.). *Handbook of Research on Teaching* (pp. 94-141). Chicago: Rand McNally.
- Gage, N.L. (1978). *The scientific basis of the art of teaching*. New York: Teachers College Press.
- Getzels, J.W. & Jackson, P.W. (1963). The teacher's personality and characteristics. In N.L. Gage (Ed.). *Handbook of Research on Teaching*. Chicago: Rand McNally.
- Good, T.L. (1979). Teacher effectiveness in elementary school: What we know about it now. *Journal of Teacher Education*, 30, 52-64.
- Heath, R.W. & Nielson, M.A. (1974). The research basis for performance-based teacher education. *Review of Educational Research*, 44, 4, 463-484.
- Jackson, P.W. (1968). *Life in classrooms*. New York: Holt, Rinehart & Winston.
- Lippitt, R. & White, K. (1943). The 'social climate' of children's groups. In R. Barker, J.S. Kounin & H.F. Wright (Eds). *Child behavior and development*. New York: McGraw Hill.
- Lowyck, J. (1980). *A process analysis of teaching (report n° 21): EDRS-ED 190513*. Leuven: K.U. Leuven, Afdeling Didactiek en Psychopedagogiek.
- Lowyck, J. (1990). Teacher thinking studies: bridges between description, prescription and application. In C. Day, M. Pope & P. Denicolo (Eds). *Insights into teachers' thinking and practice* (pp.85-103). London: The Falmer Press.
- Lowyck, J. & Clark, C.M. (1989). Preface. In J. Lowyck & C.M. Clark (Eds). *Teacher thinking and professional action* (pp. 1-III). Leuven: Leuven University Press.
- Lowyck, J. & C.M. Clark (Eds). (1989). *Teacher thinking and professional action*. Leuven: Leuven University Press.
- Lowyck, J. & Pieters, J.M. (1993). *The quality of teaching. A research programming study on the quality and the functioning of teachers conducted for the Institute of Educational Research in the Netherlands*. Den Haag: SVO.
- National Institute for Education (1975). *Teaching as clinical information processing*. Washington: U.S. Department of Health, Education and Welfare.
- OECD (1990). *The teacher today. Tasks, conditions, policies*. Paris: Organisation for Economic Co-operation and Development.
- Pajares, M.F. (1992). Teachers' beliefs and educational research: cleaning up a messy construct. *Review of Educational Research*, 62, 307-332.
- Palincsar, A.S. & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction*, 1, 117-175.
- Palincsar, A.S., Brown, A.L. & Martin, S.M. (1987). Peer interaction in reading comprehension instruction. *Educational Psychologist*, 22, 231-253.
- Peterson, P.L. (1988). Teachers' and students' cognitional knowledge for classroom teaching and learning. *Educational Researcher*, 17, 5, 5-14.
- Rosenshine, B. & Furst, N. (1973). The use of direct observation to study teaching. In R.M. Travers (Ed.). *Second Handbook of Research on Teaching* (pp. 122-183). Chicago: Rand McNally.
- Scardamalia, M. & Bereiter, C. (1992). An architecture for collaborative knowledge building. In E. De Corte, M.C. Linn, H. Mandl & L. Verschaffel (Eds). *Computer-based learning environments and problem solving* (NATO/ASI Series F: Computer and System Sciences, Vol. 84) (pp. 41-66). Berlin: Springer.
- Scheerens, J. (1991). Schoolorganisatiekenmerken en condities voor effectieve instructie. In S. Dijkstra, H.P.M. Krammer & J.M. Pieters (Eds). *De onderwijskundig ontwerper* (pp. 19-35). Amsterdam/Lisse: Swets & Zeitlinger.
- Shulman, L.S. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In M.C. Wittrock (Ed.). *Handbook of Research on Teaching, Third Edition* (pp. 3-36). New York: Macmillan.

- Simon, A. & Boyer, E.G. (Eds) (1970). *Mirrors for behavior: an anthology of classroom observation instruments*. Philadelphia: Research for Better Schools.
- Slavin, R.E. (1990). *Cooperative learning: Theory, research and practice*. Englewood Cliffs, NJ: Prentice Hall.
- Smith, L.M. & Geoffrey, W. (1968). *The complexities of an urban classroom. An analysis toward a general theory of teaching*. New York: Holt, Rinehart & Winston.
- Snow, R.E. (1986). Individual differences and the design of educational programs. *American Psychologist*, 41, 10, 1029-1039.
- Wang, M.C. (1980). Adaptive instruction: building on diversity. *Theory into Practice*, 19, 122-128.
- Wittrock, M.C. (Ed.) (1986). *Handbook of Research on Teaching, Third Edition*. New York: Macmillan.

Manuscript submitted: October 13, 1993

Accepted for publication: December 22, 1993

The School-level Context of Instructional Effectiveness: A Comparison between School Effectiveness and Restructuring Models

Jaap Scheerens

*University of Twente**¹

ABSTRACT

In this article comprehensive multi-level educational effectiveness models are compared to constructivist approaches to learning and instruction within the framework of restructuring. The implications for the macro- and meso-level context of effective instruction are examined. The conclusion is that, although crucial comparative experiments between these two complex frameworks on the functioning of educational systems are feasible in theory, they will be very hard to realize in actual practice. Neither the pedagogical vision of constructivism nor the administrative philosophies on restructuring seem to add much to what is already known about factors that enhance educational effectiveness. Yet, the constructivist perspective may stimulate the expansion of educational effectiveness models in considering new goals and new methods.

INTRODUCTION

In current educational thinking there are two important developments in which organizational and contextual conditions of instruction and learning are considered. The first concerns the construction and empirical verification of "comprehensive" school-effectiveness models (Scheerens & Creemers, 1989), the second is the broad concept of "restructuring" (see for example Murphy, 1992). Although these two developments have important differences – for instance the more descriptive and empirical nature of the effectiveness models, vs the more prescriptive and ideological nature of the restructuring movement – it seems to be interesting to compare them more systematically. The objectives of this comparison are threefold:

- to examine the consequences of different instructional models (direct instruction vs constructivist perspectives on learning and instruction) for organizational and contextual arrangements;
- to see to what extent the two developments can be integrated and to identify areas where they should rather be contrasted, for instance by means of comparative experiments;
- to draw some tentative evaluative conclusions with respect to the fruitfulness of both developments for the furthering of our knowledge on "what works" in education.

1. COMPREHENSIVE EDUCATIONAL EFFECTIVENESS MODELS

Comprehensive educational effectiveness models are causal models of educational attainment that contain explanatory variables at the level of the individual student, the level of classroom teaching (micro-level), the level of school organization (meso-level) and the level of the interface between school and environment (macro-level). Examples of these models are those developed by De Corte & Lowyck, 1983; Reynolds & Walberg, 1990; Scheerens, 1990; Levine & Lezotte, 1990; Creemers, 1991 and Stringfield & Slavin, 1992).

* Address: University of Twente, Faculty of Educational Science and Technology, P.O. Box 217, 7500 AE Enschede.

The general characteristics of these models are as follows:

1. variables are categorized according to an input-process-outcome and context structure;
2. the models have a multi-level structure (see figure 1), usually including a pupil, classroom and school level and sometimes also a school-environment level;
3. they recognize causal chains, i.e. intermediate causal variables that reflect the influence of certain exogenous conditions while at the same time they are seen as antecedents to certain other variables;
4. in some cases the models also include non-recursive relationships (feedback loops) implying self-regulating causal mechanisms.

The most efficient strategy for building multi-level school effectiveness models appears to be an "inside-out" approach, that is to first model micro-processes that are closest to the actual learning processes of individual students and then consider facilitating conditions one level up. This would mean that representations of learning and decision-processes of individual students should be at the core of educational effectiveness models.

Key-variables within the framework of comprehensive school effectiveness models are:

- * At the student level *achievement scores* or *attainment indexes* [the outcome variables] need to be adjusted for background variables like *intelligence*, *cultural capital of the home* and *peer group characteristics*.
- * *Student motivation to learn* should have a place within the model as a factor that is influenced by instructional variables, while at the same time being an antecedent condition of achievement *and* a factor that is influenced by feedback on achievement.
- * The third dimension is located at classroom level and a joint index of educational time and content covered; this dimension could be called *intensity and relevance of educational exposure*.
- * The fourth dimension represents the technical quality of education. We shall refer to this dimension as *degree of structure and adaptiveness of instruction*. This dimension reflects the fact that behaviouristic instructional models, such as "mastery learning" and "direct instruction", are taken as the core of comprehensive educational effectiveness models.
- * The fifth dimension represents *teacher quality* and could be thought of as a composite rating scale used by school leaders or inspectors.
- * At school level the sixth dimension represents *a shared achievement-oriented mission* among staff and school-leadership.
- * *Coordination among sub-units* represents a more structural factor thought to be conducive of consistency in effective classroom practices represented by several of the other dimensions.
- * *Evaluation practices* (at all levels of school functioning) represents an eighth dimension which is of vital importance for the image of a school as a learning organization.
- * *Quality of allocations of pupils and teachers* is a ninth dimension, which assumes that optimal combinations of teachers and groups of pupils exist.
- * *Achievement incentives from the school environment* represents the tenth dimension (i.e. overt attainment targets set by higher administrative levels).

It should be noted that several dimensions contribute at more than one level of school functioning. For instance, "learning time" and "content covered" are shaped both at the level of the school curriculum and in the actual process of teaching and classroom management. Evaluation practices take place at school level (as for instance in so-called school-based review) and at classroom level. At the same time individual dimensions may have several "bearers". "Structure and adaptiveness", for instance, is both a characteristic of instructional media and of teacher behaviour; whilst "achievement orientation" depends on school directors and their staff. A third point worth noting with respect to this multi-level framework of variables, is that, preferably, it should be studied longitudinally, to measure the stability of its overall effectiveness and to isolate effectiveness-enhancing conditions.

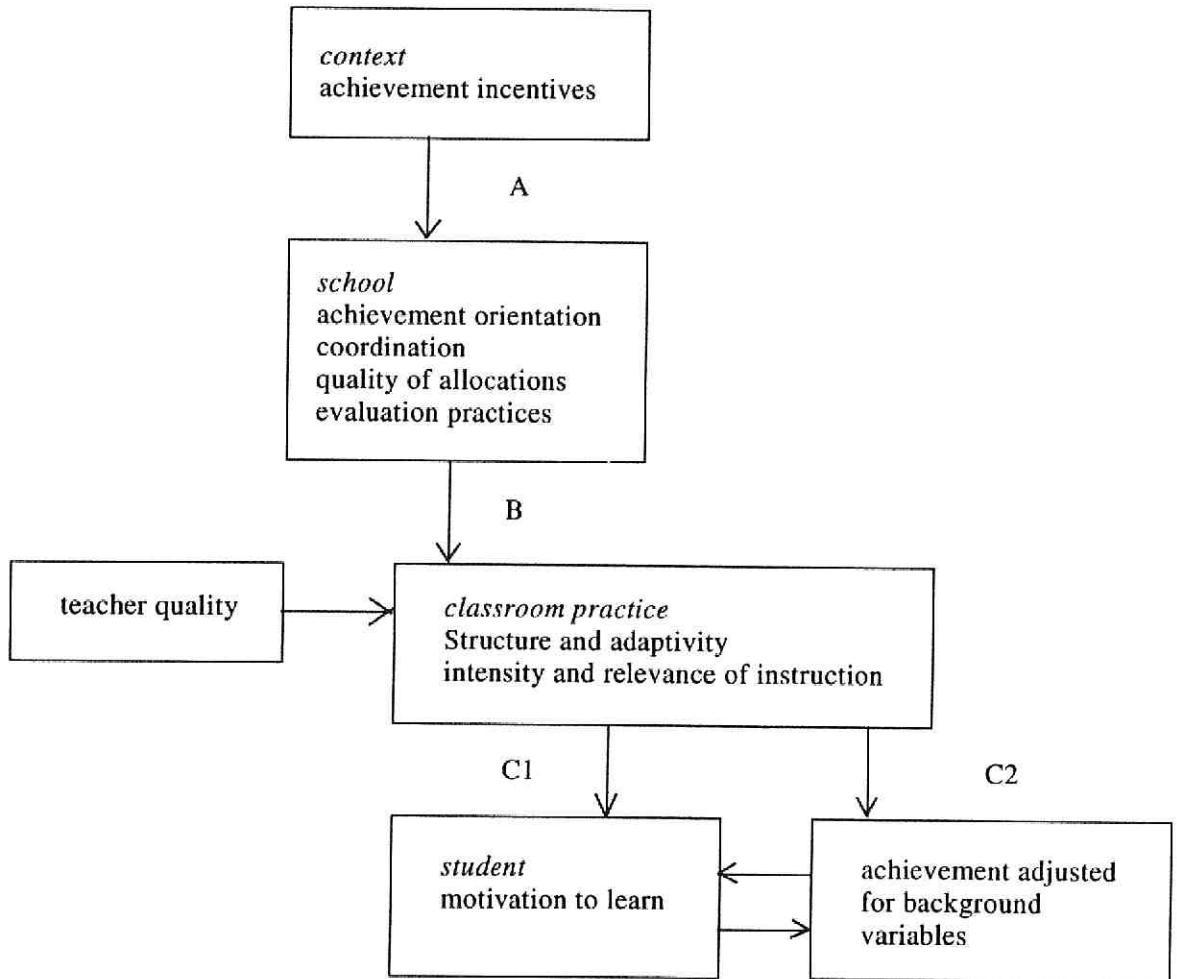


Fig. 1. Dimensions of effective schooling within a multi-level framework. Approximately the same overall framework is used as in Creemers, 1991, p.24. Likely extensions of the model could be the inclusion of a factor *teacher motivation* and a division between “classroom practice and intensity” and “relevance of instruction” to separate cause and effect as in Creemers’ model.

The three relationships of focal interest from the perspective of the nested layers image of schooling are:

- the environmental stimulation of schools to encourage them to operate effectively;
- the facilitation of effective instruction by organizational arrangements (both cultural and structural);
- the guidance of learning processes by means of instruction.

More details on these models are given by Scheerens (1992) and Bosker & Scheerens (1994), particularly with respect to alternative interpretations of cross-level facilitation.

2. AN ALTERNATIVE ORIENTATION: CONSTRUCTIVISM AND RESTRUCTURING

2.1. Constructivism

Constructivism is an offshoot of what has been called the cognitive revolution in learning theory. Not only is it opposed to behaviourism but it is also set apart from objectivist instructional applications of cognitivism, such as intelligent tutoring, as these take the expert’s knowledge structure as an external entity that is to be mapped on the students. Instead, constructivism

Table 1. Objectivism vs. Constructivism, adapted from Jonassen, 1991, p. 9.

	Objectivism	Constructivism
<i>reality</i>	External to the knower	Determined by the knower
<i>mind</i>	Abstract machine for manipulating symbols	Conceptual system for constructing reality
<i>thought</i>	Atomistic: decomposable	Gestalt properties
	Manipulates abstract symbols	Imaginative: enables abstract thought
<i>meaning</i>	Corresponds to entities and categories in the world	Does not rely on correspondence to the world (determined by understanding)
<i>symbols</i>	Represent reality	Tools for constructing reality

emphasizes the active role of the learner in constructing knowledge. The real enemy of constructivism is objectivism.

Some of the opposing claims of both positions as stated by Jonassen (1991, p. 9) are summarized in Table 1.

Constructivism claims that reality is more in the mind of the knower, but does not go as far as denying external reality altogether (solipsism), however some radical constructivists come very close to a position of complete denial. The image of *student learning* that goes with constructivism underlines the active role of the learner. Students are to be confronted with "contextual" real-world environments, or "rich" artificial environments simulated by means of interactive media. Learning is described as self-regulated with lots of opportunity for discovery and students' own interpretation of events.

Learning strategies, learning to learn and reflecting on these learning strategies (meta-cognition) are as important as mastering content. Different ways in finding a solution are as important as the actual solution itself. Terms like "active learning" (Cohen, 1988), "situated cognition" (Resnick, 1987) and "cognitive apprenticeship" (Collins et al, 1988) are used to describe student learning.

The other side of the constructivist coin are approaches to *teaching* and *instructional technology* that enable students "to construct their own meaningful and conceptually functional representations of the external world" (Duffy and Jonassen, 1992, p. 11). The teacher becomes more of a coach, who assists students in "criss-crossing the landscape of contexts", looking at the concept from a different point of view each time the context is revisited (Spiro et al., 1992, p. 8). Cohen (1988) uses the term "adventurous teaching" for this approach.

There is less emphasis on structuring goals, learning tasks and plans in advance; goals are supposed to emerge when situated learning takes place and plans are not so much supposed to be submitted to the learner as constructed in response to situational demands and opportunities.

Learning situations must be such that students are invited to engage in sustained exploration (real-life contexts, or simulated environments). Some authors writing from this perspective state that "transfer" is the most distinguishing feature (Tobias, 1991), whereas others mention argument, discussion and debate to arrive at "socially constructed meaning" (see Cunningham, 1991).

Table 2. Comparison of traditional and constructivistic instructional models.

Traditional instruction	Instruction inspired by constructivism
emphasis on basic skills	bias towards higher order skills
subject matter orientation	emphasis on learning process
structured approach	discovery-learning
. pre-specified objectives	“rich” learning environment
. small steps	
. frequent questioning/feedback	. intrinsic motivation
. reinforcement through high % of mastery	. challenging problems
abstract-generalizable knowledge	situation-specific knowledge, learning from cases
standardized achievement tests	assessment less circumscribed alternative procedures

The role of assessment and the evaluation of students' progress is hotly debated. Radical constructivists take the position that performance on an actual learning task is the only legitimate way of assessment, since distinct “external” evaluation procedures could not do justice to the specific meaning of a particular learning experience for the student.

Others (e.g. Jonassen, 1992) come to the conclusion that assessment procedures from a constructivist perspective should merely be different: goal-free, rather than fixed on particular objectives, formative rather than summative, and oriented to assessing learning processes rather than mastery of subject matter. Appraisals of samples of products, portfolios and panels of reviewers that examine authentic tasks are also mentioned as acceptable procedures.

In Table 2 some of the major distinguishing features of learning and instruction according to the constructivist position are contrasted with characteristics of more traditional instructional models like direct instruction and mastery learning.

Bipolar comparisons such as the one in Table 2 run the risk of over-simplification and polarization whilst also constructing “straw men”. It should be emphasized that less extreme constructivist views can be very well reconciled with more “objectivist” approaches (cf. Merrill, 1991). Also, more eclectic approaches are feasible, as can be seen when more teacher-controlled and learner-controlled instructional situations are used alternately (cf. Boekaerts & Simons, 1993).

2.2. Meso- and macro-aspects of restructuring

“Restructuring” has become the umbrella concept encompassing multi-dimensional educational reform initiatives in the United States. Conley (1991, p. 1) describes restructuring as “complex, multi-dimensional, and at times contradictory”. A useful division is one running according to aggregation level to attempt to order the many things that go under the label of restructuring.

The *micro-level component* (i.e. the level of teaching and learning at classroom level) can – to a certain extent – be characterized as inspired by the constructivist approach to learning and instruction discussed in the previous section.

Yet, at this level there are also initiatives that correspond more with behaviourist approaches like direct teaching, including monitoring activities to provide feedback and reinforcement.

In this section, both meso-level and the macro-level aspects will be examined. The *meso-level* is the level of school organization and management, and the *macro-level* is taken as the school environment or the interface of the school with its larger environment, extending to the school district, the state and national level.

When examining all that goes under the label of restructuring a nice consistency can be discerned as far as managerial roles are concerned. At classroom level we have already been introduced to the teacher as a *coach* of the semi-independent accomplishments of active learners. At the school-level we meet a school leader that is described as a *facilitator* (Murphy, 1992; Goldman, Dunlap & Conley, 1993) of "empowered" teachers, while the school-site managers in their turn are stimulated by superintendents, or – more generally – officials at the administrative level above the school – described as "*enablers*" (Murphy, 1992, p. 97). The consistency in this image being that the independence and autonomy of the basic processes of learning, teaching and school leadership are fully recognized at each level. An implication of this image appears to be that a lot is left to the discretion of leaders at each level and that expectations of control being exercised at a higher level are modest.

As in the case of the comprehensive educational effectiveness models described in section 1, characteristics of ideal-type micro-level processes are mirrored at the meso- and macro-level.

At the meso-level (school management and organization) it is recognized that active learning might require a reorganization of the school's daily schedule, because, for instance, longer periods would be necessary to expose students to problem-solving in a real-life setting or to simulated "rich" learning environments. Allocation processes (allocating teachers to groups of students) may also have to be redesigned as small group work and ungraded classrooms are propagated. This may be specified in order to arrive at more flexible arrangements or to improve efficiency and enhancing equity by allocating strong teachers to weaker groups of students (Monk, 1992).

Changing the basic organizational structure of the school's primary process in the sense of altered schedules and different allocation processes of teachers to groups of students requires more cooperation and coordination between teachers than is the case when these are arranged in the traditional way. But this is only one of the reasons why school restructuring puts a premium on cooperation between teachers. The other reasons being the active engagement of teachers in the policy-making process at the school level (participatory decision-making) and the envisaged new professional role of the teachers referred to as professionalism or the teacher as a "reflective practitioner" (Argyris & Schön, 1974). Mutual consultation is a key concept in this view of teacher-functioning.

Moreover, a shared sense of commitment to educational reform is expected and implied in school-restructuring, thus highlighting the importance of collaboration.

Again, a parallel can be drawn between these idealized teacher characteristics at the meso-level and the constructivist view of learning and instruction at the micro-level. Like students who are encouraged to reflect on the strategies they use in problem-solving (meta-cognition, learning to learn), teachers, as reflective practitioners, are expected to do the same with respect to their work. When thinking of the organizational setting in which reflective professionalism is encouraged the image of the "learning organization" (Senge, 1990) comes to mind (also see Simons, 1990).

Taken as a whole, restructured school organizations are seen as organic, decentralized and professionally controlled systems (Murphy, 1991). The leadership role of school management that goes with this organizational prototype emphasizes a collegial, "facilitating" role for the school leader, who is able to fully utilize the potential of collaborative teams of "empowered" teachers. Once again, in prescriptive thinking about school organizational functioning the "hu-

man" element is put above mechanistic, bureaucratic and "narrow" technological tendencies. This does not mean, however, that modern multi-media and information-technology does not have a place in this restructuring.

Instructional technologists are eager to point out the possibilities of rich, simulated, learning environments for the type of active learning that is implied in the constructivist pedagogical perspective described in the preceding section (cf. Sheingold, 1991). Indeed, constructivism use of modern information technology and restructuring are seen as three mutually reinforcing ideas (ibid, p. 22). Technology may even be taken as one of the levers to get active learning, adventurous teaching and school restructuring off the ground, since it more or less forces school work, including work schedules, locations, allocations, curriculum and assessment procedures, away from traditional arrangements (Collins, 1991).

At the "above-school level", here referred to as the macro-level, three main aspects of restructuring can be discerned:

- decentralization and devolution of authority;
- "choice", i.e. stimulating market mechanisms;
- accountability requirements.

Decentralization of discretionary power to lower levels appears to be a current tendency in most industrialized countries. (However, large differences between countries remain as to the degree to which authority is divided over administrative levels – cf. OECD, 1992). Decentralization on the one hand and the ambition to impose accountability requirements on the other may appear to be contradictory. The solution to this dilemma is to "liberate" processes and do away with all kinds of regulations, but keep control over outcomes – e.g. by demanding of schools and school districts the publication of a periodical report on their performances.

The idea of stimulating market mechanisms in the domain of educational organizations is to make schools more responsive to the demands of their clients, parents, students and local communities. In this way the "invisible hand" of parents choosing a school for their children would stimulate schools to be efficient. The combination of accountability requirements and increased opportunities for choice can be found in the practice that is being used in the UK, where schools have the obligation to publish examination and test results in local newspapers. On the other hand, there is also a clash between accountability-oriented administrative control on the one hand and choice on the other. Chubb and Moe (1990), for instance, take the position that accountability requirements ushering from higher administrative levels will inevitably lead to bureaucratic tendencies. Warnings against extending the idea of output control to the school level, as for instance in the case of merit pay, are also heard from education administration theorists (Shedd and Bacharach, 1991). They see such practices as completely counter-productive to the cooperative climate that is thought to be so essential for school restructuring.

Although constructivism on the one hand and meso and macro aspects of restructuring on the other, form a rather disjointed set of characteristics, there are also some common features. As was explained in the above, these are shared elements in the perspective on learning (particularly with respect to social aspects of learning) and a common view on control-processes in which higher level authorities are pictured as facilitators of processes where lower-level actors have a lot of discretion.

3. A COMPARISON OF MODELS: IMPLICATIONS FOR THE SCHOOL-LEVEL CONTEXT OF INSTRUCTIONAL EFFECTIVENESS

In section 1 the research literature on instructional effectiveness and school effectiveness was used to arrive at comprehensive conceptual models of educational effectiveness.

In section 2 an alternative model was sketched: constructivism within the context of meso- and macro-aspects of restructuring. Now these two types of models will be compared, where the main objective is to see what can be learned with respect to the meso- and macro-level context of instructional effectiveness. If both types of models were internally consistent and at all comparable one would expect a different set of critical contextual factors to the degree that the perspectives on the instructional process differed and similarities at the contextual level to the degree the instructional processes had similar features.

In Figure 2 a global comparison between the two types of models is made.

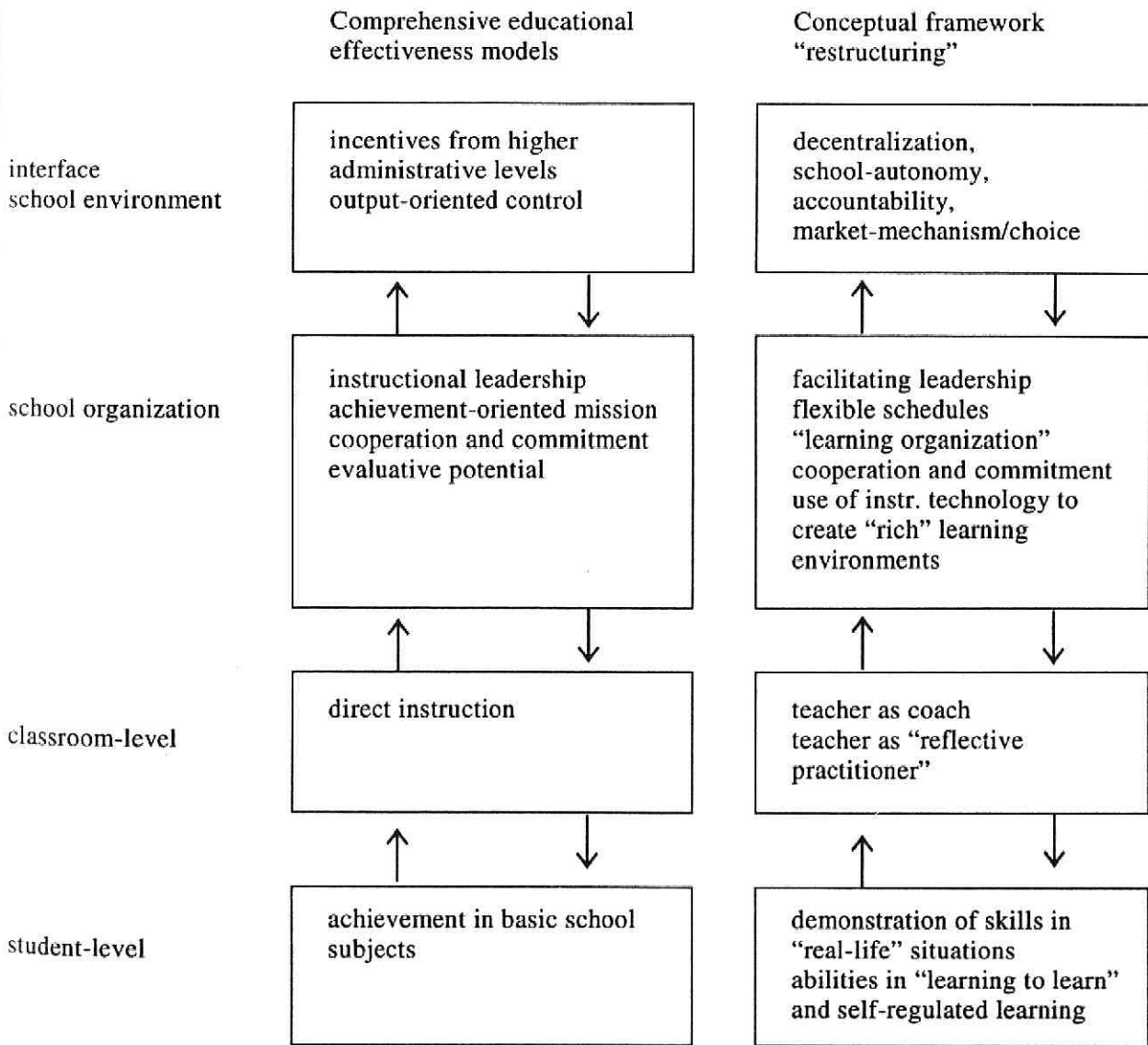


Fig. 2. Global comparison of comprehensive school effectiveness models and conceptual framework of restructuring. Arrows in solid lines indicate the direction for design (bottom up); arrows in dotted lines indicate the actual functioning of the system where higher level processes are seen as facilitative conditions for lower levels.

Before going into the comparison of these two types of models in a more detailed way something more should be said about the *comparability*.

3.1. Comparability of the two frameworks

First of all the position implied in this article and made visible in Figure 2 is that both frameworks can be moulded into structures that share basic ordering principles, namely:

- a multi-level structure which implies an ambition to study educational phenomena “in context”;
- the suggestion that conjectures as to what factors are important at higher levels should start from an analysis of micro-level processes, i.e. the level of learning and instruction (inside-out, or bottom-up principle of design);
- the suggestion that in the actual functioning of the system higher level conditions facilitate processes at lower levels;
- finally, both frameworks share the orientation of goal-attainment, where the ultimate goals are defined at the level of the individual students.

The second of the above points (inside-out analysis) is quite important for the “cross-level” internal consistency of each of the two types of models. Although at first sight constructivism within the context of restructuring may appear (to a European) as a multi-level framework of US-fashions, the plausibility that “active learning” and “adventurous teaching” would require, among other things, flexible schedules and team-work cannot be denied.

Secondly, the empirical status of comprehensive educational effectiveness models on the one hand and the restructuring framework on the other is quite different. The former are the product of an inductive process of making sense of literally hundreds of empirical studies conducted within various traditions of effectiveness-oriented educational research (cf. Scheerens, 1991). Constructivism and restructuring, on the other hand, are like a more or less accidental joining of three different agendas: constructivist learning and instruction, multi-media and interactive educational technology, and the propagation of enlarging opportunities for choice and the operation of market-mechanism in education (Sheingold, 1991). The claims of the constructivist pedagogical view have been criticized for their lack of empirical support (Merrill, 1991). The stronger empirical basis of school effectiveness models also reflects the fact that these are much closer to educational practice, whereas constructivism and restructuring are still to a considerable degree the ideals of educational theorists.

For our purposes the considerably stronger empirical corroboration of educational effectiveness models is only partly relevant, since the careful modelling of contextual conditions of effectiveness enhancing instructional conditions is still very much in its infancy (see Bosker & Scheerens, 1994) and new ideas coming from the restructuring framework might well be incorporated in future modelling studies.

In the third place the question arises as to whether “effectiveness” and “restructuring” as described should not be considered as *different educational approaches aimed at different educational objectives* to such an extent that they are really incomparable. When considering the radical constructivist approach it would indeed be hard to find a common reference point with respect to the definition of an ultimate student achievement assessment criterion. As soon as more moderate constructivist views are at play and alternative assessment criteria are mentioned (such as portfolios, learning to learn-tests, etc. – cf. Jonassen, 1992) comparison is possible – at least from a research perspective. One could conceive of a comparative study in which the effects of direct instruction vs. constructivist teaching would be examined on the union of the two sets of desirable learning outcomes propagated from both positions.

A fourth and final point that is relevant to the question of the comparability of the two frameworks in question concerns the distinction between “steady state” models of effective school functioning on the one hand and dynamic school improvement schemes on the other. Within the framework of school effectiveness, research and improvement have developed as

distinct and even separate fields. Although this logic is by no means compelling, improvement projects could be seen as the application of the school effectiveness research findings (cf. Reynolds, Hopkins & Stoll, 1993).

In the case of restructuring, the impression is given that these two phases (steady state functioning and dynamic school improvement) are really indistinguishable because restructuring is often pictured as a continuous process. (Also compare the image of the learning organization).

Although this would indeed complicate comparative research, two types of solutions are possible. In the case of cross-sectional "snap shot" comparisons, one would somehow have to try and incorporate the "innovation history" of schools in each of the conditions and, secondly, in the case of longitudinal studies, one could follow innovation processes more closely and be in a better position to incorporate them in the comparison. As far as the latter is concerned one could think of a comparative school improvement project, designed as a quasi-experiment.

The conclusion is that, despite major differences in history and general orientation the two frameworks have enough in common to make comparative analysis and research feasible.

3.2. Comparison: agreements and differences on macro- and meso-aspects

At the macro-level, that is the interface between school and environment there is a lot of common ground between the two frameworks under comparison.

Accountability requirements from higher administrative levels and incentives that would force schools to be more responsive to demands from external constituencies (like parents) have a place in both approaches.

In restructuring, the premium is put on market-mechanisms rather than to formal requirements (such as standards) – although it is hard to conceive of accountability requirements without such standards. Next, restructuring emphasizes decentralization, deregulation and enlarging school-autonomy, where this is not a critical feature in the macro-parts of comprehensive school-effectiveness models. Both perspectives may be united on the principle of "liberating" educational process or throughput and exercising a form of output control. Although, as mentioned before, the consistency of the multi-level constructivism/restructuring framework would be lost to the more radical constructivists once accountability requirements were introduced.

It should also be noted that this compromise solution (deregulated processes and output control) masks the inherent tension between market-based and administrative control (cf. Chubb and Moe, 1990). Technical problems and ethical considerations of the mixture of accountability requirements and market-mechanisms in education are left undiscussed here, although the British experience should be singled out as quite relevant in this respect (cf. Scheerens, Korevaar & De Rijcke, 1991).

At the meso-level of school-organizational functioning the importance of cooperation between teachers, teamwork and a commitment to the school's mission are common in both frameworks. The perspective on leadership roles differ, however. In educational effectiveness models the construct of instructional leadership carries elements of achievement-press and hierarchical bureaucratic control, for instance when record-keeping on classroom practice and teacher assessment are emphasized. In restructuring, these bureaucratic and rational control elements are absent. Instead, the facilitative and pedagogical role of school leaders is emphasized, which also corresponds to the focus on teacher empowerment and participatory decision-making in school policy matters.

Although creating conditions for adaptive instruction is an issue for both perspectives, empirical work in the area of educational effectiveness has led to a certain modesty in aspirations with respect to differentiation and individualization, recognizing that instructional time, as a crucial mediating variable may be used more effectively when simpler organizational arrangements, such as whole-class teaching, are used. Constructivism and restructuring are far less inhibited in this respect and much more radical in recommending the redesigning of schedules and allocations.

Within the framework of school-effectiveness thinking, the evaluation potential of schools, (that is the use of evaluation mechanisms to monitor school work, within the framework of the cybernetic principle of feedback and reinforcement), is seen as an important effectiveness-enhancing vehicle (cf. Scheerens, 1992). Restructuring with constructivism at its core is likely to denounce these behaviouralistic procedures. A similar note could be made with respect to the employment of monetary incentives (as in merit pay) which, although being a debatable issue on both sides, would be likely to be abolished altogether from the constructivism/restructuring agenda.

Restructuring, on the other hand, is more open to the scope of modern information technology for education. From this feature it is again evident that constructivism and restructuring are more future-oriented ideals when compared to the pragmatic nature of the outlook of educational effectiveness thinking on "what works" in current practice.

At the meso-level the differences between the educational effectiveness and constructivism/restructuring outlooks appear to come down to some old antagonistic positions in organizational theory. With respect to the organization-structural dimension (this is the distinction between mechanistic and organic structures (cf. Mintzberg, 1979) and for the procedural dimensions, this is the distinction between rational/empirical vs more interactive strategies. In the educational effectiveness outlook there is definitely room for mechanistic structure and rational-empirical strategies although "softer" aspects like a shared mission and cooperative climates are also considered important. Restructuring swings the pendulum back to the human relations approach in its emphasis on teacher autonomy, the reflective practitioner, facilitative leadership and organic structure.

Although it is not the main theme of this article to appraise both perspectives, the present author is somewhat concerned by the assumption that the constructivism/restructuring position contains sufficient levers for improvement, when vehicles for achievement-pressure such as the evaluation-feedback-reinforcement cycle are shielded away from.

CONCLUSION: IMPLICATIONS FOR RESEARCH

In theory, constructivism within the framework of restructuring is a viable alternative to comprehensive educational effectiveness models. To the degree that the constructivist/restructuring alternative is amenable to assessment in terms of student achievement (of whatever kind), crucial experiments concerning the micro-level variants (direct teaching vs. constructivist learning and instruction) are feasible. In fact such "experiments of nature" have been conducted before, within the framework of a large-scale compensatory programme (Bereiter & Kurland, 1982). In that case direct instruction appeared to be more effective, for disadvantaged learners.²

Comparative experiments of nature that include meso- and macro-aspects of restructuring are even harder to conceive of. In fact, such experiments would require no less than the comparison of two large-scale educational improvement programmes inspired on the two frameworks that have been discussed in this article. Of course, research would not be the primary purpose for organizing such alternative programmes. It is therefore quite clear that the feasibility of such a comparison is quite small; although it might be realized in cross-district, cross-state or even international comparisons.

In the mean time it would seem that the ideas and theoretical policy-agendas implied in the constructivist/restructuring framework add rather little in terms of new ideas concerning crucial factors and cross-level relationships on "what works" in education. Exceptions are the enforcement of the importance of collaborative planning, the as yet unrealized potential of modern educational technology to create manageable adaptive instruction and more flexible organizational arrangements and work-schedules. Apart from this the "hard core" of effectiveness-enhancing factors in education: time, opportunity to learn, commitment, teacher quality and

experience, leadership (including press for achievement and control) and reality testing (or the evaluation-feedback-reinforcement function) remain largely unaffected. It should be recognized, however, that the bulk of educational effectiveness research is based upon conditions that have been shown to be positively associated with achievement in basic school subjects, most often at the primary school level and with a bias towards disadvantaged learners. The constructivist/restructuring framework may help in expanding the educational effectiveness research-agenda, by stimulating investigations of the attainment of "higher" and more "formal" cognitive goals (like learning to learn) and attitudes to learning also considering more open instructional processes. At the same time the effectiveness research approach might stimulate studies directed at finding the critical conditions that determine the effects of constructivistic learning environments.

NOTES

1. The author is indebted to Prof.dr. Jules Pieters for helpful suggestions and comments on an earlier version.
2. In this respect Murphy's statement that constructivist instruction is particularly beneficial to disadvantaged students, which he does not corroborate with any research evidence (Murphy, 1993), is not convincing.

REFERENCES

- Argyris, C. & Schön D.A. (1974). *Theory in Practice: Increasing Professional Effectiveness*. San Francisco: Jossey-Bass.
- Bereiter, C. & Kurland, M. (1982). A constructive look at Follow-Through results. *Interchange*, 12, 1-22.
- Boekaerts, M. & Simons, P.R.J. (1993). *Leren en Instructie: Psychologie van de Leerling en het leerproces*. (Learning and Instruction: Psychology of the student and of the learning process) Assen: Dekker & Van de Vegt.
- Bosker, R.J. & Scheerens, J. (1994). Alternative Models of School Effectiveness Put to the Test. (To appear in: *International Journal of Educational Research*).
- Chubb, J.E. & Moe, T.M. (1990). *Politics, Market and American Schools*. Washington D.C.: Brookings Institution.
- Cohen, D.K. (1988). Teaching Practice ... Plus ça Change ... In Ph. Jackson (ed.), *Contributing to Educational Change: Perspectives on Research and Practice*. Berkely, Ca: McCutchan.
- Collins, A. (1991). The role of computer technology in restructuring schools. *Phi Delta Kappan*, 73, 1, 28-36.
- Collins, A., Brown, J.S. & Newman, S.E. (1988). Cognitive Apprenticeship: Teaching the Craft of Reading, Writing and Mathematics. In L.B. Resnick (ed.), *Cognition and Instruction: Issues and Agendas*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Conley, D. (1991). *Restructuring Schools: Educators Adapt to a Changing World*. Eugene, OR: ERIC Clearinghouse on Educational Management.
- Corte, E. de & Lowyck, J. (1983). Heroriëntatie in het onderzoek van het onderwijzen. (Research on teaching reconsidered). *Tijdschrift voor Onderwijsresearch*, 8, 6, 242-261.
- Creemers, B.P.M. (1991). *Effectieve instructie: een empirische bijdrage aan de verbetering van het onderwijs in de klas*. (Effective instruction: an empirical contribution to the improvement of classroom teaching). The Hague: SVO (Balansreeks).
- Cunningham, D.J. (1991). In Defense of Extremism (1991) *Educational Technology*, 31, 9, 26-27.
- Duffy, Th.M. & Jonassen, D.H. (1992). *Constructivism and the Technology of Instruction: A Conversation*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Goldman, P., Dunlop, D.M. & Conley, D.T. (1991). Facilitative Power and Nonstandardized Solution to School Site Restructuring. *Educational Administration Quarterly*, 29, 1, 69-92.
- Jonassen, D.H. (1991). Objectivism versus Constructivism: Do We Need a New Philosophical Paradigm? *Educational Technology Research & Development*, 39, 3, 5-14.
- Jonassen, D.H. (1992). Evaluating Constructivist Learning. In Th.M. Duffy & D.H. Jonassen, *Constructivism and the Technology of Instruction: A Conversation*. (p. 138-148). Hillsdale, NJ: Lawrence Erlbaum Associates.

- Levine, D.U. & Lezotte, L.W. (1990). *Unusually Effective Schools: A Review and Analysis of Research and Practice*. Madison, Wise: National Center for Effective Schools Research and Development.
- Merril, M.D. (1991). Constructivism and Instruction Design. *Educational Technology*, 31, 5, 45-53.
- Mintzberg, H. (1979). *The Structuring of Organizations*. Englewood Cliffs: Prentice Hall.
- Monk, D.H. (1989). The Education Production Function: It's Evolving Role in Policy Analysis. *Educational Evaluation and Policy Analysis*, 11, 1, 31-45.
- Murphy, J. (1992). School Effectiveness and School Restructuring: Contributions to Educational Improvement. *School Effectiveness and School Improvement*, 3, 2, 90-109.
- Murphy, J. (1993). Restructuring Schooling: The Equity Infrastructure. *School Effectiveness and School Improvement*, 4, 2, 111-130.
- OECD (1992). *Education at a Glance*. OECD Indicators. Paris: OECD.
- Resnick, L.B. (1987). *Education and Learning to Think*. Washington D.C.: National Academic Press.
- Reynolds, A.J. & Walberg, H.J. (1990). *A structural model of educational productivity*. Unpublished manuscript. Northern Illinois University.
- Reynolds, D., Hopkins, D. & Stoll, L. (1993). Linking School Effectiveness Knowledge and School Improvement Practice: Towards a Synergy. *School Effectiveness and School Improvement*, 4, 1, 37-58.
- Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School Effectiveness and School Improvement*, 1, 1, 61-80.
- Scheerens, J., Korevaar, J. & De Rijcke, F.J.M. (1991). Onderwijsindicatoren in Engeland, oftewel de vele valkuilen voor een evaluatie-georiënteerde beleidsvoering opnieuw geïllustreerd. (Education Indicators in the U.K.) *Tijdschrift voor Onderwijswetenschappen*, 21, 3/4, 221-226.
- Scheerens, J. (1992). *Effective Schooling: Research, Theory and Practice*. London: Cassell.
- Senge, P.M. (1990). *The Fifth Discipline: The Art and Practice of the Learning Organization*. New York: Doubleday Currency.
- Shedd, J.B. & Bacharach, S.B. (1991). *Tangled Hierarchics: Teachers as Professionals and the Management of Schools*. San Francisco: Jossey-Bass.
- Sheingold, K. (1991). Restructuring for Learning with Technology: The Potential for Synergy. *Phi Delta Kappan*, 73, 1, 17-27.
- Simons, P.R.J. (1990). Leren Leren in een Lerende Organisatie. (Learning to learn in a learning organization) In C. Aarnoutse & M. Voeten (red.), *Gaat en onderwijst: liber amicorum ter gelegenheid van het afscheid van dr. M. Mommers*. Tilburg: Zwijsen.
- Spiro, R.J., Feltowich, P.J., Jacobson, M.J. & Caulson R.L. (1992). Cognitive Flexibility, Constructivism and Hypertext: Random Acces Instruction for Advanced Knowledge Acquisition in Ill-Structured Domains. In Th.M. Duffy & D.H. Jonassen, *Constructivism and the Technology of Instruction: A Conversation*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Stringfield, S.C. & Slavin, R.E. (1992). A hierarchical longitudinal model for elementary school effects. In B.P.M. Creemers & G.J. Reezigt (eds), *Evaluation of Educational Effectiveness* (p. 35-69). Groningen: ICO
- Tobias, S. (1991). An Eclectic Examination of some Issues in the Constructivist-ISP Controversy. *Educational Technology*, 31, 9, 41-43.

Manuscript submitted: October 13, 1993

Accepted for publication: December 23, 1993

School and Instruction Characteristics and Pupils' Arithmetic Achievements at the End of Primary Education

Greetje van der Werf

*GION, Institute for Educational Research, University of Groningen**

ABSTRACT

To establish the effects of school and instruction characteristics on pupils' arithmetic achievements at the end of primary education a secondary analysis was carried out on the dataset of the National Evaluation of the Educational Priority Program in the Netherlands. Multilevel analyses were applied to the data of 183 schools and 2953 pupils. Differences between schools in pupils' arithmetic attainment in grade 8 and pupils' arithmetic achievement gain between grade 6 and 8 were analyzed, taking into account pupils' social-ethnic background, gender and intelligence. Between-schools differences were related to school and instruction variables. The results show that schools differ remarkably in mean arithmetic attainment as well as in achievement gain. The school and instruction characteristics that explain the between-schools difference in attainment are the same as those that are related to the difference between schools in achievement gain. School characteristics that explain these differences are denomination, percentage of ethnic minorities in the school population and the primary school-leaving test of the National Institute for Educational Measurement (CITO). Instruction characteristics that account for these differences are whole-class instruction, homework assignment and regular feedback.

OBJECTIVES

In this article the effectiveness of schools and teachers on the arithmetic achievements of pupils in the final year of primary school is explored. Secondary analyses were applied on the dataset of the National Evaluation of the Educational Priority Program in the Netherlands. Besides data about specific priority activities of schools this dataset contains information on virtually all relevant characteristics that are to be found in literature on school effectiveness and effective instruction (see Scheerens, 1989; Creemers, 1991).

A comprehensive school effectiveness model was developed and tested for its empirical value. The research questions to be answered are:

1. Do schools differ in the average arithmetic achievement level of their pupils, taking into account individual pupils' characteristics?
2. What school characteristics are related to the differences in effectiveness between schools?
3. What instruction characteristics are related to the differences in effectiveness between schools?
4. What is the relative contribution of school and teacher characteristics to the effectiveness of schools?

THEORY

To answer the research questions we developed a comprehensive model of school effects on pupils' learning outcomes, by linking the Dutch tradition of longitudinal school career research with the tradition of research into the effectiveness of schools and instruction.

In the tradition of longitudinal school career research much work has been done in developing causal models of educational achievement based on correlational data (e.g. Peschar, 1975; Meijnen, 1977; 1984; Meester & De Leeuw, 1983; Tesser, 1986). The basic model is a longitudinal one at pupil level in which the effects of social background on the achievements at a certain point in time (t_2) are mediated by intelligence and earlier achievements (t_1). This leads to the lower part of the model in Figure 1.

We extended the pupil-level model to a three-stage multilevel one in which variables at school level are assumed to have effects on teacher-level variables and, whether or not mediated by the teacher-level variables, on the parameters of the individual model. As the individual level model is basically a regression model, school effects are effects on the intercepts and slopes of the regressions of achievements at a certain point in time on achievements at an earlier moment, intelligence and social background within schools.

To extend the model to a multilevel one we took the integrated school effectiveness model of Creemers (1991) as a starting point. The core of the model is the teaching-learning process within classrooms. The first reason for this is that variables at classroom level account for most of the variance in pupil achievement, compared to variables at school level. The second reason is that causal relationships between variables at classroom level and those at pupil level can be

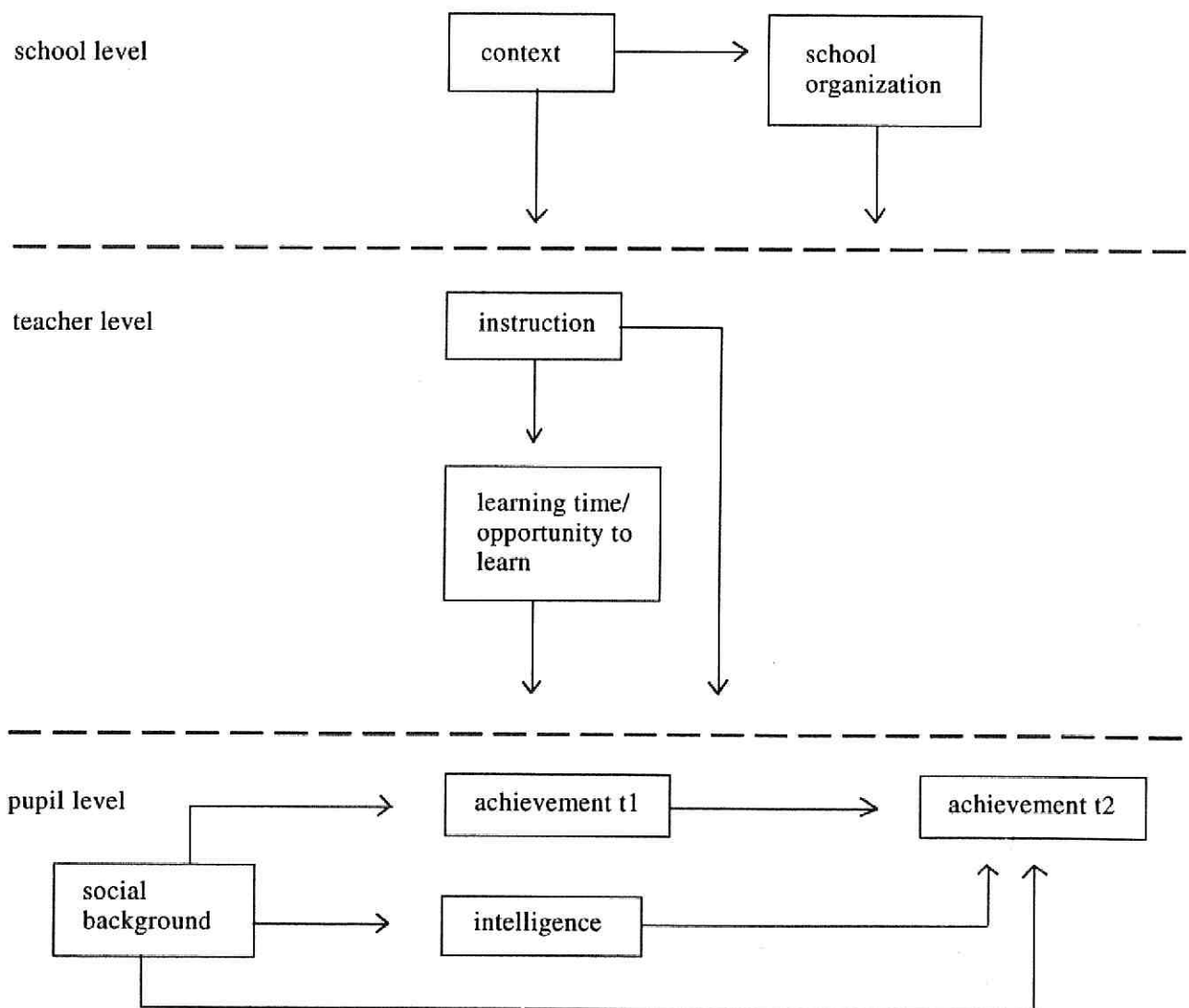


Fig. 1. Multilevel model for differences in pupils' achievement.

specified rather clearly from a theory about learning at school: i.e. the models of Bloom (1976) and Carroll (1963; 1989).

In specifying the variables at teacher level we made a distinction, as Carroll (1965) also does, between learning time and opportunity to learn on the one hand and the quality of instruction on the other. Learning time is defined as the time given to pupils to learn relevant knowledge and skills (e.g. Carroll, 1963; Stallings, 1980; Bloom, 1976; Denham & Lieberman, 1980). Usually 'time allowed' implies the time pupils are allowed to learn within the classroom. Teachers, however, can extend this time by assigning learning outside the classroom (homework). Another important aspect of learning time is 'time-on-task', the time pupils are actually involved in learning (see e.g. Veenman, Lem, Voeten, Winkelmolten & Lassche, 1986).

Opportunity to learn is mainly described as the curriculum covered by teachers in their lessons (Pelgrum, 1989; Creemers, 1991). Both opportunity to learn as well as learning time are assumed to be influenced by the quality of the teaching process (Bloom, 1976; Slavin, 1987; Creemers, 1991). In meta-analyses of research some consistent findings about effective teaching variables are reported (e.g. Brophy & Good, 1986; Fraser, Walberg, Welch & Hattie, 1987; Rosenshine & Stevens, 1986). Creemers (1991) categorized these variables into three groups: the curriculum, grouping of pupils and instructional behaviour. Indicators for the quality of the instruction are goal setting, structuring the lessons, clear presentation plus evaluation and feedback.

The Carroll model of school learning also offers possibilities to specify variables at school level and their relation to instruction and learning at classroom level. Therefore school-level factors are considered to be conditions for classroom level factors. According to Creemers (1991), school-level factors should promote cohesion between teachers or control what is going on in classrooms. School-level factors that could do this include an evaluation system, rules about time use and opportunity to learn, as well as intervision and teacher supervision. Also traditional school effectiveness variables like educational leadership, high aspirations, an orderly climate etcetera (see e.g. Edmonds, 1979; Purkey & Smith, 1983; Rutter, 1983; Good & Brophy, 1986) could function like this.

Finally there are some context factors at school level that are assumed to be conditional for school organization and for the instructional process at classroom level, like school size, population of pupils etcetera.

The model is assumed to be a general model of school learning and to be able to predict all kinds of pupils' achievements, irrespective of subject matter or curriculum content. The model is tested for arithmetic achievements because other research results show that schools differ most in this subject (see e.g. Brandsma, 1993; Van der Werf, Weide & Tesser, 1991).

Moreover, the model is also longitudinal at school and teacher levels. Changes in school characteristics are assumed to have effects on changes in instruction characteristics of teachers, which in turn effects pupils' achievements.

The model will not be fully tested in this article. A three-level analysis is not possible because most primary schools only have one group (class) of pupils of a certain age, with only one teacher teaching. Thus between-class differences in pupils' achievements within schools at a certain point in time cannot be analyzed and related to within-school differences in instruction characteristics between teachers. The effects of changes in school and instruction characteristics will not be tested because only pupils' achievement data at two points in time are currently available.

What will be tested in this article are the relationships between school variables and instruction variables on the one hand and pupils' achievements on the other and the relative contribution of school and instruction variables in explaining these achievements. In fact only analyses on two levels – school and pupil – are applied. Thus instruction variables measured at teacher level, are aggregated to school level.

To explain the differences in pupils' achievements between schools with school and instruction variables, two effectiveness measures will be dealt with – arithmetic attainment at a certain point in time and arithmetic achievement gain between two different points in time. Arithmetic

attainment will be related to school characteristics and aggregated instruction characteristics of *all* teachers within a school. Arithmetic achievement gain will be also related to the instruction characteristics of *only the specific teacher(s) who taught* the pupils concerned *between the two points in time* at which pupils' achievements were measured.

METHOD

Sample

In order to answer the research questions, secondary analyses were performed on the dataset of the National Evaluation of the Educational Priority Policy (EPP). The core of this evaluation is a longitudinal study of successive cohorts of children in primary schools. Data collection takes place once in two years, each time in grades 4, 6 and 8. The children in these grades are respectively about 7, 9 and 11 years old. Until now data have been collected three times (1988, 1990 and 1992).

The evaluation study comprises a 3.5 per cent random sample of all primary schools in the Netherlands, plus stratified samples of schools with varying priority facilities. The whole sample (random plus stratified samples) consists of 700 schools, of which 250 schools are in the random sample. (For further information about the sample refer to Van der Werf, 1993)

The secondary analyses for this article were applied on the 1988 and 1990 data of the grade 6 cohort of pupils in the random sample. This sample is representative for all Dutch primary schools. Only schools where a complete set of data was available were selected. There were 183 schools and 2,953 pupils included in the analyses.

Variables and instruments

For the EPP evaluation every two years children are tested for Dutch language skills, arithmetic and intelligence using standardized tests specifically developed for the purpose (Doddema-Winsemius & Van der Werf, 1989; Van Bergen, 1989).

Information about pupil characteristics, like gender and socio-ethnic background, is collected from the school administration. The socio-ethnic background is indicated by the weight factor each pupil receives in order to determine the amount of priority facilities of schools. A Dutch low SES child counts for 1.25, an immigrant low SES child for 1.90 and a Dutch middle or higher SES child counts for 1.00.

Data on implementation of the EPP, school and instruction characteristics and specific provisions for disadvantaged children are also collected every two years. For this questionnaires are completed by principals and teachers of grades 3 to 8. Some variables in these questionnaires are measured by single items, some by means of scales (reliabilities between .60 and .90).

For the secondary analyses the following pupil variables were used: 1) gender, 2) weight factor, 3) non-verbal intelligence, 4) arithmetic achievement in grade 6 in 1988, 5) language achievement in grade 6 in 1988, 6) arithmetic achievement in grade 8 in 1990. The school and instruction variables used in the secondary analyses are summarized in Table 1.

Two main categories of variables are discerned; those measured at school level and those measured at teacher level.

The school variables are divided into those relating to context, school policy and school organization. Context variables include the denomination of the school (State, Roman Catholic or Protestant), school size (the number of pupils) and the social-ethnic composition of the pupil population as indicated by the percentage of Dutch low SES children (% 1.25 pupils) and the percentage of immigrant children (% 1.90 pupils). School organization variables relate to the school climate, the presence of explicit rules to maintain order, cohesion within teaching staff and educational leadership. These variables are assumed to be important for consistency and control of the teaching and instruction process of individual teachers. School variables that are considered to be more directly important for the pupil achievements include the policy concerning the goals of education (the relative importance of high achievements in basic skills com-

Table 1. Review of school and instruction variables.

School variables	Teacher variables
<i>Context</i>	<i>Curriculum</i>
Denomination (RK, Pr, State)	Using method appropriately
School size	<i>Grouping of pupils</i>
% 1.25 pupils	Whole-class instruction
% 1.90 pupils	
<i>School organization</i>	<i>Quality of Teaching</i>
School climate	Minimum goals
Rules	Controlling learning content
Educational leadership	Repeating learning content
Team cohesion	Evaluation
Feedback	Remedial teaching
<i>School policy</i>	<i>Learning time</i>
Aspirations	Scheduled time for arithmetics
Emphasis on high achiev.	Time allowed within lessons
Evaluation	Homework
.monitoring	
.central registration	
.CITO test	<i>Opportunity to learn</i>
	% of content covered

pared to the emphasis on social skills, attitudes etcetera), aspirations regarding the school careers of pupils and having an appropriate evaluation policy as shown by pupil monitoring, the elementary school-leaving test of the National Institute for Educational Measurement (CITO) and a central registration of pupils' achievements.

The teacher variables are divided into five categories: 1) curriculum: the degree to which teachers use their arithmetic method appropriately; 2) the grouping of pupils: the degree to which teachers give whole-class instruction; 3) the quality of instruction as indicated by effective components such as the use of minimum goals, the frequency of controlling whether subject matter has been mastered, the frequency of repeating learning content not mastered, the frequency of evaluating pupils' progress and giving extra help to pupils where needed; 4) learning time, divided into set periods for arithmetic, time allowed for learning within lessons and time allowed for homework; 5) opportunity to learn or content covered as indicated by the percentage arithmetic curriculum covered by the teacher at the end of the school year.

Analyses

To answer the research questions the data were analyzed using the VARCL program for multi-level analysis (Longford, 1988). The program divides the variance in the dependent variable into variance at pupil level (within-schools variance) and variance school level (between-schools variance). The program takes into account that in some schools there are more pupils present than in others (see also Bosker, 1990).

Two models were analyzed. The first model is 'arithmetic attainment'. In this model the dependent variable is arithmetic achievement in grade 8 (1990), while gender, socio-ethnic background, indicated by the weight factor, and non-verbal intelligence, measured two years earlier, are the co-variates at pupil level. The between-schools variance that is left after taking the pupil co-variates into account expresses the difference in arithmetic attainment between schools. The next step is to add school and aggregated instruction variables (the average scores of all teachers within a school on these variables) blockwise to the model in order to analyze which of these make a significant contribution to explaining the between-schools variance.

The second model is 'arithmetic learning gain'. In this model the dependent variable is arithmetic achievement in grade 8 (1990), while gender, socio-ethnic background, non-verbal intelligence and arithmetic and language achievement in grade 6 (1988) are the co-variates at pupil level. The between-schools variance that is left expresses the difference in arithmetic learning gain between schools. To explain this variance, firstly the school and aggregated instruction characteristics significant to the earlier described attainment model were included. In the next block the average scores of the instruction variables of the specific teacher(s) who taught the pupils between 1988 and 1990 (grade 6 and 7 teachers) were included.

RESULTS

Firstly the results of the arithmetic attainment model will be discussed followed by the results of the arithmetic achievement gain model.

Arithmetic attainment model

In Table 2 the results of the empty model and three models with only pupil covariates, of which the coefficients were modeled fixed, are presented. The percentage of school-level variance left after taking into account the pupil variables of gender, weight factor and intelligence expresses the difference between schools in average arithmetic attainment at the end of primary school. Modelling the coefficients of the pupil co-variates randomly did not deliver a better fit of the models.

Table 2 shows that in the empty model the school-level variance is 18 per cent of the total variance. The pupil co-variates of gender, intelligence and weight factor together account for nearly a quarter of the total variance and about a third of the between schools variance. Comparing models 1 and 2 it is clear that intelligence accounts for more between-pupil variance than the

Table 2. Percentages of variance at pupil and school level for arithmetic attainment in grade 8 (models with only pupil co-variates).

	Pupil level	School level	Total	Accounted for
Empty model	82	18	100	
Model 1: gender + weight factor	75	14	89	11
Model 2: gender + intell.	65	13	78	22
Model 3: gender + intell. + w.f.	62	12	74	26

Table 3. Percentages of school-level variance for arithmetic attainment in grade 8 (models with pupil co-variates and school and instruction variables).

	School level	Accounted for
Model 1	12	—
Model 2	10	17
Model 3	8	33
Model 4	7	42

Model 1: only pupil co-variates included

Model 2: pupil co-variates and context variables included

Model 3: pupil co-variates, context variables and school variables included

Model 4: pupil co-variates, context variables, school variables and instruction variables included

weight-factor. Intelligence has a positive relationship with arithmetic achievement, the effects of the weight factor 1.25 and 1.90 (dummies contrasted with the weight factor 1.00) are negative; the effect of gender is also negative: girls achieve less well than boys. In the final model a 12 per cent variance at school level remains unexplained, which means that the differences in arithmetic attainment at the end of primary school between schools are rather substantial.

In order to discover which school and instruction characteristics explain the between-schools variance, these were included blockwise in the model. In the first block the context variables of denomination, school size, percentage of 1.25 pupils and percentage of 1.90 pupils were included. In the second block the other school variables were added, while in the last one the aggregated instruction variables were included (see Table 1). Table 3 shows the percentages of variance accounted for by each block of variables.

The context variables reduce the unexplained school-level variance by two percent. The variables of denomination and percentage of 1.90 pupils have a significant relation with achievement. Concerning denomination there is only a positive effect of Roman Catholic schools, scoring better than State ones. Protestant schools score the same as State schools. The effect of the percentage of 1.90 pupils is negative.

Adding the other school variables reduces the unexplained between-school variance by another two per cent. The only variable from this category that has a significant relation with average arithmetic achievement is the elementary school-leaving test of the National Institute for Educational Measurement (the CITO test). Schools that use this test every year achieve better results than those that do not.

When the aggregated scores of all teachers on the instruction variables are also included in the model the unexplained variance between schools is reduced by a further one per cent. Three instruction variables have a significant relationship with the arithmetic achievement: 'whole-class instruction', 'the frequency of assigning homework' and 'regular feedback'. All three variables have a positive effect.

In the final model the unexplained variance between schools is still seven per cent. School and instruction variables together account for only 42% of the between-schools variance in the model with only pupil co-variates (model 1 of Table 3) when this percentage (12%) is set to 100 per cent.

Arithmetic achievement gain model

In Table 4 the results of three arithmetic achievement gain models are presented. For comparison purposes the model with only the pupil co-variates of gender, intelligence and weight factor (model 3 of Table 2) is repeated here as model 1. The second model is the one with the earlier arithmetic achievements in grade 6 (1988) added to the three other pupil co-variates of gender, intelligence and weight factor. In the third model the language achievements in grade 6 are included as we assume that these achievements also predict arithmetic achievements in grade 8.

Table 4. Percentages of variance at pupil and school level for arithmetic in grade 8, taking into account arithmetic and language achievement in grade 6.

	Pupil level	School level	Total	Accounted for
Model 1: gender + intelligence + weight factor	62	12	74	26
Model 2: gender + intelligence + weight factor + arithmetic t1	39	9	48	52
Model 3: gender + intelligence + weight factor + arith. t1 + language t1	38	9	47	53

Table 4 shows that including the earlier arithmetic achievements in grade 6 in the model reduces the variance at pupil level by 23 per cent and the variance at school level by three per cent, compared with the model with only the pupil co-variables of gender, intelligence and weight factor. More than half of the total variance and half of between-schools variance is now explained. Adding the earlier language achievements in grade 6 to the model accounts for another one per cent of the variance at pupil level, but does not account for between-schools variance. Comparing the effects of the pupil co-variables in the final model with those in the model with only gender, intelligence and weight factor, it appears that the effects of all three co-variables have become smaller, especially the negative effect of the weight factor 1.90, though they are still present (see Table 5).

It is important to notice that in the most complete model with only pupil co-variables 9 per cent between-schools variance is still left unexplained. This means there is also a rather substantial difference between primary schools in students' arithmetic achievement gain in grades 6 to 8.

Table 5. Regression coefficients for the regressions of arithmetic achievement in grade 8 on the pupil co-variables.

	Model 1	Model 3
Intelligence	1.6	.5
Weight factor 1.25	-1.9	-.8
Weight factor 1.90	-3.8	-1.0
Gender (girl)	-2.2	-1.5
Arithmetic t1		.5
Language t1		.1

Model 1: gender, intelligence and weight factor included

Model 3: also arithmetic and language achievement t1 included

To explain the between-schools difference in learning gain, school and instruction variables are again included blockwise in the model. In the first block the context variables that were significant in the attainment model were included, in the second and third block the significant school variables and the aggregated instruction scores of all teachers. In the fourth block the average scores of only the teachers of grade 6 and 7 on all instruction variables were included. In Table 6 the percentages of variance accounted for by each block in the analysis are presented.

The context variables reduce the unexplained between-schools variance by two per cent. Both variables in this category – denomination and percentage of 1.90 pupils – have a significant effect. Of the school variables only the use of the CITO test was included. This variable reduces the unexplained variance by a further one per cent and has a positive effect on achievements. The three aggregated instruction variables of whole-class instruction, frequency of homework and feedback do not reduce the unexplained variance but deliver a significant better fit of the model. All three variables have a significant positive relation with achievements. Including the grade 6 and 7 teachers' scores on the instruction variables does not reduce the unexplained variance, nor leads to a better fit of the model. None of these variables has a significant effect on achievement, so this block of variables was left out of the model. The final best fitting model, model 4, accounts for 33% of the between-schools variance in arithmetic achievement gain after taking only pupil co-variables into account, when this variance (9%) is set to 100 per cent. Notwithstanding, six per cent of the between-schools variance is left unexplained.

Table 6. Percentages of school-level variance for arithmetic achievement gain in grade 8 (models with pupil co-variates and school and instruction variables).

	School level	Accounted for
Model 1	9	—
Model 2	7	22
Model 3	6	33
Model 4	6	33
Model 5	6	33

Model 1: only pupil co-variates included

Model 2: pupil co-variates and context variables included

Model 3: pupil co-variates, context variables and school variables included

Model 4: pupil co-variates, context variables, school variables and aggregated instruction variables included

Model 5: pupil co-variates, context variables, school variables, aggregated instruction variables and instruction variables of the grade 7 and 8 teachers included

CONCLUSION AND DISCUSSION

In this article an attempt was made to answer the question which school and instruction characteristics are effective for pupils' arithmetic achievements at the end of primary school. To answer this, the variance in arithmetic achievements between schools was first computed, taking into account relevant pupil characteristics like intelligence, gender and social-ethnic background. Two models were applied, an arithmetic attainment one and an arithmetic achievement gain one. The arithmetic attainment model can be considered as the one by which differences in final attainment at the end of primary school between schools can be established. It can be considered as a model of *cumulative school* effects. Attainment in grade 8 is assumed to be the result of school characteristics and the quality of instruction by all teachers throughout the childrens' school career. School variables and instruction data of all teachers within the school were included to explain these effects. The arithmetic achievement gain model can be considered more or less as a comprehensive model of *school and teacher* effects. This model was tested in order to find out to what degree learning gain is the result of school characteristics and the instruction quality of all teachers and to what degree the quality of instruction of specific teachers additionally contributes to learning gain. For this purpose differences between schools in achievement gain between grades 6 and 8 were related to school variables and instruction data of all teachers as well as to the data of specific teachers who taught the pupils concerned between 1988 and 1990.

The results show that schools differ substantially in both arithmetic attainment and arithmetic achievement gain. Respectively 12 and 9 per cent of the total variance is between schools after pupil co-variates have been included in the models. These percentages are consistent with the findings of other research into differences between schools in arithmetic achievements (see e.g. Brandsma & Knuver, 1989; Reezigt, 1993; Brandsma, 1993).

The results also show that the differences between schools in attainment are explained by the same school and instruction variables as the differences in achievement gain.

Of the school and instruction variables only a few account for the between-schools variance in attainment and achievement gain. Two of these include the context variables of denomination and the percentage of ethnic minorities within the schools. The positive effect of Roman Catholic schools has been found earlier in Dutch school effectiveness research (e.g. Hofman, 1993) and is sometimes explained by the fact that the boards of these schools are more directly involved with the quality of education within the schools than the boards of state schools. The

negative effect of the percentage of ethnic minorities in schools was also found earlier (e.g. Tesser, Mulder & Van der Werf, 1991; Brandsma, 1993). Probably the number of ethnic minorities in schools is linked to lower academic expectations by teachers, which in turn can negatively influence pupils' progress.

The only school process variable that has a positive effect is the use of the CITO test. This effect has also been reported by Brandsma (1993). There are two explanations for this. Firstly schools that apply the CITO test prepare children between grades 6 and 8 for this and thus give their pupils more opportunity to learn the relevant skills that are tested. Because the CITO test is very similar to the grade 8 test for evaluating EPP these children also are better prepared for the EPP test. Another explanation could be that children in schools using the CITO test are more used to testing conditions and thus have less test anxiety to make the EPP test. This explanation is only assumed when schools that do not use the CITO test neither do apply another school-leaving test.

There were three instruction variables that have a positive effect on arithmetic achievement: whole-class instruction, assigning homework and feedback. The positive effect of whole-class instruction contradicts the results found in other studies about effective instruction; these other studies mainly reached the conclusion that ability grouping or individualized instruction were more effective than whole-class instruction. However, the latter conclusion was never confirmed in Dutch studies (see e.g. Reezigt, 1993). The positive effects of homework and feedback are in agreement with literature on effective instruction (see e.g. Creemers, 1991; Walberg, 1986; Bloom, 1984). The effect of homework was also found in the first measurement of the EPP evaluation (Van der Werf, Weide & Tesser, 1993).

Finally, it is remarkable that the difference in achievement gain is not explained by the quality of instruction of the specific teachers who taught the pupils between the two moments of achievement measurement, but only by the instructional quality of the entire teaching staff. This leads to the conclusion that the attainment of pupils at the end of primary school as well as pupils' progress in achievements between two points in time are a result of a cumulation of school and teacher effects throughout the entire progress through primary school.

REFERENCES

- Bergen, J.B.A.M. van (1989). *Verantwoording constructie toetsen voor de evaluatie van het onderwijsvoorrangsbeleid*. (Account of the construction of tests for the evaluation of the EPP.) Arnhem: CITO.
- Bloom, B.S. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.
- Bloom, B.S. (1984). The 2 sigma problem: the search for methods of group instruction as effective as one-to-one tutoring. *Educational Researcher*, 41, 4-16.
- Bosker, R.J. (1990). *Extra kansen dankzij de school?* (Extra opportunities due to the school?) Nijmegen: ITS, OOMO reeks.
- Brandsma, H.P. (1993). *Basisschoolkenmerken en de kwaliteit van het onderwijs*. (Primary school features and the quality of education.) Groningen: RION.
- Brandsma, H.P., & Knuver, J.W.M. (1989). *Basisschoolkenmerken als determinanten van het functioneren van leerlingen. Deelrapport II. Resultaten*. (Primary school features as decisive factors of pupil functioning. Report II. Results.) Groningen: RION.
- Brophy, J., & Good, T.L. (1986). Teacher behavior and student achievement. In M.C. Wittrock (Ed.), *Handbook of Research on Teaching*, 3rd ed. (p. 328-375). New York: Macmillan.
- Carroll, J.B. (1963). A model of school learning. *Teachers College Record*, 64, 723-733.
- Carroll, J.B. (1989). The Carroll model: a 25-year retrospective and prospective view. *Educational Researcher*, 18, 26-31.
- Creemers, B.P.M. (1991). *Effectieve instructie. Een empirische bijdrage aan de verbetering van het onderwijs in de klas*. (Effective instruction. An empirical contribution to the improvement of classroom education.) The Hague: SVO Balansreeks.
- Denham, C., & Lieberman, A. (Eds.) (1980). *Time to learn: A review of the beginning teacher evaluation study*. Washington: National Institute of Education.

- Doddema-Winsemius, H., & Werf, M.P.C. van der (1989). *Selectie/constructie van toetsen voor het meten van sociale redzaamheid en intelligentie ten behoeve van de evaluatie OVB*. (Selection/construction of tests for measuring social cognition and intelligence to be used in the EPP evaluation.) Groningen: RION.
- Edmonds, R. (1979). *A discussion of the literature and issues related to effective schooling*. Cambridge: Harvard Graduate School of Education.
- Fraser, B.J., Walberg, H.J., Welch, W.W., & Hattie, J.A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research*, 11, 145-252.
- Good, Th.L., & Brophy, J. (1986). School effects. In M.C. Wittrock (Ed.). *Handbook of Research on Teaching*, 3rd ed. (p. 570-605). New York: Macmillan.
- Hofman, R.H. (1993). *Effectief schoolbestuur*. (Effective school board.) Groningen: RION.
- Longford, N. (1988). *VARCL-manual*. Princeton, N.J.: Educational Testing Service.
- Meester, A., & Leeuw, J. de (1983). *Intelligentie, sociaal milieu en de schoolloopbaan*. (Intelligence, SES and the educational career.) Leiden: Faculty of Social Sciences.
- Meijnen, G.W. (1977). *Maatschappelijke achtergronden van intellectuele ontwikkeling*. (Social background and intellectual development.) Groningen: Wolters-Noordhoff.
- Meijnen, G.W. (1984). *Van zes tot twaalf. Een longitudinaal onderzoek naar de milieu- en schooleffecten van loopbanen in het lager onderwijs*. (From six to twelve. A longitudinal research into the SES- and school effects of educational careers in elementary education.) Harlingen: SVO reeks no. 79.
- Pelgrum, W.J. (1989). *Educational assessment: monitoring, evaluation and the curriculum*. Enschede: Febo.
- Peschar, J.L. (1975). *Milieu, school en beroep*. (SES, school and occupation.) Groningen: Tjeenk Willink.
- Purkey, S.C., & Smith, M.S. (1983). Effective schools: a review. *The Elementary School Journal*, 83, 427-252.
- Reezigt, G.J. (1993). *Effecten van differentiatie op de basisschool*. (Effects of differentiation in primary school.) Groningen: RION.
- Rosenshine, B.V., & Stevens, R. (1986). Teaching functions. In M.C. Wittrock (Ed.). *Handbook of Research on Teaching*, 3rd ed. (p. 376-392). New York: Macmillan.
- Rutter, M. (1983). School effects on pupil progress: research findings and policy implications. *Child Development*, 54, 1-29.
- Scheerens, J. (1989). *Wat maakt scholen effectief? Samenvatting en analyse van onderzoeksresultaten*. (What makes schools effective? Summary and analysis of research results.) The Hague: SVO Balansreeks.
- Slavin, R. (1986). *Quality, appropriateness, incentive and time: elements of effective instruction*. San Francisco: AERA paper.
- Stallings, J. (1980). Allocated academic learning time revisited, or beyond time on task. *Educational Researcher*, 9, 11-16.
- Tesser, P.Th.M. (1986). Sociale herkomst en schoolloopbanen in het voortgezet onderwijs. (Social background and schoolcareers in secondary education). Nijmegen: ITS.
- Veenman, S. Lem, P., Voeten, M., Winkelmolen, M. & Lassche, H. (1986). *Onderwijs in combinatieklassen* (Education in combination groups). The Hague: SVO.
- Walberg, H.J. (1986). Synthesis of research on teaching. In M.C. Wittrock (Ed.). *Handbook of Research on Teaching*, 3rd ed. (p. 214-230). New York: Macmillan.
- Werf, M.P.C. van der, Weide, M.G., & Tesser, P. (1991). *Het onderwijsvoorrangsbeleid in de school en in de klas. De eerste meting in het basisonderwijs*. (The EPP in the school and in the classroom. The first measurement in primary education.) Groningen/Nijmegen: RION/ITS.

Manuscript submitted: November 23, 1993

Accepted for publication: January 3, 1994

School Effects on Drop Out

A multi-level logistic approach to assessing school-level correlates of drop out of ethnic minorities

Roel J. Bosker

*University of Twente, Department of Education**

W.H. Adriaan Hofman

Erasmus University of Rotterdam, RISBO

ABSTRACT

Ethnic minorities drop out of school before certification more often than native pupils. From three theoretical frameworks hypotheses are derived concerning which school factors may increase the chances of ethnic minorities staying on in school. These theories are: empowerment theory, rational choice theory, and social comparison theory. Using a pretest-posttest design at pupil-level, and a cross-sectional design at school-level, 2,523 pupils in 71 schools were involved in this study. Comparisons between drop-outs and stay-ons show that among ethnic minorities the drop-out rate is three times as high as for native pupils. Moreover, differences between drop-outs and stay-ons concerning individual, family, and peer group characteristics were demonstrated. Using a two-level logistic linear model in which probabilities to drop out were predicted, some support for each of the three theories was found, indicating that counselling, and the non-offering of (desintegrating) pull-out programs are positively associated with less drop out, while this is also less in schools where there are many ethnic minorities.

1. INTRODUCTION

In EC member states educational achievement of pupils from recently immigrated ethnic groups systematically lags behind that of indigeneous children (Fase, 1994). Ethnic minorities leave school earlier, drop out more often from the educational system, and obtain lower school qualifications (CERI/OECD, 1987, 1989; De Jong, 1987; Boulot & Boyzon-Fradet, 1988; Smith & Tomlinson, 1989).

In the US finishing high school had become almost universal among young white men and women by 1980: more than 87 per cent reported they had completed high school. Although there was a very rapid growth among blacks finishing high school, about one-quarter of young black adults still did not complete high school by 1980: 76 per cent of black women and 74 per cent of black men reported finishing high school (Jaynes & Williams, 1989). It is difficult however, as Jaynes & Williams (1989) state, to reconcile these kind of estimates with common reports of black high school drop-out rates approaching 50 per cent. Such discrepancies may occur because of significant variation in drop-out rates between schools and districts/regions or it may be caused by the absence of standardization of concepts or methods for measuring high school drop out. These discrepancies may also occur because of response bias in social surveys (drop-outs being less likely to respond).

In the Netherlands about 80,000 youngsters enter the labour market lacking adequate educational qualifications. Some 15 per cent of these do not even have minimum qualifications

*Address: P.O. Box 217, 7500 AE Enschede.

(Education Ministry, 1993). While drop-out figures were decreasing in the eighties, in recent years the percentage of early schoolleavers has increased. Major differences occur in drop-out rates between different ethnic subgroups. Social surveys show drop-out rates for Surinamese and Antillean pupils twice as high, and for Moroccan and Turkish pupils between three and four times as high when compared to native Dutch pupils (van Langen & Jungbluth, 1990).

In this article the role of schools in this respect will be studied, thus adding a new topic to the research agenda for school effectiveness.

2. THEORETICAL FRAMEWORK

2.1. Introduction

Several levels of explanation are potentially important with respect to truancy and drop out in secondary education. At the societal level, low chances on the labour market often lead youngsters to break off their school career. Shifts in the outcomes of competition among schools and businesses for the youth may also possibly explain increasing drop-out rates. Moreover, youngsters more than ever have a say in their school attendance. At secondary school-level continuing selection procedures can contribute to increasing drop-out rates. Family characteristics are seen as important factors in school careers, e.g. socio-cultural background variables like low formal education and occupational status of parents, family income, family size, ethnic minority status, or a native language other than the vernacular. Peer group behaviour has significant effects on achievements and drop out while individual characteristics also play an important role. Among these are instrumental qualifications such as non verbal and verbal intelligence and functional qualifications that include the presumed lower achievements of drop-outs (Chaikind, 1987). Aptitude, self-esteem, motivation, and the way the school is perceived as an organisation and social system are decisive factors when being on the point of dropping out (De Vries, 1992).

This article focuses on school-level factors associated specifically with ethnic-minorities dropping out, though possible individual, peer group, and family factors will be considered as potentially compounding school effects. Accordingly, the theoretical framework chosen concentrates on the potential school effects on ethnic-specific drop out.

2.2. Empowerment theory

At school organisation level theories about socialization processes within the school are the point of departure for constructing school effect models. These capitalize on the opportunities of improving the school careers of minorities by cultural and linguistic integration. One explanation regarding ethnic specific school effects that offers clues for building a school model is Cummins' empowerment thesis.

Cummins (1990) introduced his empowerment theory as a means of explaining ethnic specific school effects. From an international perspective, patterns of educational failure of minorities support the assumption that power and status relations between minority and majority groups exert a major influence on school performance. A few examples illustrate this (Cummins, 1986). Academic failure of Finnish pupils in Sweden, where they constitute a low status group, compared to their success in Australia where they are regarded as a high status group, is a case in point. Ogbu (1978) reports that the outcast Burukamin perform poorly in Japan but perform as well as other Japanese pupils in the United States. Cummins claims that power and status relations between minority and majority pupils have a major impact on school performance. Minorities who experience difficulties in their school career often develop, comparatively speaking, an insecure and ambivalent attitude with regard to the value of their own cultural identity as a result of their interaction with the dominant group. Therefore it is hypothesised that the occurrence of such effects will be reduced through 'empowering' ethnic pupils in the school and class situation. In short, migrant pupils are 'empowered' or 'disabled' as an immediate outcome of their interaction with educators. This interaction is mediated by implicit or explicit role defini-

tions educators assume in relation to two institutional characteristics of schools: (1) Incorporation of the language and culture of minorities into the curriculum and (2) community participation.

The extent to which pupils' language and culture are incorporated into the curriculum constitutes a significant predictor of academic success, especially for ethnic minorities that experience disproportionate levels of academic failure (Campos & Keatinge, 1984; Cummins, 1984; Appel & Muysken, 1987; Cummins, 1986; Hakuta, 1986). Cummins emphasizes that this does not require the actual teaching of minority languages. In many cases forming a minority language class may be impossible for reasons such as a small concentration of particular minority groups. Educators, however, can communicate to pupils and parents in a variety of ways to the extent that minority culture will become more valued within the school context.

Community participation, as the second relevant institutional characteristic of schools, refers to the degree that participation of minorities, particularly the parents, is stimulated as an integral part of the school program. When educators involve minority parents as partners in their children's education, these parents appear to develop a sense of efficacy that communicates itself to the children, with positive academic consequences. It seems possible to generate rather important effects on the educational attainment of minorities when school policy stimulates parental involvement in the curriculum (Brookover & Lezotte, 1979; The Haringey-project serves as an example at primary school-level; Tizard, Schofield & Hewison, 1982). Based on research in primary education, Mortimore et al. (1988) conclude that parental involvement in school policy matters correlates positively with academic performance. Tesser (1986) also reports positive influences of parental involvement on educational attainment levels in Dutch secondary education.

From empowerment theory the following hypotheses are deduced:

ethnic minorities will drop out less when:

- a) *schools pay more attention to incorporating a minority language and culture into the school program, organize multicultural projects and make serious efforts to integrate counselling and extra lessons for pupils-at-risk into the regular program and*
- b) *schools intensify contacts with the local community, especially parents.*

2.3. Rational choice theory

Rational choice theory is concerned with explaining differential selection effects. Inequal educational opportunities arise from pupils' differing social backgrounds and becomes evident in the social-specific differences of their achievement levels (primary effects) and educational choices (secondary effects). Primary effects reflect influences of different socio-cultural circumstances (e.g. education experience of the parents, their educational support and ambitions) on pupil achievement. Secondary effects occur when pupils from different social origins who have similar achievement levels systematically make different choices. A different cost-benefit analysis often causes middle and upper class pupils to opt for higher types of education and a higher course-level than working-class pupils.

Boudon (1974) introduced his rational choice theory contending that the (subjective expected) costs of choosing a more prestigious track in secondary education are increasing when the pupil has a relatively low socio-economic status. The (subjective expected) benefits of choosing and completing a more prestigious track in secondary education are larger for pupils with a high socio-economic status. From this model we derive possibilities for school policy to manipulate (lower class) pupils' educational opportunities, for example by making particular goals more attractive for these pupils, by offering professional care and counselling, by postponing selection, by formalizing selection procedures and by systematically counselling pupils and parents on matters of choice. In short, such a model offers support for formulating (parts of) a school effect theory. This approach from rational choice theory is used as a partial explanation of class specific school effects because minorities are disproportionately overrepresented in the lower class category.

From rational choice theory the following hypotheses are deduced:

ethnic minorities will drop out less when:

c) schools intensify their efforts on counselling pupils-at-risk, formalize school rules and make them more transparent, prolong the comprehensive school period and use intra-group differentiation methods.

2.4. Contextual effects

Finally, some hypotheses are related to the so-called contextual effects of schools. Social comparison theory (Festinger, 1954) pays attention to higher level conditions under which comparison processes between individuals appear. It is stated that pupils develop their competencies by comparing these with those of their fellow peers. This relative comparison then affects their academic self-concept. Hüttner (1980) reports that with a relatively high level of average achievement increasing numbers of pupils will fail to attain that level. In other words: when the distance between individual and aggregate scores increases individual performance will be less influenced by this average score. More specifically, social reference theory states that pupils will be positively affected when their own cognitive or social background is linked more closely to the background composition of the school. A study by Aitkin & Zuzovski (1992) corroborates this hypothesis. They report decreasing performances of high aptitude pupils from working-class origin when the average social origin composition of the school increases. Hofman (1993) concludes that in a Dutch school setting pupils from low socio-economic backgrounds indeed achieve more in schools with a relatively low social background composition.

From social comparison theory it is then hypothesised that the composition of the school population will have an effect on drop out:

ethnic minorities will drop out less when:

d) the pupil population of a school has a high percentage of ethnic minorities.

3. RESEARCH DESIGN

3.1. Sample

This study is based on a pretest-posttest design at pupil-level, and on a cross-sectional design at school-level. In 1990 215 Dutch secondary schools were randomly sampled, after stratification for percentage of ethnic minorities and type of education. When asked to participate in the survey 75 out of 215 schools responded positively (35%). Since the response rate was low, several checks on possible response bias were done (Hofman, 1993), the most important of which is presented in Table 1.

From the results presented in table 1 it can be readily deduced that the non-responding schools do not differ systematically from the response group with respect to the ethnic composition of the pupil populations ($\chi^2 = 4.24$, d.f. = 2, $p = .12$).

The data at school-level were collected by face-to-face interviews with the school principal. Information was collected from 3,066 eighth, and ninth grade pupils regarding aptitude, achievement, class and ethnic origin, family characteristics, motivational and peer group behaviour at the first timepoint of measurement. These data were collected in one or more classes in every school by trained test leaders. Teachers were involved in data collection procedures only when (dis)order required so. Measurements at pupil-level at the second timepoint (1992) were aimed

Table 1. Responding and non-responding schools grouped per percentage of ethnic minorities.

	non-response	response
0-10%	67 (48%)	26 (35%)
10-20%	30 (21%)	24 (32%)
> 20%	43 (31%)	25 (33%)

at retrieving information about the careers of these pupils after a two year interval. We were given access to drop-out data on 2,523 pupils (out of 3,066).

3.2. Variables

In this article the focus is on correlates at school-level of drop out. But since self-selection may cause differences between the pupil populations of schools, possible individual, peer group and

Table 2. Descriptive statistics of pupil-level variables (n=2523).

	Mean/Prop.	s.d.
pupil variables		
drop out	7%	
logical reasoning	14.96	2.88
vocabulary	5.89	2.04
arithmetic	3.21	1.27
language	3.68	1.37
ethnicity	24%	
intention to drop out		
no	78%	
yes	20%	
missing	3%	
intention to stop studies after certification		
no	65%	
yes	31%	
missing	4%	
motivation	34.12	6.65
well-being	40.97	5.75
truancy	2.05	1.42
homework behaviour	5.40	1.53
family variables		
time in the Netherlands	13.49	2.29
amount of conversation in Dutch	7.06	3.21
one-parent family		
no	82%	
yes	15%	
missing	2%	
educational support	23.57	3.99
work father		
unemployed	8%	
employed	88%	
missing	4%	
family size	3.08	1.98
no older sibling at school	63%	
peer group variables		
friends		
yes	90%	
no	7%	
missing	3%	
leisure time with peers	9.11	1.78
school time with peers	5.12	1.40
cultural composition peer group	1.85	.88
truancy of peers		
never	83%	
some or sometimes	13%	
missing	4%	

family correlates of drop out will be taken into account. At the individual-level the relative importance of intellectual aptitude, motivation, ethnicity, and truancy behaviour forms the core of the issue. Furthermore, effects of the linguistic and cultural composition of the peer group and their activities will be assessed. Family factors include socio-cultural variables such as: daily used language used within the family, family status (single parent family or not), (non-) employment of family members, educational supportive behaviour of the parents and so on. An overview of these variables and their moments is presented in Table 2.

Details on the operationalization of the variables are to be found in Hofman (1993). Logical reasoning is the PSB-3, vocabulary, arithmetic and language are CITO-tests, well-being is a scale developed by Stoel (1983), and motivation is a scale developed by De Corte et al. (1982). Intention to drop out can be viewed as the pre-test score (actual drop out being measured two years later), while other potential drop out predictors are intention to stop studies after certification, truancy behaviour (number of lessons missed in the four weeks preceding the first time-point of measurement) and attitude towards homework (daily hours spent on homework, scaled from 1 (no hours) to 8 (more than 2 hours)). Most family variables speak for themselves. Time in the Netherlands is the number of years the pupil has lived in the Netherlands, amount of conversation in Dutch is the sum of four indicators, i.e. Dutch language ability of both parents, extent of conversation in Dutch between pupil and parents and siblings, and educational support is a scale based on Jungbluth (1985). The peer group variables include the amount of time spent with peers on leisure activities, idem for educational activities, truancy behaviour in the peer-group, and the cultural composition of the peergroup (from mono- to multicultural).

Information on the school variables is presented in Table 3.

The variables are ordered in four blocks. The first block contains school variables that may have potential predictive power, according to, for instance, reviews on the impact of school climate, but that do not fit into the theoretical framework and might therefore be viewed as covariates, such as curricular track, anti-truancy policy, indicators for pupil misconduct and percentage of conflicts with a racial overtone. From empowerment theory two subsets of variables can be derived. The first subset refers to integration: adapting curriculum for ethnic minorities, number of extra-curricular lessons for ethnic minorities during regular lessons (pull-out), and home-school integration (i.e. integrating language and culture of ethnic minorities into the regular curriculum). The second subset refers to community participation: contacting parents (63% of the schools succeeding in this) and parental participation in school decision-making (i.e. the number of parent meetings organized by the school per year). From rational choice theory two subsets of variables are also deduced. The first subset relates to homework counselling, extra language or arithmetic lessons and study skills lessons (in general or for ethnic minorities in particular). The second subset has to do with postponing selection: length of orientation period, heterogenous grouping, differentiation only in secondary stage, ability grouping (i.e. setting) and possibly allowing for differences in learning rates. The last group of variables is derived from contextual effects theory: percentage of ethnic minorities, the increase in this number over the last five years, as well as the percentage of ethnic minorities enrolling for school after the first grade and the percentage of ethnic minority teachers.

3.3. The two-level logistic linear model

We assume a two-level structure (ignoring the classroom level for convenience sake). At macro-level there are 71 schools randomly sampled from an (assumed) infinite population of schools. At microlevel there are 2,523 pupils assumed to be randomly sampled per school.

The dependent variable is defined at pupil-level and is denoted p_{ij} (0 being the score of pupil i in school j not dropping out and 1 for drop-outs). Instead of this dichotomous outcome variable the logit is used, which is defined as $\text{logit}(p) = \ln(p/(1-p))$. At pupil-level the usual logistic model can

Table 3. Descriptive statistics of school-level variables (N=71).

	Mean/Prop.	s.d.
Covariates		
curricular track		
lower vocational		
home economics	17%	
technical	25%	
agri-cultural	16%	
lower general	17%	
higher general	16%	
otherwise/Middle school	10%	
misconduct at school	2.65	3.16
conflicts with racial overtones	1.18	.49
increase in misconduct	-2.23	27.23
anti-truancy policy	1.61	.93
Empowerment variables		
a)incorporation		
adapting curriculum for ethnic minorities	51%	
pull-out .35	.68	
home-school integration	2.85	1.61
b)community participation		
home-school consultation	10.87	10.56
contacting parents	63%	
Rational choice variables		
counselling	7.86	1.84
counselling for ethnic minorities	17.51	17.43
length of orientation period	1.45	.79
heterogeneous grouping	13%	
differentiation in secondary stage	28%	
ability grouping	61%	
allowing for differences in learning rate	37%	
Contextual variables		
percentage of ethnic minorities	17.47	16.35
percentage of ethnic minorities entering school after first grade	4.19	11.93
increase in ethnic minorities	17.08	31.61
percentage of ethnic minorities teaching	2.73	3.67

be applied (Snijders, 1991). For pupil i in school j a dichotomous response is observed: $Y_{ij} = 1$ (for drop-outs) and $Y_{ij} = 0$ otherwise. It is assumed that (Wong & Mason, 1985, p. 515)

$$(1) Y_{ij} | p_{ij} \sim \text{Bernoulli}(p_{ij})$$

where $p_{ij} = \Pr(Y_{ij} = 1)$ is the probability that pupil i in school j will be a drop-out.

Using the (non-linear) transformation a two-level linear model for the logits can be formulated. The independent variables at pupil-level are denoted x_{1ij} to x_{kij} and at school-level we have a set of predictor variables z_{1j} to z_{mj} . The two-level model can now be formulated as:

$$(2) \text{logit}(p_{ij}) = \beta_{0j} + \beta_{1j}x_{1ij} + \dots + \beta_{hj}x_{hij} + \beta_{(h+1)j}x_{(h+1)ij} + \dots + \beta_{kj}x_{kij}$$

Formula (1) is the pupil-level model, where the random variable $\text{logit}(p_{ij})$ is regressed on a set of pupil-level predictors and/or covariates (e.g. aptitude, achievement, ethnicity, motivation, etc.). Note that there are two types of regression coefficients: β_{0j} , and β_{1j} to β_{hj} (bold faced and indexed with the subscript j) are random coefficients and $\beta_{(h+1)}$ to β_k are fixed. The interpretation of the fixed regression coefficients is straightforward: for every unit increase in x , the logit is predicted to increase with β units. The random regression coefficients indicate the regression of y on x may vary between schools.

Next to the pupil-level model a school-level model is formulated.

$$(3a) \beta_{0j} = \beta_{00} + \gamma_{01}z_{1j} + \dots + \gamma_{0m}z_{mj} + u_{0j}$$

$$(3b) \beta_{1j} = \beta_{10} + \gamma_{11}z_{1j} + \dots + \gamma_{1m}z_{mj} + u_{1j}$$

$$\vdots$$

$$(3i) \beta_{hj} = \beta_{h0} + \gamma_{h1}z_{1j} + \dots + \gamma_{hm}z_{mj} + u_{hj}$$

In (3a) the school-specific intercept (the expected value for the logit in school j in case all the x 's are zero) is modelled as a function of the school-level predictors and/or covariates. If the x -variables all have a mean of zero, then β_{0j} is the expected logit in school j for the average pupil in the population. β_{00} is the GRAND MEAN: the expected logit for a pupil in case all the x 's and z 's are zero. γ_{01} is the regression coefficient associated with the regression of β_{0j} on the school-level variable z_1 (for each unit increase in z_1 the predicted increase in β_{0j} is γ_{01} units).

Note that, as is usual in logistic regression the pupil-level model (2) does not contain an error term. Since we are analyzing a dummy-outcome variable the variance of these errors would be confounded with the mean. At school-level residuals are u_{0j} to u_{hj} . These residuals are multivariate normally distributed, with zero mean and variance-covariance τ . The school-level residuals may thus have a non-zero covariance (see: Bosker & Snijders, 1990).

3.4. The strategy for model building

In the theoretical section of this article four hypotheses were formulated on school factors with potential predictive power. The modelling strategy will be that first of all a covariates model will be fitted on the data, containing all pupil, family, and peer group variables as well as specific school-level covariates. The hypotheses have to do with, what might be called, cross-level interaction effects of school variables with ethnic minority status on drop out. These effects, however, are tested after having taken into account the main effects of these variables: maybe all pupils benefit from these school factors. The four hypotheses are tested simultaneously. The theories are not viewed as competing but as supplementary, the falsification of which can be achieved by finding no effects at all for variables derived from one of these theories. The theories are only partially seen as possibly competing, in the sense that variables derived from one of the theories may have greater predictive power than other groups of variables.

4. RESULTS

4.1. Introduction

166 of the 2,523 pupils dropped out of school before certification, which equals seven per cent. In the on-going simple comparisons will be made between these drop-outs and the stay-ons on individual, peer group and family characteristics, before putting the school effects hypotheses to the test.

4.2. Simple comparisons of drop-outs and stay-ons

To what extent does a drop-out differ from a stay-on? Table 4 contains information on these differences.

Table 4. First order correlates of drop out at pupil-level (n(stay-ons)=2357; n(drop-outs)=166).

	stay-ons		drop-outs	
	Mean/Prop.	s.d.	Mean/Prop.	s.d.
pupil variables				
logical reasoning	15.02	2.85	14.19	3.10
vocabulary	5.94	2.02	5.07	2.25
arithmetic	3.24	1.26	2.79	1.41
language	3.71	1.34	3.21	1.58
ethnicity	22%		48%	
intention to drop out				
no	80%		57%	
yes	18%		38%	
missing	2%		5%	
intention to stop studies after certification				
no	66%		46%	
yes	30%		46%	
missing	4%		8%	
motivation	34.21	6.58	32.91	7.60
well-being	41.13	5.57	38.67	7.52
truancy	2.02	1.41	2.42	1.52
homework behaviour	5.44	1.52	4.84	1.69
family variables				
time in the Netherlands (n.s.)*	13.51	2.21	13.17	3.19
amount of conversation in Dutch	7.18	3.15	5.35	3.57
one-parent family				
no	84%		66%	
yes	14%		30%	
missing	2%		5%	
educational support	23.64	3.92	22.60	4.71
work father				
unemployed	8%		13%	
employed	88%		78%	
missing	4%		10%	
family size	3.04	1.93	3.59	2.54
no older sibling at school (n.s.)	63%		61%	
peer group variables				
friends (n.s.)				
yes	91%		84%	
no	7%		8%	
missing	3%		8%	
leisure time with peers (n.s.)	9.09	1.79	9.29	1.61
school time with peers (n.s.)	5.13	1.38	5.01	1.56
cultural composition peer group	1.83	.88	2.07	.91
truancy of peers				
never	84%		66%	
some or sometimes	12%		27%	
missing	4%		7%	

* n.s. indicates non-significant differences between drop-outs and stay-ons ($\alpha < .01$).

The first important finding is that ethnic minorities drop out of school more often than other pupils. Since the figure of 48 per cent may easily lead to false conclusions, it may be more useful to present the finding as: 14 per cent of ethnic minorities drop out against only four per cent of other pupils. Two years before the actual drop out was assessed many of these drop-outs already had no intention of finishing school, or if they had, they had no intention of further study. Furthermore drop-outs score lower on aptitude tests, achievement tests, motivation, well-being, truancy behaviour and homework behaviour than the stay-ons.

The following family differences between drop-outs and stay-ons were found. Drop-outs are more often living in single parent families, have more siblings, an unemployed father, have less conversation in Dutch with family members and receive less educational support than stay-ons. No differences are found between drop-outs and stay-ons with respect to length of stay in the Netherlands or having an older sibling at the same school.

As to peer group characteristics, only two differences between drop-outs and stay-ons were found. Peers of drop-outs more often play truant, and their peer group is more multicultural than that of stay-ons. No (significant) differences were found with respect to having friends and the kind of activities they engaged in.

4.3. School effects on drop out

After having assessed differences between drop-outs and stay-ons at individual-level, the testing of the four school effect hypotheses will be done after taking into account initial differences between drop-outs and stay-ons. Although many differences were found, maybe only a few of these should be taken into account because of multicollinearity (e.g. that drop-outs score lower on achievement tests may be explained by their lower scores on aptitude tests). For this reason a parsimonious model will be fitted to the data before testing the school effect hypotheses. Covariates will be entered in blocks: first a block of pupil variables, then one of family variables, followed by a block of peer-group variables, and then one with school-level covariates. Only then will the effects of those variables derived from the hypotheses be estimated (first their main effects and then their interaction with ethnicity). Potential effects of these variables should thus be interpreted as: the effects of schools on drop out in as far as this cannot be accounted for by initial differences between drop-outs and stay-ons concerning future plans, aptitude, achievement, family circumstances, peer group characteristics, etc.

The results of the multilevel logistic regression analyses are presented in Table 5.

The first model fitted was the empty model. The most important finding resulting from this was the significant between school variance (.808) in drop-out rates, which indicates that the drop-out rates of schools in terms of a ten per cent prediction interval vary from .01 to .30. The second model contains the pupil variables. Ethnicity still has an effect once the intention to drop out, aptitude (the vocabulary test) and homework behaviour are taken into account. Of the family variables, only living in a single parent family adds to the prediction of drop out. Leisure time spent with peers and truancy behaviour of peers are the two peer group variables associated with drop out. Once all of these covariates are in the model, none of the school covariates (like curricular track, conflicts, etc.) have a significant (additional) effect on drop out. At this point the hypotheses can be put to the test: can the effects of integration, community participation, counselling, selection and the composition of the school population on drop out be demonstrated? The column under the 'school' model gives the results. Pull-out and counselling have effects on drop out, although including these variables in the prediction model does not lead to a better fit of the model to the data (the improvement of the model is: $\chi^2 = 4.35$, d.f. = 2, $p = .14$). These variables, however, being collinear with covariates have significant effects on drop out. The collinearity then should be interpreted for example as: pull-out programs are more likely to be in schools where relatively many pupils live in single parent families and have peers playing more than average truant. Empowerment theory as well as rational choice theory have provided variables that indeed correlate with drop out. Contextual effects come in only at the cross-level interaction model: the size of the effect of ethnicity on drop out is dependent on the composition

Table 5. Results of the logistic multilevel regression (standard errors between brackets).

	empty	pupil	family	peers	school	cross-level
grand mean	-2.629	-1.829	-2.077	-3.131	-1.957	-2.145
vocabulary		-0.093 (.044)	-0.095 (.044)	-0.102 (.044)	-0.092 (.044)	-0.092 (.044)
ethnic minority		1.050 (.184)	1.025 (.185)	1.036 (.187)	1.045 (.187)	1.390 (.282)
intention to drop out*		0.988 (.183)	0.989 (.184)	0.911 (.188)	0.972 (.184)	0.952 (.188)
homework behaviour		-0.190 (.058)	-0.175 (.058)	-0.165 (.050)	-0.148 (.055)	-0.141 (.058)
one-parent family*			0.784 (.196)	0.744 (.199)	0.698 (.193)	0.715 (.199)
leisure time spent with peers				0.101 (.050)	0.103 (.049)	0.107 (.050)
truancy of peers*				0.631 (.208)	0.609 (.202)	0.608 (.207)
pull-out					0.268 (.109)	0.320 (.162)
counselling					-0.224 (.060)	-0.203 (.078)
percentage of ethnic minorities						0.014# (.010)
interaction of ethnicity and percentage of ethnic minorities						-0.019 (.011)
school-level variance	.808 (.197)	.547 (.132)	.538 (.132)	.519 (.128)	.329 (.082)	.282 (.071)
deviance	1181	1087	1070	1057	1053	1050
d.f.		5*	1	3*	2	2
p-value		.000	.000	.005	.114	.259

* an extra degree of freedom was needed to model the category "missing" for the variables 'intention to drop out', 'one-parent family', and 'truancy of peers'.

effect not significant ($\alpha < .05$).

of the school population (in terms of percentage of ethnic minorities). Once again, however, the fit of the model does not improve, indicating that the variables added to the logistic regression equation are collinear with variables already in the model. To facilitate interpreting the effects of the variables Table 5 is transformed into Table 6, where the effects are presented as predicted probabilities of drop out.

The result under the 'empty' column indicates the estimated probability for an "average" pupil to drop out is .067. The 'pupil' model contains the results for pupils with particular scores on predictor variables. For ethnic minorities with an average score on the vocabulary test and average homework behaviour, who initially had no intention to drop out, the estimated drop-out probability is .086 (= .032 + .054), whereas native pupils with the same initial situation have

Table 6. Results in terms of likelihood to drop out (for dichotomous variables: the increased probability to drop-out for category 1 (when compared to category 0) and for continuous variables the increased probability for a pupil who has a score of + 1 s.d. or - 1 s.d. on the predictor variable (when compared to the "average" pupil, i.e. a pupil with a '0'-score on the dichotomous variables and an average score on the continuous variables)).

	empty	pupil	family	peers	school	cross-level
"average" pupil	.067	.032	.027	.024	.018	.024
- 1 s.d. vocabulary		+0.006	+0.006	+0.005	+0.004	+0.005
ethnicity		+0.054	+0.045	+0.041	+0.031	+0.065
intention to drop out		+0.049	+0.043	+0.034	+0.028	+0.036
- 1 s.d. homework behaviour		+0.010	+0.008	+0.007	+0.004	+0.006
one-parent family			+0.030	+0.025	+0.017	+0.024
+ 1 s.d. leisure time spent with peers				+0.005	+0.003	+0.005
truancy of peers				+0.020	+0.014	+0.019
+ 1 s.d. pull-out					+0.003	+0.006
- 1 s.d. counselling					+0.009	+0.010
+ 1 s.d. percentage of ethnic minority pupil for ethnic minority pupils						-0.002
+ 1 s.d. percentage of ethnic minority pupil for other pupils						+0.005

an estimated probability of .032. Moreover, it can be seen that pupils who were initially inclined to leave school before certification have an increased probability of dropping out of .049 (*ceteris paribus*). A lower aptitude score increases this probability by .006, while paying less attention to homework results in an increase of .010. Being a child from a single parent family increases the probability of dropping out by three per cent, whereas having peers who play truant leads to an increase of two per cent.

The effects of the school variables are very modest. Less counselling, as deduced from rational choice theory, is indeed correlated with the probability of dropping out (an increase of 0.9 per cent). Providing pull-out programs, i.e. providing lessons for ethnic minorities during regular lessons instead of combining both is also correlated with drop out. The prediction of

empowerment theory that desintegration will lead to an increased probability of dropping out finds some support in the data (an increase of 0.3 per cent). What should be noted, however, is that both variables are correlated with drop out irrespective of the origin of the pupil. Moreover, it should be pointed out that the other predictions (concerning community participation, selection, contextual effects) up till this point find no support in the data. Since the hypotheses are focused on drop out of ethnic minorities (and although they benefit as much from counselling, for example, as other pupils), special attention should be paid to the cross-level interaction effects model, since it allows for estimating differences in drop out between ethnic minorities and native pupils within schools. The results presented under the 'cross-level' column show that the composition of the school population, as was predicted on the basis of contextual effects theory, is differentially correlated with the drop out of pupils, depending on them being from ethnic origin or not. An ethnic minority in a school where 34 per cent of the pupils are also ethnic minorities (being a school 1 s.d. above average on this variable) has a somewhat smaller probability of dropping out when compared to a pupil of native origin in the same school (0.7 per cent). No other cross-level interaction effects find support in the data.

5. DISCUSSION

Ethnic minorities are more likely to drop out of school before certification than other pupils. This phenomenon is once again found in the present study, conducted on 71 schools with over more than 2,500 pupils. Fourteen per cent of ethnic minorities against only four per cent of pupils of native origin dropped out. Although it is sometimes argued that the implications of drop out should not be exaggerated, since drop-outs may just as well return to education later (De Vries, 1993), it is clear from both a macro-level and a individual-level viewpoint that drop out is, if not the cause, then at least, associated with certain problems. The societal one is that educational investment in these drop-outs did not pay off, that ethnic-specific inequality in status attainment will follow ethnic-specific drop out, while the individual problem is that without a certificate job earnings will be less (Meesters, 1992).

As a byproduct of the analyses it was found, that drop-outs differed from stay-ons with respect to aptitude, achievement, homework behaviour, truancy behaviour, initial intentions to finish school or to continue studies after certification, motivation, and well-being. Moreover it was found that drop-outs are more likely than stay-ons to come from single parent families, have unemployed fathers, receive less educational support, have less conversation in Dutch with family members and have more siblings. Next to this their peers more often play truant than peers of stay-ons, while the peer group of drop-outs is somewhat more multicultural than that of stay-ons.

Following three theoretical frameworks it was predicted that ethnic minorities would drop out less in schools where pupils are empowered, where schools help pupils in making rational choices and where there were other minorities in school. Sixteen variables were tested for their main effects on drop out as well as for their interaction effect with ethnicity on drop out. Only three effects could be demonstrated and each effect was predicted by one of the three theories. According to empowerment theory the integration of language and culture into the regular curriculum, as well as stimulating community participation in school activities should lead to reduced likelihood of dropping out. Desintegration in terms of pull-out programs for ethnic minorities is indeed associated with increased drop-out. Rational choice theory lead to the hypothesis that counselling and postponing selection decisions should lead to less drop out. There was indeed an association between counselling and drop out. Finally contextual effects on drop out should occur according to social comparison theory. Indeed, ethnic minorities dropped out less than pupils of native origin in schools where there were relatively many ethnic minorities.

Although a pretest-posttest design was used, the partial corroboration of the hypotheses cannot of course lead to causal attribution, albeit because of the correlation between these

school variables and the scores of drop-outs on the pre-test. Nevertheless, it was demonstrated that schools differ in their drop-out rates, and that there are possible linking points for increasing the probability of pupils -ethnic minorities in particular- staying on. We will not dwell on the contextual effect, for which it is hard to see the policy implications, and the counselling effects, for which the reverse is true. The association between pull-out programs and drop out should at least raise doubts on the usefulness of these programs. They certainly do not meet the prerequisites that can be deduced from empowerment theory and, more practically, their detrimental effects can also be easily understood as stemming from ethnic minorities missing parts of the regular curriculum because of them.

REFERENCES

- Aitkin, M. & Zuzovsky, R. (1992). *Multilevel interaction models and their use in the analysis of large-scale school effectiveness studies*. Tel Aviv University and Australian National University.
- Appel, R. & Muysken, P. (1987). *Language contact and bilingualism*. London: Edward Arnold.
- Bosker, R.J. & Sniijders, T.A.B. (1990). Statistische aspecten van multi-niveau onderzoek. *Tijdschrift voor Onderwijsresearch*, 15, 317-330.
- Boudon, R. (1974). *Education, opportunity and social inequality*. New York/London: John Wiley.
- Boulot, J. & Boyzon-Fradet, D. (1987). *Les Immigrés et l'école: une course d'obstacles*. Paris: L'Harmathon/CEIMI.
- Brookover, W.B. & Lezotte, L.W. (1979). *Changes in school characteristics coincident with changes in pupil achievement*. East Lansing, Michigan, Institute for Research on Teaching, College of Education, Michigan State University.
- Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical linear models. Applications and data analysis methods*. Newbury Park/London/New Delhi: Sage publications.
- Campos, G.J. & Keatinge, H.R. (1988). *Language minority pupils experience: from theory to practice, to succes*. Philadelphia: Falmer.
- Centre for Educational Research and Innovation (1987). *Immigrants' children at school*. Paris: OECD.
- Centre for Educational Research and Innovation (1989). *One school, many cultures*. Paris: OECD.
- Chaikind, J. (1987). *College Enrollment Patterns of Black and White Pupils*. Washington D.C.: Decision Resources Corporation.
- Cummins, J. (1984). Heritage languages and Canadian school programs. In J.R. Mallea & J.C. Young (Eds.), *Cultural Diversity and Canadian Education*. (p. 477-500). Ottawa: Carleton University Press.
- Cummins, J. (1986). Empowering minority pupils: a framework for intervention. *Harvard Educational Review*, 56, 18-36.
- Cummins, J. (1990) Multilingual and multicultural education: Evaluations of underlying theoretical constructs and consequences for curriculum development. In P. Vedder (Ed.), *Fundamental studies in Educational Research*. Lisse, Swets & Zeitlinger.
- De Corte, E., Van der Velde, D. & Van Wijnsberge, M.J. (1982). Leermotivatie bij leerlingen in het Algemeen Vormend Secundair Onderwijs. *Pedagogische Studiën*, 59, 157-174.
- Education Ministry (1993). *Rapportering van de adviseurs tweede fase voortgezet onderwijs*. Utrecht.
- Fase, W. (1994). *Ethnic divisions in Western European Education*. Münster/New York: Waxman.
- Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, 7, 117-140.
- Hakuta, K. (1986). *Mirror of Language. The debate on bilingualism*. New York: Basic Books.
- Hofman, W.H.A. (1993). *Effectief onderwijs aan allochtone leerlingen*. Delft: Eburon.
- Hüttner, H.J.M. (1980). *Multilevel onderzoek: de operationalisatie van omgevingskenmerken*. (Ph.D. thesis). Nijmegen: KUN.
- Jaynes, G.D. & Williams, R.M. (1989). *A common destiny: Black and American Society*. Committee on the Status of Black American, Commission on Behavioural and Social Sciences and Education, Washington D.C., National Research Council, National Academy Press.
- Jong, M.J. de (1987). *Herkomst, kennis en kansen. Allochtone en autochtone leerlingen tijdens de overgang van basis- naar voortgezet onderwijs*. Dissertatie. Lisse: Swets & Zeitlinger.
- Jungbluth, P. (1985). *Verborgene differentiatie. Leerlingbeeld en onderwijsaanbod op de basisschool*. Nijmegen: ITS.

- Langen, A. van & Jungbluth, P. (1990). *Onderwijskansen van migranten. De rol van sociaal-economische en culturele factoren*. Amsterdam/Lisse: Swets & Zeitlinger.
- Meesters, M.J. (1992). *Loopbanen in het onderwijs en op de arbeidsmarkt*. Nijmegen: ITS/OoMO.
- Mortimore, P., Sammons, P., Stoll, P., Lewis, L. & Ecob, R. (1988). *School matters: the junior years*. Somerset: Open Books.
- Ogbu, J.U. (1978). *Minority Education and Caste: The American System in Cross Cultural Perspective*. New York: Academic Press.
- Smith, D.J. & Tomlinson, S. (1989). *The school effect: A study of multi-racial comprehensives*. London: Policy Studies Institutes.
- Snijders, T.A.B. (1991). *Some remarks on multilevel models for binary dependent variables*. Groningen: RUG, ICS/VSM.
- Stoel, W.G.R. (1983). *De relatie tussen de grootte van scholen voor voortgezet onderwijs en het welbevinden van leerlingen*. Groningen: RION.
- Tesser, P. Th.M. (1986). *Sociale herkomst en schoolloopbanen in het voortgezet onderwijs*. Dissertatie. Nijmegen: ITS.
- Tizard, J., Schofield, W.N. & Hewison, J. (1982). Collaboration between teachers and parents in assisting children's reading. *British Journal of Educational Psychology*, 52, 1-15.
- Vlug, I. (1985). *Schoolverzuim van Turkse en Marokkaanse meisjes*. Rotterdam: EUR.
- Vries, A.M. de (1992). *Hoe breder hoe beter? De effecten van heterogeen groeperen in het voortgezet onderwijs in Nederland*. Groningen: RION.
- Vries, G.C. de (1993). *Het pedagogisch regiem: groei en grenzen van de geschoolde samenleving*. Amsterdam: Meulenhoff.
- Wong, G.Y. & Mason, W.M. (1985). The hierarchical logistic regression model for multilevel analysis. *Journal of the American Statistical Association*, 80, 513-524.

Manuscript submitted: December 17, 1993

Accepted for publication: January 5, 1994

A Coupled Process of Conceptualizing a Model of School Effectiveness and Developing an Indicator System for Monitoring Effectiveness

Ruth Zuzovsky and

*School of Education, Tel Aviv University, Israel**

Murray Aitkin

Department of Mathematics, University of Western Australia, and

Department of Statistics, Tel Aviv University

ABSTRACT

In this article we present an indicator system which was developed in line with a theoretical approach aimed toward explaining schooling phenomenon. The study illustrates the coupled process of developing an indicator system and conceptualizing a model of school effectiveness.

Findings supported the claim that the modelling of school effectiveness studies must be both multi-level and interactive. Achievement is found to be dependent in a very sensitive, non-additive way, on the particular combination of pupil's home background, his or her general ability, the teaching style and other characteristics of the teacher, the class and the school context in which the pupil learns.

By adopting a general interactive model as the theoretical framework of an indicator system, the meaning of the latter dramatically changes. Indicators can no longer be regarded as simple measures of the "health of a system" since their effect is dependent on the values of other indicators that interact with them. The interpretation of their values is not straightforward – their effects are dependent on the values of other indicators. Nevertheless, indicators derived from an interactive model are better at describing the richness of the educational reality. The change in the specification of indicators representing the schooling phenomenon over years causes another difficulty in using indicators for monitoring purposes. It is becoming evident that changes in society which are reflected in schools' demographic features, and changes in the reality of life in schools themselves are, in turn, reflected in different school effectiveness models.

THEORETICAL FRAMEWORK

Foreword

Thirty years of research on school effectiveness have resulted in two distinct movements in the field of education: the school improvement movement and the indicator system movement.

The indicator system movement was in a sense a reaction to the growing number of school reforms which were initiated by the school improvement movement and stood in need of evaluation and monitoring. Tendencies toward school autonomy and accountability demands supported this need for monitoring and feedback mechanisms. Important reports on the failure of these educational reforms and on the declining achievements in the United States, and in other countries, stimulated and promoted the construction of indicator systems worldwide.

This growing interest in monitoring systems in education reflected explicitly a genuine need for comparative and diagnostic information but implicitly indicators and monitoring systems were sometimes used politically in order to exert control and to ensure schools meet some predetermined standards. Tied in with mistaken beliefs in causality, the indicators that were

developed were considered as "policy variables" which could be subject to administrative manipulation in order to improve the productivity of schools. "This is what makes indicators useful for locating problems and diagnosing what is wrong. They reflect the causal links we believe exist in the educational system" (Selden, 1990, p. 384).

Indicator systems regarded in this way represent a pragmatic approach which is oriented toward solving problems and toward fulfilling policymakers' needs. Even such pragmatic indicator systems are based implicitly on theoretical models but these models are fairly simple and address in the first place demands for practicality and feasibility.

Another approach to the development and use of indicator systems is more theoretical. It stems from a conceptual framework for school effectiveness which specifies the important variables of the schooling phenomenon, the relations between them and what is regarded in such frameworks as measures of effectiveness. Such a conceptual framework also dictates the statistical modelling of the schooling phenomenon, the analysis of data and interpretation of findings.

Indicator systems that are based on such an approach aimed not only toward pragmatic purposes but also at theoretical contributions. Usually, compared with pragmatic systems, these are comprehensive if not overloaded with variables. Thus, screening and selecting meaningful indicators is part of the process of developing such systems. Alternative sets of competitive indicators are tested for the best minimal explanatory combination. The result of this process is twofold: the development of the explanatory model that is most economical for monitoring purposes and a better understanding of the schooling phenomenon.

In the present paper we present an indicator system that was developed in line with the second approach, illustrating the coupled process of developing an indicator system and conceptualizing a model of school effectiveness.

The adoption of an initial theoretical framework

The starting point of the theoretical approach to developing an indicator system is adopting a suitable conceptual framework of school effectiveness.

There have been at least three distinct historical frameworks for school effectiveness research since the 1960's: The "input-output" framework, the "process-product" framework, and the "organizational" framework (Madaus, Airasian & Kellaghan, 1980; Miller, 1983, 1985; Cuttance, 1985; Scheerens & Creemers, 1989).

The most advanced and adequate among these is the organizational framework. It can be traced back to the dominant paradigm in sociology during the fifties and sixties, the structural/functional paradigm (Merton, 1957; Parsons, 1959), which emphasizes the importance of organizational and structural features of large systems in shaping their functioning. This paradigm also acknowledged the hierarchical nature of such systems. The failure to relate the analysis in previous studies to this hierarchical structure resulted, according to the advocates of the approach (Barr & Dreeben, 1983; Dreeben, 1983; Bidwell & Kasarda, 1975, 1980; Bidwell, 1983), in the small size of school effects.

Like other frameworks, the organizational paradigm defined effectiveness as the maximization of educational output, this time at each level of the educational organization and in relation to specific contexts. The range of explanatory variables was broadened to include organizational features as well as traditional input or output variables which were recognized as operating at different levels of the educational system.

The organizational framework enabled the definition of a representative set of indicators, established their relationship and enabled their categorization into: input, process, output, class contextual: (class composition) and out of school contextual: (societal, cultural, economic) variables. A schematic representation of this basic model is illustrated in Figure 1.

This scheme enabled us to sort almost 400 single or composite variables obtained from the second (1984) IEA study on science studies in Israeli elementary schools (Zuzovsky, 1987). A further selection of those variables that were found to be highly correlated with science achievement, and had a high internal homogeneity, resulted in 95 indicators. Their distribution, according to the conceptual scheme, is represented in Table 1.

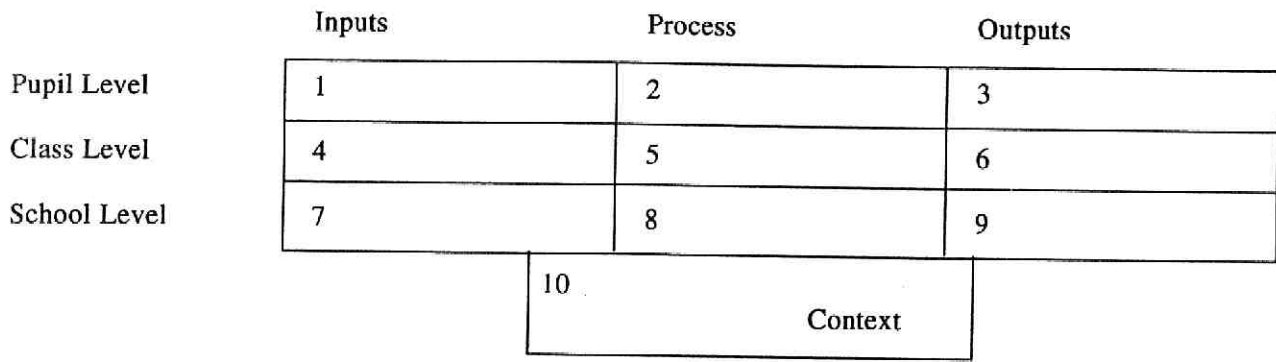


Fig. 1. Schematic representation of the conceptual model.

Table 1. Distribution of a selected set of educational indicators according to the conceptual model.

Level	Content	No. of Variables
Pupil	Pupils characteristics	4
	Home characteristics	1
	Pupils aptitude for the study of science	1
	Pupils perception of instruction	7
	Pupils achievement and attitudinal scores	4
Class	Student body composition	3
	Number of students in class	1
	Teachers experience and training	3
	Teachers preference of science objectives	3
	Teachers general characteristics	7
	Instructional variables (teachers report)	5
	Instructional variable (aggregated student report)	11
Mean class achievement and attitude scores	4	
School	School conditions	8
	Process at school level	26
	Outputs at school level	4
Context	Contextual variable operating at the school level	5

In this study, as in similar large-scale national and international studies, a large number of potential explanatory variables was measured at all levels. These variables described pupil and home characteristics, class composition (class means of pupil-level variables), class size, teachers' training and experience, their preferences toward science objectives, teachers' report of their instructional approach, and aggregated pupil perceptions of the way science was taught, school level variables describing the implementation of the science curriculum, physical and school conditions as perceived by teachers. The outcome variable reported here is percentage correct on the 45 "core" items of the science test. In later analyses, a new but similar test which included 33 items of the first one served also for measuring science outcomes. A reading comprehension test score was used as a proxy for pupil ability. More details of the data base for this study are given in Zuzovsky, Chen and Tamir (1987), and in Zuzovsky and Aitkin (1990, 1991).

This pool of variables (indicators) served an exploratory phase of model fitting and the delineation of the most explanatory set of indicators which represent the best model of school effectiveness.

METHOD

The model fitting exploratory phase

In model fitting procedures there are both logical and computational constraints. Given the sample size (1587 pupils from 69 classes – one class per school), only a small number of class/school variables could reasonably be included in one analysis. Since the data included a wide range of possible explanatory school or class indicators, the process of modelling at the school level had to be iterative with an initial selection of a relatively small number of school-level indicators which were screened so that redundant indicators were omitted and replaced with new ones until a final set of indicators remained.

In our study, three iterations led us to a final set of five pupil-level variables, three class-level aggregates of pupil-level variables, 11 class-level teacher variables and three school-level variables, a total of 17 upper-level variables.

The five pupil-level variables were:

SEX (1=male, 2= female);

RCOMP, a reading comprehension test score (effective range 50-100) used as a surrogate for ability;

GEN, the number of generations the pupil's family had lived in Israel (0-3);

ORIGIN, a dummy variable for Ashkenazi (0) or Sephardic (1) origin; and

SES, family socioeconomic status index scaled to have mean 0 and sd 1.0.

Three of these variables were aggregated to the class level, giving (mean) class context variables CRCOMP, CORIGIN and CSES.

Other variables at the class level were:

INSERV, a teacher input dummy variable taking the value of 1 if the teacher had participated more than 10 hours per year in in-service training, and zero otherwise;

TRAINING, a three-item composite index of the teacher's experience and preparation for science teaching (0-1);

REASON, a composite indicator of teacher's attitude measuring the extent to which the teacher prefers to foster reasoning in teaching science (1-3);

POSGROUP, a composite indicator of positive teacher attitudes towards group work (1-3);

LECTURE, the extent to which the teacher used a formal lecturing style in teaching science (1-4);

UNITS, the number of curriculum units studied in the class until the end of the fifth grade (1-16);

CLASSEXP, an aggregate pupil perception index of the extent to which pupils do experiments in class (1-3);

DIVERSITY, teachers' report of a diversified and rich mode of instruction in their class (1-4);

ATMOSPHERE, a composite indicator of pupil's perception of a liberal atmosphere in class (free movement of pupils and absence of punishment) (1-3);

LABTIME, time allocated to laboratory work according to teacher's report (1-4);

GROUPCHANGE, a dummy variable indicating whether teachers changed the composition of pupil groups, reflecting flexible expectations of pupil progress (1 = no, 2 = yes).

The remaining variables were dummies describing school conditions and processes:

- ANIMALC : whether the school had an animal corner (0=no, 1=yes);
 CURRICULUM : whether the new science curriculum had been introduced (0=no, 1=yes);
 CONDITIONS : teacher assessment of whether the teaching conditions in school were adequate (0=no, 1=yes).

It will be noted that many of these variables are simple binary indicators, or composites of indicators, of latent variables. For example, ANIMALC is an indicator of the seriousness with which the school regards science studies: the presence of an animal corner in a school does not by itself directly affect achievement – it is the positive attitude towards science studies represented by a value of 1 on this variable which is presumed to have the causal effect on achievement.

The statistical modelling of the conceptual framework and the shift from additive models to interactive ones

Related to the conceptual development of the organizational framework was the methodological development of the multilevel model for the analysis of hierarchical structures (Aitkin, Bennett & Hesketh, 1981; Aitkin, Anderson & Hinde, 1981). This development for the first time put the analysis of such studies on a sound methodological basis which was consistent with the theoretical assumptions.

Applying multilevel linear models for the analysis of hierarchical data enabled the simultaneous representation of outcome variation at each hierarchical level in one model, and the identification of the important variables “explaining” this variation at each level in terms of the magnitudes of their regression coefficients.

An important feature of the multilevel model is the ability to allow the effect of certain pupil level variables to vary randomly over classes or schools. This random variation in regression coefficients or slopes introduces yet another source of variation which can be represented or “explained” by appropriate “variables”, namely interactions between these pupil-level variables and class- or school-level variables.

In many reported studies, varying slopes of pupil-level variables have been identified and attracted attention to the possible existence of interactions, not only between pupil-level variables and school- or class-level variables, but also between variables at the same level. Of particular importance are interactions of pupil-level variables with each other, since their existence may affect the class-level regression coefficients and may change the whole specification of the model.

Supported by common sense and empirical evidence on the existence of interactions among individual traits and between individual traits and school treatment (Cronbach & Snow, 1977; Bloom, 1976) a very gradual shift toward interactive modelling of schooling phenomena has begun. Interactions have been used occasionally also in multilevel studies (Mason, Wong & Entwisle, 1983; Lee & Bryk, 1989) but were not recognized as essential in school-effectiveness studies. In the present study this approach to modelling schooling phenomena is completely embraced.

Briefly, we argue that the modelling of school effectiveness studies must be both multilevel and interactive in the most general sense. We argue that achievement cannot be considered the outcome of additive main effects of variables at each level of the hierarchy. It is not like the value added to each unit of a product manufactured from raw material which increases consistently through the manufacturing processes (as in the “value added” approach in the production function framework). Rather, achievement is a “product” of simultaneous processes which may be inconsistently effective over all types of children, and may depend in a very sensitive, non-additive way on the particular combination of the pupil’s home background, his or her general ability, the teaching style and other characteristics of the teacher, the class, and the school context in which the pupil learns. Statistically, this non-additive conditional dependence of achievement on combinations of variables is represented by interactions between the variables that appear in the regression model. It is a basic assumption of this framework that such

interactions occur routinely, and should be allowed for, in any routine analysis. If they are not included in the model, we can only learn about "global" or "average" features of variation in achievement. If such interactions are large, the interpretation of the global or average features of variables involved may be seriously misleading. We conjecture that the early claims of "no school effects" are due at least in part to the failure to incorporate interactions in the model representing school effects.

The logic of the statistical modelling

In this section we will briefly describe the logic of the statistical modelling procedure:

- 1) For any given outcome variable, a "null" variance component model with no explanatory variables is fitted, to provide estimates of the components of the total variation at each level. This step is important because the estimated variance components at each level provide an upper limit to the explanatory power of variables at each level.
- 2) Next a main effect model is fitted in which all explanatory variables at all levels are included. The explanatory variables are derived from a combination of educational theory and prior empirical evidence. Output from the analysis of this fitted model includes the estimated variance components s^2 , the regression coefficient estimate and standard error for each explanatory variable, and the "deviance" ($-2 \times$ the log maximized likelihood), a measure of "lack of support" of the model by the data: the larger the deviance, the poorer the agreement between model and data.

The reduction in the estimated variance components at each level after fitting the explanatory variables provides a measure of the explanatory power of the model at that level: specifically, at any level the ratio $s^2_{\text{model}} / s^2_{\text{null-model}}$ at that level is equal to $1 - R^2$ at that level.

- 3) The main effect model may then be reduced by omitting variables according to their explanatory power as deduced from their regression coefficients by some form of variable elimination on statistical grounds. In "backward elimination" for example (see e.g., Draper & Smith, 1981), it is the smallest "t" or "F" value. In models with large numbers of variables, the critical value for omission may be larger than the usual value of 2 (corresponding to a 5% level test). This process of elimination is continued until no redundant variables remain and a "minimal" model is obtained. The reduced model contains only those variables at each level whose effects justify interpretation, and it serves as a "baseline" model for the interpretation of "global" or "average" effects of the variables.

- 4) After the full main effect model is fitted, analysis proceeds directly to a general two-way interaction model, involving all two-way interactions between each pupil-level variable and all other pupil- and class- or school-level variables. This model should not include all two-way interactions between class-level variables, unless the number of classes is large enough to allow this number of interactions. If the number of classes is not large, some class \times class-level interactions may be included, but they will have to be specified a priori on strong theoretical grounds.

It may seem contradictory to restrict the number of class-level variables but to allow interactions between these variables and pupil-level variables. These "cross-level" interactions however have the properties of pupil-level, not class-level variables, and so can be included with no difficulty, given the usually large number of pupils in the study.

- 5) We then reduce this "full" interaction model by backward elimination in order of the smallest t-statistic in magnitude, until all remaining interactions (and main effects not involved in these interactions) have substantial t-values which may be as large as 2.5 instead of 2. This more conservative stopping rule is required to control the otherwise high Type I error rate (Aitkin & Zuzovsky (1994) give a discussion of this issue).

For multilevel analysis and especially for the large interaction models we used the large-model version of the VARCL computer program (Longford, 1987; Aitkin & Longford, 1986) which estimates by maximum likelihood the components of variance of the outcome variable at the different levels of the hierarchy, and allows the inclusion of up to 312 regression parameters in one model. This version enables us to look for different effects over schools of certain pupil

characteristics, and it also allows us to incorporate into the model all possible two-way interactions between pupil-level variables, and between pupil-level variables and class-level variables.

Results findings from the analysis of additive and interactive models – a comparison

A comparison between the reduced main effect model and the reduced interactive model supports the claim for general multilevel interaction models. The reduced models reflect the parsimonious selection of the most important indicators to be used for monitoring purposes according to the two modelling approaches. This comparison deals with both the explanatory power of the models as a whole and of each of their components separately. We will also look for differences in educational conclusions drawn from these two models.

Two comparisons will be presented, the first based on data collected in 1984, and the second based on data collected from the same schools eight years later using identical indicators.

Table 2 presents the comparison based on the 1984 data and Table 3 presents the comparisons from the 1992 data.

When comparing the explanatory power of the two models, similarity is evident. The explanatory power of the interaction model is slightly smaller, a consequence of the more conservative criterion for including variables in the model ($t > 2.5$ in the interaction model).

The conservatism required in the testing of a large number of interactions (larger t -values) led in the analysis of the interactive model to the omission of several variables included in the main effect models. Of considerable importance are the pupil level variables which are found to interact with class level indicators, thus changing the estimated regression coefficients of class level indicators and leading to a different interpretation of their importance.

In both sets, pupil ability is involved in many interactions with pupil socio-economic status, ethnic origin and the number of generations his/her family lives in Israel. Also, this indicator is involved with many class and school level indicators. Thus, this indicator is considered a central mediating indicator that determines the effects of others on the achievement of students. Both data sets provide interesting results concerning the impact of class composition.

The contextual effects of class socio-economic and ethnic composition which have a differential effect on students with a different ability were revealed in the analysis of the first data set. There are three reference values for individual SES which is a standardized measure (Low – 1, Medium 0, and High + 1) and three reference values for class SES (CSES) (Low – 0.5, Medium 0 and H + 0.5). These effects are derived from the following equation:

$$\text{ACHIEV} = \text{const} + 0.65 \text{ RCOMP} - 5.4 \text{ SES} + 14.2 \text{ CSES} + 0.10 \text{ SES.RCOMP} - 0.15 \text{ CSES.RCOMP}$$

Three sets of regression lines illustrate these combined effects. They are shown separately in Figure 2a and are combined in Figure 2b.

These regression lines have direct interpretation: For low RCOMP (50) mean achievement increases steadily with CSES but is unaffected by individual SES. For high RCOMP (100), achievement increases steadily with SES but is unaffected by CSES. Low ability pupils seem to benefit from being in high SES classes, while high ability pupils tend to be resistant to CSES effects and are more sensitive to their individual SES background.

These findings shed additional light on the debate about the power of the classroom composition contextual effect, an issue that has been extensively studied in Israel during the last 15 years. Three types of contextual effects at the class-level were explored: the ethnic, the socio-economic and the cognitive. Most studies showed consistent, positive and steady effects of heterogeneous composition on achievement of weak students. However a debate continues regarding the preferred type of heterogeneity. While early studies focused more on ethnic composition, later studies (Minkowich, Davis & Bashi, 1977; Klein & Eshel, 1980; Egozi, 1980) shifted to SES composition. Dar and Resh (1986, 1988) in a review of educational

Table 2. Comparison of Models 1984 Data.

Variance components	Reduced main effect model		Reduced general interaction model	
	Unexp. variance	% explained	Unexp. variance	% explained
Total	174	34	178	32
Pupils	149	25	148	26
Classes / Schools	25	62	30	53

Model Specification

	Reduced main effect model			Reduced general interaction model		
	Estimate	SE	t	Estimate	SE	t
<u>Pupil</u>						
SEX	-3.72	0.70	-5.7	-3.68	0.65	-5.7
RCOMP	0.46	0.024	18.8	0.645	0.075	
GEN	1.00	0.40	2.2			
ORIGIN	-1.85	0.70	-2.4	-1.97	0.75	-2.6
SES	2.02	0.40	5.3	-5.36	2.04	
<u>Class context</u>						
CSES				14.22	4.20	3.4
CORIGIN	-8.10	2.94	-2.8			
<u>Teacher / School</u>						
ATMOSPHERE	10.39	4.58	2.3			
GROUPCHANGE	3.72	1.60	2.3			
LECTURE				5.53	2.56	2.2
ANIMALC	4.33	1.49	2.9	5.29	1.58	3.3
<u>Pupil X Pupil</u>						
SES.RCOMP				0.100	0.027	3.7
<u>Pupil X Context</u>						
RCOMP.CSES				-0.145	0.053	-2.7
<u>Pupil X Teacher / School</u>						
RCOMP.LECTURE ...				-0.078	0.032	-2.5

Note: Empty t-value spaces correspond to variables involved in interactions.

Table 3. Comparison of Models 1992 Data.

Variance components	Reduced main effect model		Reduced general interaction model	
	Unexp. variance	% explained	Unexp. variance	% explained
Total	162.0	35.5	160.2	36.0
Pupils	134.0	31.4	134.5	31.0
Classes	5.6	27.0	0	0
Schools	22.4	53.3	25.7	46.5

Model Specification

	<u>Estimate</u>	<u>SE</u>	<u>t</u>	<u>Estimate</u>	<u>SE</u>	<u>t</u>
<u>Pupil</u>						
SEX	-4.62	0.43	-10.8	-0.95	1.16	
RCOMP	0.44	0.01	32.9	0.36	0.03	
GEN	0.59	0.27	2.2	-2.52	0.91	
ORIGIN	-1.23	0.47	-2.6	3.91	1.81	
SES	1.53	0.23	6.7	4.72	1.16	
<u>Class context</u>						
CSES				2.07	1.53	
CORIGIN	-4.56	2.25	-2.0	0.29	2.42	
<u>Teacher / School</u>						
UNITS	-0.35	0.17	-2.0	-0.40	0.15	-2.5
DIVERSITY	5.61	1.43	3.9	5.71	1.22	
LAB	-2.43	0.78	-3.1	-2.88	0.68	-4.2
CURRICULUM	5.05	1.34	3.8	0.09	2.3	
<u>Pupil X Pupil</u>						
RCOMP.GEN				0.047	0.013	3.6
RCOMP.ORIGIN				-0.074	0.026	-2.9
<u>Pupil X Context</u>						
SEX.CORIGIN				-8.43	2.53	-3.3
SEX.CSES				-5.26	1.45	-3.6
<u>Pupil X Teacher / School</u>						
RCOMP.CURRICULUM				0.072	0.027	2.7
SES.DIVERSITY				-1.22	0.43	-2.9

Note: Empty t-value spaces correspond to variables involved in interactions.

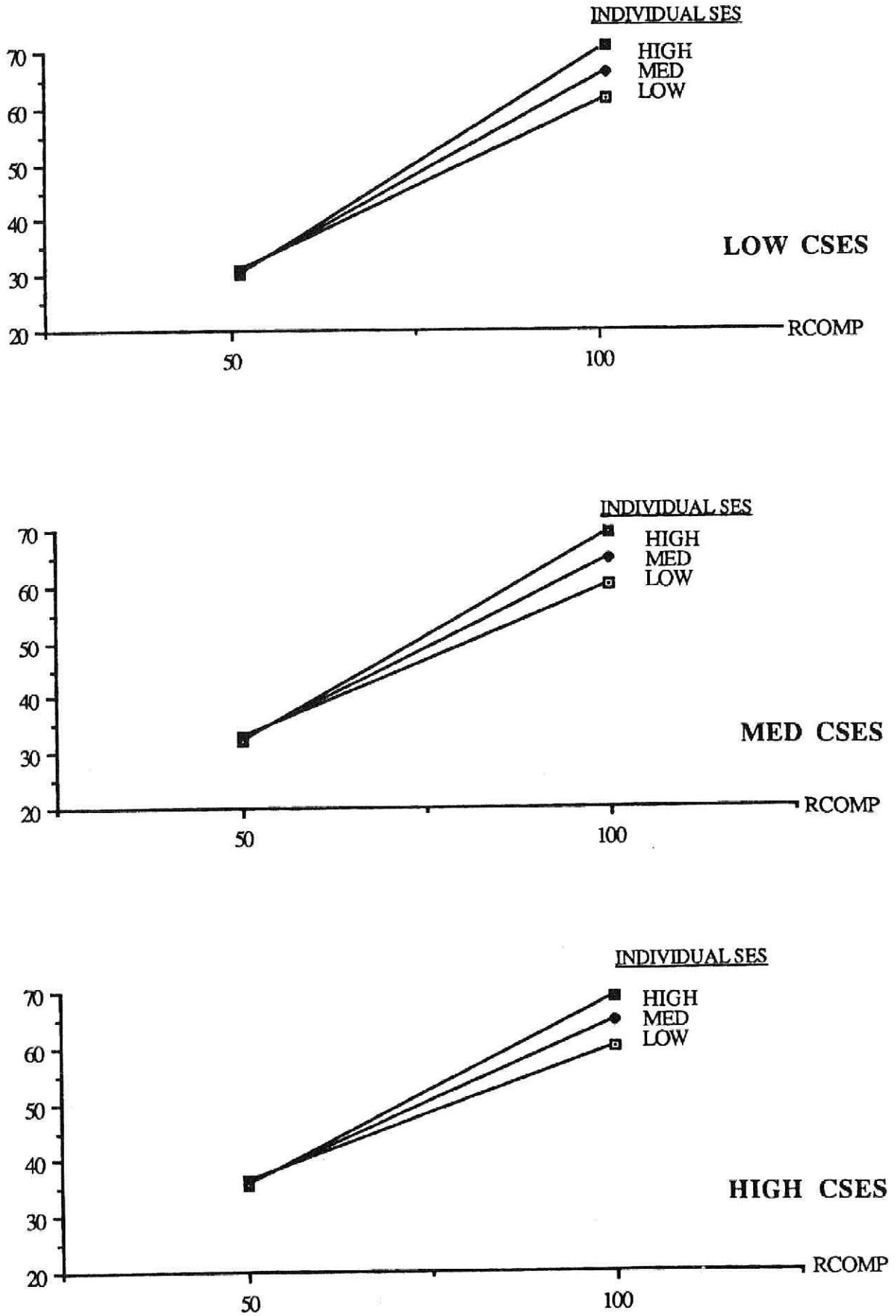


Fig. 2a. Interaction Between Individual SES and Reading Comprehension in Three Different Contexts of Class SES (CSES).

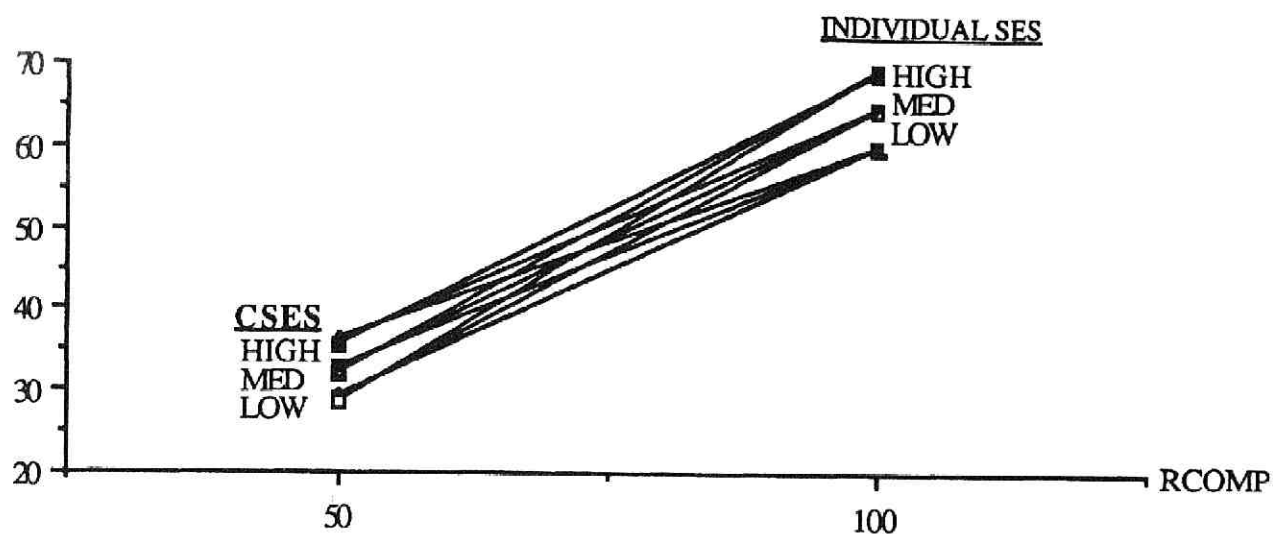


Fig. 2b. An Overlap of the Graphs in Fig. 2a.

integration conclude that it is the cognitive composition of the class that is the most effective factor interacting with individual characteristics and affecting achievement. On the basis of this study, we regard this conclusion as a result of a misspecified model and point back towards the important, differential role of socio-economic context on student achievement as mediated by their individual cognitive abilities.

Other differential effects of the class ethnic and socio-economic composition were found in the analyses of the 1992 data.

Class ethnic composition – CORIGIN is an aggregate of a dummy variable indicating Sephardic origin. This indicator ranges between 0 - 1. The cut off the score of high and low levels of CORIGIN was determined according to the median of the distribution.

Sex differences (girls = 0, boys = 1) are moderated by class SES and CORIGIN. The sex differences in favor of boys are: $0.95 + 8.43 \text{ CORIGIN} + 5.26 \text{ CSES}$.

Thus, girls suffer a greater relative disadvantage (-6.8 points) in classes of low CSES/high CORIGIN, than in high CSES/low CORIGIN (only - 3.6 points).

The different indicators specified in the interaction model and the existence of significant interactions as part of the model lead towards a different appreciation of the values of school indicators and to better understanding of the complexity of school effects.

Several school indicators such as the introduction of a new science curriculum, using a diversified mode of instruction or the use of more inquiry-oriented approaches in teaching science seem to benefit more high socio-economic status or high ability pupils.

In sum, though interaction models compared with main effect models explain little more or even slightly less of the between school variance, they supply a much more accurate, rich and sensitive picture of school effects and point toward conditional interpretation of the values of educational indicators that are involved in school effectiveness models.

Validation of an indicator system

The validity of a measurement tool is judged in the light of its intended use. The indicator system that was developed for monitoring purposes was meant to consist of reliable measures of relevant features of science education, sufficiently predicting or explaining the variability in science teaching outcomes and adequately representing a model of school effectiveness in science studies.

These demands were to a great extent met through the process of developing the indicator system.

Another demand is related to the use of an indicator system as an ongoing monitoring device, i.e., the indicator system should retain its relevancy and explanatory power over time.

Time as another source of variation in educational outcomes was recognized by Aitkin (1988) and Raudenbush (1989). Both researchers recognize the dynamic nature of school effects and suggest several statistical models for studying this phenomenon. They propose several research designs for measuring time effects. One of the designs, repeated measurement in same schools (containing new student samples) at different points of time was used in the study.

A repeated survey in the same school sample that served the development of the indicator system in 1984 was conducted after eight years. On the basis of the 1984 IEA instruments, reduced questionnaires consisting only of relevant items and two forms of the science achievement test with 33 anchor items identical to the 1984 ones were developed. The reliabilities of the two forms of this new test were satisfactory ($R=.89, .88$). These instruments served for the collection of data, the construction of the indicators and the standardized calculation of their values in a similar way to that for the 1984 survey. Twenty-one indicators, identical to the previous 22 with the omission of one on the ground of missing data, were constructed and analyzed in the same way the 1984 analysis was carried out.

The full sample contained 86 schools. Complete data in 1984 were obtained from 69 schools – 69 classes (one from each school) and from 1528 pupils. In 1992 complete data were obtained from 78 schools – 151 classes (two from each school, if possible) and 3216 pupils. The enlarged class sample was aimed toward representing the effects of class level variables that were confounded with school level variables in the previous study.

A comparison between the reduced interactive models of 1984 and 1992 can give an answer to the questions of the consistency of the indicator system. This comparison will deal with the consistency of the explanatory power of the models, their specification and their resulting educational implications. Table 4 presents this comparison.

The explanatory power of the 1992 interactive model is greater than that of the 1984 model.

With two classes in most schools, separate class and school variance components could be estimated. The class variance component which was found was very small, and consequently the effects of class-level variables were quite limited, though their interaction effects with pupil-level variables could be large. This is probably the reason for the increase in the explained portion of the between pupil variance in 1992 (since the interaction is a pupil-level variable). However, the 1992 model explains less of the between school variance and large differences among schools remained unexplained by the model which had only a few school-level variables. No random slope variation was found in pupil-level variables, over either classes or schools.

The model structures found in the two surveys are rather different. Figure 3 illustrates these differences.

Overlapping of circles or two-headed arrows represent interaction effects, positive and negative signs represent their direction. While all the pupil level variables and reading comprehension in particular, were important in both surveys, the class and school-level variables changed completely between the two surveys, presumably reflecting changed educational practices in schools over this period. The reappearance of GEN – number of generation the pupil's family living in Israel – is undoubtedly the effect of recent immigration from the USSR and Ethiopia to Israel. All pupil-level indicators are interacting with other variables. In cases where they interact with class or school indicators, those latter are conditioned by them.

As in 1984, student ability is the most important determinant of achievement. It interacts with other pupil-level indicators and class-level indicators, thus serving as a mediating variable.

While in 1984 it was the effect of only the individual SES that was conditioned by student ability, in 1992 it is the effect of the number of generations the pupil's family had been in Israel (GEN) and his/her ethnic origin (ORIGIN) that are mediated by ability.

Table 4. Comparison Between Reduced Two-Way Interaction Models of 1984 and 1992 Data.

Variance Components	1984			1992		
	unexp. variance	% exp.		unexp. variance	% exp.	
Total	178	32		161	36	
Pupil	148	26		135	31	
School / Class	30	53		26	47	
Model Specification	<u>Estimate</u>	<u>SE</u>	<u>t</u>	<u>Estimate</u>	<u>SE</u>	<u>t</u>
Pupil						
SEX	-3.68	0.65	-5.69	-0.95	1.16	-
RCOMP	0.63	0.075	-	0.36	0.031	-
ORIGIN	-1.97	0.75	-2.64	3.91	1.81	-
SES	-5.36	2.04	-	4.72	1.16	-
GEN				-2.52	0.91	-
Context						
CSES	14.22	4.19	-	2.07	1.53	-
CORIGIN				0.29	2.42	-
Teacher / Class						
LECTUR	5.53	2.56	-			
DIVERSITY				5.71	1.22	-
LABTIME				-2.88	0.68	-4.24
UNITS				-0.40	0.15	-2.59
School						
ANIMALC	5.29	1.58	3.34			
CURRICULUM				0.09	2.30	-
Pupil X Pupil						
RCOMP X SES	0.10	0.03	3.67			
RCOMP.GEN				0.047	0.013	3.58
RCOMP.ORIGIN				-0.074	0.026	-2.89
Pupil X Context						
RCOMP.CSES	-0.145	0.05	-2.75			
SEX.CSES				-5.26	1.45	-3.64
SEX.CORIGIN				-8.43	2.53	-3.33
Pupil X Teacher						
LECTUR.RCOMP	0.078	0.032	2.48			
RCOMP.CURRICULUM				-0.072	0.027	2.66
SES.DIVERSITY				-1.22	0.43	-2.86
Intercept	18.31	0.73		18.29	0.61	

Note: Empty t-value spaces correspond to variables involved in interactions.

Teachers' and school indicators' effects were very limited in 1984; in 1992 there are more variables at these levels that affect science achievement. Some of them have only main effects and some of them have interactions with pupil-level indicators.

In 1984 these were the class SES composition and teachers' frequent lecturing that were both mediated by pupils' ability. In 1992 pupils' ability continued to serve as mediating variables, this time mediating the introduction of the new science curriculum to schools (CURRICULUM).

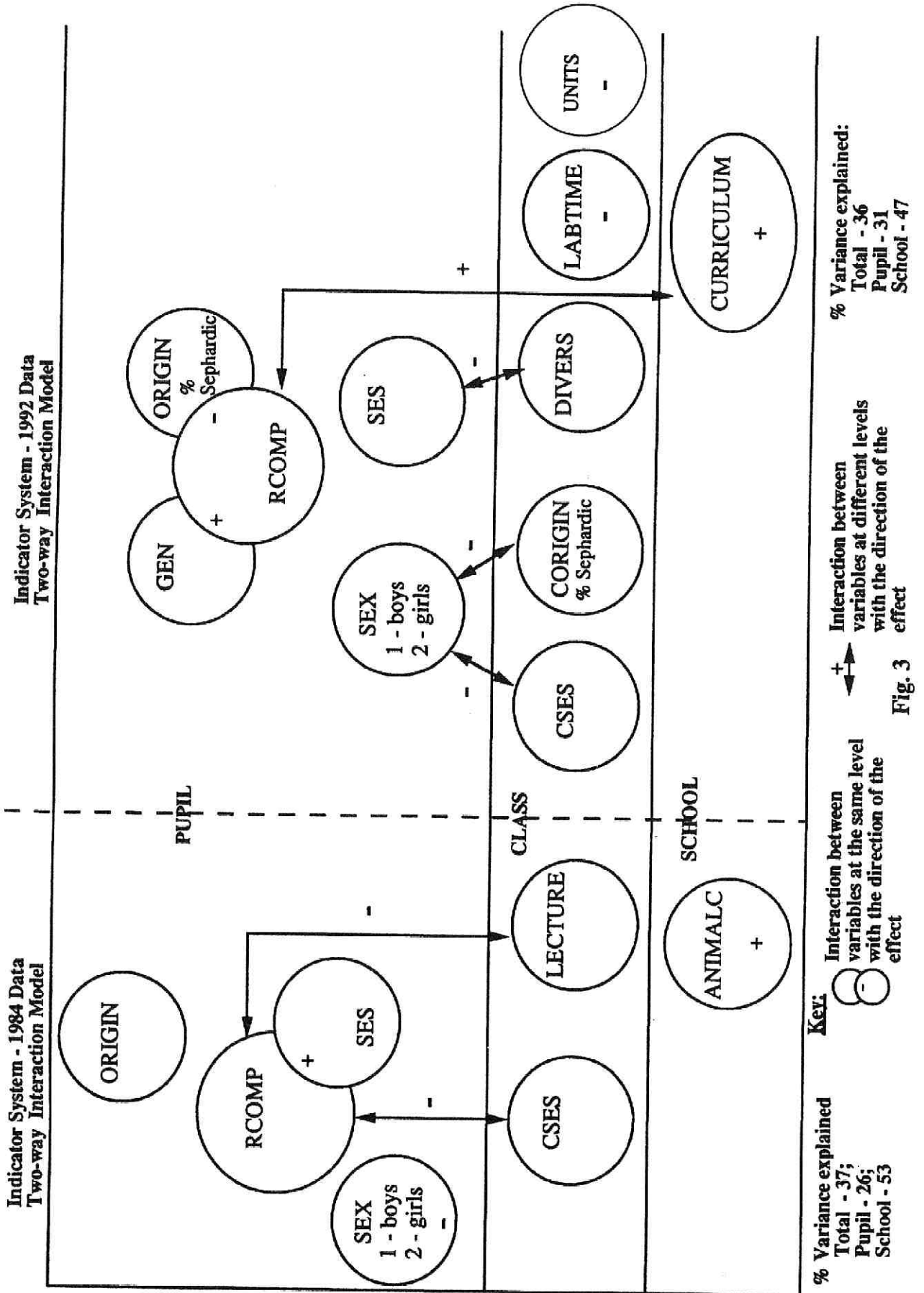


Fig. 3.

Yet pupils' SES also plays a role and mediates another class-level variable – diversified mode of instruction (DIVERSITY). Some of the negative results of new class and school indicators (LABTIME and UNITS) that appear in 1992 are difficult to explain but unquestionably real. The new curriculum clearly benefitted achievement, as did the existence of an animal corner in 1984.

CONCLUSIONS

The development of an indicator system was a combined effort of conceptualizing a model of school effectiveness, reducing it to its essential elements, testing its validity, and learning how to interpret data obtained in monitoring activities.

By adopting a general interactive model as the theoretical framework of an indicator system, the meaning of the latter dramatically changes. Indicators can no longer be regarded as simple measures of the "health of a system" since their effect is dependent on the values of other indicators that interact with them. The interpretation of their values is not straightforward – their effects are dependent on the values of other indicators. Nevertheless, indicators derived from an interactive model are better at describing the richness of the educational reality.

The change in the specification of indicators representing the school phenomenon over years causes another difficulty in using indicators for monitoring purposes.

It is becoming evident that changes in society which are reflected in schools' demographic features, and changes in the reality of life in schools themselves, are, in turn, reflected in different school effectiveness models.

The main conclusions of this study are the following:

1. The use of an indicator system for monitoring purposes is dependent on the inclusion of all necessary interactions in which these indicators are involved. The meaningful interpretation of the values of the indicators measured may be dependent on the values of the other indicators.

If such interactions are not included in the model, only 'global' or 'average' effects are estimated and they may mask or distort the real effects.

2. An indicator system tends to change over time. Part of an ongoing monitoring effort should be the updating of conceptual schemes that underlie indicator systems. Structural changes in the model can be identified by repeated measurement over a not more than two years interval. Larger intervals would not allow the tracing of dynamic regular changes.

3. Structural changes in the indicator system over short periods may be represented by a longitudinal variance component model using time as another level (Raudenbush, 1989).

No new methodological problems arise. In the present study the eight-year period is so long that major changes in the model occurred.

4. Indicator systems should reflect not only the hierarchical structural nature of the educational system, but also its dynamic nature and evolution over time. The process of delineating models of school effectiveness that represent the functioning of schools at different points in time should be regarded as a monitoring activity which is essential for the understanding and interpreting of information on important features of the system.

REFERENCES

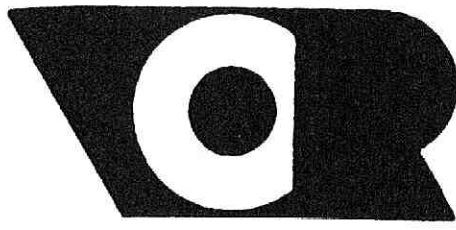
- Aitkin, M. (1988). *Multilevel models for educational systems*. Paper presented at the Educational Testing Service, Princeton, New Jersey.
- Aitkin, M.A., Anderson, D. & Hinde, J. (1981). Statistical modelling of data on teaching styles (with discussion). *Journal of the Royal Statistical Society, Series A*, 144, 148-161.
- Aitkin, M.A., Bennett, N.S. & Hesketh, J. (1981). Teaching styles and pupil progress: A reanalysis. *British Journal of Educational Psychology*, 51, 170-186.

- Aitkin, M.A. & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *The Journal of the Royal Statistical Society, Series A*, 149, 1-43.
- Aitkin, M.A. & Zuzovsky, R. (1994). Multilevel interaction models and their use in the analysis of large scale school-effectiveness studies. *School Effectiveness and School Improvement*, 5, 1.
- Barr, R. & Dreeben, R. (1983). *How schools work: A study of reading instruction*. University of Chicago.
- Bidwell, C.E. (1983). *Discussion of papers' symposium on school effects research*. Annual Meeting of the American Educational Research Association, Montreal.
- Bidwell, C.E. & Kasarda, J.D. (1975). School district organization and students achievement. *American Sociological Review*, 40, 55-70.
- Bidwell, C.E. & Kasarda, J.D. (1980). Conceptualizing and measuring the effects of school and schooling. *American Journal of Education*, 88, 401-430.
- Bloom, B.S. (1976). *Human characteristics and school learning*. New York: McGraw Hill.
- Cronbach, L.J. & Snow, R.E. (1977). *Aptitudes and instructional methods*. New York: Irvington Publishers, Inc.
- Cuttance, P. (1985). Frameworks for research on the effectiveness of schooling. In D. Reynolds (Ed.), *Studying school effectiveness*. London: The Falmer Press.
- Dar, Y. & Resh, N. (1986). *Classroom, composition and pupil achievement*. London: Gordon & Breach.
- Dar, Y. & Resh, N. (1988). Educational integration and pupil achievement. Conclusion and evaluation of research in Israel. *Megamot*, 31, 180-207. (Hebrew)
- Draper, N. & Smith H. (1981). *Applied Regression Analysis (2nd edition)*. New York: Wiley.
- Dreeben, R. (1983). *School production and school effects*. Paper presented at a symposium on school effects research. Annual Meeting of the American Educational Research Association, Montreal.
- Egozi, M. (1980). *The effect of class social composition on achievement of studies from different social classes*. Jerusalem: Ministry of Education.
- Klein, Z. & Eshel, Y. (1981). Integration and fostering educational aims in Israel. *Megamot*, 26, 271-282. (Hebrew)
- Lee, V. & Bryk, A.S. (1989). A multilevel model of the social distribution of high school achievement. *Sociology of Education*, 62, 172-192.
- Longford, N.T. (1987). A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested random effects. *Biometrika*, 74, 817-827.
- Madaus, G.F., Airasian, P.W. & Kellaghan, T. (1980). *School effectiveness: A reassessment of the evidence*. New York: McGraw Hill.
- Mason, W.M., Wong, G.Y. & Entwisle, B. (1984). Contextual analysis through the multilevel linear model. In S. Leinhardt (Ed.), *Sociological Methodology (pp. 72-103)*. San Francisco: Jossey-Bass.
- Merton, R.K. (1957). *Social theory and social structure (Rev. Ed.)*. Glencoe Ill.: The Free Press.
- Miller, S.K. (1983). *The history of effective schools research: A critical review*. Paper presented at a symposium on research on effective schools, state of art. Annual Meeting of the American Educational Research Association, Montreal. (Eric Document Reproduction Service No. Ed. 231, 818).
- Miller, S.A. (1985). Research on exemplary schools: An historical perspective. In G. Austin & H. Garber (Eds.), *Research on Exemplary Schools (pp. 3-30)*. Academic Press.
- Minkowich, A., Davis, D. & Bashi, J. (1977). *An evaluation study of Israeli elementary schools*. Jerusalem: The Hebrew University of Jerusalem, School of Education. (Hebrew)
- Parsons, T. (1959). The school class as a social system: Some of its functions in American society. *Harvard Educational Review*, 29, 297-381.
- Raudenbush, S.W. (1989). The analysis of longitudinal, multilevel data. *International Journal of Educational Research*, 13, 721-739.
- Scheerens, J. & Creemers, B.P.M. (1989). Conceptualizing school effectiveness. *International Journal of Educational Research*, 13, 691-706.
- Selden, R. (1990). Developing Educational Indicators. A State National Perspective. *International Journal of Educational Research*, 14, 4, 383-393.
- Zuzovsky, R. (1987). *The elementary school in Israel and science achievements*. Ph.D. dissertation. Jerusalem: Hebrew University. (Hebrew)
- Zuzovsky, R. & Aitkin, M. (1990). Using a multi-level model and an indicator system in science education to assess the effect of school treatment on student achievement. *School Effectiveness and School Improvement*, 1, 2, 121-138.
- Zuzovsky, R. & Aitkin, M. (1991). Curricular change and science achievement in Israeli elementary schools. In S.W. Raudenbush & J.D. Willms (Eds.), *Schools, classrooms and pupils (pp. 25-36)*. San Diego: Academic Press.

Zuzovsky, R., Chen, D. & Tamir, P. (1987). *Science teaching in Israel in the eighties: Instruction, learning, achievement and attitudes*. Jerusalem: The Amos de Shalit Israeli Science Teaching Center. (Hebrew)

Manuscript submitted: December 16, 1993

Accepted for publication: January 10, 1994



BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie-adres
Dr. P.J.J. Stijnen, Open Universiteit,
Postbus 2960, 6401 DL Heerlen,
tel. 045-762293

Jrg. 18, nr. 1

1994

Inhoudsopgave

Verslag VOR-ledenvergadering

83

Verslag VOR-ledenvergadering 16 december 1993: 13.30-14.30 uur in het Stiltecentrum te Utrecht

Aanwezig: leden: J. Thijssen, H. Wijffels

bestuursleden: N. Lagerweij (voorzitter), F.J.J. van Bussel (penningmeester), J. Dronkers, W.Th.J.G. Hoeben, J.M. Pieters, P.J.J. Stijnen (secretaris)

1. Opening

De voorzitter opent de vergadering om 13.40 uur. Bericht van verhindering is ontvangen van: J. Ax, J. Claessen, M. Mulder, H. van Hout, L. de Klerk, P. van den Broek, J. Demand, J. Slavenburg, J. Giesbers, F. van Vught, L. van der Kamp, J. Scheerens.

2. Mededelingen

De voorzitter deelt mede dat er het laatste jaar een aanzienlijke uitbreiding van het ledenbestand heeft plaatsgevonden

- Er komt een statutenwijziging in die zin dat de algemene ledenvergadering in de toekomst 1 keer per jaar zal plaatsvinden, tijdens de ORD (zie ook de geringe opkomst van leden vandaag). In januari zal het DB spreken over doelstellingen en statuten van de vereniging. Het resultaat wordt voorgelegd aan het VOR-bestuur en later aan de leden (volgende algemene ledenvergadering is op woensdag 25 mei 1994 tijdens de ORD in Utrecht)
- In overleg met de TOR-redactie is besloten TOR in het vervolg 4 keer per jaar te laten verschijnen. Het aantal pagina's blijft gelijk
- Namens de VOR heeft de voorzitter, de heer Lagerweij, in oktober deelgenomen aan een tweedaagse conferentie in Aarau (Zwitserland). Daar is besloten tot oprichting van een 'European Educational Research Association'. De statuten voor dit samenwerkingsverband komen waarschijnlijk in juni 1994 tot een afronding. Een verslag van de conferentie zal verschijnen in het VOR-bulletin (achter in TOR)
- Blijkens schriftelijke mededeling van de heer Scheerens zal het bestuur van de divisie Methodologie en Evaluatie in februari/maart 1994 een ledenvergadering organiseren. Het inhoudelijk thema is waarschijnlijk 'basisonderwijs'.

3. Verslag van de VOR-ledenvergadering van 27 mei 1993

Het verslag wordt onder dankzegging aan de secretaris ongewijzigd vastgesteld.

4. Begroting 1994

De begroting wordt goedgekeurd. Omdat er soms personen zijn die wel lid zijn van een divisie maar niet van de VOR (hetgeen eigenlijk niet mogelijk is) wordt afgesproken dat de secretaris, de heer Stijnen, hierover een brief zal sturen aan alle secretarissen van de divisies. Hij zal hen verzoeken in het vervolg bij nieuwe inschrijvingen alleen gebruik te maken van het inschrijfformulier dat opgenomen is in de folder van de VOR. Indien dit formulier goed wordt ingevuld is het probleem uit de wereld.

5. Ere-lidmaatschappen te verlenen tijdens de ORD'94

Er worden enkele namen genoemd van personen die in aanmerking komen voor een ere-lidmaatschap. De heren Pieters en Hoeben zullen dit verder afwerken.

6. Wat verder ter tafel komt

- Gevraagd wordt door de heer Thijssen of er in 1994 afzonderlijke onderzoeksvisitatiecommissies komen voor onderwijskunde en pedagogiek. De voorzitter beaamt dit.

- Er wordt kort van gedachten gewisseld over de mogelijkheid bij diverse instellingen een propaedeuseverklaring onderwijskunde te verstrekken
- Er komt een themanummer van het blad 'Opleiding en Ontwikkeling' over omscholing. De aanwezigen wordt gevraagd suggesties te doen voor namen en voor voorbeelden van omscholing. Suggesties doen toekomen aan de heer Thijssen.

7. Rondvraag

Hiervan wordt geen gebruik gemaakt.

8. Sluiting

Niets meer aan de orde zijnde sluit de voorzitter onder dankzegging om 14.30 uur de vergadering.

N.B.: De volgende Algemene Ledenvergadering vindt plaats op woensdag 25 mei 1994 tijdens de ORD in Utrecht.

Taakkenmerken die bijdragen aan de moeilijkheid van functionele leestaken¹

Wendy C. de Geus en Pieter Reitsma*
Paedologisch Instituut Amsterdam/Duivendrecht

ABSTRACT

In everyday life people encounter a variety of reading materials, such as recipes, forms, indexes, manuals, etc. This study tries to identify critical components of performance on these so-called functional reading tasks or document literacy tasks. The data of a nation-wide literacy survey were used to explain the difficulty of document literacy items by formal aspects of the task. Two sets of variables were used as predictors of item difficulty in a multiple regression analysis. One set was taken from a study by Kirsch and Mosenthal (1990) and was based on a parsing of the document by using a relational grammar. The alternative set was designed to tap the same sources of variance, but without the time-consuming propositional analysis. Both sets appeared to be equally good in explaining variance in item difficulty, but the specific item set seems to limit the amount of variance that can be accounted for by formal aspects of the task.

INLEIDING

Men kan in het dagelijks leven allerlei soorten leesmaterialen tegenkomen. Denk bijvoorbeeld aan de handleiding bij de videorecorder, het formulier voor het terugkrijgen van een deel van reeds betaalde belasting, het recept voor het kerstdiner, de bijsluiter bij een geneesmiddel of de gebruiksaanwijzing op een tube lijm. Men spreekt hierbij ook wel van functionele leestaken. De mate waarin iemand zich in het dagelijks leven kan redden voor wat betreft het uitvoeren van allerhande voor hem of haar van belang zijnde leestaken, wordt zijn of haar functionele leesvaardigheid genoemd.

Mensen verschillen in de mate waarin ze functioneel leesvaardig zijn en afhankelijk van de moeilijkheidsgraad van de taak zal een wisselend percentage van de bevolking een taak met succes kunnen volbrengen. In de hier gerapporteerde studie wordt de vraag gesteld waarom de ene functionele leestaak moeilijker is dan de andere. Dit zal vermoedelijk vooral afhangen van de cognitieve operaties die nodig zijn om de leestaak uit te voeren. Er kunnen op basis van taakanalyse variabelen geïdentificeerd worden die gerelateerd zijn aan deze cognitieve operaties. Deze variabelen kunnen vervolgens empirisch onderzocht worden als predictoren van de prestatie op functionele leestaken. In het huidige onderzoek wordt als invalshoek gekozen voor een formele analyse van aspecten die de moeilijkheid van functionele taken bepalen, dat wil zeggen dat gepoogd wordt de relatieve moeilijkheid van een set verschillende functionele leestaken te verklaren met een aantal variabelen, die zijn gebaseerd op onderscheiden kenmerken van het leesmateriaal en de bijbehorende vragen. Meer in het bijzonder wordt uitgegaan van een studie van Kirsch en Mosenthal (1990). Zij verklaarden met slechts vijf variabelen 89% variantie in de moeilijkheid van 61 "document literacy"-items, afkomstig uit de Survey of Young Adult Literacy (Kirsch & Jungeblut, 1986). Deze taken waren tamelijk gevarieerd, maar door gebruik te maken van een propositionele analyse die speciaal ontworpen is voor documenten, konden Kirsch en Mosenthal de verschillende taken toch onderling vergelijkbaar maken. Het onderzoek dat Kirsch en Mosenthal deden, is vooral belangrijk omdat ze aantoonde dat over een groot aantal qua formaat verschillende documenten dezelfde kenmerken van materiaal en items gebruikt kunnen worden om de moeilijkheid van de items met succes te voorspellen.

* Adres: Paedologisch Instituut, Postbus 303, 1115 ZG Duivendrecht.

Het doel van het huidige onderzoek is na te gaan of de methode van Kirsch en Mosenthal ook te gebruiken is bij een set Nederlandse functionele leestaken. Als ook de moeilijkheid van Nederlandse taken op dezelfde manier goed beschreven kan worden met slechts vijf variabelen dan is dat een aantrekkelijke reductie van het aantal potentiële parameters van de prestatie op dergelijke alledaagse leestaken. De variabelen kunnen dan vervolgens gebruikt worden om andere, soortgelijke taken te beschrijven en met elkaar te vergelijken en om de prestatie van groepen personen op deze taken te voorspellen. Behalve dit replicatie- of kruisvalidatie-onderzoek is er echter ook een alternatieve predictorset gemaakt. Zoals namelijk in een volgende paragraaf zal blijken, is het uitzetten van tekst in proposities tamelijk tijdrovend werk en is het bovendien over het algemeen "quite complex and far from objective" (Rayner & Pollatsek, 1989, p.292). Naar aanleiding van deze opmerking van Rayner en Pollatsek is het van belang na te gaan of de beoordeling op betrouwbare wijze kan geschieden. Tevens kan de vraag worden gesteld of met eenvoudiger middelen, namelijk door het gebruik van beoordelingsschalen, niet eenzelfde predictie-resultaat behaald kan worden. Voor dit doel werden alternatieve variabelen geselecteerd die dezelfde variantiebronnen dekken als de aan de relationele grammatica ontleende variabelen, maar waarvoor geen propositionele analyse van het materiaal vereist is.

De relationele grammatica van Kirsch en Mosenthal

Mosenthal en Kirsch (1991) ontwierpen een model voor de cognitieve verwerking van leestaken waarbij informatie uit documenten opgezocht of gebruikt moet worden. Volgens het model bestaat het zoekproces uit vijf stappen: Eerst formuleert de gebruiker een doel om het document te gebruiken. In een testsituatie wordt het doel doorgaans bepaald door de vraag of opdracht die gegeven wordt. Als tweede stap wordt de gegeven informatie en de gevraagde informatie uit de vraag of opdracht bepaald. Als stap 3 wordt vervolgens de gegeven informatie als zoekcode gebruikt: men doorzoekt het document tot de in de vraag gegeven informatie gevonden wordt. Daar zoekt men als stap 4 een overeenkomst tussen informatie die in het document staat en de informatie die in de vraag of opdracht gevraagd wordt en bepaalt men als vijfde stap of deze overeenkomst voldoet aan het criterium van de vraag. Om dit model te relateren aan leestaken, ontwikkelden Kirsch en Mosenthal (1990) een systeem om de structuur en inhoud van materiaal en opdracht van document-taken te beschrijven: een relationele grammatica. Aangezien ook in het huidige onderzoek gebruikt gemaakt zal worden van deze grammatica, wordt beknopt enige toelichting gegeven.

De grammatica begint met de aanname dat in een document een hiërarchie van taal-, beeld- of getals-informatie gerepresenteerd wordt. Deze hiërarchie bestaat uit drie niveaus: (1) *het betekeniskenmerk* (2) *de eenheid van specifieke informatie* (SPE), (3) *de organiserende categorie* (OC). Het niveau van het betekeniskenmerk is het laagste niveau in de hiërarchie en wordt gevormd door argumenten en relationele termen. Argumenten zijn meestal zelfstandig naamwoorden. Bijvoorbeeld, in de zin "Dosering: volwassenen - elke 4 uur 2 theelepels" zijn dosering, volwassenen en theelepels argumenten. Relationele termen beschrijven de argumenten en de relaties tussen argumenten; het zijn meestal werkwoorden, bijvoeglijk naamwoorden, bijwoorden, enz. Bijvoorbeeld, het getal 2 staat in een attribuut-relatie (ATT) met het zelfstandig naamwoord theelepels. De afgeleide persoonsvorm in deze zin ("neem") relateert volwassenen als een agens (AG) aan theelepels als een object (OBJ).

Op het volgende niveau vormen de argumenten, relationele termen en werkwoorden samen één eenheid van specifieke informatie (SPE). Elke SPE vormt een micro-propositie en SPE's kunnen ook gebruikt worden om een andere SPE of een direct voorafgaande organiserende categorie (OC) te verduidelijken.

Een organiserende categorie (OC), de hoogste analyse-eenheid, bestaat uit een algemene term of categorie die enkele SPE's samenvat. De SPE's die samen een OC uitmaken delen één of meer gelijke betekeniskenmerken die door de OC gerepresenteerd worden. Stel dat het eerder gegeven voorbeeld een zin is uit een medicijnbijsluiter, waarin ook nog doseringsvoorschriften voor kinderen opgenomen zijn, dan is "Dosering" in bovengenoemde voorbeeldzin een OC.

De voorbeeldzin uiteengezet volgens de relationele grammatica ziet er dan als volgt uit:

```

\OC Doscring
  \SPE *ncem
    \AG volwassenen
    \OBJ theelepels
      \ATT 2
    \TEMP uur
      \ATT 4
      \ATT elke

```

Dergelijke uiteenzettingen in organiserende categorieën, eenheden van specifieke informatie en betekeniskenmerken geven samen de hiërarchie van de informatie in documenten weer. Uit de onderlinge relaties tussen de OC's en SPE's kan de macrostructuur van het document worden afgeleid.

Kirsch en Mosenthal (1990) analyseerden eerst de structuur van de NAEP-documenten en van de bijbehorende 61 vragen en opdrachten (items) door ze volgens de bovenbeschreven relationele grammatica uiteen te zetten. In termen van de SPE's en OC's onderscheidde ze vervolgens twee typen variabelen die aan de moeilijkheid van de items gerelateerd zouden zijn, namelijk documentvariabelen en taakvariabelen. De hoeveelheid informatie en de complexiteit van het document werd vastgesteld met zes *document variabelen*. In het document werden geteld: (1) *het aantal OC's*, (2) *het aantal ingebedde OC's*, (3) *het diepste niveau van inbedding van een OC*, (4) *het aantal SPE's*, (5) *het aantal ingebedde SPE's*, (6) *het diepste niveau van inbedding van een SPE*. Alle items (vragen of opdrachten) bij één bepaald document krijgen dezelfde scores op deze materiaalvariabelen

De structurele relaties tussen de vraag en het document werden vastgesteld met vier *taakvariabelen*, te weten: (1) *het aantal OC's*, (2) *het diepste niveau van inbedding van een OC*, (3) *het aantal SPE's* en (4) *het diepste niveau van inbedding van een SPE* dat verwerkt moet worden voor de taak. Het gaat bij deze variabelen om de aard en de hoeveelheid informatie uit het materiaal die voor het vinden van het juiste antwoord verwerkt moet worden.

Ook werden drie *procesvariabelen* onderscheiden. Deze zijn gebaseerd op de cognitieve processen die volgens het hierboven besproken model verlopen bij het opzoeken van informatie in documenten:

(1) *mate van overeenkomst* (letterlijk, synoniem, eenvoudige afleiding nodig, complexe afleiding nodig, voorkennis nodig) tussen de in de vraag *gegeven* informatie en de corresponderende informatie in het document;

(2) *type informatie*: dit is een maat voor de verwerking die nodig is voor het localiseren, identificeren, genereren van de *gevraagde* informatie. Het is een samengestelde maat waarin tot uitdrukking komt hoe letterlijk het antwoord in de tekst staat, hoeveel verschillende informatie-eenheden betrokken zijn bij het antwoord en hoe verspreid de benodigde informatie in het document aanwezig is;

(3) *plausibiliteit van distractors*: hierin komt tot uitdrukking hoeveel informatie in het document, behalve het goede antwoord, overeenkomende kenmerken heeft met de informatie die in de vraag op opdracht gegeven en gevraagd wordt en hoe dicht bij het goede antwoord in de informatie-hiërarchie deze distractors zich bevinden.

De procesvariabelen werden op een vijfpuntsschaal gescoord. In totaal zijn zo dertien variabelen onderscheiden waarvan verwacht werd dat ze met de moeilijkheid van de items samen zouden hangen. De moeilijkheid van de 61 "document literacy" items uit de Young Adult Literacy Study was bepaald door de p-waarde. In een multiple regressieanalyse met de p-waarde als criteriumvariabele bleek maar liefst 89% van de variantie in de itemmoeilijkheid verklaard te kunnen worden door vijf van deze dertien variabelen. De vijf significante predictoren waren: één van de documentvariabelen, namelijk (1) het aantal SPE's in het document, twee van de taakvariabelen: (2) het aantal OC's dat verwerkt moet worden voor de taak en (3) het aantal SPE's dat verwerkt moet worden voor de taak en twee van de procesvariabelen: (4) de mate van overeenkomst en (5) type informatie. Aparte regressieanalyses op basis van ras (blanken, zwar-

ten en "hispanics") en opleidingsniveau lieten in alle subgroepen behalve één, overeenkomstige hoeveelheden verklaarde variantie in itemmoeilijkheid zien. Kirsch en Mosenthal replicateerden hun bevindingen dus in verschillende steekproeven, maar ze gebruikten wel steeds dezelfde set leestaken. Het is dus nog onbekend of hun methode tot dezelfde conclusies leidt wanneer andere leestaken gebruikt worden.

Eigen vooronderzoek

In een vooronderzoek is al getracht om ook Nederlandse leestaken in moeilijkheidscomponenten uiteen te zetten (De Geus, Reitsma & De Gloppe, 1992). Er is toen echter niet gekozen voor het uiteenzetten van de documenten in proposities, omdat dit zeer veel tijd kost en omdat de bijdrage van de op de propositionele analyse gebaseerde documentvariabelen aan de predictie eigenlijk maar gering bleek in het onderzoek van Kirsch en Mosenthal (1990). Daarom werden in het pilotonderzoek predictoren van itemmoeilijkheid geselecteerd op basis van andere relevante literatuur (o.a. Wesdorp, 1978; Duffy, 1985; Kreeft, 1985; Kirsch & Jungeblut, 1986; Van den Bergh, Baltzer, Marschall, Triesscheijn & Wesdorp, 1986; Guthrie, 1988). Alleen de variabelen mate van overeenkomst en type informatie werden overgenomen uit de variabelenset van Kirsch en Mosenthal (1990).

In het pilotonderzoek werden uit drie recente peilingsonderzoeken (Wesdorp et al., 1986; Kuhlemcier & Van den Bergh, 1989; Zwarts et al., 1990) 53 teksten en documenten met in totaal 227 items geselecteerd waarvan het materiaal in alledaagse, buitenschoolse situaties voor kon komen. Deze taken werden door twee beoordelaars gescoord op de volgende variabelen: gemiddelde zinslengte, aantal moeilijke woorden, hoeveelheid tekstinformatie, organisatie van de tekst, itemvorm, verwijzing naar de tekst, benodigde tekstinformatie, inhoud van de vraag, benodigde mentale verwerking, aantal te matchen kenmerken, mate van overeenkomst en type informatie.

Uit de resultaten van multiële regressieanalyses bleek dat verschillende variabelen als significante predictoren naar voren kwamen voor elk van de drie steekproeven. Uit een aparte analyse van een subset van taken die in twee van de peilingsonderzoeken afgenomen was ("Schoolreis naar Waddenoog"), bleek daarentegen dat de regressie-vergelijkingen hier nauwelijks verschilden over de twee respondentgroepen. Deze uitkomst was aanleiding voor de veronderstelling dat over de drie datasets verschillen in de specifieke set taken belangrijker zijn voor de predictie dan verschillen in de groepen respondenten waarop de p-waarde gebaseerd is.

De bevindingen van dit vooronderzoek gaven dus nog geen uitsluitsel over de toepasbaarheid van de opzet van Kirsch en Mosenthal op Nederlandse leestaken. In het hier beschreven vervolgonderzoek zijn in tegenstelling tot het eerste onderzoek alleen taken in de dataset opgenomen die met Kirsch en Mosenthal's takenset vergelijkbaar zijn, zodat de kans vergroot wordt dat ook een vergelijkbaar resultaat gevonden wordt. Dit wil zeggen dat er geen prozataken of traditionele tekstbegriptaken opgenomen worden. Een andere reden om deze taken buiten te sluiten is dat op deze soorten taken andere moeilijkheidsvariabelen van toepassing zouden kunnen zijn dan op document-taken.

ONDERZOEK MET DE RELATIONELE GRAMMATICA

Methode

Selectie van taken

Onder een taak verstaan we een document of tekst met daarbij één of meer vragen of opdrachten (items). De selectie van taken voor het replicatieonderzoek moest uiterst zorgvuldig plaatsvinden, omdat deze zoveel mogelijk moest lijken op de door Kirsch en Mosenthal gebruikte taken uit de NAEP Young Adult Literacy Study (hierna te noemen NAEP-studie). Een grote overeenkomst in taken is noodzakelijk omdat resultaten van regressie-analyses altijd afhankelijk zijn van de steekproef en in de hier gebruikte opzet geldt dat voor de steekproef aan taken. Overeen-

komst in taken werd gezocht op het soort tekst of document (materiaal van de taak), het soort items, de moeilijkheidsverdeling van de items en het aantal items per taak. Met het oog op het aantal predictoren dat Kirsch en Mosenthal hadden geïdentificeerd (13), moeten er verder zoveel mogelijk items in de dataset zijn. Hoewel een hele reeks eisen geformuleerd is, waren voor de selectie slechts drie Nederlandse datasets beschikbaar waren, namelijk: de Voorstudie PPO (Wesdorp e.a., 1986); de Proefpeiling Nederlands (Kuhlemeier & Van den Bergh, 1989) en de PPO-Balans van het taalonderwijs (Zwarts e.a., 1990). Voor vergelijking en selectie zijn eerst de proza-taken, d.w.z. traditionele tekstbegripte taken en samenvattingsopdrachten, uit de drie studies uit de bestanden verwijderd. Slechts de functionele taken, die het opzoeken en combineren van informatie uit documenten behelzen, bleven over. Van deze geselecteerde taken worden enkele relevante gegevens in Tabel 1 weergegeven.

Zoveel mogelijk aspecten in aanmerking nemend, zijn de taken uit de Voorstudie PPO de beste keuze voor dit replicatieonderzoek. De Voorstudie is een in 1983-'85 in het basisonderwijs uitgevoerd peilingsonderzoek, waarbij o.a. functionele leestaken (zie Tabel 2) werden afgenomen. Deze waren verspreid over drie toetsboekjes die elk aan gelijkwaardige steekproeven 12-jarigen (8e-groepers) zijn afgenomen. Elke leestaak is door ongeveer 700 leerlingen

Tabel 1. Enkele gegevens betreffende de verschillende datasets na selectie van document-items.

	NAEP	Voorstudie	Proefpeiling	Balans
Aantal items	61	72	84	45
Aantal taken	37	9	5	37
Aantal taken (aantal items) per soort document				
- brief	?	1 (13)	1 (13)	-
- naslag	ja	3 (25)	-	9 (9)
- gebruiksaanwijzing	ja	2 (11)	1 (5)	2 (8)
- tabel	ja	2 (11)	2 (11)	2 (4)
- formulier	ja	1 (12)	1 (55)	-
- grafiek	?	-	-	24 (24)
Gemiddelde moeilijkheid in				
p-waarde	.82	.73	.60	.70
(SD)	(.15)	(.20)	(.28)	(.17)
Verhouding items-documenten	1.6	8.0	16.8	1.2

Noot: Van de NAEP-studie zijn niet alle gebruikte taken openbaar gemaakt. Vandaar dat alleen voor zover het bekend was, aangegeven is of een bepaalde documentsoort aanwezig was in het databestand.

Tabel 2. Beschrijving geselecteerde Voorstudie-taken.

taaknr.	naam van de taak	materiaaltype	aantal items
1	tv-gids	naslag	11
2	inschrijfformulier	brief + formulier	13
3	vlierbes	gebruiksaanwijzing	7
4	omelet	gebruiksaanwijzing	4
5	aanmeldingsformulier	formulier	12
6	dienstregeling	tabel	6
7	kampeerboerderij	tabel	5
8	pictogrammen	naslag	7
9	telefoonboek	naslag	7

gemaakt. De taken uit dit bestand zijn redelijk goed gevarieerd, er zijn genoeg items en de moeilijkheidsverdeling komt nog het best overeen met die van de NAEP-studie (Kolmogorov-Smirnov $Z = 1.321$, two-tailed $p = .06$). Alleen de verhouding items-documenten is eigenlijk hoger dan wenselijk.

In strikte zin is overigens geen sprake van een replicatieonderzoek aangezien weliswaar de takensets zoveel mogelijk overeenkomen op relevante kenmerken, maar de data afkomstig zijn van verschillende respondentgroepen. Om van een echte replicatie te spreken, zou er een taken-set afgenomen moeten zijn bij dezelfde respondentgroep. Dit soort data zijn voor Nederland echter niet beschikbaar. We beschouwen het onderzoek toch in zekere zin als een replicatie, omdat het onderzoek erop gericht is de bruikbaarheid van de methode van Kirsch en Mosenthal te bepalen. Om deze bruikbaarheid te bepalen is het ook nodig aandacht te besteden aan de betrouwbaarheid van scoring.

Propositionele analyse

Op basis van de verschenen publicaties over de relationele grammatica (Mosenthal, 1985; Kirsch & Mosenthal, 1990; Mosenthal & Kirsch, 1991; Sheehan & Mislevy, 1990) is een handleiding samengesteld om de wijze waarop het materiaal uiteen gezet moest worden in proposities duidelijk vast te leggen. Als training zetten de eerste auteur en een onafhankelijke beoordelaar afzonderlijk vijf verschillende typen materialen uiteen in proposities. Verschillen van mening werden besproken waardoor het gebruik van de grammatica werd opgehelderd. Hierna zijn door iedere beoordelaar apart alle documenten uit de Voorstudie uiteengezet in proposities.

De betrouwbaarheid is als volgt bepaald. Per taak waren er twee uiteenzettingen in proposities beschikbaar en hiervan is geteld hoeveel OC's, ingebedde OC's, SPE's en ingebedde SPE's elke uiteenzetting had. Daarnaast is bepaald hoeveel van die OC's etc., identiek waren over de beide uiteenzettingen. Als maat voor de betrouwbaarheid van uiteenzetting in proposities geldt nu het percentage identieke eenheden ten opzichte van het totale aantal unieke geïdentificeerde eenheden gesommeerd over beide beoordelaars. Een nadeel van de gevolgde methode is dat één beoordelaar het percentage overeenkomst kan beïnvloeden door zeer veel unieke eenheden te onderkennen en hiermee de noemer van de verhouding op te schroeven. Dit verschijnsel doet zich bijvoorbeeld voor als de ene beoordelaar een informatie-eenheid als SPE benoemd en de andere als OC. Dit heeft gevolgen voor de overeenkomst in aantal SPE's en in aantal OC's, hetgeen te zien is bij de taak "telefoonboek" (taak 9, zie Tabel 3).

Inspectie van de gegevens in Tabel 3 laat zien dat afgezien van de taak "telefoonboek" (taak 9) de overeenkomst redelijk tot goed is. Het grootste deel van de propositionele analyse heeft tot identieke resultaten geleid. In het geval er verschil van mening was, werd dit achteraf besproken totdat er overeenstemming bereikt was over zowel de micro- als de macro-structuur van het document. In een paar gevallen leidde dit tot een oplossing die geen van beide beoordelaars in

Tabel 3. Percentage identieke eenheden van het totaal aantal onderscheiden eenheden per taak.

variabele	taak								
	1	2	3	4	5	6	7	8	9
# OC's	76	100	100	100	100	89	50	100	8
# ingebedde OC's	75	100	100	100	100	88	52	100	11
diepste niveau ingebedde OC	60	100	100	100	100	100	100	100	66
# SPE's	86	92	100	100	100	100	99	88	0
# ingebedde SPE's	40	50	100	50	100	0	100	100	0
diepste niveau ingebedde SPE	100	66	100	100	100	50	100	100	50

Noot: afgeleide OC's en SPE's zijn niet in de beoordeling betrokken

Tabel 4. Percentage taken (9) of items (72) waarbij exacte overeenkomst was tussen beoordelaars.

Variabele	Aantal beoordelaars die het eens zijn (max.3):		
	3	2	2 of 3 (som)
1 # OC's	100	0	100
2 # ingebedde OC's	78	11	89
3 diepste niveau van ingebedde OC	89	11	100
4 # SPE's	78	22	100
5 # ingebedde SPE's	56	44	100
6 diepste niveau van ingebedde SPE	89	11	100
7 # OC's dat verwerkt moet worden voor de vraag	36	43	79
8 diepste niveau van ingebedde OC nodig voor de vraag	58	42	100
9 # SPE's dat verwerkt moet worden voor de vraag	24	50	74
10 diepste niveau van ingebedde SPE nodig voor de vraag	72	26	98
11 mate van overeenkomst	46	47	93
12 soort informatie	60	22	82
13 plausibiliteit van distractors	44	44	88

eerste instantie gevonden had. Op deze wijze werd voor elke taak één versie tot stand gebracht, die kon dienen als basis voor de beoordeling van de taken.

Beoordeling van taken

De Voorstudie-taken dienden vervolgens beoordeeld te worden op de dertien variabelen die Kirsch en Mosenthal (1990) onderscheiden hebben. De eerste auteur en twee onafhankelijke beoordelaars scoorden alle 9 documenten en 72 items op deze variabelen. De beoordelaars kregen een handleiding bij de scoring, waarin ook een toelichting over de grammatica en het daaraan verbonden cognitieve verwerkingsmodel van Kirsch en Mosenthal gegeven werd. Voorafgaand aan de eigenlijke scoring werden als oefening drie van de vijf al eerder gebruikte oefentaken gescoord. Bij elke oefentaak hoorden twee items. Na elke oefentaak werden de onderlinge verschillen besproken en werden de scoringsregels waar nodig verduidelijkt.

De betrouwbaarheid is bepaald door te berekenen in hoeveel van de 9 (materiaal) dan wel 72 (items) gevallen alle drie de beoordelaars dezelfde scoring hadden gegeven (een verschil $\leq 1\%$ bij de materiaalvariabelen werd als identiek beschouwd) en hoe vaak slechts 2 van de drie dezelfde mening hadden. In Tabel 4 zijn de gevonden percentages weergegeven.

Zoals in Tabel 4 te zien is en uit navraag bleek, hebben de beoordelaars de scoring van enkele vraagvariabelen (aantal OC's en aantal SPE's nodig voor de vraag) anders opgevat. Nadere inspectie van de scoringsregels en voorbeelden van Kirsch en Mosenthal (Kirsch & Mosenthal, 1990; Mosenthal & Kirsch, 1991; Sheehan & Mislevy, 1990) bleek geen uitsluitel te geven. Er is daarom in overleg met de beoordelaars en op basis van de beschikbare informatie een nieuwe scoringsversie voor de vraagvariabelen geschreven.

Door iedere beoordelaar zijn in totaal 558 verschillende scoringen gegeven gemaakt, nl. 9 taken werden op 6 materiaalvariabelen gescoord plus 72 items werden op 7 variabelen gescoord. Van deze 558 beoordelingen is in 64 gevallen door elk van de drie beoordelaars iets anders gescoord (11%). Deze 64 gevallen zijn door twee beoordelaars nader bekeken en hier is alsnog naar overeenstemming gezocht door onderlinge discussie. Van de vraagvariabelen is tevens nagegaan of de scoring houdbaar bleef in het licht van de nieuwe opvatting over hoe deze gescoord moeten worden. Door deze discussie achteraf is voor alle variabelen tot een eensluidende beoordeling gekomen.

RESULTATEN EN DISCUSSIE

In Tabel 5 zijn de gemiddelde scores over alle items van de beoordelingsvariabelen weergegeven. De meeste variabelen blijken niet normaal verdeeld (Kolmogorov-Smirnov Z toetsen). Alleen de verdeling van de p-waarde lijkt uit een normaal-verdeelde populatie te komen². Ook bij de gegevens van Kirsch en Mosenthal (1990; Sheehan & Mislevy, 1990) blijken niet-normale verdelingen overigens eerder regel dan uitzondering.

Beoordelingen van de taken en items op de dertien verschillende variabelen werden ingevoerd als predictoren in een multipel regressie-analyse met de p-waarde als afhankelijke variabele. In Tabel 6 worden de correlaties tussen de variabelen gegeven.

In navolging van Kirsch en Mosenthal zijn de significant met de p-waarde samenhangende predictoren in de regressievergelijking gedwongen. De resultaten van deze zogenaamde "forced

Tabel 5. Verdeling van de aan Kirsch en Mosenthal ontleende variabelen.

Variabele	gemiddelde	s.d.	min	max
p-waarde	72.99	19.72	26	100
# OC's	46.13	56.07	0	168
# ingebedde OC's	42.08	51.91	0	151
diepste niveau van inbedding van een OC	1.86	1.7	0	4
# SPE's	127.68	161.18	11	540
# ingebedde SPE's	31.53	69.44	0	238
diepste niveau van inbedding van een SPE	2.44	1.37	1	5
# OC's dat verwerkt moet worden voor de vraag	6.24	20.41	0	137
diepste niveau van inbedding van OC nodig voor de vraag	1.06	1.44	0	4
# SPE's dat verwerkt moet worden voor de vraag	8.25	25.28	0	165
diepste niveau van inbedding van SPE nodig voor de vraag	.83	.53	0	2
mate van overeenkomst	2.03	1.34	1	5
soort informatie	2.47	1.55	1	5
plausibiliteit van distractors	1.97	.75	1	3

Tabel 6. Correlaties tussen de dertien predictoren en de criteriumvariabele.

	document					vraag					proces		
	1	2	3	4	5	6	7	8	9	10	11	12	13
(1) # OC's	-												
(2) # ingebedde OC's	1.00	-											
(3) diepste inbedding OC	.84	.84	-										
(4) #SPE's	.72	.70	.56	-									
(5) # ingebedde SPE's	.31	.27	.29	.82	-								
(6) diepste inbedding SPE	-.20	-.19	.20	-.06	.30	-							
(7) # OC's voor taak	.53	.52	.33	.12	-.11	-.28	-						
(8) diepste OC voor taak	.86	.87	.75	.39	-.09	-.28	.54	-					
(9) # SPE's voor taak	.42	.40	.25	.07	-.12	-.29	.97	.44	-				
(10) diepste SPE voor taak	.18	.16	.27	.20	.22	.12	.09	.07	.12	-			
(11) mate v overeenkomst	.45	.48	.31	.17	-.21	-.30	.35	.55	.30	.13	-		
(12) type informatie	.34	.34	.18	.07	-.28	-.60	.38	.40	.40	.17	.53	-	
(13) distractors	.12	.10	.16	.26	.23	-.12	.01	.00	.04	.34	.15	.45	-
p-waarde	-.37	-.36	-.23	-.33	-.13	.25	-.36	-.30	-.34	-.24	-.44	-.61	-.31

Noot: dikgedrukt is significant op $\leq .05$ ($r \geq .232$) of $\leq .01$ ($r \geq .302$) niveau

Tabel 7. Betacoëfficiënten en multiële correlatie bij vier multiële regressie analyses met aan Kirsch en Mosenthal ontleende predictoren.

predictor	methode			
	1	2	3	4
	enter (preselectie)	stapsgewijze selectie	enter (bewezen variabelen gedwongen)	voorwaartse selectie
# OC's	1.433	—	—	.641**
# ingebedde OC's	-.767	—	—	—
diepste inbedding OC	—	—	—	—
# SPE's	-.689**	—	-.262**	-.641**
# ingebedde SPE's	—	-.324**	—	—
diepste inbedding SPE	—	—	—	-.226*
# OC's voor taak	-.1088*	—	-.174	-1.156*
diepste OC voor taak	-.008	—	—	—
# SPE's voor taak	.758	—	.072	.823
diepste SPE voor taak	—	—	—	—
mate van overeenkomst	-.134	—	-.090	-.156
type informatie	-.622**	-.696**	-.502**	-.720**
plausibiliteit van distractors	.055	—	—	—
Multiple R	.73	.68	.68	.75
R ²	.53	.46	.47	.56
Adjusted R ²	.46	.45	.43	.51
SE	14.45	14.64	14.91	13.91

Noot: significantie beta-gewichten: * $\leq .05$, ** $\leq .01$. Een liggend streepje geeft aan dat de variabele niet in de vergelijking is opgenomen bij de regressie-analyse.

entry"-methode staan in Tabel 7 in de eerste kolom. Hieruit komt ongeveer hetzelfde beeld naar voren als bij Kirsch en Mosenthal, hoewel in zwakkere vorm. De verklaarde variantie is minder groot (53% versus 89%) en het aantal significante predictoren is ook iets kleiner (drie versus vijf), maar deze maken wel deel uit van de vijf significante predictoren van Kirsch en Mosenthal (aantal SPE's, aantal OC's dat verwerkt moet worden voor de vraag, aantal SPE's dat verwerkt moet worden voor de vraag, mate van overeenkomst en type informatie). Deze procedure (preselectie van predictoren) leidt echter tot een dubbele kapitalisatie op kans, want er wordt tweemaal op significantie getoetst met dezelfde gegevens. Daarom is ook de stapsgewijze selectieprocedure toegepast (zie de tweede kolom in Tabel 7). Bij deze methode van predictorenselectie zijn er maar twee variabelen die een significante bijdrage leveren aan de predictie van de p-waarde.

De bovenbeschreven methoden van predictorenselectie leiden tot een steekproef-afhankelijke regressievergelijking, omdat het geheel van de dataset afhangt welke predictoren opgenomen worden. Er kan voor het selecteren van predictoren ook gebruik gemaakt worden van een theorie of van eerder verworven kennis. De variabelen die zich in een eerder onderzoek "bewezen" hebben worden hierbij in de regressievergelijking gedwongen. Daarna worden eventueel aanvullende predictoren toegelaten. Als de bovengenoemde vijf variabelen van Kirsch en Mosenthal in de regressievergelijking gedwongen worden, leidt dat tot de resultaten die in de derde kolom van Tabel 7 staan. De resultaten van het additioneel toepassen van het voorwaarts selectiecriteria op de resterende variabelen staat in de vierde kolom.

In alle analyses blijkt type informatie een significante predictor. In de stapsgewijze procedure komt het aantal ingebedde SPE's als tweede predictor. In de overige drie analyses komt aantal SPE's daarvoor in de plaats. Deze twee variabelen zijn overigens sterk gecorreleerd (.82), dus het is te verwachten dat als één van beide opgenomen wordt in de regressievergelijking, de andere variabele niet wordt opgenomen. Het aantal SPE's en type informatie zijn de twee meest robuuste predictoren. Een minder stabiele predictor is het aantal OC's dat verwerkt moet worden voor de vraag. Alleen in de voorwaartse selectie-procedure komen aantal OC's en diepste niveau van inbedding van een OC nog naar voren.

Men kan zich afvragen in hoeverre de variabelen geacht worden additief bij te dragen aan de voorspelling van de p-waarde. Het is bijvoorbeeld mogelijk dat het effect van het aantal SPE's dat verwerkt moet worden voor de vraag afhangt van het totaal aantal SPE's in het document. Daarom is in een exploratieve analyse de verklarende bijdrage van interacties getoetst. In een stapsgewijze selectieprocedure deden alle tweeweg-interacties mee als potentiële predictor, naast alle oorspronkelijke variabelen. De belangrijkste predictor was type informatie en daarnaast werden drie interactievariabelen opgenomen. Samen verklaarden deze vier variabelen 58% van de variantie in itemmoeilijkheid, terwijl in de analyse zonder interacties 46% werd verklaard door twee predictoren. Helaas zijn de gevonden interactievariabelen nogal lastig te interpreteren (aantal ingebedde SPE's * aantal ingebedde OC's; aantal ingebedde SPE's * type informatie; mate van overeenkomst * type informatie).

Zoals weergegeven is in Tabel 7 blijft de door de oorspronkelijke variabelen verklaarde hoeveelheid variantie in itemmoeilijkheid in alle vier analyses beperkt tot rond de 50 procent. Soms is dit in de sociale wetenschappen al een zeer aardig resultaat. Op basis van de resultaten van Kirsch en Mosenthal (89%) is echter de te verwachten hoeveelheid verklaarde variantie in een onafhankelijke steekproef 87% volgens Wherry's adjusted R^2 . Het is duidelijk dat dit niet gehaald wordt in onze analyse van Nederlandse gegevens. Er zijn verschillende mogelijke oorzaken voor het feit dat de hoeveelheid verklaarde variantie minder groot is.

Ten eerste bestaat de mogelijkheid dat de scoring van de variabelen anders opgevat is dan zoals Kirsch en Mosenthal het bedoeld hadden. De verwarring omtrent de scoringsregels van de vraagvariabelen is al genoemd. Een ander problematische categorie zijn de procesvariabelen. Het onafhankelijk scoren van mate van overeenkomst en type informatie bleek soms lastig. Bijvoorbeeld, staat de in de vraag gegeven informatie letterlijk in de tekst en moet daarna het antwoord afgeleid worden, of moet de overeenkomst met de tekst afgeleid worden en is het antwoord daarna eenvoudig te identificeren.

Uit de correlaties tussen alle variabelen blijkt dat ook multicollineariteit een probleem is. Ook uit de *variance inflation* factoren blijkt dat de variabelen onderling sterk samenhangen: wanneer een predictor voor meer dan 90% voorspeld kan worden uit alle andere predictoren, dan zou men moeten overwegen of deze predictor niet beter verwijderd kan worden (Stevens, 1992, p.77). Volgens dit criterium zou verwijdering overwogen moeten worden voor de volgende variabelen: aantal OC's, aantal ingebedde OC's, diepste niveau van inbedding van een OC, aantal SPE's, aantal ingebedde SPE's, diepste niveau van inbedding van een OC dat verwerkt moet worden voor de vraag en diepste niveau van inbedding van een SPE dat verwerkt moet worden voor de vraag, onder voorbehoud dat er waarschijnlijk bij verwijdering van één van deze variabelen een ander beeld ontstaat voor de overgebleven variabelen. De afhankelijkheid lijkt vooral te gelden voor materiaalvariabelen, hetgeen ook niet verwonderlijk is aangezien bijvoorbeeld het aantal ingebedde OC's een deelverzameling is van het totaal aantal OC's. Bij verwijdering van de genoemde predictoren neemt de totale verklaarde variantie in de p-waarde natuurlijk niet veel af, omdat de unieke bijdrage van de betreffende predictor maar zeer klein kon zijn. Bijvoorbeeld na verwijdering van aantal ingebedde OC's en aantal ingebedde SPE's uit de predictorset wordt in een stapsgewijze regressie 45% variantie ($R=.67$) verklaard door type informatie en aantal SPE's (vergelijk $R=.68$ zonder verwijdering).

Stevens (1992, p.77) noemt onder andere twee manieren om multicollineariteit op te lossen: 1) hoog gecorreleerde ($r \geq .80$) predictoren combineren door ze op te tellen; 2) door middel van principale componenten analyse een kleiner aantal (ongecorreleerde) predictoren vormen; de

regressie wordt dan op de factorscores gedaan. Deze twee mogelijkheden zijn beide beproefd. Een principale componenten analyse op de predictoren levert vier factoren op die in een multi-pele regressie samen 37% variantie in de p-waarde verklaren. De interpretatie van deze factoren levert echter problemen op, terwijl wij juist geïnteresseerd zijn in de bijdragen van de verschillende predictoren. Als gekozen wordt voor optelling van acht sterk samenhangende variabelen ontstaan drie nieuwe variabelen: OCsom (aantal OC's, aantal ingebedde OC's, diepste niveau van inbedding van een OC en diepste niveau van inbedding van een OC dat moet worden verwerkt), SPEsom (aantal SPE's en aantal ingebedde SPE's) en taaksom (aantal OC's dat verwerkt moet worden en aantal SPE's dat verwerkt moet worden). Samen met de vijf overgebleven oorspronkelijke variabelen worden deze samengestelde variabelen geanalyseerd. De preselectie-methode met alle acht variabelen in de regressie levert een multi-pele correlatie van .73 (53% verklaarde variantie) op met OCsom, SPEsom en type informatie als significante predictoren. Een stapsgewijze selectie van predictoren leidt tot 46% ($R = .68$) verklaarde variantie met als significante predictoren type informatie en SPEsom. Deze oplossingen leveren dus weinig verandering in de resultaten op.

Het feit dat er in dit onderzoek door de gezamenlijke variabelen minder variantie verklaard wordt dan bij de studie van Kirsch en Mosenthal kan ook samenhangen met de specifieke itemset die hier gebruikt is. Hierop zal verder worden ingegaan nadat het onderzoek naar de predictie met de alternatieve variabelenset hieronder is gerapporteerd.

Als voorlopige conclusie kan worden gesteld dat ondanks de veelbelovende resultaten van Kirsch en Mosenthal en de waarschijnlijkheid dat de door hen gevonden variabelen van belang zijn voor de moeilijkheid van functionele leestaken, in het huidige onderzoek maar in beperkte mate bevestiging gevonden kon worden voor deze verwachting.

ONDERZOEK MET BEOORDELINGSSCHALEN

Uit de voorgaande gegevens blijkt dat de moeilijkheid van de Voorstudie-items maar voor de helft verklaard kon worden door de variabelen die in de studie van Kirsch en Mosenthal zo veelbelovend leken. In aanmerking genomen de hoeveelheid werk die het uiteenzetten in proposities met zich meebrengt, is het daarom de moeite waard na te gaan of met een andere operationalisatie niet een even goede of betere voorspelling van de moeilijkheid gegeven kan worden. Als dat het geval zou blijken te zijn dan is een moeizame uiteenzetting van taken met behulp van een relatieve grammatica overbodig.

Ten behoeve van dit onderzoek is eerst een literatuuronderzoek gedaan naar mogelijke variabelen die specifiek de moeilijkheid van documenttaken verklaren. Er zijn al veel pogingen gedaan om de moeilijkheid van items te voorspellen vanuit allerlei tekst- en itemkenmerken. Er is echter weinig onderzoek gedaan waarbij de moeilijkheid van *verschillende* taken vergeleken wordt. Meestal wordt juist de takenset zo homogeen mogelijk gemaakt, bijvoorbeeld meerkeuzevragen bij één alinea informatieve tekst (o.a. Drum, Calfee & Cook, 1981; Embretson & Wetzel, 1987; Davey, 1988; Scheuneman & Gerritz, 1990). Dit soort onderzoek is niet op voorhand uitgesloten, maar voor de keuze van predictoren is echter in eerste instantie gezocht naar variabelen die in eerder onderzoek van belang bleken voor functionele leestaken (Kirsch & Guthrie, 1980; Kirsch & Jungeblut, 1986; Kirsch & Mosenthal, 1990; Sheehan & Mislevy, 1990; Kirsch, Jungeblut & Campbell, 1992). Daarnaast zijn er allerlei classificatie-systemen voor items en teksten waarbij een relatie met de moeilijkheid slechts verondersteld wordt, maar geen empirische gegevens hierover worden verstrekt (e.g. Davis, 1968; Wesdorp, 1978; Duffy, 1985; Kadar-Fülop, 1985; Kreeft, 1985; Van den Bergh et al., 1986).

Uit het literatuuronderzoek zijn in eerste instantie de volgende, specifiek voor opzoektaken geschikte variabelen geselecteerd om de itemmoeilijkheid te voorspellen (met daarachter de bronvermelding): lengte en dichtheid van de tekst (Kirsch & Jungeblut, 1986; Kirsch, Jungeblut & Campbell, 1992), aantal kenmerken/features die de lezer moet verwerken (Kirsch & Jungeblut, 1986), overeenkomst tussen de informatie die in de vraag gegeven wordt en de informatie

in het document (Kirsch & Jungeblut, 1986), wat vereist de vraag: localiseren, matchen, integreren, genereren (Kirsch & Jungeblut, 1986), distractors/plausibele doch foutieve antwoorden (Kirsch & Jungeblut, 1986), volgorde-congruentie van informatie in vraag en document (Kirsch & Guthrie, 1980), opvallendheid van de gevraagde informatie (Kirsch & Guthrie, 1980), hoeveelheid woorden/getallen in het document (Kirsch & Guthrie, 1980), correspondentie tussen woorden en typografie in de vraag en het document (Kirsch & Guthrie, 1980), aantal informatie-categorieën in vraag en document die verwerkt moeten worden (Kirsch & Guthrie, 1980)

De bovenstaande variabelen uit de literatuur en de eerder genoemde variabelen die in het pilotonderzoek geselecteerd en uitgeprobeerd waren (De Geus, Reitsma & De Glopper, 1992) zijn globaal uitgetest op een aantal documenttaken. Hierbij is gelet op discriminerend vermogen, toepasbaarheid op verschillende soorten materiaalvormen en eenduidigheid in het scoren. Tevens moest rekening gehouden worden met de wens inhoudelijk dezelfde variabelen te gebruiken als Kirsch en Mosenthal. Er zijn uiteindelijk negen variabelen gekozen waarop de Voorstudie-taken gescoord moesten worden: vier materiaalvariabelen (die de hoeveelheid informatie in het document en de organisatie ervan aangeven) en vijf itemvariabelen (die de moeilijkheid van het zoeken van de gegeven en gevraagde informatie aangeven). Hieronder wordt van de variabelen een summier omschrijving gegeven. Een uitgebreidere beschrijving van de scoring is weergegeven in Bijlage 1.

- 1) *de structuur van het materiaal* (afgeleid uit Kirsch & Jungeblut, 1986): Een combinatie van lengte en organisatie van het materiaal.
- 2) *de hoeveelheid informatie* (klassiek): Het aantal woorden en getallen in het materiaal
- 3) *het aantal zinnen* (klassiek, maar aangepast voor tabellen)
- 4) *de organisatie van het document* (De Geus, Reitsma & De Glopper, 1992): In hoeverre wordt er door de presentatie en indeling duidelijkheid gecreëerd, of in hoeverre vergemakkelijkt de schrijver het begrip door middel van opbouw en hulpmiddelen.
- 5) *benodigde documentinformatie* (vergelijk Drum, Calfee & Cook, 1981; Kádár-Fülöp, 1985; Van den Bergh et al., 1986): Hoeveel document-informatie moet verwerkt worden om het antwoord te vinden?
- 6) *mate van correspondentie tussen woorden en typografie* (Kirsch & Guthrie, 1980): Als de organisatie, kopjes of algehele typografie van het document gerelateerd zijn aan de vraag is het gemakkelijker om de benodigde informatie te vinden.
- 7) *mate van overeenkomst* (vergelijk Kirsch & Mosenthal, 1990): Deze variabele verwijst naar de mate van overeenkomst tussen de informatie die in een vraag of opdracht gegeven wordt en de corresponderende informatie in het materiaal.
- 8) *verwerkingsdiepte* (vergelijk Kirsch & Mosenthal, 1990, type informatie): De verwerking die nodig is om het antwoord te geven bestaat uit het localiseren, matchen, vergelijken of integreren van een of meer kenmerken.
- 9) *het aantal en de plausibiliteit van de distractors* (vergelijk Kirsch & Mosenthal, 1990): De mate waarin plausibele antwoorden die niet het goede antwoord zijn, afleidend zijn.

Beoordeling van de taken met beoordelingsschalen

Zes beoordelaars (waaronder de eerste auteur) scoorden alle negen documenten en 72 items op deze variabelen. De beoordelaars kregen een handleiding bij de itemscore, waarin ook een korte uiteenzetting van het cognitieve verwerkingsmodel van Kirsch en Mosenthal gegeven werd. Voor iedere variabele was elk schaalpunt omschreven (behalve voor variabele 2 en 3 waarbij geteld moest worden). Voorafgaand aan de eigenlijke scoring werden vier trainingstaken gescoord met elk twee items. Na elke trainingstaak werden de verschillen besproken en werden de scoringsregels waar nodig verduidelijkt.

Voor het bepalen van de mate van overeenkomst tussen alle zes beoordelaars is gekozen voor de concordantiemaat W van Kendall, een maat die lineair gerelateerd is aan de gemiddelde rangcorrelatie-coëfficiënt (Siegel & Castellan, 1988). In Tabel 8 is deze overeenstemmingsmaat per variabele weergegeven in kolom 1. Voor een aantal variabelen laat deze overeenstemming nog te wensen over. Een probleem van overeenstemmingsmaten is overigens dat er alleen een

uitspraak wordt gedaan over de mate waarin de beoordelaars op dezelfde manier gerangschikt hebben. Er vloeit niet automatisch uit voort wat de "juiste" rangschikking is. Men zou het gemiddelde rangnummer kunnen nemen, als niet het idee bestond dat er een ware score is die echter niet door alle beoordelaars onderkend is. Vandaar dat er in plaats van de gemiddelde score, de modale score genomen wordt. Berekend is in hoeveel van de negen (materiaal) dan wel 72 (items) gevallen een absolute meerderheid, dus vier of meer van de zes beoordelaars exact dezelfde scoring had gegeven (bij hoeveelheid informatie en aantal zinnen werd een verschil van $\pm 1\%$ als identiek beschouwd). In Tabel 8 is per variabele dit percentage onder het kopje absolute meerderheid weergegeven. Er bleek een lage overeenstemming te zijn over het aantal zinnen en de organisatie. Dit heeft te maken met de moeilijkheden die het scoren van deze variabelen blijkbaar opleverde bij minder "tekstuele" documenten. Bij het aantal zinnen is het gebrek aan identieke overeenkomst echter voornamelijk te wijten aan telfouten; Kendall's W geeft aan dat de scoringen van de beoordelaars voor wat de volgorde van taken betreft zeer goed overeenkomen.

Daarnaast is in Tabel 8 opgenomen het percentage overeenkomst indien ook items meegerekend worden waarbij drie beoordelaars het eens waren mits de andere drie het niet ook eens waren (genoemd "relatieve meerderheid"). Het uitgaan van een absolute of relatieve meerderheid blijkt niet in alle gevallen tot een eenduidige beslissing te leiden. In 12 van de 36 (4*9) beoordelingen van materiaalvariabelen moest op andere wijze bepaald worden wat de juiste

Tabel 8. De interbeoordelaarsovereenkomst voor de variabelen in de alternatieve beoordeling.

Variabele	Kendall's W	meerderheid	
		absoluut	relatief
structuur (5)	.71	100	100
hoeveelheid informatie	.99	55	77
aantal zinnen	.99	33	56
organisatie (6)	.62	33	33
benodigde documentinformatie (6)	.72	61	76
correspondentie (5)	.69	56	79
mate van overeenkomst (5)	.48	54	88
verwerkingsdiepte (6)	.63	57	78
distractors (7)	.76	77	90

Noot: bij elke variabele staat tussen haakjes het aantal keuzemogelijkheden op de beoordelingsschaal

Tabel 9. Gemiddelden en standaarddeviaties van de variabelen in de alternatieve beoordeling.

Variabele	Gemiddelde	SD
structuur	2.58	.67
hoeveelheid informatie	429.11	424.25
aantal zinnen	143.99	198.53
organisatie	2.78	.83
benodigde informatie	2.90	1.39
correspondentie	3.07	1.00
mate van overeenkomst	1.94	1.34
verwerkingsdiepte	2.81	1.90
distractors	5.14	2.43
p-waarde	72.99	19.72

score is en idem bij 59 van de 360 (4*72) itembeoordelingen. De betreffende gevallen zijn daarom nogmaals kritisch beoordeeld, waarbij bij elke door de beoordelaars gegeven scoring zorgvuldig is nagegaan hoe men tot deze score gekomen kon zijn en zo is tot een eindbeoordeling gekomen, die gebruikt werd in de voorspelling van de itemmoeilijkheid.

Resultaten

In Tabel 9 zijn de gemiddelde scores over alle items van de beoordelingsvariabelen weergegeven en de intercorrelaties in Tabel 10. Om enigszins te controleren of de alternatieve predictoren variabelen zoals verwacht dezelfde variantiebronnen dekken als de variabelen van Kirsch en Mosenthal, zijn canonische correlatie analyses gedaan. Hierbij wordt uit elk van de twee groepen variabelen een zodanige lineaire combinatie gezocht dat de correlatie tussen deze combinaties maximaal is. Een analyse op de materiaalvariabelen apart leverde een canonische correlatie van 1.00 op tussen twee lineaire combinaties waarvan aantal zinnen en hoeveelheid informatie enerzijds en aantal SPE's en aantal ingebedde SPE's anderzijds de meest bijdragende variabelen zijn. Een afzonderlijke analyse op de taak- en procesvariabelen gezamenlijk leverde een canonische correlatie van .99 op tussen twee lineaire combinaties waarbij verwerkingsdiepte enerzijds en type informatie anderzijds de meest bijdragende variabelen zijn en een canonische correlatie van .97 tussen twee lineaire combinaties waarvan mate van overeenkomst en diepste niveau van inbedding van een SPE's dat verwerkt moet worden voor de taak enerzijds en mate van overeenkomst uit de alternatieve variabelen en benodigde tekstinformatie anderzijds de meest bijdragende variabelen zijn. Op basis van deze resultaten kan gesteld worden dat de alternatieve predictoren inderdaad dezelfde bronnen van variantie aanspreken en kunnen fungeren als alternatieve operationalisatie van de door Kirsch en Mosenthal onderscheiden bronnen van variantie.

Voor de multiële regressie analyses zijn dezelfde procedures gevolgd als in de eerder beschreven studie. In kolom 1 (van Tabel 11) staan de resultaten vermeld voor het geval dat de significant met p-waarde samenhangende predictoren in de regressievergelijking gedwongen worden³. In kolom 2 van dezelfde tabel staan de resultaten van de stapsgewijze selectieprocedure.

Een andere gebruikte methode was om allereerst de variabelen die zich in een eerder onderzoek al "bewezen" hebben in de regressievergelijking te dwingen en dan voorwaarts aanvullende predictoren te selecteren. Nu is, alhoewel ze wel dezelfde variantiebronnen aanspreken als de voorspellende variabelen van Kirsch en Mosenthal, voor de nieuwe set variabelen geen eerder bewijs van hun importantie geleverd. Als de predictoren worden opgenomen die het best vergelijkbaar zijn met die van Kirsch en Mosenthal, dan kunnen de volgende in de vergelijking gedwongen worden: aantal zinnen (vergelijkbaar met aantal SPE's in het materiaal), benodigde

Tabel 10. Correlaties tussen de variabelen in de alternatieve beoordeling.

variabele	1	2	3	4	5	6	7	8	9
1 structuur	-								
2 hoeveelheid info	.46	-							
3 aantal zinnen	.42	.91	-						
4 organisatie	.60	.41	.09	-					
5 benodigde info	.32	.12	.32	.01	-				
6 correspondentie	-.59	-.57	-.65	-.36	-.46	-			
7 mate v. overeenkomst	.13	-.15	.10	-.27	.70	-.07	-		
8 verwerkingsdiepte	-.15	-.16	-.06	-.09	.65	-.08	.49	-	
9 distractors	.11	.39	.40	.23	.63	-.39	.32	.52	-
p-waarde	-.07	-.14	-.28	.16	-.54	.16	-.51	-.59	-.47

Noot: dikgedrukt is significant op $\leq .05$ ($r \geq .232$) of $\leq .01$ ($r \geq .302$) niveau

Tabel 11. Beta-coëfficiënten en multiële correlatie bij drie multiële regressie met de alternatieve predictoren.

predictor	method		
	1	2	3
	enter (preselectie)	stapsgewijs	enter (bewezen variabelen gedwongen)
Structuur	—	—	—
Hoeveelheid informatie	—	—	—
Aantal zinnen	-.302**	-.286**	-.328**
Organisatie	—	—	—
Benodigde informatie	-.177	—	-.138
Correspondentie	—	—	—
Mate van overeenkomst	-.325**	-.246*	-.308*
Verwerkingsdiepte	-.517**	-.485**	-.546**
Distractors	-.088	—	—
Multiple R	.71	.70	.70
R ²	.50	.49	.50
Adjusted R ²	.46	.47	.47
SE	14.47	14.39	14.42

Noot: significantie beta-gewichten: * $\leq .05$, ** $\leq .01$. Een liggend streepje geeft aan dat de variabele niet in de vergelijking is opgenomen bij de regressie analyse.

tekstinformatie (vgl. aantal SPE's en OC's nodig voor de vraag), en mate van overeenkomst en verwerkingsdiepte. Het resultaat van deze analyse is weergegeven in kolom 3. Er bleken met de voorwaartse selectie methode geen variabelen te worden toegevoegd die aanvullende variantie verklaren. De resultaten van deze analyse komen dus exact overeen met de gegevens in kolom 3.

Uit de resultaten van de verschillende analyses komt het volgende consistente beeld naar voren: steeds zijn er dezelfde predictoren, namelijk aantal zinnen, mate van overeenkomst en verwerkingsdiepte die gezamenlijk 50% van de variantie in de itemmoeilijkheid verklaren.

Ook zijn bij deze set variabelen in een stapsgewijze regressieanalyse alle tweeweg interactie-variabelen naast de oorspronkelijke variabelen als predictoren gebruikt. De enige predictoren die opgenomen werden in de regressievergelijking zijn drie interactie-variabelen en die verklaren samen 53% variantie in de p-waarde. Dit zijn mate van overeenkomst * verwerkingsdiepte, aantal zinnen * aantal distractors en correspondentie * verwerkingsdiepte. De interpretatie van deze variabelen is echter bijzonder lastig.

ALGEMENE CONCLUSIES EN DISCUSSIE

In deze studie is getracht de moeilijkheid van een aantal functionele leestaken uit een recente leesvaardigheidsstudie te verklaren op grond van een aantal formele aspecten van de materialen en de gestelde vragen. Aspecten die werden onderscheiden waren bijvoorbeeld de hoeveelheid informatie in het materiaal, de organisatie van deze informatie, hoeveel van deze informatie nodig is om het antwoord te geven, hoeveel verwerking nodig is om het antwoord te geven, enzovoort. Het is alleszins aannemelijk dat dit relevante parameters voor de vraagmoeilijkheid zijn, maar onderzoek waarin de empirische significantie wordt aangetoond, was tamelijk schaars.

In het huidige onderzoek zijn twee predictorsets gebruikt. Eén was afkomstig uit een studie van Kirsch en Mosenthal, waarin een propositionele analyse van de materialen werd verricht, en de andere was een set variabelen, waarvan verwacht werd dat dezelfde bronnen van variantie gemeten werden maar waarvoor het niet nodig was de documenten uiteen te zetten in proposities. Het uiteenzetten in proposities is namelijk veel werk, terwijl uit de resultaten van de studie van Kirsch en Mosenthal bleek dat het alleen een maat voor de hoeveelheid informatie in het document als significante predictor opleverde.

Uit de huidige resultaten blijkt dat er zowel overeenkomsten als verschillen zijn tussen de twee predicties van de itemmoeilijkheid. Zowel met een analyse op de manier van Kirsch en Mosenthal, als met een analyse met behulp van een set alternatieve variabelen wordt door drie à vier variabelen ongeveer de helft van de variantie in de p-waarden verklaard. Van deze variabelen is in elk geval type informatie/verwerkingsdiepte van belang. Daarnaast blijkt ook elke keer een maat voor de hoeveelheid informatie in het materiaal van belang, namelijk het aantal (ingebodde) SPE's of het aantal zinnen. Doordat de samenstelling van de set aan predictoren verschilt bij beide benaderingen, verschillen ook de in de regressievergelijking opgenomen variabelen. Zo is bij de benadering volgens Kirsch en Mosenthal soms de vraagvariabele "aantal OC's dat verwerkt moet worden voor de taak" een voorspeller. Bij de alternatieve variabelen blijkt de mate van overeenkomst tussen vraag en tekst een robuuste predictor.

Ook al zijn er verschillen in de significante predictoren per variabelenset, de totale hoeveelheid variantie in de itemmoeilijkheid, die door dit soort formele aspecten verklaard wordt, bedraagt in beide gevallen ongeveer de helft. De beide predictorsets zijn dus even goed in het voorspellen van de itemmoeilijkheid. Dit is ook niet zo verwonderlijk, aangezien de alternatieve variabelen gemodelleerd zijn naar de variabelen van Kirsch en Mosenthal en betrekking hebben op dezelfde aspecten van de taken (zoals inderdaad blijkt uit de canonische correlaties). Uit het huidige onderzoek blijkt echter dat het niet cruciaal is welke van beide predictorsets gebruikt wordt. De specifieke steekproef van taken is wellicht van groter belang voor de hoeveelheid variantie in de itemmoeilijkheid die door de predictoren verklaard wordt. Dat was overigens ook al de reden waarom er geprobeerd was de takenset zo veel mogelijk te doen gelijken op de takenset van de NAEP-studie. Zo was er onder andere voor gekozen de spreiding van p-waarden zo veel mogelijk hetzelfde te laten zijn. Een verschil in de te verklaren hoeveelheid variantie tussen beide takensets kan daarom geen verklaring zijn voor het verschil in predictie tussen het huidige onderzoek en de resultaten van Kirsch en Mosenthal.

Samenhangend met het vorige punt is de spreiding van de predictoren in de set. Er waren in de huidige studie minder taken en meer items per taak dan bij Kirsch en Mosenthal. De spreiding in materiaalvariabelen is daardoor in de huidige verzameling items minder dan bij Kirsch en Mosenthal. Toch is er een significante bijdrage van de materiaalvariabelen in onze itemset: evenals bij Kirsch en Mosenthal is het aantal SPE's in het document of het aantal zinnen een significante voorspeller en tevens de enige robuust voorspellende materiaalvariabele.

De waarden die de predictoren aannemen, liggen vast voor een bepaalde takenset, maar de criterium-variabele, de p-waarde, hangt af van de vaardigheidsverdeling binnen de groep respondenten. Het verschil in predictie van de NAEP-taken en de Voorstudie-taken kan misschien voor een deel verklaard worden door de respondentgroep waarbij de p-waarden verkregen zijn. De gebruikte itemset in de huidige studie bestaat uit gegevens die bij 12-jarigen aan het eind van de basisschool verzameld zijn. De spreiding in p-waarden voor deze taken zal bij volwassenen vermoedelijk anders zijn en dat zou de voorspelling kunnen beïnvloeden. Deze veronderstelling wordt aannemelijk door de volgende observatie. Drie van de taken die hier zijn gebruikt, zijn in een later onderzoek onder 15-jarigen nogmaals gebruikt (Kuhlemeier & Van den Bergh, 1989). Dit zijn taak 2, 6 en 7 uit Tabel 2. Voor de in totaal 24 items bij de 3 taken was de correlatie tussen de p-waarden in de twee groepen leerlingen .82. Dit is vrij hoog, maar het geeft toch aan dat slechts 67% van de variantie in p-waarden gedeelde variantie is.

Naast een algemene toename in kennis en ervaring met verschillende documenten met de leeftijd, kan scholing ook tot specifieke kennis en ervaring leiden die tot een verschuiving in p-waarden kan leiden. Te denken valt bijvoorbeeld aan opgedane bekendheid met bepaalde woor-

den. In het NAEP onderzoek was een substeekproef aanwezig van personen met relatief weinig (maximaal 8 jaar) scholing en bij een aparte analyse op deze groep werd slechts 56% van de variantie in de p-waarden voorspeld door slechts één significante predictor, namelijk type informatie. Niet qua leeftijd, maar wel qua scholing is onze steekproef van 8e-groepers het best vergelijkbaar met deze substeekproef en dan blijkt de predictie opmerkelijk goed overeen te komen.

Bij de groep 12-jarigen die de Voorstudie taken gedaan heeft, wordt de helft van de variantie in de itemmoeilijkheid verklaard door de gebruikte formele aspecten van de taak. De andere helft van de variantie in itemmoeilijkheid in deze taken hangt klaarblijkelijk af van andere dan de in onze analyses betrokken factoren. Hiervan zijn hier boven al voorbeelden gegeven. Van belang kan zijn de bekendheid met moeilijke woorden in de tekst of de bekendheid met het soort document of zelfs de aantrekkelijkheid van de taak. Dit zijn dus eigenschappen van de populatie die de taken krijgt voorgelegd. Tenzij al dit soort variabelen kunnen worden gecontroleerd in de populatie of als al deze mogelijke bronnen van variantie worden opgenomen in de predictorenset, zal de hoeveelheid verklaarde variantie in de itemmoeilijkheid in onbekende mate structureel beperkt zijn.

Ofschoon de resultaten van de Amerikaanse studie zeer veelbelovend leken en de moeilijkheidsparameters die Kirsch en Mosenthal vonden aannemelijk geacht werden, zijn er blijkbaar nog andere belangrijke bronnen van variantie waarmee rekening gehouden moet worden, voordat er conclusies getrokken kunnen worden over de kenmerken die de moeilijkheidsgraad van functionele leestaken bepalen.

Het belang van verder onderzoek op dit terrein is erin gelegen dat pas als men de variabelen die de prestatie op functionele leestaken bepalen geïdentificeerd heeft, bepaald kan worden of de op school onderwezen vaardigheden hier in voldoende mate op aansluiten. In het onderwijs komt het namelijk niet zo vaak, of helemaal niet, voor dat aan functionele leestaken expliciet aandacht wordt besteed. Op school krijgt men leesvaardigheid aangeleerd die vooral van belang is in een schoolse context en voor een goed vervolgen van het onderwijs. Na de instructie in technisch lezen wordt voornamelijk de vaardigheid in het kritisch lezen van informatieve teksten en het maken van toetsen voor tekstbegrip geoefend. Dit soort leesvaardigheid wordt in contrast met functionele leesvaardigheid soms wel schoolse leesvaardigheid genoemd. Het is natuurlijk van het grootste belang na te gaan in hoeverre deze op school aangeleerde vaardigheden van belang zijn voor het maken van in het dagelijkse en na-schoolse leven voorkomende, functionele leestaken. Dit is de vraag naar de transferwaarde van op school geleerde leesvaardigheden. Als er van uitgegaan wordt dat de potentiële transfer van vaardigheden over taken afhangt van de overeenkomst tussen die taken wordt duidelijk dat men allereerst dient na te gaan wat de kenmerkende aspecten van de prestatie op allerlei functionele leestaken zijn. Pas als deze geïdentificeerd zijn, kan worden bepaald in welke mate deze vaardigheden ook in "schools" lezen aan bod komen of aan bod zouden moeten komen.

NOTEN

1. Dit onderzoek werd uitgevoerd in het kader van een NWO-project (PEDON nr. 590-291-001).
2. Omdat er sprake is van niet-normaal verdeelde variabelen zijn er enkele normaliserende transformaties toegepast op de meest scheef verdeelde variabelen. Zo is de vierkantswortel genomen uit het aantal OC's in het document, het aantal OC's dat verwerkt moet worden voor de vraag en het aantal SPE's dat verwerkt moet worden voor de vraag en het logaritme is genomen uit het aantal SPE's in het document. Het aantal ingebedde OC's en SPE's uit het document is eerst omgezet als proportie van het totaal aantal OC's en SPE's in het document en daarna is hier het natuurlijk logaritme van genomen. Ook is de boogsinus-transformatie van de p-waarde genomen. Met deze deels getransformeerde variabelenset zijn dezelfde selectieprocedures genomen als voor de niet-getransformeerde variabelen, te weten preselectie, stepwise, "bewezen variabelen" geforced en forward. De resultaten kwamen zeer sterk overeen met de uitkomsten van de analyses op de niet-getransformeerde variabelen.

3. Ook bij de alternatieve variabelen zijn transformaties toegepast. De vierkantswortel is genomen van de hoeveelheid woorden en het aantal zinnen en van de p-waarde is weer de boogsinus-transformatie genomen. De resultaten van de regressieanalyses zijn vrijwel identiek aan de resultaten van de analyses op de ongetransformeerde variabelen.

LITERATUUR

- Davey, B. (1988). Factors affecting the difficulty of reading comprehension items for successful and unsuccessful readers. *Journal of Experimental Education*, 56(2), 67-76.
- Davis, F.B. (1968). Research in comprehension in reading. *Reading Research Quarterly*, 3, 499-545.
- Van den Bergh, H., Baltzer, J.E., Marschall, F.T., Triesscheijn, B., & Wesdorp, H. (1986). *Instrumenten ter peiling van de taalvaardigheid van leerlingen uit de derde klas V.O.* (SCO-rapport nr. 71). Amsterdam: SCO.
- Drum, P.A., Calfee, R.C., & Cook, L. K. (1981). The effects of surface structure variables on performance in reading comprehension tests. *Reading Research Quarterly*, 16, 486-514.
- Duffy, T. M. (1985). Readability formulas: what's the use? In: T.M. Duffy, & R. Waller (Eds.), *Designing usable texts*. London: Academic Press.
- Embretson, S.E., & Wetzel, C.D. (1987). Component latent trait models for paragraph comprehension tests. *Applied Psychological Measurement*, 11(2), 175-193.
- Geus, W.C. de, Reitsma, P. & Glopper, K. de (1992). Task characteristics contributing to difficulty of functional reading tests. In: Tj. Plomp, J.M. Pieters, & A. Feteris. *European Conference on Educational Research. Book of Summaries*. (pp. 298-301). Enschede: University of Twente.
- Guthrie, J.T. (1988). Locating information in documents: examination of a cognitive model. *Reading Research Quarterly*, 23(2), 178-199.
- Kádár-Fülöp, J. (1985). *The CTD reading study*. Manuscript for evaluation in education: An international review series, 9(2).
- Kirsch, I.S., & Guthrie, J.T. (1980). Construct validity of functional reading tests. *Journal of Educational Measurement*, 17(2), 81-93.
- Kirsch, I.S., Jungeblut, A., & Campbell, A. (1992). *Beyond the school doors. The literacy needs of job seekers served by the U.S. Department of Labor*. Princeton, NJ: Educational Testing Service.
- Kirsch, I.S., & Jungeblut, A. (1986). *Literacy: Profiles of America's Young Adults*. (Report no. 16-PL-02). Princeton, N.J.: Educational Testing Service.
- Kirsch, I.S., & Mosenthal, P.B. (1990). Exploring document literacy: variables underlying the performance of young adults. *Reading Research Quarterly*, 25(1), 5-30.
- Krceft, H.P.J. (1985). *Een beschrijvingsschema voor tekstbegripvragen*. (Interne documentatie nr 152). Arnhem: Cito.
- Kuhlemeier, J.B., & Van den Bergh, H. (1989). *De Proefpeiling Nederlands: een onderzoek naar de haalbaarheid van peilingsonderzoek in het voortgezet onderwijs*. (Specialistisch bulletin nr.74). Arnhem: Cito.
- Mosenthal, P.B. (1985). Defining the expository discourse continuum. Towards an taxonomy of expository text types. *Poetics*, 14, 387-414.
- Mosenthal, P.B., & Kirsch, I.S. (1991). Toward an explanatory model of document literacy. *Discourse Processes* 14, 147-180.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall.
- Scheuneman, J.D., & Gerritz, K. (1990). Using differential item functioning procedures to explore sources of item difficulty and group performance characteristics. *Journal of Educational Measurement*, 27(2), 109-131.
- Sheehan, K., & Mislevy, R.J. (1990). Integrating cognitive and psychometric models to measure document literacy. *Journal of Educational Measurement*, 27(3), 255-272.
- Siegel, S., & Castellan, N.J. (1988). *Nonparametric statistics for the behavioral sciences*. New York: McGraw Hill.
- Stevens, J. (1992). *Applied multivariate statistics for the social sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Wesdorp, H. (1978). *Evaluatietechnieken voor het moedertaalonderwijs. Een inventarisering van beoordelingsmethoden voor de stelvaardigheid, het begrijpend lezen, de spreek-, luister- en discussievaardigheid*. Amsterdam: Stichting Research Instituut voor toegepaste psychologie aan de U.v.A.
- Wesdorp, H., Van den Bergh, H., Bos, D.J., Hoeksma, J.B., Oostdam, R.J., Scheerens, J., & Triesscheijn, B. (1986). *De haalbaarheid van periodiek peilingsonderzoek. Een voorstudie op het gebied van het taalonderwijs in de lagere school*. (SVO-reeks 85). Lisse: Swets & Zeitlinger.

Zwarts, M., Janssens, F., Rijlaarsdam, G., Veldhuijzen, N., Wesdorp, H., & Wolfhagen, I. (1990). *Balans van het taalonderwijs aan het einde van de basisschool*. Uitkomsten van de eerste taalpeiling einde basisonderwijs. (PPON-reeks nr. 2). Arnhem: Cito.

BIJLAGE 1 SCORING VAN DE BEOORDELINGSSCHALEN

1. *De structuur van het materiaal (1-5):*
 - 1) korte tekst met organisatie (zoals kopjes, drukstijlen etc.) of simpele tabel of grafiek
 - 2) korte tekst zonder organisatie
 - 3) lange tekst zonder organisatie of tabel of grafiek met 1 geneste rij of kolom
 - 4) lange tekst met organisatie
 - 5) complexe tabel of grafiek
2. *De hoeveelheid informatie*
Tel het aantal woorden en getallen
3. *Het aantal zinnen* (klassiek, maar aangepast voor tabellen):
Tel het aantal zinnen. Ook kopjes en losse woorden in tekst en "de vakjes" in tabellen of de kruisingen in grafieken worden als zin aangemerkt.
4. *De organisatie van het document (0-5):*
 - 0) weinig of geen tekst, bijvoorbeeld bij een tabel of grafiek
 - 1) doorlopende tekst zonder alinea-indeling of korte zinnen zonder ordening/opbouw
 - + 1: Bij aanwezigheid van alinea-indeling of ordening (bijvoorbeeld inspringing of nummering, of tabelvorm)
 - + 1: Bij aanwezigheid van kopjes behalve de titel of kolom- en rijnamen
 - + 1: Bij aanwezigheid van verschillende drukstijlen (cursief, dik, in hoofdletters of onderstreept gedrukte woorden)
 - + 1: Bij aanwezigheid van voorbeelden, schema's en/of tekeningen die de organisatie of het gebruik van het document toelichten
5. *Benodigde documentinformatie (1-6):*
 - 1) Alleen eigen kennis van de lezer is nodig
 - 2) Het antwoord staat in een zin of zinsdeel of binnen een rij of kolom
 - 3) Het antwoord staat in meerdere aansluitende zinnen in de tekst of meerdere aansluitende rijen of kolommen van een tabel óf zowel rij- als kolominformatie moet verwerkt worden om het antwoord te bepalen
 - 4) Een combinatie van meerdere niet-aansluitende gedeelten van het document moet verwerkt worden om het antwoord te bepalen
 - 5) De tekst als geheel of het document als geheel moet verwerkt worden om het antwoord te bepalen
 - 6) De vraag kan alleen beantwoord worden door het verwerken van meerdere tekst- of documentdelen in combinatie met vooraf aanwezige kennis die alleen ingewijden hebben
- 6) *Mate van correspondentie tussen woorden en typografie (1-5):*
 - 1) De belangrijkste kopjes in het document lijken sterk op de sleutelwoorden in de vraag en zijn ook onderscheiden in locatie en/of typografie
 - 2) De belangrijkste ordening van het document is gerelateerd aan de informatie in de vraag
 - 3) Er zijn ondergeschikte ordeningen, kopjes en typografie en die zijn gerelateerd aan de vraag
 - 4) Het document bevat ordening en kopjes, maar die helpen niet om de vraag te beantwoorden
 - 5) Er zijn geen kopjes of ruimtelijke ordeningen aanwezig

7. *Mate van overeenkomst (1-5):*

- 1) (bijna) letterlijke overeenkomst
- 2) synonieme overeenkomst
- 3) overeenkomst die voornamelijk gemaakt wordt door een eenvoudige afleiding op basis van het materiaal
- 4) overeenkomst die gemaakt wordt doordat de gegeven informatie afgeleid moet worden waarbij zowel probleemoplossen en/of voorkennis als informatie uit het materiaal vereist worden óf overeenkomst die gemaakt moet worden door een complexe afleiding op basis van het materiaal
- 5) overeenkomst wordt bepaald op basis van speciale voorkennis

8. *Verwerkingsdiepte (1-6):*

- 1) antwoord binnen dezelfde informatie-eenheid (zin, rij, kolom) expliciet genoemd óf persoonlijke kennis invullen/geven
- 2) achtereenvolgens localiseren en matchen van steeds een letterlijk of synoniem kenmerk
- 3) eenvoudige afleiding op basis van de tekst maken om het juiste antwoord te bepalen óf het vergelijken van twee of meer kenmerken die zich binnen dezelfde informatie-eenheid bevinden
- 4) twee of meer kenmerken uit verschillende, maar aangrenzende informatie-eenheden identificeren en deze vergelijken; óf het antwoord op de vraag bepalen door op basis van de tekst een complexere afleiding te maken van een kenmerk dat expliciet vermeld is binnen dezelfde informatie-eenheid
- 5) een hele serie matches maken óf twee of meer kenmerken vergelijken die in verschillende en niet-aansluitende document-delen gevat zijn óf een groep verschillende expliciete kenmerken identificeren in één bron (bijvoorbeeld geheugen of een eerste document) en onder verschillende informatie-eenheden invullen in een tweede bron (bijvoorbeeld een tweede document) waarbij een serie eenvoudige afleidingen op basis van de tekst gemaakt moet worden
- 6) een serie matches maken en daarna kiezen tussen twee of meer antwoorden door rekening te houden met een voorwaarde die in een ander stuk tekst gevonden wordt óf als het antwoord op de vraag alleen gegeven kan worden op basis van speciale voorkennis

9. *Het aantal distractors (1-7):*

- 1) geen distractors
- 2) 1 distractor ver verwijderd van de zin of paragraaf, rij of kolom waar het juiste antwoord in staat
- 3) 1 distractor dichtbij het juiste antwoord (in dezelfde zin of paragraaf, rij of kolom)
- 4) 2 of 3 distractors ver verwijderd van de zin of paragraaf, rij of kolom waar het juiste antwoord in staat
- 5) 2 of 3 distractors dichtbij het juiste antwoord (in dezelfde zin of paragraaf, rij of kolom)
- 6) 4 of meer distractors ver verwijderd van de zin of paragraaf, rij of kolom waar het juiste antwoord in staat
- 7) 4 of meer distractors dichtbij het juiste antwoord (in dezelfde zin of paragraaf, rij of kolom)

Leereffecten van stages in sociaal-wetenschappelijke studierichtingen

R.F. Poell en M.J.M. Voeten

*Vakgroepen Algemene Pedagogiek en Onderwijskunde, Katholieke Universiteit Nijmegen**

ABSTRACT

Evaluating learning effects of practical training periods in university education has gained little attention in educational research so far. The aim of this study is to identify the factors that contribute to learning effects of practical training in higher education. On the basis of research on practical training in secondary and higher vocational education a conceptual model consisting of possible learning effects and their determinants was set up. Questionnaires derived from this model were completed by students in educational science (N=87) at the end of their practical training period and by their supervisors (N=16). Learning effects were measured by comparing post-test scores on 55 goal statements to retrospective pre-test scores. Data about the determinants of learning effects included the students' assessments of themselves, of characteristics of organisations offering practical training and of the preparation and guidance provided by the university. The study resulted in a list of learning goals for practical training deemed important by students as well as their supervisors. Statistically significant learning effects were reported for all these goals. Learning effects were correlated with a large number of factors in the three domains of determinants. The questionnaires appeared to be a valuable instrument for assessing learning effects of practical training periods in university education.

INLEIDING

De stage bevindt zich op het snijvlak van het onderwijssysteem en het arbeidsbestel, twee maatschappelijke systemen die tegenwoordig duidelijk onderscheiden zijn. In vroeger tijden was deze scheiding minder pregnant. De stage is ontstaan als antwoord op het steeds verder uit elkaar groeien van onderwijs en arbeid, met name na de industriële revolutie die steeds verdergaande arbeidsdeling en ongeschoolde arbeidstaken met zich meebracht. De Vries (1988, p. 99-126) beschrijft hoe theoretisch onderwijs hierdoor steeds meer institutioneel los is komen te staan van de beroepspraktijk. Hij ziet de stage als een corrigerende reactie op de tekortkomingen die deze ontwikkeling met zich meebracht, als een terugkeer naar de praktijk. De centrale vraag, waarop in de loop der tijden verschillende antwoorden worden gegeven, luidt in welke mate de voorbereiding op een beroep in de beroepspraktijk zelf plaats moet of kan vinden en in welke mate dat in afzonderlijke instituties het geval moet of kan zijn. Duidelijk is in ieder geval dat zowel een beroepsopleiding die uitsluitend uit theoretisch onderwijs bestaat als een die uitsluitend uit leren door werken bestaat, belangrijke nadelen kennen. In Nederland is de combinatie van beide via schools onderwijs met stages dan ook dominant.

Ook in het universitair onderwijs heeft de stage een belangrijke plaats. In vele studierichtingen komt de stage in een of andere vorm voor als een verplicht studie-onderdeel. In deze bijdrage richten wij ons op de stage zoals die feitelijk voorkomt in twee verwante studierichtingen, namelijk onderwijskunde en pedagogiek. Kenmerkend van deze stages is dat zowel de voorbereiding op de beroepspraktijk als het uitvoeren van wetenschappelijk onderzoek deel ervan kunnen uitmaken.

De stage is niet de enige, maar wel de meest voorkomende vorm van buitenschools leren. Van belang is dat de stage moet bijdragen aan het realiseren van doelstellingen in het kader van een opleiding, dat die functie boven het belang van de stagebiedende organisatie moet gaan en

* Adres: Postbus 9103, 6500 HD, Nijmegen.

dat er sprake dient te zijn van begeleiding, wat blijkt uit onderstaande hanteerbare definitie van stage (Academische Raad, 1981, p. 4): "Een stage in het wetenschappelijk onderwijs wordt gekenmerkt door:

- werkzaamheden gedurende een al dan niet aaneengesloten periode
- door een student van een universiteit of hogeschool
- in een stagebiedende organisatie die een andere is dan de faculteit of afdeling en die medeverantwoordelijkheid draagt voor de stage
- op grond van afspraken tussen de onderwijsinstelling (faculteit of afdeling), stagebiedende organisatie en stagiair t.a.v. onder meer inhoud, begeleiding, verslaglegging en rechtspositionele consequenties (werktijd, aansprakelijkheid, betaling) van de stage
- in het kader van het door de faculteit of afdeling vastgestelde onderwijsprogramma en onder eindverantwoordelijkheid van de faculteit of afdeling inzake de onderwijskundige doelmatigheid van de stage
- begeleid door zowel de stagebiedende organisatie als de onderwijsinstelling (faculteit of afdeling)."

Onze studie heeft betrekking op opleidende stages aan het einde van een wetenschappelijke opleiding, die zowel intern (overwegend onderzoekstages) als extern (overwegend praktijkstages) kunnen plaatsvinden en tot doel hebben: "het al-doende-leren van een bepaald beroep of vak, waarbij het vooral gaat om het leren toepassen in de praktijk van kennis, inzichten en vaardigheden die op school zijn geleerd." (Ministeries van Onderwijs en Wetenschappen en Sociale Zaken, 1981, p. 21). De opleidende stage onderscheidt zich hierin van de oriënterende en de voorbereidende stage, die veeleer bedoeld zijn om kennis te maken met (aspecten van) een nog tamelijk onbekend beroep of werkterrein.

In het wetenschappelijk onderwijs wordt de stage in veel gevallen, echter lang niet altijd, gekoppeld aan het schrijven van een onderzoeksscriptie over het onderwerp van de stage. De scriptieperiode wordt in deze studie beschouwd als onderdeel van de stage, die om deze reden zou kunnen worden betiteld als 'afstudeerproject'. Vanwege het dominante gebruik van de term 'stage' handhaven wij deze echter. De vraag of stage en scriptie aan elkaar gekoppeld dan wel van elkaar losgekoppeld zijn is als variabele in het onderzoek opgenomen.

Doelstelling

In het curriculum van veel wetenschappelijke studierichtingen is een stage opgenomen (Van den Bijtel, 1988). Over het algemeen neemt deze een aanzienlijke periode in beslag, die kan oplopen tot acht à negen maanden binnen een vierjarig programma. In schril contrast met de aandacht die de stage in deze zin ten deel valt, staat de mate waarin aandacht wordt besteed aan de vraag naar de leeropbrengst van de stage. Het gaat daarbij om de vraag wat studenten feitelijk leren van een stage. Komen die leereffecten wel overeen met hetgeen de opleiding beoogt met het opnemen van een stage in het curriculum? Aan deze vragen wordt nauwelijks aandacht geschonken. Mogelijk komt de vraag informeel aan de orde bij de afsluiting van de stage door de begeleider(s) en de stagiair. Systematisch onderzoek naar deze materie is nauwelijks verricht, althans wat betreft het wetenschappelijk onderwijs (De Vries & Van Wel-Berretty, 1986). Wel wordt bij uitstroom- en aansluitingsonderzoek soms gevraagd naar de functie die de stage heeft vervuld bij het vinden van een baan. De leerinhoud van de stage komt daarbij niet aan bod (Veen & Voorthuis, 1989; Van der Velden, 1990; Bevers & Verrijt, 1989). De Vries (1988, p. 9) komt tot dezelfde conclusie, met dien verstande dat er in MBO en HBO wel degelijk onderzoek naar leereffecten van stages is verricht.

Wat het wetenschappelijk onderwijs betreft, is onderzoek naar de stage het meest te vinden op het gebied van de lerarenopleidingen en de 'professional education' (Dinham & Stritter, 1986), met name het medisch onderwijs. Het overzichtsartikel van Dinham en Stritter is vrijwel geheel gewijd aan de praktijkcomponent van opleidingen met betrekking tot de professies. Ook voor deze terreinen geldt echter dat er nog veel onderzoek nodig is naar wat studenten leren van hun stage-ervaringen en naar de optimale onderwijskundige vormgeving van de stage. Dinham en Stritter eindigen hun artikel dan ook met aanbevelingen voor een research-agenda. Daarbij

gaat het niet zozeer om theorievorming, als wel om de voorwaarden daartoe: "define the essential elements or the essential factors in the young professional's education" (Dinham & Stritter, 1986, p. 965). Het recent verschenen proefschrift van Wolfhagen (1993) handelt over de constructie en validering van een meetinstrument om de kwaliteit te meten van stages in de tweede fase van de medische opleiding (de co-assistentenschappen) aan de Rijksuniversiteit Limburg. Dit instrument is uitsluitend gericht op het onderwijsleerproces, op de voorwaarden waaronder leren plaatsvindt, en niet op de leereffecten. Zij constateerde dat het gebruik van evaluatiegegevens om klinische stages te verbeteren nog in de kinderschoenen staat (Wolfhagen, 1993, p. 143). Op basis van haar onderzoek komt zij tot de conclusie dat systematische evaluatie van stages mogelijk is en kan bijdragen aan kwaliteitsbewaking en -verbetering.

Wij willen in deze explorerende studie een bijdrage leveren aan de ontwikkeling van een instrument voor de systematische analyse van stages in het wetenschappelijk onderwijs. Met een dergelijk instrument kan kritisch worden gekeken naar dit belangrijke onderdeel van de wetenschappelijke opleiding. Om dit doel te bereiken is gepoogd de ontwikkelde theorievorming omtrent stages in het beroepsonderwijs toe te passen op de situatie in het wetenschappelijk onderwijs. We gaan hierbij uit van de bestaande opleidingssituatie: de stage als onderdeel van de opleiding. Meer algemene vragen naar de wijze waarop de praktijkcomponent van een wetenschappelijke opleiding zou moeten worden ingericht, blijven hier buiten beschouwing.

Probleemstelling

De probleemstelling van het onderzoek is: Welke factoren beïnvloeden het optreden van leereffecten van stages binnen een wetenschappelijke opleiding? Ter beantwoording van deze hoofdvraag kunnen meerdere deelvragen worden geformuleerd: Wat zijn de belangrijkste leerdoelen van een stage? In welke mate worden de leerdoelen van een stage door stagiairs bereikt? In hoeverre hebben kenmerken van de stageplaats, van de stagiair en van de opleiding invloed op de mate waarin de leerdoelen worden bereikt?

THEORETISCH KADER

Het antwoord op de eerste deelvraag, naar de belangrijkste leerdoelen van de stage, wordt gekleurd door de opvatting die betrokkenen hanteren in het kader van de aansluitingsproblematiek tussen onderwijs en arbeid. Geurts (1989, p. 14) schetst in dit verband wat hij noemt de 'enge' en de 'brede' visie op de aansluitingsproblematiek. De enge visie, die overheersend is, wordt gekenmerkt door:

- de aanname dat oorzaken van en oplossingen voor de discrepanties tussen vraag en aanbod exclusief aan de kant van het onderwijs te vinden zijn;
- de veronderstelling dat er een stelsel van min of meer exclusieve en rechtlijnige relaties tussen opleiding en beroepen bestaat. Iedere opleiding wordt geacht een eigen domein van beroepen te hebben;
- het hanteren van een beperkt kwalificatiebegrip. Kwalificaties worden vooral in technisch-instrumentele zin opgevat en beschouwd als een produkt van uitsluitend het onderwijs;
- een statische kijk op aansluitingsproblemen. De wijze waarop opgeleiden momenteel in het arbeidsproces worden ingeschakeld, wordt vaak als norm gehanteerd. Veranderingen in vraag en aanbod en aanpassingsprocessen tussen beide worden zo buiten beschouwing gelaten."

In deze visie richt het onderwijs zich volledig op de behoeften van de arbeidsmarkt. Dat kan via aanpassing van curricula en dus ook via de keuze van bepaalde (leer)doelen voor de stage binnen de opleiding. De functie van de stage is in deze visie met name het opdoen van technisch-instrumentele kennis, inzichten en vaardigheden passend bij een specifiek beroep(enveld). De opleiding dient zich, ook via de stage, aan te passen aan de bestaande gang van zaken in het arbeidsproces. Er is weinig ruimte voor kritische en innovatieve kwalificaties.

Aan de andere kant van het continuüm staat de brede visie op aansluitingsproblematiek. Deze visie erkent dat aansluitingsproblemen voortvloeien "uit eigen ontwikkelingen van het onder-

wijs, eigen ontwikkelingen van de arbeid en ook uit eigen ontwikkelingen van de arbeidsmarkt, de ontmoetingsplaats van vraag en aanbod". Dit model is complexer dan de enge visie. De eisen die vanuit het arbeidssysteem worden gesteld aan de kwalificaties van werknemers (en afgestudeerden) veranderen onder invloed van technologische en organisatorische ontwikkelingen zo snel, dat er steeds meer een beroep wordt gedaan op het aanpassingsvermogen en de flexibiliteit van die personen. In deze visie heeft de opleiding (en daarbinnen de stage) de taak om de flexibiliteit van toekomstige arbeidskrachten te versterken. Dat wil zeggen dat er meer nadruk wordt gelegd op sociaal-normatieve kwalificaties, die van toepassing kunnen zijn op meer dan één beroep(endomein). De gedachte is dat technisch-instrumentele kwalificaties relatief eenvoudig en snel via functiegerichte bedrijfsopleidingen kunnen worden bijgebracht, wanneer een degelijke sociaal-normatieve basis is verworven. Erkend wordt dat onderwijs en arbeid elkaar wederzijds (dienen te) beïnvloeden. Het onderwijs kan via het bijbrengen van kritische en innovatieve kwalificaties invloed uitoefenen op het arbeidssysteem. Bij uitstek in de stage komen deze aspecten tot uiting. De Vries (1988, p. 73) noemt de HBO-V als sprekend voorbeeld van een opleiding die actief poogt via de stages de beroepspraktijk te vernieuwen. Uiteraard heeft dit verregaande invloed op datgene wat de studenten van een dergelijke opleiding tijdens hun stage zullen ervaren en leren.

Opleidingen die de brede visie toepassen zullen zich met name richten op het bijbrengen van breed toepasbare beroepskwalificaties (De Jong, Moerkamp, Onstenk & Babeliowsky, 1990). Het gaat dan om beroepsmatige kennis, inzichten, vaardigheden en houdingen die in een veelheid van situaties zijn toe te passen, die met andere woorden transferwaarde bezitten naar andere functiesituaties en naar de verdere competentie-ontwikkeling van de beginnend beroepsbeoefenaar. Men kan daarbij denken aan sleutelkwalificaties als zelfstandigheid, probleemoplossen, initiatief nemen, besluitvaardigheid, creativiteit, uitdrukkingsvermogen, planmatig kunnen denken etcetera (Van Zolingen, 1991; Van Zolingen & Van Onna, 1992). Tegenover deze sociaal-normatieve kwalificaties, gericht op het kunnen uitvoeren van werkzaamheden die los kunnen worden gezien van één specifiek beroepenveld, staan de technisch-instrumentele kwalificaties, die verwijzen naar het in staat zijn werkzaamheden binnen één specifiek beroepenveld te verrichten. Overigens merken Van den Berg en Van den Tillaart (1986, p. 18) op dat de grens tussen technisch-instrumentele en sociaal-normatieve kwalificaties niet altijd scherp te trekken is. De overlap zit met name op het gebied van sociale en communicatieve vaardigheden, die in bepaalde (met name leidinggevende) beroepen en functies tot de technisch-instrumentele, en in andere tot de sociaal-normatieve kwalificaties moeten worden gerekend. In navolging van Van den Berg en Van den Tillaart worden zij hier in het sociaal-normatieve domein geplaatst (zie voor verdere discussie Berretty, 1981, p. 9; Van den Dool, 1987, p. 40).

Bij de beantwoording van de tweede deelvraag naar de mate waarin leerdoelen van stages worden bereikt komt het begrip 'leereffect' centraal te staan. Er is een nauwe relatie tussen de termen 'leerdoelen' en 'leereffecten'; in de literatuur worden de twee termen vaak zelfs door elkaar heen gebruikt (Steinen, 1985, p. 4). Toch is er een wezenlijk verschil. In een leerdoel drukt men meer of minder specifiek uit wat men verwacht dat een student op een gegeven moment zou moeten kunnen of weten. Wanneer de student dat leerdoel heeft bereikt is er sprake van een leereffect. Leerdoelen zijn beoogde of verwachte leereffecten; leereffecten zijn bereikte leerdoelen. Uit deze beschrijving spreekt een intentionele opvatting: men stelt een leerdoel vast en via een gericht leerproces komt men tot een leereffect. In het algemeen zijn vier kanttekeningen bij dit intentionele model te plaatsen. Ten eerste kunnen zich niet-intentionele (neven)effecten voordoen, leereffecten die niet waren gepland. Ten tweede is het goed te bedenken dat bepaalde leerdoelen om wat voor reden dan ook niet zullen worden bereikt, zodat van sommige geplande leereffecten geen sprake zal zijn. Ten derde kan er sprake zijn van leerdoelen (en leereffecten) uit een 'verborgen curriculum' (Meijnen, 1987, p. 58), waarbij niet-geëxpliciteerde waarden, normen en opvattingen worden overgedragen. Ten vierde kan er sprake zijn van ongewenste leereffecten, ook te omschrijven als 'verleren' (Kraayvanger & Van Onna, 1985, p. 28). Deze negatieve leereffecten kunnen zich voordoen wanneer op bepaalde aanwezige vaardigheden, inzichten en kennis tijdens de stage weinig of geen beroep wordt gedaan.

Weliswaar kunnen deze effecten zich in elk leerproces voordoen, de stage neemt in dit opzicht een bijzondere plaats in aangezien zij slechts in beperkte mate als een gericht leerproces te beschouwen is. Hier komt de derde deelvraag in beeld, namelijk die naar de invloed van opleidings-, stageplaats- en stagiairkenmerken op leereffecten. De stage vindt plaats in een arbeidscontext, die vaak weinig vanuit onderwijskundige principes structureerbaar is. Hier doet zich een interessante paradox voor: enerzijds wordt de stage als onderdeel van de opleiding geacht bij te dragen aan het bereiken van de opleidingsdoelen bij de student, anderzijds heeft de opleiding weinig mogelijkheden om de gang van zaken tijdens de stage te sturen in de gewenste richting. De Vries (1988, p. 150) stelt zelfs vast dat pogingen om leren in een arbeidscontext te stimuleren en te sturen via schoolse principes ertoe kunnen leiden dat dit leren veel minder of helemaal niet plaatsvindt. Teveel sturing vanuit onderwijskundige principes kan dus een averechts effect opleveren. Naast deze opleidings- en stageplaatskenmerken worden kenmerken van de individuele stagiair genoemd als van invloed op het bereiken van leerdoelen. Ook op deze kenmerken kan de opleiding slechts beperkte invloed uitoefenen, hoogstens bij de bepaling of studenten 'klaar zijn' voor de stage. Dit gebeurt naar alle waarschijnlijkheid weinig expliciet; de stage lijkt dan ook een tamelijk autonoom proces dat met name door arbeidsgestructureerde principes wordt gestuurd.

Voorlopig analysemodel

Onstenk, Moerkamp, Voncken en Van den Dool (1990) rapporteren onderzoek naar het verloop van leerprocessen bij stagiairs in het MBO. Hun onderzoek leidde tot een conceptueel model waarin stageplaats- en opleidingskenmerken tezamen met stagiairkenmerken in verband worden gebracht met leerprocessen en leereffecten. Aangezien dit het meest uitgewerkte model is dat we in de literatuur aantreffen, namen we dit als uitgangspunt. Het werd aangevuld met bevindingen uit andere onderzoeken (Kok, 1981; De Vries, 1988; Nieuwenhuis, 1991; Mertens, 1981; Van den Dool, 1987; De Jong et al, 1990; Van Puffelen, Van Paridon & Van Ree, 1985; Steinen, 1985; Van Onna, 1981). Daarbij pasten we het model aan voor de opleidende stage in het wetenschappelijk onderwijs. De leerprocessen werden grotendeels buiten beschouwing gelaten; zij komen deels voor in de vorm van stageplaats-, opleidings- en stagiairkenmerken. Het resultaat was een lijst van 44 leerdoelen, ingedeeld in 12 categorieën, en een aanzienlijk aantal relevant geachte opleidingskenmerken, stagiairkenmerken en stageplaatskenmerken. Tevens werden de leerdoelen verdeeld in sociaal-normatieve en technisch-instrumentele leerdoelen. Zo ontstond een voorlopig analysemodel (Poell, 1992), dat als basis diende voor de constructie van een vragenlijst. Op basis van de onderzoeksresultaten werd dit analysemodel bijgesteld (zie voor dit bijgestelde model Figuur 1).

METHODE

Steekproef

De *onderzoekseenheden* in dit onderzoek waren studenten in het wetenschappelijk onderwijs die in het kader van hun opleiding een stage uitvoeren of hebben uitgevoerd. De steekproef bestond uit alle studenten van de studierichtingen onderwijskunde en algemene pedagogiek van de Katholieke Universiteit Nijmegen, die niet vóór januari 1990 zijn afgestudeerd. Gegevens werden niet alleen bij stagiairs verzameld, maar ook bij hun begeleiders vanuit de opleiding. Doel hiervan was allereerst om antwoord te krijgen op de eerste deelvraag naar het belang van bepaalde leerdoelen voor de wetenschappelijke stage. Daarnaast werd elke begeleider gevraagd om van één recentelijk door hem begeleidde stagiair aan te geven in welke mate deze relevante leerdoelen heeft bereikt (tweede deelvraag). Doel hiervan was om de inschatting wat betreft leereffecten van de stagiair zelf te kunnen vergelijken met die van zijn begeleider, zodat kon worden nagegaan hoe valide het is om uit te gaan van het oordeel van de stagiair alleen.

Het *onderzoeksontwerp* kan aangemerkt worden als een survey onder studenten in het wetenschappelijk onderwijs die een stage hebben uitgevoerd. Het was niet mogelijk een vergelijkbare

controlegroep zonder stage in het onderzoek te betrekken. Een vergelijkbare studierichting zonder stageperiode was niet beschikbaar. Het is overigens niet aannemelijk dat studenten die géén stage uitvoeren leerdoelen bereiken die specifiek zijn voor deze vorm van buitenschools leren. Het gaat hier niet om de vraag hoe de stage als afzonderlijk onderdeel van het curriculum zich verhoudt tot andere manieren om vorm te geven aan de praktijkcomponent van een opleiding. Het geheel ontbreken van een praktijkcomponent is wel een denkbaar maar geen levensvatbaar alternatief.

Meting

De *vragenlijst* bestond uit een aantal open en gesloten vragen om stageplaats-, stagiair- en opleidingskenmerken vast te stellen. Stagiairs en begeleiders werd daarnaast gevraagd van alle leerdoelen het belang aan te geven op een Likert-schaal met de waarden 1 ('niet belangrijk'), 2 ('enigszins belangrijk'), 3 ('belangrijk') en 4 ('zeer belangrijk'). Tot slot werd stagiairs gevraagd aan te geven hoe zij op elk van de leerdoelen voor de stage en na de stage scoorden op een Likert-schaal, met waarden variërend tussen 1 ('zeer slecht') en 7 ('zeer goed'). Een voorbeeld van een leerdoel uit de vragenlijst is: "Aan het einde van de stage moet de stagiair theorie - praktijk verschillen kunnen aangeven."

In februari 1992 zijn interviews met alle 16 docenten verricht. De *respons* onder de begeleiders was 100 %. Aan het begin van april is de vragenlijst verstuurd naar 107 (ex-)stagiairs. Aan het einde van mei 1992 waren, na één herinneringsronde, 87 vragenlijsten ingevuld geretourneerd: een respons van 81 %.

In het onderzoek zijn vier hoofdgroepen van *variabelen* betrokken: drie groepen van onafhankelijke variabelen, namelijk stageplaats-, stagiair- en opleidingskenmerken, en één groep van afhankelijke variabelen, namelijk verschillende soorten leereffecten. In eerste instantie is een meting bij de begeleidende docenten van de stagiairs verricht. Hen is gevraagd het belang aan te geven van alle leerdoelen uit het model. De docenten konden daarnaast relevante leerdoelen aanvullen die zij in de lijst misten. Op grond hiervan is de aanvankelijke lijst van 44 leerdoelen uitgebreid tot een lijst van 55, voornamelijk door de toevoeging van specifieke leerdoelen omtrent onderzoeksvaardigheden. Zoals eerder vermeld is de begeleiders bovendien gevraagd om van één recentelijk door hen begeleidde stagiair aan te geven hoe deze vóór en na de stage op elk van de leerdoelen scoorde. Het verschil in score tussen de voor- en de nameting beschouwen we als het leereffect. Ten slotte beantwoordden de begeleiders een aantal vragen met betrekking tot opleidingskenmerken, zoals het stageverslag en de stagebegeleiding.

De belangrijkste meting vond plaats bij de stagiairs. Hen is eveneens gevraagd aan te geven hoe belangrijk zij de leerdoelen vonden en hoe zij vóór en na de stage op elk van de leerdoelen scoorden (leereffect-meting). Bovendien is hen een aantal vragen voorgelegd om stageplaats-, stagiair- en opleidingskenmerken vast te stellen. Leereffecten werden vastgesteld in de vorm van oordelen van de stagiairs zelf over hun eigen functioneren. De situatie vóór het uitvoeren van de stage werd retrospectief vastgesteld. Bij een retrospectieve voormeting wordt aan personen gevraagd terug te denken naar de situatie vóór een bepaalde gebeurtenis c.q. treatment en een score toe te kennen die van toepassing is op die situatie. Een 'echte' voormeting was om praktische redenen onmogelijk.

Het is evident dat zelf-rapportage een kwetsbare methode is voor het vaststellen van leereffecten, vanwege de mogelijkheid van subjectieve vertekening van de resultaten. In verschillende studies is niettemin aangetoond dat zelf-rapportage een bruikbaar meetinstrument kan zijn. Zie White (1988) voor een overzicht van de literatuur. Dit betekent niet dat geen statistische vertekening te verwachten zou zijn. Met name kan men in onze toepassing verwachten dat de respondenten de behaalde leereffecten te hoog zullen inschatten. Hiervoor konden we in onze onderzoeksopzet niet controleren. Een beperkte check was beschikbaar in de vorm van het oordeel van de begeleider. Dit oordeel betreft echter slechts een klein aantal van de onderzochte stagiairs (N=13). Bovendien kunnen ook de begeleider-oordelen in positieve zin zijn vertekend. Overigens liggen gedragsmaten voor het evalueren van effecten van wetenschappelijke stages niet voor de hand en zijn ook gedragsmaten onderhevig aan vertekening (Sprangers, 1988a, p.

98). Het gebruik van een retrospectieve voormeting in plaats van een 'echte' voormeting heeft als voordeel dat de 'response-shift bias' kan worden vermeden (zie onder meer Hoogstraten, 1982; Sprangers, 1988a, 1988b). Wanneer men leereffecten van stages via zelfrapportage vaststelt, kan zich het probleem voordoen dat de maatstaf waarmee een student zichzelf meet verandert als gevolg van ervaringen opgedaan tijdens de stage. Het is bijvoorbeeld mogelijk dat iemand zich door de stage beter bewust wordt van zijn sterke en zwakke kanten. Het verschil tussen een nameting en een retrospectieve voormeting kan daarom een meer valide meting van het leereffect opleveren dan verschillcores bepaald ten opzichte van een echte voormeting. Overigens zijn alle zelf-rapportages retrospectief, want gebaseerd op vroegere ervaringen (De Meijer, Hoogstraten & Sprangers, 1986). Opname van een 'echte' voormeting naast de retrospectieve voormeting zou hier gewenst zijn geweest, omdat de studenten dan een beoordeling hadden kunnen geven op basis van hun ervaringen tijdens het theoretisch deel van de opleiding. Zoals gezegd was dit om praktische redenen niet mogelijk.

RESULTATEN

Over de beantwoording van de eerste deelvraag: Wat zijn de belangrijkste leerdoelen van een stage? kunnen we kort zijn. Tevoren werd gesteld dat leerdoelen die gemiddeld lager dan 2.50 op de vierpuntsschaal zouden scoren als 'niet belangrijk' uit het onderzoek zouden worden weggelaten. Aan dit criterium gemeten bleken bijna alle leerdoelen voor de stage belangrijk te worden gevonden, zowel door stagiairs als begeleiders. Alleen de leerdoelen 'Beschikken over inzicht in kenmerken van leidinggeven' en 'Zich kunnen identificeren met het beroep' werden onbelangrijk geacht. De categorieën van leerdoelen die als meest belangrijk werden gezien zijn 'Beroepsmatige probleemoplossingsvaardigheden', 'Arbeidssocialisatie', 'Onderzoeksvaardigheden' en 'Persoonlijke ontwikkeling'. Sociaal-normatieve leerdoelen worden zowel door docenten als studenten significant belangrijker geacht dan technisch-instrumentele.

Het feit dat bijna alle leerdoelen belangrijk werden geacht maakt nieuwsgierig naar de antwoorden op de tweede deelvraag: In welke mate worden de leerdoelen van een stage door stagiairs bereikt? Men mag immers verwachten dat op leerdoelen die belangrijk worden gevonden een leereffect wordt geboekt. Leereffecten zijn voor elk leerdoel bepaald door het groeps-gemiddelde vóór de stage af te zetten tegen het groeps-gemiddelde na de stage. Via t-toetsen voor gepaarde waarnemingen werd duidelijk dat voor elk leerdoel een statistisch significant leereffect optrad (zelfs op de leerdoelen die in eerste instantie als onbelangrijk terzijde leken te kunnen worden geschoven). Dit laat natuurlijk onverlet dat individuele stagiairs op individuele leerdoelen géén leereffect of zelfs een negatief leereffect aangaven. Zo werd bij 15 % van de respondenten een negatief leereffect vastgesteld op het leerdoel 'Beschikken over leerplezier en motivatie' en bij 11 % op 'Zelfvertrouwen hebben'. De grootste leereffecten werden gerapporteerd op de categorieën van leerdoelen 'Kennis van de arbeidsorganisatie' en 'Opleiding en beroepspraktijk', zoals blijkt uit Tabel 1. In deze tabel zijn de gemiddelde oordelen van stagiairs over categorieën van leerdoelen weergegeven, waarbij in de meest rechtse kolom het gemiddelde leereffect wordt weergegeven als verschillscore tussen de na- en de voormeting. Tevens is de effectgrootte vermeld, die berekend is door de verschillscore te delen door de standaarddeviatie van de voormeting. De effectgrootte is voor alle categorieën van leerdoelen hoog.

Middels t-toetsen voor gepaarde waarnemingen werd slechts een beperkt aantal verschillen vastgesteld tussen de inschatting van leereffecten door de stagiairs zelf en door hun begeleider. Op zeven van de 55 leerdoelen kwamen op 5%-niveau significante verschillen tussen beide groepen naar voren. Op 10%-niveau werden dat er twaalf. Opvallend was dat begeleiders in alle gevallen een groter leereffect aangaven dan stagiairs zelf. De docenten schatten de situatie vóór de stage op een klein aantal leerdoelen lager in dan de studenten. Ten aanzien van de scores op de nameting was er op 5%-niveau geen enkel significant verschil tussen de studenten en hun begeleider; op 10%-niveau werden 2 significante verschillen gevonden.

Een ander opvallend resultaat is dat het leereffect op sociaal-normatieve leerdoelen signifi-

Tabel 1. Leereffecten op categorieën van leerdoelen.

categorie van leerdoelen	score vóór stage	score na stage	verschilscore na minus vóór
	gemiddelde, sd	gemiddelde, sd	gemiddelde, effectgrootte
Arbeidssocialisatie	4.84 .64	5.47 .51	.59 .94
Inzicht in het beroep	3.92 .95	4.92 1.09	1.00 1.05
Kennis van de arbeidsorganisatie (proces en organisatie)	3.54 .93	5.18 .90	1.62 1.74
Praktische beroepsvaardigheden	4.30 .76	5.32 .55	1.01 1.31
Onderzoeksvaardigheden	4.03 .86	5.19 .81	1.05 1.27
Beroepsmatige omgangsvaardigheden	4.81 .84	5.36 .66	.53 .63
Beroepsmatig inzicht	4.17 .75	4.94 .68	.76 1.01
Beroepsmatige probleemoplossingsvaardigheden	4.56 .91	5.40 .75	.86 .94
Expertise	3.93 .77	4.87 .75	.94 1.22
Eigen toekomst	4.02 1.19	5.12 1.18	1.08 .94
Persoonlijke ontwikkeling	4.77 .63	5.41 .55	.65 1.08
Ontwikkeling van het leervermogen	4.97 .69	5.44 .57	.46 .67
Opleiding en beroepspraktijk	4.17 .77	5.34 .63	1.19 1.53

Noot: N varieert tussen 55 en 85.

cant kleiner is dan dat op de, belangrijker geachte, technisch-instrumentele leerdoelen. Dit gegeven kan worden verklaard doordat de score vóór de stage op sociaal-normatieve leerdoelen hoger is dan die op technisch-instrumentele; het leereffect op de eerste groep neigt daardoor kleiner te zijn.

De derde deelvraag was in hoeverre kenmerken van de stageplaats, van de stagiair en van de opleiding invloed hebben op de mate waarin de leerdoelen worden bereikt. Voor de beantwoording van deze deelvraag is de samenhang geanalyseerd tussen alle afzonderlijke kenmerken enerzijds en de grootte van alle leereffecten anderzijds. Ook werd de samenhang bekeken tussen afzonderlijke kenmerken aan de ene kant en de grootte van leereffecten op categorieën van leerdoelen aan de andere kant, waarvoor het gemiddelde van alle leerdoelen in een categorie werd vastgesteld. Samenhang werd in de meeste gevallen bepaald middels Pearson-correlaties. In het geval van nominale onafhankelijke variabelen werd univariate variantie-analyse toegepast. Vanwege het grote aantal getoetste verbanden worden de resultaten hier alleen globaal per categorie gepresenteerd. Voor een volledig overzicht zie Poell (1992).

Een aantal kenmerken bleek met geen enkel leereffect samen te hangen; deze zijn uit het analysemodel verwijderd. De andere kenmerken vertoonden in meer of mindere mate samenhang met de grootte van leereffecten. De belangrijkste beïnvloedende factoren waren: 1) voldoende voorbereiding vanuit de opleiding op de stage; 2) voldoende voorbereiding vanuit de

opleiding op het maken van een apart stageverslag; 3) het al dan niet uitvoeren van een extra stage voorafgaand aan de doctoraalstage: stagiairs die een extra stage hadden gedaan rapporteerden in veel categorieën grotere leereffecten; 4) de vraag of stage en scriptie gekoppeld zijn in één afstudeerproject: stagiairs bij wie dit het geval was, die in veel gevallen een interne stage hadden gedaan, gaven grotere leereffecten aan in 'Onderzoeksvaardigheden' en 'Beroepsmatige probleemoplossingsvaardigheden', terwijl zij kleinere leereffecten boekten in 'Inzicht in het beroep', 'Expertise' en 'Eigen toekomst'; 5) de mate waarin de stagewerkzaamheden van de stagiair uitdagend, interessant en afwisselend waren: hoe meer dit het geval was, des te groter het leereffect; 6) het nivo van de stage-werkzaamheden: stagiairs met werkzaamheden op of net boven hun eigen niveau rapporteerden grotere leereffecten dan stagiairs met werkzaamheden onder hun niveau of ver boven hun niveau.

Vermeldenswaard is nog dat tamelijk weinig verschillen werden aangetroffen tussen stagiairs van de twee onderzochte studierichtingen: slechts op de leerdoelcategorieën 'Onderzoeksvaardigheden' en 'Arbeidssocialisatie' geven onderwijskunde-studenten meer leerwinst aan dan respondenten van algemene pedagogiek, terwijl de uitgangssituatie vóór de stage vergelijkbaar is. Op andere leerdoelcategorieën bereikten stagiairs van de beide studierichtingen in

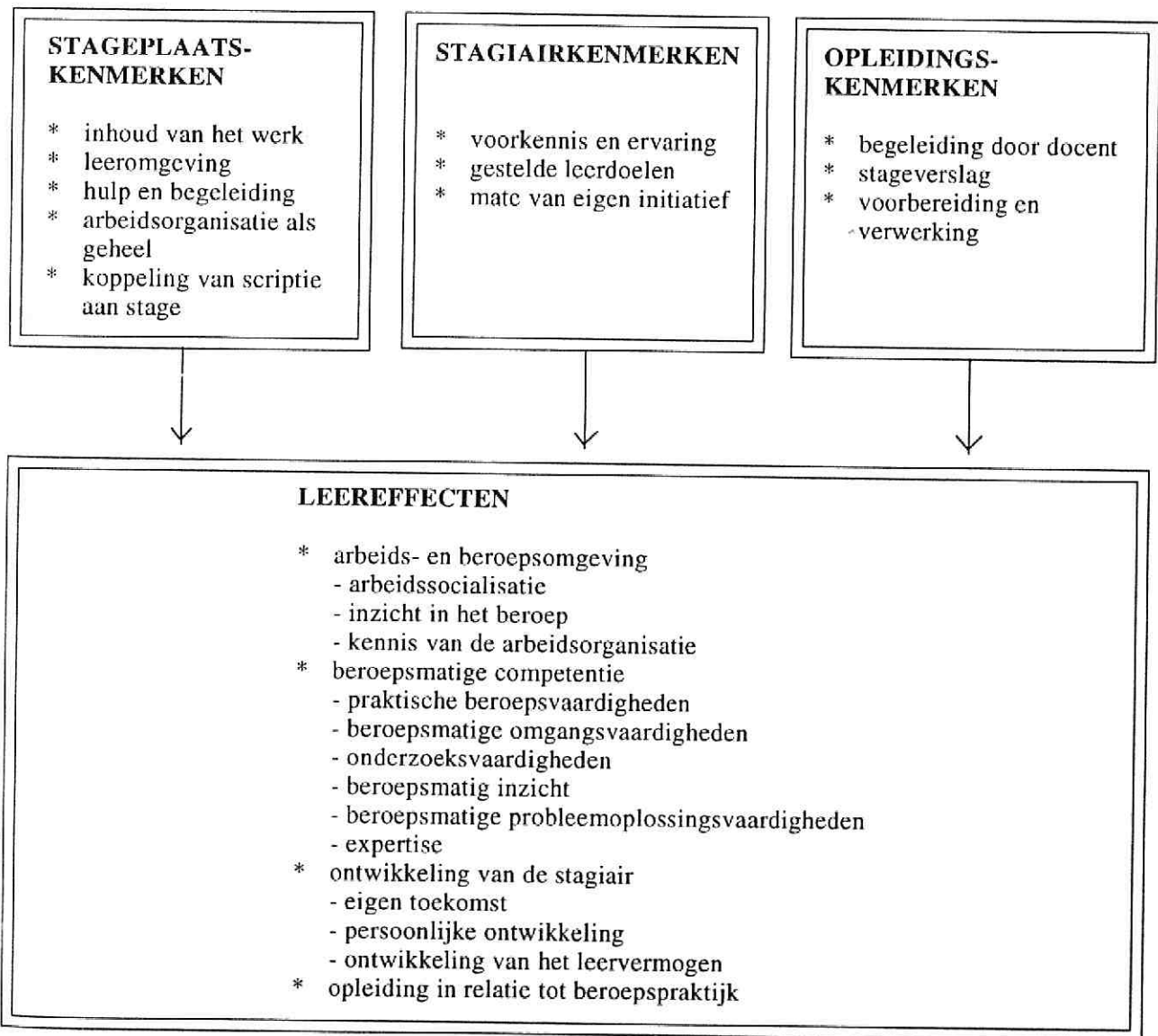


Fig. 1. Analysemodel voor systematische evaluatie van leereffecten van stages in het (sociaal-)wetenschappelijk onderwijs en factoren die daarop van invloed zijn.

gelijke mate leereffecten. Overigens is de geldigheid van dit resultaat beperkt tot de genoemde opleidingen aan de universiteit van Nijmegen.

In Figuur 1 wordt het 'definitieve' analysemodel getoond dat op grond van de gepresenteerde resultaten is samengesteld.

CONCLUSIES EN DISCUSSIE

We kunnen uit deze resultaten een aantal conclusies trekken. Ten eerste blijkt de stage een onderdeel van de sociaal-wetenschappelijke opleiding te zijn waaraan een zeer breed scala van leerdoelen wordt toegekend. Ten tweede blijkt de stage in de inschatting van stagiairs en hun begeleiders tot statistisch significante leereffecten te leiden, althans gemiddeld over de gehele onderzoeksgroep heen. In individuele gevallen doet een stage echter in bepaalde domeinen (bijvoorbeeld zelfvertrouwen en motivatie) meer kwaad dan goed. Op sociaal-normatieve leerdoelen boeken stagiairs minder leereffect dan op technisch-instrumentele, hoewel de eerste groep belangrijker wordt geacht. Er zijn betrekkelijk weinig verschillen tussen de onderzochte studierichtingen in de grootte van het leereffect op de verschillende leerdoelen. Ten derde zijn er meerdere factoren op de stageplaats, bij de stagiair en in de opleiding aan te wijzen die duidelijk invloed uitoefenen op de mate waarin bepaalde (categorieën van) leerdoelen worden bereikt. Gezien het feit dat het onderzoek slechts bij twee sociaal-wetenschappelijke studierichtingen is uitgevoerd, lijkt het raadzaam om de onderzoeksresultaten niet verder te generaliseren dan naar studenten van vergelijkbare studierichtingen.

Op deze plaats is het goed om op te merken dat de resultaten gekleurd zouden kunnen zijn door sociale wenselijkheid: van stagiairs (en begeleiders evenzo), die een behoorlijke tijd hebben geïnvesteerd in de stageperiode, kan worden verwacht dat ze aan die periode aanzienlijke leereffecten toeschrijven. Het gegeven dat op alle leerdoelen significante leereffecten werden gevonden kan aan dit beeld bijdragen. Dit gegeven betekent echter bij lange na niet dat alle stagiairs op alle leerdoelen leereffecten aangaven. Inventarisatie van het individuele scorepatroon over alle leerdoelen heen leert dat stagiairs gemiddeld op 1 tot 48 van de 55 leerdoelen een positief leereffect boekten, met een gemiddeld aantal van 25; op 1 tot 44 van de 55 leerdoelen werd géén leereffect aangegeven, met een gemiddeld aantal van 21; negatieve leereffecten kwamen voor op 0 tot 7 van de 55 leerdoelen, met een gemiddeld aantal van minder dan 1. De scores van de individuele respondenten waren derhalve tamelijk evenwichtig verdeeld over geen of een positief leereffect; negatieve leereffecten kwamen beduidend minder voor. De leereffecten mogen in positieve richting vertekend zijn, we konden niettemin verschillen in effecten op de diverse leerdoelen vaststellen. Hoewel voor elk leerdoel over alle respondenten heen een positief effect werd gevonden, was er per stagiair bovendien een aanzienlijke spreiding in de aantallen leerdoelen met positief effect. Er was dus zeker geen sprake van eenvormige scorepatronen. Sociale wenselijkheid mag een rol hebben gespeeld, maar heeft het scorepatroon slechts in beperkte mate beïnvloed.

De gepresenteerde resultaten hebben geleid tot een aanpassing van het voorlopige analysemodel (zie Figuur 1). Hierna worden de onderdelen van het 'definitieve' model besproken. Voor een meer gedetailleerde weergave verwijzen we naar Poell (1992).

De leereffecten uit het model zijn onder te verdelen in 13 categorieën, die hieronder globaal worden beschreven. 'Arbeidssocialisatie' verwijst naar samenwerking en communicatie met anderen, en naar het inzicht hebben in arbeidscultuur: kortom, het kunnen werken. 'Inzicht in het beroep' en 'Kennissen van de arbeidsorganisatie' spreken voor zich. 'Praktische beroepsvaardigheden' omvat computer-, onderhandelings-, presentatie-, vergader-, adviserings-, literatuur- en analytische vaardigheden. 'Onderzoeksvaardigheden' heeft betrekking op alle fasen in het uitvoeren van een onderzoek. 'Beroepsmatige omgangsvaardigheden' gaat over de omgang met anderen dan collega's in het werk. 'Beroepsmatig inzicht' draait om het inzicht in de kernproblemen van het beroep. Met 'Beroepsmatige probleemoplossingsvaardigheden' wordt geduid

op systematisch werken en situationele vaardigheden (kunnen improviseren). 'Expertise' verwijst naar het ontwikkelen van routine en ervaringskennis in de werkuitoefening. Onder 'Eigen toekomst' valt het inzicht in eigen arbeidsmarkt- en beroepsperspectieven en in eigen toekomstplannen en -mogelijkheden. 'Persoonlijke ontwikkeling' duidt op zaken als zelfkennis, zelfvertrouwen, zelfstandigheid, verantwoordelijkheid, drukbestendigheid, kritische houding, identificatie met het beroep, opkomen voor eigen belangen, zich handhaven in een groep, initiatief nemen, professioneel werken en doorzettingsvermogen. 'Ontwikkeling van het leervermogen' bestaat uit het herkennen van leermomenten, van leermogelijkheden en leerbelemmeringen, van de 'eigen leerstijl', uit het ontwikkelen en uitvoeren van leerstrategieën en het stimuleren van de motivatie. 'Opleiding en beroepspraktijk' tot slot, draait om het inzicht in de verschillen tussen theorie en praktijk, in de sterke en zwakke kanten van de opleiding en in de stage als onderwijsmiddel.

De eerste hoofdgroep van beïnvloedende factoren is die van stageplaatskenmerken. Onder 'Inhoud van het werk' vallen aspecten die met de verrichte stage-werkzaamheden te maken hebben, zoals de invloed die de activiteiten op de organisatie hadden, de mate waarin de stagiair zelfstandig functioneerde, de veelzijdigheid, uitdagendheid en afwisseling in de werkzaamheden, het niveau ervan, de samenwerking met anderen en de mentale taakbelasting. 'Leeromgeving' omvat de mogelijkheden die de stageplaats bood om te leren, zoals ruimte voor oriëntatie en experimenteren, de echtheid van het werk, het ontwikkelen van specifieke leerstrategieën door de stagiair en de mate waarin sprake was van een open cultuur en van *critical incidents*. 'Hulp en begeleiding' verwijst naar het commentaar en de ondersteuning die begeleiders en collega's gaven bij de uitvoering van werkzaamheden en het oplossen van problemen. Met 'Arbeidsorganisatie als geheel' wordt bedoeld op het onderscheid tussen een interne en een externe stage, en op specifieke kenmerken van of gebeurtenissen in de bredere organisatiecontext van de stagewerkplek, die hun invloed deden gelden. Tot slot is van belang of stage en scriptie in één project waren gekoppeld, of dat deze van elkaar waren losgemaakt.

De tweede hoofdgroep van beïnvloedende factoren omvat stagiairkenmerken. Hiervan maakt 'Voorkennis en ervaring' deel uit: het uitgevoerd hebben van een extra stage vóór de doctoraalstage en van werkzaamheden tijdens de huidige opleiding, buiten het kader van de opleiding, of tijdens (een stage binnen) een eerdere opleiding. Onder 'Gestelde leerdoelen' valt de vraag in hoeverre de stagiair expliciete leerdoelen voor de stage heeft gesteld, in hoeverre hij op basis van een stageplanning heeft gewerkt en of hij een duidelijk beeld heeft van de werkzaamheden die hij in het kader van de stage kan uitvoeren en zou willen uitvoeren. Tot 'Mate van eigen initiatief' behoren de factoren vragen stellen en informatie verzamelen, en motivatie.

De laatste hoofdgroep van factoren die de grootte van leereffecten beïnvloeden wordt gevormd door kenmerken van de opleiding waarvan de stage deel uitmaakt. 'Begeleiding door de docent' is hierin van groot belang: zijn bemoeienis met de inhoudelijke invulling van het stagewerkplan vóór de stage en bij afwijking daarvan tijdens de stage, de kwaliteit en kwantiteit van de vakmatige begeleiding, de aandacht voor de bevordering van het leerproces bij de stagiair, het initiatief tot het maken van begeleidingsafspraken, het bezoeken van de stageplaats en de ervaring van de docent in het begeleiden van stagiairs spelen hierbij een rol. De verplichting tot het schrijven van een apart stageverslag naast de scriptie, de mate waarin de inhoud daarvan wordt gespecificeerd door de opleiding en waarin de stagiair op het schrijven ervan wordt voorbereid zijn eveneens belangrijke factoren. Tot slot is van belang of er binnen de opleiding voldoende aandacht wordt besteed aan voorbereiding (voorlichting, opstellen stagecontract) en verwerking (terugkomdagen) van de stage.

De resultaten van het hier gerapporteerde onderzoek dragen bij aan verdere theorievorming op het terrein van leereffecten van stages in het wetenschappelijk onderwijs. Het onderzoek verschaft indicaties omtrent de onderwijskundige voorwaarden waaronder stages in het kader van een opleiding tot leereffecten kunnen leiden. Deze voorwaarden zijn uiteraard met name gelegen in opleidings- en stageplaatskenmerken. Hiernaast kunnen de resultaten door stagiairs en begeleiders worden gebruikt als checklist ter voorbereiding op de stage; het model kan dan

worden beschouwd als een lijst van topics waaraan door hen gezamenlijk aandacht kan worden besteed vóór de stage definitief wordt ingevuld.

LITERATUUR

- Academische Raad (1981). *Stages in het wetenschappelijk onderwijs. Een verkenning in het licht van de Beleidsnota Stages*. Den Haag: Commissie Onderwijsvraagstukken Wetenschappelijk Onderwijs, Werkgroep Stages, Academische Raad.
- Berg, J. van den & Tillaart, H. van den (1986). *Nieuwe werknemers. Een onderzoek naar recrutering en introductie van nieuwkomers in bedrijven*. Nijmegen: ITS.
- Berretty, J., *Buitenschoolse leerplaatsen, een alternatief. Literatuurstudie naar de betekenis van leerplaatsen in arbeidsorganisaties vanuit het oogpunt van emancipatie*. Nijmegen: ITS, 1981.
- Beyers, J.A.A.M. & Verrijt, A.H.M. (1990). *Bedrijfscommunicatie: uitval, stage en onderwijsprogramma*. Nijmegen: IOWO.
- Bijtel, J.A.L. van den (1988). *Analyse van het onderwijsaanbod van eerste-fase opleidingen. Deelrapport 4: evaluatie-onderzoek Wet Tweefasenstructuur*. Amsterdam: SCO.
- Dinham, S.M. & Stritter, F.T. (1986). Research on professional education. In: M.C. Wittrock (ed.), *Handbook of research on teaching. Third edition*. (pp. 952-970). New York: Macmillan.
- Dool, P.C. van den (1987). *Beoordelen van stages. Verslag van een literatuurstudie*. Specialistisch bulletin, nr. 59. Arnhem: CITO.
- Geurts, J. (1989). *Van niemandsland naar beroepenstructuur. Een studie over de aansluiting tussen onderwijs en arbeid op het niveau van aankomend vakmanschap*. Proefschrift Nijmegen. Nijmegen: ITS.
- Hoogstraten, J. (1982). The retrospective pretest in an educational training context. *Journal of Experimental Education*, 50, 200-204.
- Jong, M.W. de, Moerkamp, T., Onstenk, J.H.A.M. & Babeliowsky, M. (1990). *Breed toepasbare beroepskwalificaties in leerplan en beroepspraktijk. Een probleemverkenning en begripsanalyse*. Amsterdam: SCO.
- Kok, J. (1981). *De stage als intermediair tussen opleiding en maatschappij. Een pleidooi voor een betere sociologiestage*. Afstudeerscriptie sociologie. Groningen: Rijksuniversiteit.
- Kraayvanger, G. & Onna, B. van (1985). Arbeid en leren: ontwikkelingsfasen en begripsvorming. In: G. Kraayvanger & B. van Onna (red.), *Arbeid en leren*. Bijdragen tot de volwasseneneducatie. Baarn: Nelissen.
- Meijer, E. de, Hoogstraten, J. & Sprangers, M. (1986). De status van retrospectie. *Kennis en Methode*, 2, 191-203.
- Meijnen, G.W. (1987). Onderwijssociologie. In: N.A.J. Lagerweij & J.F. Vos (red.), *Onderwijskunde, een inleiding*. Groningen: Wolters-Noordhoff.
- Mertens, F.J.H. (1981). *Stages in een beroepsopleiding. Het praktijkjaar van het HTO*. Proefschrift Universiteit van Amsterdam. Nijmegen: ITS.
- Ministeries van Onderwijs en Wetenschappen en Sociale Zaken (1981). *Beleidsnota stages*. Den Haag: Staatsuitgeverij.
- Nieuwenhuis, A.F.M. (1991). *Complexe leerplaatsen in school en bedrijf. Een studie naar de implementatie en effecten van participerend leren in het MBO*. Proefschrift Rijksuniversiteit Groningen.
- Onna, B. van (1981). Arbeid - leren - educatie. *Tijdschrift voor Agologie*, 6, 396-428.
- Onstenk, J., Moerkamp, T., Voncken, E. & Dool, P.C. van den (1990). *Leerprocessen in stages*. Amsterdam: SCO.
- Poell, R.F. (1992). *Leereffecten van stages in het wetenschappelijk onderwijs. Een empirisch onderzoek bij twee sociaal-wetenschappelijke studierichtingen*. Doctoraalscriptie onderwijskunde, Katholieke Universiteit Nijmegen.
- Puffelen, F. van, Paridon, W.C. van & Ree, K. van der (1985). *De werking van de stagemarkt in Nederland. Een onderzoek naar de vraag- en aanbodfactoren op de stagemarkt van het middelbaar en hoger beroepsonderwijs*. Amsterdam: SEO.
- Sprangers, M.A.G. (1988a). *Response shift and the retrospective pretest. On the usefulness of retrospective pretest-posttest designs in detecting training related response shifts*. Den Haag: SVO.
- Sprangers, M.A.G. (1988b). A further note on the necessity of including a retrospective pretest in self-report pretest-posttest designs to detect training effectiveness. *Tijdschrift voor Onderwijsresearch*, 13, 353-355.

- Steinen, H. (1985). *Stages. Een onderzoek naar de relatie tussen stagedoelen en stagecondities*. Doctoraalscriptie onderwijskunde, Rijksuniversiteit Utrecht.
- Veen, A. & Voorthuis, M. (1989). *PAOW en arbeidsmarkt. Nieuwe mogelijkheden op de arbeidsmarkt voor pedagogen, andragogen en onderwijskundigen*. Amsterdam: SCO.
- Velden, R. van der (1990). Letterenstudenten op de arbeidsmarkt. *Tijdschrift voor Hoger Onderwijs*, 3, 122-128.
- Vries, B. de (1988). *Het leven en de leer. Een studie naar de verbinding van leren en werken in de stage*. Proefschrift Katholieke Universiteit Nijmegen. Nijmegen: ITS.
- Vries, B. de & Wel-Berretty, J. van (1986). *Tien jaar buitenschools leren*. Nijmegen: ITS.
- White, P. (1988). Knowing more about what we can tell: 'Introspective access' and causal report accuracy 10 years later. *British Journal of Psychology*, 79, 13-45.
- Wolfhagen, H.A.P. (1993). *Kwaliteit van klinisch onderwijs*. Proefschrift Rijksuniversiteit Limburg, Maastricht.
- Zolingen, S.J. van (1991). *Sleutelkwalificaties van afgestudeerden uit het MBO (MTS-E en MEAO-C) in startfuncties*. Paper gepresenteerd op de WESWA-onderzoekscongresdag (Nijmegen, 14 november 1991). Nijmegen: Vakgroep Algemene Pedagogiek.
- Zolingen, S.J. van & Onna, B. van (1992). *Key qualifications and senior secondary vocational education (MBO)*. Paper presented at the European Conference on Educational Research (Enschede, 22-25 juni 1992). Nijmegen: Vakgroep Algemene Pedagogiek.

Manuscript ontvangen 28-7-1993

Definitieve versie ontvangen 10-1-1994

Diagnose van cognities als interventie in de training van leraren

H.H. Tillema

*Vakgroep Onderwijsstudies, Rijksuniversiteit Leiden**

ABSTRACT

Pre-existing knowledge and beliefs of teachers are important factors in the acquisition of new domain-specific knowledge (as is the case in the training of teachers). There are, however, few training-designs to date that explicitly take into account the cognitions of teachers.

In an experimental study, it was investigated whether making available diagnostic information about teachers to the trainer in a course would benefit the subsequent acquisition of new information of those teachers. It was hypothesized that a close match (congruence) of new information to the already existing knowledge and beliefs of teachers would help in the integration of that information; diagnosis of pre-existing knowledge is regarded as a device to help the trainer to deliver new information more in correspondence with already existing knowledge and beliefs. Two training designs were used that differentially take into account diagnosis of teacher cognitions: conceptual-based training and experience based training.

The results of the study give support to the contention that the more congruent teacher cognitions are with the to-be-acquired knowledge, the more learning takes place. However, informing the trainer about existing teacher cognitions in the course, in order to be able to adapt to those cognitions and arrive at a closer correspondence of teacher cognitions and training content, did not prove to be an effective strategy.

INLEIDING

De vraag hoe in trainingen rekening kan worden gehouden met aanwezige cognities van leraren is een probleem dat steeds meer lijkt te worden erkend (Corporaal, 1988; Wilson & Colc, 1992). Echter, er zijn nog nauwelijks adequate theoretische benaderingen over het verwerven van bekwaamheden door leraren voorhanden of concrete aanvullingen of aanzetten tot interventies uitgewerkt ten behoeve van de trainingsopzet. Daardoor is het moeilijk om met leerkrachtcognities om te gaan in trainingen (Zeichner & Gore, 1990). Toch wordt verondersteld dat het meer in overeenstemming zijn van leerkrachtcognities met de trainingsinhoud van de cursus van positieve invloed is op trainingsresultaten (Tillema, De Jong & Mathijssen, 1990). Een studie werd uitgevoerd om de invloed van leerkrachtcognities bij de opname van nieuwe kennis via een training te toetsen. Als probleem werd in beschouwing genomen dat de bovengenoemde veronderstelling weliswaar het belang van cognities impliceert, maar constaterend van aard is; het zegt nog weinig over hoe de aansluiting bij ofwel het rekening houden met cognities in een training kan worden gerealiseerd. Ter aanvulling werd een tweede, meer construerende veronderstelling als hulphypothese gehanteerd, nl. de beschikbaarheid van diagnostische informatie over aanwezige cognities van leraren bij de trainer heeft een positief effect op de trainingsresultaten omdat een trainer die de beschikking heeft over diagnostische informatie de leerstofbehandeling meer, dan wel beter, kan afstemmen op de aanwezige cognities. De volgende toelichting kan daarbij gegeven worden.

Aansluiting bij aanwezige cognities

Leerkrachten worden hier opgevat als 'professionals' die beschikken over geëlaboreerde kennisstructuren (Bennett & Clive, 1992). In de verwerving van nieuwe kennis door deze professio-

*Adres: Vakgroep Onderwijsstudies, Postbus 9555, 2300 RB Leiden.

nals spelen aanwezige cognities (inhoudende zowel kennis als opvattingen en beliefs – Carter, 1990) in belangrijke mate een sturende en selecterende rol bij de opname van nieuwe informatie (Kagan, 1992; Pajares, 1992). Leerkrachten beschikken over specifieke kennisdomeinen voor hun handelen, die al dan niet eenvoudig expliciteerbaar te maken zijn. Veel van deze kennis is 'tacit' en gebonden aan of geëliciteerd in het handelen. Shulman (1986) en Berliner (1987) onderscheiden verschillende typen van kennis bij leraren. Waar het gaat om propositionele kennisdomeinen (zoals kennis van de leerling, kennis van pedagogische acties etc.) zijn deze vaak in hoge mate contextspecifiek. Door ervaring (toename van professionele kennis) is beschikbare kennis tevens meer persoonlijk en narratief ofwel 'geval' gebaseerd (Elbaz, 1983) en minder abstract of regelgebaseerd (Gudmundottir, 1991). Daarbij valt op dat het denken en handelen, zeker bij 'goed onderwijzen' (Clark, 1989), in belangrijke mate geleid wordt door morele overwegingen. Beschikbare kennis van leraren is bovendien perspectief- en opvattingen-gebonden – zie ook Goodlad, Soder en Sirotnik als het gaat om het belang van 'teacher beliefs' (1990). Daarmee vormen de aanwezige cognities een belangrijke conditie voor nieuw leren.

Bij het expliciet aanbieden en verwerven van nieuwe kennis, zoals dat in trainingen beoogd wordt, is het noodzakelijk op enigerlei methodische wijze aansluiting te vinden bij deze reeds aanwezige leerkrachtcognities. Dat dit laatste allerminst eenduidig is laat onderzoek naar verschillen tussen beginner en expert zien; cognities van meer of minder ervaren leraren kunnen aanmerkelijk verschillen. Ook vormt een 'probleem' dat cognities niet alleen betrekking hebben op expliciteerbare, op onderdelen omschrijfbaar kennis maar, zeker bij professionals, betrekking heeft op beliefs, opvattingen en oriëntaties op beroepsmatig handelen. In al deze gevallen geldt als veronderstelling dat kennisverwerving pas dan geslaagd is wanneer er overeenstemming is of is bereikt (congruentie) tussen leerkrachtcognities en de nieuwe kennis (Vosniadou & Brewer, 1987). We noemen dit de congruentie-hypothese.

Echter, in de gebruikelijke trainingsopzetten is meestal volstaan met weinig specifieke, d.w.z. theoretisch nauwelijks onderbouwde maatregelen om de 'beginsituatie' te analyseren; zo dit al gebeurt (zie Cruickshank & Metcalf, 1990). Er geldt in het algemeen een gemis aan diagnose-maatregelen in trainingen. In recente, cognitief-georiënteerde trainingsbenaderingen (Wilson & Cole, 1992) worden evenwel, meer dan tot nu toe in de trainingsliteratuur gebruikelijk is, expliciete verwijzingen aangetroffen over de cognitieve aanbieding en verwerking van informatie bij ervaren professionals. Evenwel, de verwijzingen zijn allerminst gelijklopend, zodat het waardevol is ze in onderzoek te nemen.

Gliesman (1984) is een der eersten die heeft gewezen op het belang van conceptuele instructie als een vorm van trainen waarin de leerstofbehandeling expliciet gericht is op het aanbrenge van eenduidige, essentiële begrippen. Dit gebeurt op basis van een analyse van de relaties tussen concepten. Uitgaande van de structuur van begrippen in aan te bieden informatie is het de bedoeling een nieuwe en hechte kennisstructuur aan te bieden, ter vervanging van de bestaande. De veronderstelling wat betreft de invloed van cognities in deze trainingsbenadering is dat de conditie waaronder concept-acquisitie plaats vindt in feite de enig bepalende factor is voor conceptverwerving. Het is zelfs bepalend voor de daaropvolgende concepttoepassing, d.i. handelen. In deze (vrij extreme) benadering wordt in zoverre 'rekening' gehouden met leerkrachtcognities dat nieuwe kennis zo volledig mogelijk en in hoge mate gestructureerd wordt aangeboden als een 'pakket, dat bedoeld is ter vervanging van bestaande kennis. Met andere woorden in de training vindt een confrontatie met nieuwe informatie plaats die bedoeld is om de aanwezige cognities opnieuw in te vullen; aansluiting bij aanwezige cognities ligt bij de leerkracht als verwerker van die informatie. In deze benadering is er sprake van accommodatie en vervanging i.p.v. assimilatie met bestaande kennisstructuren. Verschillen tussen beginners en meer ervaren leraren worden als zodanig niet onderkend. Varianten van de door Gliesman voorgestane benadering zijn te vinden in cognitief georiënteerde trainingsopzetten van Carpenter e.a. (1990), Tillema e.a. (1990) en Verloop (1989).

Een andere oplossing om rekening te houden met beschikbare kennis van leerkrachten is te vinden bij enkele ervaringsgeoriënteerde leerkrachtrainingen (Richardson, 1990; Joyce,

Murphy, Showers & Murphy, 1989; Zeichner & Gore, 1990) waar meer aandacht wordt geschonken aan het vooraf expliciteren van bestaande leerkrachtcognities en reflectie op aanwezige opvattingen van cursisten. Deze ervaringsgerichte trainingen voorzien daarin door ruimte te creëren voor discussie en cursist-geleide studie en een vergrote inbreng in de training. De concept-verwerving sluit in sterke mate aan bij of gaat zelfs uit van de bestaande kennisorganisatie of schemata (zogenaamde 'personalized knowledge') van leraren (Elbaz, 1983). Er bestaat meer aandacht in de training voor assimilatie van nieuwe informatie met bestaande kennisstructuren dan de accommodatie of herstructurering van de kennis; dit laatste is onder controle van de cursist zelf. Een mogelijk probleem daarbij is dat weliswaar in principe rekening kan worden gehouden met verschillen in ervaring en kennisorganisatie, maar hoeft dat niet perse, vanwege de geringe sturing door de trainer en de controleerbaarheid van de over te dragen leerstofinformatie in de training.

Concluderend blijkt, dat volgens de eerste benadering aanwezige cognities dienen te worden vervangen door (correcte) nieuwe informatie en er weinig of geen rekening wordt gehouden met aanwezige leerkrachtcognities (geen assimilatie, wel accommodatie). Cognities worden in feite alleen aangesproken wanneer ze passend zijn in de leerstofbehandeling. In de andere benadering zijn aanwezige cognities weliswaar het aangrijpingspunt voor de behandeling van concepten, maar er is geen controle of zodanige sturing dat alle beoogde nieuwe kennis werkelijk aan bod komt (assimilatie, maar geen accommodatie).

Diagnose van cognities

Het overbruggen van de afstand tussen aanwezige cognities en nieuwe kennis in een training kan worden bereikt door specifieke maatregelen te treffen in de trainingsopzet, waardoor op directe wijze rekening kan worden gehouden met aanwezige cognities. Dit kan door het ter beschikking stellen van diagnostische informatie aan de trainer (Peterson, 1988). Onder diagnose wordt hier begrepen het ter beschikking hebben van informatie over de beschikbare kennis en 'beliefs' van cursisten door de trainer om in de leerstofbehandeling er rekening mee te houden. We noemen dit de diagnose-hypothese.

Het ter beschikking hebben van diagnostische informatie kan op verschillende manieren doorwerken in de beide bovengenoemde trainingsbenaderingen en zelfs een belangrijk bevorderend effect aan de training leveren. Diagnostische informatie in handen van de trainer kan helpen de 'match' (een term die Glaser hanteert) ofwel de congruentie te bevorderen tussen de 'beliefs' en kennis die de leraar meebrengt in een cursus enerzijds en de kennis en perspectiefgebonden informatie (attituden, oriëntaties) die de training beoogt over te dragen anderzijds. Met andere woorden, diagnose van cognities is een hulpmiddel om cognities van leraren in overeenstemming te brengen met de informatie en oriëntaties in de training. In algemene zin is het effect van opname van diagnose als onderdeel van een training-design (Tillema & Veenman, 1987) dat de voorwaarden gecreëerd worden voor betere aansluiting van de training bij cognities van leraren. Dit werkt echter verschillend uit, al naar gelang de gekozen trainingsbenadering. Inpassing van diagnose in een conceptuele training betekent meestal dat geprobeerd wordt bij de introductie van nieuwe concepten de aanwezige voorwaardelijke of begin-kennis op te sporen. Deze begin-kennis kan dan dienen als aangrijpingspunt om nieuwe kennis aan toe te voegen of eraan te koppelen. Bij ervaringsgerichte training houdt diagnose eerder in een activering van relevante ervaringen en preconcepties alsmede een poging tot sturing/bijstelling van de aanwezige cognities waardoor de trainer de leerstofbehandeling beter in relatie kan brengen met de leerstofstructuur/doelen.

Carpenter et al. (1990) noemen deze 'matching', d.w.z. de koppeling via diagnose van training en trainee een basisconditie voor het ontwerpen van instructie. Hoewel diagnose een voor de hand liggende maatregel is, wordt het in trainingen (en wat dit aangaat in onderwijsleersituaties überhaupt) weinig gehanteerd (Stallings 1982). Ervaren trainers gebruiken diagnostische informatie niet of nauwelijks (Bromme, 1987; Carter, 1990; Bennett, 1992) maar baseren zich eerder op een eigen ervaringsbasis over aannemelijke cognities bij cursisten; bovendien worden vaak expliciete diagnoseprocedures gemist. Anderzijds moet worden erkend dat in cognitieve

trainingsbenaderingen (Wilson & Cole, 1992) het belang van diagnose juist wordt onderstreept. Diagnose als interventie in training van leraren krijgt daarmee op zijn minst een controversieel karakter.

In deze studie is onderzocht of de diagnose-interventie, afhankelijk van de trainingsbenadering effect heeft, d.w.z. heeft het ter beschikking hebben van expliciete diagnostische informatie, welke de trainer in staat stelt de leerstofbehandeling tijdens de training meer in overeenstemming te brengen met de aanwezige cognities van de leraren in training, een invloed op de kennisverwerving van die leraren.

METHODE

Onderzoeksdesign

De voorgestelde diagnose-hypothese als aanvulling op de congruentie-hypothese is onderzocht bij ervaren leraren die beschikken over voorkennis (via opleiding) en ervaring (via de beroepspraktijk) met de te trainen onderwerpen. De onderzoeksgroep vormde een homogene groep leraren, die een aanvullende opleiding volgden ten behoeve van het speciaal onderwijs. Zij werden opgeleid door de Opleiding Speciaal Onderwijs van de Katholieke Leergangen in Tilburg. De onderzoeksgroep werden 2 verschillende trainingsonderwerpen aangeboden. Onderzocht werden de verschillen tussen een ervaringsgerichte en conceptuele training, waarbij al dan niet is gevarieerd op de beschikbaarheid van diagnostische informatie over de cursist bij de trainer. Geanalyseerd werd op meerdere trainingseffecten.

De trainingscondities zijn a) conceptueel zonder diagnose; b) conceptueel met diagnose; c) ervaringsgericht. De condities werden door eenzelfde trainer verzorgd. De twee verschillende onderwerpen werden beschouwd als een replicerende controlemaatregel in het onderzoeksdesign. De training werd gegeven op het terrein van sociale agogiek: vaardigheid in herkenning en hantering van gesprekstechnieken, en op het terrein van orthodidactiek: vaardigheid in registratiemethodieken van leer- en gedragsproblemen.

De voortraining van trainers (docenten)

De docenten/trainers van de 2 onderwerpen zijn geïnstrueerd volgens een voortraining bestaande uit:

Stap 1. Een serie van zeven gezamenlijke planningsbijeenkomsten voor alle trainers (n=5) met de onderzoekers waarin de trainingsaanpak werd beoefend. Dit gebeurde aan de hand van uitleg en voorbeelden bij situaties in het eigen onderwijs,

Stap 2. Met elke trainer afzonderlijk, werd de zgn. leerstofkaart voor zijn/haar onderdeel behandeld, bestaande uit een bespreking van het netwerk van te behandelen begrippen en relaties alsmede bespreken van de toetsen,

Stap 3. Er vonden gezamenlijke lesvoorbereidingsessies plaats voor elke trainingsbijeenkomst met een nabespreking van deze bijeenkomst door de bezoekende onderzoeker.

Voor elke bijeenkomst is de verkregen toetsinformatie over de cursisten doorgenomen met de trainer, in het geval van de conceptueel-diagnostische conditie. Dit is gebeurd aan de hand van de leerstofkaart; de daarin genoemde concepten zijn toegelicht met informatie uit de cursistvoortoetsen, waardoor de trainer de beschikking heeft over een directe terugkoppeling van beschikbare concepten bij de cursist in relatie tot de aan te bieden concepten.

Trainingsmaterialen ten behoeve van de trainers

De trainers kregen ten behoeve van de leerstofbehandeling in de bijeenkomst de beschikking over een leerstofkaart in de vorm van een uitgewerkte propositielijst met daarin alle concepten en relaties daartussen aangegeven (als een hiërarchie van geordende leerstofbegrippen). Deze propositielijst wordt verkregen met een techniek voor 'conceptmapping' van leerstof (uitgewerkt in Kok & Tillema, 1982).

De propositielijst omvat de grafische weergave van essentiële begrippen. Dit zijn de 15

hiërarchisch hoogst geplaatste begrippen en hun relaties. Daarnaast werd de lijst uitgeschreven in een draaiboek van te behandelen essentiële begrippen en relaties, tezamen met toelichting over de wijze van aanbieding van de kernrelaties; voor elk van de condities werd een draaiboek opgesteld waarin het verloop van bijeenkomsten werd gespecificeerd.

Conditie

In de conceptuele trainingsconditie kreeg de trainer de leerstofkaart tijdens de planningsbijeenkomsten voorgelegd (stap 2 in de voortraining) met daarin een voorgestelde volgorde voor de leerstofbehandeling van essentiële kernbegrippen en hun specificaties in hiërarchie van geordende deelbegrippen. Op basis van deze informatie is tijdens de planningsbijeenkomst samen met de trainer de leerstofbehandeling/route uitgewerkt volgens het draaiboek. Het beschikbaar stellen van diagnostische informatie is gebeurd door tijdens de voorbespreking per bijeenkomst (stap 3 in de voortraining) toetsinformatie van cursisten te plaatsen in de leerstofkaart.

Er zijn (zie par. toetsen hierna) 6 verschillende informatiebronnen beschikbaar omtrent de kennis en 'beliefs' van cursisten die aan de trainer werden aangeboden en deze werden besproken in relatie tot de leerstofkaart. Dit is gebeurd door het signaleren van discrepanties, tekorten, missende informatie, misvattingen. Voor iedere cursusbijeenkomst zijn steeds de consequenties van de toetsinformatie voor de leerstofbehandeling besproken.

In de ervaringsgerichte conditie heeft de trainer ook de beschikking over de leerstofkaart. Dit is gedaan om gelijkheid in de beoogde, te behandelen leerstof te krijgen. De planningsbijeenkomst in deze conditie richt zich op het selecteren en vinden van cases en voorbeelden als toelichting bij de leerstofkaart die de cursist inbreng kunnen verhogen, via inbreng van ervaringen, opvattingen en dergelijke.

Geen expliciete diagnostische informatie werd de trainer ter beschikking gesteld, maar de trainer werd gevraagd zich een beeld te vormen van de opvattingen, misvattingen en aanwezige kennis bij cursisten. Dit werd beschouwd als een impliciete oriëntering op aanwezige cognities bij de cursisten. In deze conditie kreeg de trainer niet de beschikking over door de onderzoekers verzamelde toetsinformatie (de voortoetsen).

Toetsen en hun betrouwbaarheid

Er is een onderscheid gemaakt tussen kennis, opvattingen ('beliefs') en gedrag als verschillende aspecten die leerkrachtcognities kunnen vastleggen.

De kennistoetsen bestaan uit

– Probleemvignettes (PROB): dit zijn vier open, niet voorgestructureerde situatieschetsen waarin door de cursist oplossingshandelingen moeten worden beschreven voor een probleem uit de leerstof. De door de cursist genoemde concepten uit de behandelde leerstof zijn via een conceptmapping techniek (Kok & Tillema, 1982) grafisch weergegeven in termen van relaties tussen concepten en geanalyseerd op voorkomen van behandelde leerstofconcepten, op het voorkomen van behandelde relaties en de overeenkomst met de normbeantwoording door de trainer, waarbij is gelet op de discrepantie met het trainer-antwoord. De trainer werd gevraagd aan te geven hoe de beantwoording bij voorkeur zou moeten luiden. De interne consistentie van deze test was (Cronbach's alpha) .76 (over n=146).

– Begrippenparentoets (REL): een paarswijze aanbieding van ± 15 belangrijke begrippen uit de leerstofkaart met een aanduiding van de sterkte van het verband tussen beide (5-puntschaal). Deze toets is geanalyseerd met multidimensionale schaaltechnieken (SMACOF-1B) en vergeleken met het normantwoord van de trainer. Een maat voor interne consistentie kon niet worden berekend over dit type data.

– Multiple choice toets (MCT): een gesloten gestructureerde toets met 15 items over beschikbare kennis uit de leerstof op basis van goed/fout gescoord. De interne consistentie voor de pretest data voor de twee topics was respectievelijk .84 en .89.

De 'belief' c.q. opvattingentoetsen bestaan uit:

– Leerverwachtingentoets (EXP): een open, door drie startvragen gestructureerde opdracht om een 'learner report' te schrijven over het belang, de relevantie en de verwachte leereffecten

van de cursus. De verkregen tekst is op voorkomen van categorieën uitgewerkt en vergeleken met een categoricënljst over essentiële leerervaringen, -doelen die door de trainer over zijn cursus is aangelegd. De interne consistentie voor deze test was (α) .57, hetgeen als redelijk voor dit type data werd beschouwd.

– Argumentatie-toets (ARC): Voor een drietal cases is gevraagd te beschrijven, in een korte tekst, welk gedrag adequaat, professioneel is binnen de beschreven situatie. De cases zijn door de trainer aangewezen professionele situaties. De cursist is gevraagd te beschrijven wat professioneel handelen of oplossingen voorstelt binnen de situaties. De tekst is eveneens via de gehanteerde conceptmapping techniek vergeleken met de normbeantwoording van de trainer. De interne consistentie over de gecombineerde pre- en posttest data was .78.

– Attitude-schaal (ATT): Via een semantische differentiaal zijn de houdingen ten aanzien van professioneel handelen in de beide topics bevraagd op een 10-tal items. De alpha coëfficiënten voor de twee topics waren .73 en .89.

De gedragstoetsen die in het onderzoek zijn opgenomen dienen om het gebruik en de inzet van de leerstof/kennis te beoordelen in op de praktijk van het leerstofgebied lijkende situaties. Deze toetsvorm is gekozen om de toepassing van het geleerde op meer gecontroleerde wijze dan in de praktijk zelf mogelijk is, te toetsen. Gehanteerd is de assessment center-methode. Deze methode lijkt meer recht te doen aan meting van bekwaamheid bij leraren dan directe vormen van kennistoetsing (Tillema, 1993). De volgende vormen van assessment zijn in dit onderzoek opgenomen.

– Portfolio (POR): Dit is een opdracht om een verzameling aan te leggen van producten, specimen, outputs uit de eigen praktijk van de leraar. Een aan de leerstof ontleende taak/opdracht dient uitgewerkt voor de eigen onderwijssituatie. In dit geval is gekozen voor: een leerlingobservatie, opstellen van een interviewverslag, overzicht maken van leerlingmateriaal. Het verzamelde dossier is op een tienpuntschaal beoordeeld door de trainer van de cursus, en wel op volledigheid, reflectie op de leerstof en adequaat gebruik van de leerstof.

– Assessment-center (ASS): Op een speciale bijeenkomst zijn praktijkgelijkende taken als oefening voorgelegd aan de cursisten.

a: een videokritiek van een klassesituatie – doel is het geven van onderbouwd commentaar op een praktijksituatie vanuit de geleerde theorie; de genoemde theorie-elementen zijn gescoord.

b: expert-analyse: doel is een onderbouwing, commentaar te geven op expert-voorbeeld gedrag ten aanzien van een probleem. Gevraagd is om onderbouwde redenen te geven voor dit gedrag; het aantal redenen is als frequentie gescoord.

c: een plannings- of probleembehandeling aan de hand van een videosituatie. Doel is het geven van een ontwerp voor de oplossing van een praktijkprobleem vanuit de theorie. De oplossingen zijn beoordeeld op een tienpuntschaal door de betreffende trainer.

De schriftelijke reflectieverslagen van de cursist zijn beoordeeld door de trainer (vrije beoordeling) en via een categorieën-instrument voor het meten van reflectie door de onderzoeker, ontleend aan Van Manen (1977).

De ASS-toets en scoringsprocedure zijn eerst bij een kleine deelpopulatie van leraren die geen deel uit zouden maken van de onderzoeksgroep vooraf beproefd. De interbeoordelaar-betrouwbaarheid (kappa) van drie beoordelaars/onderzoekers was gemiddeld over de drie maten .69. Ondanks deze lage waarde, welke wijst op 'ruis' tussen de assessoren, werd, gelet op het belang van de toets, de afname niet gewijzigd. De toets fungeerde als een eindopdrachtbeoordeling van de training door de trainer.

Afname en betrouwbaarheid van de procedure

Voor de drie condities (met 2 verschillende onderwerpen) werden 6 cursusgroepen geformeerd met in totaal 146 cursisten. De onderwerpen waren over gemiddeld 6 bijeenkomsten verspreid behandeld. De condities zijn gecontroleerd op feitelijk correcte implementatie van de treatments (trainingsaanpak) door:

– registratie van de behandelde leerstof aan de hand van audio-opnames van de trainer-

Tabel 1. Mate van implementatie van de te behandelen leerstof (in percentages) en concepten.

Conditie	percentage behandelde leerstof (%)		ratio 15 meest centrale concepten	
	topic 1	topic 2	topic 1	topic 2
Conceptuele conditie	91	96	.76	.42
Conceptuele-diagnostische conditie	98	94	.69	.57
Ervaringsgerichte conditie	89	85	.58	.43

cursist-interactie. Het gaat hierbij om controle op de 'dekkingsgraad' van de behandelde leerstof vergeleken met de leerstofkaart. Scoring vond plaats op het aan de orde komen van concepten uit de leerstofkaart en de vergelijkbaarheid van leerstofconcepten tussen de condities. Dit gebeurde op grond van a: percentage van behandelde concepten volgens de concept map en b: de ratio van de meest centrale begrippen (15 van deze begrippen zijn geselecteerd) ten opzichte van het totaal aantal concepten in de leerstofkaart; dit werd als een meer inhoudsgebonden maat dan a beoordeeld.

Tabel 1 geeft een weergave van de resultaten. Per treatment werden tevens zgn. conditie definiërende kenmerken bepaald. Dit is een reeks van observatiecategorïeën waarmee de wijze van leerstofbehandeling kan worden gescoord per conditie. Deze scoring volgt in welke mate de specificatie zoals gegeven in het draaiboek ook daadwerkelijk gevolgd is tijdens de bijeenkomsten. Via een 'time-sampling' categorieënsysteem werden frequenties gescoord (zie tabel 2 voor de categorieën).

Toetsing op stabiliteit van de conditie over de verschillende topics levert (middels Chi kwadraat toetsing) als resultaat dat de condities qua mate van implementatie niet significant van elkaar verschillen ($\chi^2 = 32.53$ met $df = 33$; $p = .39$ en $\chi^2 = 10.63$ met $df = 13$; $p = .45$ voor de conceptuele en ervaringsgerichte conditie).

Tabel 2. Lijst van definiërende kenmerken per conditie.

Conceptuele training (met en zonder diagnose)	Ervaringsgerichte training
Rol van de trainer	
. presenteert overzicht, geeft doelformulering	. presenteert startprobleem
. definieert concept en legt uit	. stimuleert/leidt discussie
. noemt uitwerking en hantering van concept	. brengt informatie in
. relateert concept en representeert concept	. structureert discussie, geeft hints
. vraagt om begrip, checkt aandacht	. recapituleert
. herhaalt en grijpt terug	. vraagt om meer verduidelijking
. geeft kleine opdracht/vraag	. introduceert nieuw punt
. vraagt om specifieke informatie	. checkt voor begrip
Rol van de leraar	
. reageert op vragen	. interacteert met anderen zonder trainer
. vraagt om toelichting, brengt punt in	. vraagt om toelichtende informatie (1 op 1)
. geeft toelichting en geeft informatie	. levert probleem (elaboratie/oriëntatie)
. geeft mening	. stelt nieuw probleem
	. checkt voor relevantie

Voor de afname van de condities heeft bij alle cursisten de voortoetsing (met de kennis en opvattingentoetsen) plaatsgevonden op een daarvoor bestemde bijeenkomst, geruime tijd voor de eerste cursusbijeenkomst.

Voor de eerste cursusbijeenkomst zijn de toetsgegevens geanalyseerd en besproken met de trainer in de diagnosconditie.

Na de (gemiddeld 6) bijeenkomsten is de natoets afgenomen, met gebruikmaking van parallel versies van de voortoetsen. Tevens is op dat moment de portfoliotoets verstrekt.

Na een maand -met een uitloop tot 3 maanden vanwege organisatorische redenen- is de assessmentbijeenkomst met de drie situatietests gehouden; de portfoliotoetsen zijn daarbij ingeleverd.

ANALYSE

De onderzoek is een veldexperiment met een factorieel design met een pretest-posttest. De meting vindt plaats op drie typen criteriumvariabelen: kennis, opvattingen en gedrag. De factoren in het variantie-analytisch design zijn: a) trainingsaanpak en b) topic of onderwerp.

Ten aanzien van de trainingsaanpak geldt dat in plaats van deze variabele te beschouwen als een 'fixed factor' het met behulp van de scoring op de observatiecoderingsschaal (de mate van implementatie van de treatment) ook gemeten kan worden als random factor. Echter, vanwege de geringe verschillen in implementatie is dit niet uitgevoerd.

Ten aanzien van diagnose geldt dat het als een conditie en daarmee als een fixed factor is beschouwd. Met behulp van implementatiecontrole is beoordeeld of het als conditie feitelijk is ingevoerd. De voorkennis van de cursisten is daarmee niet als random factor meegenomen, dit vanwege onderzoeksdesignoverwegingen.

De scoring van de toetsen is op tweeërlei wijze gebeurd: a) de behaalde totaalscore op een toets en b) de discrepantiescore op de toets in vergelijking met de bij de trainer bepaalde score als 'norm' d.w.z. gewenste beheersingsscore. Van de kennistoetsen zijn tevens de verwachte scores verzameld bij de trainer, nl. de discrepantie tussen de eindscores van de leraar en de inschatting van die score door de trainer. Dit werd beschouwd als een maat voor de kennis van de trainer over zijn cursisten. Afname van deze toetsen bij de trainers vond plaats tijdens de treatmentperiode, en voor de natoetsafname bij de cursisten.

RESULTATEN

In de beschrijving van resultaten is eerst van belang vast te stellen of er verschillen zijn ten aanzien van kennisverwerving (leerwinst) en verschuiving in opvatting of gedrag, gemeten als verschillscores. Vervolgens is aan de orde de verschillen op de trainingscondities gemeten ten aanzien van de effectmaten, die verband houden met de congruentie-hypothese: hoe kleiner de aanwezige verschillen tussen leraar 'beliefs' en trainingsinhouden, des te meer leerwinst. Tenslotte zijn de verschillen geanalyseerd onder de diagnosehypothese: kennis over de 'leerder' bevordert de effectiviteit van de trainingspresentatie.

Verschillen tussen meetmomenten

In tabel 3 zijn de gemiddelden, standaarddeviaties en toetsingen opgenomen voor de verschillen tussen meetmomenten per conditie en de verschillen tussen condities gemeten over voor- en nameting, voorzover betrekking hebbend op de kennistoetsen.

De tabel laat zien dat er weinig kennisverwerving c.q. nieuw leren heeft plaatsgevonden. Alleen in de conceptuele conditie (topic 2) heeft er significante leerwinst plaatsgevonden, en alleen voorzover het de MCT-test betreft. In de ervaringsgerichte conditie heeft zelfs een zekere mate van 'unlearning' plaatsgevonden.

De F-ratios (d.w.z. de verschillen) tussen de condities op de voortoetsen waren niet signifi-

Tabel 3. Gemiddelden, verschillen, standaarddeviaties and t-waarden voor de experimentele condities (N=146) in winstcores (absolute verschillen tussen pre- and posttest scores).

	Kennis toetsen (verschil scores)				
	PROB		REL	MCT	
	M (sd)	t (df)		M (sd)	t (df)
topic 1					
-conceptueel-diagnostische treatment (n=28)	.45 (3.21)	.52 (17)	n.b.	.13 (1.32)	.76 (26)
-conceptuele treatment (n=31)	.22 (3.36)	.23 (17)	n.b.	.12 (1.57)	.10 (17)
-ervaringsgerichte treatment (n=12)	.66 (1.58)	1.26 (8)	n.b.	.41 (2.23)	.65 (11)
topic 2					
-conceptueel -diagnostische treatment (n=29)	.13 (2.61)	.25 (21)	n.b.	.04 (2.32)	.08 (28)
-conceptuele treatment (n=25)	1.50 (1.87)	1.96 (15)	n.b.	2.08 (2.94)	3.54* (24)
-ervaringsgerichte treatment (n=21)	.10 (4.57)	.11 (18)	n.b.	2.95 (2.81)	.48 (21)
Verschillen tussen condities					
posttest	F				
	df				
	p				
		3.86 (2,85)		8.57 (2,145)	
		.02*		.00**	

** p<.01; * p<.05; n.b.= niet berekend -REL test levert geen somscores

cant, ook voor wat betreft de 'belief' toetsen, met uitzondering van de MCT-pretest; hier werd een significant effect tussen topics gevonden met een $F(1,145)$ ratio van 5.31, $p<.001$. De gevonden natoets verschillen op de MCT-test kunnen daar deels aan toe te schrijven zijn. In het geval van de MCT-test bleek dat leraren de moeilijkheidsgraad van de toets sterk hebben onderschat; de items waren specifiek geformuleerd en gingen gedetailleerd op de stof in. De relatief hogere prestaties van de leraren op de PROB-test indiceert dat zij wel conceptueel betrokken waren bij de topics die ze leerden.

In Tabel 4 staan de gemiddelden, standaarddeviaties en toetsingen voor de belief-tests. De leerverwachtingen (EXP) van leraren veranderden significant in bijna elke conditie op de posttest. De verandering is in de richting van de leerverwachting die de trainer als ideaal c.q. gewenste eindtoestand beschouwt. In dit opzicht zijn al de condities succesvol in het meer 'realistisch' maken van de beliefs van leraren en het bijstellen van de vaak hoge verwachtingen bij aanvang. De resultaten op de ATT(itude)tests werpen een betekenisvol licht op deze verandering in beliefs. In bijna elke conditie is er een verandering in de attitudetest waar te nemen, en wel zodanig dat de discrepanties toenemen met de door de trainer als ideaal geachte score. Bijvoorbeeld in topic 2 is een van de attitude-dimensies: communicatief – directief. Het blijkt dat leraren verschuiven in hun attitude, en wel weg van de (communicatieve) attitude van de

Tabel 4. Gemiddelden, verschillen, standaarddeviatie en toetsing van de experimentele condities (N=146) voor de belieftests in verschillscores.

	Belief tests (verschil scores)					
	EXP		ARG		ATT	
	M (sd)	t (df)	M (sd)	t (df)	M (sd)	t (df)
topic 1						
-conceptueel - diagnostische treatment (n=28)	2.25 (2.06)	2.18 (24)	-.25 (1.66)	-.42 (24)	1.66 (2.78)	1.80 (27)
-conceptuele treatment (n=31)	3.43 (2.95)	4.95** (19)	.41 (1.97)	.86 (18)	-.39 (3.66)	-.45 (17)
-ervaringsgerichte treatment (n=12)	4.17 (1.42)	5.26** (8)	.44 (1.42)	.94 (8)	-.67 (2.18)	-.60 (12)
topic 2						
-conceptueel - diagnostische treatment (n=29)	2.37 (2.16)	4.77** (25)	-.76 (1.92)	-1.98* (28)	.17 (6.18)	.15 (28)
-conceptuele treatment (n=25) (2.85)	3.83 (18)	3.29* (2.43)	.55 (18)	.74 (.81)	-2.32 (24)	-1.48
-ervaringsgerichte treatment (n=21)	1.84 (2.54)	2.58* (20)	-.54 (1.91)	-.94 (18)	-3.33 (4.85)	-2.82** (20)
Analyse tussen condities						
- posttest	F	4.76	2.14	24.32		
	df	(2,122)	(2,90)	(2,145)		
	p	.01**	.12	.00**		

trainer, hetgeen niet de bedoeling kan zijn geweest van de training. De ervaringsgerichte conditie geeft de grootste discrepantie te zien. Een congruente reactie vertoont de diagnostische conditie.

De andere belieftest (ARG) vertoont geen verschillen tussen condities, ook is er geen toename tussen pre- en posttests in de geëlaboreerdheid van gebruikte argumenten. In dit verband is informatie over de betrouwbaarheid van de toetsen van belang: De product-momentcorrelatie tussen PROB en ARG is hoog ($r=.84$) hetgeen erop wijst dat probleemoplossing in een domein en argumentatie over correct, professioneel handelen sterk samenhangen. De onderlinge correlaties binnen de toetsengroepen (Kennis, Beliefs) zijn hoger dan tussen de toetsengroepen; met als range van kennis-toetsen: $r= .52 - .81$; en belief-toetsen: $r= .65 - .94$; met uitzondering van de attitude test die laag correleert met de andere.

Gedragstoetsen

De gedragstoetsen konden alleen bij topic 2 afgenomen worden en bestonden uit de portfolio en het assessment center. De leraar-prestaties op beide toetsen waren redelijk hoog (gem: 20,48 op een maximum van 30). De verschillen tussen condities waren evenwel niet significant (voor de Portfolio: $F(11,30)= .41$; $p= .53$ en voor de Assessmenttoefeningen: $F(1,17)= 1.31$ met $p= .26$). De correlaties met kennistoetsen en de belieftoetsen was laag (voor POR liggend in de range van .03 - .10 en voor ASS in de range van .21-.26). Dit wijst erop dat de gedragstoetsen niet rechtstreeks afhangen van belief- en kennistoetsen.

Tabel 5. Regressies van belief-toetsen op kennis en gedrag.

op de gecombineerde kennisscore (natoets-somscore)	Verklaarde variantie R ²
leerverwachting	
argumentaties	.24
voorkennis (pretest MCT)	.23
	.03
op gedrag	
gecombineerde voortoets: kennis	.13
gecombineerde voortoets: beliefs	.28

Regressie-analyses werden uitgevoerd om de impact van kennis en belief zowel op elkaar als op gedrag te onderzoeken. Tabel 5 geeft de resultaten.

De belief-toetsen tonen een incongruentie-relatie met kennisacquisitie, d.w.z. er is een negatief verband tussen verandering in beliefscores en de verschillen, d.w.z. toename op de kennistoetsen. Hoe groter de incongruentie in beliefs tussen training en leraar bij voormeting, hoe minder nieuwe kennis wordt opgenomen bij de nameting. De impact (gemeten in β) van de belief-tests op gedrag en ook op nakennis onderstreept de betekenis ervan. Uit dit gegeven blijkt dat beliefs een grotere invloed hebben dan voorkennis. In dit verband is de relatie van de attitude-test met de andere belief-tests van belang: veranderingen in attitudes hangen in negatieve zin samen met de verschuivingen in de andere belief-tests ($r = -.22^*$). Dit lijkt erop te wijzen dat attitude-vorming naast leerstofoverdracht een rol speelt in de training. Gevonden is verder dat de belief-tests negatief samenhangen met de posttest kennistoetsen ($r = -.38^{**}$), hetgeen doet vermoeden dat hoe meer incongruentie in beliefs aanwezig is, des te minder leren optreedt. Van belang is verder het gegeven dat de ARG-test een impact op de gedragstoetsen heeft van $R^2 = .25$. Wat betreft de congruentie-hypothese werd getest of verschillen tussen de trainer's ideale score en actuele leraar-score op de belieftests een verband heeft met kennisacquisitie; deze regressie-analyse leverde een R^2 van .21. Een dergelijke analyse kon door gebrek aan vrijheidsgraden niet voor de gedragstoetsen worden uitgevoerd.

Trainingseffecten

Op de posttests gaven Newman-Keuls vergelijkingen geen significante verschillen te zien van de conceptuele met de ervaringsgerichte condities op de MCT-test ($t(144) = 1.91$ en $p = .063$). Op de posttestscores laat de EXP-test een significant verschil zien van de ervaringsgerichte met de beide conceptuele condities ($p = .021$) ten voordele van de laatste. Op de ARG-test was er geen significant verschil ($p = .71$) tussen de twee trainingstypen. Scores op de ATT-test toonden een significant verschil van de beide conceptuele condities met de ervaringsgerichte conditie ($p = .001$); ook was er een significant verschil tussen de beide conceptuele condities onderling ($t(87) = -4.75$, $p = .001$). Voor de REL-tests zijn afwijkingsscores van de 'similarity' matrix van de leraar met die van de trainer berekend. De somscores per leraar toonden geen significant verschil van de beide conceptuele condities met de ervaringsgerichte conditie ($F(1,68) = .049$, $p = .856$). Er werden geen verschillen tussen de ideale (trainer) score en de actuele leerkracht-score op de belief-tests gevonden tussen de condities (voor EXP, $p = .27$; voor ARG, $p = .27$ en voor ATT, $p = .27$).

Tabel 6. Gemiddelden van de relatieve verschillen tussen de verkregen scores bij leraren en de trainer's verwachting van de score (E) voor wat betreft de kennistoetsen, getoetst op verschillen tussen condities.

	Kennis toetsen - verschil met verwachte score (E)							
	PRO pre	post	t	post	REL F	pre	MCT post	F
voor alle condities	1.51	1.70	-.89	1.05	7.70**	1.56	1.76	-.72
conceptueel-diagnostisch	1.53	1.71		.97		*	*	
conceptueel	1.48	1.69		1.17		*	*	
ervaringsgericht	1.30	1.60		1.00		*	*	

* = niet berekend

Diagnose

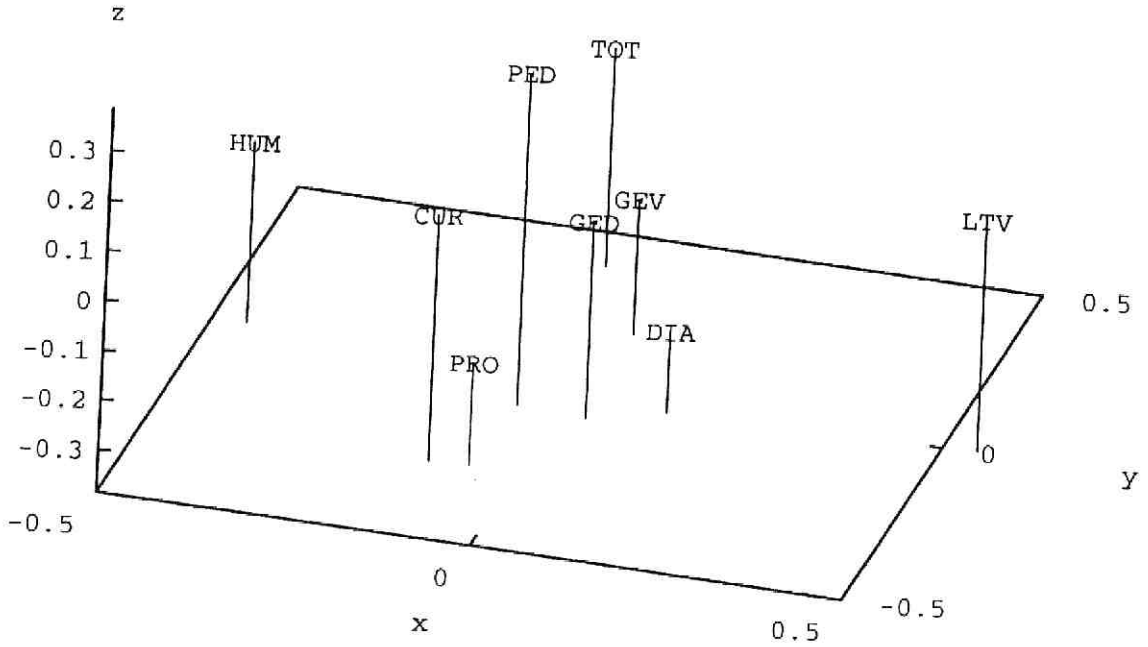
Het ter beschikking stellen van diagnostische informatie aan de trainer leverde geen significante effecten tussen condities ($t(144)=1.91, p=.063$). Het is echter interessant na te gaan in hoeverre de trainers correcte voorspellingen geven van de eindscores van de leraren op de kennistoetsen; dit kan beschouwd worden als indicator van hun 'kennis van de leerkracht'. Tabel 6 geeft de resultaten weer voor de door de trainer verwachte scores en de feitelijke leraar-scores (verschil- of discrepantiescores). De tabel laat geen substantiële verschillen tussen condities zien. Dit wijst erop dat er geen verschillende voorkennis over de leraarcognities bij de trainers aanwezig is, hetgeen doet vermoeden dat de diagnostische informatie wellicht weinig heeft toegevoegd aan de reeds aanwezige 'knowledge of the pupil' bij de trainer. De geringe verschillen tussen feitelijk en verwachte score (de lage discrepantie van de E-score) wijst erop dat de trainers redelijk accurate verwachtingen hebben ten aanzien van de eindscores van de leraren, ongeacht de conditie. Kennelijk hebben trainers zich een redelijk beeld van de cursisten gevormd op het moment van de toetsen.

Diagnostische informatie had wel een differentieel effect op de attitudescores. Het geven van informatie resulteerde in verandering van attitude, en wel in de richting van de training voor de diagnostische conditie ($t(1,44)=4.51; p=.001$); een tegenovergesteld effect werd voor de andere condities gevonden.

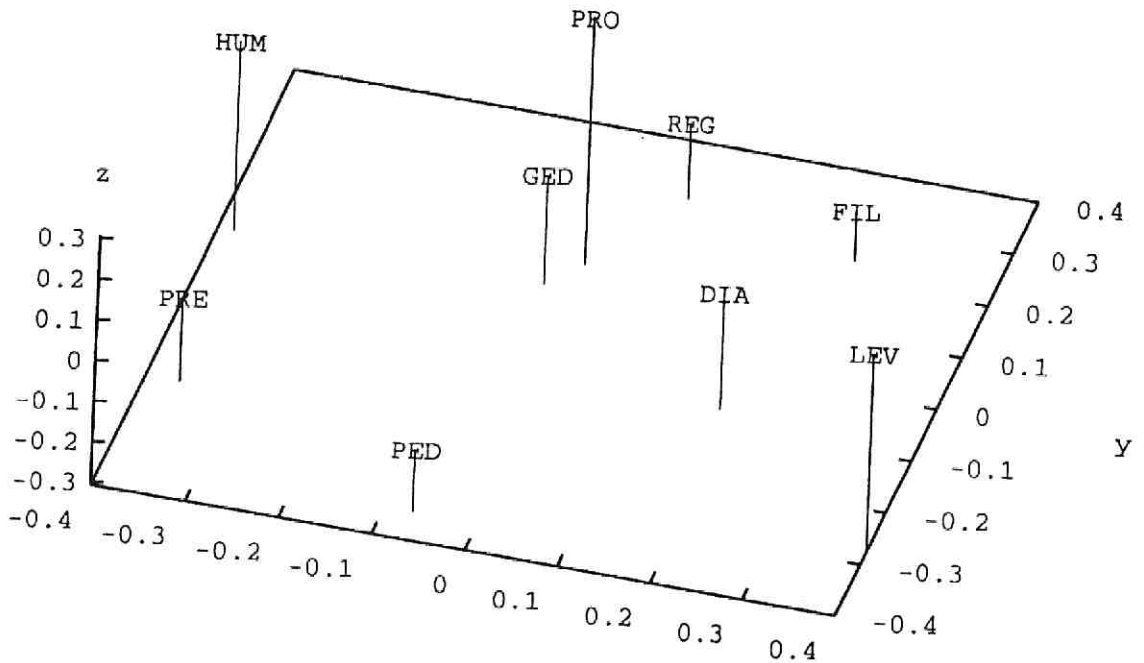
Multidimensionale schalings-technieken bieden de mogelijkheid specifiek te kijken naar de overeenstemming in concepten. De procedure SMACOF - 1b (Stoop & Van der Leeuw, 1982) werd gebruikt voor een niet-metrische analyse van de similarity matrix. Op deze manier is een 'concept'-ruimte direct interpreteerbaar en te vergelijken met de leerstofkaart zoals gehanteerd in de training, d.w.z. een vergelijking is mogelijk op concept niveau van de behandelde met de geleerde leerstof.

Figuur 1 geeft de oplossingen voor de conceptueel diagnostische en de ervaringsgerichte conditie, gekozen als diegenen die de grootste verschillen te zien gaven.

De stress-waarden voor beide condities zijn .12 en .15. Interpretatie van de driedimensionale ruimten is in vergelijking met de leerstofkaart als volgt: dezelfde dimensies keren terug in beide. Voorzover het de diagnostische conditie betreft: dimensie z is te benoemen als 'probleem vs behandeling', handelend over de aard en inzet van het toetsgebruik, dimensie y is te benoemen als: 'curatief vs preventief', handelend over de bedoeling van toetsgebruik; dimensie x is te benoemen als 'humanistisch vs leertheoretische benadering', handelend over het theoretisch kader van toetsgebruik. Deze dimensies vormen een getrouwe afspiegeling van de doelen en



Conceptueel-diagnostische conditie (topic 1)



Ervaringsgerichte conditie (topic 1)

Fig. 1. Configuratie van concept relaties in de conceptueel-diagnostische en ervaringsgerichte conditie.
 Legenda: PRO = probleem; DIA = diagnose; HUM = humanistische benadering; PED = pedagogische behandeling; REG = registratietechniek; GED = gedragsobservatie; PRE/CUR = preventie/curatief; LTV = leertheoretische benadering; DIA = diagnose tests; GEV = gevalsbeschrijving; FIL/TOT = toetsgegevens.

leerstofbehandeling van de trainer. De dimensies in de ervaringsgerichte conditie sporen niet volledig met die in de leerstofkaart, c.q. de doelen vooraf van de training. De dimensies zijn te benoemen als: Dimensie x = humanistische vs leertheoretische benadering; deze dimensie is

goed plaatsbaar. Dimensie z lijkt op die in de diagnostische conditie maar de plaatsing van concepten op deze dimensie laat zich minder goed interpreteren (bijv. het bijeen plaatsen van preventie en humanistische benadering is niet volgens de leerstofstructuur). Dimensie y = behandeling vs registratie is niet goed plaatsbaar omdat het niet overeenkomt met een hoofdthema maar wel kwam het tijdens de ervaringsgerichte bijeenkomsten aan de orde. De concepten zijn bij deze conditie meer verstrooid over de ruimte dan in de conceptueel diagnostische conditie. Ook als gelet wordt op de onderlinge plaatsing van concepten is de diagnostische conditie meer in overeenstemming met de leerstofkaart dan de ervaringsgerichte conditie. Daarbij moet bedacht worden dat de leerstofbehandeling in de laatste conditie minder onder controle van de trainer is dan in de diagnostische conditie. In een dergelijk geval is het ter beschikking stellen van informatie (vooraf) dan ook niet opportuun.

DISCUSSIE

De hypothese die in onderzoek is genomen, luidt dat het meer in overeenstemming zijn van leerkrachtcognities met de trainingsinhoud van de cursus, c.q. hoe meer aansluiting er is bij aanwezige leerkrachtcognities, van positieve invloed is op trainingsresultaten (de congruentiehypothese). Voor dit doel werden een conceptuele en ervaringsgerichte trainingsaanpak tegenover elkaar gesteld. In relatie daarmee staat een tweede meer constructivistische hypothese: Diagnose van aanwezige leerkrachtcognities, in de vorm van beschikbaarheid van diagnostische informatie over beliefs en kennis bij de trainer, heeft een positieve invloed op het meer in overeenstemming brengen van trainingsinhoud en aanwezige cognities bij leraren. De interactie van beide verklaringen is erin gelegen dat ervaringsgerichte training als methode reeds rekening houdt met het expliciteren van aanwezige opvattingen (beliefs), bijvoorbeeld door meer inbreng van cursisten en onderlinge discussie toe te laten. De conceptuele training gericht op 'expliciete uitleg' zou bij uitstek geholpen zijn met diagnostische informatie ten behoeve van de trainer over al of niet beheerste concepten bij de cursist om zo in de leerstofbehandeling ermee rekening te houden.

Verondersteld werd dan ook:

1. Beschikbaarheid van diagnostische informatie bij de trainer heeft een positief effect op de trainingsresultaten in tegenstelling tot geen diagnostische informatie;
2. De conceptuele trainingsaanpak profiteert meer van de beschikbaarheid van diagnostische informatie bij de trainer dan de ervaringsgerichte trainingsaanpak.

De resultaten met betrekking tot de kennisverwerving laten zien dat bijna in elke conditie er weinig nieuw leren heeft plaatsgevonden, m.a.w. geen uitbreiding van kennisstructuren heeft plaats gevonden (Norman, 1982). In een conditie heeft zelfs enige mate van 'afleren' plaatsgevonden. Dit hoeft niet noodzakelijk te betekenen dat er niet geleerd is in de trainingen; herstructurering en 'verfijnen' van kennisstructuren kan hebben plaats gevonden als resultaat van de training. In dit verband is de PROB-test relevant. Leerwinst werd gevonden waar het gaat om probleemoplossingen in het kennisdomein, hetgeen suggereert dat leraren althans vooruit gaan wat betreft het denken en vinden van oplossingen binnen het domein. Deze manier van leren is typisch voor professionals (Boshuizen & Schmidt, 1992; Bennett & Clive, 1993) waar eerder sprake is van kennisherstructurering dan kennistoevoeging (uitbreiding).

In tegenstelling tot de geringe leerwinst in kennis zijn er substantiële verschillen in 'beliefs'-verandering gevonden. Een significante verschuiving vond in elke conditie plaats wat betreft de leerverwachtingen van leraren. De discrepantie tussen de ideaal geachte leeruitkomsten en datgene wat de training feitelijk aanbood, is door de leraren in realistische zin bijgesteld aan het einde van de trainingsperiode. Leraren hebben kennelijk niet-intentioneel geleerd hun aanvankelijke opvattingen bij te stellen. Interessant is de gevonden relatie met de attitudetest. Het is gebleken dat de verschuiving in leerverwachting ten koste is gegaan van de aanvankelijke overeenstemming in attitudes met de trainer. In bijna elke conditie, behalve de diagnostische,

zien we een verschuiving in negatieve zin, d.w.z. weg van de ideaal geachte attitude, zoals beoogd door de training. Dit doet erop wijzen dat, vaak onbedoelde, verschuivingen plaats kunnen vinden als gevolg van de presentatiewijze van de training, hetgeen verband houdt met hoog gestelde leerverwachtingen bij leraren. Wat betreft de meer inhoud-gebonden 'beliefs' (de ARG-test) zien we dat de training geen verandering bewerkstelligt. Het denken over moreel juist geachte, professionele handelwijzen in een domein is niet beïnvloed door de condities. Uit de resultaten van de regressie-analyses komt naar voren dat beliefs een sterke impact vertonen op zowel kennis als, zij het in mindere mate, gedrag; en daarbij de invloed van kennis op gedrag overschaduwet. Dit was met name het geval voor de leerverwachtingen en de argumentatie-test.

Twee veronderstellingen waren geformuleerd met betrekking tot de congruentie-hypothese: 1) hoe dichter de leraaropvattingen liggen bij die van de trainer, c.q. de ideaal geachte opvattingen, hoe meer leren er plaats vindt en 2) hoe dichter de kennisscores van de leraar bij die van de trainer's verwachtingen liggen, hoe meer leren er plaats vindt.

Het verschil tussen de ideale scores en de beliefs van leraren heeft een aanzienlijke impact op de kennisacquisitie. Dit effect werd voornamelijk veroorzaakt door verschillen ten aanzien van EXP en verschillen ten aanzien van ARG. De relatie was invers, d.w.z. hoe groter de verschillen tussen ideaal en feitelijke beliefs des te minder leren vindt er plaats. Deze bevinding is ondersteuning voor de congruentie hypothese. De verwachting van de trainer over leraar(na)kennis in relatie tot de feitelijke (na)kennis heeft niet een direct zichtbaar verband met variatie in leerwinst.

De diagnose – hypothese is niet zondermeer ondersteund door dit onderzoek. Er werd verondersteld dat de conceptuele conditie meer van diagnostische informatie zou profiteren dan de ervaringsgerichte. Hoewel de tendens in deze richting wijst, zijn geen significante verbanden gevonden. Dit wil niet zeggen dat de trainers niet op de hoogte waren van aanwezige opvattingen en kennis bij de leraren, integendeel de gegevens wezen uit dat deze kennis beschikbaar was, maar niet of nauwelijks leidde tot verandering van presentatie ofwel het rekening houden met deze diagnostische informatie in de leerstofbehandeling waardoor een 'matching' van instructie en leerproces kan ontstaan.

De REL-data laten echter zien dat de diagnostische conditie duidelijk meer profiteerde van de diagnostische informatie dan de meer impliciet verkregen diagnose in de ervaringsgerichte conditie. In de laatste conditie konden belangrijke afwijkingen met de leerdoelen worden geconstateerd. De leerstofbehandeling c.q. dekking van centrale concepten in de diagnostische conditie was zoals bedoeld; de ervaringsgerichte conditie had daarentegen rekening te houden met eigen inbreng van leerinhouden door de cursist.

Concluderend kan gesteld worden dat deze studie ondersteuning geeft voor de bewering dat veranderingen in kennisstructuren samenhangen en beïnvloed worden door veranderingen in opvattingen van leraren. Het bevorderen van de toegankelijkheid en hanteerbaarheid, en eventueel uitbreiding, van kennisstructuren bij professionals is, zo laat zich afleiden, niet enkel een zaak van het presenteren van de correcte informatie of concepten ('cold conceptual change'; Pintrix, Marx & Boyle, 1993) maar kan eerder gezien worden als het toevoegen aan kennisstructuren die zijn "beveiligd" door evaluaties over de waarde en praktische bruikbaarheid van nieuw aangeboden kennis in het licht van bestaande kennis.

In deze studie werd gevonden dat veranderingen in leerverwachtingen kunnen worden beïnvloed door de training (omdat de training een context vormt waaraan de leraar zich via deelname aanpast, dient aan te passen). Inhoud-gerelateerde beliefs zijn echter moeilijker te veranderen dan door context of situatie alleen (Palincsar, 1989; Nuttin, 1974) of zelfs door gerichte instructie alleen. Een trainingsaanpak die dit laatste bewerkstelligt ligt niet direct voorhanden (zie echter Tillema & Imants, in press); de in het onderzoek gehanteerde trainingsaanpakken hadden althans geen directe invloed op inhoudsgerelateerde opvattingen.

Ook is gebleken dat – als een voor de handliggende tactiek – het geven van informatie over de cursisten aan de trainer niet het gewenste effect heeft. Maar voordat een oordeel hierover wordt geveld is het wellicht zinvol de impact van de treatment eerst te vergroten. Diagnostische

informatie is wellicht niet voldoende om de presentatie van de trainer te beïnvloeden; directe matching van presentatie aan (diagnose) van het leerproces is daarvoor nodig. Blijkbaar is alleen informatie over de discrepantie in attitudes in ogenschouw genomen door de trainers en niet die over kennisargumentaties, probleemoplossingen in het vakgebied.

De studie geeft aan dat een investering om kennisstructurering van professionals te bevorderen rekenschap moet geven aan het professionele belief-systeem van leraren en de manier waarop zij kennis beoordelen (Clark, 1989); kennis is niet enkel conceptueel van aard maar verbonden en vermengd met evaluatieve, motivationele betekenissen en actiepotenties (Strike & Posner, 1992) Kennisacquisitie is, voor professionals althans, meer een kwestie van kennisherstructurering dan kennistoevoeging.

NOOT

Deze studie werd mogelijk gemaakt door een subsidie van de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) te Den Haag.

LITERATUUR

- Berliner, D.C. (1987). Ways of thinking about students and classrooms. In J. Calderhead, *Exploring teacher's thinking*. London: Cassell.
- Bennet, N. (1992). Perspectives on knowledge bases for teaching. In Tj. Plomp (ed.), *Proceedings of the European Conference on Educational Research*. Enschede: University of Twente, Department of Education.
- Bennett, N. & Clive, J. (1993). *Learning to Teach*. London: Routledge.
- Bromme, R. (1987). Teachers' assessment of student difficulties and progress in understanding. In J. Calderhead, *Exploring teachers' thinking*. London: Cassell.
- Boshuizen, H.P. & Schmidt H.G. (1992). The role of biomedical knowledge in clinical reasoning by experts, intermediates and novices. *Cognitive science*, 16, 153-184.
- Carpenter, T.P., Fennema, E., Peterson, P.L., Chiang, C-P. & Loef, M. (1989). Using knowledge of children's mathematics thinking in classroom teaching: an experimental study. *American Educational Research Journal*, 26, 4, 499-531.
- Carter, K. (1990). Teachers' knowledge and learning to teach. In W.R. Houston, *Handbook of Research on Teacher Education* (p. 291-310). New York: Macmillan.
- Clark, C.M. (1989). Asking the right questions about teacher preparation. In J. Lowyck & C.M. Clark, *Teacher thinking and professional action*. Leuven: Leuven University Press.
- Corporaal, A.H. (1988). *Bouwstenen voor een opleidingsdidactiek*. Proefschrift Rijksuniversiteit Leiden. De Lier: Academisch Boeken Centrum.
- Cruickshank, D.R. & Metcalf, K.N. (1990). Training within teacher preparation. In W.R. Houston, *Handbook of Research on Teacher Education* (p. 469-497). New York: Macmillan.
- Darling-Hammond, L. (1990). Teachers and teaching: signs of a changing profession. In W.R. Houston, *Handbook of Research on Teacher Education* (p. 267-290). New York: Macmillan.
- Doyle, W. (1990). Themes in teacher education research. In W.R. Houston, *Handbook of Research on Teacher Education* (p. 3-24). New York: Macmillan.
- Elbaz, F. (1983). *Teacher Thinking, a Study of Practical Knowledge*. London: Croom Helm.
- Gliesman D.H. (1984). Changing teacher performance. In L.G. Katz & J.D. Raths, *Advances in teacher education (I)* (p. 95-111). Norwood: Ablex.
- Goodlad, J.I., Soder R. & Sirotnik, K.A. (1990). *The moral dimensions of teaching*. San Francisco: Jossey Bass.
- Gudmundottir, S. (1991). Ways of seeing are ways of knowing. *Journal of Curriculum Studies*, 23, 5, 409-422.
- Guskey, T.R. (1986). Staff development and the process of teacher change. *Educational Researcher*, 15, 5, 5-12.
- Joyce, B.R., Murphy, C., Showers, B. & Murphy, J. (1989). School renewal as cultural change. *Educational Researcher*, 48, 70-77.

- Kagan, D.M. (1990). Ways of evaluating teacher cognitions, inferences concerning the Goldilock principle. *Review of Educational Research*, 60, 3, 419-470.
- Kok, W. & Tillema, H.H. (1981). Instrument voor de analyse en representatie van leerstofstructuren. *Tijdschrift voor Onderwijsresearch*, 6, 4, 182-199.
- Nuttin, J. (1974). *The illusion of attitude change*. New York: Academic Press.
- Palincsar, A.S. (1989). Response to Brown Collins and Duguid's: Situated cognition and the culture of learning. *Educational Researcher*, 18, 4, 5-7.
- Prawat, R.S. (1989). Promoting access to knowledge strategy and disposition in students: a research synthesis. *Review of Educational Research*, 59, 1, 1-41.
- Pajares, M.F. (1992). Teacher beliefs and educational research: cleaning up a messy construct. *Review of Educational Research*, 62, 307-332.
- Peterson, P.L. (1988). Teachers and students cognitional knowledge for classroom teaching and learning. *Educational Researcher*, 17, 5, 5-14.
- Pintrich, P.R., Marx, R.W. & Boyle, R. (1993). Beyond cold conceptual change. *Review of Educational Research*, 63, 2, 167-200.
- Richardson, V. (1990). Significant and worthwhile change in teaching practice. *Educational Researcher*, 19, 7, 10-18.
- Schon, D.A. (1983). *The reflective practitioner. How professional think in action*. New York: Basic Books.
- Shulman, L.S. (1986). Those who understand: knowledge growth in teaching. *Educational Researcher*, 15, 2, 4-14.
- Smith, D.C. & Neale, D.C. (1989). The construction of subject matter knowledge in primary science teaching. *Teaching and Teacher Education*, 5, 1, 1-20.
- Stallings, J. (1982). Follow through: a model for inservice teacher training. *Curriculum Inquiry*, 9, 163-181.
- Stoop, I. & De Leeuw, J. (1982). *How to use SMACOF-1B*. Leiden: Department of Datatheory, University Leiden.
- Strike, K.A. & Posner, G.J. (1992). A revisionist theory of conceptual change. In R. Duschl & R. Hamilton, *Philosophy of science, cognitive psychology and educational theory and practice*. Albany: SUNY.
- Tillema, H.H. (1993). *Assessment en opleiden in organisaties*. Capita selecta 14. Deventer: Kluwer Bedrijfswetenschappen.
- Tillema, H.H., De Jong, R. & Mathijssen, C. (1990). Conceptual or experience based learning of teachers. *Teaching and Teacher Education*, 6, 2, 165-172.
- Tillema, H.H. & Imants, J. (in press). Training in context; professional development of teacher in the school. In T. Guskey & M. Huberman, *Perspectives on professional development in education*. London: Falmer Press
- Tillema, H.H. & Veenman, S. (1987). Conceptualising training methods in teacher education. *International Journal of Educational Research*, 11, 5, 519-529.
- Van Manen, M. (1977). Linking ways of knowing with ways of being practical. *Curriculum Inquiry*, 6, 205-228.
- Verloop, N. (1989). *Interactive cognitions of student teachers*. Doctoral dissertation, University of Leiden, The Netherlands.
- Vosniadou, S. & Brewer, W.F. (1987). Theories of knowledge restructuring in development. *Review of Educational Research*, 57, 51-67.
- Wilson, B. & Cole, P. (1992). A review of cognitive teaching models. *Educational training & development*, 39, 4, 47-64.
- Zeichner, K. & Gore, J.M. (1990). Teacher socialization. In W.R. Houston, *Handbook of Research on Teacher Education*. New York: Macmillan.

Manuscript ontvangen 4-10-1993

Definitieve versie ontvangen 14-2-1994

Eindexamens in de tijd vergeleken: het einde van de wet van Posthumus

H.D. Webbink

Stichting voor Economisch Onderzoek van de Universiteit van Amsterdam*

ABSTRACT

In this article an intertemporal comparison of the difficulty of final exams of students of the highest levels of secondary education is carried out. A model of passing the final test is estimated for 1982 and 1991. The parameters of this model are an expression of the difficulty of the exam. By using the parameters for predicting the number of students of different years passing the final test we can identify changes over time. Moreover, the quality of students of different years can be compared. We find that the final exam of 1982 was more difficult than the exam of 1991. The quality of the student cohort of 1982 was comparable with the quality of the 1991 cohort.

1. INLEIDING

Volgens de wet van Posthumus (1942) wordt van elke groep leerlingen of studenten steeds een kwart weggeselecteerd. Omdat er geen objectieve beoordelingscriteria van de geschiktheid van leerlingen en studenten zijn worden de zwakste leerlingen afgewezen. Dit selectie criterium heeft tot gevolg dat de zwaksten uit de groep als ongeschikt worden aangemerkt. Studenten kunnen hierdoor het slachtoffer worden van een toevallige verdeling van de capaciteiten in hun groep. Ook is het mogelijk dat studenten die toevallig in een zwakke groep terecht zijn gekomen naar een te hoge opleiding doorstromen. Uiteraard is de klassieke wet van Posthumus ook van toepassing op het eindexamen, het belangrijkste selectie criterium voor leerlingen in het Voortgezet Onderwijs. De selectie door middel van het eindexamen is onderhevig aan een zekere willekeur omdat enerzijds de moeilijkheidsgraad van eindexamen in de tijd niet constant hoeft te zijn en anderzijds er verschillen kunnen bestaan in de geschiktheid van cohorten eindexamenkandidaten.

In dit artikel wordt een methode beschreven (en toegepast) waarmee veranderingen in de geschiktheid van cohorten leerlingen en veranderingen in de zwaarte van examens geïsoleerd worden. Door een econometrische simulatie met een model voor het slagen voor het eindexamen laten we als het ware twee cohorten eindexamenkandidaten (van 1982 en 1991) elkaars examen maken. Hierdoor wordt het mogelijk om de zwaarte van eindexamens van verschillende jaren te vergelijken. Ook, en zeker niet minder belangrijk, kan door deze simulatie de geschiktheid van verschillende cohorten eindexamenkandidaten worden vergeleken¹.

Met de in dit artikel besproken methode wordt een betere selectie mogelijk die kan leiden tot het einde van de wet van Posthumus. Ook kunnen de analyseresultaten relevant zijn voor de huidige beleidsdiscussies over selectie en kwaliteit in het Hoger Onderwijs (zie HOOP '94). Het gaat dan onder meer over de mogelijke daling van de kwaliteit van de instroom in het Hoger Onderwijs, de aansluiting tussen Vwo/Havo en het Hoger Onderwijs en de rendementen in het Voortgezet en Hoger Onderwijs.

Op voorhand willen we er echter op wijzen dat dit artikel in de eerste plaats gaat over de methode. De empirische toepassing dient vooral gezien te worden als illustratie van de methode. Bij de beleidsconclusies is enige voorzichtigheid geboden. In de eerste plaats omdat de data zijn verzameld door verschillende groepen onderzoekers, met verschillende steekproeven en soms enigszins afwijkende vraagstellingen. In de tweede plaats bestaat er met de specificatie van het

* Adres: Roetersstraat 11, 1018 WB Amsterdam.

model voor het slagen voor het eindexamen weinig ervaring. In de literatuur zijn hierover weinig voorbeelden te vinden.

De opzet van het artikel is als volgt. In paragraaf 2 wordt de methode uitgelegd en wordt het model voor het slagen voor het eindexamen besproken. Paragraaf 3 bespreekt de data. De schattingsresultaten worden vervolgens in paragraaf 4 besproken. De conclusies zijn te vinden in paragraaf 5.

2. HET VERGELIJKEN VAN EINDEAMENS VAN VERSCHILLENDE JAREN

Het vergelijken van eindexamens is lastig omdat zowel de populatie van eindexamenkandidaten kan veranderen als de moeilijkheidsgraad van de eindexamens. Een direct gevolg is dat eindexamencijfers in de tijd niet vergelijkbaar zijn, te meer daar ze soms aangepast worden op grond van slagingspercentages. Voor een goede vergelijking van de moeilijkheidsgraad van eindexamens (en van eindexamencijfers) zou een bepaald cohort leerlingen twee of meer eindexamens moeten maken. Als van dit cohort een groter deel leerlingen voor een bepaald eindexamen zou slagen dan kunnen we aannemen dat dit examen minder moeilijk was dan de andere examens. In werkelijkheid maakt elk cohort eindexamenkandidaten echter maar één examen. We kunnen echter wel *simuleren* dat een bepaald cohort leerlingen een eindexamen van een ander jaar maakt. Hiervoor maken we gebruik van een model voor het slagen voor het eindexamen. In dit model wordt het slagen voor het eindexamen verklaard uit individuele kenmerken van de leerlingen zoals de capaciteiten en de motivatie. In hoeverre bepaalde individuele kenmerken van invloed zijn op de slaagkans is afhankelijk van de moeilijkheidsgraad van het eindexamen. Bijvoorbeeld de capaciteiten van leerlingen zullen een sterkere invloed hebben op de slaagkans bij een moeilijk eindexamen dan bij een gemakkelijk eindexamen. Bij het schatten van het model zullen we dan een grotere coëfficiënt (parameter) vinden bij een moeilijk examen dan bij een gemakkelijk examen. Dit betekent dat de parameters van het model (de geschatte coëfficiënten) een uitdrukking vormen van de zwaarte van het examen. Wanneer we een dergelijk 'slaagmodel' voor twee verschillende jaren (en dus voor twee cohorten leerlingen) schatten krijgen we voor elk jaar een set parameters die de moeilijkheidsgraad van het examen vastlegt. Met deze parameters kunnen we vervolgens voorspellen welk deel van een bepaald cohort leerlingen zou slagen voor een bepaald eindexamen, aangezien we de individuele kenmerken van deze cohorten kennen. Aan de hand van de slagingspercentages op de gesimuleerde eindexamens kunnen we vervolgens de moeilijkheidsgraad van de eindexamens en de geschiktheid van de cohorten leerlingen vergelijken.

Een model voor het slagen voor het eindexamen van Vwo of Havo

Welke leerlingen slagen voor het eindexamen en welke leerlingen halen het niet? In de omvangrijke literatuur over studiesucces en -uitval is over deze specifieke vraag weinig bekend. Kodde (1984) geeft een verklaringsmodel van de slaagkans voor het eindexamen gebaseerd op algemene noties over studiesucces. De slaagkans is hierin afhankelijk van de capaciteiten en de motivatie/inzet van de leerlingen. De capaciteiten van de leerlingen worden niet direct gemeten maar benaderd door variabelen als zittenblijven en studiemotieven/studieplannen. Verwacht wordt dat de studieplannen afhankelijk zullen zijn van de gepercipieerde capaciteiten van de leerlingen. Op soortgelijke wijze wordt ook de motivatie van de leerlingen benaderd door een aantal factoren, bijvoorbeeld door het belang dat men toekent aan zelfontplooiing en in welk stadium de studieplannen gemaakt waren. Verwacht wordt dat leerlingen die gaan studeren omdat ze het onderwerp van de studie interessant vinden waarschijnlijk sterk gemotiveerd zijn voor de komende studie en voor het slagen voor het eindexamen. Er wordt dus een positief effect verwacht op het slagen voor het eindexamen. Daarentegen verwachten we een negatief effect voor leerlingen die gaan studeren omdat ze nog niet willen werken, omdat dit een negatieve motivatie is. Ook zijn in het model de variabelen schooltype, geslacht en regio opgenomen. Wat betreft de

regionale invloed op de slaagkans wordt de mogelijkheid genoemd dat er regionale verschillen in het schooladvies aan het eind van de lagere school bestaan. Dit kan leiden tot selectie en verschillen in capaciteiten van leerlingen op eenzelfde schoolniveau.

3. DE DATA

In de analyse wordt gebruik gemaakt van gegevens van twee cohorten eindexamenkandidaten van Havo en Vwo. Het eerste cohort leerlingen is in 1982 in het kader van het project *'De vraag naar Hoger onderwijs'* twee keer bevraagd; voor en na het eindexamen (Kodde en Ritzen, 1985). Deze eerste twee zogenaamde 'Belangstellingsenquêtes' kenden een respons van ruim 15 duizend leerlingen. In de enquêtes werden relatief weinig vragen gesteld. In 1983 volgden nog twee meer uitvoerige 'Belangstellingsenquêtes', één onder leerlingen en één onder de ouders van de eindexamenkandidaten, waaraan door ongeveer 4000 leerlingen is deelgenomen. In de navolgende analyse wordt gebruik gemaakt van de gegevens van de eerste twee 'Belangstellingsenquêtes', dus van een bestand van ruim 15 duizend respondenten.

Het tweede cohort leerlingen is het zogenaamde pré-HO panel uit het project *'Verder Studeren'*. Dit is een longitudinaal onderzoek onder scholieren en studenten dat in 1991 van start is gegaan. Het is de bedoeling om de scholieren en studenten minstens tot eind 1995 jaarlijks te bevragen over hun onderwijservaringen en -plannen². Het pré-HO panel betreft eindexamenkandidaten Vwo, Havo, Mbo, Mavo en Lbo van 1991. Uit dit pré-HO panel zijn alleen de eindexamenkandidaten Vwo en Havo geselecteerd omdat alleen deze groep in 1982 bevraagd is.

4. SCHATTINGSRESULTATEN

De kans op het slagen voor het eindexamen is geschat met een logitmodel³ waarin het al of niet slagen voor het examen de afhankelijke variabele is. De individuele kenmerken van de leerlingen zijn de verklarende variabelen in het model. Het hiervoor beschreven model kan vrijwel geheel worden toegepast op het datamateriaal dat in 1982 en in 1991 verzameld is. Er zijn evenwel een paar verschillen tussen de data van deze jaren. In de eerste plaats is de indeling naar schooltypen in 1991 minder uitvoerig gevraagd dan in 1982⁴. In de tweede plaats zijn de studieplannen in 1982 meer gedetailleerd gevraagd, dat wil zeggen er waren meer antwoordcategorieën. In de derde plaats is in 1982 door Kodde een proxy voor het zittenblijven genomen (op grond van leeftijd en eerdere diploma's). In 1991 is het zittenblijven direct gevraagd.

Ook dient een kanttekening gemaakt te worden bij het toepassen van het model van 1982 op het cohort leerlingen uit 1991. Als gevolg van sociaal-culturele veranderingen is het mogelijk dat de betekenis van bepaalde variabelen in de tijd is veranderd. Een voorbeeld hiervan zou de variabele 'studieplan' kunnen zijn. Deze variabele diende in 1982 als operationalisatie van de capaciteiten maar ook van de motivatie van de leerlingen. Tussen 1982 en 1991 is er ten aanzien van de keuze al of niet verder te studeren nogal wat veranderd. In 1982 was verder studeren alleen weggelegd voor de betere leerlingen, in 1991 studeert vrijwel iedereen verder⁵. Dit betekent dat de variabele 'studieplan' in 1991 niet meer hetzelfde meet als in 1982 en derhalve minder goed bruikbaar is in het model van de slaagkans.

Gezien het bovenstaande is ervoor gekozen om voor 1982 en voor 1991 een model te schatten dat rekening houdt met de genoemde verschillen en kanttekeningen en dat dus op een aantal plaatsen afwijkt van het eerder door Kodde geschatte model.⁶ Een beschrijving van de in de analyses gebruikte variabelen en de gemiddelden en de standaarddeviaties in beide jaren zijn gegeven in de appendix. De schattingsresultaten van het model voor het slagen voor het eindexamen zijn gegeven in Tabel 1. Gegeven zijn de coëfficiënten en de standaardfouten⁷; een positief effect betekent een grotere kans om te slagen voor het eindexamen.

Tabel 1. De slaagkans voor het eindexamen in 1982 en in 1991.

	1982		1991	
	parameter	st. fout	parameter	st. fout
constante	2,16	0,26	1,03	1,14
geslacht (man=1)	0,31	0,06	0,41	0,29
schooltype t.o.v. Vwo-B				
Havo	-0,31	0,09	-0,81	0,47
Vwo-A	-0,02	0,10	-0,40	0,53
gedoubleerd	-0,36	0,09	-0,42	0,27
aantal exacte vakken	0,01	0,02	0,13	0,13
aantal talen	0,16	0,03	0,11	0,20
eerdere diploma's	-0,16	0,06	-0,32	0,31
vroege keuze	-0,10	0,06	2,38	0,27
studieplannen t.o.v. niet studeren				
voltijd studeren	0,48	0,09	1,37	0,52
deeltijd studeren	0,38	0,15	1,02	0,73
weet niet	-0,13	0,13	1,99	1,00
regio t.o.v. zuid				
noorden	-0,17	0,10	-0,50	0,44
oosten	-0,03	0,09	0,35	0,43
westen	-0,27	0,07	-0,57	0,35
motieven				
zelfontplooiing	0,18	0,07	0,32	0,28
kans op baan	-0,03	0,09	-0,10	0,49
nog niet willen werken	-0,03	0,07	0,11	0,26
interesse in onderwerp van studie	0,30	0,13	0,31	0,46
bepaald beroep willen uitoefenen	-0,42	0,11	-0,24	0,42
loglikelihood	-4.470,7		-215,7	
likelihood ratio test	255,7		143,2	
vrijheidsgraden	19		19	
aantal geslaagden	9737		742	
aantal waarnemingen	11.329		833	

De schattingsresultaten van het model van 1982 komen overeen met hetgeen a priori verwacht werd. Dit geldt ook voor de resultaten van 1991. Een verschil is evenwel dat een aantal coëfficiënten niet significant is hetgeen waarschijnlijk veroorzaakt wordt door het kleinere aantal waarnemingen in 1991. De likelihood ratio test geeft voor beide jaren een significant resultaat. In beide jaren slagen jongens vaker dan meisjes. Havo'ers hebben een kleinere slaagkans dan Vwo'ers en de slaagkans van Vwo-A leerlingen verschilt niet significant van de slaagkans van Vwo-B leerlingen.

Ook bij de indicatoren van de **capaciteiten** van de leerlingen vinden we vergelijkbare resultaten voor 1982 en 1991. Leerlingen die al eens gedoubleerd zijn hebben een kleinere slaagkans voor het eindexamen. Dit geldt ook voor leerlingen met diploma's van lagere schooltypen. Zij

hebben een significant lagere slaagkans dan leerlingen die direct naar het betreffende schooltype zijn gegaan. Aannemelijk is dat dit te maken heeft met selectie naar schooltype op jongere leeftijd op basis van verschillen in capaciteiten. Bij de variabelen over de samenstelling van het vakkenpakket zien we dat in 1982 leerlingen vaker slagen naarmate ze meer talen in hun pakket hebben, in 1991 wordt dit niet gevonden. Het aantal exacte vakken is niet van invloed op de kans om te slagen voor het eindexamen.

Studieplannen zijn in het model opgenomen als indicatoren van de *motivatie* van de leerlingen. We zien dat leerlingen die in voltijd of in deeltijd willen gaan studeren een grotere kans hebben om te slagen voor het eindexamen dan leerlingen die van plan zijn om niet te gaan studeren. Dit effect kan volgens Kodde (1984) verklaard worden door de percepties van de capaciteiten van de leerlingen. Een verschil tussen 1982 en 1991 wordt gevonden bij de groep leerlingen die nog geen studieplannen heeft. In 1991 heeft deze groep een grotere kans om te slagen dan de leerlingen die niet willen gaan studeren. Een ander verschil is dat leerlingen die, al in een vroeg stadium namelijk voor het examen wisten wat te gaan doen na het eindexamen, in 1991 wel een grotere slaagkans hebben maar in 1982 niet.

De *regio-variabelen* laten zien dat leerlingen in het Westen van het land een lagere slaagkans hebben dan leerlingen in de rest van Nederland. Regionaal verschillende zelf-selectieprocessen spelen hierbij wellicht een rol⁸. Opvallend is dat dit effect in beide jaren gevonden wordt.

Tot slot zijn nog een aantal motieven om te gaan studeren in het model opgenomen. Degenen die zelfontplooiing belangrijk vinden slagen vaker dan andere leerlingen, in 1982 en in 1991. Nog niet willen gaan werken heeft een negatief effect op de slaagkans maar het effect is niet significant. Ditzelfde geldt voor het motief 'kans op een baan'. Leerlingen die geïnteresseerd zijn in het onderwerp van de (komende) studie hebben een grotere kans om te slagen, dit geldt alleen in 1982. Sterk beroepsgerichte leerlingen slagen daarentegen minder vaak (niet significant in 1991).

Simulatieresultaten

Het is nu mogelijk om met de parameters van het slaagkansmodel van 1982 te voorspellen welk deel van het cohort leerlingen van 1991 zou slagen voor het eindexamen van 1982 en vice versa. In feite wordt de totale verandering in slaagpercentages tussen 1982 en 1991 uiteen gelegd in een gedeelte veroorzaakt door een verandering in de samenstelling van de populatie en een gedeelte veroorzaakt door veranderingen in het eindexamen. Gomulka & Stern (1990) geven de volgende uitdrukking voor de decompositie van een verandering in de tijd bij een binomiale afhankelijke variabele:

$$\hat{y}^{91} - \hat{y}^{82} = \{P(\beta^{91}, X^{91}) - P(\beta^{82}, X^{91})\} + \{P(\beta^{82}, X^{91}) - P(\beta^{82}, X^{82})\}$$

Links staat de verandering in slaagpercentages voor het eindexamen tussen 1982 en 1991, en $P(\beta^i, X^j)$ is het gemiddelde over de steekproef X^j van de voorspelde slaagkans voor het eindexamen met de parameters van jaar i . De eerste term tussen haakjes aan de rechter kant geeft het effect van een verandering in de parameters (de zwaarte van de examens), de tweede term tussen haakjes geeft het effect van een verandering van de verdeling van de steekproefkenmerken (de leerlingenpopulatie).

De resultaten van deze decompositie zijn gegeven in tabel 2.

Tabel 2. Geslaagden in 1982 en 1991 voor beide examens (percentages).

populatie	parameters	
	1982	1991
1982	85,9	88,0
1991	85,1	89,1

In 1982 slaagde bijna 86 procent van de Havo- en Vwo-leerlingen die deelnamen aan het eindexamen. In 1991 slaagde ruim 89 procent van de leerlingen. Hadden de leerlingen uit 1991 het examen van 1982 gemaakt dan zou iets meer dan 85 procent geslaagd zijn. Wanneer de leerlingen van het cohort 1982 het examen van 1991 gemaakt hadden dan was 88 procent geslaagd. We kunnen hieruit twee dingen afleiden. In de eerste plaats zijn er duidelijke verschillen in de slaagpercentages voor de twee eindexamens. Het deel van de leerlingen van beide cohorten dat slaagt voor het examen van 1982 is kleiner dan het deel dat slaagt voor het examen van 1991. Dit is een aanwijzing dat het examen van 1982 moeilijker was dan het examen van 1991. De verschillen tussen de populaties bij het slagen voor een bepaald eindexamen zijn niet groot. We kunnen hieruit afleiden dat de populaties overeenkomen qua geschiktheid. Dit wil evenwel niet zeggen dat er geen kwaliteitsverschillen bestaan wat betreft de instroom in het Hoger Onderwijs. Aangezien een veel groter deel van de geslaagde leerlingen in 1991 gaat studeren dan in 1982 zijn verschillen in de kwaliteit van de instroom populaties zeer goed mogelijk⁹.

De verschillen in slaagpercentages tussen de examens van 1982 en 1991 zijn echter niet dramatisch. Of het hier om incidentele dan wel structurele verschillen gaat is op grond van de beschikbare data niet te zeggen; hiervoor zijn gegevens van meer cohorten leerlingen noodzakelijk.

Het einde van de wet van Posthumus?

Stel een eindexamencommissie beschikt over de resultaten van tabel 2 (of over soortgelijke resultaten voor andere jaren). In dat geval lijkt het mogelijk om de willekeur van de wet van Posthumus te beperken. Wanneer bijvoorbeeld het eindexamen van 1982 als ijkpunt genomen zou worden, zou de commissie kunnen adviseren om de normering voor het examen van 1991 zodanig aan te passen dat het percentage geslaagden zou dalen. Ook is het denkbaar om de hier toegepaste methode te gebruiken voor afzonderlijke vakken. Hiervoor zouden dan voor verschillende jaren (regressie)modellen geschat moeten worden waarin het eindexamencijfer voor een bepaald vak verklaard wordt uit individuele kenmerken van leerlingen. De verandering in het eindexamencijfer van een bepaald vak kan dan weer op dezelfde wijze uiteen gelegd worden in een gedeelte veroorzaakt door een verandering in de samenstelling van de populatie en een gedeelte veroorzaakt door een verandering in het eindexamen. Dit kan vervolgens een belangrijk hulpmiddel zijn bij aanpassingen van de normering van een bepaald vak.

5. CONCLUSIES

De voorgaande analyses laten zien dat van de cohorten leerlingen van 1982 en 1991 een groter deel slaagt voor het eindexamen van 1991 dan voor het eindexamen van 1982. Het is daarom aannemelijk dat het eindexamen van 1982 moeilijker was dan dat van 1991. Dit betekent dat het in 1991 gemakkelijker was een diploma te verwerven dat recht geeft op deelname aan het Hoger Onderwijs. Gezien de stijging van de deelname aan het Hoger Onderwijs in de afgelopen jaren is het zeer goed mogelijk dat de kwaliteit van de instroom daarom lager is dan 10 jaar geleden. Wat betreft de geschiktheid van de cohorten eindexamenkandidaten van 1982 en 1991 worden nauwelijks verschillen gevonden.

Met deze resultaten is het mogelijk de willekeur van de wet van Posthumus te beperken. Wanneer 1982 als ijkpunt wordt genomen zou hiervoor de normering van het eindexamen van 1991 zodanig moeten worden aangepast dat het aantal geslaagden enigszins zou dalen. De in dit artikel beschreven analysemethode zou daarmee richting kunnen geven bij de aanpassing van normeringen.

NOTEN

1. In de Verenigde Staten wordt de geschiktheid van cohorten scholieren al vele jaren systematisch gemeten. Veel universiteiten gebruiken de Scholastic Aptitude Test, de SAT, als toelatingsexamen. Deze test meet niet de intelligentie of de kennis van scholieren maar het leervermogen. De resultaten van de afgelopen jaren laten een duidelijke afname van de SAT-scores zien hetgeen betekent dat het leervermogen van de scholieren en dus ook de kwaliteit van de instroom in het hoger onderwijs afneemt (zie onder andere A Nation at Risk (1984) en J. Bishop (1990)).
2. Over dit onderzoek zijn inmiddels twee rapporten verschenen in de serie 'Verder Studeren' van het ministerie van Onderwijs en Wetenschappen (zie de Jong, e.a. (1992) en Webbink, e.a. (1993)).
3. Zie o.a. Cramer (1991).
4. In 1991 is slechts gevraagd of men op Havo of Vwo zit. Daarentegen werd in 1982 onderscheid gemaakt tussen Gymnasium-A, Gymnasium-B, Gymnasium-ongedeeld, Atheneum-A, Atheneum-B, Atheneum-ongedeeld, Vwo-ongedeeld en Havo.
5. Zie de Jong, e.a. (1992) en Webbink, e.a. (1993).
6. In Webbink e.a. (1993) zijn drie modellen geschat die in meerdere of minder mate rekening houden met de genoemde verschillen tussen 1982 en 1991. Het hier gepresenteerde model bevat alle variabelen uit de drie afzonderlijke modellen en geeft een betere verklaring voor het slagen voor het eindexamen.
7. Vaak worden niet de coëfficiënten gepresenteerd maar de afgeleiden, deze geven namelijk het directe effect op de slaagkans. Hier geven we de coëfficiënten omdat deze voor de latere simulatie belangrijk zijn.
8. Zie ook de Jong (1992) over verschillen in schooladviezen tussen regio's.
9. In Webbink (1993) wordt in dit verband gewezen op een afname van de zelfselectie op grond van eindexamencijfers bij de keuze voor Hoger Onderwijs.

LITERATUUR

- Bishop, J.H. (1989). Is the test score decline responsible for the productivity growth decline? In: *American Economic Review*, 79, 178-197.
- Cramer, J.S. (1991). *The LOGIT-model: An introduction for economists*. Edward Arnold, Londen.
- Gomulka, J. & Stern, N. The employment of married women in the United Kingdom 1970-1983, in: *Economica*, 57, 171-99.
- Jong, U. de, Oosterbeek, H., Roeleveld, J. & Webbink, H.D. (1992). *Voornemens van eindexamenkandidaten 1991*, Ministerie van Onderwijs en Wetenschappen, Serie Verder Studeren deel 1, een panelstudie onder scholieren en studenten, Den Haag.
- Jong, U. de (1992). *De loopbaan doorlopen*, proefschrift, SCO/Kohnstamm Instituut, Amsterdam.
- Kodde, D.A. (1984). *Wie kiest er voor Hoger Onderwijs*. Ministerie van Onderwijs en Wetenschappen, Serie beleidsgerichte studies Hoger Onderwijs en Wetenschappelijk onderzoek, Den Haag.
- Kodde, D.A. (1985). *Microeconomic analysis of demand for education*. Ph.D.-thesis, Erasmus Universiteit Rotterdam.
- Kodde, D.A. & Ritzen, J.M.M. (1986). *Vraag naar Hoger Onderwijs. Eindrapport*. Ministerie van Onderwijs en Wetenschappen, Serie beleidsgerichte studies Hoger Onderwijs en Wetenschappelijk onderzoek, nr. 6, Den Haag.
- Posthumus, K. (1942). Middelbaar onderwijs en schifting. In: *De Gids*, 104, no. 2, p. 24-42.
- Webbink, H.D., Jong, U. de, Oosterbeek, H. & Roeleveld, J. (1993). *Studiekeuze van scholieren en studenten in 1991*. Ministerie van Onderwijs en Wetenschappen, Serie Verder Studeren deel 2, een panelstudie onder scholieren en studenten, Den Haag.

Manuscript ontvangen 26-1-1994

Definitieve versie ontvangen 18-4-1994

APPENDIX

Gemiddelden en standaardafwijkingen van variabelen in 1982 en in 1991.

	1982		1991	
	gemid.	st. dev.	gemid.	st. dev.
geslacht (man=1, vrouw=1)	0,50	0,50	0,38	0,49
schooltype t.o.v. Vwo-B				
Havo (Havo=1, anders=0)	0,47	0,50	0,49	0,50
Vwo-A (Vwo-A=1, anders=0)	0,26	0,44	0,27	0,45
gedoubleerd (ja=1, nee=0)	0,08	0,27	0,35	0,47
aantal exacte vakken	1,97	1,59	2,03	1,41
aantal talen	1,85	0,94	1,90	0,80
eerdere diploma's (=1 bij Lbo/Mavo of Havo diploma)	0,24	0,43	0,18	0,39
vroege keuze (=1 als leerling al in vroeg stadium wist wat gaan doen na het examen)	0,68	0,47	0,74	0,44
studieplannen t.o.v. niet studeren				
voltijd studeren (=1 leerling wil voltijd studeren)	0,77	0,42	0,92	0,27
deeltijd studeren (=1 leerling wil deeltijd studeren)	0,05	0,21	0,03	0,18
weet niet (=1 leerling weet het nog niet)	0,06	0,25	0,01	0,11
regio t.o.v. zuid				
noorden	0,12	0,32	0,16	0,36
oosten	0,20	0,40	0,24	0,43
westen	0,44	0,50	0,39	0,49
motieven (hebben waarde 1 als leerling het belangrijk vindt bij de studiekeuze, anders 0)				
zelfontplooiing	0,68	0,47	0,73	0,45
kans op baan	0,73	0,44	0,94	0,24
nog niet willen werken	0,28	0,45	0,47	0,50
interesse in onderwerp van studie	0,81	0,39	0,92	0,28
bepaald beroep willen uitoefenen	0,75	0,43	0,85	0,35
aantal geslaagden	9737		742	
aantal niet geslaagden	1592		91	

Schoolloopbanen van Molukse leerlingen in het voortgezet onderwijs¹

A. Veen en M. Robijns*

ABSTRACT

In this article we present data on the schoolcareers of Moluccan students in the first phase of secondary education, compared with Dutch peers. In order to exclude differences in schoolcareer due to socio-economic and gender factors a matching procedure has been applied. Students were matched on age, gender, fathers occupation and educational advice. Thus data on Moluccan and comparable Dutch students were collected regarding their educational position during successive school years ('87-91), schooltype, drop-out rate and drop-out reason.

In order to bring a better understanding of the school careers we present a school-careermodel relating individual student characteristics to results in secondary education.

Results show that for these students of this first ethnic group in our country with a third generation within secondary education, significant differences in schoolcareers, although not very large, still do occur. Differences can be shown mainly where grade repeating is concerned. Furthermore differences occur in learning perception and attitude towards school.

INLEIDING

De positie van allochtone leerlingen in het onderwijs is, zo blijkt telkens opnieuw, veel minder gunstig dan die van autochtone leerlingen. Etniciteit blijkt een belangrijk onderscheidend criterium als we kijken naar het basisonderwijs, waar veel kinderen uit etnische minderheidsgroepen (met name Surinaamse en Marokkaanse kinderen) deelnemen aan het speciaal onderwijs (Teunissen en Golhof, 1987; Ledoux, 1988; Autar, 1990). Verder blijken Turkse en Marokkaanse leerlingen in het basisonderwijs gemiddeld ouder dan de autochtone leerlingen. Op twaalfjarige leeftijd volgt iets minder dan 2/3 van de Turkse en Marokkaanse leerlingen nog basisonderwijs, terwijl de meeste autochtone leerlingen op die leeftijd al zijn doorgestroomd naar het voortgezet onderwijs. Deze vertraging geldt voor zowel meisjes als jongens (Roelandt, Smeets en Veenman, 1993).

Ook in het voortgezet onderwijs doen zich verschillen tussen etnische groepen voor. Dit blijkt in de eerste plaats uit de deelname aan verschillende onderwijstypen. Op 15-jarige leeftijd zit 60% van de Marokkaanse en Turkse leerlingen op het lbo. Van de autochtone leerlingen is dat rond de 30% (Roelandt, Roijen en Veenman, 1992). In het v.o. vormt uitval een belangrijk knelpunt. Van de Turkse en Marokkaanse jongeren die een lbo- of mavo-opleiding hebben gevolgd, verlaat één op de twee die opleiding zonder diploma (Veenman, 1990).

Onder de etnische minderheidsgroepen in ons land nemen personen van Molukse herkomst een bijzondere positie in. Zij zijn al langer in Nederland en vormen de eerste etnische groep waarvan de derde generatie op dit moment in betekenisvolle aantallen in het voortgezet onderwijs te vinden is. Ook op andere punten hebben Molukkers, in vergelijking met overige etnische minderheidsgroepen een bijzondere positie in de Nederlandse samenleving. In tegenstelling tot andere groepen is de migratiereden niet gelegen in de wens tot verandering in economische of arbeidsmarktpositie, maar ligt de aanleiding voor de komst van Molukkers naar Nederland in het koloniale verleden en meer direct in de onafhankelijkheid van Indonesië van Nederland, in

* Adres: SCO-Kohnstamm Instituut/Faculteit der Pedagogische en Onderwijskundige Wetenschappen, Grote Bickersstraat 72, 1013 KS Amsterdam, 020 - 5550 328.

1949 en het hiermee verband houdende noodgedwongen vertrek van leden van de Molukse groep uit Indonesië. De speciale positie komt verder tot uiting in het -zij het afnemende- aandeel van Molukkers dat woont in 'Molukse woonwijken' verspreid over het land, meestal niet gelegen in stedelijke maar in plattelandsgebieden.

Het is moeilijk om in cijfermatige zin een beeld te krijgen van de positie van Molukkers -in het onderwijs, op de arbeidsmarkt-, zeker wanneer het de vergelijking met andere etnische minderheidsgroepen betreft. Molukkers worden, vanwege de relatief kleine omvang van de groep in grootschalige surveys doorgaans niet afzonderlijk onderscheiden. Bovendien zijn Molukkers en Molukse leerlingen niet (meer) objectief te identificeren op basis van de in registratiesystemen gebruikelijke criteria geboorteland (van ouders) en nationaliteit.

Om zicht te krijgen op de maatschappelijk positie van Molukkers wordt meestal gebruik gemaakt van de gegevens die Veenman e.a. in het afgelopen decennium in twee grootschalige survey-onderzoeken (1983 en 1990) onder de Molukse bevolkingsgroep verzamelden (Veenman, 1984 a,b,c; Veenman, 1990; Veenman en Martens, 1991). Centraal in deze onderzoeken stond de positie van Molukkers op de arbeidsmarkt, maar ook werden gegevens verzameld die op het onderwijs betrekking hebben. Gezocht werd naar factoren die de positie van Molukkers op de arbeidsmarkt zouden kunnen verklaren. De onderwijsloopbaan speelt hierbij een belangrijke rol. De onderzoeken bevatten dan ook gegevens over de onderwijspositie van Molukkers (zoals verdeling over schooltypen, behaalde onderwijsniveaus), alsmede verklarende factoren over de onderwijspositie.

Voor wat betreft het onderwijsniveau van de Molukse groep blijkt, dat dit relatief laag is. Uit de eerste survey komt bijvoorbeeld naar voren, dat 84% van de 15-30-jarigen uitsluitend lager onderwijs, lbo of mavo heeft tegenover 55% van de Nederlandse bevolking (Veenman, 1985).

Tabel 1. Gerealiseerd opleidingsniveau van personen van 15-64 jaar naar geslacht en etnische groepering, 1991 (Molukkers 1990).

	Ten hoogste basis- onderwijs	LBO	MAVO	MBO	HAVO/ VWO	HBO	WO	School- gaand	Totaal
	abs.=100%								
<i>Mannen</i>									
Molukkers	24	37	10	8	3	4	0	14	1 994
Turken	62	7	11	4	3	1	1	13	1 198
Marokkanen	69	4	5	2	2	0	1	17	1 116
Surinamers	29	16	13	7	5	4	2	24	651
Antillianen	26	19	8	10	5	5	2	25	410
Autochtonen	19	18	8	15	8	10	9	13	735
<i>Vrouwen</i>									
Molukkers	28	32	11	11	4	2	0	12	1 903
Turken	73	4	8	3	2	0	0	10	911
Marokkanen	78	3	4	1	0	0	0	14	782
Surinamers	34	12	18	10	3	4	1	18	772
Antillianen	31	13	16	10	4	3	2	21	498
Autochtonen	20	20	12	13	9	10	7	11	742

Bron: Roelandt, Smeets, Veenman, 1993 (SPVA '91, ISEO-EUR; Arbeidsmarktsurvey Molukkers 1990, ISEO-EUR).

Tabel 2. Personen die in Nederland onderwijs hebben gevolgd doch thans geen onderwijs volgen naar situatie t.a.v. opleidingen na het basisonderwijs en naar etnische groepering, 1991 (Molukkers 1990).

	Alleen BAO gevolgd	Na BAO uitsluitend afgebroken opleiding(en)	Na BAO (ook of uitsluitend) voltooide opleidingen	Totaal
	%			abs=100%
<i>Mannen</i>				
Molukkers	9	19	72	1 716
Turken	9	36	55	228
Marokkanen	16	37	48	168
Surinamers	2	18	80	266
Antillianen	4	12	84	135
Autochtonen	9	12	79	634
<i>Vrouwen</i>				
Molukkers	16	15	69	1 670
Turken	26	30	44	138
Marokkanen	29	30	42	108
Surinamers	2	18	80	287
Antillianen	4	22	74	128
Autochtonen	7	13	80	656

Bron: Roelandt, Smeets, Veenman, 1993 (SPVA '91, ISEO-EUR; Arbeidsmarktsurvey Molukkers 1990, ISEO-EUR).

Een oververtegenwoordiging van Molukkers dus in de, uit arbeidsmarktperspectief risicovolle schooltypen.

Dit beeld wordt in de meer recente survey (Veenman en Martens, 1991) weer bevestigd, al blijkt er wel enige verbetering te zijn opgetreden. Toch is er, zoals uit tabel 1 blijkt, in vergelijking met de autochtone groep, bij de Molukse groep nog steeds sprake van een sterke oververtegenwoordiging van personen met een lbo-diploma. Uit tabel 2 blijkt voorts, dat Molukkers, evenals Turken, Marokkanen, Surinamers en Antillianen/Arubanen zijn oververtegenwoordigd bij degenen met uitsluitend afgebroken opleidingen (de zogenaamde 'drop outs').

Over de oorzaken van de relatief slechte onderwijsresultaten, die leiden tot de oververtegenwoordiging van de Molukse groep in de lagere vormen van het v.o. en de hoge uitvalpercentages is weinig bekend. Aanwijzingen voor specifieke problemen in het voortgezet onderwijs vormden de uitkomsten uit een onderzoek naar Molukse leerlingen in het basisonderwijs van de Inspectie (Ministerie van Onderwijs, 1988). Een opvallende bevinding in dit onderzoek was, dat Molukse leerlingen lager scoren op de Citotoets dan Nederlandse leerlingen, maar toch een met Nederlandse leerlingen vergelijkbaar doorstroompatroon naar het voortgezet onderwijs vertonen. Het vermoeden bestond dat deze vorm van 'overadvisering' problemen zou veroorzaken in het v.o. Deze zouden vooral bestaan uit uitval uit het tweede en derde leerjaar van het v.o. Empirische gegevens om dit vermoeden te onderbouwen waren echter niet in het betreffende onderzoek verzameld. Daarom werd een onderzoek verlangd, waarin de schoolloopbanen van Molukse leerlingen in de eerste fase van het voortgezet onderwijs zouden worden vergeleken met die van Nederlandse leerlingen. Op deze wijze zouden mogelijkerwijs specifieke problemen van Molukse leerlingen in het v.o. kunnen worden opgespoord. Dit onderzoek hebben wij uitgevoerd in de jaren 1990 en 1991.¹

METHODE

Onderzoeksvragen

In het onderzoek stonden vier onderzoeksvragen centraal, waarvan er in dit artikel drie worden behandeld, namelijk:

1. Hoe verloopt de schoolloopbaan van Molukse leerlingen in de eerste fase van het voortgezet onderwijs en welke verklaringen zijn er voor dit verloop te geven vanuit de sekse van de leerling, het sociaal-economisch milieu, de resultaten op de Citotoets, het schooladvies aan het einde van de basisschool?
2. Treden er verschillen op in schoolloopbanen van Molukse leerlingen en vergelijkbare Nederlandse leerlingen?
3. Welke kennis, houdingen, verwachtingen en aspiraties hebben Molukse en vergelijkbare Nederlandse leerlingen en hun ouders voor zichzelf/hun kind ten aanzien van onderwijs en arbeid en hoe hangt dit samen met de schoolloopbanen van de leerlingen?

De vierde onderzoeksvraag had betrekking op de maatregelen die scholen treffen om een ongunstig verloop van de schoolloopbanen van Molukse leerlingen te voorkomen.

Onderzoeksofzet en uitvoering

Om de in dit artikel te behandelen onderzoeksvragen te beantwoorden zijn de volgende onderzoeksactiviteiten verricht.

Begonnen is met een interviewronde onder sleutelpersonen om nader inzicht te krijgen in de specifieke problematiek van Molukse leerlingen in het onderwijs. Vervolgens zijn, specifiek ter beantwoording van de vraag naar het loopbaanverloop in de eerste fase van het v.o., gegevens verzameld voor de reconstructie van de schoolloopbanen van Molukse en vergelijkbare Nederlandse leerlingen in het voortgezet onderwijs. Deze gegevensverzameling heeft plaatsgevonden op scholen voor voortgezet onderwijs met een voldoende aantal Molukse leerlingen². We moesten in dit onderzoek zelf gegevens verzamelen en konden geen gebruik maken van -in principe beschikbare- bestaande gegevensbestanden, zoals in het bestand *Schoolloopbanen Voortgezet Onderwijs -SLVO-*, het bestand "schoolloopbanen etnische groepen" (SVO-project 6113) of het gegevensbestand dat momenteel door het CBS wordt opgebouwd ten behoeve van het OVB en de evaluatie van het voortgezet basisonderwijs (SVO-project 8075). In het eerste zijn Molukse leerlingen niet te identificeren, het tweede bevat slechts gegevens over één leerjaar en het derde bevatte in de periode van gegevensverzameling voor ons onderzoek nog te weinig loopbaangegevens voor de beantwoording van de onderzoeksvragen.

Om zicht te krijgen op factoren die van belang zouden kunnen zijn voor het verloop van de schoolloopbanen van Molukse en vergelijkbare Nederlandse leerlingen zijn de leerlingen (en hun ouders) op een groot aantal topics bevraagd met behulp van schriftelijke vragenlijsten. Hierbij is rekening gehouden met verschillende typen leerlingen: leerlingen die zich ten tijde van het onderzoek nog wel en leerlingen die zich niet meer op de school bevonden waarop zij in het cursusjaar 1987-1988 instroomden. Deze tweede groep bestaat uit leerlingen die naar een andere school vertrokken zijn en leerlingen die voorgoed de school de rug toe hebben gekeerd. Voor al deze groepen (nog aanwezig op de aanvangsschool, vertrokken naar een andere school en voorgoed vertrokken) zijn verschillende versies van de vragenlijst opgesteld. Daarbinnen is nog weer onderscheid gemaakt tussen de Molukse en de Nederlandse groep, dat wil zeggen de Molukse leerlingen kregen enkele extra vragen voorgelegd.

Schoolloopbanen en controlecohort

In het onderzoek is in de eerste plaats het verloop van de schoolloopbanen van Molukse leerlingen onderling vergeleken. Onder het begrip schoolloopbaan kan een reeks van aspecten of gebeurtenissen worden verstaan die zich in het (voortgezet) onderwijs kunnen voordoen: overgaan, doubleren, al dan niet behalen van het examen, afstroom naar lagere schooltypen, dropout, de keuze van vakkenpakketten en vakrichtingen en eindexamenniveaus. Voor dit onderzoek

onderscheidde we de volgende afhankelijke variabelen: schooladvies van de basisschool, onderwijspositie in opeenvolgende cursusjaren en schooltype. Als schooltypen onderscheiden we: ibo, lbo, mavo, havo, vwo. Voor de onderwijsposities gebruiken we een kwantificering, waarin zowel het schooltype als het leerjaar wordt meegewogen³.

Voor de reconstructie van schoolloopbanen is het noodzakelijk dat voldoende posities van de leerlingen in het onderwijs bekend zijn. Daarom is ervoor gekozen om gegevens te verzamelen van leerlingen die in 1987-1988 in het v.o. zijn ingestroomd. Van dit cohort waren in 1990 namelijk maximaal vier onderwijsposities oftewel drie overgangen tijdens de schoolperiode bekend. Omdat de resultaten van opeenvolgende cohorten nogal eens een zekere instabiliteit vertonen (Roeleveld en De Jong, 1989) is ervoor gekozen de leerlingen die instroomden in 1988-1989 op te nemen als controlecohort. Dit cohort levert weliswaar een jaar minder gegevens op, maar kan toch zicht geven op de vraag of de conclusies die over het eerste cohort getrokken worden ook geldig geacht mogen worden voor andere cohorten. Gezien de functie van dit controlecohort zijn alleen de schoolloopbaangegevens verzameld van Molukse leerlingen.

Matching

Om de vraag naar eventuele verschillen in schoolloopbanen tussen Molukse en Nederlandse leerlingen te kunnen beantwoorden is bij de opbouw van het onderzoeksbestand gebruik gemaakt van de methode van paarsgewijze vergelijking, oftewel matching. In deze procedure worden leerlingparen gezocht die gelijkgeschakeld zijn op zoveel mogelijk factoren waarvan bekend is of vermoed wordt dat ze het verloop van de schoolloopbanen beïnvloeden. De matching is bedoeld om de invloed van 'storende' factoren zoveel mogelijk te neutraliseren en alleen de invloed te meten van de variabele 'etnische achtergrond' op het verloop van de schoolloopbaan. In ons onderzoek is gekozen de leerlingen te matchen op de in schoolloopbaanonderzoek gebruikelijke variabelen: sekse, leeftijd, sociaal-economische status⁴, advies voor het voortgezet onderwijs en score op de Citotoets.

Matching is een zeer arbeidsintensieve procedure die een grote zorgvuldigheid vereist. Hierbij is als volgt te werk gegaan. Op basis van de leerlinglijsten van de betreffende school hebben we in de leerlingadministraties van de deelnemende v.o.-scholen Molukse leerlingen geselecteerd die instroomden in de cursusjaren 1987-1988 (basiscohort) en 1988-1989 (controlecohort). Als Molukse leerlingen zijn die leerlingen beschouwd die ten minste één ouder hebben van Molukse herkomst. Van deze leerlingen werden vervolgens de leerlingkaarten opgezocht; de daarop geregistreerde leerlinggegevens die nodig waren voor het onderzoek werden vervolgens overgenomen op speciaal hiervoor ontworpen leerlingformulieren. De verzameling van gegevens van leerlingen uit het controlecohort was hiermee afgerond. Bij de Molukse leerlingen uit het basiscohort werden vervolgens Nederlandse leerlingen met vergelijkbare matchkenmerken gezocht. Om te voorkomen dat, als gevolg van non-respons in de groep van gematchte Nederlanders in het vragenlijst-survey onder de leerlingen (en ouders) de onderzoekspopulatie teveel zou reduceren, is geprobeerd om bij elke Molukse leerling uit het basiscohort drie Nederlandse leerlingen te matchen. Dit gebeurde door per geselecteerde Molukse leerling de leerlinglijst van het instroomjaar 1987-1988 door te lopen en de eerste, tweede en zo mogelijk derde, Nederlandse leerling te selecteren met dezelfde sekse en leeftijd en een overeenkomend v.o.-advies en sociaal-economische achtergrond. De gegevens van deze leerlingen werden overgenomen op het leerlingformulier. De Citoscore is, zo bleek, een gegeven dat maar zelden geregistreerd is in de leerlingadministraties van v.o.-scholen. Deze heeft dan ook niet als matchvariabele kunnen functioneren.

Onafhankelijke variabelen

Aan onderwijsachterstanden ligt een zeer complexe problematiek ten grondslag die zich uiteraard niet binnen het kader van één onderzoek uitputtend laat behandelen. De diverse literatuur waarin mogelijke achtergronden van de problematiek van Molukse leerlingen besproken wordt (Molony en Pechler, 1982; Pille, 1989; Veenman, 1984, 1985, 1990) alsmede de interviews met

deskundigen op het gebied van onderwijs en Molukse leerlingen die in het kader van het onderzoek werden geraadpleegd wijzen op cultuur-, milieu en sekse-specifieke oorzaken. Geprobeerd is om uit de veelheid aan informatie determinanten te distilleren die door middel van schriftelijke vragenlijsten bevestigd konden worden. Van belang voor dit artikel zijn in de eerste plaats de variabelen die bevestigd zijn door middel van attitudeschalen en die in het schoolloopbaanmodel zijn opgenomen: de variabelen 'aspiraties met betrekking tot onderwijs en arbeid', 'houding ten aanzien van onderwijs en arbeid', 'academisch zelfbeeld' en 'sociaal zelfbeeld'.

Aspiraties van leerlingen met betrekking tot onderwijs en arbeid zijn bevestigd met behulp van de subschaal 'aspiraties', afgeleid uit een onderzoek van De Vries e.a. (1990) naar attitude ten aanzien van en beleving van school. Verder is de leerlingen in dit verband gevraagd naar het hoogste onderwijsniveau waarvan ze een diploma denken te halen, de plannen direct na de huidige opleiding (werken of leren) en verwachtingen van arbeid (al of niet hebben van werk, beroepsbeeld). Om een beeld te krijgen van de houding ten aanzien van onderwijs en arbeid zijn eveneens attitudeschalen afgeleid van subschalen uit het onderzoek van De Vries e.a., (1990). Het gaat om de subschalen 'schoolhouding' en 'schoolbeleving'. 'Kennis over onderwijs en arbeid' is vooral toegespitst op kennis over vervolgoopleidingen na de huidige opleiding en de trajecten die opleiden tot bepaalde beroepen. Zoals de labels 'academisch zelfbeeld' en 'sociaal zelfbeeld' al doen vermoeden heeft het eerste te maken met het beeld dat de leerlingen hebben van hun eigen prestaties en het tweede met het beeld dat zij hebben van hun sociaal functioneren. Opnieuw is gebruik gemaakt van de attitudeschalen van De Vries e.a., namelijk de subschalen 'zelfbeeld leren/school' en 'beleving klas'.

Voorts is in het onderzoek de variabele 'culturele oriëntatie' opgenomen. De achtergrond hiervan is de aanwijzing uit eerder onderzoek (De Jong en Verkuyten, 1989) en uit de voorbereidende interviews met sleutelpersonen dat één van de factoren die van invloed zou zijn op schoolprestaties van allochtone leerlingen in het voortgezet onderwijs de cultuur is waarop zij zich primair oriënteren. Enerzijds wordt verondersteld dat de schoolloopbanen van allochtone leerlingen die zich het waardepatroon eigen maken van de in het onderwijs dominante 'blanke middenklasse cultuur' gunstiger verlopen. Anderzijds wordt verondersteld dat een sterke oriëntatie op de thuiscultuur juist zou leiden tot betere schoolloopbanen van allochtone leerlingen: dit zou leiden tot een positiever zelfbeeld en zo een gunstige uitwerking hebben op de schoolloopbaan. Over de relatie tussen culturele oriëntatie en schoolloopbanen en het belang van zelfbeeld daarbij is nog weinig duidelijkheid. Wat de Molukse bevolkingsgroep betreft blijkt uit onderzoek van Veenman en Martens (1991) dat de culturele oriëntatie, of anders geformuleerd, de mate van integratie in en deelname aan de Nederlandse samenleving een belangrijke factor is in de verbetering van de maatschappelijke positie. Uit een vergelijking van de uitkomsten uit de eerder besproken surveys blijkt bijvoorbeeld, dat tussen 1983 en 1990 sprake is van een substantiële afname van het werkloosheidspercentage (genoemd wordt een afname van 60%). Deze wordt niet alleen toegeschreven aan een verhoging van het onderwijspeil onder de hele groep of een algehele verbeterde economische situatie die de kansen van Molukkers heeft vergroot. De onderzoekers vonden namelijk, dat de afname van de werkloosheid met name gerelateerd is aan de mate waarin men 'geïntegreerd' is in de Nederlandse samenleving. Factoren die hierbij vooral van belang waren zijn het al dan niet gemengd gehuwd of samenwonend zijn en de mate en intensiteit van contacten met Nederlanders. Het werkloosheidscijfer bleek bij gemengde partnerschappen 14% en bij Molukse 30%, terwijl bij de mate van contacten met autochtonen werd gevonden dat bij Molukkers die vaker en intensiever met Nederlanders omgaan de werkloosheid zo'n 10% lager ligt (Veenman en Martens, 1991). De auteurs doen geen uitspraken over de richting van het verband en we kunnen uit de gegevens dus niet afleiden of integratie leidt tot een kleinere kans op werkloosheid, of dat werk integratie bevordert. De Jong en Verkuyten (1989) deden onderzoek naar de relatie tussen culturele oriëntatie en zelfbeeld bij Turkse jongeren. Zij vonden een verband tussen oriëntatie op de Nederlandse cultuur en schoolprestaties: Turkse leerlingen die zich sterker oriënteren op (of: meer vertrouwd zijn met) de Nederlandse cultuur, presteren beter in het onderwijs.

'Culturele oriëntatie' is een moeilijk te definiëren begrip omdat het staat voor een complex

van uitingen, gedragingen, normen, waarden en gewoonten. In dit onderzoek hebben we slechts twee uitingsvormen nagevraagd, namelijk de taal die thuis (met ouders, broers/zussen) en met vrienden/vriendinnen wordt gesproken en de omgang met anderen (Molukkers, Nederlanders) in de vrije tijd en op school/het werk. Beide onderdelen zijn aan de leerlingen voorgelegd in de vorm van uitspraken met items als: 'met mijn moeder spreek ik Maleis', 'met mijn broer(s) en zus(sen) spreek ik Nederlands', 'op school ga ik met Molukse jongens om' en 'in mijn vrije tijd ga ik met Nederlandse jongens om'. Leerlingen konden op een vijfpuntsschaal aangeven in hoeverre de uitspraak op de eigen situatie van toepassing was.

Bestandsopbouw en respons

Voor het onderzoek werden 51 scholen voor voortgezet onderwijs benaderd. Hiervan bleken er 34 bereid medewerking aan het onderzoek te verlenen. Tot deze groep behoren diverse soorten scholen voor voortgezet onderwijs, variërend van categorale lbo-scholen en mavo's tot brede scholengemeenschappen. Op deze scholen zijn de schoolloopbaangegevens verzameld van 123 Molukse en 246 Nederlandse leerlingen die in het cursusjaar 1987-1988 in het voortgezet onderwijs begonnen. Het controlecohort (Molukse leerlingen die in 1988 begonnen op deze scholen voor voortgezet onderwijs) betrof in totaal 99 leerlingen.

Van de Molukse leerlingen uit het cohort 1987-1988 retourneerden 80 de vragenlijst (65%), van de Nederlandse leerlingen respondeerden er 180 (73%). In totaal konden uit deze respons 69 matchparen worden samengesteld.

In het volgende worden de onderzoeksresultaten besproken. Eerst presenteren we enkele schoolloopbaangegevens van de Molukse groep en enkele verschillen binnen deze groep op de variabelen 'seks' en 'culturele oriëntatie'. Vervolgens gaan we in op de onderzoeksresultaten van de gematchte paren.

RESULTATEN

Molukse leerlingen onderling vergeleken

De 123 Molukse leerlingen uit het cohort 1987-1988⁵ zijn vergeleken op het schooltype waarop de leerlingen terecht zijn gekomen. Verder zijn vergelijkingen gemaakt op enkele achtergrondkenmerken, bijvoorbeeld verschillen in culturele oriëntatie en sekseverschillen.

Schooltype

Omdat de Molukse leerlingen verdeeld zijn over nogal wat verschillende typen scholen (van categorale lbo- en mavo-scholen tot brede scholengemeenschappen (lbo t/m vwo of mavo-havo-vwo), maken we alleen een vergelijking tussen leerlingen op avo-scholen (met in elk geval een avo-afdeling) en leerlingen op lbo-scholen (met alleen maar lbo en eventueel ibo). Uit tabel 3 blijkt, dat op de lbo-scholen leerlingen wat vaker de school verlaten hebben dan op de avo-scholen: (23% tegenover 16%); dit verschil is echter te klein om significant te zijn. Voorts blijkt dat -in alle leerjaren- de Molukse leerlingen op de avo-scholen iets vaker blijven zitten. Als gevolg hiervan zitten de leerlingen die op de avo-scholen begonnen na vier jaar wat vaker in een derde klas. Ook hier zijn de verschillen te klein om significant te zijn.

Uiteraard zijn er grote verschillen in het onderwijstype waarin de leerlingen zich in het 4e leerjaar bevinden. Van de leerlingen die op een lbo-school begonnen is niemand 'opgestroomd' naar een avo-school. Van de leerlingen die begonnen op een avo-school zit nu 11% op een lbo- of lbo-afdeling (of -school).

Uit de analyses blijkt verder, dat Molukse leerlingen over het algemeen na de basisschool terecht komen op een v.o.-school die correspondeert met het v.o.-advies. Slechts 4 leerlingen blijken op een schooltype te komen dat niet 'past' bij hun advies: 3 leerlingen met lbo-advies (eventueel ibo-advies, dit onderscheid wordt in het eerste leerjaar niet gemaakt) en 1 leerling met havo/vwo-advies beginnen op een mavo-school.

Tabel 3. Gegevens loopbanen Molukse leerlingen uit cohort 1987 op twee schooltypen.

		avo (N=76)	lbo (N=47)	P			avo (N=76)	lbo (N=47)	P
Weg van start school		16%	23%	ns	Onderwijstype 4e jaar				
Zittenblijven	1e klas	11%	6%	ns	ibo	3%	6%		
	2e klas	17%	11%	ns	lbo	8%	74%		
	3e klas	13%	11%	ns	mavo	57%	-		
Klasnr 4e jaar	2e klas	1%	-		havo	16%	-		
	3e klas	36%	26%		vwo	7%	-		
	4e klas	53%	55%		vertrokken	11%	19%	***	
	onbekend	11%	19%	ns					

Significante verschillen worden aangeduid met asterisken.

*** : $p < .01$

** : $p < .05$

* : $p < .10$

ns : $p > .10$ (ns = niet significant)

Sekse

Uit de analyses naar de relatie tussen het geslacht van de leerlingen en de schoolloopbaangeburtentissen blijkt het volgende. De onderwijsresultaten van de Molukse meisjes zijn over het geheel genomen iets positiever. Er is een tendens dat meisjes iets minder vaak op het lbo zitten en iets meer op het avo dan jongens. Verder blijven meisjes iets minder vaak zitten en zijn minder vaak vertrokken uit de school waarop zij in '87 begonnen. Dit resulteert al met al in een iets hogere onderwijspositie (Op) bij de meisjes. De resultaten staan in Tabel 4.

Culturele oriëntatie

'Culturele oriëntatie' is geoperationaliseerd als de taal die thuis (met ouders, broers/zussen) en met vrienden/vriendinnen wordt gesproken en de omgang met anderen (Molukkers, Nederlanders) in de vrije tijd en op school/het werk. Enkele uitkomsten uit de vragen naar het spreken van Nederlands en Maleis uit de Molukse groep leerlingen in ons onderzoek zijn, dat meer dan de helft van de leerlingen aangeeft regelmatig Maleis te spreken in de thuissituatie (met vader, moeder, broers en zussen). Een derde spreekt met de ouders nooit Maleis. Jongens zeggen wat vaker dan meisjes Maleis te spreken met vrienden en vriendinnen. Iets meer dan de helft van de leerlingen geeft aan op school en in de vrije tijd altijd of meestal met Molukse jongens of meisjes om te gaan. Slechts rond de 20% doet dit nooit.

Nagegaan is vervolgens in hoeverre de scores op de beide schalen ('taal' en 'omgang') correleerden en er dus een samengevoegde schaal 'culturele oriëntatie' geconstrueerd kon wor-

Tabel 4. Loopbaanmerken van Molukse meisjes (m) en jongens (j); (kolompercentages).

		m	j	P	hyp			m	j	P	hyp
type 1990	lbo	39%	48%			Vertrek niet	83%	79%			
	mavo	43%	39%			wel	17%	21%	ns	+	
gemiddelde	havo + vwo	19%	14%	ns	+	Zittenbl. niet	68%	62%			
	Op90	52.7	49.6	ns	+	wel	32%	38%	ns	+	

Tabel 5. Gemiddelde scores op oriëntatieschaal Molcult naar verschillende loopbaan kenmerken. Correlaties scores op oriëntatieschaal Molcult met loopbaanvariabelen.

		gem	P	hyp
Vertrek	niet	2.3		
	wel	2.6	ns	+
Zittenblijven	niet	2.2		
	wel	2.6	ns	+
Schooltype 1990	lbo	2.1		
	mavo	2.6		
	havo + vwo	2.3	ns	

De antwoordscores betreffen gemiddelden op een schaal die liep van 1 t/m 5, waarbij 1 een geringe oriëntatie op de eigen cultuur representeert en 5 een hoge.

den. De correlaties bleken inderdaad zodanig, dat schaalconstructie mogelijk was. De schaal bestaat uit 9 items met een alpha van .88.

De vraag die we geprobeerd hebben te beantwoorden is, of er een samenhang is tussen de mate waarin Molukse leerlingen georiënteerd zijn op de Molukse cultuur en eventuele onderwijsproblemen.

We zien in tabel 5 dat voor zowel vertrek als zittenblijven de veronderstelling in de verwachte richting gaat, dat wil zeggen: Molukse leerlingen die meer georiënteerd zijn op de eigen cultuur hebben vaker de v.o.-school waar ze aanvankelijk heengegaan waren verlaten en zijn vaker blijven zitten. Verder is te zien dat van de leerlingen op de verschillende schooltypen in 1990 de mavo-leerlingen gemiddeld het meest op de eigen cultuur georiënteerd zijn. De verschillen zijn niet significant.

We zien dus, samenvattend, dat Molukse leerlingen op lbo-scholen iets vaker de school hebben verlaten dan leerlingen op de avo-scholen. De leerlingen op de lbo-scholen blijven iets minder vaak zitten. 'Opstroom' naar een hoger schooltype (van lbo naar avo) heeft niet plaatsgevonden. 'Afstroom', van avo naar lbo wel, zij het slechts bij 11% van de leerlingen. Molukse leerlingen die -volgens de in het onderzoek gehanteerde definiëring- meer georiënteerd zijn op de eigen cultuur hebben vaker de aanvangsschool verlaten en zijn vaker blijven zitten. Uit de analyses naar de relatie tussen het geslacht van de leerlingen en de schoolloopbaangebeurtenissen komt naar voren dat de onderwijsresultaten van de meisjes over het geheel genomen iets positiever zijn. Er is een tendens dat meisjes iets minder vaak op het lbo zitten en iets meer op het avo dan jongens, iets minder vaak blijven zitten en minder vaak de school waarop zij in '87 begonnen hebben verlaten. Dit resulteert al met al in een iets hogere onderwijspositie bij de meisjes. Geen van de genoemde verschillen is significant.

Molukse en Nederlandse leerlingen vergeleken

In deze paragraaf vergelijken we de onderzoeksgegevens van de 69 gematchte paren met elkaar⁶. We beginnen met een beeld van het verloop van de schoolloopbanen van Molukse en Nederlandse leerlingen. Daarna vergelijken we door middel van vier attitudeschalen het academisch en het sociaal zelfbeeld en de aspiraties en houding van ten aanzien van school en werk van beide groepen leerlingen.

De onderwijsresultaten van de 69 gematchte paren zijn vergeleken aan het begin van het schooljaar 1990-1991, vier jaar nadat zij ingestroomd zijn in het voortgezet onderwijs. De resultaten van deze vergelijking vormen geen ondersteuning voor de verwachting dat Molukse leerlingen systematisch lagere onderwijstypen bezoeken dan hun Nederlandse matchpartners⁷. Het overgrote deel van de matchparen bezoekt hetzelfde onderwijstype. Voor slechts vier match-

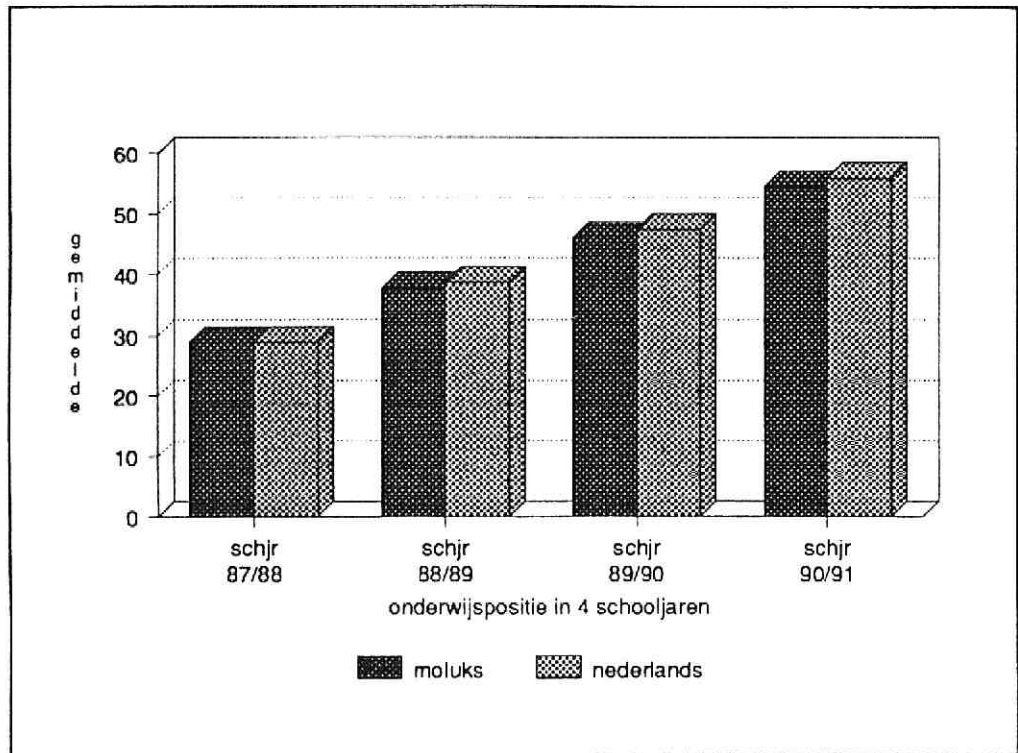


Fig. 1. Gemiddelde onderwijspositie in 4 schooljaren voor gematchte groepen Molukse en Nederlandse leerlingen.

paren geldt dat de Nederlandse leerling zich op een hoger schooltype bevindt, onder hen 2 paren waarvan de Molukse leerling op het lbo terecht is gekomen en terwijl de Nederlandse leerling zich op de havo respectievelijk vwo bevindt. Bij 5 matchparen geldt dat de Molukse leerling zich op een (aangrenzend) hoger schooltype bevindt.

Wel blijken Molukse leerlingen vaker te doubleren dan Nederlandse⁸. Vier jaar nadat zij ingestroomd zijn in het voortgezet onderwijs bevinden zij zich vaker in de derde klas dan hun Nederlandse matchpartners.

In de onderstaande figuur is de gemiddelde onderwijspositie in de vier schooljaren voor de gematchte groepen weergegeven. Indicatoren voor schoolsucces waren onderwijstype en klasnummer, gecombineerd tot de samengestelde maat Onderwijspositie.

De figuur toont aan dat Molukse leerlingen na de gelijke start in het schooljaar 87-88 (een direct gevolg van de manier van dataverzameling) al snel, namelijk in het tweede jaar gaan achterlopen bij de Nederlandse leerlingen.

Hoewel zittenblijven een (negatieve) voorspeller is voor de toekomstige loopbaan, kan gesteld worden dat afstroom naar een lager schooltype in het algemeen negatievere consequenties heeft, aangezien er dan meer toekomstmogelijkheden worden afgesneden. We concluderen dan ook dat de onderwijsresultaten van Molukse leerlingen minder gunstig zijn dan die van Nederlandse. De achterstand is echter aanzienlijk minder groot dan op voorhand verwacht.

Om inzicht te verkrijgen in de factoren die samenhangen met de minder succesvol verlopende schoolloopbanen van Molukse leerlingen zijn de reeds genoemde attitudeschalen in het schoolloopbaanmodel opgenomen. Het gaat om de schalen 'aspiraties ten aanzien van onderwijs', 'houding ten opzichte van en beleving van school', 'sociaal zelfbeeld' en 'academisch zelf-

beeld. Wanneer we de antwoordpatronen van Molukse en Nederlandse leerlingen op deze schalen vergelijken dan komt het beeld naar voren dat Molukse en Nederlandse leerlingen zich even prettig voelen op school en dat hun aspiraties even hoog zijn: beide groepen leerlingen willen graag 'iets bereiken in de maatschappij'. Het lijkt er echter wel op dat de Molukse groep minder duidelijk voor ogen staat hoe deze latere maatschappelijke positie er uitziet: een concreet beroepsbeeld is bij hen in mindere mate aanwezig dan bij de Nederlandse groep en hun toekomstplannen lijken minder omljnd. De minder duidelijk omljnde toekomstplannen van de Molukse leerlingen zouden te maken kunnen hebben met onzekerheid over de eigen mogelijkheden, hun 'eigen kunnen'. Hoe schatten de leerlingen bijvoorbeeld hun kansen in op het behalen van een diploma van de huidige school en op het verkrijgen van werk, later? Uit de antwoorden op de 7 items van de schaal 'academisch zelfbeeld' blijkt dat Molukse leerlingen in het algemeen een minder positief beeld over zichzelf en de eigen prestaties hebben dan de Nederlandse leerlingen ($p < .05$). Negatiever bijvoorbeeld schatten de Molukse leerlingen hun kans in op het halen van een diploma. Verder nemen zij, meer dan de Nederlandse leerlingen, bij zichzelf leerproblemen waar in die zin, dat zij het lestempo moeilijker kunnen bijhouden.

Op basis van de uit 10 items bestaande schaal 'houding ten opzichte van en beleving van school' kan geconcludeerd worden dat Molukse en Nederlandse leerlingen een verschillende houding en beleving ten aanzien van school hebben ($p < .01$). Het lijkt erop dat Molukse leerlingen nog sterker dan Nederlandse leerlingen de school zien als instrument dat toegang geeft tot de voor later gewenste, goede maatschappelijke positie. Beide groepen zijn het erover eens zijn dat zij later veel aan hun huidige school zullen hebben. Bij Molukse leerlingen speelt het plezier in naar school gaan echter een minder sterke rol. Het besef dat het nu eenmaal moet en primair nodig is om een diploma te behalen is sterker ontwikkeld in deze groep.

Schoolloopbaanmodel

We concludeerden hiervoor dat de onderwijsresultaten van Molukse leerlingen iets minder gunstig zijn dan die van Nederlandse leerlingen. De onderwijspositie die Molukse en Nederlandse leerlingen in vier jaar behaald hebben vertoont echter geen significante verschillen. Om meer licht te werpen op de vraag in hoeverre Molukse en Nederlandse leerlingen via soortgelijke of juist uiteenlopende processen de nageoeg gelijke onderwijspositie in het vierde leerjaar bereiken is een schoolloopbaanmodel geconstrueerd. Meer specifiek is nagegaan of er een identiek model van toepassing is op de gegevens van zowel de Molukse als de Nederlandse groep. Om de analyse uit te voeren is gebruik gemaakt van Lisrel-VII (Jöreskog & Sörbom, 1988). De bereikte onderwijspositie in het vierde leerjaar (Op4) is gemodelleerd afhankelijk van:

- *achtergrondvariabelen:*
 - Sex sekse (gecodeerd als 0: meisje en 1: jongen)
 - Ber beroep van de vader
- *resultaten in het basisonderwijs:*
 - Zit zittenblijven
 - Adv advies voor voortgezet onderwijs
- *startmoment in het voortgezet onderwijs:*
 - Op1 onderwijspositie in het eerste jaar
- *attitudeschalen:*
 - Asp aspiraties
 - Hou houding
 - Azb Academisch zelfbeeld
 - Szb Sociaal zelfbeeld

Het vinden van een identiek model passend op de data van beide groepen bleek onmogelijk. De verschillen tussen de variantie-covariantie matrices van de tien variabelen bleken hiervoor te sterk. De hypothese dat deze matrices identiek zijn voor de Molukse en de Nederlandse groep moest verworpen worden (chi-kwadraat = 146.54, bij 110 vrijheidsgraden; $p = .011$). Vervolgens is, met het vrijlaten van zo min mogelijk effecten tussen beide groepen, gezocht naar modellen die voor beide groepen wel bij de data pasten. Deze uiteindelijke modellen zijn weergegeven in figuur 2 de bijbehorende coëfficiënten worden weergegeven in tabel 6.

Het algemene schoolloopbaanmodel vertoont een vertrouwd beeld dat we ook terugvinden in ander onderzoek (zie bijv. Dronkers, 1989, voor een overzicht): Zittenblijven (Vtlo) in het basisonderwijs heeft een direct nadelig effect op het advies (Advies) en op de startpositie in het voortgezet onderwijs (Op1). Daarnaast heeft het advies voor het voortgezet onderwijs sterke rechtstreekse invloed op de startpositie en bovendien op de onderwijspositie in het vierde jaar (Op4). De startpositie in het voortgezet onderwijs heeft bovendien een positief en direct effect de onderwijspositie in het vierde jaar (Op4). Alhoewel een goed begin blijkbaar het halve werk is, blijkt het effect niet zo groot te zijn als verwachten zou kunnen worden. Dit is misschien toe te schrijven aan het feit dat vrij veel leerlingen in een brugklas (met meerdere schooltypen) beginnen.

Tabel 6. Gestandaardiseerde coëfficiënten in het schoolloopbaanmodel van Molukse en Nederlandse leerlingen (69 matchparen). Differentiële effecten voor Molukse (m) en Nederlandse (n) zijn vet gedrukt.

effect van:	op:	Zit	Adv	Op1	Azb	Asp	Hou	Szb	Op4
Sekse	m	.16							
	n	0	0	0	0	0	0	0	0
Beroep		0	0	0	0	0	0	0	.18
Zittenblijven		-	-.40	-.20	0	0	0	0	0
Advies	m						.36		
	n	-	-	.72	0	0	0	0	.51
Onderwijspositie jaar 1	m				.36		0	0	.35
	n				-.13				
Academisch zelfbeeld	m					.18	0	.24	-.16
	n						.36		0
Aspiraties		-	-	-	-	-	0	0	0
Houding		-	-	-	-	0	-	0	0
Sociaal zelfbeeld		-	-	-	-	0	0	-	0

model fit: chi-kwadraat = 74.34 met 82 vrijheidsgraden; $N = 2 \times 69$ ($p = .714$)
verklarede variantie in op 4: 67%.

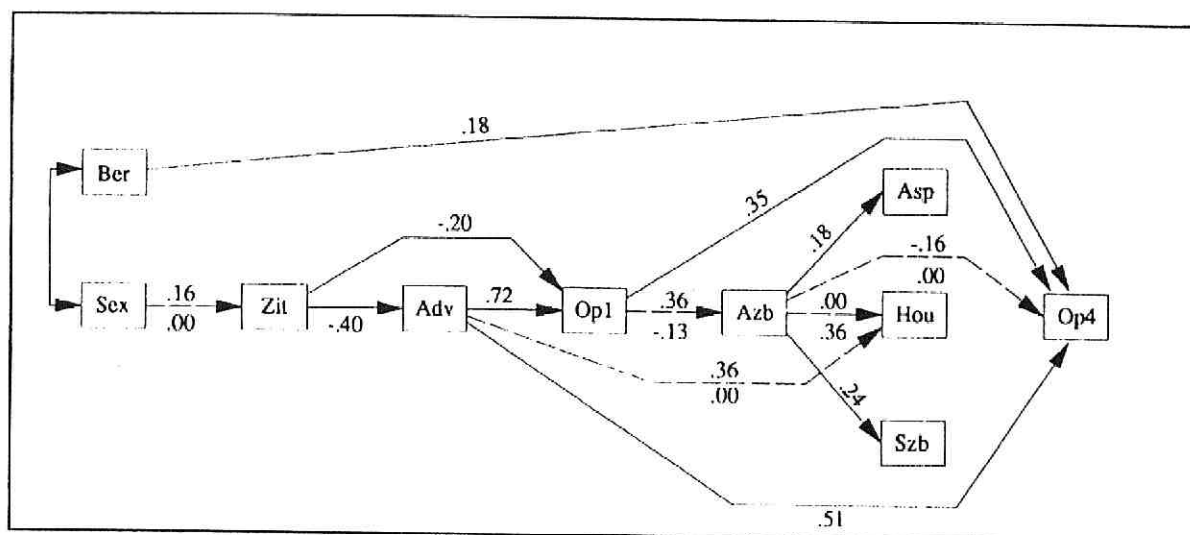


Fig. 2. Schoolloopbaanmodellen voor Molukse en Nederlandse leerlingen (gematchte paren).

Van de exogene variabelen heeft het beroep van de vader (Beroep) meestal een indirect effect, via advies of via de startpositie. In dit model heeft deze variabele een klein, maar significant direct effect op de bereikte onderwijspositie (Op 4). Sekse heeft een klein direct effect op zittenblijven: meisjes hebben minder vertraging opgelopen in het basisonderwijs dan jongens. Dit effect, ook bekend uit andere schoolloopbaanonderzoeken vinden we echter alleen bij de Molukse groep en niet bij de Nederlandse. We hebben geen verklaring voor dit verschijnsel, maar we veronderstellen dat het gaat om een toevallig resultaat bij de Nederlandse groep.

De grootste verschillen tussen de Nederlandse en de Molukse groep vinden we rondom de attitudeschalen, die we introduceerden om meer inzicht te krijgen in achtergronden van de schoolloopbaan in het voortgezet onderwijs. De effecten van deze schalen op de uiteindelijke positie in het voortgezet onderwijs (Op4) zijn echter zeer minimaal: voor de Molukse groep is er alleen een effect van academisch zelfbeeld (azb) en bij de Nederlandse leerlingen wordt helemaal geen effect gevonden van deze attitude op de uiteindelijke onderwijspositie. Onze conclusie is dan ook dat, hoewel er enkele differentiële effecten zijn voor Molukse en Nederlandse leerlingen, hun uiteindelijke onderwijspositie (de positie in het vierde leerjaar) op dezelfde manier beïnvloed wordt: door effecten van advies en startpositie en in mindere mate door het beroep van de vader.

De verschillen die optreden tussen de Molukse en de Nederlandse groep zijn moeilijk te interpreteren. Het academisch zelfbeeld van Molukse leerlingen wordt positief beïnvloed door hun startpositie; voor de Nederlandse leerlingen is dit (kleine) effect negatief. Hierbij moet bedacht worden dat academisch zelfbeeld geen absolute maat is, maar de waarneming van de leerling van de eigen prestaties reflecteert, relatief ten opzichte van het bereikte niveau in het voortgezet onderwijs. In zekere zin reflecteert deze maat het vertrouwen van een leerling in de eigen mogelijkheden tot presteren op een bepaald niveau. Het lijkt erop dat Nederlandse leerlingen die op een hoog niveau beginnen op dit gebied 'gemiddeld' zijn, terwijl Molukse leerlingen die op een hoog niveau starten ertoe neigen hun leerprestaties over te waarderen en te veel zelfvertrouwen hebben. Deze tendens zou ook het kleine negatieve effect van de Molukse groep van het academisch zelfbeeld op hun positie in het vierde jaar kunnen verklaren. Tot slot wordt de houding ten opzichte van onderwijs (Hou) positief beïnvloed door advies (voor Molukse leerlingen) en academisch zelfbeeld (voor Nederlandse leerlingen). We hebben geen kant-en-klare verklaring voor dit verschil, maar zoals genoemd heeft het maar weinig consequenties, omdat er geen effect is van de houding ten opzichte van onderwijs op de onderwijspositie in het vierde jaar.

CONCLUSIES EN DISCUSSIE

De Molukse bevolkingsgroep is de eerste allochtone groep in Nederland waarvan de kinderen van de derde generatie in betekenisvolle aantallen het voortgezet onderwijs bezoeken. Ondanks het feit dat Molukkers dus al lang in het land wonen bevindt de groep als geheel zich nog steeds in een achterstandssituatie, afgemeten aan de positie van Molukkers in het onderwijs en op de arbeidsmarkt. De werkloosheid onder Molukkers was bijvoorbeeld in het midden van de tachtiger jaren bijna twee keer zo hoog als die onder de totale bevolking. In een uitgebreide survey onder de Molukse groep werden het relatief lage opleidingsniveau van de Molukkers en het grote aantal voortijdig beëindigde opleidingen gevonden als belangrijke verklarende factoren. Alhoewel inmiddels is aangetoond dat Molukse jongeren éér de laatste jaren beter in slagen een hoger opleidingsniveau te realiseren dan de oudere Molukkers en bovendien de arbeidsmarktpositie de laatste jaren iets verbeterd is, blijft de opwaartse onderwijsmobiliteit van de Molukse groep nog steeds achter bij die van de autochtonen. In een tweede grootschalige survey in 1990 bleek namelijk opnieuw, zij het in geringere mate, sprake van een hoog percentage afgebroken opleidingen (drop-outs), een lange opleidingsduur en een relatief laag opleidingsniveau (Veenman 1990; Veenman en Martens, 1991).

Reeds geruime tijd voordat deze survey-onderzoeken de bestaande achterstand aantoonde, bestond de indruk dat Molukse leerlingen met name in de onderbouw van het voortgezet onderwijs problemen tegen kwamen. Als gevolg daarvan zou hun schoolloopbaan in deze fase van het v.o. aanmerkelijk ongunstiger verlopen dan de schoolloopbaan van Nederlandse leerlingen. Het onderhavige onderzoek zou de empirische gegevens moeten leveren om dit vermoeden te onderbouwen. Kern van het onderzoek werd gevormd door de vraag naar het verloop van de schoolloopbaan van Molukse leerlingen in de onderbouw van het v.o., de factoren die hiermee in verband gebracht kunnen worden en de mate waarin de schoolloopbaan van Molukse- en Nederlandse leerlingen onderling afwijkt.

In verschillende recente onderzoeken onder allochtone leerlingen komt naar voren dat het sociaal milieu waarin leerlingen opgroeien een cruciale variabele is in het verloop van schoolloopbanen (Tesser e.a., 1991; Van 't Hof en Dronkers, 1991; Van Langen en Jungbluth, 1989). Er is dan ook voor gekozen het onderzoeksbestand op te bouwen op basis van de methode van paarsgewijze vergelijking ofwel matching. Omdat met deze methode belangrijke variabelen constant gehouden worden is het mogelijk de invloed van 'storende' factoren – zoals in dit geval de sociaal economische achtergrond van leerlingen – meer te neutraliseren. Op deze wijze wordt het mogelijk de invloed van bijvoorbeeld etniciteit op het verloop van de schoolloopbaan nauwkeuriger te meten. In het onderzoek zijn Molukse leerlingen gematcht met Nederlandse leerlingen op de kenmerken sekse, leeftijd, beroep van de vader en advies voor het voortgezet onderwijs.

Uit de verzamelde schoolloopbaangegevens blijkt dat de schoolloopbanen van Molukse en vergelijkbare Nederlandse leerlingen in de eerste leerjaren van het v.o. een vrijwel gelijk verloop kennen. In ieder geval doen zich onder Molukse leerlingen de op voorhand verwachte verzuim-, uitval-, en afstroomproblemen in het tweede en derde leerjaar niet opvallend veel meer voor dan onder vergelijkbare Nederlandse leerlingen. Wel blijken Molukse leerlingen vaker te doubleren dan vergelijkbare Nederlandse leerlingen, een gegeven waaraan niet zonder meer voorbij gegaan kan worden. Doubleren kan namelijk leiden tot voortijdig schoolverlaten en zodoende voor problemen zorgen in de latere (beroeps)loopbaan. Aanwijzingen voor deze gedachte levert een eveneens recent uitgevoerd onderzoek van de Inspectie voor het Onderwijs, waarin Molukse leerlingen met Nederlandse leerlingen werden vergeleken op variabelen als zittenblijven, zakken voor het examen en schooluitval. Gevonden werd dat vooral op havo en vwo Molukse leerlingen 1.4 maal zo vaak doubleren in het pre-examenjaar, twee keer zo vaak zakken voor het examen en ruim twee keer zo vaak ongediplomeerd het onderwijs verlaten als hun Nederlandse leeftijdsgenoten (Inspectie van het Onderwijs, 1991).

Uit het gegeven dat onderwijsloopbanen van de Molukse en de Nederlandse leerlingen bij

– op onder meer sociaal milieu – gematchte groepen minder uiteenlopen dan op voorhand verwacht, mag niet geconcludeerd worden dat aan factoren die als ‘etnische factoren’ bestempeld kunnen worden (zoals cultuurverschillen, gerichtheid op de eigen groep/culturele oriëntatie, discriminatie en racisme, tweede taal-problemen) voorbijgegaan kan worden. Verschillen tussen beide groepen hebben we in dit onderzoek namelijk ook gevonden. Het gaat dan niet alleen om het feit dat Molukse leerlingen uit de derde generatie vaker blijven zitten dan vergelijkbare Nederlandse, maar ook om verschillen in scores op andere variabelen, bijvoorbeeld de in dit artikel besproken attitudeschalen ‘academisch en sociaal zelfbeeld’, ‘aspiraties’ en ‘houding ten aanzien van onderwijs’. Alhoewel van deze schalen slechts geringe invloed aangetoond werd op de onderwijspositie in het vierde leerjaar, is het toch belangrijk stil te staan bij de resultaten. De houding ten opzichte van school, de inschatting van leerlingen van hun eigen capaciteiten kunnen namelijk van belang zijn voor de verdere (school)loopbaan. Met name daar waar het gaat om het bereiken van hogere opleidings- en beroepsniveaus. Ook om een andere reden blijft het belangrijk om in onderzoek naar achterstanden en maatregelen ter bestrijding daarvan niet alleen rekening te houden met het sociaal milieu maar ook met de etnische groep waaruit leerlingen afkomstig zijn. Er is immers nog weinig bekend over de inhoud van ‘sociaal milieu’ bij allochtone en autochtone leerlingen. Zijn de sociale verschillen tussen ‘hoge’ en ‘lage’ milieus in beide groepen gelijk? Is de kloof tussen school en thuis voor allochtone en autochtone kinderen uit lagere sociaal-economische milieus van dezelfde aard en omvang? Zijn de belemmeringen die deze leerlinggroepen ondervinden, op sociaal en cognitief gebied, hetzelfde?

Zolang nog geen helder inzicht bestaat in de inhoud van variabelen als ‘sociaal milieu’ en ‘etniciteit’, hun samenhang en de wijze waarop in het onderwijs met uit het sociaal milieu of de etniciteit van leerlingen voortkomende cultuurverschillen wordt omgegaan, lijkt voorlopig het betrekken van deze variabelen in schoolloopbaanonderzoek van belang.

NOTEN

1. A. Veen, M. Robijns & J. Roeleveld, *Schoolloopbanen van Molukse leerlingen in het voortgezet onderwijs*. SCO-rapport 317. Amsterdam: SCO-KI.
2. Bij de benadering van v.o.-scholen is gebruik gemaakt van een overzicht van het Ministerie van Onderwijs en Wetenschappen van de v.o.-scholen die voor het schooljaar 1989-1990 faciliteiten voor allochtonen aanvroegen voor de op school ingeschreven Molukse leerlingen. Dit overzicht bevat 308 scholen. Hieruit zijn scholen geselecteerd die in het schooljaar 1989-1990 minimaal 10 leerlingen voor extra faciliteiten aanmeldden. Dit bleken 51 scholen te zijn. Deze scholen zijn voor het onderzoek benaderd.
3. Deze kwantificering van onderwijsposities (combinatie van niveau en klas) volgens de Formele Evaluatie Schaal (FES; Van den Eeden, De Jong, Koopman en Roeleveld, 1992) verloopt als volgt. Elke overgang naar een hogere klas van hetzelfde niveau krijgt een score van 10 punten en de afstanden tussen de opeenvolgende niveaus bedragen ook 10 punten. In de hier gehanteerde scoring krijgt Ibo de waarde 5, Lbo de waarde 10, Mavo de waarde 20, Havo de waarde 30 en Vwo de waarde 40. Een eerste klas krijgt de waarde 10, een tweede klas de waarde 20, etc. Een specifieke onderwijspositie krijgt nu als waarde de som van niveau en klas. Variaties tussen leerlingen op deze variabelen worden onder meer veroorzaakt door: schoolverlaten zonder diploma (drop-out), zittenblijven, afstroom of opstroom naar een ander schooltype, gediplomeerd verlaten van de school. Deze laatste situatie komt bij onze leerlingpopulatie niet voor, aangezien we de gegevens in het vierde leerjaar verzamelden. De range van posities loopt voor de onderzochte groep dan ook van 15 (eerste leerjaar laagste Ibo-niveau) tot 80 (vierde leerjaar hoogste Vwo-niveau).
4. De sociaal-economische status van de leerlingen is gemeten via de variabele *Beroep*, en wel het beroep van de vader, een gegeven dat doorgaans in leerlingadministraties van scholen aanwezig is. Uitgangspunt bij de codering is in dit onderzoek de CBS-indeling, zoals die weergegeven is in de Beroepenklassificatie 1984 (CBS, 1985). Een indikking hiervan naar 83 categorieën heeft plaatsgevonden met behulp van een indeling, gemaakt door deskundigen (Bakker, 1987; in enigszins gewijzigde vorm in Bakker, Jonker en Oud, 1988). In het onderzoeksproject “Schoolloopbanen in Amsterdam” (De Jong en Koopman, 1987; Van den Eeden, De Jong, Koopman en Roeleveld, 1992) zijn ‘optimal scaling’ technieken gebruikt om aan de oorspronkelijke 83 categorieën van de indeling van Bakker (1987) schaalwaarden toe te kennen.

Op basis van het SMVO-bestand (CBS, 1982) is met het programma HOMALS een homogeniteitsanalyse uitgevoerd op de 83-deling van beroep, de opleiding van de vader en de opleiding van de moeder. De categoriekwantificaties voor de beroepsindeling uit deze analyse worden in dit onderzoek gebruikt als scores op de variabele *Beroep*. Omdat een indeling van beroepen in 83 categorieën onwerkbaar leek bij de dataverzameling op de scholen, in het bijzonder bij het vinden van voldoende op het beroep van de vader gematchte leerlingen, is voor de matching-procedure gebruik gemaakt van een verdere indikking van de beroepsindeling tot 21 categorieën. In de kwantitatieve analyses is echter gebruik gemaakt van de verfijndere variabele *Beroep*.

5. Zoals eerder genoemd zijn de loopbaangegevens van dit cohort vergeleken met gegevens van cohort 1988-1989. De onderwijsresultaten van beide groepen leerlingen verschillen niet (significant) van elkaar. De conclusies die in dit onderzoek worden getrokken over de onderwijsloopbanen van het cohort 1987 mogen daarom met meer vertrouwen gegeneraliseerd worden.
6. Een gevaar van matching is het optreden van zogenaamde 'regressie-effecten'. (Zie bijvoorbeeld de discussie in het Tijdschrift voor Onderwijsresearch over het proefschrift van Peschar (Peschar, 1975). De meting op een bepaald tijdstip (het eerste tijdstip, t1 genoemd) zal, wanneer het om extreme scores gaat, gemiddeld een systematische onder- of overschatting zijn van de 'ware' score. Bij de tweede meting (t2) zal dan het verschijnsel 'regressie naar het gemiddelde' optreden: de extreem lage scores gaan omhoog en de extreem hoge scores gaan omlaag, louter als statistisch artefact en onafhankelijk van de 'experimentele condities' tussen de tijdstippen t1 en t2 (in dit onderzoek: onafhankelijk van de etniciteit van de leerlingen).
In de hier gehanteerde onderzoeksopzet is het optreden van dergelijke regressie-effecten echter niet te verwachten. In de eerste plaats zijn gegevens verzameld van alle Molukse leerlingen op de onderzochte scholen: er is dus geen extreem deel van deze groep geselecteerd. Verder vormen de bij deze leerlingen gematchte Nederlandse leerlingen weliswaar geen afspiegeling van de hele populatie van Nederlandse leerlingen die instromen in het voortgezet onderwijs, maar er is ook geen sprake van een extreem lage groep. Ook bij deze groep zijn er bijvoorbeeld leerlingen met Vwo-adviezen en met vaders in hogere beroepen. Tenslotte, indien er al regressie-effecten zouden optreden doordat de gematchte Nederlandse leerlingen een extreem 'lage' groep zouden zijn, dan zou het gevolg zijn dat hun tweede meting beter is dan op basis van de eerste meting verwacht kon worden. Het algemene beeld uit de gepresenteerde schoolloopbaangegevens, namelijk dat er weinig substantieel verschil is in onderwijsresultaten tussen Molukse en vergelijkbare Nederlandse leerlingen, zou daarmee alleen maar versterkt worden.
7. Resultaten zijn met de Wilcoxon Matched-pairs signed-ranks test eenzijdig getoetst op basis van de veronderstelling dat de schoolloopbanen van Molukse leerlingen ongunstiger verlopen dan die van de Nederlandse leerlingen (Siegel, 1956, p.75-83).
8. Bij een alfa van .1, hetgeen passend lijkt te zijn bij deze kleine aantallen.

LITERATUUR

- Autar, K. (1990). Uitstoting van leerlingen uit etnische minderheidsgroepen. De overrepresentatie van Surinaamse kinderen op scholen voor het speciaal onderwijs. In: V.Th. Tjon-A-Ten & K. Autar (red.), *Schoolloopbaanverbetering van Surinaamse leerlingen* (pp. 101-121). Amsterdam/Lisse: Swets & Zeitlinger.
- Bakker, B.F.M., Jonker, J.K. & Oud, J.A. (1988). *Een alternatieve aggregatie van beroepen op basis van de beroepenclassificatie 1984*. Voorburg: CBS.
- Centraal Bureau voor de Statistiek (CBS) (1985). *Beroepenclassificatie 1984*. Voorburg: CBS.
- Eeden, P. van den, Jong, U. de, Koopman, P. & Roeleveld, J. (1992). The effect of social status on educational attainment in Amsterdam: school differences and stability. *International Journal of Adolescence and Youth*, vol. 3, no. 1.
- Jong, U. de & Koopman, P. (1987). Stabiliteit van schoolloopbaanprocessen van leerlingen op zes Amsterdamse scholengemeenschappen voor HAVO en VWO tussen 1969 en 1980. In: J. Scheerens & W. Stoel (eds.), *Effectiviteit van onderwijsorganisaties*. Lisse: Swets & Zeitlinger.
- Jong, W. de & Verkuyten, M. (1989). Culturele oriëntatie en leerprestaties Nederlandse taal van Turkse kinderen. *Pedagogische Studiën*, 66, 296-306.
- Jöreskog, K.G. & Sörbom, D. (1988). *Lisrel VII: a guide to the program and applications*. Chicago, SPSS Inc.
- Ledoux, G. (1988). Kinderen uit etnische groepen in het speciaal onderwijs: verwijzing, toelating en begeleiding. In: A.G. Bus & S.J. Pijl (red.), *Diagnostiek en leerlingbegeleiding* (pp. 103-115). Amsterdam/Lisse: Swets & Zeitlinger.

- Ministerie van Onderwijs en Wetenschappen (1988). *Onderwijs aan Molukse leerlingen*. Inspectierapport 26. 's-Gravenhage: DOP.
- Molony, C. & Pechler, H. (1982). *Molukse leerlingen in het Nederlandse basisonderwijs*. Den Haag: SVO.
- Peschar, J.L. (1975). *Milieu, school, beroep: een achteraf-experiment over de periode 1958-1973 naar de invloed van het sociaal milieu op school- en beroepsloopbaan*. Groningen: Tjeenk Willink.
- Pille, L.J. (1989). *Kansloos, kansarm, kansrijk? De levenskansen van Molukkers in het onderwijs en op de arbeidsmarkt verkend*. Nijmegen: ISO.
- Roelandt, Th., Smeets, H.M.A.G. & Veenman, J. (1993). *Jaarboek minderheden 1993*. Houten/Zaventem: Bohn Stafleu Van Loghum.
- Roelandt, Th., Roijen, J.H.M. & Veenman, J. (1992). *Minderheden in Nederland; statistisch vademecum 1992*. Den Haag: CBS/ISEO-EUR.
- Roeleveld, J. & Jong, U. de (1989). *Evaluating effectiveness of secondary schools in The Netherlands: models and stability*. Paper 2nd International Congress for School Effectiveness, January 1989. Amsterdam: SCO.
- Siegel, S. (1956). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill Series in Psychology.
- Teunissen, F., & Golhof, A. (1987). Etnische minderheden en speciaal onderwijs. In: K. Doornbos & L.M. Stevens (red.), *De groei van het speciaal onderwijs* (pp. 166-198). Den Haag: Staatsuitgeverij.
- Veen, A., Robijns, M. & Roeleveld, J. (1992). *Schoolloopbanen van Molukse leerlingen in het voortgezet onderwijs*. SCO-rapport 317. Amsterdam: SCO-KI.
- Veenman, J. (1984). *De werkloosheid van Molukkers, deelrapport 1*. Rotterdam: EUR.
- Veenman, J. (1984a). "Het arbeidsmarktbeleid voor allochtonen. Twee passen vooruit, één achteruit?", *Economisch Statistische Berichten*, jrg. 69, nr. 3476, 10 oktober 1984a, p. 952-956.
- Veenman, J. (1984b). *De werkloosheid van Molukkers, deelrapport 2*. Rotterdam: EUR.
- Veenman, J. (1984c). *De werkloosheid van Molukkers, deelrapport 3: Beleidsaanbevelingen*. Rotterdam: EUR.
- Veenman, J. (1985). *De arbeidsmarktproblematiek van Molukkers, eindrapport*. Rotterdam: EUR.
- Veenman, J. (1990). *De arbeidsmarktpositie van allochtonen in Nederland, in het bijzonder van Molukkers*. Proefschrift. Groningen: Wolters-Noordhoff B.V.
- Veenman, J. & Martens, E.P. (1991). *Een lange weg. Molukkers op de arbeidsmarkt 1990*. Rotterdam: Erasmus Universiteit / ISEO.
- Vries, G.C. de, Monsma, G., Mellink, E., m.m.v. Ruijven, N. van (1990). *Als een slak, ongaarne, schoolwaarts*. Een onderzoek naar schoolverzuim van leerlingen en invloeden van de school. Amsterdam: SCO.

Manuscript ontvangen 3-2-1993

Definitieve versie ontvangen 31-3-1994

Effecten van schoolkenmerken en steldidactiek op functionele stelvaardigheid Nederlands in het MAVO¹

Hans Kuhlemeier*

Instituut voor Toetsontwikkeling (Cito), Arnhem

ABSTRACT

In this article relationships between functional writing ability, writing instruction in class, and school characteristics are analyzed. The data are gathered in a national assessment study on language proficiency of students in grade 11 (age 15). In this study 343 students from intermediate general education (MAVO) completed two functional writing tasks (a letter of application and a letter of apology). Multilevel analysis shows that 10 out of 36 instructional characteristics are related to functional writing ability. Among the effective instructional characteristics are: instruction in writing for a specific purpose, tailoring to a particular audience, focus on functional writing tasks, global rating of writing products by the teacher, and frequent evaluation of Dutch language proficiency through written assignments. No effects were found for the rather popular subskill exercises on idiom, spelling, punctuation and grammar, and for pre-writing activities, text revisions and peer-review. MAVO-students in schools with MAVO and vocational training (LBO) perform less well compared with MAVO-students in categorial MAVO-schools. Students in protestant and public schools outperform their catholic counterparts. It is concluded that these adverse effects of school-type and denomination are mainly due to differences in writing instruction.

1. INLEIDING

Stellen is een bij uitstek schoolse vaardigheid. In het dagelijks leven schrijven leerlingen weinig. Recent onderzoek toont aan dat het niveau van stelvaardigheid in het voortgezet onderwijs sterk van school tot school verschilt (De Glopper, 1987; 1990; Kuhlemeier & Van den Bergh, 1989b). Vanwege deze grote verschillen tussen scholen vormt stelvaardigheid een interessant object voor onderzoek naar instructie-effectiviteit. Tegelijkertijd moeten we constateren dat stelvaardigheid nog maar weinig als criterium in dergelijk onderzoek gebruikt is. Het enige ons bekende onderzoek in Nederland op dit terrein is dat van De Glopper (1990). Stelvaardigheid operationaliseerde hij als de algemene kwaliteit van drie verschillende schrijfproducten¹. De praktijk van het stelonderwijs werd geïnventariseerd met een vragenlijst. Slechts één instructiekenmerk bleek op de stelprestaties van derdeklassers van invloed te zijn. Het betrof onderwijs in de kenmerken van verschillende typen teksten (b.v. argumenterende en verhalende teksten, fictionele en zakelijke teksten). Als docenten hieraan aandacht schonken, kwam dat de stelvaardigheid van de derdeklassers ten goede. Leerlingen profiteerden evenwel niet van oefeningen in de verschillende onderdelen van het schrijfproces. Ook hadden zij geen baat bij de verschillende vormen van feedback op hun schrijfproducten. Maar wellicht nog meer dan om de bevindingen is het onderzoek van De Glopper van belang vanwege de gehanteerde analysetechniek. Hij maakte namelijk gebruik van multilevel-analyse, een techniek die het mogelijk maakt de invloed van leerling-, instructie- en schoolkenmerken simultaan te bestuderen (Goldstein, 1987).

In dit artikel staat de relatie tussen stelvaardigheid, schoolkenmerken en steldidactiek centraal. De effectmaat bestaat uit scores op twee korte schrijftaken: een sollicitatiebrief en een

1. Met dank aan Huub van den Bergh, Leijn Melse en Johan Wijnstra voor hun commentaar op een eerdere versie van dit artikel.

* Adres: Instituut voor Toetsontwikkeling (Cito), Postbus 1034, 6801 MG Arnhem.

excuusbrief. Dit zijn voorbeelden van functionele schrijftaken die van belang zijn voor het dagelijks leven en/of voor de latere beroepspraktijk. In de huidige onderwijspraktijk worden dit soort schrijftaken nog maar weinig onderwezen (Damhuis, De Glopper & Wesdorp, 1983; Kuhlemeier & Van den Bergh, 1989a). In de basisvorming Nederlands wordt voor functionele stelvaardigheid daarentegen een belangrijke plaats ingeruimd (Besluit kerndoelen, 1993). Docenten Nederlands moeten leren dit soort schrijftaken te onderwijzen. In dit onderzoek staat de vraag centraal naar het effect van het huidige stelonderwijs op de functionele stelvaardigheid. De praktijk van het stelonderwijs is in kaart gebracht met een vragenlijst, ingevuld door docenten Nederlands. Deze vragenlijst is samengesteld op basis van een review van onderzoek naar effectief stelonderwijs (Wesdorp, 1983) en een beschrijving van de praktijk van het schrijfonderwijs (Damhuis, De Glopper & Wesdorp, 1983). Het merendeel van de vragen gaat over stelopdrachten en -oefeningen die mogelijk een bijdrage leveren aan de stelprestaties van de leerlingen.

Reviews van onderzoek naar effectieve steldidactiek (Wesdorp, 1983; Hillocks, 1986) geven een overzicht van instructievariabelen waarvan een (positieve) invloed op de stelprestaties verwacht mag worden. Opmerkelijk is dat er nauwelijks aandacht wordt geschonken aan interacties tussen instructie en leerlingkenmerken. Er wordt (stilzwijgend) aangenomen dat de instructie voor alle leerlingen even effectief is. Een van de verworvenheden van multilevel analyse is nu dat een dergelijke aanname niet noodzakelijk is. Er kan een onderscheid gemaakt worden tussen effecten die gelden voor alle leerlingen en effecten die gelden voor bepaalde typen leerlingen. Zowel het uniforme als het niet-uniforme effect van de instructie kan simultaan geschat worden. In deze studie zijn twee leerlingkenmerken gebruikt: sekse en zittenblijven. Dit biedt de mogelijkheid om na te gaan of meisjes of zittenblijvers meer van bepaalde instructiekenmerken profiteren dan jongens of niet-zittenblijvers (en vice versa). Omdat functionele stelvaardigheid met twee schrijftaken gemeten is, kunnen we tevens nagaan in hoeverre het effect van de instructie van taak tot taak verschilt. Met andere woorden: worden de prestaties op de ene schrijftaak sterker door instructie beïnvloed dan die op de andere?

Als de ene leerling of school hogere stelprestaties behaalt dan de andere, hoeft dat niet per se aan de kwaliteit van het stelonderwijs te liggen. Of instructie effectief is, zal niet alleen afhangen van het type leerling, maar ook van het type school. Het is daarom van belang het effect van instructie te bepalen onder constanthouding van zoveel mogelijk andere factoren. In deze studie is een beperkte correctie voor de invloed van schoolkenmerken toegepast. Er is rekening gehouden met onder meer de structuur van de school (categoriaal versus scholengemeenschap), de grootte van de school, de denominatie (katholiek, protestants of openbaar), de bevolkingsdichtheid van het gebied waarin de school gevestigd is en de 'fusiehistorie' van de school. Deze schoolkenmerken beschouwen we als covariaten, om het effect van de instructie zuiverder te kunnen vaststellen. Tegelijkertijd krijgen we hiermee inzicht in de prestatieverschillen tussen deze typen scholen. Gegevens over het effect van schoolstructuur, schoolgrootte, denominatie, bevolkingsdichtheid en fusiehistorie op de leerprestaties in het voortgezet onderwijs zijn in Nederland schaars en vaak niet geheel eenduidig. Ten aanzien van de structuur van de school constateren Kleintjes & Kremers (1992) onder meer hogere leerprestaties van categoriale MAVO's in vergelijking met LBO/MAVO-scholengemeenschappen. Deze voorsprong van de categoriale MAVO's bleek zich wel voor te doen bij de vakken Engels, wiskunde en biologie, maar niet bij Nederlands. De schoolstructuur lijkt ook van belang vanwege de mogelijke relatie met het instroomniveau van de schoolbevolking. Zo is het denkbaar dat MAVO-leerlingen van LBO/MAVO-scholengemeenschappen aan het einde van de basisschool vaker een MAVO-advies met twijfel LBO krijgen, terwijl MAVO-leerlingen van scholengemeenschappen met HAVO en/of VWO vaker een MAVO-advies met twijfel HAVO ontvangen (vgl. De Jong & Roeleveld, 1989). De uitkomsten van onderzoek naar denominatie zijn vaak tegenstrijdig (vgl. Dijkstra, 1990). Voor zover er in het voortgezet onderwijs al kwaliteitsverschillen tussen scholen van een verschillende denominatie geconstateerd zijn, pakken deze uit in het voordeel van de katholieke sector (Van Laarhoven, Bakker, Dronkers & Schijf, 1987; Van Laarhoven & Dronkers, 1988; De Jong & Roeleveld, 1989) of in het voordeel van de protestantse sector

(Kleintjes & Kremers, 1992). Het belang van schoolgrootte voor de leerprestaties is in het voortgezet onderwijs waarschijnlijk niet erg groot. Zo konden Kleintjes en Kremers (1992) in het derde leerjaar voor geen van de vakken Nederlands, Engels, wiskunde en biologie een effect van schoolgrootte aantonen. Ook Luyten (1993) vond in het tweede leerjaar geen effect van schoolgrootte op de leerprestaties voor wiskunde, natuurkunde, scheikunde en biologie. Van de bevolkingsdichtheid is wel een effect aangetoond, al lijkt dat te verschillen naar vakgebied. In het onderzoek van Kleintjes en Kremers (1992) behaalden leerlingen in dunbevolkte gebieden hogere prestaties voor Nederlands, wiskunde en biologie, maar bij Engels hadden leerlingen in dichtbevolkte gebieden een voorsprong. Dit laatste verklaarden zij vanuit de grotere blootstelling aan de Engels taal in stedelijke gebieden, bij voorbeeld via buitenlandse televisiezenders. De regio binnen Nederland lijkt van belang in verband met de mogelijke interactie met het instroomniveau en de bevolkingsdichtheid. Er zijn aanwijzingen dat het instroomniveau van de MAVO's in plattelandsgebieden, met name in het noorden van Nederland, verhoudingsgewijs hoog is. Dit hangt samen met de slechte bereikbaarheid van de HAVO/VWO-scholen in deze gebieden. Daardoor zou een deel van de potentiële HAVO/VWO-leerlingen noodgedwongen voor een meestal wel goed bereikbare MAVO kiezen (Blank e.a., 1990). Het laatste schoolkenmerk is de fusiehistorie van de school. Sinds de uitvoering van ons onderzoek, in 1988, zijn nogal wat scholen in het kader van de schaalvergrotingsoperatie opgeheven dan wel opgegaan in een fusie. Opheffingen en fusies zijn ingrijpende processen (Van Kuijk & Van Kessel, 1986). Docenten die een opheffing of fusie 'overleefden', rapporteren onder meer een verzwaring van hun taken, een minder positief werkklimaat, minder plezier in het werk en meer onzekerheid (Pouwels, Van Esch & Vrieze, 1990). Ten aanzien van de gevolgen voor de leerprestaties vermoeden deze docenten een daling van het algehele onderwijsniveau op school (al lijkt het inhoudelijke niveau van het eigen vak er volgens hen niet onder te lijden). Volgens deze hypothese zouden de gemiddelde prestaties van de inmiddels opgeheven dan wel gefuseerde scholen relatief lager zijn dan van de scholen die daar in 1988 (nog) niet mee te maken hadden. Al met al staan in deze studie vier onderzoeksvragen centraal:

- Wat is het effect van de schoolkenmerken?
- Wat is het effect van de instructie met betrekking tot het stellen?
- In hoeverre verschilt het effect van de instructie van taak tot taak?
- In hoeverre is de effectiviteit van de instructie afhankelijk van de sekse en de doubleerstatus van de leerling?

In het vervolg van dit artikel komt eerst de opzet en uitvoering van het onderzoek aan bod (paragraaf 2). Daarna gaan we in op de verschillen tussen scholen (paragraaf 3.1), het effect van de schoolkenmerken (paragraaf 3.2) en het effect van het stelonderwijs (paragraaf 3.3). In paragraaf 3.4 trachten we de verschillen tussen effectieve en minder effectieve scholen te interpreteren in het licht van het gegeven stelonderwijs. Het artikel besluit met een samenvatting van de belangrijkste bevindingen en een discussie (paragraaf 4).

2. METHODE VAN ONDERZOEK

2.1. Dataverzameling en proefpersonen

De gegevens over de stelprestaties en het stelonderwijs zijn afkomstig uit een peilingsonderzoek naar de inhoud en het niveau van het vak Nederlands in het voortgezet onderwijs (Kuhlemeier & Van den Bergh, 1989a). De dataverzameling vond plaats in het schooljaar 1987-1988, aan het einde van het derde leerjaar. Van de 358 voor deelname benaderde opleidingen voor VWO, HAVO, MAVO, LTO of LHNO hebben er in totaal 186 (52%) aan het onderzoek deelgenomen.

De sollicitatie- en excuusbrief zijn voorgelegd aan een random steekproef van acht leerlingen uit alle derdeklassers van de opleiding waarvoor de school aan het onderzoek deelnam. Uit een verkennende analyse kwam naar voren dat het effect van de instructiekenmerken deels afhankelijk is van het opleidingstype in kwestie. In het ene opleidingstype bleken soms andere

instructiekenmerken een rol te spelen dan in het andere. Daarom is besloten de analyse tot één opleidingstype te beperken: het MAVO. In totaal zijn 343 MAVO-derdeklassers van 43 MAVO-opleidingen in de analyse betrokken (zie Bijlage 1). In dit opleidingstype bevinden zich de meeste derdeklassers en is de verdeling naar schoolstructuur het meest evenwichtig (dertig categoriale MAVO's en dertien MAVO-opleidingen binnen scholengemeenschappen).

Het effect van de schoolkenmerken wordt in deze studie vastgesteld op basis van een beperkt aantal MAVO-opleidingen ($N = 43$). Een relevante vraag is die naar de generaliseerbaarheid naar de populatie van MAVO-opleidingen in Nederland. Omdat de MAVO-opleidingen getrokken zijn met een trekkingskans evenredig aan het aantal derdeklassers en er binnen een opleiding een vast aantal derdeklassers getrokken is, is de steekproef van *leerlingen* zelfwegend ten opzichte van de populatie van MAVO-derdeklassers in Nederland (non-respons daargelaten). Echter, voor de steekproef van MAVO-opleidingen is dat niet noodzakelijkerwijs het geval (vgl. Moors & Van Muilwijk, 1985). Ter controle is de proportionaliteit van de responsgroep ten opzichte van de populatie nagegaan (zie Bijlage 1). Uit deze analyse blijkt dat de verdeling van de respons niet ernstig afwijkt van die in de populatie. Desalniettemin kan – gegeven een non-respons van bijna 50% – over de generaliseerbaarheid van de bevindingen naar de populatie van MAVO-opleidingen in Nederland geen uitsluitend worden gegeven. Overigens is dat voor de onderhavige studie slechts tot op zekere hoogte relevant. Het gaat immers vooral om het schatten van de invloed van de instructiekenmerken onder constanthouding van het effect van zoveel mogelijk andere variabelen (waaronder het effect van eventuele disproportionaliteit van de responsgroep van MAVO-opleidingen ten opzichte van de populatie van MAVO-opleidingen in Nederland).

2.2. De stelprestaties

Zoals gezegd maken wij voor deze studie gebruik van de scores op twee korte, functionele schrijftaken: een sollicitatiebrief en een excuusbrief. In de kerndoelen van de basisvorming wordt aan dit type schrijftaken een groot belang gehecht (Besluit kerndoelen, 1993). De sollicitatiebrief schrijft de leerling, naar aanleiding van een advertentie waarin om een vakantiehulp gevraagd wordt, aan de personeelschef van een postorderbedrijf. De excuusbrief wordt geschreven aan de directeur van de school. De leerling moet zich verontschuldigen voor feit dat hij of zij een afspraak wegens ziekte niet kan nakomen.

In totaal zijn tien kwaliteitsaspecten door twee jury's van negen beoordelaars onafhankelijk van elkaar beoordeeld. Van deze tien aspecten hebben er vijf betrekking op stelvaardigheid en vijf op technische conventies van interpunctie, spelling en grammatica. Wat betreft de stelvaardigheid is elk schrijfprodukt per aspect door een aselecte steekproef van drie uit negen beoordelaars beoordeeld (vgl. Kuhlemeier & Van den Bergh, 1989a). Omdat de scorebetrouwbaarheid van de vijf technische aspecten te wensen over laat, maken we – zowel voor de excuus- als de sollicitatiebrief – uitsluitend gebruik van de scores op de volgende vijf aspecten van stelvaardigheid:

1. Globale kwaliteit. Hierbij gaven de beoordelaars hun eerste indruk over de kwaliteit van het schrijfprodukt na één keer lezen.
2. Inhoud algemeen. Bij dit aspect gaven de beoordelaars een algemeen oordeel over de inhoudelijke kwaliteit van het briefje.
3. Stijl. Beoordeeld is in hoeverre het taalgebruik passend en begrijpelijk is; daarbij is tevens gelet op de treffendheid of vaagheid van de woordkeus.
4. Organisatie. Het oordeel over de organisatie weerspiegelt de wijze waarop de leerling de inhoud gepresenteerd heeft. Warrige en onoverzichtelijke schrijfproducten scoren op dit aspect laag.
5. Inhoud specifiek. Hier telden de beoordelaars met behulp van een checklist het aantal relevante inhoudselementen dat de leerling in zijn of haar brief aanroerde.

De beide schrijftaken zijn voldoende betrouwbaar beoordeeld. De jurybetrouwbaarheid varieert van .71 voor het oordeel over de organisatie van de excuusbrief tot .89 voor het oordeel over de

inhoud van de sollicitatiebrief, met een gemiddelde van .77 (vgl. Kuhlemeier & Van den Bergh, 1989a).

2.3. School- en leerlingkenmerken

Voor de analyse hebben we de beschikking over zes schoolkenmerken en twee leerlingkenmerken. De schoolkenmerken zijn ontleend aan het bestand 'Integrale Leerlingentelling in het Voortgezet Onderwijs 1987-1988' van het Ministerie van Onderwijs en Wetenschappen. De leerlingkenmerken zijn afkomstig uit een door de leerlingen ingevulde achtergrondvragenlijst. Voor de verdeling van deze variabelen in de responsgroep zij verwezen naar Bijlage 1.

Schoolstructuur. Een onderscheid is gemaakt in categoriale MAVO's die alleen tot het MAVO-diploma opleiden en MAVO-opleidingen binnen LBO/MAVO-scholengemeenschappen².

Schoolgrootte. De grootte van de school is simpelweg gemeten als het totale aantal leerlingen op school. Overigens zijn schoolgrootte en schoolstructuur positief gecorreleerd. Vanzelfsprekend tellen categoriale MAVO's doorgaans minder leerlingen dan scholengemeenschappen (vgl. Bijlage 1).

Denominatie. Een onderscheid is gemaakt tussen protestants-christelijke, rooms-katholieke en openbare MAVO's³.

Bevolkingsdichtheid. De bevolkingsdichtheid is gemeten als het totaal aantal inwoners per vierkante kilometer van het gebied waarin de school gevestigd is.

Regionale spreiding. Deze variabele geeft aan of de MAVO-opleiding zich in het noorden, oosten, midden, zuiden of westen van Nederland bevindt.

Opgeheven of gefuseerd. De dataverzameling vond plaats aan het einde van het schooljaar 1987-1988. Sindsdien zijn in het kader van de schaalvergrotingsoperatie veel MAVO-opleidingen opgeheven dan wel opgegaan in een fusie. Een beleidsmatig interessante vraag is die naar de prestatieverschillen tussen enerzijds de MAVO-opleidingen die nog steeds bestaan en anderzijds de MAVO-opleidingen die zijn opgeheven dan wel gefuseerd. Voor de 43 MAVO-opleidingen is daartoe nagegaan of het schoolnummer of de schoolnaam nog voorkomt in het bestand 1991-1992 van het Ministerie van Onderwijs en Wetenschappen. Van niet te traceren scholen mag worden aangenomen dat deze zijn opgeheven dan wel zijn gefuseerd met een andere school.

Sekse en doubleerstatus. Meisjes zijn gemiddeld betere stellers dan jongens (vgl. Kuhlemeier & Van den Bergh, 1991a). Een interessante vraag is bij voorbeeld of de voorsprong van meisjes in alle MAVO-opleidingen even groot is. Bij de doubleerstatus gaat het erom of de leerling in het voortgezet onderwijs al dan niet is blijven zitten.

2.4. Instructiekenmerken

De gegevens over de praktijk van het schrijfonderwijs zijn ontleend aan een vragenlijst over het onderwijsaanbod bij Nederlands (Kuhlemeier & Van den Bergh, 1989a; Van den Bergh & Kuhlemeier, 1991). Deze vragenlijst gaat niet alleen over het schrijfonderwijs. Er is ook gevraagd naar zaken als onderwijsopvattingen, gebruikte leergangen, onderwijsleermiddelen, werkvormen, differentiatie en evaluatie. Tevens zijn vragen gesteld over het onderwijs in de vakonderdelen luisteren, spreken en lezen. De vragen zijn voorgelegd aan de docenten Nederlands die lesgeven aan de derdeklassers die aan de peiling deelnamen. Van de 43 potentiële MAVO-respondenten hebben er 39 (91%) de vragen over het schrijfonderwijs ingevuld.

De rubriek over de praktijk van het stelonderwijs is gebaseerd op een review van onderzoek naar effectief stelonderwijs (Wesdorp, 1983). Een deel van de vragen betreft instructiekenmerken waarvoor empirische evidentie gevonden is (zie ook Hillocks, 1986): creatief schrijven ('free writing'), oefeningen in syntactische variantie, zinsbouw en structureren (o.a. 'sentence combining'), oefeningen in het voorbereiden en plannen van een schrijfprodukt, oefeningen in tekstrevisie, en oefeningen waarbij leerlingen commentaar geven op elkaars schrijfproducten ('peer review'). Andere vragen betreffen instructiekenmerken die in de literatuur als niet of nauwelijks effectief te boek staan, maar die in de huidige lespraktijk relatief populair zijn (vgl. Damhuis, De Gloppe & Wesdorp, 1983; Kuhlemeier & Van den Bergh, 1989a). Het betreft onder meer: evaluatie van schrijfproducten door de leerkracht ('teacher comments'), oefenin-

gen ter uitbreiding van de woordenschat en aandacht voor de meer technische aspecten van het schrijven, zoals oefeningen in spelling en interpunctie. Daarnaast bevat dit gedeelte van de vragenlijst een aantal vragen over instructiekenmerken waarover voor zover ons bekend nog geen empirische gegevens voorhanden zijn, maar die met name voor het schrijven van functionele teksten van belang lijken. Het gaat daarbij om het schrijven van korte, functionele schrijftaken (zoals een briefje aan een vriend of een familielid, een briefje aan een instantie of bedrijf, een korte instructie of een korte overredende tekst) en om oefeningen in het schrijven van zakelijke, doelgerichte en publiekgerichte teksten. Ten slotte zijn voor deze studie enkele vragen geselecteerd die niet uit de rubriek 'Schrijfonderwijs' afkomstig zijn, maar die toch voor de stelprestaties van belang lijken. Deze vragen betreffen de schriftelijke evaluatie van de leervorderingen bij Nederlands in zijn algemeenheid. Een onderscheid is gemaakt in evaluatie via extern ontwikkelde toetsen en evaluatie via zelf gemaakte proefwerken, schriftelijke overhoringen of werkstukken. Deze kenmerken van de algemene evaluatie hebben weliswaar slechts zijdelings met stelvaardigheid te maken. Ze zijn echter in de analyse betrokken omdat zij alle een sterk beroep doen op de schrijfvaardigheid van de leerlingen. Bovendien blijkt frequente evaluatie één van de kenmerken van buitengewoon effectieve scholen (Levine & Lezotte, 1990; Creemers, 1992; Veenman, Lem, Roelofs & Nijssen, 1992). In totaal gaat het om 36 instructiekenmerken, verdeeld over de volgende rubrieken:

Vormen van schrijven. Gevraagd is naar de mate waarin de leerlingen geconfronteerd worden met creatief schrijven (vrij, fictioneel of expressief schrijven) en met zakelijk schrijven (gericht, factioneel of transactioneel schrijven).

Deeloefeningen. Hier gaat het onder meer om de frequentie van oefeningen in handschriftverbetering, spelling- en interpunctie, grammaticale oefeningen op zins- en tekstniveau, voorbereidings- en planningsactiviteiten, oefeningen in doel- en publiekgerichtheid en tekstrevisie.

Instructie in teksttypen. De docenten Nederlands is gevraagd naar de mate waarin een aantal teksttypen in het schrijfonderwijs aan bod komt. Het betreft zowel langere teksten, zoals het schrijven van een beschouwend opstel of een zelfverzonnen verhaal, als kortere teksten, zoals het schrijven van een brief aan vrienden of familie, een brief aan een officiële instantie of bedrijf of een korte instructie.

Evaluatie van stelvaardigheid. Gevraagd is naar de wijze waarop de docent feedback geeft op de schrijfproducten van de leerlingen en naar de criteria die de docent bij de beoordeling aanlegt.

Lezen van zakelijke teksten. Deze vraag betreft de frequentie waarmee zakelijke teksten in het leesonderwijs aan bod komen. Zakelijke teksten worden immers niet alleen in het schrijfonderwijs gebruikt.

Algemene schriftelijke evaluatie bij Nederlands. De docenten Nederlands is gevraagd naar de mate waarin zij de leerprestaties van hun leerlingen schriftelijk evalueren.

De vragen zijn van het Likerttype. Bij elke vraag moest de docent Nederlands aangeven hoe vaak het in de uitspraak gestelde in zijn of haar lessen voorkomt. Een overzicht met enkele beschrijvende statistische gegevens is opgenomen in Bijlage 2.

2.5. Statistische analyse

Voor elk leerlingprodukt beschikken we over de oordelen van drie beoordelaars over de globale kwaliteit, inhoud, stijl en organisatie alsmede over de tellingen van twee beoordelaars van het aantal relevante inhoudselementen. De scores van een leerling zijn per schrijftaak per aspect over beoordelaars gesommeerd tot één score per leerling. Na eliminatie van de leerlingen waarvan de sekse en/of de doubleerstatus onbekend waren, resteerden in totaal 3361 van deze leerlingsscores, afkomstig van 343 MAVO-derdeklassers van 43 scholen. Het aantal leerlingen per school varieert van zes tot negen, met een gemiddelde van 7.97. Het aantal scores per leerling bedraagt maximaal tien: vijf voor de excuusbrief en vijf voor de sollicitatiebrief. Het werkelijke aantal scores per leerling varieert van vijf tot tien, met een gemiddelde van 9.89.

Als analysetechniek maken we gebruik van multilevel analyse, uitgevoerd met het programma ML3 (Prosser, Rasbash, & Goldstein, 1991). Overeenkomstig de structuur van de dataset

- scores genest binnen leerlingen die zelf weer genest zijn binnen scholen - onderscheiden we drie niveaus: score, leerling en school. De totale variantie wordt daarbij opgesplitst in drie delen:
- een tussen-scores-binnen-leerlingcomponent (scoreniveau);
- een tussen-leerlingen-binnen-scholencomponent (leerlingniveau);
- een tussen-scholencomponent (schoolniveau).

Zoals gezegd beschikken we voor elke leerling over tien (2 maal 5) scores. In een multilevel factoranalyse is de dimensionaliteit van deze scores op school- en leerlingniveau nagegaan (vgl. Kuhlemeier & Van den Bergh, in voorbereiding). De correlaties tussen de scores voor globale kwaliteit, inhoud algemeen, stijl, organisatie en inhoud specifiek waren dermate hoog dat het onderscheid in deze vijf stelvaardigheidsaspecten empirisch gezien als betekenisloos beschouwd moest worden. Zowel op school- als op leerlingniveau was er sprake van één factor die we interpreteerden als de algemene kwaliteit van het stellen. Het onderscheid tussen de beide schrijftaken vond daarentegen wel steun in de data. Voor de modellering van de afhankelijke variabele betekent dit dat we met twee factoren kunnen volstaan: één voor de sollicitatiebrief en één voor de excuusbrief.

Om de onderzoeksvragen te beantwoorden onderscheiden we drie statistische modellen. Het eerste model is het zogeheten onconditionele of nulmodel, omdat de verklarende leerling-, school- en instructievariabelen daarin nog geen rol spelen (model 1). In het tweede en derde model voegen we eerst de school- en leerlingkenmerken (model 2) en daarna de instructiekenmerken toe (model 3). Omdat de instructiekenmerken na de school- en leerlingkenmerken in de regressievergelijking zijn ingebracht, wordt het effect van de instructie geschat onder constant-houding van het effect van de school- en leerlingkenmerken. We willen immers weten wat het effect van de instructie is nadat de invloed van andere variabelen is uitgeschakeld. Hieronder lichten we de modellen toe.

Onconditionele model (model 1)

In het onconditionele model zijn er twee stelvaardigheidsfactoren: één representeert de sollicitatiebrief en de ander de excuusbrief. Omdat de 2 maal 5 toetsscores genest zijn binnen leerlingen modelleren we de varianties van deze toetsscores op het eerste analyseniveau. Op dit 'tussen-scores-binnen-leerlingniveau' schatten we voor elke taakfactor een afzonderlijke residuele variantie. Deze residuele variantie representeert dat deel van de totale variantie dat, gegeven het model, als 'ruis' te beschouwen is.

De verhouding tussen de varianties op school- en leerlingniveau geeft inzicht in de verdeling van de systematische variantie over beide niveaus. Omdat de scores per taak per aspect gestandaardiseerd zijn, is de totale variantie van een factor gelijk aan één. De intraklassecorrelatie, gedefinieerd als de ratio van de tussen-scholenvariantie ten opzichte van deze totale variantie, is dan gelijk aan de tussen-scholenvariantie. Omdat de totale variantie een deel 'ruis' bevat, is de hoogte van de intraklassecorrelatie een onderschatting van de 'ware' intraklassecorrelatie (Muthén, 1992). Deze 'error free' intraklassecorrelatie is eenvoudig te berekenen door de tussen-scholenvariantie te delen door de totale systematische variantie (de som van de tussen-scholenvariantie en de tussen-leerlingenvariantie). In de analyse gaan we ervan uit dat de twee taakfactoren met elkaar samenhangen (oblique factoren). Beide hebben immers betrekking op het schrijven van functionele brieven. Zowel op school- als op leerlingniveau schatten we behalve de varianties van beide taakfactoren ook hun covariantie. Op schoolniveau representeert de covariantie de mate waarin scholen op beide schrijftaken even hoge (gemiddelde) prestaties behalen; op leerlingniveau wordt de samenhang tussen de (gemiddelde) leerlingcores binnen scholen weergegeven.

De betrouwbaarheid van de gemiddelden⁴ op schoolniveau (ρ_s) bedraagt voor de sollicitatiebrief .51 en voor de excuusbrief .26; voor de meting op leerlingniveau bedragen de overeenkomstige betrouwbaarheden .78 en .81.

Leerling en school (model 2)

In het tweede model wordt het effect van de school- en leerlingkenmerken geschat. Daartoe voegen we de desbetreffende verklarende variabelen als zogeheten fixed effecten aan model 1 toe. Merk op dat de 3-levelstructuur van het model hierbij intact wordt gelaten; ook nu worden met andere woorden variantiecomponenten op score-, leerling- en schoolniveau onderscheiden (Goldstein, 1987, p. 51 e.v.). In de analyse is expliciete aandacht geschonken aan interacties tussen school- en leerlingkenmerken, aan interacties tussen schoolkenmerken onderling en aan interacties met de schrijftaak (sollicitatiebrief versus excuusbrief).

Leerling, school en instructie (model 3)

Vervolgens zijn de instructiekenmerken aan model 2 toegevoegd. Daarbij is speciaal gelet op interacties met de schrijftaak, de sekse en de doubleerstatus. Vanwege de beperkte omvang van de dataset zijn de interacties met deze drie variabelen in afzonderlijke analyses bestudeerd: model 3a (schrijftaak), 3b (sekse) en 3c (doubleerstatus).

De geringe omvang van de dataset in aanmerking nemend – 43 MAVO-scholen met circa acht leerlingen per school – is 36 verklarende variabelen rijkelijk veel. Als het aantal predictoren in verhouding tot het aantal waarnemingen te groot wordt, dreigt het gevaar van overparametrisering. Teneinde tot een hanteerbaar aantal predictoren te komen, zijn van te voren enkele statistische analyses uitgevoerd. Met behulp van homogeniteitsanalyse en exploratieve factoranalyse is getracht het aantal verklarende variabelen enigszins in te dikken. Dit laatste was echter maar zeer ten dele mogelijk (voor een vergelijkbare bevinding, zie Damhuis, De Gloppe & Wesdorp, 1983). Zelfs tussen inhoudelijk verwante instructiekenmerken bleek de samenhang soms laag. Slechts in enkele gevallen was het verantwoord om afzonderlijke instructiekenmerken te sommeren tot composietvariabelen. Dit betekent dat toch met een groot aantal afzonderlijke predictoren gewerkt moet worden, met alle nadelen vandien. Om het gevaar van overparametrisering tegen te gaan zijn de verklarende variabelen niet alle tegelijkertijd in de analyse ingebracht. Er is een stapsgewijze aanpak gehanteerd, een combinatie van 'forward' en 'backward' selectie. Volgens de forward-selectieprocedure zijn de verklarende variabelen eerst één voor één – met teruglegging – in de regressievergelijking ingebracht. Vervolgens is de sterkste predictor geselecteerd, d.w.z. de verklarende variabele met de grootste positieve of negatieve (partiële) regressie. Als het desbetreffende regressiegewicht significant van nul verschilde ($p < .05$) zijn vervolgens de resterende verklarende variabelen volgens dezelfde procedure geanalyseerd, waarbij opnieuw de sterkste (significante) kandidaat gekozen is. Volgens de backward-selectieprocedure zijn na toevoeging van een predictor steeds de reeds geselecteerde verklarende predictoren opnieuw op significantie getoetst en zo nodig (vooral nog) uit het model verwijderd. De selectieprocedure is beëindigd op het moment dat er geen nieuwe significante predictoren meer naar voren kwamen.

De regressiegewichten laten zien wat de invloed is op de effectmaat van elke verklarende variabele afzonderlijk (rekening houdend met alle overige predictoren in het model). Ze geven geen antwoord op de vraag in hoeverre de niveauverschillen tussen scholen aan de instructie dan wel aan andere kenmerken zijn toe te schrijven. Deze vraag beantwoorden we aan de hand van het percentage verklaarde variantie⁵. Een eenduidige maat voor het interpreteren van de grootte van effecten is het percentage van de tussen-scholenvariantie dat verklaard wordt: R^2_{school} . In de evaluatie van de verschillende modellen staat de toename van R^2_{school} centraal.

De grootte van het gezamenlijke effect van de school- en leerlingkenmerken evalueren we aan de hand van de toename van R^2_{school} van model 2 ten opzichte van het onconditionele model (model 1). Omdat dit laatste model geen verklarende variabelen bevat, is het percentage verklaarde variantie uiteraard nul. Het percentage onverklaarde variantie uit model 2 vormt de bovengrens van hetgeen we – gegeven ons model – met alle instructiekenmerken kunnen verklaren (vgl. Willms & Raudenbush, 1989; Bosker, 1991; Bryk & Raudenbush, 1992). De additieve invloed van alle instructiekenmerken gezamenlijk evalueren we aan de hand van de toename van de R^2_{school} van model 2 naar model 3.

De significantietoetsingen van de (verschillen tussen) varianties en regressiegewichten zijn

uitgevoerd met behulp van een multiple vergelijkingsprocedure (Goldstein, 1987). Deze procedure levert een toetsingsgrootte op die bij benadering Chi-kwadraat verdeeld is. Bij de statistische toetsingen is steeds het 5%-significantie criterium gehanteerd. In het geval van de interacties met schrijftaak, sekse of doubleren spreken we pas van een differentieel effect als aan twee voorwaarden voldaan is. Ten eerste moeten beide regressiegewichten van een instructiekenmerk significant afwijken van nul. En ten tweede moet het verschil tussen beide regressiegewichten significant zijn.

3. BEVINDINGEN

3.1. Verschillen tussen de MAVO-opleidingen

Een cruciale vraag is die naar de mate waarin de MAVO-opleidingen verschillen in het niveau van stelvaardigheid. Als er geen verschillen tussen opleidingen zijn, heeft onderzoek naar instructie- en schooleffectiviteit op voorhand weinig zin. De proporties variantie op school-, leerling- en scoreniveau (model 1) zijn weergegeven in Tabel 1.

Allereerst blijkt uit Tabel 1 dat het scoreniveau zo'n 50% van de totale variantie naar zich toetrekt, zowel bij de sollicitatiebrief als bij de excuusbrief. De helft van de variantie van de 2 maal 5 scores wordt dus niet door de beide taakfactoren verklaard en moet – gegeven het model – als residuele variantie of 'ruis' worden beschouwd. Dit vestigt weer eens de aandacht op het gegeven dat produktieve taalvaardigheid moeilijk te meten is (o.a. Wesdorp, 1981; Meuffels, 1994).

Tussen de MAVO-opleidingen, zo blijkt uit Tabel 1, bestaan aanzienlijke verschillen in het gemiddelde prestatieniveau op de sollicitatie- en excuusbrief. Voor de sollicitatiebrief trekt het schoolniveau 30% [$.15/ (.15+.35)$] van de systematische variantie naar zich toe. Op de ene MAVO schrijven de leerlingen dus relatief goede sollicitatiebrieven en op de andere MAVO relatief slechte. Voor de excuusbrief liggen de gemiddelde prestaties van de MAVO-opleidingen veel dichter bij elkaar. Hier hangt slechts 12% [$(.06/ (.06+.45))$] van de systematische variantie samen met de MAVO-opleiding die de leerling bezoekt.

Dat de MAVO-opleidingen sterker verschillen in de kwaliteit van de sollicitatiebrief dan van de excuusbrief lijkt goed te verklaren. Volgens een enquête-onderzoek van Damhuis, De Gloppe & Wesdorp (1983) wordt de sollicitatiebrief op sommige MAVO's nog wel eens geoefend. Voor de excuusbrief is dat veel minder aannemelijk. Daar wordt waarschijnlijk helemaal geen aandacht aan besteed.

Als een MAVO-opleiding goed presteert op de ene taak, doet het dat ook op de andere; de correlatie tussen de schoolgemiddelden voor de sollicitatie- en de excuusbrief is echter vrij laag ($r = .36$). Ook op leerlingniveau is het verband tussen de leerlinggemiddelden voor de sollicitatie- en de excuusbrief zwak ($r = .46$). Wederom blijkt dat de ene taak de andere niet is (zie o.a. Kuhlemeier & Van den Bergh, 1991b).

3.2. Invloed school en leerling

De stapsgewijze analyse van de regressie van stelvaardigheid op de school- en leerlingkenmerken brengt drie invloedrijke variabelen aan het licht: de structuur van de school (MAVO-

Tabel 1. Variantiedecompositie en correlaties tussen de factoren voor de beide schrijftaken (model 1).

Niveau	Factor I Sollicitatiebrief	Factor II Excuusbrief	$r_{(F1,F2)}$
School (N = 43)	.15	.06	.36
Leerling (N = 343)	.35	.45	.46
Score (N = 3361)	.50	.50	

Tabel 2. Verschil tussen de gemiddelden voor schoolstructuur, denominatie en sekse (berekend op grond van model 2; onderstreepte verschillen wijken significant af van nul op 5%-niveau).

	verschil
Schoolstructuur	
categoriale MAVO – MAVO in LBO/MAVO-SG	<u>.40</u>
Denominatie	
katholiek – protestant	-.19
katholiek – openbaar	-.25
protestant – openbaar	-.06
Denominatie * sekse	
meisjes – jongens (katholiek)	.19
meisjes – jongens (protestant)	<u>.35</u>
meisjes – jongens (openbaar)	<u>.52</u>

categoriaal versus LBO/MAVO-scholengemeenschappen), de denominatie van de school (openbaar, protestant of katholiek) en de sekse van de leerling. De verschillen tussen de gemiddelden van de onderscheiden groepen scholen zijn weergegeven in Tabel 2. Omdat de afhankelijke variabelen gestandaardiseerd zijn (met een gemiddelde van nul en een variantie van één) en de onafhankelijke variabelen zogeheten dummies zijn, kunnen de weergegeven getallen tot op zekere hoogte geïnterpreteerd worden als effectgroottes (d.w.z.: het verschil tussen twee gemiddelden gedeeld door de standaarddeviatie; vgl. Cohen, 1977).

Uit Tabel 2 blijkt dat de structuur van de school van invloed is op de stelprestaties. MAVO-leerlingen behalen binnen LBO/MAVO-scholengemeenschappen lagere stelprestaties dan binnen categoriale MAVO's (.40 sd).

Het tweede schoolkenmerk dat van belang is, is de denominatie van de school. Op katholieke MAVO's behalen de leerlingen lagere stelprestaties dan op protestantse (.19 sd) of openbare MAVO's (.25 sd). Het verschil tussen de protestantse en openbare MAVO's is niet significant.

Het derde belangrijke kenmerk is de sekse van de leerling. Significantietoetsing laat zien dat meisjes gemiddeld genomen beter zijn in stellen dan jongens. Opvallend is de interactie met de denominatie van de school. De kloof tussen de prestaties van meisjes en jongens blijkt niet onafhankelijk van de richting van de school. In verband hiermee zijn in Tabel 2 de gemiddelden voor jongens en meisjes binnen elke denominatie afzonderlijk weergegeven (geschat onder constanthouding van het effect van de structuur van de school). In de katholieke MAVO's, waar de leerlingen de laagste stelprestaties behaalden, is de voorsprong van meisjes niet significant (.19 sd). In de protestantse MAVO's en openbare MAVO's is het verschil in het voordeel van meisjes wel significant (respectievelijk .35 en .52 sd).

Nadat de schoolstructuur, de denominatie en de interactie van denominatie * sekse in de regressievergelijking zijn ingebracht kan van de overige school- en leerlingvariabelen geen additief effect meer worden aangetoond. Gegeven deze variabelen zijn de doublecerstatus, de grootte van de school, de bevolkingsdichtheid, de vestigingsregio en de fusiehistorie voor de stelprestaties van geen of ondergeschikt belang. Onder meer kunnen we dan concluderen dat grote MAVO's het niet beter of slechter doen dan kleine MAVO's, dat MAVO's in dichtbevolkte gebieden niet lager presteren dan MAVO's in dunbevolkte gebieden en dat het zeker niet alleen de slechtste MAVO's zijn die in de schaalvergrotingsoperatie het eerst het loodje leggen.

3.3. Invloed stelonderwijs

Het effect van de instructiekenmerken en hun interacties met schrijftaak, sekse en doubleren zijn in afzonderlijke modellen geanalyseerd (respectievelijk model 3a, 3b en 3c). Hieronder bespreken we de bevindingen.

Instructie en schrijftaak

De uitkomsten van de stapsgewijze analyse van regressie van de stelprestaties op de 36 instructiekenmerken (model 3a) zijn samengevat in Tabel 3. Wellicht ten overvloede zij opgemerkt dat de regressiegewichten berekend zijn na uitschakeling van het effect van schoolstructuur, denominatie en de interactie van denominatie * sekse.

Van de 36 instructiekenmerken, zo blijkt uit Tabel 3, zijn er negen voor de stelprestaties van belang. Gegeven deze negen, leveren de overige 25 instructiekenmerken geen additieve bijdrage aan de voorspelling.

Slechts twee instructiekenmerken hebben een positief effect op de prestaties voor zowel de sollicitatie- als de excuusbrief. Naarmate leerlingen vaker brieven aan instanties en bedrijven schrijven, is de kwaliteit van beide brieven hoger. Hetzelfde geldt voor een regelmatige evaluatie van de leervorderingen bij Nederlands met behulp van proefwerken en schriftelijke overhoringen.

Voor drie instructiekenmerken geldt dat het positieve effect slechts bij één van beide schrijftaken aantoonbaar is. Als er in het onderwijs brieven aan vrienden en familie worden geschreven, komt dit de kwaliteit van de sollicitatiebrieven ten goede. Ook het evalueren van de leervorderingen aan de hand van schriftelijke werkstukken heeft alleen een positief effect op de sollicitatiebrief. Het omgekeerde beeld zien we bij de globale beoordeling van schrijfprodukten als geheel. Als docenten de schrijfprodukten van hun leerlingen op een globale wijze beoordelen, behalen de leerlingen hogere prestaties op de excuusbrief.

In één geval is er sprake van een positief effect voor de ene schrijftaak en een negatief effect voor de andere schrijftaak. Als zakelijke brieven, waartoe ook de sollicitatiebrief behoort, in het leesonderwijs aan bod komen, komt dit de kwaliteit van de sollicitatiebrief ten goede; tegelijkertijd blijkt het lezen van zakelijke brieven gepaard te gaan met lagere prestaties op de excuusbrief.

Voor drie instructiekenmerken is het teken van het regressiegewicht negatief. Als docenten veel tijd besteden aan beschouwende opstellen, gaat dit ten koste van de prestaties op de sollicitatiebrief. Ook het gebruik van extern ontwikkeld toetsmateriaal heeft een negatief effect en ook nu geldt dit alleen voor de sollicitatiebrief. Veel aandacht voor de verbetering van het handschrift gaat daarentegen ten koste van de prestaties op de excuusbrief.

Tabel 3. Effect van de instructiekenmerken op de stelprestaties voor de sollicitatie- en de excuusbrief (model 3a; onderstreepte regressiegewichten verschillen significant van nul op 5%-niveau).

	Sollicitatiebrief	Excuusbrief
Effectief		
brief aan officiële instantie of bedrijf	<u>.11</u>	<u>.11</u>
brief aan vrienden of familie	<u>.11</u>	.03
evaluatie van de leervorderingen via proefwerken en schriftelijke overhoringen	<u>.09</u>	<u>.09</u>
evaluatie van de leervorderingen via werkstukken	.10	.02
globale beoordeling van de schrijfprodukten door de docent	-.06	<u>.08¹</u>
Onbeslist		
lezen zakelijke brieven	.09	<u>-.09¹</u>
Ineffectief		
beschouwend opstel	-.11	-.01
handschriftverbeteren	-.02	<u>-.08</u>
evaluatie van de leervorderingen via extern toetsmateriaal	-.14	<u>-.02¹</u>

¹ het effect van de instructie is significant verschillend voor de sollicitatie- en de excuusbrief op 5%-niveau.

Tabel 4. Regressiegewichten voor het effect van de instructiekenmerken op de stelprestaties van jongens en meisjes (model 3b) en van zittenblijvers en niet-zittenblijvers (model 3c) (onderstreepte regressiegewichten zijn significant op 5%-niveau).

	Jongens	Meisjes	Doubleurs	Niet-doubleurs
Instructie * sekse				
doel- en publiekgerichtheid	<u>.16</u>	-.06 ¹		
Instructie * doubleren				
evaluatie via proefwerken			-.06	<u>.14¹</u>
evaluatie via werkstukken			-.05	<u>.12¹</u>

¹ het effect van de instructie is significant verschillend voor jongens en meisjes of voor zittenblijvers en niet-zittenblijvers op 5%-niveau

Instructie en leerlingkenmerken

Nagegaan is in hoeverre de instructie een differentieel effect heeft voor meisjes en voor jongens (model 3b) of voor zittenblijvers en niet-zittenblijvers (model 3c). De belangrijkste uitkomsten zijn weergegeven in Tabel 4.

Er zijn drie interacties tussen instructie- en leerlingkenmerken gevonden. Een daarvan is een interactie van doel- en publiekgerichtheid met de sekse van de leerling. Jongens profiteren wel van de oefeningen in het schrijven van doel- en publiekgerichte teksten, meisjes daarentegen niet. De overige twee zijn interacties tussen de meer algemene evaluatie van de leervorderingen voor het vak Nederlands en doubleren. Het evalueren van de leervorderingen met behulp van proefwerken en schriftelijke overhoringen heeft alleen een positief effect op de prestaties van de niet-zittenblijvers; de zittenblijvers profiteren daar niet van. Voor de andere vorm van interne evaluatie – beoordeling middels proefwerken en schriftelijke overhoringen – is er een vergelijkbaar effect. Alleen degenen die niet zijn blijven zitten hebben daar baat bij. Voor de overige instructiekenmerken kunnen we niet aantonen dat de zittenblijvers daar meer of juist minder van profiteren dan degenen die niet zijn blijven zitten.

Relatief belang instructie

In het voorgaande rapporteerden we over het effect van elk van de leerling-, school- en instructiekenmerken afzonderlijk. Nu stellen we de vraag naar de gezamenlijke invloed van de (in)effectieve instructiekenmerken, rekening houdend met de bijdrage van de school- en leerlingkenmerken. Deze vraag naar het relatieve belang van de instructie beantwoorden we aan de hand van de toename van het percentage verklaarde tussen-scholenvariantie van model 2 naar model 3a (zie Tabel 5).

Kijken we in Tabel 5 naar het gezamenlijk effect van de school- en leerlingkenmerken dan blijken deze een aanzienlijk deel van de prestatieverschillen tussen de MAVO-opleidingen te verklaren. Voor de sollicitatiebrief is dat overigens relatief gezien minder dan bij de excuusbrief het geval is (resp. 26% versus 55%).

Tabel 5. Percentage verklaarde variantie tussen scholen (R^2_{school}) ten gevolge van de school- en leerlingkenmerken (model 2) en de instructiekenmerken (model 3a) voor de sollicitatie- en de excuusbrief.

	Sollicitatiebrief	Excuusbrief
School en leerling (model 2)	26	55
School, leerling en instructie (model 3a)	69	100

De instructie heeft een aanzienlijk effect op de stelprestaties (boven hetgeen we met de school- en leerlingkenmerken kunnen verklaren). Blijkens Tabel 5 stijgt het percentage verklaarde tussen-scholenvariantie voor de sollicitatiebrief met 43% (van 26% naar 69%). Voor de excuusbrief bedraagt de overeenkomstige toename 45% (van 55% naar 100%). Significantietoetsing van de resterende tussen-scholenvarianties wijst uit dat deze alleen in het geval van de sollicitatiebrief significant van nul afwijken. Voor de excuusbrief zijn alle systematische verschillen tussen de MAVO-opleidingen verklaard, maar voor de sollicitatiebrief is dat niet het geval⁶.

3.4. Laag versus hoog presterende MAVO's

Eerder constateerden wij dat MAVO-leerlingen van LBO/MAVO-scholengemeenschappen gemiddeld lagere prestaties behaalden dan MAVO-leerlingen van categoriale MAVO's. Tevens zagen we dat de prestaties op de katholieke MAVO's lager waren dan op protestantse of openbare MAVO's. Voor deze lagere prestaties kunnen verklaringen worden gegeven in termen van een ongelijke instroom of in termen van een verschil in instructie-effectiviteit. Over een eventueel verschil in de 'kwaliteit' van de instroom kunnen we op grond van onze gegevens geen uitspraak doen. Wel kunnen we iets zeggen over de plausibiliteit van de tweede verklaring. Wellicht worden de effectieve instructiekenmerken op de LBO/MAVO-scholengemeenschappen en de katholieke MAVO's wat minder vaak toegepast, terwijl de 'neutrale' en ineffectieve kenmerken daar juist vaker in praktijk worden gebracht. Tabel 6 toont de effectgroottes voor de verschillen tussen de desbetreffende groepen scholen in de manier van lesgeven. In de eerste kolom is de effectgrootte van het verschil tussen de scholengemeenschappen en de categoriale MAVO's aangegeven en in de tweede kolom het verschil tussen de katholieke en de overige MAVO's.

In zijn algemeenheid lijken de effectieve instructiekenmerken op de LBO/MAVO-scholengemeenschappen en op de katholieke scholen in mindere mate in praktijk gebracht te worden (zie Tabel 6)⁷. Binnen de LBO/MAVO-scholengemeenschappen krijgen de MAVO-leerlingen dus op een andere, minder effectieve manier les dan hun leeftijdsgenoten van categoriale MAVO's. Hetzelfde geldt ook, zij het in iets mindere mate, voor leerlingen van katholieke MAVO's. Deze gegevens vormen een eerste indicatie voor een verschil in de kwaliteit van het stelonderwijs. De plausibiliteit van deze verklaring kan nader worden onderzocht door te kijken naar de gemiddelde prestaties van de groepen scholen in model 2 en model 3. Als de instructie inderdaad van doorslaggevende betekenis zou zijn, dan zou de achterstand van de LBO/MAVO-scholengemeenschappen na correctie voor het verschil in de manier van lesgeven moeten ver-

Tabel 6. Effectieve en ineffectieve instructiekenmerken in LBO/MAVO-SG en categoriale MAVO's en in katholieke scholen en de overige scholen (effectgroottes).

	LBO/MAVO-SG	RK
Effectieve instructiekenmerken		
brief aan officiële instanties en bedrijven	-.57	-.49
brief aan vrienden of familie	-.99	-.75
globale beoordeling van schrijfprodukten	-.47	-.32
evaluatie via schriftelijke werkstukken	-.71	-.21
evaluatie via proefwerken en schriftelijke overhoringen	.48	.14
doel- en publiekgerichtheid	.49	.57
Ineffectieve instructiekenmerken		
beschouwend opstel	.00	-.22
handschriftverbetering	.73	-.08
evaluatie via extern toetsmateriaal	.05	.03

Tabel 7. Verschil tussen de gemiddelde stelprestaties vóór (model 2) en na (model 3a) correctie voor verschillen in instructie-effectiviteit (effectgroottes).

	Model 2	Model 3a
Schoolstructuur		
cat. MAVO's – MAVO's in LBO/MAVO-SG	<u>.40</u>	<u>.28</u>
Denominatie		
katholiek – protestant	-.19	-.10
katholiek – openbaar	<u>-.25</u>	<u>-.21</u>

dwijnen of op zijn minst een flink stuk kleiner moeten worden. Hetzelfde geldt uiteraard voor het verschil tussen de gemiddelde leerlingprestaties van de katholieke en de overige denominaties. Dit kunnen we nagaan door de gemiddelde prestaties van de groepen scholen in model 2 te vergelijken met die in model 3. In model 3 zijn de gemiddelde prestaties immers gecorrigeerd voor het verschil in instructie-effectiviteit. Tabel 7 toont de effectgroottes van de prestatieverschillen tussen de groepen scholen vóór en na correctie voor de manier van lesgeven.

Zoals verwacht worden de prestatieverschillen tussen de groepen scholen na correctie voor de instructie inderdaad kleiner (zie Tabel 7). Dit betekent dat de achterstand van de LBO/MAVO-scholengemeenschappen en de katholieke MAVO's deels te maken heeft met de effectiviteit van het gegeven onderwijs.

Na correctie wordt de achterstand van de LBO/MAVO-scholengemeenschappen aanzienlijk kleiner (van .40 sd naar .28 sd). Maar ook nu is het verschil tussen de categoriale MAVO's en de LBO/MAVO-scholengemeenschappen nog steeds significant. We moeten derhalve concluderen dat de instructie geen afdoende verklaring vormt voor de achterstand van de LBO/MAVO-scholengemeenschappen. Kennelijk zijn er hier nog andere variabelen in het geding waarover wij overigens geen uitspraak kunnen doen.

De achterstand van de katholieke op de protestantse MAVO's wordt na correctie bijna gehalveerd (van .19 sd naar .10 sd). Het verschil tussen de gemiddelden is nu niet meer significant. Hier lijkt het prestatieverschil dus grotendeels op een verschil in instructie-effectiviteit te berusten.

Op de achterstand van de katholieke MAVO's op de openbare MAVO's is de uitgevoerde correctie nauwelijks van invloed (van .25 sd naar .21 sd). Ook hier spelen er kennelijk nog andere variabelen een rol.

4. SAMENVATTING EN DISCUSSIE

In zijn algemeenheid vestigen de uitkomsten van deze studie de aandacht op het belang van taakeffecten in onderzoek naar instructie- en schooleffectiviteit. MAVO-opleidingen blijken sterker van elkaar te verschillen in de gemiddelde prestaties op de sollicitatiebrief dan op de excuusbrief. Een plausibele verklaring is onzes inziens gelegen in het onderwijsaanbod. Aangenomen mag worden dat zakelijke teksten zoals een sollicitatiebrief op meer scholen geïnstrueerd en geoefend worden dan een excuusbrief. Verder blijkt het effect van de instructiekenmerken bij de ene taak vaak anders dan bij de andere. Ook de relatieve bijdrage van de school- en instructiekenmerken aan de verklaring van de prestatieverschillen tussen de MAVO-opleidingen verschilt van taak tot taak.

Van de 36 bestudeerde instructiekenmerken zijn er tien van invloed op de stelvaardigheid van de derdeklassers: zes positief, één onbeslist en drie negatief.

Drie van de positieve effecten hebben betrekking op het schrijven van functionele brieven.

Als leerlingen in de lessen Nederlands brieven aan instanties of bedrijven moeten schrijven komt dat de prestaties op de sollicitatie- en de excuusbrief ten goede. En als zij daarnaast brieven aan vrienden en familie schrijven, heeft dat een extra effect op de stelprestaties. En als zij ook nog les krijgen in het analyseren en schrijven van doel- en publiekgerichte teksten dan heeft dat weer een positief effect. Deze drie effecten lijken goed verklaarbaar. Alle drie hebben ze namelijk betrekking op functionele schrijfvaardigheid. Net als de sollicitatie- en de excuusbrief zijn brieven aan instanties, bedrijven, vrienden of familie voorbeelden van functionele schrijftaken waarbij zowel het doel als het publiek helder zijn. Predictor en criterium lijken in dit geval sterk op elkaar. Mogelijk is er sprake van een soort transfer binnen dezelfde vaardigheid en binnen hetzelfde taaktype.

De overige drie kenmerken van een effectieve instructie hebben betrekking op de evaluatie. Als docenten de leervorderingen vaker evalueren via proefwerken, schriftelijke overhoringen of werkstukken komt dat de stelprestaties ten goede. Deze bevinding sluit aan bij een van de best gedocumenteerde bevindingen van onderzoek naar effectieve instructie: frequente evaluatie is een van de meest succesvolle voorspellers van leerprestaties (o.a. Levine & Lezotte, 1990; Creemers, 1992; Veenman, Lem, Roelofs & Nijssen, 1992). Opmerkelijk is het positieve effect van de globale beoordeling van de schrijfprodukten door de docent. Uit twee belangrijke reviews van onderzoek naar effectieve steldidactieken blijkt dat de hoeveelheid en aard van de feedback van de docent maar zelden voor de stelprestaties van belang is (Wesdorp, 1983; Hillocks, 1986).

Bij deze bevindingen is een relativiserende kanttekening te maken. De opzet van de studie laten een causale interpretatie van de gevonden relaties niet toe. Interpretatie van de richting van een effect berust geheel en al op additionele argumentatie waarvan de plausibiliteit ter discussie staat. We lichten dit toe aan de hand van de ineffectieve instructiekenmerken. Uit de analyse komt naar voren dat het (algebraïsche) teken van de regressiegewichten voor het effect van het schrijven van beschouwende opstellen en oefening in handschriftverbetering op de stelprestaties negatief is. Causaal geïnterpreteerd zou dit betekenen dat het schrijven van opstellen en oefening in handschriftverbetering lagere stelprestaties tot gevolg zouden hebben. Een dergelijke oorzaak-gevolgrelatie lijkt ons moeilijk voorstelbaar. Wellicht is er iets anders aan de hand. Het is bekend dat de onderwijstijd beperkt is, zeker voor schrijven als een van de vele onderdelen van het vak Nederlands. Mogelijk besteden de docenten die veel nadruk leggen op het schrijven van opstellen en het verbeteren van het handschrift maar weinig aandacht aan het type taken dat voor onze effectmaten van belang is, met als gevolg lagere prestaties op de sollicitatie- of excuusbrief. Een vergelijkbaar interpretatieprobleem doet zich voor bij het negatieve effect van evaluatie met behulp van toetsen die de docent niet zelf gemaakt heeft maar die extern ontwikkeld zijn. Ook hier is het teken van het regressiegewicht negatief. Dit zou betekenen dat de stelprestaties afnemen naarmate er meer met extern toetsmateriaal gewerkt wordt. Deze verklaring staat evenwel op gespannen voet met de eerder gerapporteerde positieve effecten van zelf ontwikkelde toetsen. Een meer voor de hand liggende interpretatie is dat het vooral de docenten van de slecht presterende leerlingen zijn die hun heil zoeken in extern toetsmateriaal.

De effectiviteit van de instructiekenmerken blijkt niet alleen afhankelijk van de taak in kwestie, maar ook van het type leerling. De sekse van de leerling is van belang in de zin dat alleen jongens wat opsteken van oefeningen in het schrijven van doel- en publiekgerichte teksten; meisjes lijken daar geen baat bij te hebben. Tevens profiteren alleen degenen die niet zijn blijven zitten van een frequente evaluatie van de leervorderingen via proefwerken, schriftelijke overhoringen en werkstukken. Bij gebrek aan theorie is een verklaring moeilijk te geven. Waarom hebben meisjes geen baat bij oefeningen in het analyseren en schrijven van doel- en publiekgerichte teksten? En waarom heeft evaluatie geen effect op de prestaties van de zittendblijvers? Mogelijk dat het geringe onderscheidingsvermogen ten gevolge van het betrekkelijk kleine aantal waarnemingen ons hier parten speelt.

Voor het merendeel van de bestudeerde instructiekenmerken kan er geen (additief) effect op de prestaties worden aangetoond. Dit geldt voor vrijwel alle deeloefeningen op het terrein van idioom, interpunctie, spelling en grammatica die in het schrijfonderwijs zo populair zijn. Gezien

de keuze van de effectmaat – globale kwaliteitsoordelen over functionele stelvaardigheid – is deze uitkomst wellicht niet zo verrassend. Weliswaar hebben de beoordelaars dit soort aspecten in hun oordelen over de kwaliteit van de brieven mee laten wegen, met name bij globale kwaliteit en stijl. Het is echter de vraag in hoeverre het effect van deze deelvaardigheidsoefeningen van invloed is op het schrijven van een *volledige* tekst. Er spelen dan niet alleen geïsoleerde deelvaardigheden een rol, maar het gaat dan ook om de coördinatie van al deze deelvaardigheden tegelijkertijd (Flower & Hayes, 1980; Broekkamp & Van den Bergh, in druk). En met name deze coördinatie lijkt een karakteristieke moeilijkheid van stellen. Zo citeren Bereiter & Scardamalia (1987) een zestienjarige schrijver: 'I have all my thoughts in my mind, but when I come to a word I can't spell it throws me off my writing' (p. 154).

Ook van schrijfactiviteiten die volgens recente reviews (Wesdorp, 1983; Hillocks, 1986) wel een relatie met de prestaties zouden kunnen vertonen kan geen positief effect worden vastgesteld. We noemen hier: oefeningen in het voorbereiden en plannen van schrijfprodukten, oefeningen in het verzamelen, selecteren en gebruiken van materiaal, oefeningen in het reviseren van teksten en oefeningen waarbij leerlingen commentaar geven op elkaars schrijfprodukten.

Zoals gezegd hangt een groot deel van de instructiekenmerken niet samen met de stelprestaties. En van de instructiekenmerken die dat wel doen, is het effect nogal eens klein. Een mogelijke verklaring ligt wellicht in het gegeven dat de praktijk van het stelonderwijs gemeten is met een vragenlijst waarvan de validiteit aan twijfel onderhevig is. Indicatief is dat wij niet in staat waren om de veelvormige lespraktijk met een beperkt aantal clusters ofwel steldidactieken te beschrijven. Zoals Damhuis, De Glopper & Wesdorp (1983, p. 232) terecht opmerken, hoeft dit niet per se aan de vragenlijst te liggen. Wellicht zijn er in het huidige stelonderwijs inderdaad geen 'duidelijk afgetekende dimensies of stromingen' te onderscheiden.

Een andere verklaring verwijst naar de geringe frequentie van de onderscheiden instructiekenmerken in de vigerende onderwijspraktijk (vgl. De Glopper, 1987). Met de meeste schrijfactiviteiten komt de doorsnee derdeklassers op school hooguit één keer per maand in aanraking (zie ook Bijlage 2). Mogelijk is dit te weinig om veel effect te kunnen sorteren. Dit betekent uiteraard niet dat deze kenmerken, gegeven voldoende instructie, niet effectief zouden kunnen zijn. We veronderstellen alleen dat de presentie in de huidige onderwijspraktijk onvoldoende groot is om een effect te kunnen aantonen.

Op LBO/MAVO-scholengemeenschappen schrijven MAVO-leerlingen kwalitatief gezien minder goede sollicitatie- en excuusbrieven dan op categoriale MAVO's. Evenzo presteren leerlingen op katholieke MAVO's lager dan op protestantse en openbare MAVO's. Voor deze relatief lage stelprestaties zijn verschillende verklaringen te geven. Een eerste verklaring verwijst naar een verschil in instroom. Zo wezen we reeds op de mogelijkheid dat MAVO-leerlingen binnen scholengemeenschappen met LBO aan het einde van de basisschool wat vaker een LBO-advies hebben gekregen dan hun leeftijdsgenoten op categoriale MAVO's (vgl. De Jong & Roeleveld, 1989). Omdat we niet beschikken over instroomgegevens, kunnen we hier geen uitspraak over doen. Een tweede verklaring verwijst naar een verschil in instructie-effectiviteit. Op de LBO/MAVO-scholengemeenschappen en de katholieke MAVO's zou de instructie wel eens van een wat mindere kwaliteit kunnen zijn, minder effectief met het oog op de functionele taken die in deze studie centraal staan. Geconcludeerd wordt dat deze achterstand inderdaad goeddeels is toe te schrijven aan de wijze waarop het stelonderwijs gegeven wordt. Met name worden de effectieve instructiekenmerken op de LBO/MAVO-scholengemeenschappen en de katholieke MAVO's minder vaak in praktijk gebracht.

Meisjes zijn gemiddeld genomen beter in stellen dan jongens, een bevinding waarover elders uitvoeriger gerapporteerd is (Kuhlemeier & Van den Bergh, 1991a). Veel interessanter is echter de interactie met de denominatie van de school. De kloof tussen meisjes en jongens blijkt niet onafhankelijk van de richting van de school. In de katholieke MAVO's, waar de leerlingen de laagste stelprestaties behaalden, is de voorsprong van meisjes niet significant. In de protestantse MAVO's en openbare MAVO's is het verschil in het voordeel van meisjes wel significant. Dit verschijnsel kunnen we op ten minste twee manieren interpreteren. Een eerste interpretatie stelt dat het de katholieke MAVO's kennelijk gelukt is om de kloof tussen meisjes en jongens te

dichten. De andere interpretatie is wellicht wat realistischer en verwijst naar het zogeheten Mattheus-effect. Als er ergens minder goed onderwijs gegeven wordt, is dat even ineffectief voor alle typen leerlingen, dus niet alleen voor de betere stellers (meisjes) maar ook voor de wat minder goede stellers (jongens). Minder effectief onderwijs zou met andere woorden een nivellerend effect hebben op de prestatieverschillen tussen leerlingen. Bij effectief onderwijs zouden de prestaties daarentegen uiteenlopen in de zin dat de betere leerlingen naar verhouding meer opsteken dan de minder goede leerlingen.

De grootte van de school, de vestigingsregio en de fusiehistorie blijken voor de stelprestaties van geen belang. Grote MAVO's doen het niet beter of slechter dan kleine, in dichtbevolkte gebieden zijn de prestaties niet lager dan in dunbevolkte gebieden en het zijn zeker niet de slechtste MAVO's die in de schaalvergrotingsoperatie het eerst het loodje legden. De bevindingen ten aanzien van schoolgrootte corresponderen met die van Luyten (1993) en Kleintjes en Kremers (1992), maar anders dan in laatstgenoemd onderzoek blijkt de bevolkingsdichtheid van geen belang voor de leerprestaties. Ook de door ons gesignaleerde 'achterstand' van de katholieke MAVO's staat op gespannen voet met ander onderzoek waarin geen of hooguit kleine verschillen tussen denominaties gevonden zijn (Dijkstra, 1990). En voorzover er al verschillen geconstateerd zijn, pakken die uit in het voordeel van de katholieke sector (Van Laarhoven, Bakker, Dronkers & Schijf, 1987; Van Laarhoven & Dronkers, 1988; De Jong & Roeleveld, 1989) of in het voordeel van de protestantse sector (Kleintjes & Kremers, 1992). Voor de vakken wiskunde, Engels en biologie constateren laatstgenoemde auteurs een voorsprong van de categoriale MAVO's in vergelijking met LBO/MAVO-scholengemeenschappen. Anders dan in ons onderzoek, bleek dit verschil bij het vak Nederlands echter afwezig. Over de generaliseerbaarheid van de door ons gevonden prestatieverschillen qua denominatie en schoolstructuur naar de populatie van MAVO-opleidingen in Nederland mag men dan ook geen al te grote illusies hebben (zie ook paragraaf 2.1). De bevinding dat de achterstand van de katholieke MAVO's en de LBO/MAVO's goeddeels is toe te schrijven aan de wijze waarop het stelonderwijs op deze scholen gegeven wordt, wijst eveneens in deze richting.

NOTEN

1. De Glopper (1990) construeerde één samenvattende maat voor stelvaardigheid: het gemiddelde van oordelen over globale kwaliteit, inhoud, stijl en opbouw voor drie verschillende schrijftaken. Merk op dat in deze definitie de meer technische aspecten van het schrijven (o.a. spelling, interpunctie, grammaticale en idiomatische correctheid) niet tot de stelvaardigheid gerekend worden.
2. De stelprestaties van MAVO-leerlingen binnen scholengemeenschappen met HAVO en/of VWO bleken niet significant te verschillen van die van de MAVO-leerlingen van categoriale MAVO's. Om deze reden zijn deze twee groepen in de analyse samengevoegd.
3. De gemiddelde prestaties van de openbare en de 'overige' MAVO's bleken niet significant van elkaar te verschillen. Deze twee categorieën zijn in de analyse dan ook samengenomen.
4. De betrouwbaarheid van de factorgemiddelden op schoolniveau (ρ_S) is berekend op grond van de volgende formule (Rowan, Raudenbush & Kang, 1991):

$$\rho_S = \sigma_S^2 / [\sigma_S^2 + (\sigma_L^2 / N_i) + (\sigma_R^2 / N_i)] \quad (1)$$

waarbij:

- σ_S^2 : de tussen-scholenvariantie;
- N_i : het aantal leerlingen per school;
- σ_L^2 : de tussen-leerlingenvariantie;
- N_i : het aantal indicatoren;
- σ_R^2 : de residuele variantie.

De betrouwbaarheid van de factorgemiddelden op leerlingniveau (ρ_L) berekenen we als volgt:

$$\rho_L = \sigma_L^2 / [\sigma_L^2 + (\sigma_R^2 / N_i)] \quad (2)$$

5. Het percentage verklaarde variantie (R^2) is het kwadraat van de correlatie tussen de geobserveerde scores en de op grond van het model voorspelde scores.

6. Een relativerende kanttekening is hier op zijn plaats. Het additieve effect van de instructie is geëvalueerd als de toename van de verklaarde tussen-scholenvariantie na uitzuivering van het effect van de school- en leerlingkenmerken. Hoewel theoretisch juist, geeft deze procedure misschien een te optimistisch beeld van het effect van de instructie in de onderwijspraktijk. We hadden dit effect namelijk ook kunnen evalueren aan de hand van de veel kleinere toename van de verklaarde *totale* variantie (d.w.z. op school-, leerling- en scoreniveau gezamenlijk). Aangezien de instructie alleen van invloed is op de tussen-scholenvarianties (aptitude-treatment-interacties daargelaten) kan deze slechts een beperkt deel van de totale variantie in de stelprestaties naar zich toe trekken [de tussen-scholenvariantie bedraagt immers slechts 15% (sollicitatiebrief) respectievelijk 6% (excuusbrief) van de totale variantie (zie Tabel 1)].
7. Opvallend is tevens dat de LBO/MAVO-scholengemeenschappen veel meer aandacht schenken aan deeloefeningen en schrijfactiviteiten die geen aantoonbaar effect op de stelprestaties hebben dan de categoriale MAVO's (de zogeheten neutrale instructiekenmerken). Het betreft onder meer oefeningen in spelling en interpunctie (1.38 sd), aandacht voor 'moeilijke' woorden en het uitbreiden van de woordenschat (.52 sd) en oefening in het schrijven van korte instructies (1.37 sd).

LITERATUUR

- Bergh, H. van den & Kuhlemeier, J.B. (1990). De haalbaarheid van eindtermen Nederlands voor de basisvorming. *Pedagogische Studiën*, 67, 1, 1-15.
- Bergh, H. van den & Kuhlemeier, J.B. (1991). Relaties tussen aspecten van het onderwijsaanbod en tijdsbesteding bij het vak Nederlands in de derde klas voortgezet onderwijs. *Spiegel*, 9, 1, 55-68.
- Bereiter, C. & Scardamalia, M. (1987). *The psychology of written composition*. Hillsdale, N.J.: Lawrence Erlbaum.
- Besluit kerndoelen en adviesurentabel basisvorming 1993-1998*. Zoetermeer: Ministerie van Onderwijs en Wetenschappen.
- Blank, J.T.L., Boef-Van der Meulen, S., Bronneman-Helmers, H.M., Herweijer, L.J., Kuhry, B. & Scheurs, R.A.H. (1990). *School en schaal*. Tilburg: Sociaal Cultureel Planbureau.
- Bosker, R.J. (1991). De consistentie van schooleffecten in het basisonderwijs. *Tijdschrift voor Onderwijsresearch*, 16, 4, 206-218.
- Broekkamp, H. & Bergh, H. van den (in druk). Aandachtsstrategieën bij het reviseren in een vreemde taal: de invloed van de taakomgeving. *Tijdschrift voor Onderwijsresearch*.
- Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical linear models: application and data analysis methods*. Newsbury Park: Sage Publications.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Creemers, B.P.M. (1992). *Effectieve instructie: een empirische bijdrage aan de verbetering van het onderwijs in de klas*. 's-Gravenhage: Instituut voor Onderzoek van het Onderwijs.
- Damhuis, R., Gloppe, K. de, & Wesdorp, H. (1983). *Het opstelonderwijs. Stelvaardigheid in het voortgezet onderwijs: theorie en praktijk*. Amsterdam: Stichting Centrum voor Onderwijsonderzoek.
- Dijkstra, A.B. (1990). Doet Godsdienst leren? Schoolloopbanen in de gereformeerde sector van het Nederlandse onderwijs. *Tijdschrift voor Onderwijsresearch*, 15, 1, 67-72.
- Flower, L.S. & Hayes, J.R. (1980). The dynamics of composing: making plans and juggling with constraints. In: L.W. Gregg & E.R. Steinberg (Eds.). *Cognitive processes in writing* (p. 31-50). Hillsdale, N.J.: Lawrence Erlbaum.
- Gloppe, K.D. (1987). *Schrijven beschreven. Inhoud, opbrengsten en achtergronden van het schrijfonderwijs in de eerste vier leerjaren van het voortgezet onderwijs*. 's-Gravenhage: SVO.
- Gloppe, K.D. (1990). Schrijfvaardigheid en schooleffectiviteit. *Spiegel*, 8, 3, 71-88.
- Goldstein, H. (1987). *Multilevel models in educational and social research*. London: Griffin.
- Jong, U. de & Roeleveld, J. (1989). Public and private secondary schools in Amsterdam. In: B.F.M. Bakker, J. Dronkers & G.W. Meijnen (Eds.). *Educational opportunities in the welfare state. Longitudinal studies in educational and occupational attainment in the Netherlands* (p. 151-166). Nijmegen: ITS.
- Kleintjes, F.G.M. & Kremers, E.J.J. (1992). *School size, characteristics and outcomes in Dutch secondary education*. Paper presented at the ECER Conference, Enschede, June 22-25 (ongepubliceerd paper).
- Kuhlemeier, J.B. & Bergh, H. van den (1989a). *De Proefpeiling Nederlands: een onderzoek naar de haalbaarheid van peilingsonderzoek in het voortgezet onderwijs*. Specialistisch Bulletin Nr. 74. Arnhem: Cito.

- Kuhlemeier, J.B. & Bergh, H. van den (1989b). *De Proefpeiling Nederlands: een onderzoek naar de haalbaarheid van peilingsonderzoek in het voortgezet onderwijs*. Bijlagen bij Specialistisch Bulletin Nr. 74. Arnhem: Cito.
- Kuhlemeier, J.B. & Berg, H. van den (1991a). Sekseverschillen in het voortgezet onderwijs: taalprestaties, taalattituden en taalactiviteiten op school en in de vrije tijd. *Pedagogische Studiën*, 68, 3, 101-113.
- Kuhlemeier, J.B. & Bergh, H. van den (1991b). De correlatieve structuur van taalvaardigheid: een exploratie. *Tijdschrift voor Onderwijsresearch*, 16, 3, 143-159.
- Kuhlemeier, J.B. & Bergh, H. van den (in voorbereiding). *De structuur van schrijfvaardigheid: een toepassing van multilevel factoranalyse*.
- Kuijk, J. van & Kessel, N. van (1987). *Ophef over opheffing of fusie*. Nijmegen: ITS.
- Laarhoven, P. van, Bakker, B., Dronkers, J. & Schijf, H. (1987). Richting van de school en schoolloopbanen in het voortgezet onderwijs. *Tijdschrift voor Onderwijsresearch*, 12, 1, 23-40.
- Laarhoven, P. van & Dronkers, J. (1988). Verschillen de onderwijsresultaten in openbaar en bijzonder onderwijs? In: A.M.L. van Wieringen, S. Karsten & J. Ax (Red.). *Gemeenten en onderwijs* (p. 75-88). Lisse: Swets & Zeitlinger.
- Levine, D.U. & Lezotte, L.W. (1990). *Unusually effective schools. A review and analysis of research and practice*. Madison: National Center for Effective Schools Research and Development.
- Luyten, L.W. (1992). Schoolgrootte en leerlingprestaties. In: R. Bosker (Red.). *Schoolgrootte, effectiviteit en de basisvorming* (p. 21-42). Twente: OCTO.
- Meuffels, B. (1994). *De verguisde beoordelaar: opstellen over opstelbeoordeling*. Amsterdam: Thesis Publisher (in druk).
- Moors, J.J.A. & Muilwijk, J. van (1975). *Steekproeven: een inleiding tot de praktijk*. Amsterdam: Agon Elsevier.
- Muthén, B.O. (1992). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, 28, 338-354.
- Pouwels, J., Esch, J. van & Vrieze, G. (1990). *Scholen en fusies. Een onderzoek bij schoolleiders en docenten naar de gevolgen van fusies in het voortgezet onderwijs*. Lisse: Swets & Zeitlinger.
- Prosser, R., Rasbash, J., & Goldstein, H. (1991). *Software for three-level analysis. Users' guide for V.2*. London: Institute of Education.
- Rowan, B., Raudenbush, S.W., & Kang, S.J. (1991). School climate in secondary schools. In: S.W. Raudenbush & J.D. Willms (Eds.). *Schools, classrooms, and pupils. International studies of schooling from a multilevel perspective* (pp. 203-223). San Diego: Academic Press.
- Veenman, S., Lem, P., Roelofs, E., & Nijssen, J. (1992). *Effectieve instructie en doelmatig klasmanagement*. Amsterdam: Swets & Zeitlinger.
- Wesdorp, H. (1981). *Evaluatietechnieken voor het moedertaalonderwijs*. Den Haag: SVO.
- Willms, J.D., & Raudenbush, S.W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement*, 26, 3, 209-232.

Manuscript ontvangen 28-2-1994

Derfinitieve versie ontvangen 21-4-1994

BIJLAGE 1

Enkele beschrijvende statistische gegevens voor de school- en leerlingkenmerken

In deze bijlage wordt nadere informatie versterkt over de verdeling van de school- en leerlingkenmerken in de respons en, voorzover daarover gegevens voorhanden zijn, in de populatie (ten tijde van het onderzoek).

Schoolstructuur. De scholenrespons (N = 43) bestaat uit dertig categoriale MAVO's (70%), zes scholengemeenschappen voor MAVO met een HAVO-en/of VWO-opleiding (14%) en zeven scholengemeenschappen voor LBO/MAVO (16%). Gegeven de beperkte omvang van de steekproef is de verdeling naar schoolstructuur redelijk representatief ten opzichte van de populatie. De verdeling in de populatie van MAVO-opleidingen in Nederland is: 61% categoriale MAVO's, 25% binnen scholengemeenschappen van MAVO met HAVO en/of VWO en 14% binnen scholengemeenschappen van MAVO met LBO (zie ook Blank e.a., 1990, p. 165).

Denominatie. De verdeling van de respons naar denominatie is als volgt: zeventien (40%) katholieke opleidingen, vijftien (35%) protestantse en elf (26%) openbare en overige opleidingen. De verdeling in de responsgroep komt redelijk overeen met die in de populatie: 33% katholiek, 29% protestants en 38% openbaar (zie ook Blank e.a., 1990, p. 165).

Regionale spreiding. De verdeling naar regio is als volgt: vijf uit het Noorden (12%), vijftien uit het Oosten (35%), vijf uit het Midden (12%), twaalf uit het Westen (28%) en zes uit het Zuiden (14%) van Nederland.

Opgeheven dan wel gefuseerd. Van de 43 onderzoeksscholen ten tijde van de dataverzameling in het schooljaar 1987-1988 waren er aan het begin van het schooljaar 1991-1992 negen (21%) opgeheven dan wel opgegaan in een fusie met een andere school.

Schoolgrootte en bevolkingsdichtheid. De gegevens over het aantal derdeklassers, het totaal aantal leerlingen op de school (exclusief IBO-leerlingen) en de bevolkingsdichtheid in de responsgroep en de populatie van MAVO-opleidingen in Nederland zijn hieronder weergegeven.

	gemiddelde	standaard- deviatie	laagste waarde	hoogste waarde	N
Responsgroep					
aantal derdeklassers in het derde leerjaar van de MAVO-opleiding	86	44	27	224	43
aantal leerlingen op de gehele school	421	304	57	1595	43
bevolkingsdichtheid nodaal gebied	724	692	149	2546	43
Populatie					
aantal derdeklassers in het derde leerjaar van de MAVO-opleiding	76	35	7	224	937
aantal leerlingen op de gehele school	553	418	45	2284	937
bevolkingsdichtheid nodaal gebied	803	677	88	2546	937

Sekse en doubleren. De verdeling van de 343 MAVO-derdeklassers naar sekse is: 51% meisjes en 49% jongens; 32% was in het voortgezet onderwijs al eens blijven zitten.

BIJLAGE 2

Enkele beschrijvende statistische gegevens voor de instructiekenmerken

Instructiekenmerken	gemiddelde	standaard-deviatie	laagste score	hoogste score	N
Vormen van schrijven ¹					
creatief schrijven	2.46	.69	1	3	37
zakelijk schrijven	2.57	.65	1	4	35
Deeloefeningen van schrijven ¹					
handschriftverbetering	1.21	.47	1	3	38
spelling en interpunctie	3.10	.79	1	4	39
woordenschat/moeilijke woorden	3.26	.91	1	5	39
syntactische variatie/zinsbouw	2.82	.85	1	5	39
structureren	2.65	.59	1	4	37
verzamelen/selecteren materiaal	2.03	.64	1	3	37
plannen schrijfprodukt	2.21	.78	1	4	33
doelgerichtheid	2.25	.72	1	4	32
publiekgerichtheid	2.03	.85	1	4	33
revisie eigen schrijfprodukt	2.00	.75	1	4	33
Instructie van teksttypen ¹					
brief aan vrienden of familie	1.83	.38	1	2	35
brief aan instantie of bedrijf	2.00	.49	1	3	35
beschouwend opstel	2.00	.53	1	3	36
zelfverzonnen verhaal	2.23	.60	1	3	35
samenvatting	2.15	.66	1	3	34
beschrijving iets concreets	1.94	.65	1	3	34
korte instructie	1.77	.55	1	3	35
korte overredende tekst	1.70	.53	1	3	33
verslag	1.94	.68	1	3	35
aantekeningen	2.25	.94	1	5	36
Evaluatie van schrijfprodukten ²					
via globale beoordeling	2.97	1.14	1	5	37
via analytische beoordeling	3.03	1.17	1	5	37
via voorbeeld-schrijfprodukten	1.66	.83	1	4	32
via objectieve schrijftoetsen	1.50	.72	1	3	32
via 'peer-review'	3.31	1.36	1	5	39
Evaluatiecriteria schrijfprodukten ³					
organisatie	4.44	.77	2	5	36
communicatie	3.89	1.02	1	5	37
stijl	4.33	.66	3	5	39
inhoud	4.49	.60	3	5	39
Lezen van zakelijke brieven ¹	2.11	.66	1	4	37
Evaluatie met extern toetsmateriaal ¹					
externe toetsen/tests (Cito)	1.71	.83	1	3	35
objectieve toetsen bij leertaak	2.32	.91	1	3	37
Evaluatie met intern toetsmateriaal ¹					
proefwerken en overhoringen	3.35	.59	2	4	37
schriftelijke werkstukken	1.97	.82	1	3	35

¹ alternatieven: 1 = nooit, 2 = jaarlijks; 3 = maandelijks; 4 = wekelijks; 5 = vrijwel elke les;

² alternatieven: 1 = nooit, 2 = soms; 3 = vaak; 4 = meestal; 5 = bijna altijd;

³ alternatieven: 1 = volstrekt niet; 2 = niet of nauwelijks; 3 = soms wel, soms niet; 4 = in lichte mate; 5 = in sterke mate.

Boekbesprekingen

J.W. Holleman (1993). *Over studielast en studeerbaarheid*
Proefschrift R.U. Utrecht. Utrecht: Uitgeverij Lemma (ISBN 90-5189-284-5).

In dit omvangrijke proefschrift (268 pagina's) wordt verslag gedaan van een onderwijstechnologisch onderzoek gericht op interpretatie en operationalisering van de wetsvoorschriften over studielast, cursusduur en studeerbaarheid in het Nederlandse hoger onderwijs. De auteur geeft aan dat het doel van zijn onderzoek tweeledig is. Het eerste doel is de wettelijke bepalingen over studielast en cursusduur interpreteren ten behoeve van faculteiten. Met andere woorden aan welke eisen moeten universitaire curricula sinds de invoering van de Wet Tweefasenstructuur (1982) voldoen. Het tweede doel is hulpmiddelen te ontwikkelen om de probleemsituatie het hoofd te bieden die door de wettelijke bepalingen is ontstaan. Kortom, op welke wijze kan een probleemsituatie in kaart gebracht worden en hoe kan een faculteit haar studieprogramma's beschrijven, optimaliseren en evalueren als zij wil zorgen dat deze zo goed mogelijk aan de wettelijke bepalingen voldoen.

Deze twee doelen zijn geoperationaliseerd in vijf onderzoeksvragen, die in de diverse hoofdstukken uitgewerkt worden. De eerste twee onderzoeksvragen hebben betrekking op de interpretatie van de beoogde toestand van studieprogramma's en de probleemverheldering. De beoogde toestand wordt uitgewerkt door aandacht te besteden aan de eisen waaraan universitaire studieprogramma's krachtens de wettelijke bepalingen inzake studielast en cursusduur moeten beantwoorden. Tevens wordt nagegaan of deze beoogde toestand toetsbaar, bestaanbaar en haalbaar is voor elk studierichting. Om hierop antwoord te kunnen geven wordt aandacht besteed aan de relatie tussen studiesucces en cursusduur (hoofdstuk 4). Vervolgens wordt een model ontwikkeld voor het beschrijven van criteria en aspecten van studeerbaarheid, dat vervolgens getoetst wordt (hoofdstuk 8 en 9). Voorts zijn de wettelijke bepalingen inzake de studielast geïnterpreteerd en worden tegenstrijdigheden in wet- en regelgeving gesignaleerd (hoofdstuk 10 en 11).

De derde onderzoeksvraagstelling heeft betrekking op de voorschriften waaraan een programma-beschrijving van een instelling moet voldoen. Hiertoe werd een model ontwikkeld voor het formuleren van opleidingsdoelen (hoofdstuk 2). Daar opleidingsdoelen slechts een aspect van het studieprogramma zijn, wordt het model verder aangevuld met een model ter beschrijving van studieprogramma's (hoofdstuk 3).

In de vierde onderzoeksvraagstelling staan de voorschriften voor programma-ontwikkeling centraal. Aandacht wordt besteed aan de factoren die een faculteit kan manipuleren en met welke zij rekening moet houden om een studieprogramma dusdanig in te vullen dat het aan de gestelde eisen beantwoordt. Daartoe is een model ontworpen voor het bevorderen van studiesucces voor een bepaalde student in een gegeven studieduur (hoofdstuk 5). Dit model wordt vervolgens getoetst (hoofdstuk 6 en 7).

De vijfde onderzoeksvraagstelling heeft betrekking op de voorschriften voor programma-evaluatie. Nagegaan werd aan de hand van welke operationele criteria en met welke operationele methoden de faculteit kan evalueren of een studieprogramma aan die beoogde toestand voldoet. Hiertoe zijn voorschriften opgesteld om te evalueren of de beoogde toestand bereikt is. De wettelijke bepalingen worden geoperationaliseerd in evaluatiecriteria en -procedures om te bepalen of een studieprogramma aan de bedoelingen van de wetgever voldoet. Hierbij werden ook de criteria voor numeriek opleidingsrendement en doorstroming betrokken, die door de faculteit zelf moeten worden opgesteld (hoofdstuk 12).

Afgesloten wordt met een concluderend overzicht van de antwoorden op de onderzoeksvraagstellingen (hoofdstuk 13). Tevens resumeert dit hoofdstuk problemen die aan de wettelijke voorschriften kleven, waarbij ook enige beleidsalternatieven worden voorgesteld om deze problemen te ondervangen.

Terugblikkend naar de doelen van het onderzoek kan gesteld worden dat Holleman de gestelde doelstellingen heeft bereikt. Ten aanzien van het eerste doel worden alle regelingen en mogelijke interpretaties gedegen op een rijtje gezet en krijgt de lezer inzicht in de "wirwar" van wet- en regelgeving. Verhelderend is de wijze waarop de geschiedenis van de wet- en regelgeving beschreven wordt en de inconsistenties die Holleman daarin signaleert.

Ten aanzien van de tweede doelstelling is Holleman er in geslaagd een aantal zeer bruikbare modellen te ontwikkelen die binnen de instellingen gehanteerd kunnen worden om de studeerbaarheid nader onder de loep te nemen. Het zou verhelderend zijn geweest als een volledig uitgewerkt voorbeeld zou zijn opgenomen waarbij ingegaan werd op de knelpunten die de praktijk ondervindt met deze modellen en waarbij aangetoond wordt dat het modelmatig werken een "meerwaarde" heeft die ook door de practicus ervaren wordt.

Holleman brengt duidelijk tot uiting dat studeerbaarheid een complexe materie is, waarbij een rationale benadering van belang is. Uit de complexiteit van dit proefschrift blijkt al dat het verbeteren van studeerbaarheid niet gemakkelijk is en dat hier tal van haken en ogen aan kleven. Alles hangt met alles samen, zaken interfereren en zijn wel te onderscheiden maar niet te scheiden. Dit is er waarschijnlijk mede de oorzaak van dat het proefschrift niet makkelijk leesbaar is en dat het moeilijk is om de rode draad vast te houden. De auteur heeft getracht dit op te lossen door gebruik te maken van (te) veel schema's, tabellen, samenvattingen en verwijzingen. Ook wordt te vaak volstaan met het uitsluitend opnemen van verwijzingen naar reeds verrichte vooronderzoeken, zonder dat beknopt de kern van deze onderzoeken wordt beschreven. De toevoeging van de gehanteerde begrippenlijst als appendix is zeer handig en functioneel.

Gezien de actualiteit van deze problematiek, de veranderde wetgeving en de eisen die heden ten dage aan instellingen gesteld worden is het van belang dat ook de practicus op de hoogte is van de inhoud van dit boek. Ik denk dat deze groep geen baat heeft bij het uitsluitend lezen van de kernhoofdstukken (de auteur heeft in de leeswijzer aangegeven wat de kernhoofdstukken zijn). Aan te bevelen is om op basis van dit onderzoek een bruikbare handleiding te schrijven voor de onderwijsinstellingen die de studeerbaarheid van hun programma's willen verbeteren. Voor degenen die geïnteresseerd zijn in de onderliggende theorieën van de modellen en de gehele procesgang biedt dit proefschrift aanvullende relevante achtergrondinformatie.

Ineke Wolfhagen
Vakgroep Onderwijsontwikkeling en -research
Rijksuniversiteit Limburg

M. Boekaerts & P.R.J. Simons, *Leren en instructie: psychologie van de leerling en het leerproces*. Assen: Dekker & Van de Vegt, 1993.

W. Tomic & P. Span (redactie), *Onderwijspsychologie: beïnvloeding, verloop en resultaten van leerprocessen*. Utrecht/Heerlen: Lemma/Open Universiteit, 1993.

Het bespreken van studieboeken is moeilijk, immers zij hebben als doel de lezer (lees: de leerling, de student) in te leiden in het onderwerp. De auteurs van studieboeken zijn verplicht om keuzen te maken en de keuzen die gemaakt worden zijn afhankelijk van de auteurs. In een studieboek leidt een auteur de lezer in in een gebied, waarop hij deskundig is, op de manier zoals hij het gebied ziet. Hoewel iedere auteur er vanuit zijn visie op het vakgebied tegen aan kijkt, zijn er tussen auteurs, uiteraard, grote overeenkomsten. Vanuit deze vaststelling wordt er naar de beide studieboeken gekeken.

Het inhoudelijke en redactionele concept

Naar het concept van de beide boeken kan op twee manieren gekeken worden: het inhoudelijk concept en het redactionele concept.

Het inhoudelijk concept van de beide boeken verschilt: bij Boekaerts & Simons (verder aangeduid met B&S) is er gekozen voor een procesmatige opvatting van de onderwijspsychologie, terwijl Tomic & Span (verder aangeduid met T&S) kozen voor een produktmatige opvatting. Bij B&S neemt het (*onderwijs*)leerproces en alles, dat hierbij een rol speelt, een centrale rol in. Centrale delen in het boek zijn dan ook deel II ("aandacht voor het leerproces") en deel III ("aandacht gericht op gedragscontrole"). Bij T&S nemen de *resultaten van het leerproces* de centrale positie in. Het centrale concept in het boek, hoewel pas expliciet aan de orde in deel IV ("leerresultaten van leerlingen"), is "transfer".

Aan het verschil in inhoudelijk concept liggen twee verschillende "scholen" onderwijspsychologie ten grondslag. Aan het concept van B&S ligt een Tilburgse opvatting ten grondslag (de "school" van De Klerk). In De Klerk (1979, p. 9) wordt onderwijspsychologie omschreven als "een wetenschappelijke discipline binnen de psychologie, die zich bezighoudt met de systematische bestudering van onderwijsleerprocessen en van de factoren die daarop van invloed zijn". Aan het concept van T&S ligt een Utrechtse opvatting aan de basis: onderwijsproceskunde. Onderwijsproceskunde is gericht op het optimaliseren van leerprocessen, en wel van die leerprocessen die zich in onderwijssituaties voordoen (Schouten-Van Parreren, 1981, p.5). Centrale persoon in de Utrechtse "school" was Van Parreren. In 1960 verscheen van Van Parreren het boek "Psychologie van het leren I" met als ondertitel "verloop en resultaten van leerprocessen"; in dit boek werd het begrip "wendbaarheid" (=transfer) geïntroduceerd. De ondertitel van het boek van Van Parreren vinden we terug als ondertitel van T&S.

Een tweede verschil met betrekking tot het inhoudelijke concept betreft de historische dimensie. In het boek van B&S is deze (bijna) afwezig en in het boek van T&S wordt aan het ontstaan van begrippen en theorieën en de doorwerking ervan naar het heden prominente aandacht gegeven. Wellicht is dit laatste terug te voeren op de omvang van de boeken: T&S 643 pagina's, B&S 301 pagina's.

Ook het redactionele concept van de beide boeken verschilt. Het boek van B&S is gebaseerd op een eerder boek van Boekaerts uit 1987 ("Psychologie van de leerling en het leerproces") en bevat ook onderdelen die teruggaan op haar boek uit 1982 ("Onderwijsleerprocessen organiseren, hoe doe je dat ..?"). In het redactionele concept is duidelijk de expertise van de beide auteurs (Boekaerts: emotionele en motivationele handelingscontrole; Simons: zelfstandig leren van leerlingen) aan te geven. Rond de expertises van de auteurs is het boek opgebouwd. Het boek van T&S zou men de benadering van het vakgebied via deskundigen kunnen noemen. Voor ieder onderwerp, dat in het door de redacteuren opgestelde concept voor het boek past, is een deskundige (uit Nederland) aangezocht. Als criterium voor deskundigheid geldt (voor de hoofdauteur) gepromoveerd op het onderwerp van het hoofdstuk.

Overeenkomsten en verschillen

Een boek over onderwijspsychologie behoort aandacht te schenken aan "leren" en "leertheorieën". Beide boeken kennen dit onderwerp als hoofdstuk 1. In de benadering van beide boeken zijn echter verschillen. In T&S wordt, zoals eerder genoemd, het onderwerp (associationisme, behaviorisme, cognitivisme, handelingstheorie) veel historischer behandeld, maar anderzijds gaan zij meer dan B&S in op de betekenis van de leertheorieën voor het onderwijs en de kritiek erop. B&S maken onderscheid tussen leertheorieën en onderwijspsychologische theorieën. Alle leertheorieën worden kort en kernachtig behandeld. Door hen wordt expliciet aandacht gegeven aan de "meta-cognitieve theorie" en het "constructivisme". De behandelde onderwijspsychologische theorieën hebben de namen van hun belangrijkste vertegenwoordiger (Gagné, Ausubel, Van Parreren, Galperin, Spiro). T&S kiezen het model van Bloom als ordeningskader voor leren op school.

In het onderstaande gebruiken we het (behavioristische) model van Bloom met als belangrijkste variabelen "kenmerken van leerlingen", "instructie" en "leeruitkomsten" om de overeenkomsten en verschillen tussen beide boeken aan te geven.

Kenmerken van leerlingen

In T&S wordt er in speciale hoofdstukken aandacht gegeven aan onderwerpen als "verschillen in begaafdheid", "voorkennis", "leren en geheugen", "motivatie en emotie", terwijl dit bij B&S niet het geval is. De benadering van B&S is een andere. In de hoofdstukken over "domeinspecifieke kennis en vaardigheden" en "(meta)cognitieve strategieën en leerstrategieën" worden de onderwerpen behandeld. De overlap is groot, de diepgang en de insteek zijn verschillend.

Instructie

De boeken behandelen drie thema's binnen "instructie". Het eerste thema is het onderwijsleerproces tijdens de instructie, het tweede is de rol van de leerkracht en het derde is instructietheorieën.

De benadering van het eerste thema door T&S is een vakspecifieke. Via de leer- en denkprocessen binnen de vakken "aanvankelijk lezen", "begrijpend en studerend lezen", "wiskunde" en de "exacte vakken". B&S hebben voor een andere benadering gekozen, namelijk die van de "controle" (motivationale, handelings-, emotionele en sociale controle) in de zin van (zelf)sturing van het leerproces. Ook aan het vakspecifieke wordt aandacht geschonken, maar B&S beperken zich tot lezen, rekenen en schrijven. Er is, uiteraard, weer overlap, maar er zijn grotere verschillen. Het boek van B&S is breder qua scope en vanwege het feit dat het gehele boek door dezelfde auteurs is geschreven, is het een coherenter geheel en door aandacht te hebben voor het onderwerp "controle" heeft het boek iets unieks, dat het andere boek niet heeft.

De rol van de leerkracht is in beide boeken een hoofdstuk waard, echter de benadering is totaal verschillend. Bij B&S heeft de leerkracht een rol in de begeleiding van het leerproces van de leerling, zodat deze zelfstandig leert leren en denken. Vanuit de opvatting over onderwijspsychologie en het accent op het leerproces, wordt de taak van de leerkracht afgeleid en geformuleerd: leerlingen moeten zelfstandig leren leren. Als instrument om de leerkracht te begeleiden, presenteren B&S PRO-AL. PRO-AL staat voor procesgericht onderwijs dat als doel heeft adaptief leren te bevorderen. "Onder proces-gericht onderwijs verstaan we vormen van instructie waarbij expliciete aandacht wordt besteed aan de wijze waarop het leer- en informatieverwerkingsproces verloopt of zou moeten verlopen. Met de term adaptief leren verwijzen we naar leerprocessen waarin diverse vormen van gedragscontrole expliciet worden aangesproken en verder worden geoefend" (B&S, p. 245). Voor de uitvoering van PRO-AL beschrijven B&S drie uitvoeringsregimes (docent-gestuurd, gedeelde sturing en leerling-sturing) en voor ieder uitvoeringsregime de onderwijsfuncties uit het BTES-model (uitleggen, vragen stellen, opdrachten geven, feedback, evalueren en toegevoegd is waarnemen van niet-verbaal gedrag. Hoe anders is de benadering van T&S. Vanuit een proces-produkt benadering (als onderdeel van de descriptieve-correlational-experimental loop design, zoals o.a. beschreven door Rosenshine & Furst, 1973) wordt effectief leerkrachtgedrag beschreven.

Het derde thema wordt gevormd door de instructietheorieën. Alleen door B&S wordt hier expliciet aandacht aan besteed. Het gebruik van het begrip "instructietheorieën" kan de deskundige lezer in verwarring brengen. Het gaat in dit hoofdstuk vooral om het model dat in de BTES-studie is geformuleerd, nader uitgewerkt door Good, Grouws & Ebmeier (1983) en in Nederland gepopulariseerd door Veenman (1990). Het gaat niet om een benadering van de instructietheorieën à la Reigeluth (1983, 1987), maar om een model voor onderwijs in de classesituatie. Onder het BTES-model leggen B&S het BIT-model (begrijpen, integreren, toepassen). De opsomming/bespreking van didactische werkvormen is in min of meer dezelfde vorm in het boek van Boekaerts uit 1982 beschreven. In T&S krijgt dit (BTES)model eveneens in hetzelfde kader aandacht.

Leeruitkomsten

In T&S ligt de nadruk op het begrip "transfer". Hieraan worden drie hoofdstukken besteed en zonder de andere hoofdstukken te kort te doen, moet hier worden opgemerkt, dat dit het unieke van dit boek is. In het register van T&S kent het begrip "transfer" 44 verschillende ingangen, in B&S slechts één. "Transfer" is een van de meest centrale thema's in leersituaties, vooral omdat deze het toepassingskarakter van het geleerde in verschillende situaties betreft. Hierbij is de kernvraag hoe we de onderwijsleersituaties moeten inrichten om het verschijnen van positieve transfer te optimaliseren en daarbij de mogelijkheid tot het induceren van negatieve transfer te elimineren. Het onderwerp wordt uitgebreid en vanuit de historie (beginnend bij Thorndike) behandeld. Met name zijn de hoofdstukken 13 ("Het ontwerp van transfersituaties") en 15 ("Transfer: oefening baart kunst") interessant en ter lezing aanbevolen. De hoofdstukken over transfer (geschreven door Kingma en Ten Vengert) zijn een uitstekende aanvulling op o.a. het thema-nummer Transfer (redactie Simons & Verschaffel, 1992) van het TOR. Als zodanig raad ik alle onderzoekers op dit gebied aan men name van deze hoofdstukken kennis te nemen.

Slotopmerkingen

Het boek van T&S bevat twee hoofdstukken die m.i. niet in het boek thuis horen: het hoofdstuk over testleer (hoofdstuk 16) en het hoofdstuk over constructie van beoordelingsinstrumenten (hoofdstuk 17). Het hoofdstuk over testleer behandelt de klassieke onderwerpen en gezien het karakter van de andere hoofdstukken had ik hier een "modernere" benadering van de testtheorie verwacht (Rasch, latente trek theorieën, etc.). Het hoofdstuk over beoordelingsinstrumenten is van een ander abstractie-niveau dan de rest van het boek.

Twee andere hoofdstukken in T&S die men niet snel verwacht in een boek over onderwijspsychologie en die nog een opmerking behoeven, betreffen het onderwerp "cognitieve ontwikkeling". Vanwege de koppeling van dit onderwerp met "representatie" passen deze hoofdstukken goed in het concept en wordt de lezer ingeleid in Piaget, Bruner en Galperin.

Een laatste punt waar ik aandacht aan wil geven, betreft de doelgroep: voor wie zijn nu deze boeken bestemd. Beide boeken zijn inleidingen, maar inleidingen van verschillende niveaus. Het boek van B&S is m.i. geschikt voor twee doelgroepen van studenten: (1) studenten die interesse hebben in de onderwijspsychologie en (2) studenten van de lerarenopleiding. Het is m.i. geen boek voor doktoraalstudenten. Het boek van T&S is primair bedoeld voor doktoraalstudenten; de hoofdstukken in T&S geven in kort bestek de stand van zaken op een bepaald terrein (maar dat is inherent aan de benadering die door de redacteuren is gekozen) en daarmee denk ik dat het ook als naslagwerk voor onderzoekers kan/zal gaan fungeren.

Beide boeken "lezen" goed en er is zorg aan de vormgeving besteed. Beide boeken zullen hun eigen specifieke rol voor de eigen doelgroep in het universitaire onderwijs gaan innemen en ik hoop, dat de gebruikers/docenten zich bij de keuze laten leiden door de benadering van de onderwijspsychologie die zij aan hun studenten willen overdragen: een proces- of een produktbenadering. De keuze moet een bewuste zijn.

REFERENTIES

- Good, T.L., Grouws, D.A. & Ebmeier, H. (1983). *Active mathematics teaching*. New York: Longman.
- De Klerk, L.F.W. (1979). *Inleiding in de onderwijspsychologie*. Deventer: Van Loghum Slaterus.
- Van Parreren, C.F. (1960). *Psychologie van het leren*. Deventer: Van Loghum Slaterus.
- Van Parreren, C.F. (1981). *Onderwijsproceskunde*. Groningen: Wolters-Noordhoff.
- Van Parreren-Schouten, M.C. (1981). Voorwoord. In C.F. van Parreren, *Onderwijsproceskunde*. Groningen: Wolters-Noordhoff.
- Reigeluth, C.M. (1983). *Instructional-design theories and models*. Hillsdale: LEA.
- Reigeluth, C.M. (1987). *Instructional theories in action*. Hillsdale: LEA.
- Rosenshine, B. & Furst, N. (1973). The use of direct observation to study teaching. In R.M.W. Travers (Ed.), *Second handbook of research on teaching*. Chicago: Rand McNally.
- Simons, P.R.J. & Verschaffel, L. (Red.) (1992). Themanummer Transfer. *Tijdschrift voor Onderwijsresearch*, 17 (1), 3-66.
- Veenman, S., Lem, P. & Nijssen, F. (1990). *Omgaan met combinatieklassen*. Den Haag: Staatsuitgeverij.

P.C. van der Sijde
Universiteit Twente

Swets & Zeitlinger BV

Postbus 825 •
2160 SZ Lisse •
tel. 02521-35111 •
fax. 02521-15888 •



**swets &
zeitlinger**

publishers

Ouders en schoolsucces

Activiteiten om ouders bij het onderwijs te betrekken

B. Hogeboom

Ouders kunnen veel bijdragen aan het schoolsucces van hun kind. De kwaliteit van deze ondersteuning is in belangrijke mate afhankelijk van de opleiding van de ouders. Lager opgeleide ouders, waartoe veel allochtone ouders behoren, beschikken vaak niet over de kennis en de vaardigheden om hun kinderen te helpen.

In Nederland en het buitenland worden activiteiten ondernomen om deze groep te helpen bij de ondersteuning van hun kinderen. Aan de hand van beschrijvingen van projecten in binnen- en buitenland geeft de auteur praktische aanbevelingen voor scholen en instellingen die zelf activiteiten in deze richting willen ontplooien.

Daarnaast komen theoretische aspecten van de ouderlijke ondersteuning aan bod. Uitgegeven in samenwerking met SARDES: het Samenwerkingsverband SVO-CPS voor Advisering, Research en Dienstverlening in de Educatieve sector.

ISBN 90 265 1376 3, Reeks Kind in Ontwikkeling,
ca. 136 pagina's, f 29,50

Bestelbon

Stuur mij:

..... ex. Ouders en schoolsucces

via boekhandel

te

rechtstreeks naar onderstaand adres

naam: _____

straat: _____

postcode: _____

plaats: _____

datum: _____

handtekening: _____

Stuur deze bon in een open envelop zonder postzegel naar:
Swets & Zeitlinger, antwoordnummer 202, 2160 VB Lisse

bij rechtstreekse levering worden verzendkosten in rekening gebracht

Themanummer: Ongelijkheid in onderwijskansen

Inleiding

J. Dronkers¹, J.A.M. Heijke² en R.K.W. van der Velden²

¹SCO-Kohnstamm Instituut, Universiteit van Amsterdam

²Researchcentrum voor Onderwijs en Arbeidsmarkt, Rijksuniversiteit Limburg

De ongelijkheid van onderwijskansen is een thema dat reeds sinds de jaren '60 een prominente plaats heeft op de onderzoeksagenda van onderwijssociologen en onderwijskundigen. Het beginpunt van deze onderzoekstraditie wordt vaak gelegd bij het befaamde 'Talentenproject' dat onder leiding van de Leidse hoogleraar Van Heek tussen 1961 en 1967 werd uitgevoerd. Het eindverslag van dit project¹ heeft een belangrijke rol gespeeld in de verdere wetenschappelijke debatten over onderwijsongelijkheid.

Had het onderzoek van Van Heek nog uitsluitend betrekking op de achterstandssituatie van arbeiderskinderen, in de jaren daarna komen ook andere achterstandsgroepen in beeld. Met name de positie van meisjes en van allochtone kinderen in het onderwijs worden belangrijke onderzoeksthema's evenals veranderingen in de onderwijskansen en de betekenis van schooleffecten.

Ook op dit moment is de ongelijkheid van onderwijskansen nog steeds een belangrijk onderzoeksthema, getuige de bijdragen in dit themanummer. De artikelen zijn bewerkingen van enkele papers die tijdens de Onderwijs Research Dagen van mei 1993 te Maastricht zijn gepresenteerd binnen het thema 'Onderwijs en Samenleving'.

De artikelen in dit themanummer geven een goede illustratie van 'the state of the art' op het onderzoeksterrein van de ongelijke onderwijskansen. In vergelijking met de beginperiode zien we een aantal belangrijke verschuivingen. Ten eerste is een belangrijk accent komen te liggen op de analyse van causale modellen in plaats van tabellen. Ten tweede is er een verschuiving merkbaar naar de meer procesmatige elementen achter de relatie tussen achterstandsgroep en schoolsucces. Ten derde zijn er belangrijke verbeteringen geboekt op het gebied van de toegepaste statistische methoden, zoals multilevel analyse.

Het artikel van *Bakker en Cremers* stelt de vraag aan de orde in hoeverre het effect van sociale herkomst op de keuze van voortgezet onderwijs is veranderd. Zij concluderen dat de invloed van sociale herkomst in de loop der tijd enigszins is afgenomen. Deze afname kan echter niet verklaard worden door een meritocratisering van het selectieproces in het onderwijs. Het effect van de prestaties op het advies en de keuze voor voortgezet onderwijs is namelijk eerder afgenomen dan toegenomen.

Het artikel van *Van der Hoeven-van Doornum* geeft een (gedeeltelijke) verklaring waarom deze meritocratisering achterwege blijft. Leerkrachten hanteren voor leerlingen uit verschillende milieus verschillende streefniveaus. Op grond van het verwachte leervermogen en de sociale herkomst van leerlingen wordt door leerkrachten reeds vroegtijdig een inschatting gemaakt van het verwachte niveau van voortgezet onderwijs. Dit streefniveau heeft een langdurige werking op de schoolloopbaan in het basis- en zelfs het voortgezet onderwijs.

Ook in het artikel van *Bosker en Dekkers* is de aandacht gericht op de school als mogelijke verklaringsbron voor verschillen in onderwijskansen. Hun aandacht gaat uit naar de invloed van de school op de verschillen tussen jongens en meisjes in de keuze van wiskunde B. Hieruit blijkt dat de verschillen tussen jongens en meisjes in wiskunde prestaties gedurende de eerste vier

1. F. van Heek e.a. (1968). Het verborgen talent. Milieu, schoolkeuze en schoolgeschiktheid. Meppel: Boom.

leerjaren steeds groter worden. Scholen verschillen niet in deze seksspecifieke ontwikkeling van de wiskundeprestaties. Ze verschillen echter wel in de wijze waarop de prestaties uiteindelijk omgezet worden in de keuze voor wiskunde B. Met andere woorden, bij een gelijk prestatieniveau zijn meisjes op de ene school meer geneigd om voor wiskunde B te kiezen dan op een andere school. De invloed van de school is weliswaar niet erg groot, maar vanuit beleidsmatig oogpunt wel relevant.

De laatste twee artikelen hebben betrekking op de schoolprestaties van allochtone leerlingen. In het artikel van *Latuheru en Hessels* is de vraag aan de orde of het effect van etnische herkomst toegeschreven kan worden aan verschillen in sociaal milieu of verschillen in intelligentie. Uit hun analyse komt naar voren dat van een zelfstandig effect van etnische herkomst geen sprake meer is na controle voor intelligentie en sociaal milieu. Het omgekeerde blijkt echter ook het geval te zijn. Na controle voor etnische herkomst, is er geen significant effect meer van sociaal milieu. De auteurs concluderen dan ook dat de effecten van etnische herkomst en sociale herkomst zodanig onderling samenhangen, dat een schatting van de afzonderlijke bijdrage van etnische herkomst niet mogelijk is.

Ook in de bijdrage van *Koeslag en Dronkers* blijkt dat de onderwijsachterstand van allochtone leerlingen voor een groot deel verklaard kan worden door de lage maatschappelijke positie van allochtonen. Dat betekent overigens niet dat de schoolloopbanen van allochtone leerlingen zich op dezelfde manier voltrekken als bij autochtone leerlingen. Er worden duidelijke verschillen aangetroffen in de effecten van ouderlijk milieu en prestaties op de bereikte onderwijspositie. Dat geldt niet alleen voor de verschillen tussen allochtonen en autochtonen, ook binnen de groep allochtonen zijn duidelijke verschillen aanwijsbaar in de manier waarop de schoolloopbanen zich voltrekken. Deze procesmatige verschillen zijn vermoedelijk een gevolg van verschillen in migratiegeschiedenissen en uiteenlopende culturen.

Gelijke kansen in het onderwijs?

Een vergelijking van vier cohorten leerlingen in hun overgang naar het voortgezet onderwijs

B.F.M. Bakker en P.G.J. Cremers*

ABSTRACT

In the Netherlands the school system has been reformed to reduce the effects of parent's socio-economic position on the educational attainment of their children. The main educational reform is the so-called Mammoth-law, in which a more meritocratic selection process is proposed. By selecting pupils on their scholastic achievement instead of on their background, it was assumed that pupils from lower socio-economic backgrounds would be more enabled to achieve higher educational levels. In this article, the educational careers of four cohorts of Dutch pupils are compared, born around 1953, 1965, 1971 and 1977. Criterion for educational attainment is their choice for a particular level of secondary education. We show that the educational opportunities of pupils from lower socio-economic backgrounds have improved. However, this is not caused by meritocratisation in education: the effect of scholastic achievement has not changed during the last 25 years.

1. INLEIDING

De positie van achterstandgroepen in het Nederlandse onderwijs mag zich verheugen in een grote belangstelling van zowel wetenschappers als politici. Er zijn een aantal maatschappelijke ontwikkelingen geweest die mogelijk de relaties tussen sociaal milieu en onderwijskansen hebben beïnvloed. De groei van de economie heeft de noodzaak doen toenemen het aanwezige talent optimaal te benutten. Een 'screening device' voor de mate van talent was het bereikte opleidingsniveau. Voor een betere benutting van het aanwezig talent was het noodzakelijk ook talentvolle kinderen uit lagere milieus te scholen. Dit kon worden bereikt door het onderwijs te meritocratiseren, dat wil zeggen dat meer op prestaties dan op afkomst of geslacht werd geselecteerd. Hierdoor zou de kans toenemen dat kinderen uit lagere sociale milieus ten opzichte van hun ouders sociaal stijgen. Dit wordt wel de moderniseringstheorie genoemd.

Dit moderniseringsdenken wordt onder andere teruggevonden in de hervormingen in het onderwijs na de tweede wereldoorlog. Zo is het onderwijsbeleid in de jaren zestig en zeventig onder andere gericht op de verbetering van onderwijskansen van leerlingen afkomstig uit lagere milieus (bijvoorbeeld leerlingen waarvan de vader arbeider is of waarvan de ouders geen diploma in het voortgezet onderwijs hebben behaald) en meisjes. De bekendste onderwijsvernieuwing is de Mammoetwet geweest, die vanaf 1968 effectief is, waarin is geregeld dat een objectieve prestatietoets aan het eind van de basisschool en het advies voor voortgezet onderwijs bepalend zijn voor de feitelijke toegang tot het voortgezet onderwijs. Tevens moeten deze adviezen mede worden gebaseerd op de prestaties van de leerlingen. Dit kwam in de plaats van de toelatingsexamens die tot op dat moment werden gebruikt als middel voor selectie.

Naast deze Mammoetwet zijn er andere voorbeelden van overheidsingrijpen, die gericht waren op het beïnvloeden van loopbanen van achterstandsgroepen in het lager onderwijs. De

* B.F.M. Bakker en P.G.J. Cremers zijn werkzaam bij het Centraal Bureau voor de Statistiek. De eerste op de hoofdafdeling Sociale Rekeningen, de tweede op de hoofdafdeling Statistiek van Onderwijs en Wetenschappen. Dit artikel is op persoonlijke titel geschreven en geeft als zodanig geen opvattingen of keuzen van het CBS weer. Correspondentie richten aan: P.G.J. Cremers, CBS, Hoofdafdeling Statistiek van Onderwijs en Wetenschappen, Postbus 4481, 6401 CZ Heerlen.

belangrijkste daarvan zijn het onderwijsstimuleringsbeleid en het onderwijsvoorrangsbeleid, die respectievelijk vanaf 1974 en 1985 effectief zijn geworden. In dit beleid wordt gestreefd naar de verbetering van de onderwijskansen van arbeiderskinderen door de ontwikkeling van nieuwe leermiddelen en -methodes. Daarnaast krijgen scholen die veel leerlingen uit achterstandsgroepen telden additionele financiële middelen voor het aanstellen van extra leerkrachten (Tesser, Mulder en Van der Werf, 1991). In het onderwijsvoorrangsbeleid wordt tevens aandacht geschonken aan allochtone leerlingen. In het bijzonder voor Turkse en Marokkaanse kinderen geldt dat hun onderwijsprestaties sterk achterblijven bij hun autochtone leeftijdsgenoten. Hoewel dit onderwerp nog sterk in discussie is, lijkt de conclusie gerechtvaardigd dat dit niet alleen te wijten is aan het feit dat ze allochtoon zijn, maar voor een goed deel aan de zwakke sociaal-economische positie van hun ouders.

Samengevat richt het beleid zich op twee punten:

- Het verbeteren van de leerprestaties van leerlingen uit lagere milieus, hetgeen zou moeten resulteren in betere schoolprestaties aan het eind van de basisschool.
- Het versterken van het belang van de schoolprestaties aan het eind van de basisschool voor de keuze in het voortgezet onderwijs. Dit wordt ook wel meritocratisering van het onderwijs genoemd. Door het advies voor voortgezet onderwijs onder meer te baseren op 'objectieve' prestaties worden de leerlingen uit lagere milieus met goede prestaties gestimuleerd om hogere vormen van voortgezet onderwijs te kiezen.

Naast de opkomst van de moderniseringsgedachte hebben echter ook andere maatschappelijke ontwikkelingen mogelijk invloed gehad op de relatie tussen sociaal milieu en schoolsucces. Zo wijzen onder andere Parkin (1979) en Bourdieu en Paaseron (1977) erop dat aanwezige talenten niet de enige kwaliteiten zijn die nodig zijn voor schoolsucces. Omdat in het onderwijs juist aandacht gegeven zou worden aan de "formal culture", dat wil zeggen de cultuur van de hogere en middenklassen, hebben kinderen afkomstig uit hogere sociale milieus grotere kansen op een succesvolle school- en beroepsloopbaan dan kinderen uit lagere sociale milieus. Naarmate leerlingen afkomstig zijn uit hogere sociale milieus hebben ze meer kennis van de formele cultuur (cultureel kapitaal) en kunnen ze beter voldoen aan de eisen die in het onderwijs aan hen worden gesteld. Omdat de overheidssector sterk gegroeid is, is ook het aantal ouders met relatief veel cultureel en weinig materieel kapitaal gegroeid. Dit kan ertoe leiden dat de relatie tussen sociaal milieu en onderwijskansen in de loop der tijd is veranderd, omdat de overdracht van materieel kapitaal als land, industrieel, commercieel en financieel kapitaal eenvoudiger is dan de overdracht van cultureel kapitaal. Bij de overdracht van deze laatste vorm van kapitaal speelt intelligentie en de erfelijke component daarvan een rol. De minst intelligente kinderen uit hogere sociale milieus zullen, ondanks de steun van hun ouders, niet de leerprestaties kunnen leveren om in het onderwijs succesvol te zijn, en daarom ten opzichte van hun ouders sociaal dalen. De meest intelligente kinderen uit lagere milieus daarentegen zullen door het leveren van goede schoolprestaties een hoog onderwijsniveau bereiken en sociaal stijgen. Dat wil zeggen dat de uiteindelijke uitkomst van de selectieprocessen in het onderwijs een minder scheve verdeling van schoolsucces naar sociaal milieu te zien geeft.

In dit artikel worden de schoolloopbanen vergeleken van vier cohorten leerlingen die geboren zijn rond 1953, 1965, 1971 en 1977. De schoolloopbanen worden beschreven tot en met de keuze in het voortgezet onderwijs. De analyse spitst zich toe op de volgende vragen:

1. Zijn de leerprestaties van leerlingen uit lagere milieus aan het eind van de basisschool in de loop der tijd ten opzichte van de overige sociale milieus verbeterd?
2. Is het belang van de leerprestaties aan het eind van de basisschool voor het advies voor en de keuze in het voortgezet onderwijs toegenomen?
3. Zijn de onderwijskansen van de lagere milieus ten opzichte van de hogere milieus verbeterd? Kan een eventuele verbetering worden verklaard door een toename van het belang van de leerprestaties?

In het verleden is reeds veel methodologische onderzoek verricht naar veranderingen in schoolloopbanen. Deze onderzoeken hebben echter steeds enkele nadelen. Zo zijn de meeste studies

gebaseerd op de vergelijking van slechts twee cohorten leerlingen (De Jong, Dronkers en Saris, 1981; Willemse, 1987) en op lokale steekproeven (Bakker, Dronkers en Schijf, 1982; Meesters, Dronkers en Schijf, 1983; Faasse et al. 1986; Vrooman en Dronkers, 1986; Dronkers en Bosma, 1990). Uitzonderingen vormen de studies van Bakker en Schouten (1991) en Dronkers (1992), die drie cohorten leerlingen vergeleken die een afspiegeling vormen van de Nederlandse leerlingen in de betreffende cohorten en in totaal respectievelijk 18 en 24 jaar bestreken. Het vergelijken van meerdere cohorten heeft als voordeel dat de uitkomsten minder gevoelig zijn voor toevallige eigenaardigheden van de betreffende databestanden. Het bestrijken van een zo ruim mogelijke periode heeft als voordeel dat lange termijn veranderingen zichtbaar worden. In dit artikel wordt in deze traditie doorgewerkt. Daarbij wordt getracht om de nadelen van eerdere studies te ondervangen door vier cohorten leerlingen te vergelijken die in totaal een periode van 24 jaar bestrijken. Ten opzichte van Bakker en Schouten (1991) is het cohort geboren rond 1977 toegevoegd. Ten opzichte van Dronkers (1992) is het cohort geboren rond 1971 toegevoegd en is de operationalisering van de keuze in het voortgezet onderwijs voor het cohort geboren rond 1977 verbeterd door gebruik te maken van de onderwijspositie in het tweede en derde leerjaar in het voortgezet onderwijs. Bovendien is in tegenstelling tot Dronkers gebruik gemaakt van de gehele steekproef. Dat wil zeggen dat ook éénoudergezinnen erin zijn opgenomen.

Naast de historische vergelijkingen van schoolloopbanen is recentelijk ook met behulp van dwarsdoorsnede-onderzoek het verband tussen sociaal milieu en onderwijskansen onderzocht (Ganzeboom en De Graaf, 1989a en b; De Graaf en Luijkx, 1992). Hieruit komt naar voren dat de invloed van sociaal milieu op onderwijskansen gedurende de afgelopen eeuw ongeveer gehalveerd is. Onze analyses zullen tegen deze achtergrond worden gepresenteerd.

2. DE GEGEVENS, DE VARIABELEN, HET GEBRUIKTE SCHOOLLOOPBAANMODEL EN DE WERKWIJZE

2.1. De gebruikte gegevens

Voor de gegevens van het cohort geboren rond 1953 wordt gebruik gemaakt van het 'Van Jaar tot Jaar' (VJTJ)-onderzoek. Deze steekproef omvat 1948 waarnemingen en is representatief voor het desbetreffende cohort leerlingen. De gegevens van het cohort geboren rond 1965 zijn afkomstig van het CBS-onderzoek "Sociaal Milieu en Voortgezet Onderwijs" (SMVO). Dit betreft een per schoolsoort gestratificeerde steekproef, die na herweging representatief is voor het desbetreffende cohort (N=37280). De gegevens van het cohort geboren rond 1971 zijn afkomstig van het CBS-onderzoek "Schoolloopbaan en herkomst van leerlingen bij het Voortgezet Onderwijs" (SLVO). Dit betreft een willekeurige steekproef van lagere scholen waarvan alle leerlingen in de onderzoekspopulatie zijn opgenomen die in de zesde klas verbleven. Ook deze gegevens zijn representatief voor het desbetreffende cohort leerlingen (N=16813). De gegevens voor het cohort 1977 zijn afkomstig uit het onderzoek "Voortgezet Onderwijs Cohort Leerlingen 1977" (VOCL). Dit betreft een tweetrapssteekproef. Eerst werden er scholen voor voortgezet onderwijs en vervolgens klassen uit het eerste leerjaar getrokken. Deze steekproef is representatief voor het cohort leerlingen geboren rond 1977 (N=19524).

Voor het cohort geboren rond 1953 maken wij gebruik van de door De Jong *et al.* voor het VJTJ-onderzoek berekende correlaties, standaardafwijkingen en gemiddelden. Voor de cohorten geboren rond 1965 en 1971 maken we gebruik van de door Bakker en Schouten (1991) gepubliceerde gegevens. We hebben de gegevens van het cohort geboren rond 1977 toegevoegd.

2.2. De variabelen

In tabel 1 is een overzicht gegeven van de verdelingen, gemiddeldes en standaardafwijkingen van de in deze studie betrokken variabelen. Deze zijn in het algemeen op intervalniveau (ge-

Tabel 1. De gebruikte variabelen, gemiddelden en standaardafwijkingen.

	cohort					cohort			
	1953 %	1965	1971	1977		1953 %	1965	1971	1977
beroepsgroep vader					geboortekwartaal				
1. geen werkkring	-	20	21	18	1. oktober-december	22	24	25	23
2. arbeiders	42	32	30	31	2. januari-maart en april-juni	51	50	51	51
3. boeren	11	5	4	2	3. juli-september	27	26	24	27
4. middenstand	13	7	6	5	gemiddelde	2.05	2.01	2.00	2.03
5. lagere employés	9	10	11	8	standaardafwijking	.70	.71	.70	.70
6. middelbare employé	19	16	17	19					
7. hogere beroepen	5	10	11	16	geslacht				
gemiddelde	3.67	3.43	3.52	3.77	1. jongen	49	51	50	52
standaardafwijking	1.74	2.10	2.16	2.26	2. meisjes	51	49	50	49
opleiding vader					gemiddelde	1.51	1.49	1.50	1.49
1. alleen lager onderwijs	41	31	37	25	standaardafwijking	.50	.50	.50	.50
2. lbo	38	25	24	18					
3. mavo, vwo/havo lrtj 1-3	8	5	9	8	gezinsgrootte				
4. mbo	5	24	11	26	1. kind = leerling	6	15	7	7
5. vwo/havo	3	2	4	5	2. kinderen	16	33	47	51
6. hbo	3	10	11	12	3. kinderen	19	28	26	27
7. wo	2	3	4	6	4. of meer kinderen	59	25	20	15
gemiddelde	2.08	2.83	2.69	3.26	gemiddelde	3.32	2.64	2.58	2.50
standaardafwijking	1.40	1.78	1.88	1.90	standaardafwijking	.94	1.01	.88	.83
opleiding moeder					doubleren				
1. alleen lager onderwijs	66	48	45	30	1. klas overgeslagen	1	1	1	1
2. lbo	21	24	21	27	2. normaal	63	76	84	74
3. mavo, vwo/havo lrtj 1-3	7	10	13	13	3. 1 doublure	31	19	14	22
4. mbo	2	11	13	18	4. 2 of meer doublures	6	4	2	4
5. vwo/havo	2	2	3	4	gemiddelde	2.43	2.27	2.16	2.28
6. hbo	1	4	5	7	standaardafwijking	.62	.53	.42	.54
7. wo	0	1	1	1					
gemiddelde	1.58	2.10	2.26	2.66	prestatiescore				
standaardafwijking	1.40	1.39	1.88	1.58	1. laag	10	7	7	8
urbanisatiegraad					2.	30	28	29	25
1. plattelandsgemeenten	26	14	15	13	3.	36	33	33	33
2. versted platteland en forenzengemeenten	25	38	45	44	4.	20	25	25	29
3. plattelandsteden	3	3	3	3	5. hoog	4	8	7	7
4. kleine steden	7	9	10	10	gemiddelde	2.81	3.00	2.95	3.02
5. kleine middelgr steden	5	6	6	5	standaardafwijking	1.07	1.05	1.03	1.05
6. grote middelgr steden	12	10	7	12					
7. grote steden	22	22	15	13	gemiddelde	3.61	3.71	3.27	3.36
gemiddelde	3.61	3.71	3.27	3.36	standaardafwijking	2.39	2.26	2.08	2.07
standaardafwijking	2.39	2.26	2.08	2.07					

Tabel 1. De gebruikte variabelen, gemiddelden en standaardafwijkingen. (vervolg)

	cohort					cohort			
	1953	1965	1971	1977		1953	1965	1971	1977
	%					%			
advies voortgezet onderw.					keuze voortgezet onderw.				
1. vglo/lavo	12	1	1	0	1. vglo/lavo	7	0	0	0
2. lbo	43	34	30	29	2. lbo	42	37	32	34
3. ulo/mavo	30	36	33	35	3. ulo/mavo	34	38	38	42
4. mms/havo	4	18	22	20	4. mms/havo	2	10	13	11
5. rest vhmo/vwo	12	12	15	15	5. rest vhmo/vwo	15	15	17	13
gemiddelde	2.61	3.09	3.19	3.21	gemiddelde	2.75	3.03	3.16	3.04
standaardafwijking	1.12	1.00	1.05	1.03	standaardafwijking	1.12	1.03	1.06	.99

boortekwartaal, gezinsgrootte, doubleren, prestatiescore), dummy-variabelen (geslacht) of ordinaal waarvan verondersteld mag worden dat zij bij gebruik van rangnummers een goede interschaal benaderen (opleiding vader en moeder, urbanisatiegraad, advies en keuze). Als deze variabelen optimaal worden geschaald met behulp van correspondentie-analyse, worden de categorieën globaal op dezelfde wijze gecategoriseerd als door ons verondersteld (Meester en De Leeuw, 1983; Vrooman en Dronkers, 1986; Dronkers, 1992).

Het beroep van de vader is voor het cohort 1953 anders vastgesteld dan voor de overige cohorten. Bij het cohort 1953 is voor de niet-werkende vaders het laatste beroep gevraagd, zonder dat dit goed kan worden onderscheiden van de beroepen van werkenden. Deze keuze is begrijpelijk, omdat sprake was van vrijwel volledige werkgelegenheid toen in 1965 de enquête werd afgenomen. De meeste werklozen bleven slechts kort zonder werk en vonden in het algemeen een soortgelijke baan als vóór hun werkloosheid. Arbeidsongeschikten en éénoudergezinnen waren nog relatief schaars. Het laatste beroep kan daarom voor het cohort 1953 als een goede indicator worden beschouwd voor de sociaal-economische positie van de vader. De 2% van de vaders die geen beroep konden opgeven zijn door De Jong et al. niet in de analyse betrokken.

In de periode tussen 1960 en 1988 hebben een aantal ontwikkelingen plaatsgevonden die het percentage leerlingen zonder werkende ouders heeft doen stijgen. Door de toename van het aantal echtscheidingen is het aantal éénoudergezinnen bijna verdubbeld (CBS, 1970, 1991). De arbeidsongeschiktheid is onder invloed van de instelling van de WAO toegenomen van 221 000 in 1965 tot 803 000 in 1988 (CBS, 1970, 1991). Het aantal personen met een werkloosheidsuitkering is in dezelfde periode gegroeid van 27000 tot 590000 (CBS, 1991). Bovendien is het laatste beroep als graadmeter voor de sociaal-economische positie veel minder geschikt geworden, omdat de kans om weer aan de slag te komen voor arbeidsongeschikten en werklozen gering is. In dat verband is gesproken over het ontstaan van een nieuwe maatschappelijke onderklasse (Arts en Ultee, 1985; Bakker, 1992). Voor de laatste twee cohorten is dan ook een aparte categorie opgenomen, die onder de arbeiders is geplaatst en ongeveer éénvijfde van de leerlingen bevat. Hoewel het verschillend operationaliseren van het beroep van de vader niet onproblematisch is, is aannemelijk dat steeds een goede weergave wordt gegeven van de sociaal-economische verschillen binnen de vier onderzochte cohorten.

Bij de variabele opleidingsniveau van de vader en de moeder treden discontinuïteiten in de ontwikkelingen op. Zo wordt waargenomen dat het percentage vaders met een mbo-opleiding eerst is toegenomen van 5 tot 23%, vervolgens is afgenomen tot 11% en weer toegenomen naar 26%. Het percentage vaders zonder diploma zou eerst zijn afgenomen, vervolgens weer toegenomen en ten slotte weer afgenomen. Vergelijking met resultaten uit de Arbeidskrachtentelling 1985 (CBS, 1987) en de Enquête Beroepsbevolking 1988 (CBS, 1989) toont aan dat van de vier

cohorten de gegevens van de cohorten 1965 en 1977 het meest overeenkomen met de gegevens van de desbetreffende cohorten uit de AKT'85 en de EBB'88. De discontinuïteiten moeten toegeschreven worden aan vraagstellingseffecten. In de VJTJ-vragenlijst is een eenvoudige vraag naar het hoogst behaalde diploma gesteld. In de SMVO- en de VOCL-vragenlijst is aan de ouders gevraagd om al de door hen gevolgde opleidingen te noteren, waarna deze informatie door codeurs is verwerkt. In de SLVO-vragenlijst is een voorgecodeerde vraag opgenomen waarin de ouders zelf de door hen gevolgde opleidingen moesten aankruisen. Bovendien is de ontsnapingsmogelijkheid 'een andere opleiding, te weten' onder het kopje 'hoger beroepsonderwijs' terecht gekomen, waardoor veel ouders niet van deze ontsnapingsmogelijkheid gebruik hebben kunnen maken. Het vermoeden bestaat dat in het SLVO de wijze van vragen geleid heeft tot een onderschatting van die opleidingen die niet voorgecodeerd waren of niet duidelijk in een bepaalde categorie konden worden ondergebracht. Gezien de diversiteit aan opleidingen in het lbo en mbo is aannemelijk dat juist deelname aan deze onderwijsvormen wordt onderschat. De mbo-ers zullen dikwijls terugvallen naar het lbo-niveau, terwijl de lbo-ers zullen terugvallen naar de categorie 'alleen lager onderwijs'. Tevens is aannemelijk dat voorcoderen van opleidingstypen, zoals in VJTJ en SLVO is gebeurd, leidt tot onderschatting van het niet-reguliere onderwijs. Dit zou het relatief lage percentage vaders met een mbo-diploma in VJTJ en SLVO mede kunnen verklaren.

Bij *gezinsgrootte* valt een trend op naar kleinere gezinnen. Het percentage gezinnen met twee en drie kinderen is toegenomen onder gelijktijdige afname van het percentage grotere gezinnen. Deze trend is echter niet lineair. Zo zou het percentage leerlingen zonder broers of zusters eerst sterk zijn gestegen, om vervolgens weer te dalen. Inspectie van de diverse vragenlijsten doet vermoeden dat we ook hier te maken hebben met vraagstellingseffecten. In het SMVO wordt niet duidelijk vermeld of alle kinderen uit het gezin of alle kinderen met uitzondering van de leerling zelf mochten worden opgegeven. Dit zou betekenen dat alle categorieën met uitzondering van de categorie '4 of meer kinderen' voor het cohort 1965 overschat zijn.

Dit zijn voorbeelden van de invloed die vraagstellingen kunnen hebben op de resultaten. Het is echter niet aannemelijk dat de vertekeningen die door de veranderde vraagstelling zijn opgetreden, grote consequenties zullen hebben voor de verschillen in samenhangen. Correlaties zijn zeer robuust voor dit soort verschillen, terwijl de verschillen in standaardafwijkingen gering zijn.

De *prestatiescore* is voor de vier cohorten gestandaardiseerd in een standard-five verdeling, omdat de prestatiescores voor de drie cohorten met toetsen van verschillende omvang zijn vastgesteld en de ruwe scores daardoor niet vergelijkbaar zijn. Over verschillen in gemiddelden en standaardafwijkingen kunnen derhalve geen uitspraken worden gedaan. Voor het cohort 1953 is het betrekkelijk eenvoudig de keuze in het voortgezet onderwijs vast te stellen, omdat het eerste jaar in het voortgezet pre-mammoetonderwijs reeds sterk gedifferentieerd was. Met de komst van de mammoetwet is echter een algemene brugklas ingevoerd. Daarom wordt de schoolkeuze van de cohorten 1965, 1971 en 1977 geoperationaliseerd als het eerst bezochte schooltype na afsluiting van de brugperiode. De leerlingen die in het derde leerjaar nog in een brugklas zaten, zijn uit de steekproef verwijderd. Dit betreft een gering aantal leerlingen (ongeveer 1%) in de cohorten 1965, 1971 en 1977.

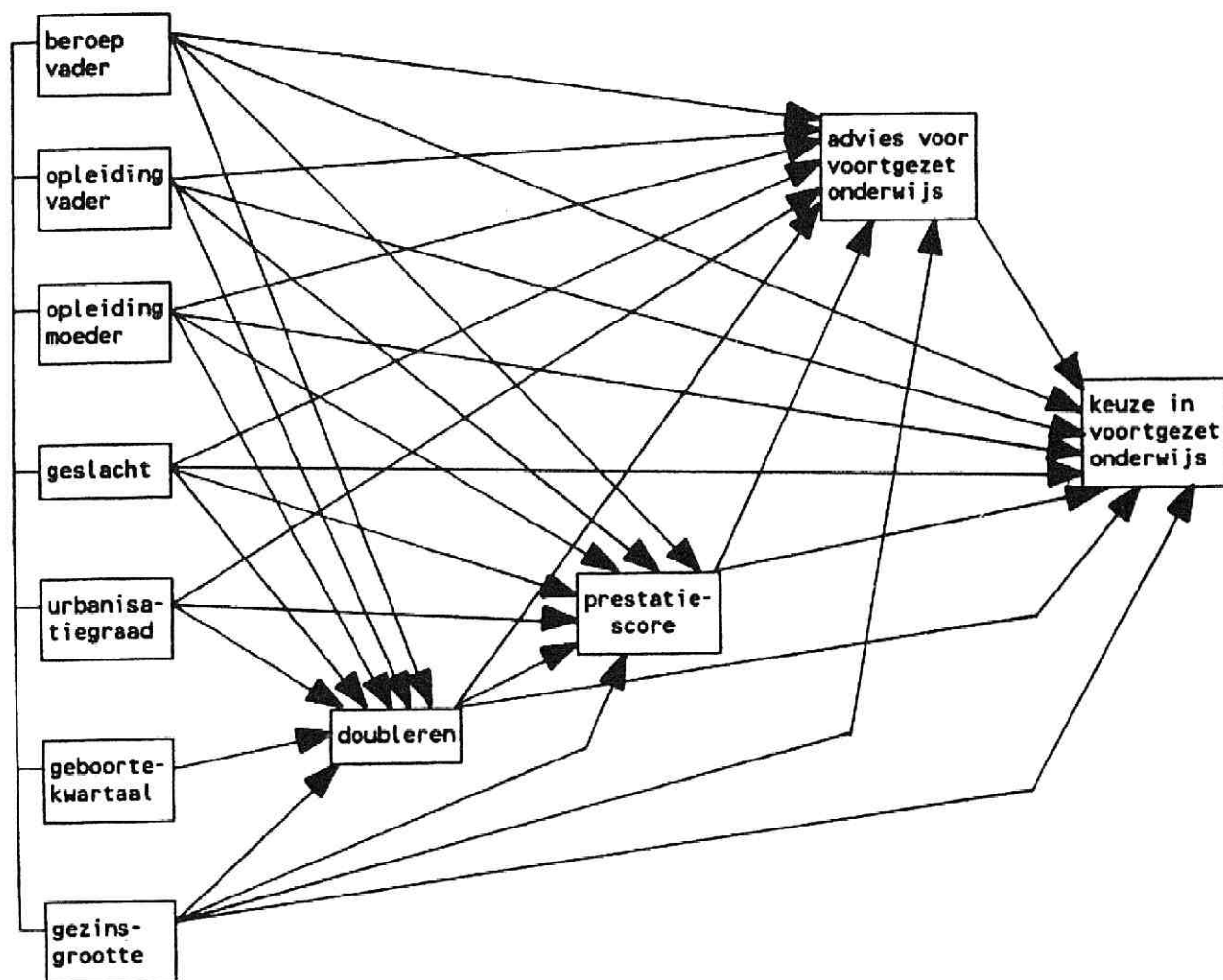
2.3. Het schoolloopbaanmodel en de werkwijze

De in paragraaf 1 gestelde vragen worden beantwoord door de multivariate samenhangen tussen de variabelen te onderzoeken met behulp van PC-LISREL 7 (Jöreskog en Sörbom, 1989). Hiermee kan onder andere worden onderzocht of een gespecificeerd model goed de relaties tussen de verschillende variabelen weergeeft. Bij de ontwikkeling van het schoolloopbaanmodel zijn we uitgegaan van het model van Bakker en Schouten (1991, blz. 10-11). Vervolgens hebben we getoetst of dit model goed de relaties tussen de variabelen weergeeft voor de vier cohorten.

Er zijn drie criteria op grond waarvan de geldigheid van het model wordt beoordeeld. Het

eerste criterium is de Chi^2 in relatie tot het aantal vrijheidsgraden. Gezien de omvang van de steekproeven en de non-normaliteit van de verdelingen is de Chi^2 -toets geen goede graadmeter voor de mate waarin een model bij de gegevens past. Dit is gedeeltelijk opgelost door het aantal waarnemingen per cohort op 1000 te stellen. Dit heeft als bijkomend voordeel dat de Chi^2 voor de vier cohorten vergelijkbaar is en dat de significantie-toetsen ook iets zeggen over de relevantie van de parameters. Het tweede criterium is de hoogte van de gestandaardiseerde residuen: dit geeft aan of en welke samenhangen niet goed door het model worden gereproduceerd. Als vuistregel is gehanteerd dat een gestandaardiseerde residu niet significant mag zijn. De interpretatie van deze gestandaardiseerde residuen is dezelfde als die van t-waarden, dat wil zeggen dat de waarde niet hoger mag zijn dan 1.96. Het derde criterium is de "Adjusted Goodness of Fit Index" (AGFI). Door Verschuren (1991, blz. 305) wordt aangegeven dat het model goed bij de gegevens past als de waarde groter of gelijk is aan .95. We gaan ervan uit dat een redelijke Chi^2 -waarde in combinatie met het afwezig zijn van significante gestandaardiseerde residuen en een AGFI van groter dan .95 tot een passend en betekenisvol model leiden.

Het uitgangsmodel past voor de cohorten geboren rond 1971 en 1977 minder goed bij de gegevens: er bleken significante residuen te bestaan voor de relatie tussen geslacht en advies. Daarom is deze relatie aan het uitgangsmodel toegevoegd. Het resulterende model is weergegeven in figuur 1. In tabel 2 zijn de verschillende criteria weergegeven voor de mate waarin het model (na wijziging) bij de gegevens van de vier cohorten past. We kunnen hieruit concluderen dat het model voor de vier cohorten goed bij de gegevens past.



Figuur 1. Het schoolloopbaanmodel.

Tabel 2. De mate waarin het model bij de gegevens past.

	cohort 1953	1965	1971	1977
chi ² bij 5 vrijheidsgraden	2.9	1.6	3.7	2.1
maximale gestandaardiseerd residu	-.99	-.68	-1.63	.54
AGFI	.99	1.00	.99	.99

Vervolgens zijn de parameters in het model berekend met behulp van de methode van maximale waarschijnlijkheid op basis van de variantie/ covariantiematrices die opgenomen zijn in tabel 5 (achterin dit artikel). De ontbrekende waarden zijn paarsgewijs weggelaten. In overeenstemming met de gangbare praktijk worden uitsluitend ongestandaardiseerde effecten vergeleken, omdat hierin tevens de verschuivingen in de verdelingen zijn opgenomen. Vanwege het gebruik van een 'standard-five'-verdeling voor de verschillende prestatiescores zijn de effecten van deze variabele 'half' gestandaardiseerd. Dit levert voor de interpretatie van de effecten geen noemenswaardige problemen op (Kim en Ferree jr., 1981). Wel wordt een deel van de feitelijke veranderingen, bijvoorbeeld een algemene stijging of grotere of kleinere spreiding van het prestatieniveau, niet in beeld gebracht. Dit kan gevolgen hebben voor de samenhang tussen milieu van herkomst en prestatiescore als de varianties binnen de onderscheiden categorieën voor milieu van herkomst zich verschillend hebben ontwikkeld. We dienen daarom voorzichtig te zijn met de interpretatie van deze samenhangen. Daarnaast worden ook de totale ongestandaardiseerde effecten voor sociaal milieu en prestatiescore op schoolkeuze gebruikt om te toetsen of de totale effecten van sociaal milieu en prestatiescore op schoolsucces over de vier cohorten zijn afgenomen.

We hebben een iets andere methode gebruikt dan onze voorgangers. Zij stelden effecten aan elkaar gelijk indien deze tussen twee cohorten niet significant van elkaar verschilden. Omdat we vier cohorten met elkaar vergelijken beschikken we over de mogelijkheid om toevallige fluctuaties beter te onderscheiden van geleidelijke ontwikkelingen. Indien we de effecten voor de verschillende cohorten aan elkaar gelijk zouden stellen, verdwijnt voor een deel het zicht op deze geleidelijke veranderingen. Daarom hebben we de ongestandaardiseerde effecten voor de vier cohorten direct met elkaar vergeleken. Om na te gaan of ontwikkelingen significant zijn, zijn de standaardfouten die behoren bij de effecten in de tabellen opgenomen.

3. DE RESULTATEN

In tabel 3 zijn de directe ongestandaardiseerde effecten weergegeven. In tabel 4 zijn de totale effecten van de drie indicatoren voor sociaal milieu op doubleren, prestatiescore, advies en keuze en van prestatiescore op advies en keuze in het voortgezet onderwijs weergegeven.

De verklaaringskracht van de variabelen in het schoolloopbaanmodel is toegenomen. De verklaarde variantie in de keuze voor voortgezet onderwijs is sterk gestegen tussen 1953 en 1965, om vervolgens ongeveer gelijk te blijven. De verklaarde variantie in doubleren, prestatiescore en advies schommelen iets, maar blijven grosso modo gelijk. De grootste veranderingen in de relaties tussen de variabelen vinden plaats tussen de cohorten geboren rond 1953 en 1965, met uitzondering van de invloeden van de opleiding van de moeder en doubleren. Daarna zijn de veranderingen marginaal. Bij de bespreking van de resultaten zullen we in eerste instantie uitgaan van de vragen zoals deze in de inleiding zijn geformuleerd.

De eerste vraag, of de leerprestaties van leerlingen uit lagere milieus aan het eind van de basisschool in de loop der tijd ten opzichte van de overige sociale milieus zijn verbeterd, kan gedeeltelijk bevestigend worden beantwoord. De invloed van het beroep van de vader is in de loop der tijd afgenomen: was deze invloed voor het cohort 1953 nog vrij sterk, voor het cohort

Tabel 3. Ongestandaardiseerde effecten (standaardfouten) in het schoolloopbaanmodel.

	doubleren				prestatiescore				advies				keuze			
	1953	1965	1971	1977	1953	1965	1971	1977	1953	1965	1971	1977	1953	1965	1971	1977
beroep vader	.00 (.01)	-.02 (.01)	-.02 (.01)	-.02 (.01)	.11 (.02)	.05 (.02)	.05 (.02)	.03 (.02)	.01 (.02)	.03 (.01)	.03 (.01)	.02 (.01)	.04 (.02)	.01 (.01)	.01 (.01)	.00 (.01)
opleiding vader	-.05 (.02)	-.03 (.01)	-.01 (.01)	-.02 (.01)	.13 (.03)	.09 (.02)	.09 (.02)	.09 (.02)	.09 (.02)	.07 (.02)	.07 (.02)	.06 (.02)	.05 (.02)	.03 (.01)	.05 (.01)	.03 (.01)
opleiding moeder	-.02 (.02)	-.02 (.01)	-.02 (.01)	-.03 (.01)	.06 (.03)	.09 (.03)	.09 (.02)	.12 (.02)	.06 (.03)	.05 (.02)	.05 (.02)	.06 (.02)	.02 (.03)	.03 (.01)	.02 (.01)	.03 (.01)
urbanisatiegraad	.04 (.01)	.02 (.01)	.01 (.01)	.03 (.01)	.03 (.01)	-.01 (.01)	-.03 (.01)	-.02 (.01)	-.01 (.01)	.03 (.01)	.03 (.01)	.03 (.01)				
geboortekwartaal	.12 (.03)	.12 (.02)	.07 (.02)	.15 (.02)												
geslacht	-.11 (.04)	-.13 (.03)	-.08 (.03)	-.13 (.03)	-.09 (.06)	-.11 (.06)	-.12 (.06)	-.10 (.06)	.05 (.05)	.08 (.04)	.15 (.05)	.12 (.04)	-.13 (.05)	.06 (.03)	.04 (.03)	.09 (.03)
gezinsgrootte	.04 (.02)	.02 (.02)	.04 (.02)	.02 (.02)	-.05 (.03)	-.05 (.03)	-.03 (.03)	-.01 (.04)	-.01 (.03)	-.06 (.02)	-.03 (.03)	.03 (.03)	-.04 (.02)	-.01 (.02)	-.02 (.02)	.01 (.02)
doubleren					-.41 (.05)	-.59 (.06)	-.50 (.07)	-.55 (.06)	-.05 (.04)	-.18 (.04)	-.28 (.06)	-.19 (.04)	-.12 (.04)	-.08 (.03)	-.06 (.04)	-.01 (.03)
prestatiescore									.65 (.03)	.57 (.02)	.59 (.03)	.58 (.02)	.22 (.03)	.14 (.02)	.14 (.02)	.12 (.02)
advies													.52 (.03)	.74 (.02)	.70 (.02)	.69 (.02)
variantie verklaard	.06	.09	.06	.11	.19	.22	.17	.24	.50	.55	.52	.54	.59	.79	.74	.75

Tabel 4. Ongestandaardiseerde totale effecten (standaardfouten) van milieu van herkomst en geslacht op doubleren, prestatiescore, advies en eerste schoolkeuze in het voortgezet onderwijs.

	doubleren				prestatiescore				advies				keuze			
	1953	1965	1971	1977	1953	1965	1971	1977	1953	1965	1971	1977	1953	1965	1971	1977
beroep vader	.00 (.01)	-.02 (.01)	-.02 (.01)	-.02 (.01)	.11 (.02)	.06 (.02)	.06 (.02)	.04 (.02)	.09 (.02)	.06 (.02)	.07 (.02)	.05 (.02)	.11 (.02)	.06 (.02)	.06 (.02)	.04 (.02)
opleiding vader	-.05 (.02)	-.03 (.01)	-.01 (.01)	-.02 (.01)	.14 (.03)	.11 (.02)	.09 (.02)	.10 (.02)	.19 (.03)	.14 (.02)	.13 (.02)	.13 (.02)	.19 (.03)	.15 (.02)	.15 (.02)	.13 (.02)
opleiding moeder	-.02 (.02)	-.02 (.01)	-.02 (.01)	-.03 (.01)	.07 (.04)	.11 (.03)	.10 (.02)	.13 (.02)	.10 (.04)	.12 (.02)	.11 (.02)	.15 (.02)	.09 (.04)	.13 (.02)	.12 (.02)	.15 (.02)
prestatiescore									.65 (.03)	.57 (.02)	.59 (.03)	.58 (.02)	.56 (.03)	.55 (.02)	.55 (.03)	.52 (.02)

1977 is deze invloed vrijwel verdwenen. Ook is de invloed van de opleiding van de vader op de prestatiescore afgenomen, zij het minder sterk dan de invloed van het beroep van de vader. Ook voor het cohort 1977 is de invloed van het opleidingsniveau van de vader nog significant. De afgenomen invloed van het beroep en de opleiding van de vader, wordt voor een deel gecompenseerd door de sterker geworden invloed van de opleiding van de moeder. Dit effect is voor het cohort 1977 het dubbele van die voor het cohort 1953. Daarbij dient aangetekend te worden dat deze resultaten voorzichtig geïnterpreteerd dienen te worden in verband met de meting van schoolprestaties in een standard-five-verdeling. Dit kan tot gevolg hebben dat de gevonden daling van de invloed van het beroep en de opleiding van de vader en de stijging van de invloed van de opleiding van de moeder iets anders ligt dan door ons gevonden.

De tweede vraag, of het belang van de leerprestaties aan het eind van de basisschool voor het advies en de keuze in het voortgezet onderwijs toegenomen is, moet ontkennend worden beantwoord. De prestatiescores aan het einde van de basisschool zijn juist van minder belang geworden voor het advies en de keuze in het voortgezet onderwijs. Daarbij moet worden aangetekend dat de keuzes voor voortgezet onderwijs in de loop der tijd sterker op het advies worden gebaseerd dan voorheen, zij het dat deze toename slechts geldt voor de eerste twee cohorten. Daarna is deze invloed nagenoeg constant gebleven. Dit heeft echter wel tot gevolg dat de totale effecten van de prestaties op de keuze in het voortgezet onderwijs nagenoeg gelijk zijn gebleven.

De derde vraag, of de onderwijskansen van de lagere milieus ten opzichte van de hogere milieus zijn verbeterd, kan gedeeltelijk bevestigend worden beantwoord. Het directe effect van het beroep van de vader op de keuze in het voortgezet onderwijs is verdwenen, terwijl de effecten van de opleidingen van de vader en de moeder nagenoeg gelijk zijn gebleven. Deze effecten zijn weliswaar significant, maar niet groot. Dat wil zeggen dat één van de indicatoren voor milieu van herkomst in belang is afgenomen en twee in belang gelijk zijn gebleven, bij gelijke doubleerstatus, prestatiescores en adviezen. In de totale effecten vinden we ook een gedeeltelijk bevestigend antwoord van de vraag. De invloed van de kenmerken van de vader zijn afgenomen en de invloed van de opleiding van de moeder is toegenomen. We kunnen hieruit concluderen dat de invloed van het milieu van herkomst in de loop der tijd enigszins is afgenomen.

Deze verbetering van de onderwijskansen van de lagere milieus kan niet worden verklaard door meritocratisering in het onderwijs. Het advies voor en de keuze in het voortgezet onderwijs worden in de loop der tijd juist minder door de prestaties bepaald. Wel is het belang van het advies voor de keuze in het voortgezet onderwijs toegenomen, waardoor een aan meritocratisering tegengestelde tendens goeddeels achterwege is gebleven. Opvallend is in dit verband dat de onderwijskansen van de lagere milieus voor het cohort 1965 (na de mammoetwet) sterk verbeterd zijn ten opzichte van het cohort 1953. Daarna zijn de veranderingen in onderwijskansen aanzienlijk geringer.

4. CONCLUSIES EN DISCUSSIE

In dit artikel is de overgang naar het voortgezet onderwijs van de cohorten geboren rond 1953, 1965, 1971 en 1977 vergeleken. Daarbij is vooral gelet op de veranderingen in de invloeden van sociaal milieu. Op grond van de resultaten wordt geconcludeerd dat de onderwijskansen van de lagere milieus ten opzichte van de hogere milieus zijn toegenomen. Dit is niet het gevolg van het verbeteren van de schoolprestaties van leerlingen uit lagere sociale milieus of van een toename van het belang van de prestaties voor het advies en de keuze. De schoolprestaties van leerlingen uit lagere milieus zijn in vergelijking met de hogere sociale milieus slechts marginaal verbeterd, en het belang van schoolprestaties voor de keuze in het voortgezet onderwijs is eerder afgenomen dan toegenomen.

In eerder onderzoek concludeerden Bakker en Schouten (1991) dat het belang van de prestatiescore voor het advies voor en de keuze van het vervolgonderwijs wel toenam. Ons inziens is

dit een gevolg van een verschil in methode. Hun methode bestond eruit om de effecten die tussen de cohorten niet significant verschilden aan elkaar gelijk te stellen. Als er slechts drie cohorten in het onderzoek zijn betrokken, is de keuze voor deze methode plausibel. Op grond van de resultaten van drie cohorten kunnen toevallige fluctuaties niet goed worden onderscheiden van geleidelijke ontwikkelingen. Voor de cohorten 1965 en 1971 vonden zij vrijwel geen significante verschillen tussen de effecten, zodat deze aan elkaar gelijk gesteld konden worden. Door nu, zoals in ons geval, het cohort 1977 toe te voegen is het minder zinvol om dergelijke beperkingen aan de veranderingen op te leggen. Hierdoor zouden immers geleidelijke ontwikkelingen die op zichzelf geen significante verschillen tussen de cohorten te zien geven onzichtbaar worden.

De resultaten van Dronkers (1992) komen voor een groot deel met de onze overeen. Een verschil is dat hij concludeert dat het advies vaker wordt opgevolgd dan vroeger. Dit is juist voor de cohorten 1953 en 1965, waarvoor blijkt dat de invloed van het advies op de keuze fors is toegenomen. Voor de latere cohorten blijkt echter de invloed weer iets te zijn gedaald.

De resultaten van de vergelijking van vier cohorten die tezamen een periode van 24 jaar beslaan, vertonen ook een grote overeenkomst met de resultaten uit dwarsdoorsnede-onderzoek waarin het beroep en de opleiding van de vader worden gerelateerd aan het opleidingsniveau van de kinderen (Ganzeboom en De Graaf, 1989a en b; De Graaf en Luijkx, 1992). Zij onderzochten een veel langere periode, vanaf ongeveer 1900, en komen tot de conclusie dat de invloed van het beroep en de opleiding van de vader op het schoolsucces van zijn kinderen is afgenomen. De invloed van de opleiding van de moeder is door De Graaf en zijn collega's niet expliciet onderzocht. Eén van de belangrijkste conclusies uit deze studie is dat de invloed van de opleiding van de moeder juist is toegenomen; hierdoor wordt de afgenomen invloed van de kenmerken van de vader gedeeltelijk gecompenseerd. Dit lijkt erop te wijzen dat selectieprocessen in het onderwijs los van de feitelijke prestaties nog steeds voor een deel afhankelijk zijn van het milieu van herkomst, maar dat leerkrachten in de loop der tijd meer belang zijn gaan hechten aan de kenmerken van de moeder. De toegenomen effecten van de opleiding van de moeder kunnen worden verklaard uit de dominantier geworden positie van moeders binnen het gezin. Dit uit zich onder andere door een grotere bemoeienis met de opleiding en schoolsituatie van hun kinderen.

De afname van de invloed van het beroep van de vader op de opleiding van de kinderen, zoals dat in dwarsdoorsnede-onderzoek wordt gevonden, is mogelijk een gevolg van andere processen dan meritocratisering. Een mogelijke verklaring is de toegenomen onderwijsdeelname, waardoor kinderen op een latere leeftijd een keuze voor vervolgonderwijs maken, die daardoor onafhankelijker van het ouderlijk milieu is. Door Dronkers (1992, blz. 68-69) wordt aangetoond dat een vermindering van het directe effect van het sociaal milieu van herkomst ontstaat doordat een groter percentage leerlingen hogere vormen van voortgezet onderwijs gaat volgen. Mare (1981) toont aan dat correlaties kunnen afnemen als gevolg van deze gegroeide onderwijsdeelname.

Met nadruk wordt gesteld dat hiermee geen definitieve conclusies over de invloed van prestaties of sociaal milieu op advies en schoolkeuze worden geveld. Uit ons onderzoek blijkt duidelijk dat de resultaten gevoelig zijn voor het gebruikte datamateriaal en de gebruikte analysemethode. Verder onderzoek is dan ook noodzakelijk om tot meer definitieve conclusies hierover te geraken. Daarbij is van belang dat zoveel mogelijk cohorten die tezamen een zo groot mogelijke tijdsspanne omvatten in het onderzoek worden betrokken, zodat toevallige fluctuaties kunnen worden onderscheiden van werkelijke ontwikkelingen. Daarnaast kunnen de eigenaardigheden van de gegevens over één cohort beter worden opgespoord als er meer cohorten in het onderzoek worden betrokken.

Tabel 5. Variantie-covariantie-matrices voor de vergelijking van de onderwijskansen van vier cohorten leerlingen.

<i>VJTJ (cohort 1953)</i>											
	keuze	advies	prest	doubl	bervad	oplvad	oplmoe	urban	gebkwrt	geslacht	gezgro
keuze	1.252										
advies	.913	1.250									
prest	.760	.821	1.143								
doubl	-.162	-.143	-.168	.381							
bervad	.676	.611	.605	-.064	3.028						
oplvad	.594	.570	.496	-.091	1.456	1.952					
oplmoe	.314	.311	.261	-.038	.724	.729	1.111				
urban	.080	.128	.207	.204	.521	.348	.305	5.731			
gebkwrt	-.009	-.026	-.019	.055	-.023	-.003	.016	-.017	.484		
geslacht	-.029	.006	-.011	-.024	-.014	.004	-.001	.047	.005	.250	
gezgro	-.081	-.052	-.063	.017	-.031	-.026	.003	-.362	-.018	-.007	.882
<i>SMVO (cohort 1965)</i>											
keuze	1.063										
advies	.902	1.000									
prest	.752	.736	1.100								
doubl	-.199	-.184	-.198	.284							
bervad	.770	.730	.640	-.187	4.427						
oplvad	.756	.710	.607	-.167	2.294	3.165					
oplmoe	.503	.463	.412	-.106	1.282	1.213	1.943				
urban	.014	.065	-.099	.106	-.081	-.036	-.082	5.094			
gebkwrt	-.035	-.038	-.024	.060	.025	.020	.022	-.003	.500		
gesl	.034	.023	-.007	-.032	.008	.010	.009	-.007	.000	.250	
gezgro	-.051	-.057	-.013	-.002	.172	.137	.117	-.336	-.011	-.003	1.024
<i>SLVO (cohort 1971)</i>											
keuze	1.114										
advies	.942	1.107									
prest	.713	.736	1.071								
doubl	-.129	-.133	-.112	.177							
bervad	.821	.772	.624	-.145	4.674						
oplvad	.835	.761	.598	-.106	2.526	3.519					
oplmoe	.539	.501	.422	-.081	1.263	1.351	2.239				
urban	-.053	-.031	-.203	.060	-.242	-.208	-.189	4.318			
gebkwrt	-.040	-.051	-.044	.031	.003	.012	.005	.002	.488		
geslacht	.028	.029	-.021	-.019	.004	-.004	-.009	.004	.009	.250	
gezgro	-.090	-.086	-.058	.031	-.158	-.089	-.081	-.121	-.015	.002	.781
<i>VOCL (cohort 1977)</i>											
keuze	.986										
advies	.881	1.069									
prest	.697	.759	1.102								
doubl	-.175	-.194	-.199	.291							
bervad	.740	.758	.657	-.201	5.128						
oplvad	.813	.817	.699	-.184	2.528	3.615					
oplmoe	.635	.629	.555	-.142	1.275	1.586	2.501				
urban	-.071	-.067	-.241	.144	-.507	-.245	-.236	4.280			
gebkwrt	-.033	-.051	-.038	.075	-.007	-.021	.002	.012	.492		
geslacht	.043	.032	-.007	-.033	.007	-.009	-.004	-.016	-.002	.250	
gezgro	.009	.008	-.007	.013	.010	.000	.015	-.107	-.000	-.001	.690

LITERATUUR

- Arts, W.A., & Ultee, W.C. (1985). Tussen radicalisering en apathie. Maatschappelijke tweedeling als sociaal-wetenschappelijk probleem, In *Economisch Statistische Berichten*, 70, nr. 3494, 183-187.
- Bakker, B.F.M. (1992). *Sociale afstand, sociale status* (Proefschrift Universiteit van Amsterdam, Nijmegen: ITS).
- Bakker, B., Dronkers, J., & Schijf, H. (1982). Veranderingen in individuele schoolloopbanen tussen 1959 en 1977 in de stad Groningen, *Mens en Maatschappij*, 57 (3), 253-267.
- Bakker, B.F.M., & Schouten, S.P. (1991). Trends in onderwijskansen. Een vergelijking van de overgang naar het voortgezet onderwijs van de generaties geboren rond 1953, 1965 en 1971, In *Sociale Wetenschappen*, 34 (1), 1-21.
- Bourdieu, P. en Passeron, J.C. (1977) [1970]. *Reproduction* Londen: SAGE.
- CBS (Centraal Bureau voor de Statistiek) (1970). *Statistisch zakboek 1970*. 's-Gravenhage: Staatsuitgeverij.
- CBS (Centraal Bureau voor de Statistiek) (1987). *Arbeidskrachtentelling 1985*. 's-Gravenhage: Staatsuitgeverij.
- CBS (Centraal Bureau voor de Statistiek) (1989). *Enquête beroepsbevolking 1988*. Voorburg/Heerlen: CBS.
- CBS (Centraal Bureau voor de Statistiek) (1991). *Statistisch jaarboek 1991*. 's-Gravenhage: SDU/CBS-publikaties.
- Dronkers, J. (1992). Komt de afname van het belang van het sociaal milieu in het onderwijs door vergroting van onderwijsdeelname of door meritocratisering?, In P. Dykstra, P. Kooij en J. Rupp (red.), *Onderwijs in de tijd. Ontwikkelingen in onderwijsdeelname en nationale curricula*. Boekaflevering Mens en Maatschappij, Houten/Zaventem: Bohn Stafleu Van Loghum.
- Dronkers, J., & Bosma, H. (1990). The consequences of educational reform for the unequal educational opportunities of children of different social classes in secondary schools, In *Zeitschrift für Sozialisationsforschung und Erziehungssoziologie*, 10, 268-278.
- Faasse, J.H., Bakker, B., Dronkers, J., & Schijf, H. (1986). Voor en na de mammoet. Vergelijking van individuele schoolloopbanen in het voortgezet onderwijs in Noord-Brabant voor en na de invoering van de Mammoetwet, *Mens en Maatschappij*, 61 (1), 75-94.
- Ganzeboom, H.B.G., & Graaf, P.M. de (1989a). Intergenerationele opleidingsmobiliteit in Nederland van geboortecohorten 1891-1960, In *Sociale Wetenschappen*, 32 (4), 263-278.
- Ganzeboom, H.B.G., & Graaf, P.M. de (1989b). Veranderingen van onderwijskansen in Nederland tussen 1900 en 1980, In I. Gadourek en J.L. Peschar (red.), *De open samenleving?* Boekaflevering Mens en Maatschappij Deventer: Van Loghum Slaterus.
- Graaf, P.M. de, & Luijkx, R. (1992). Van 'ascription' naar 'achievement'? Trends in statusverwerving in Nederland tussen 1930 en 1980, In *Mens en Maatschappij*, 67 (4), 412-433.
- Jong, U. de, Dronkers, J., & Saris, W.E. (1982). Veranderingen in de schoolloopbanen tussen 1965 en 1977: ontwikkelingen in de Nederlandse samenleving en haar onderwijs, *Mens en Maatschappij*, 57 (1), 26-54.
- Jöreskog, K., & Sörbom, D. (1989) LISREL 7. *A guide to the program and applications*. Chicago: SPSS.
- Kim, J., & Ferree, G.D. jr. (1981). Standardization in causal analysis, In *Sociological Methods and Research*, 10 (2), 187-210.
- Mare, R.D. (1981). Change and stability in educational stratification, In *American Sociological Review*, 46 (1) 72-87.
- Meester, A., & Leeuw, J. de (1983). *Intelligentie, sociaal milieu en de schoolloopbaan*. Leiden: Vakgroep Datatheorie.
- Meesters, M., Dronkers, J., & Schijf, H. (1983). Veranderde onderwijskansen? Een derde voorbeeld en afrondende conclusies, *Mens en Maatschappij*, 58 (1), 5-27.
- Parkin, F. (1979). *Marxism and class theory: a bourgeois critique* (Londen: Tavistock).
- Tesser, P., Mulder, L., & Werf, G. van der (1991). *De eerste fase van de longitudinale OVB-onderzoeken. Het leerlingenonderzoek*. Nijmegen/Groningen: ITS/RION.
- Verschuren, P.J.M. *Structurele modellen tussen theorie en praktijk*. Utrecht: Spectrum.
- Vrooman, J.C., & Dronkers, J. (1986). Changing educational attainment processes: some evidence from the Netherlands, In *Sociology of Education*, 59 (2).
- Willemse, P. (1987). Overleving in de bovenstroom van het voortgezet onderwijs: veranderde milieu-effecten voor jongens en meisjes, *Tijdschrift voor Onderwijsresearch*, 12 (6), 29-343.

Manuscript ontvangen 10-12-1993

Definitieve versie ontvangen 3-5-1994

Effecten van streefniveaus in het basisonderwijs op schoolloopbanen in het voortgezet onderwijs

A.A. van der Hoeven-van Doornum

*Instituut voor Toegepaste Sociale Wetenschappen, Nijmegen**

ABSTRACT

In a previous study it was found that parents' expectations and those of teachers influence school careers of pupils in primary schools. Children may enter higher or lower levels of secondary education depending on the aspiration levels their parents and teachers set for them. The aim of the present study was to investigate the influence of aspiration levels on the school careers of children in secondary education, using the data from the previous longitudinal study.

Analyses were performed on data from a cohort of 351 pupils (10-16 year old) who were followed during their last two years of primary education and four years later in secondary education. As indicators for expectations the levels of secondary education thought to be appropriate for the child by their parents and teachers, as well as cognitive goals set by teachers were used. The indicator for the school career was constructed by the level and the grade of secondary education.

It appeared that aspiration levels set by parents and teachers during primary education have an sustaining effect on the school careers of the children, taking into account prior learning achievement. High expectations of teachers have an indirect effect, via the advice at the end of primary education, on the level of secondary education. High expectations of parents still have a direct effect on the pupils' level and grade in secondary education. In this sense parents' and teachers' expectations represent a form of effectiveness. SES-related differences between school careers in secondary education appeared to be set during primary education. Apparently the advice at the end of primary education is the terminal determinant in SES-related school careers.

INLEIDING

Het onderzoek naar de invloed van effectief basisonderwijs op schoolloopbanen in het voortgezet onderwijs is een vervolgstudie van het onderzoek 'Effecten van leerlingbeelden en streefniveaus op schoolloopbanen' (Van der Hoeven-van Doornum, 1990). Het gaat in de studie in het voortgezet onderwijs om metingen bij dezelfde leerlingen als die in het basisonderwijs zijn onderzocht. Uit het longitudinale onderzoek in het basisonderwijs is gebleken dat in de laatste twee leerjaren van de basisschool leerkrachten verschillende streefniveaus voor hun leerlingen hanteren en daarmee de leerprestaties van hun leerlingen in de loop van de tijd beïnvloeden. Verschillen in streefniveaus berusten niet alleen op verschillen tussen leerlingen in intelligentie, maar ook op verschillen in het sociale milieu waaruit leerlingen afkomstig zijn. Streefniveaus geven een verklaring voor verschillen in schoolprestaties en vervolgadvis die verder gaat dan de tot nu toe in schoolloopbaanonderzoek geconstateerde samenhang van SES en IQ met het vervolgadvis. In een latere studie is gebleken is dat niet alleen de streefniveaus van leerkrachten invloed hebben op de schoolloopbanen van hun leerlingen, maar dat ook de voorkeur van ouders voor een bepaald type vervolgonderwijs voor hun kind bijdraagt aan de verklaring van verschillen in leerprestaties en in adviezen voor voortgezet onderwijs (Van der Hoeven-van Doornum, Voeten en Jungbluth, 1993).

* Adres: Postbus 9048, 6500 KJ Nijmegen.

Streefniveaus

Streefniveaus van leerkrachten zijn grotendeels gebaseerd op de inschattingen die leerkrachten hebben van het leervermogen en de sociale achtergrond van de leerlingen. Op grond van hun ervaring en inzicht geven leerkrachten aan welke doelen volgens hen voor de leerling bereikbaar zijn. De term 'streefniveau' verwijst zowel naar een (vroegtijdig) anticiperen door leerkrachten op het niveau van voortgezet onderwijs dat een leerling kan bereiken, als naar het niveau van leerstofbeheersing dat daarvoor nodig zou zijn, bijvoorbeeld wel of niet uitgebreid leren zinsontleden. Het gaat, in het geval van leerstofbeheersing, niet om specifieke leerdoelen voor rekenen of voor taal, maar om de mate waarin een leerkracht hoge eisen aan een leerling stelt, nadruk legt op taal en rekenen en het belangrijk vindt om een leerling voor te bereiden op hogere vormen van voortgezet onderwijs.

In dit onderzoek wordt het anticiperen op het niveau van voortgezet onderwijs opgevat als een absoluut streefniveau en de mate van nagestreefde leerstofbeheersing, aangeduid als cognitieve onderwijsdoelen, als een relatief streefniveau. Cognitieve onderwijsdoelen zijn gerelateerd aan het dagelijkse onderwijs aan en de pedagogische omgang met de leerlingen, zij hebben een pedagogisch-didactisch functie; terwijl de verwachting van leerkracht over het meest geschikte type voortgezet onderwijs aangeeft welk niveau van voortgezet onderwijs voor een leerling haalbaar is. Het onderscheid tussen absolute en relatieve prestatieniveaus is om verschillende redenen van belang (Van der Hoeven-van Doornum, Voeten en Jungbluth, 1990; Van der Hoeven-van Doornum 1990):

- Hoge cognitieve doelen gaan niet noodzakelijk gepaard met het streven naar hoge vormen van voortgezet onderwijs.
- Cognitieve onderwijsdoelen en het geanticipeerd schooltype verschillen in de mate waarin zij onderhevig zijn aan sociale bias. Cognitieve onderwijsdoelen zijn tamelijk ongevoelig voor factoren die met het sociale milieu samenhangen, terwijl het geanticipeerd schooltype wel gevoelig is voor milieu-effecten.
- Beide streefniveaus hebben invloed op latere leerprestaties. De cognitieve onderwijsdoelen in leerjaar 7 beïnvloeden echter het vervolgadvis niet, en in leerjaar 8 alleen indirect via leerprestaties. Het geanticipeerd schooltype heeft daarentegen in beide leerjaren wel directe en indirecte effecten op het vervolgadvis.

Waarschijnlijk vormen de verwachtingen die ouders over en voor hun kinderen hebben, een van de belangrijkste kenmerken van ouderschap en opvoeding. Gedurende de opvoeding geven ouders voortdurend – impliciet en expliciet, afhankelijk van hun sociale en culturele achtergrond – de verwachtingen weer die zij voor hun kind koesteren. Dit gebeurt in allerlei dagelijkse activiteiten zoals spelen, lezen, en praten met hun kinderen, waarin zij hogere of lagere eisen aan hun kinderen stellen en daarmee de cognitieve ontwikkeling van kinderen in meerdere of mindere mate stimuleren (Rupp, 1969; Tesser, 1991). De expliciete voorkeur van ouders voor een bepaald type vervolgonderwijs voor hun kind, kan worden vergeleken met het anticiperen van de leerkracht op een lager of hoger type van voortgezet onderwijs voor het kind. In eerder onderzoek is aangetoond dat, rekening houdend met intelligentie en eerdere leerprestaties van het kind, dergelijke verwachtingen van ouders invloed hebben op de schoolloopbaan van hun kind en op het schooltype waarop de leerkracht anticipeert (Van der Hoeven-van Doornum, Voeten en Jungbluth, 1993).

School- en leerkrachteffectiviteit

Het begrip schooleffectiviteit wordt gehanteerd om verschillen in variantie tussen scholen of leerkrachten aan te geven met betrekking tot cognitieve prestaties, niet-cognitieve vaardigheden en het advies voor het voortgezet onderwijs (doorstroom). Of een school meer of minder effectief is hangt samen met het effectiviteitscriterium dat wordt gekozen (Bosker, 1992). In ons onderzoek op de basisscholen bleek bijvoorbeeld dat, onder constanthouding van achtergrondkenmerken en eerdere leerprestaties van de leerlingen, verschillen in streefniveaus 15,5% en verschillen in schoolkenmerken 11,6% van de totale variantie in de scores op de Cito Eindtoets

verklaren; terwijl de streefniveaus en schoolkenmerken respectievelijk 51% en 7% van de totale variantie van het vervolgadvis verklaren. Verschillen in streefniveaus leiden tot verwijzingen van leerlingen naar hogere of lagere vormen van voortgezet onderwijs. In dit onderzoek concentreren we ons op de systematische verschillen in vervolgadvis als gevolg van de verschillen in streefniveaus, omdat de overgang van het basisonderwijs naar het voortgezet voor een kind een van de belangrijkste momenten in zijn of haar schoolloopbaan is. De verschillen in doorstroom naar het voortgezet beschouwen we als verschillen in effectiviteit tussen scholen, c.q. leerkrachten.

Onderwerp van het vervolgonderzoek is de duurzaamheid van het effect van streefniveaus. Het is bekend dat verschillen in onderwijsposities in het voortgezet onderwijs samenhangen met de sociale herkomst van leerlingen. Deze samenhang wordt grotendeels teruggevoerd op indirecte effecten van sociaal milieu via schoolprestaties en advies in het basisonderwijs. Het doel van het vervolgonderzoek is na te gaan wat de relevantie is van meetbare effectiviteitsverschillen in het basisonderwijs voor onderwijskansen in het voortgezet onderwijs. De crux van het onderhavige onderzoek ligt in de vraag of en in welke mate de werking van streefniveaus een gunstige of ongunstige invloed heeft gehad op de verdere schoolloopbaan van de leerlingen. In termen van effectiviteit gaat het om de vraag of het voor latere schoolprestaties uitmaakt of een leerling op een meer of minder effectieve basisschool heeft gezeten. Inzicht in de mate waarin deze effectiviteitsverschillen doorwerken in het voortgezet onderwijs is van belang voor de discussie over schooleffectiviteit in de komende jaren.

2. ONDERZOEKSVRAGEN

De onderzoeksvraag in het vervolgonderzoek is of de eertijds gehanteerde streefniveaus een extra bijdrage leveren aan de verklaring van meer of mindere succesvolle schoolloopbanen in het voortgezet onderwijs. Bij de verklaring van de positie van leerlingen in het voortgezet onderwijs wordt rekening gehouden met de verschillen in achtergrondkenmerken en leerprestaties in het basisonderwijs. Er zijn twee onderzoeksvragen in het vervolgonderzoek:

- Heeft de invloed van streefniveaus op leerprestaties en vervolgadvis in het basisonderwijs, ook een langdurige werking in het voortgezet onderwijs?
- Dragen streefniveaus in het basisonderwijs bij aan de verklaring van de samenhang tussen het sociale milieu en de onderwijspositie in het voortgezet onderwijs?

Afhankelijk van de beantwoording van deze twee onderzoeksvragen zal in vervolgonderzoek kunnen worden nagegaan of de effectiviteit van de bezochte basisschool een eigen (blijvende) invloed heeft op het schoolsucces in het voortgezet onderwijs.

3. DE ONDERZOEKSOPZET

De dataverzameling in het voortgezet onderwijs is tweeërlei: bij de leerlingen is informatie ingewonnen over hun onderwijspositie en bij de scholen voor voortgezet onderwijs is naar schoolkenmerken gevraagd. Voor de leerlingen is een nieuw instrument ontwikkeld, dat is toegespitst op de schoolloopbaan van 15- en 16-jarige leerlingen in het voortgezet onderwijs. Bij de ontwikkeling van de instrumenten is ook gebruik gemaakt van reeds bestaande instrumenten uit onderzoeken op het gebied van onderwijsvoorrangsbeleid en onderwijsposities van meisjes. De gegevens van het vervolgonderzoek zijn aan het bestand van het onderzoek op de basisscholen toegevoegd en zijn geanalyseerd met behulp van de Lisrel-procedure voor structurele modellen, waarbij de leerling het analyseniveau vormt. In een tweede serie multiniveau-analyses zullen ook kenmerken op schoolniveau worden betrokken.

4. DATAVERZAMELING

Van het basisonderwijs naar het voortgezet onderwijs

In 1991 is de zoekactie naar de leerlingen begonnen met het aanschrijven van de basisscholen met het verzoek om het huisadres en het adres van de school voor voortgezet onderwijs waar de leerlingen in 1988 naartoe zijn gegaan. Van de 53 basisscholen hebben 46 scholen (87%) meegewerkt aan deze zoekactie door ons de adressen van de leerlingen te geven en/of de adressen van de scholen waar de leerlingen in 1988 naartoe zijn gegaan. Het totale aantal leerlingen dat in de periode 1986-1988 aan het onderzoek heeft meegedaan is 1149. In de loop van deze twee schooljaren zijn er de nodige wisselingen geweest in de onderzoeksgroep. Het aantal leerlingen dat op *alle* relevante metingen in groep 7 en groep 8 een score heeft is 701.

In 1992 is de dataverzameling voor het vervolgonderzoek gestart. De dataverzameling is in twee rondes verlopen. Gezien het verloop van leerlingen gedurende twee jaar basisonderwijs, is het niet verwonderlijk dat we vier jaar later bij de dataverzameling in het voortgezet onderwijs met een nog groter aantal verhuizingen en wisselingen van scholen werden geconfronteerd. De respons van de leerlingen bedroeg 64% (N=588).

Van 10-16 jaar

Voor de beantwoording van de hierboven genoemde onderzoeksvragen is de variabele 'positie in het voortgezet onderwijs' gemaakt. De variabele is geconstrueerd volgens de Groningse leerjarenladder (Bosker, 1990). Door toevoeging van de gegevens over het voortgezet onderwijs aan de gegevens uit de basisschoolperiode krijgen we een overzicht van de schoolloopbanen van de leerlingen van 10-16 jaar. Vanwege de non-respons in het voortgezet onderwijs is het niet zo dat voor alle 701 leerlingen van wie we een 'volledig' databestand uit het basisonderwijs hebben, ook de gegevens uit het voortgezet onderwijs compleet zijn. Zo is bijvoorbeeld van 438 leerlingen uit het bestand van 701 de onderwijspositie in het voortgezet onderwijs bekend. Vergelijking van gemiddelden en standaarddeviaties (t-toets) van de oorspronkelijke steekproef van 701 (b.o.) en de steekproef van 438 (v.o.) liet zien dat de laatste groep op alle variabelen hoger scoort. De verschillen tussen de eerste en de laatste steekproef zijn significant, behalve voor de variabele die de sociale herkomst van de leerlingen aanduidt (SES). Dit betekent dat de non-respons in het voortgezet onderwijs geen selectie naar sociale afkomst heeft opgeleverd, maar wel naar leerlingen die een minder succesvolle schoolloopbaan achter de rug hebben.

Voorkeur van de ouders

Vervolgens is het databestand gereed gemaakt voor de uit te voeren Lisrel-analyses. Bij controle van de data bleek dat van 87 leerlingen de ouders indertijd, begin leerjaar 7, geen uitgesproken voorkeur hadden voor een bepaald type voortgezet onderwijs voor hun kind. Deze leerlingen zijn uit het databestand verwijderd. Bij vergelijking (t-toets) van de gemiddelden en standaarddeviaties van de in de analyses te gebruiken variabelen, bleek de steekproef van 351 leerlingen steeds significant hoger te scoren dan de steekproef van 701 leerlingen aan het einde van de basisschool en de steekproef van 438 leerlingen in het voortgezet onderwijs, vergelijk Tabel 1 en Tabel 2.

In Tabel 1 staat een overzicht van de verdelingskenmerken van de variabelen uit het basisonderwijs en het voortgezet onderwijs die in dit onderzoek worden gebruikt (N = 351). In Tabel 2 staat een overzicht van de verdelingskenmerken van dezelfde variabelen voor de steekproeven van 701 leerlingen en 438 leerlingen. Behalve op de intelligentiemeting, de leerprestaties en leerkrachtverwachtingen onderscheidt de te onderzoeken groep van 351 leerlingen zich ook van de oorspronkelijke onderzoeksgroep in het basisonderwijs en de responsgroep in het voortgezet onderwijs vanwege een hogere sociale afkomst. Door de grotere vertegenwoordiging van leerlingen uit hogere milieus en met hogere leerprestaties in de responsgroep, vergeleken met de oorspronkelijke onderzoeksgroep, zal de kans om milieuspecifieke verschillen in schoolloopbanen te vinden afnemen.

Tabel 1. Verdelingskenmerken van de variabelen op leerlingniveau. (N=351).

	gem.	sd.	min.	max.	range
SES	51.39	10.50	39.12	76.77	38- 77
IQ	13.46	2.33	5.29	18.43	1- 20
Begin leerjaar 7					
Rekenen	24.80	7.80	7.00	39.00	0- 40
Taal	25.72	5.10	6.00	34.00	0- 36
Voorkeur Ouders	26.82	8.88	10.00	40.00	0- 40
Geant. schooltype	24.49	9.50	10.00	40.00	5- 40
Begin leerjaar 8					
Rekenen	12.74	5.37	2.50	24.00	1- 24
Taal	12.70	3.37	2.50	21.00	1- 24
Cognitieve doelen	3.75	.68	1.00	5.00	1- 5
Geant. schooltype	25.11	9.46	10.00	40.00	5- 40
Eind leerjaar 8					
Cito-toets	537.30	9.65	508.00	550.00	500-550
Advies	25.47	9.99	5.00	40.00	5- 40
Na vier jaar v.o.					
Onderwijspositie	8.49	1.25	6	10	6- 10

Tabel 2. Gemiddelden en standaarddeviaties van de leerlingen die het basisonderwijs in 1988 verlieten en de responsgroep in 1992, na vier jaar voortgezet onderwijs.

	1988 (N=701)		1992 (N=438)	
	gem.	sd.	gem.	sd.
SES	50.14	10.04	50.61	10.52
IQ	12.87	2.49	13.18	2.47
Begin leerjaar 7				
Rekenen	22.85	7.84	24.11	7.74
Taal	24.28	5.53	25.16	5.27
Voorkeur ouders	20.49	13.55	21.56	13.30
Geant. schooltype	22.13	9.60	23.34	9.56
Begin leerjaar 8				
Rekenen	11.81	3.43	12.20	5.32
Taal	11.39	5.36	12.36	3.44
Cognitieve doelen	3.61	.71	3.68	.69
Geant. schooltype	22.82	9.70	23.95	9.58
Eind leerjaar 8				
Cito-toets	534.43	10.79	535.95	10.35
Advies	22.95	10.20	24.12	10.15

5. MODELSPECIFICATIE

Vertrekpunt voor de beantwoording van de onderzoeksvragen was een model dat is ontleend aan de eerdere studie naar de invloed van ouderlijke verwachtingen op schoolloopbanen in het basisonderwijs. In dit model zijn de effecten geanalyseerd van sociaal milieu (SES), aanleg (IQ), streefniveaus van ouders en leerkrachten en leerprestaties op het advies voor voortgezet onderwijs. In dit model worden SES en IQ als gecorreleerde exogene variabelen beschouwd. Uitgaande van dit model zijn modellen ontwikkeld waarin de invloed van streefniveaus op de positie van leerlingen in het voortgezet onderwijs is opgenomen. De effecten van de streefni-

veaus in leerjaar 7 en in leerjaar 8 zijn in afzonderlijke modellen bestudeerd. In de modellen voor leerjaar 8 zijn drie streefniveaus opgenomen (voorkeur van ouders, cognitieve doelen en geanticipeerd schooltype van leerkrachten); en in leerjaar 7 zijn twee streefniveaus opgenomen (voorkeur van ouders en geanticipeerd schooltype). Dit is gedaan omdat uit de eerdere studies is gebleken dat de cognitieve doelen in leerjaar 7 geen effect hebben op het vervolgadvis. In Figuur 1 is een vereenvoudigd uitgangsmodel voor beide leerjaren weergegeven. De werkelijke uitgangsmodellen zijn uitgebreider.

De hypothese in dit onderzoek is dat streefniveaus van ouders en leerkrachten in het basisonderwijs een langdurige werking hebben op de schoolloopbanen in het voortgezet onderwijs. Nadat voordat beide leerjaren passende modellen zijn gevonden, is nagegaan wat de invloed van streefniveaus is op de positie in het voortgezet onderwijs en of (en in welke mate) de streefniveaus bijdragen aan de verklaring van de samenhang tussen SES en de onderwijspositie in het voortgezet onderwijs.

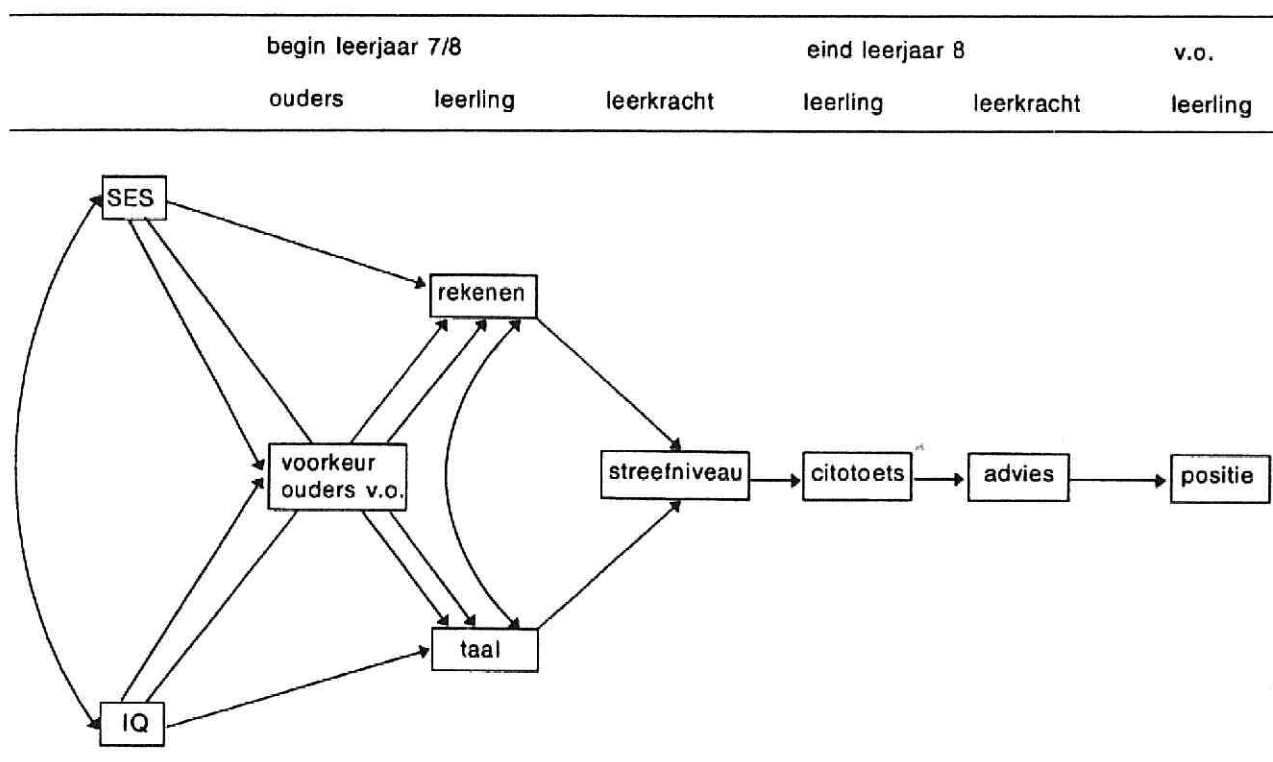


Fig. 1. Uitgangsmodel. De invloed van streefniveaus van ouders en leerkrachten op de onderwijspositie van leerlingen na vier jaar voortgezet onderwijs.

6. RESULTATEN

Voor beide leerjaren is een passend model gevonden, waarin de positie van leerlingen na vier jaar voortgezet onderwijs voor 70% wordt verklaard door variabelen in het basisonderwijs. In het model voor leerjaar 7 ($\chi^2=7.54$, $df=7$, $p=.48$, $AGF=.97$) vinden we directe effecten van beide streefniveaus op de onderwijspositie van leerlingen na vier jaar voortgezet onderwijs (Figuur 2). In het model voor leerjaar 8 ($\chi^2=16.28$, $df=14$, $p=.30$, $AGF=.97$) zien we alleen een direct effect van de voorkeur van de ouders op de onderwijspositie (Figuur 3).

In Tabel 3 en Tabel 4 staan de directe en indirecte effecten van de (leerling)kenmerken in respectievelijk leerjaar 7 en leerjaar 8 van het basisonderwijs op de latere onderwijspositie. Opmerkelijk is dat in de modellen van beide leerjaren geen directe effecten van SES op de onderwijspositie in het v.o. worden gevonden, terwijl de totale effecten van SES op de positie in

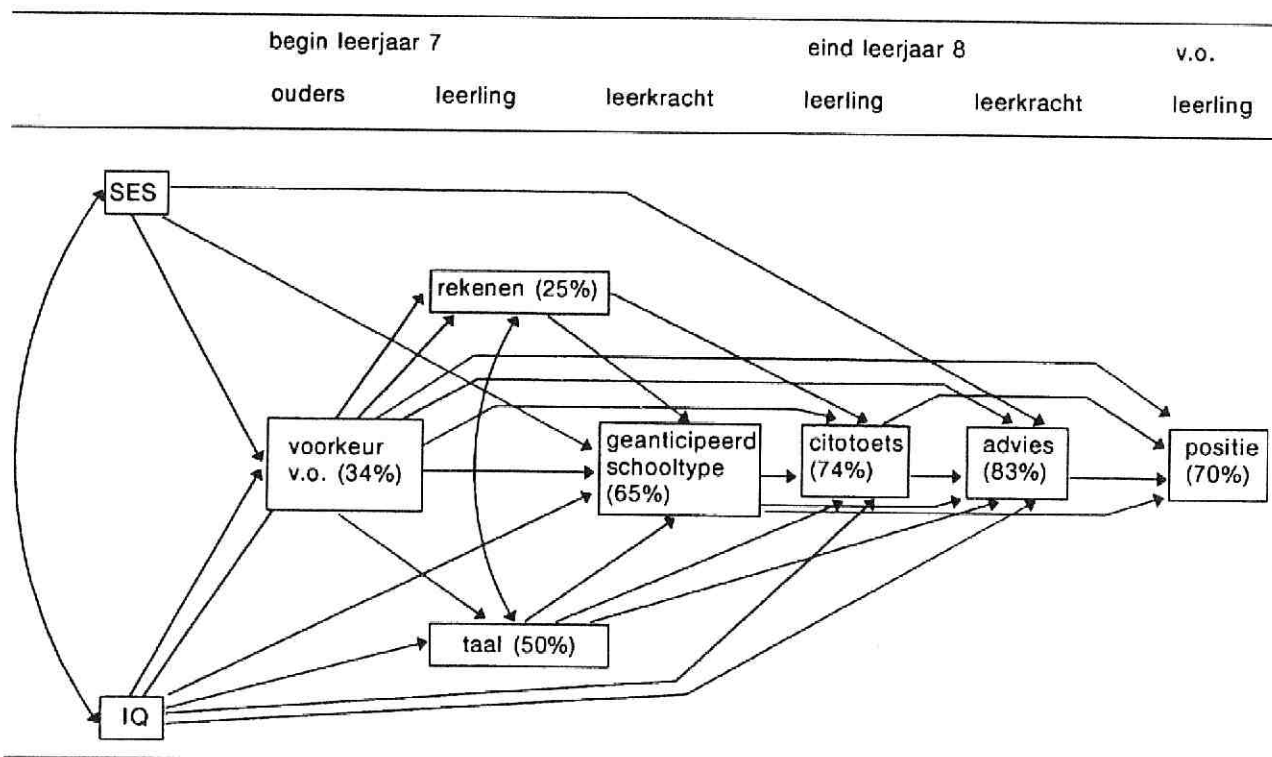


Fig. 2. Leerjaar 7. De invloed van streefniveaus van ouders en leerkrachten op de onderwijspositie van leerlingen na vier jaar voortgezet onderwijs. $N=351$ ($\chi^2=7.54$, $df=8$, $p=.48$, $AGF=.97$).

Tabel 3. Leerjaar 7. Totale, directe en indirecte effecten van (leerling)kenmerken in het basisonderwijs op de onderwijspositie van leerlingen na vier jaar voortgezet onderwijs.

effecten	totaal	direct	indirect	r
SES	.21	geen	.21	.44
IQ	.55	geen	.55	.64
Voorkeur ouders	.43	.16	.27	.66
Rekenen	.11	geen	.11	.44
Taal	.12	geen	.12	.62
Geant. schooltype	.31	.11	.20	.72
Cito-toets	.38	.23	.15	.74
Advies	.42	.42	-	.81

Tabel 4. Leerjaar 8. Totale, directe en indirecte effecten van (leerling)kenmerken in het basisonderwijs op de onderwijspositie van leerlingen na vier jaar voortgezet onderwijs.

effecten	totaal	direct	indirect	r
SES	.18	geen	.18	.44
IQ	.56	geen	.56	.64
Voorkeur ouders	.43	.18	.25	.66
Rekenen	.10	-.13*	.23	.58
Taal	.17	.08	.09	.60
Cognitieve doelen	.03	geen	.03	.49
Geant. schooltype	.44	geen	.44	.77
Cito-toets	.38	.30	.08	.74
Advies	.50	.50	geen	.81

* multi-collineariteit

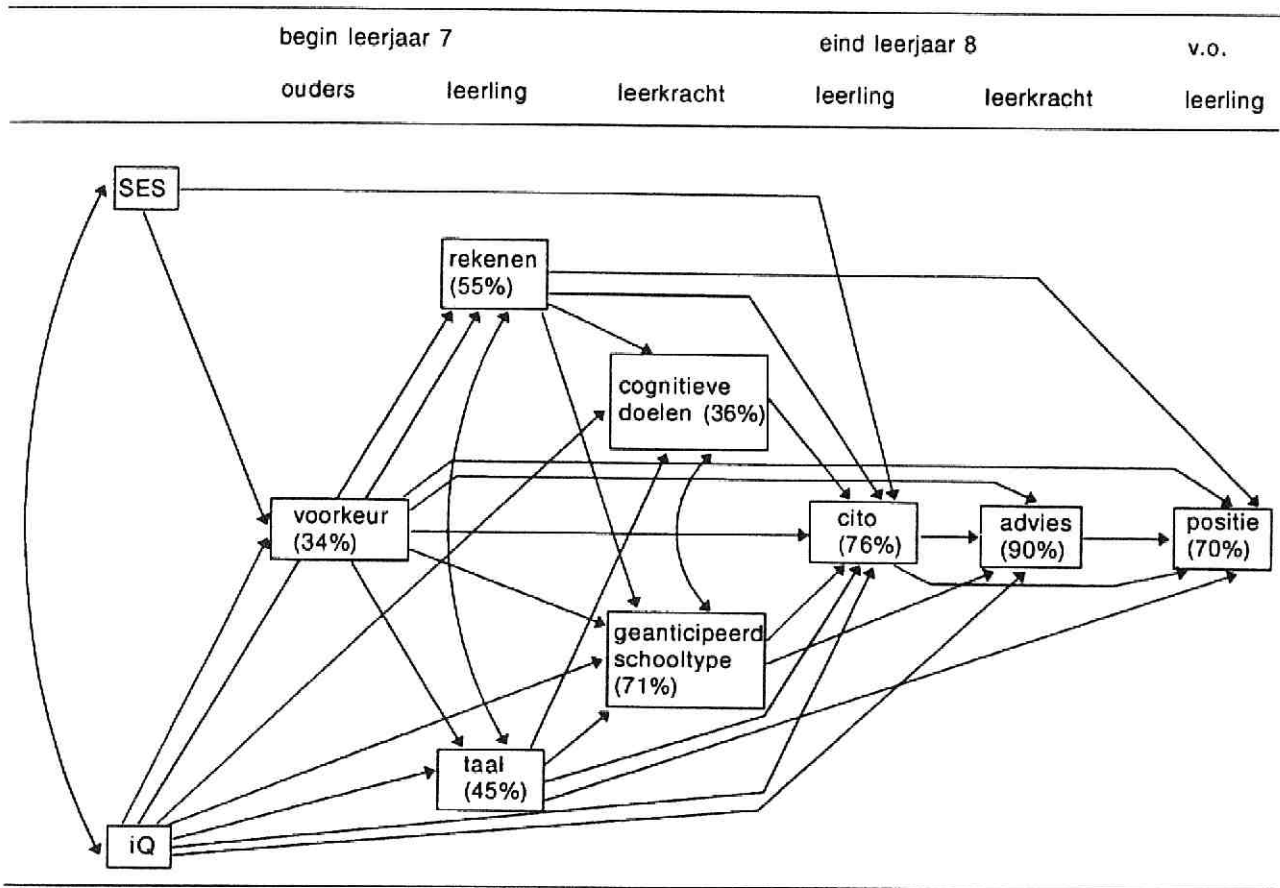


Fig. 3. Leerjaar 8. De invloed van streefniveaus van ouders en leerkrachten op de onderwijspositie van leerlingen na vier jaar voortgezet onderwijs. $N=351$ ($\chi^2=16.28$, $df=14$, $p=.30$, $AGF=.97$).

Tabel 5. Totale, directe en indirecte effecten en correlaties van SES op de (leerling)kenmerken in het basisonderwijs en de onderwijspositie na vier jaar voortgezet onderwijs.

effecten	leerjaar 7			leerjaar 8		
	totaal	direct	indirect	totaal	direct	indirect
Voorkeur ouders	.35	.35	geen	.35	.35	geen
Rekenen	.05	geen	.05	.06	geen	.06
Taal	.08	geen	.08	.10	geen	.10
Cognitieve doelen				.04	geen	.04
Geant. schooltype	.23	.10	.13	.15	geen	.15
Cito-toets	.10	geen	.10	.07	geen	.07
Advies	.25	.07	.19	.20	.06	.15
Positie v.o.	.21	geen	.21	.18	geen	.18

het v.o. (.21 en .18) geen aanzienlijke verschillen vertonen met de totale effecten op het vervolgadvis (.25 en .20), zie Tabel 5. Alle SES-gebonden variantie in de schoolloopbanen van de leerlingen van 10-16 jaar ligt in de periode van het basisonderwijs, hetzelfde geldt overigens ook voor IQ. Toetsing van de hypothese dat streefniveaus bijdragen aan de verklaring van de samenhang tussen SES en de onderwijspositie van leerlingen na vier jaar, laat zien dat de indirecte effecten van SES nauwelijks verandering ondergaan wanneer de streefniveaus uit de modellen worden weggelaten. Dit is in tegenstelling tot de bijdrage van streefniveaus aan de verklaring van de samenhang tussen sociaal milieu en vervolgadvis, zie Tabel 6.

Tabel 6. De invloed van SES op het vervolgadvis en de onderwijspositie na vier jaar voortgezet onderwijs. Toetsing hypothesen in modellen zonder streefniveaus (1) en met streefniveaus (2 tot en met 4).

model leerjaar 7	SES – vervolgadvis		SES – positie v.o.	
	direct	indirect	direct	indirect
1. zonder streefniveaus	.19	.08	geen	.20
2. streefniveau ouders	.09	.18	geen	.23
3. streefniveau leerkracht	.11	.15	geen	.20
4. streefniveaus ouders/leerkracht	.07	.19	geen	.21
model leerjaar 8	SES – vervolgadvis		SES – positie v.o.	
	direct	indirect	direct	indirect
1. zonder streefniveaus	.17	.10	geen	.20
2. streefniveau ouders	.09	.18	geen	.23
3. streefniveau leerkracht	.08	.19	geen	.21
4. streefniveaus ouders/leerkracht	.06	.15	geen	.18

7. DISCUSSIE

Het doel van het onderzoek is een bijdrage te leveren aan de verklaring van verschillen in schoolloopbanen, die niet kunnen worden toegeschreven aan verschillen in aanleg. Factoren zoals streefniveaus die samenhangen met het sociale milieu, dragen bij aan verschillen in schoolloopbanen. Door de longitudinale opzet van het onderzoek was het mogelijk het effect van streefniveaus in het basisonderwijs over een langere periode te bekijken. Verschillen in streefniveaus berusten niet alleen op verschillen in intelligentie, maar ook op verschillen in het sociale milieu waaruit de leerlingen afkomstig zijn. Het resultaat van de longitudinale opzet is dat de langdurige werking van streefniveaus is aangetoond. Dit resultaat behoeft ook nadere beschouwing. Aangetoond is dat, rekening houdend met verschillen in intelligentie en leerprestaties, de verwachtingen die ouders voor hun kinderen hebben, een relevante bijdrage leveren aan de verklaring van verschillen in onderwijsposities van kinderen in het voortgezet onderwijs. Wat de streefniveaus van leerkrachten betreft is dit slechts ten dele het geval. Voor de streefniveaus van leerkrachten in leerjaar 7 is wel een direct effect op de latere onderwijspositie gevonden en voor streefniveaus van de leerkrachten in leerjaar 8 niet. Het ontbreken van een direct effect in leerjaar 8 kan worden toegeschreven aan de nabijheid van het vervolgadvis in dit leerjaar. Naarmate de tijdstippen waarop de streefniveaus en het vervolgadvis zijn vastgesteld, dichter bij elkaar liggen, zal de overeenkomst tussen streefniveaus en advies toenemen; met als gevolg dat het advies domineert in de verklaring van de latere onderwijspositie.

Hoe bepalend de rol van het advies is, blijkt ook uit het ontbreken van directe effecten van sociaal milieu (en intelligentie) op de latere onderwijspositie. Alle milieu-gebonden variantie in de schoolloopbanen van de leerlingen van 10-16 jaar ligt in het basisonderwijs. Ook de streefniveaus, die in het basisonderwijs wel bijdragen aan de verklaring van de invloed van sociaal milieu op het vervolgadvis, hebben nauwelijks betekenis als het gaat om de verklaring van de samenhang tussen sociaal milieu en de onderwijspositie na vier jaar voortgezet onderwijs. Die verklaring ligt namelijk bij het vervolgadvis. Het vervolgadvis aan het einde van de basisschool is de laatste en bepalende factor in het determinatieproces van milieuspecifieke schoolloopbanen. Het advies, de 'terminal determinant', is de uitdrukking van een onderliggend dynamisch proces dat zich in de dagelijkse klaspraktijk afspeelt tussen sociaal milieu, aanleg, streefniveaus en leerprestaties. De resultaten van dit onderzoek samenvattend, komen we tot drie conclusies:

- . Streefniveaus in het basisonderwijs, van ouders en leerkrachten, hebben een langdurige werking op schoolloopbanen in het voortgezet onderwijs.
- . Alle milieu-gebonden variantie in schoolloopbanen (in de eerste vier jaar van het voortgezet onderwijs) ligt in het basisonderwijs.
- . Het vervolgadvis is de laatste en bepalende factor in het proces van milieuspecifieke schoolloopbanen (tot zo'n 16 jaar).

In dit onderzoek is nagegaan of streefniveaus in het basisonderwijs bijdragen aan de verklaring van de samenhang tussen het sociale milieu en de onderwijspositie in het voortgezet onderwijs. Niet is nagegaan of er binnen de posities in het voortgezet onderwijs zelf sprake is van een andere vorm van milieu-specifieke selectie die verband houdt met de vakkenpakketkeuzes van jongens en van meisjes. Uitsplitsing naar vakken in het voortgezet onderwijs kan zicht bieden op het verband tussen sekse en richting of vakkenpakketkeuze en de mate waarin daarvoor de basis wordt gelegd in het basisonderwijs. De vraag naar verschillen tussen jongens en meisjes in de samenstelling van hun vakkenpakketten wordt nog nader onderzocht. Hetzelfde geldt voor de vraag of het voor latere schoolprestaties uitmaakt of een leerling op een meer of minder effectieve basisschool heeft gezeten.

LITERATUUR

- Bosker, R.J. (1990). *Extra kansen dankzij de school?* Dissertatie. Nijmegen: ITS.
- Bosker, R.J. (1992). *De stabiliteit en consistentie van schooleffecten in het basisonderwijs*. Enschede: OCTO.
- Hoeven-van Doornum, A.A. van der (1990). *Effecten van leerlingbeelden en streefniveaus op schoolloopbanen*. Dissertatie. Nijmegen, ITS.
- Hoeven-van Doornum, A. A. van der, Voeten, M.J.M. & Jungbluth, P. (1989). The effect of Aspiration Levels Set by Teachers for their Pupils on Learning Achievement. In B. Creemers, T. Peters & D. Reynolds (Eds.), *School Effectiveness and School Improvement*. Lisse: Swets & Zeitlinger.
- Hoeven-van Doornum A.A. van der, Voeten M.J.M. & Jungbluth, P. (1990a). De invloed van klas- en schoolkenmerken op leerprestaties en vervolgadvis in het basisonderwijs. I: C.A.C. Klaassen & P.L.M. Jungbluth (Eds.), *Onderwijs Research Dagen 1990. Onderwijs en Samenleving*. Nijmegen: ITS.
- Hoeven-van Doornum A.A. van der, Voeten M.J.M. & Jungbluth, P. (1990b). Het effect van leerkrachtverwachtingen en streefniveaus op schoolloopbanen. *Tijdschrift voor Onderwijs Research*, 1, 23-42.
- Hoeven-van Doornum, A.A., Voeten, M.J.M. & Jungbluth, P. (1993). The influence of parents' expectations on school careers. *Tijdschrift voor Onderwijs Research*, 6, 369-379.
- Rupp, J.C. (1969). *Opvoeding tot schoolweerbaarheid*. Groningen: Wolters-Noordhoff.
- Tesser, P. (1986). *Sociale herkomst en schoolloopbanen in het voortgezet onderwijs*. Nijmegen: ITS.

Manuscript ontvangen: 10-12-1993

Definitieve versie ontvangen: 3-5-1994

School- en sekseverschillen in vakkenkeuzen in het VWO

R.J. Bosker¹ en H.P.J.M. Dekkers²

¹ *Faculteit der Toegepaste Onderwijskunde, Universiteit Twente**

² *Instituut voor Toegepaste Sociale wetenschappen, Universiteit Nijmegen*

ABSTRACT

There are still considerable gender differences in the choice of mathematics as a subject in the second stage of secondary education. This article reports the results of a longitudinal study into the causes of these gender differences. It concerns the role of the school in the development of gender related mathematics achievement and the choice of mathematics B in pre-university education. To cope with the hierarchical structure of the data a series of multilevel models are applied: two-level models to assess differences between schools, three-level models to assess differences between schools in the development of mathematics achievement, and a multilevel logistic model to assess differences between schools in the choice of mathematics as a subject. The results show that schools do not differ in the development of gender specific mathematics achievement. Schools, however, do differ in gender related choice behaviour of their students, especially at the moment when the choice of subjects becomes final (at the end of grade four). Furthermore, the results show that some schools are more meritocratic than other schools: the gender gap in choice behaviour, after taking into account achievement differences between boys and girls, is at some schools almost twice as large as at other schools.

INLEIDING

Ondanks ruime aandacht voor het fenomeen en vele maatregelen om de situatie te keren is er nog steeds sprake van aanzienlijke sekseverschillen in sector- en vakkenkeuzen in het voortgezet onderwijs. Niet alleen in Nederland, maar ook in de andere Westerse landen zijn er ondanks een inhaalrace van meisjes qua onderwijsniveau nog grote achterstanden in hun deelname aan en prestaties in exacte en technische vakken (Hanna, ter perse; Dekkers, 1993a). Alhoewel soms een afname van de verschillen in de laatste decennia wordt gerapporteerd (o.a. in een meta-analyse van Hyde, Fennema en Lamon, 1990) moet aan de hand van gegevens van het CBS worden geconstateerd dat ook in Nederland de verschillen nog erg groot zijn. Onderstaande tabel toont als voorbeeld de percentages meisjes en jongens die exacte vakken kiezen in het VWO in de laatste tien jaar.

De cijfers voor wiskunde A laten een lichte relatieve verbetering zien, die echter voornamelijk is te danken aan het in grote getale kiezen van het nieuwe wiskunde B ten koste van wiskunde A door jongens na de programmatische wijzigingen in 1986 (HEWET). Nog steeds kiest echter een tweemaal zo groot percentage van de jongens als van de meisjes wiskunde B en natuurkunde (en in mindere mate scheikunde). Deze situatie heeft gevolgen voor hun kansen in het vervolgonderwijs en op de arbeidsmarkt. Door bijvoorbeeld geen wiskunde en/of natuurkunde te kiezen in het VWO sluiten leerlingen 49 procent van de mogelijke vervolgstudies in het wetenschappelijk onderwijs uit (Bosker, 1990).

In de afgelopen jaren is veel onderzoek gedaan naar de oorzaken van sekseverschillen in vakkenkeuzen in het voortgezet onderwijs. Het onderzoek vond plaats vanuit verschillende invalshoeken. Sociologische, psychologische en onderwijskundige verklaringen werden aange-

* Adres: Postbus 217, 7500 AE Enschede

Tabel 1. Keuze van exacte vakken door jongens (δ) en meisjes (♀) in het VWO in percentages (bron: CBS-statistieken van het AVO).

	1983		1985		1987		1989		1991	
	δ	♀	δ	♀	δ	♀	δ	♀	δ	♀
wiskunde A	82	57	82	59	62	54	61	58	61	61
wiskunde B	25	5	27	6	63	30	62	31	63	32
natuurkunde	64	28	64	28	63	28	62	28	62	29
scheikunde	57	35	56	35	52	32	48	29	46	27

voerd. De meeste onderzoeken hadden betrekking op slechts één mogelijke oorzaak, zoals leerkrachtgedrag, keuzebegeleiding, curriculum of attitudes van leerlingen. Er is nauwelijks onderzoek gedaan vanuit een theoretisch kader waarin meer dan één (soort) verklarende variabele(n) was opgenomen. Een ander manco in eerder onderzoek is, dat er geen sprake was van longitudinale verzameling van de data. Gegeven het feit dat de grote sekseverschillen in keuzen op 15-jarige leeftijd de min of meer gelijke prestaties waarvan op het eind van de basisschool sprake is slecht weerspiegelen, lijkt longitudinale dataverzameling echter zeer nuttig.

Om aan deze manco's tegemoet te komen is in het project 'Meisjes en exacte vakken: een geleidelijk proces van vakkenkeuze' besloten een cohort van leerlingen te volgen vanaf de brugklas tot aan het moment van vakkenkeuze aan de hand van diverse variabelen die uit eerder onderzoek als relevant naar voren kwamen. Uitgaande van de informatie dat er op 12-jarige leeftijd nauwelijks sekseverschillen in rekenen zijn, maar dat er op 15/16-jarige leeftijd grote seksegebonden verschillen zijn in zowel vakkenkeuze als in prestatie, is de periode tussen 12 en 15 jaar meer gedetailleerd onderzocht. Met name de rol van de school in de sekse specifieke ontwikkeling van leerprestaties en vakkenkeuzen komt in dit artikel aan de orde.

Theoretisch kader

Zoals aangegeven betreft het onderhavige onderzoek een groot aantal factoren die (in onderlinge samenhang) sekseverschillen in vakkenkeuzen zouden kunnen beïnvloeden.

Het betreft aan de ene kant kenmerken op leerlingniveau, aan de andere kant kenmerken op schoolniveau. De desbetreffende leerlingkenmerken passen in een sociaal-psychologisch verklaringsmodel dat is ontwikkeld om de factoren te beschrijven die sekseverschillen in schoolse keuzen kunnen verklaren (Eccles, 1986; Van Eck & Volman, 1990). Het betreft naast de gezinsachtergrond (opleidingsniveau en sociaal-economische achtergrond van de ouders) van de leerling (de ontwikkeling in) factoren als: al dan niet realistische toekomstverwachtingen met betrekking tot werk en gezin, attitudes ten opzichte van diverse vakken (interesse, waardetoe- kenning, nut), de inschatting van de eigen geschiktheid voor een vak door de leerling, de invloed van relevante anderen zoals ouders, leerkrachten etc. en, niet het minst belangrijk, voorafgaande leerprestaties. Voor een meer uitvoerige theoretische beschrijving van deze factoren (en bijbehorende literatuurreferenties) wordt verwezen naar rapportage over het onderhavige onderzoek waarin de nadruk ligt op de rol van leerlingkenmerken (Dekkers, 1993b).

Op het belang van (voorafgaande) leerprestaties van leerlingen gaan we in dit artikel meer uitvoerig in. Dat er naast keuzeverschillen ook sekseverschillen zijn in leerprestaties in exacte vakken is ruim gedocumenteerd. Hanna (ter perse) concludeert in een heranalyse van de data uit de 'Second International Mathematics Study' van de IEA dat in de meeste landen (12 van de 15) jongens op 17-jarige leeftijd beter presteren in wiskunde dan meisjes. Uit de in 1988 uitgevoerde 'uitgangsmeting' naar de prestaties van ongeveer 6000 leerlingen in de vakken Nederlands, Engels, biologie en wiskunde (Kremers, 1990) blijkt dat de sekse specifieke verschillen in prestaties over het algemeen gering zijn, maar dat jongens gemiddeld een 5 procent hogere score halen bij wiskunde. In een eerdere studie werd hetzelfde aangetoond (Pelgrum & Plomp, 1986).

Het verband tussen leerprestaties en voorgenomen vakkenkeuzen is op het niveau van de individuele leerling zwakker dan verwacht (Dekkers, 1992), een reden om andere verklarende variabelen (zoals attitudes of schoolfactoren) op te sporen. Daarnaast blijkt dat aan het eind van het voortgezet onderwijs, overeenkomstig hun prestaties, meisjes minder dan jongens voor een technische opleiding kiezen (Dekkers, 1990). Deze meisjes hadden minder wiskunde en natuurkunde in hun pakket en presteerden bovendien minder dan hun mannelijke klasgenoten. Iets soortgelijks vonden Kuiper & Meulenbeld (1989) ten aanzien van de feitelijke vakkenkeuze in het voortgezet onderwijs. Er is dus reden om de relatie tussen prestaties en keuzen nader te onderzoeken. En in dit verband met name de rol van de school in dit proces.

In schooleffectiviteitsonderzoek (nationaal en internationaal) is de outputvariabele zelden 'sekselijkheid'. En in emancipatie-onderzoek wordt zelden gekeken naar schooleffecten op sekseverschillen. Er is kortom nog weinig bekend over schoolverschillen in seksegebonden leerprestaties en keuzen. Wel is de invloed onderzocht van diverse schoolse factoren zoals keuzebegeleiding, leerkrachtgedrag en inhoud van curriculum en lesmateriaal op sekseverschillen in prestaties en keuzen. Ten behoeve van dergelijk onderzoek zijn in diverse Nederlandse onderzoeksprojecten scholen gerangordend van hoog naar laag in populariteit van exacte vakken. In deze outlier-studies wordt ervan uitgegaan dat de grote verschillen die worden gevonden (op sommige scholen kiest slechts één meisje op de drie jongens wiskunde, op andere is de ratio één op één) significante verschillen zijn, en wordt gezocht naar relevante factoren waarop deze scholen van elkaar verschillen. Schoolkenmerken worden op leerlingniveau lineair in verband gebracht met leerlingoutput. Dit geschiedt ook in andere dan outlierstudies.

Omdat veel van bovengenoemde onderzoeken uitgevoerd werden in opdracht van de overheid, zijn vooral aspecten onderzocht waarop de overheid zich in haar beleid eventueel kan richten. Een belangrijk aanknopingspunt was de keuzebegeleiding en de rol van de decaan daarin. Op scholen waar meisjes relatief vaak exacte vakken kiezen bleek sprake van een systematisch, goed georganiseerd keuzebegeleidingssysteem (voor jongens én meisjes) en van decanen die zich in hun advisering meer lieten leiden door de leerprestaties dan door de eigen wensen en attitudes van de meisjes (Dekkers, 1985; Kristensen & Jenneskens, 1991).

Andere schoolvariabelen in het onderzoek naar sekseverschillen in vakkenkeuzen zijn in een vroeg stadium ontleend aan publikaties van Rutter (1979) over karakteristieken van de school als sociale institutie, en van Brookover (1979) over het effect van schoolklimaat op leerprestaties van leerlingen. Later zijn factoren die als mogelijk relevant uit meer recent schooleffectiviteitsonderzoek naar voren kwamen (Bosker & van der Velden, 1989; Scheerens, 1989) naast factoren die voortvloeiden uit de resultaten van eerder emancipatie-onderzoek (Dekkers, 1985; Kuiper & van der Werf, 1987) gerelateerd aan schoolloopbanen van jongens en meisjes. Uit deze onderzoeken kwam naar voren dat specifieke maatregelen ten behoeve van de emancipatie van meisjes geen invloed hebben op de mate waarin meisjes exacte vakken kiezen (Kristensen & Jenneskens, 1991; van der Velden, 1989; Bosker, 1990). Ook een positief effect van een mogelijke voorbeeldfunctie van vrouwelijke wiskundedocenten kan niet worden aangetoond (Kuiper & van der Werf, 1987). Wel lijkt er sprake van niet-bewust emancipatoire aspecten die de onderwijspositie van meisjes begunstigen. Dat zijn met name het ontbreken van een dominerende maatschappijvisie (conservatief of progressief), vernieuwing die is geconcentreerd op organisatorische aspecten en niet op leermiddelen, beargumenteerde (redelijke) disciplinehandhaving, een (inhoudelijk) betrokken schoolleiding, een afstandelijk, vriendelijk contact met leerlingen dat meer gericht is op prestaties dan op sociaal-emotionele problemen (Dekkers, 1985). Ook persoonlijk contact met de ouders en hoge verwachtingen ten aanzien van leerlingen komen meer voor op scholen waar veel meisjes exact kiezen (Kristensen & Jenneskens, 1991). Bosker (1990) laat in één van de weinige longitudinale studies op dit terrein zien dat er schoolverschillen zijn in keuzegedrag van leerlingen en ook schoolverschillen in sekseverschillen in keuzegedrag: op 25 procent van de scholen zijn de keuzeverschillen tussen jongens en meisjes groter dan één standaarddeviatie, terwijl op andere scholen slechts van een halve standaarddeviatie sprake is. De schoolfactoren die samenhangen met deze verschillen (doeloriëntatie en een coherent leerkrachtteam) verklaren echter alleen keuzeverschillen tussen jongens.

In het onderhavige artikel zal de invloed van de school op seksespecifieke leerprestaties in en keuzen van exacte vakken centraal staan. Anders dan in de beschreven studies echter, zal nadrukkelijk het seksespecifieke segregatieproces centraal staan, dat wil zeggen dat de rol van de school bij het ontstaan van verschillen tussen jongens en meisjes op het terrein van de exacte vakken wordt onderzocht.

Vraagstelling

In het project 'Meisjes en exacte vakken: een geleidelijk proces van vakkenkeuze' is gepland, in een design waarbij onderzoeksscholen gerangschikt worden naar de proportie meisjes die exacte vakken kiezen, te zoeken naar schoolfactoren die de schoolverschillen (in sekseverschillen) kunnen verklaren. Gezien de vaak grote verschillen tussen de verschillende schooltypen ligt concentratie op één schooltype voor de hand, in dit geval het VWO. De keuze voor wiskunde B differentieert op dat schooltype het meest tussen de seksen. In een eerste analyse lijken dergelijke schoolverschillen inderdaad aanwezig: in grafiek 1 is van de 20 scholen in het onderhavige onderzoek aangegeven welk percentage meisjes respectievelijk jongens uiteindelijk dit vak aan het eind van het vierde leerjaar opneemt in het pakket. We zien dat er enerzijds scholen zijn waar de verschillen in het percentage meisjes dat wiskunde B kiest ten opzicht van de jongens zo'n 80 procent bedragen (in school 1 kiest 0 percent van de meisjes wiskunde B tegenover 86 percent van de jongens; in school 2 zijn de percentages respectievelijk 29 en 100), maar dat er anderzijds ook scholen zijn waar meisjes vaker dan jongens dit vak kiezen (in school 19 zijn de percentages voor meisjes en jongens respectievelijk 38 en 14; in school 20: 80 en 38).

Alvorens te trachten verklarende factoren op te sporen lijkt het echter nuttig precies vast te stellen hoe groot de verschillen in seksespecifieke leerprestaties en keuzen (en de relatie daartussen) tussen scholen zijn en wanneer die verschillen ontstaan.

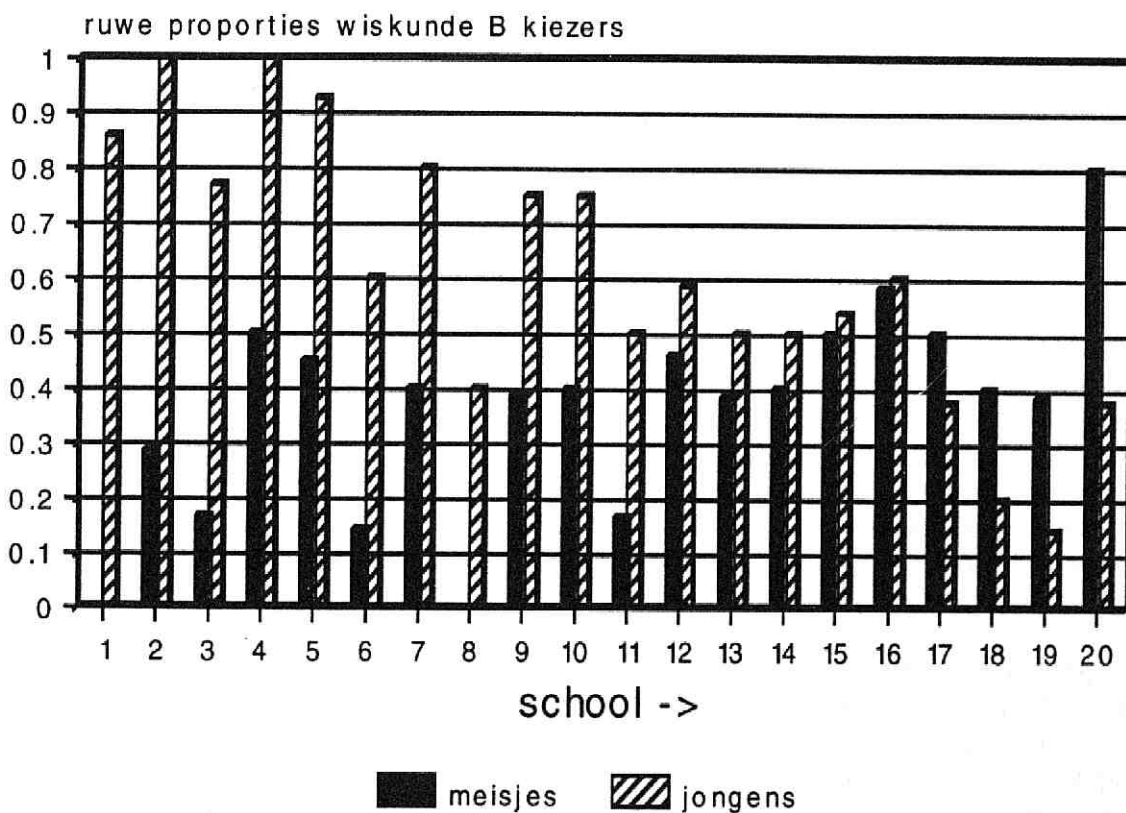


Fig. 1. Percentages meisjes en jongens dat wiskunde B kiest per school.

De volgende vragen zullen in dit artikel aan de orde komen:

1. In welke periode komen seksspecifieke verschillen ten aanzien van de prestaties in en de keuze van wiskunde B aan het licht?
2. Zijn er tussenschoolse verschillen in de ontwikkeling van wiskundeprestaties en in de ontwikkeling van sekseverschillen in wiskundeprestaties?
3. Verschillen scholen in keuzen die leerlingen zich voornemen en de facto maken ten aanzien van de vakkenpakkettenamenstelling, en verschillen scholen in de seksspecificiteit van deze keuzen?
4. Zijn sommige scholen meritocratischer dan andere, dat wil zeggen is de relatie tussen prestaties en keuzen in sommige scholen minder seksspecifiek dan in andere?

Data en methode

In het project is een cohort van 2500 leerlingen van 25 aselect getrokken scholen voor voortgezet onderwijs gevolgd vanaf de brugklas tot aan het moment van vakkenpakketkeuze (eind tweede klas voor MAVO, eind derde klas voor HAVO, eind vierde klas voor VWO). In 1993 vonden de laatste metingen bij VWO-leerlingen plaats. Ieder jaar zijn naast de (voorgenomen) vakkenkeuzen verschillende soorten variabelen (zoals gezinsachtergrond, toekomstplannen, waarde- en nutinschatting van wiskunde, attributie van succes en falen in wiskunde) bij alle leerlingen gemeten via een schriftelijke vragenlijst. Daarnaast hebben de leerlingen ieder jaar een door het CITO ontwikkelde toets voor wiskunde gemaakt. Zoals reeds aangegeven beperkt de analyse zich in dit artikel tot de gegevens van de VWO-leerlingen.

Vijf variabelen staan daarbij centraal: 1) de sekse; 2) de school die door de leerling bezocht wordt; 3) wiskundeprestaties; 4) de (voorgenomen) pakketkeuze, met name of daarin wiskunde B wordt opgenomen en 5) het tijdsaspect. De operationalisatie van de variabelen is als volgt:

- 1) sekse: 0=jongens; 1=meisjes.
- 2) school: uiteindelijk zijn er gegevens beschikbaar over 20 scholen, daar enkele scholen uitvielen en voor enkele andere scholen te weinig gegevens over VWO-4 leerlingen resteerden om zinvol sekseverschillen binnen deze scholen te kunnen onderzoeken.
- 3) wiskunde prestaties: wiskundetoetsen voor leerjaar 1, 2 en 3, ontwikkeld door het CITO, met als betrouwbaarheden respectievelijk .70, .73 en .71.
- 4) wiskundekeuze: geneigdheid om wiskunde als vak te kiezen gemeten op een vijf-puntschaal: 1 staat dan voor zeker niet en 5 voor zeker wel. Voor alle duidelijkheid zij opgemerkt dat het bij de eerste drie metingen dus gaat om het *voornemen* wiskunde B al dan niet in het pakket op te nemen. De feitelijke keuze van de VWO-leerlingen aan het einde van het vierde leerjaar is een dichotome variabele (0: niet gekozen; 1: wel gekozen).
- 5) tijd: deze variabele wordt geïntroduceerd om de ontwikkeling in wiskundeprestaties in de eerste drie leerjaren te kunnen modelleren. De codering is: 0 voor leerjaar 1, 1 voor leerjaar 2 en 2 voor leerjaar 3.

Het hier gerapporteerde onderzoek betreft enkel de onvertraagde leerlingen, dat wil zeggen leerlingen die tot aan het vierde leerjaar niet zijn blijven zitten, noch zijn afgestroomd naar het HAVO, noch ongediplomeerd het onderwijssysteem hebben verlaten. Het gaat hier dus om een selectief deel van het oorspronkelijk cohort (namelijk de betere leerlingen), maar deze selectie verstoort vermoedelijk niet het onderzoek naar de sekseverschillen tussen de leerlingen. Correlaties zijn in het algemeen niet erg gevoelig voor dergelijke uitval.

De volgende tabel bevat de gegevens over de VWO-leerlingen, waartoe we ons, zoals gezegd, in dit artikel zullen beperken. De gemiddelden vermeld achter de dichotome variabelen (sekse en keuze in leerjaar 4) zijn proporties.

In het vervolg van dit artikel zullen we nagaan of de verschillen zoals gepresenteerd in grafiek 1 systematisch zijn danwel toe te schrijven aan toevalsfluctuaties. We passen daartoe het hiërarchisch lineair model toe (Bryk & Raudenbush, 1992), zodat we in staat zijn steekproevenvariantie (een gevolg van het feit dat we binnen elke school een steekproef van leerlingen

Tabel 2. Beschrijvende gegevens.

	Gem.	s.d.	n
seks (0: meisje; 1: jongen)	.47	.50	392
wiskunde toets in leerjaar 1	58.65	14.84	392
wiskunde toets in leerjaar 2	65.26	15.26	392
wiskunde toets in leerjaar 3	71.37	16.04	392
keuze van wiskunde B in leerjaar 1	2.87	1.10	392
keuze van wiskunde B in leerjaar 2	3.15	1.35	392
keuze van wiskunde B in leerjaar 3	3.56	1.39	392
keuze van wiskunde B in leerjaar 4	.51	.50	392

tijd (0: leerjaar 1; 1: leerjaar 2; 2: leerjaar 3)

hebben getrokken) te scheiden van ware parameter variantie ('echte' verschillen tussen scholen). Waar verschillen in voorgenomen keuzen worden geanalyseerd zal een twee-niveau model worden gehanteerd: leerlingen genest onder scholen. Op het niveau van de leerlingen worden dan geslacht en toetsprestaties (en de interactie daartussen) als voorspellers opgenomen. De vraag of op sommige scholen de jongens-meisjes verschillen groter zijn dan op andere scholen kan dan positief beantwoord worden als de variantie tussen scholen in het schoolspecifieke effect van geslacht op voorgenomen keuzen systematisch groter is dan nul. De vraag of scholen verschillen in de seksespecifieke ontwikkeling van wiskundeprestaties wordt beantwoord door een dubbele toepassing van het multilevel model. Om te beginnen worden de tijdstippen (leerjaar 1, 2 en 3) beschouwd als genest onder de leerlingen. Door nu tijd als een voorspeller op leerling-niveau in het model op te nemen kan worden nagegaan of leerlingen verschillen in de ontwikkeling van hun wiskunde prestaties (het effect van tijd op leerprestaties varieert dan systematisch van leerling tot leerling). Op het niveau van de leerling kan de sexe-variabele vervolgens worden opgenomen als voorspeller van verschillen tussen leerlingen in hun wiskundeprestaties, maar tevens als voorspeller van die ontwikkelingsverschillen (in de praktijk gaat het hier dan om geslacht*tijd interactie). Bovendien kunnen we het schooleffect in kaart brengen door een onderscheid aan te brengen in een leerlingniveau en een schoolniveau: op het niveau van de leerlingen zijn de afhankelijken dan de initiële wiskundekennis en de groei in wiskundekennis, met als onafhankelijke variabele geslacht. Voor deze afhankelijke variabelen geldt wederom dat de variantie gesplitst kan worden in een binnen- en tussenschools gedeelte en bovendien kan worden nagegaan of het effect van seks van school tot school varieert. Deze twee multilevel modellen worden als het ware in elkaar geschoven tot één drie-niveau model: tijdstippen genest onder leerlingen genest onder scholen. Op tijdniveau is tijd de voorspeller; op leerlingniveau is geslacht de voorspeller; er kan nu worden nagegaan in welke opzichten scholen van elkaar verschillen. Een aparte benadering tenslotte is nodig om de feitelijke keuze van het vak wiskunde B aan het einde van het vierde leerjaar te onderzoeken. Het gaat hier immers om een dichotome afhankelijke variabele. Hiertoe wordt een multilevel logistisch model gehanteerd (Longford, 1989), waarbij in plaats van de afhankelijke de logit daarvan ($\ln(p/(1-p))$) wordt geanalyseerd. Het splitsen van variantiecomponenten is nu echter niet meer mogelijk, daar gemiddelde en variantie bij dichotomen van elkaar afhankelijk zijn. Wel kan op schoolniveau de variantie in proporties worden geschat en bovendien kan worden nagegaan of het effect van geslacht op de keuze van wiskunde B tussen de scholen varieert. Ook hiervoor geldt dat steekproevenvariantie (een gevolg van het trekken van een steekproef van scholen en een steekproef van leerlingen binnen de scholen) en ware parameter variantie gesplitst worden.

RESULTATEN

Sekseverschillen in keuzegedrag en prestaties

De eerste vraag waarop we een antwoord zoeken betreft de grootte van de verschillen tussen jongens en meisjes in de geneigdheid om wiskunde B als examenvak te kiezen en hun prestaties in het vak wiskunde op diverse momenten in de schoolloopbaan.

Tabel 3. Sekseverschillen in prestaties en keuzen.

	meisjes	jongens
wiskunde toets in leerjaar 1	55.15	62.60
wiskunde toets in leerjaar 2	63.54	67.20
wiskunde toets in leerjaar 3	69.45	73.53
keuze van wiskunde B in leerjaar 1	2.78	2.98
keuze van wiskunde B in leerjaar 2	3.08	3.23
keuze van wiskunde B in leerjaar 3	3.26	3.89
keuze van wiskunde B in leerjaar 4	.38	.65

De gegevens in tabel 3 laten zien, dat ten aanzien van prestaties in het vak wiskunde de verschillen tussen jongens en meisjes in de loop van de eerste drie leerjaren lijken af te nemen. Is het verschil bij aanvang van de carrière ruim 7 punten, aan het einde van leerjaar 3 is het geslonken naar nog slechts 4 punten (terwijl de variantie in de toetsresultaten wel lichtelijk toeneemt). Ten aanzien van het keuzegedrag treedt het omgekeerde op: zijn de verschillen tussen jongens en meisjes ten aanzien van het voornemen wiskunde B uiteindelijk in het vakkenpakket op te nemen aanvankelijk zeer klein, naarmate het moment van de feitelijke keuze naderbij komt worden de verschillen steeds groter. Uiteindelijk kiest op de 20 scholen in het onderzoek 38% van de meisjes en 65% van de jongens wiskunde B.

In het navolgende zullen we eerst nagaan of scholen verschillen in de seksspecifieke ontwikkeling van wiskundeprestaties, waarna we vervolgens onderzoeken of scholen effect hebben op sekseverschillen in de keuze van wiskunde B als examenvak. Tenslotte wordt het meritocratisch gehalte van de scholen onderzocht: dat wil zeggen of scholen verschillen in de mate waarin meisjes en jongens bij gelijke prestaties verschillen in de keuze van wiskunde B.

Schoolverschillen in de seksspecifieke ontwikkeling van wiskundeprestaties

Om de schoolverschillen in de ontwikkeling van wiskundeprestaties in kaart te brengen zal, zoals gezegd, een drie-niveau analyse worden uitgevoerd. De essentie van het model is, dat prestaties worden gemodelleerd als een functie van de verstreken tijd. Anders gezegd: dit is een herhaalde metingen ofwel binnen subjecten analyse, waarin prestaties worden gemodelleerd in een eerste orde polynoom. Het belangrijkste aspect van deze analyse is, dat elke leerling zijn of haar eigen groei-traject kan hebben: sommigen zullen zich wellicht in de tijd gezien sneller ontwikkelen dan anderen. Daarnaast kunnen we nagaan of de gemiddelde groeicurves van de scholen verschillen. En tenslotte kan nagegaan worden of de mate waarin de ontwikkeling van wiskundeprestaties tussen jongens en meisjes verschilt op zich weer van school tot school varieert. Het gaat hier met andere woorden om een derde orde interactie-effect tussen school, sekse en tijd.

Schooleffecten op wiskundeprestaties kunnen aldus op meerdere manieren zichtbaar worden, namelijk als verschillen tussen scholen in: 1) aanvangsniveau, 2) ontwikkeling, 3) sekseverschillen in aanvangsniveau, 4) sekseverschillen in ontwikkeling. De resultaten van de analyse staan vermeld in tabel 4.

Tabel 4. Schoolverschillen in de ontwikkeling van wiskundeprestaties.

intercept (meisjes)	55.00	
seks effect (jongens)	6.55	
tijd effect	7.15	
seks * tijd effect	-1.78	(n.s.)
var. binnen leerlingen	115.40	
var. binnen scholen	85.03	
var. in tijd effect binnen scholen	16.85	
var. tussen scholen	25.17	
var. in seks * tijd effect tussen scholen	6.23	(n.s.)

De resultaten laten zien dat jongens aanvankelijk 6.55 punten hoger scoren op de wiskundetoets in leerjaar 1 dan de meisjes. Reeds in het begin van het VWO zijn er dus al grote seksverschillen in wiskundeprestaties. Daarnaast zien we, dat leerlingen een ontwikkeling doormaken in hun wiskundekennis: per jaar neemt hun score met ruim 7 punten toe. De verschillen in aanvangsniveau tussen leerlingen binnen scholen zijn met een variantie van 85.03 aanmerkelijk groter dan de verschillen tussen scholen (een variantie van 25.17). Deze laatste parameterwaarde suggereert dat tussen de top 5 procent en de onderste 5 procent scholen een verschil in aanvangsniveau wordt waargenomen van $2 * 1.64 * \sqrt{25.17}$, ofwel 16.46 punten. Leerlingen verschillen in hun ontwikkeling van wiskundekennis. Het effect van tijd varieert namelijk tussen de leerlingen met een geschatte variantie van 16.85. Dit impliceert dat de top 5 procent leerlingen groeien ($2 * 1.64 * \sqrt{16.85}$) 13.46 punten sneller per jaar dan de onderste 5 procent leerlingen. Tussen scholen treedt er echter geen (significant) verschil in gemiddelde ontwikkelingssnelheid op. Voorts zien we dat er geen verschillen optreden tussen jongens en meisjes in de ontwikkeling van hun wiskunde prestaties (het seks*tijd effect is namelijk niet significant). Ondanks de afwezigheid van dit effect zou het natuurlijk nog wel zo kunnen zijn dat op sommige scholen de jongens zich relatief gunstig ontwikkelen waar op andere scholen de meisjes zich relatief gunstig ontwikkelen. We zien echter dat de variantie in het seks*tijd effect tussen de scholen niet systematisch van nul afwijkt: scholen verschillen met andere woorden niet in de seksespecifieke ontwikkeling van wiskunde prestaties.

Schoolverschillen in seksespecifiek keuzegedrag

Initiële analyse van verschillen tussen scholen in de keuze van wiskunde laten zien dat er aan de top scholen zijn waar bijna alle jongens wiskunde B kiezen tegenover slechts enkele van de meisjes, en dat er ook scholen zijn waar meisjes relatief vaker dan de jongens dit vak kiezen (vgl. grafiek 1). We zullen nagaan in hoeverre dergelijke tussen-schoolse verschillen toevallig danwel systematisch zijn. Aangezien de afhankelijken qua meetniveau van elkaar verschillen (voor de eerste drie leerjaren is er een continue variabele voor de voorgenomen keuze van wiskunde B, terwijl voor leerjaar 4 er een dichotome variabele is), zullen de analyses per leerjaar worden uitgevoerd. De resultaten staan vermeld in tabel 5.

Tabel 5. Sekse- en schoolverschillen in keuzegedrag.

	leerjaar 1	leerjaar 2	leerjaar 3	leerjaar 4
intercept (meisjes)	2.779	3.081	3.264	-.462
seks effect (jongens)	0.199(n.s.)	0.136(n.s.)	0.627	1.016
var. binnen scholen	1.193	1.717	1.837	1.000
var. tussen scholen	0.000(n.s.)	0.043(n.s.)	0.000(n.s.)	0.076
var. in seks-effect tussen scholen	0.000(n.s.)	0.151(n.s.)	0.000(n.s.)	0.626

In het eerste en tweede leerjaar blijken jongens en meisjes nog niet (significant) van elkaar te verschillen in het voornemen wiskunde B al dan niet te kiezen. Ook scholen vertonen in dit opzicht geen onderlinge verschillen. In leerjaar 3 echter, treden er wel sekse-verschillen op: meisjes weten in het algemeen nog niet of ze wiskunde B als vak zullen kiezen, waar jongens gemiddeld genomen geneigd zijn het vak op te nemen in het pakket. Wederom echter verschillen scholen niet van elkaar in de voorgenomen keuzen van hun leerlingen, noch in de sekseverschillen die hierin optreden. In leerjaar 4 (het moment van de feitelijke keuze) treden er echter meerdere verschillen op. Aangezien de data met een multilevel logistisch model zijn geanalyseerd, waarbij gewerkt is met de logit van de afhankelijke moeten de resultaten terugvertaald worden naar de oorspronkelijke variabele. In de analyses is gewerkt met de afhankelijk variabele y' ($=\ln(y/(1-y))$). We kunnen nu y terugvinden met behulp van de volgende transformatie: $y=e^{y'}/(1+e^{y'})$. De kans dat een meisje wiskunde B kiest is $e^{-.462}/(1+e^{-.462})=.38$. Voor jongens wordt die kans geschat op $e^{-.462+1.016}/(1+e^{-.462+1.016})=.64$. De variantie tussen scholen wordt geschat op .076, hetgeen betekent dat verschillen tussen de scholen in de populariteit van wiskunde B bij meisjes volgens het negentig-procents predictieinterval liggen tussen $e^{-.462-1.64*\sqrt{.065}}/(1+e^{-.462-1.64*\sqrt{.065}})$ en $e^{-.462+1.64*\sqrt{.065}}/(1+e^{-.462+1.64*\sqrt{.065}})$, ofwel tussen de .30 en de .49. Voor de jongens zijn de tussen-schoolse verschillen aanmerkelijk groter: $e^{-.462+1.016-1.64*\sqrt{.626}}/(1+e^{-.462+1.016-1.64*\sqrt{.626}})$ en $e^{-.462+1.016+1.64*\sqrt{.626}}/(1+e^{-.462+1.016+1.64*\sqrt{.626}})$, ofwel tussen de .32 en de .86. Om deze schoolverschillen in sekse-specifieke keuzegedragingen te verduidelijken wordt nogmaals grafiek 1 gepresenteerd, maar nu met uitzuivering van de steekproevenvariantie. De rationale achter deze posterior proporties zou men kunnen samenvatten als: we hebben een steekproef van leerlingen binnen elke school getrokken; naarmate de omvang van deze steekproef toeneemt zijn we er zekerder van dat we het echte schooleffect te pakken hebben; naarmate de steekproefomvang afneemt zijn we daar minder zeker van en maken we gebruik van de (geschatte) gegevens over de populatie van leerlingen (waar deze leerlingen immers ook uit afkomstig zijn). De resultaten staan vermeld in Figuur 2.

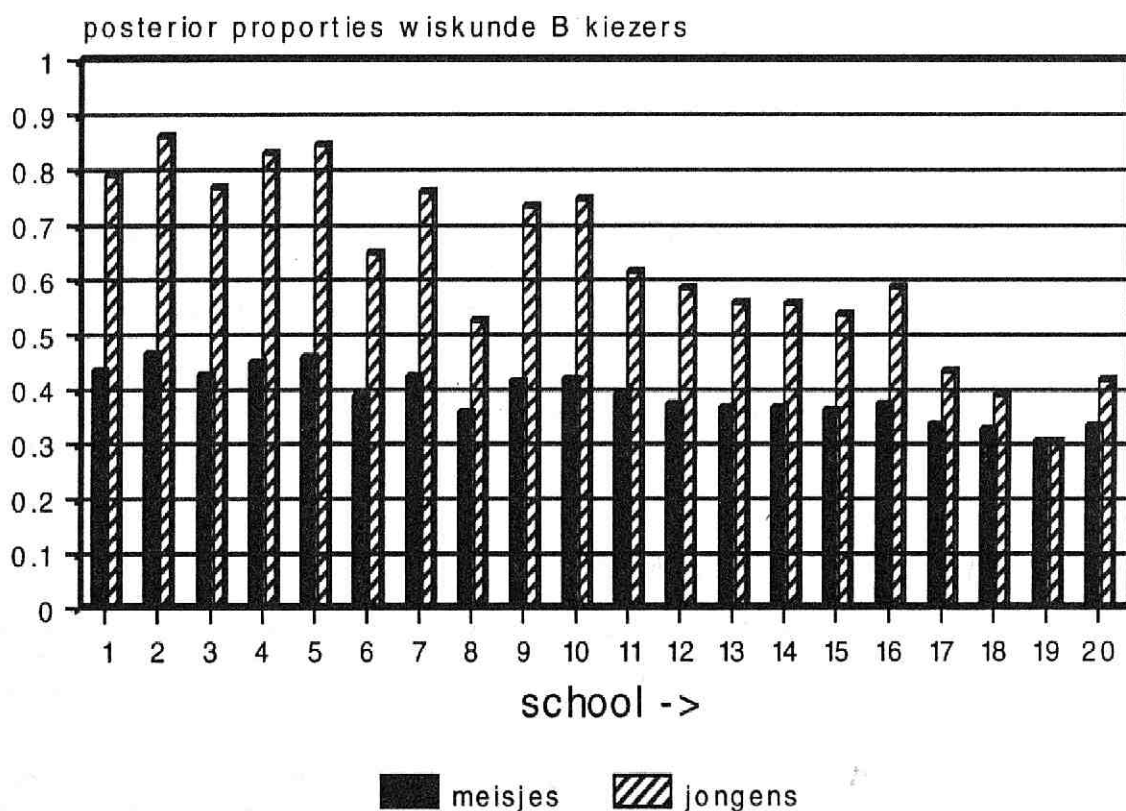


Fig. 2. Geschatte percentages meisjes en jongens dat wiskunde B kiest per school.

Op geen van de 20 scholen nemen meisjes vaker dan jongens wiskunde B op in het pakket. De eerder in grafiek gepresenteerde ruwe resultaten blijken dus in dit opzicht op toeval te berusten.

Het meritocratisch gehalte van scholen

De factor school lijkt dus niet van groot belang te zijn om sekseverschillen in voorgenomen keuzen te kunnen verklaren, maar wel in feitelijke keuzen. Daarnaast hebben we gezien dat scholen noch verschillen vertoonden in seksepecifieke wiskundeprestaties noch in de seksepecifieke ontwikkeling van die prestaties. Het lijkt er dus op dat bij gelijke prestatieniveaus scholen verschillen in de keuzen die door jongens en meisjes worden gemaakt. Ook voor de eerste drie leerjaren zou het zo kunnen zijn dat er wel verschillen tussen scholen zijn waar het de omzetting van wiskundeprestaties in keuzegedrag betreft: mogelijkwijze zijn er scholen waar leerlingen eerder wiskunde B kiezen dan op grond van hun prestaties verwacht zou worden. En mogelijkwijze ligt dit verband voor jongens en meisjes anders. Wellicht ligt op sommige scholen de mate waarin het verband tussen prestaties en keuzen tussen jongens en meisjes verschilt anders dan op andere scholen. Mogelijkwijze wordt op sommige scholen puur gekozen op basis van prestaties, terwijl op andere scholen wellicht een voor meisjes stimulerend beleid wordt gevoerd om wiskunde B te kiezen ongeacht het behaalde prestatieniveau. We onderzoeken daarom of het interactie-effect van sekse*prestaties (het betreft hier de initiële prestaties) op keuzen tussen scholen verschilt. Wederom onderzoeken we een en ander per leerjaar. Om te zorgen dat varianties tussen scholen geïnterpreteerd kunnen worden als verschillen tussen scholen voor meisjes met een gemiddeld prestatieniveau, worden de scores op de wiskundetoets gecentreerd rond het gemiddelde (vgl. Bosker & Snijders, 1990).

Tabel 6. Het derde orde interactie effect van school*sekse*prestatie op keuzegedrag.

	leerjaar 1	leerjaar 2	leerjaar 3	leerjaar 4
intercept (meisjes)	2.809	3.135	3.336	-.336
sekse effect (jongens)	0.164(n.s.)	0.008(n.s.)	0.495	0.847
prestatie	0.010(n.s.)	0.011(n.s.)	0.020	0.057
interactie effect van prestatie * sekse	-.012(n.s.)	0.004(n.s.)	-.006(n.s.)	-.021(n.s.)
var. binnen scholen	1.133(n.s.)	1.668	1.727	1.000
var. tussen scholen	0.032(n.s.)	0.076(n.s.)	0.000(n.s.)	0.01557(n.s.)
var. in prestatie * sekse effect tussen scholen	0.000(n.s.)	0.001(n.s.)	0.000(n.s.)	0.00223

Het eerste opvallende feit is dat, naast de reeds eerder geconstateerde afwezigheid van sekse-effecten in de eerste twee leerjaren, de seksegebonden verschillen in keuzegedrag in het derde leerjaar niet verdwijnen als de voorafgaande toetsprestaties als voorspeller in het model worden opgenomen. Sekseverschillen in het derde leerjaar in de neiging om wiskunde B als examenvak te kiezen kunnen dus niet 'verklaard' worden vanuit prestatieverschillen tussen jongens en meisjes. In geen van de drie leerjaren blijkt er een interactie-effect van prestatie*sekse op voorgenomen keuzen op te treden, noch treden in dit opzicht verschillen op tussen scholen. Maar in het vierde leerjaar geldt dit niet: meisjes kiezen ook feitelijk minder dan jongens wiskunde B in het examenpakket, zelfs als de invloed van hun mindere wiskundeprestaties op het keuzegedrag is verdisconteerd. Voor een jongen en een meisje van gemiddeld prestatieniveau bedragen de geschatte proporties respectievelijk .62 en .42. Het tweede opmerkelijke resultaat is, dat de samenhang tussen prestaties en keuzegedrag voor jongens en meisjes hetzelfde is, zoals we kunnen aflezen aan het niet significante interactie-effect tussen prestatie en sekse op de feitelijke keuze. Wel treden er in dit opzicht verschillen tussen scholen aan het licht. Het 90%-predictie-interval voor het effect van sekse*toets op feitelijke keuzen is $-.021 \pm 1.64 * \sqrt{.01557}$,

ofwel $-.098$ tot $.056$. Aangezien sekse gecodeerd is als 0 voor meisjes en 1 voor jongens, dient dit effect als volgt geïnterpreteerd te worden: voor meisjes geldt dat het effect van prestaties op uiteindelijke keuzen voor alle scholen hetzelfde is (nl. 0.057). Voor jongens daarentegen maakt het bij een gegeven prestatieniveau wel degelijk uit op wat voor school ze zitten. Voor jongens is het predictie-interval voor het toets-effect $\langle .057; .098 \rangle$, ofwel op sommige scholen is het toetseffect voor jongens -0.041 en op andere scholen $.113$. In Figuur 3 zijn de geschatte kansen voor jongens met een lage en hoge toetsscore (respectievelijk één standaarddeviatie onder het gemiddelde voor jongens en één er boven) op de keuze van wiskunde B per school weergegeven.

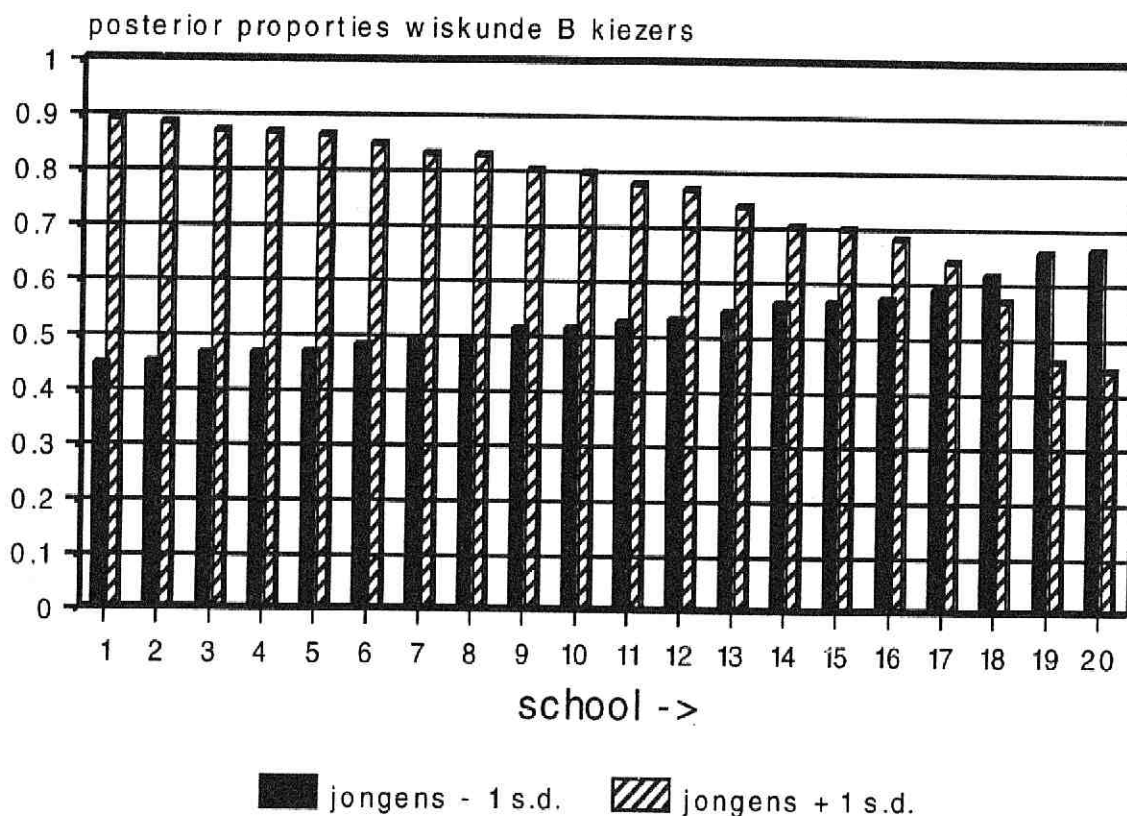


Fig. 3. Geschatte percentages jongens met een lage en een hoge toetsscore dat wiskunde B kiest per school.

DISCUSSIE

De resultaten samenvattend, kunnen we stellen dat jongens eerder dan meisjes geneigd zijn wiskunde B als examenvak te kiezen. De verschillen nemen bovendien toe naarmate het moment suprême nadert: het feitelijke keuzemoment in leerjaar 4. De rol die de school speelt in dit proces kan als volgt worden samengevat: aangezien scholen niet verschillen in de seksspecifieke ontwikkeling van de wiskundeprestaties van hun leerlingen, maar wel in het uiteindelijk keuzegedrag ten aanzien van wiskunde B in leerjaar 4, bevreemdt het niet dat scholen verschillen in de wijze waarop prestatieverschillen tussen leerlingen worden omgezet in keuzeverschillen. Sommige scholen zijn meritocratischer dan andere in de wijze waarop ze prestaties seksspecifiek omzetten in vakkenkeuze. We moeten constateren dat de schoolverschillen in seksspecifiek keuzegedrag weliswaar kleiner zijn dan op het eerste gezicht leek, maar dat ook een

meer geeigende (multi-level) analyse resulteert in de bevinding dat de school er in dit opzicht toe doet. Opgemerkt dient te worden dat vergelijkbare analyses voor het HAVO (Bosker & Dekkers, 1994) dergelijke verschillen niet te zien geven.

Aangezien het vak wiskunde B een belangrijk selectievak is met het oog op vervolgstudies in het hoger onderwijs lijkt het zinvol in (zorgvuldig ontworpen) schooleffect studies schoolfactoren op te sporen die tot seksespecifiek keuzegedrag leiden. De relatie tussen prestatie en keuze loopt blijkbaar gedeeltelijk via de school. Zoals reeds eerder opgemerkt (vgl Bosker, 1990) blijkt echter ook nu weer dat schooleffecten met name optreden voor jongens. Dit geldt zowel de keuze van wiskunde B als de kans op de keuze van wiskunde B bij een gegeven prestatieniveau; in dit laatste geval blijken schooleffecten met name op te treden voor jongens met een aanvankelijk hoog prestatieniveau in het vak wiskunde. Tenslotte zij opgemerkt, dat niet iedere school alle potentiële B-kiezers onder de meisjes zo ver krijgt, dat ze het vak dat hun kansen op vervolgopleidingen en op de arbeidsmarkt vergroot, ook daadwerkelijk kiezen.

LITERATUUR

- Bosker, R.J. (1990). *Extra kansen dankzij de school?* Nijmegen: ITS/OoMO.
- Bosker, R.J. & Dekkers, H.P.J.M. (1994). School differences in producing gender related subject choices. To appear in: *School Effectiveness and School Improvement*.
- Bosker, R.J. & Snijders, T.A.B. (1990). Statistische aspecten van multi-niveau onderzoek. *Tijdschrift voor Onderwijsresearch*, 15, 5, 317-329.
- Bosker, R.J. & Velden, R.K.W. van der (1989). Schooleffects on educational perspectives. In B. Bakker, J. Dronkers & G.W. Meijnen (Eds.): *Educational opportunities in the welfare state. Longitudinal research in educational and occupational attainment in the Netherlands*. Nijmegen: ITS/OoMO.
- Brookover, W. (1979). *School social systems and student achievement*. New York: Praeger publishers.
- Bryk, A.S. & Raudenbush, S.W. (1992). *Hierarchical Linear Models*. Newbury Park/London/New Delhi: Sage publications.
- Dekkers, H.P.J.M. (1985). *Soms kiezen meisjes anders*. Nijmegen: ITS.
- Dekkers, H.P.J.M. (1990). *Seksespecifieke studiekeuzen in het wetenschappelijk onderwijs*. Nijmegen: ITS.
- Dekkers, H.P.J.M. (1992). *Determinanten van seksespecifieke vakkenkeuze in het voortgezet onderwijs (reader)*. Nijmegen: ITS (interne publicatie no. 2040).
- Dekkers, H. (1993a). *Onderwijsemancipatiebeleid en economische zelfstandigheid van vrouwen in Noordwest Europa*. Nijmegen: ITS.
- Dekkers, H. (1993b). Determinanten van seksespecifieke vakkenkeuzen in HAVO en VWO. *Pedagogische Studiën*, 70, 445-459.
- Eccles, J.S. (1984). Sex differences in achievement patterns. *Nebraska Symposium on Motivation*, 32, 97-132.
- Eccles, J.S. (1986). *Gender roles and women's achievement*. San Francisco: AERA-paper.
- Eck, E. van & Volman, M. (1990). *Determinanten van seksespecifieke keuzen en prestaties*. Symposiumverslag. Amsterdam: SCO.
- Hanna, G. (1994) Cross cultural differences in mathematics education. To appear in: *International Journal of Educational Research*.
- Hyde, J.S., Fennema, E., & Lamon, S.J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, 107, 139-155.
- Kremers, E.J.J. (red.) (1990). *Overzicht van leerresultaten aan het einde van de eerste fase voortgezet onderwijs*. Arnhem: CITO.
- Kristensen, D. & Jenneskes, A. (1991). *Emancipatiebeleid op schoolniveau*. Lisse: Swets & Zeitlinger.
- Kuyper, H. & Meulenbeld, J. (1989). *De invloed van het gedrag van docenten op de prestaties in, keuze van, en attitudes ten opzichte van wiskunde door meisjes in het AVO/VWO*. Groningen: RION.
- Kuyper H. & Werf, M.P.C. van der (1987). *De invloed van het gedrag van docenten op de prestaties in keuzen van en attitudes ten opzichte van wiskunde door meisjes in het AVO/VWO. Verslag van fase I*. Groningen: RION.
- Longford, N.L. (1989). A quasi-likelihood adaptation for variance component analysis. In P. van den Eeden, J. Hox & J. Hauer (Eds.): *Theory and model in multilevel research*. Amsterdam: SISWO.
- Pelgrum, W.J. & Plomp, T.J. (1986). *Second International Science Study. Beschrijving van uitkomsten en analyse*. Enschede: Universiteit Twente.

- Prosser, R., Rasbash, J., & Goldstein, H. (1991). *ML3: Software for three-level analysis*. London Institute of Education, University of London.
- Rutter, M., Maughan, B., Mortimer, P., & Ouston, J. (1979). *Fifteen thousand Hours. Secondary schools and their effects on children*. London: Open Books.
- Scheerens, J. (1989). *Wat maakt scholen effectief? Samenvatting en analyse van resultaten*. s'Gravenhage: SVO.

Manuscript ontvangen 10-12-1993

Definitieve versie ontvangen 3-5-1994

Schoolprestaties van allochtone leerlingen: de invloed van etnische herkomst

E.J. Latuheru en M.G.P. Hessels
*RISBO, Erasmus Universiteit Rotterdam**

ABSTRACT

Research into the educational position of pupils shows that allochtonous pupils are behind in comparison with indigenous pupils. This article attempts to answer the question whether differences in pupils' school records can be attributed to their ethnic descent, or whether they are caused by the intervening effect of their socio-economic descent and/or intelligence. It is determined that, after controlling for socio-economic descent and intelligence, the pupils' ethnic descent does not have a significant effect on their school record. However, there is also no significant effect of social-economic descent after controlling for ethnic descent and intelligence. The effect of both variables appears to be highly interchangeable. In conclusion it can be stated that, due to the fact that ethnic and social-economic descent are mutually contaminating, it cannot be determined whether the pupils' ethnic descent contributes to an explanation of the differences in school records.

INLEIDING: SCHOOLPRESTATIES VAN ALLOCHTONE LEERLINGEN

Nederland wordt gekenmerkt door een multi-etnische samenstelling van haar bevolking. Dit kenmerk is een gevolg van de emigratie en immigratie die Nederland heeft gekend in haar geschiedenis (Entzinger, 1984). Deze migratie-bewegingen werden op de eerste plaats in belangrijke mate bepaald door de activiteiten van Nederlanders op het gebied van handel, nijverheid en zeevaart. Daarnaast kent Nederland een klimaat van relatieve geestelijke vrijheid en tolerantie, waardoor veel religieuze en politieke vluchtelingen naar Nederland zijn gekomen. Een toename in de migratie naar Nederland in de twintigste eeuw werd hoofdzakelijk veroorzaakt door de onafhankelijkheidswording van voormalige Nederlandse koloniën én door een tekort aan arbeidskrachten in de jaren zestig en zeventig als gevolg van een periode van economische groei. Hoewel er thans sprake is van een afnemende migratie naar Nederland, blijft de omvang van de allochtone bevolkingsgroep toenemen (Roelandt, Roijen & Veenman, 1991). Dit is niet alleen het gevolg van de nog steeds bestaande migratie van met name politieke vluchtelingen, maar ook van de gezinshereniging en gezinsvorming waartoe de zich hier gevestigde migranten overgingen (Van der Erf, 1988). Gezien deze ontwikkelingen ligt het niet in de lijn der verwachting dat het aantal allochtonen of de groei van dat aantal in de komende jaren zal afnemen.

De naar Nederland gemigreerde personen en hun gezinnen nemen deel aan de samenleving, maar zij blijken daarbij op een aantal maatschappelijke terreinen minder succesvol te zijn dan de autochtone Nederlanders (Penninx, 1988; Roelandt & Veenman, 1988). Tussen allochtonen en autochtonen enerzijds en tussen verschillende allochtone bevolkingsgroepen anderzijds blijken onder andere verschillen te bestaan in deelname aan het onderwijs, positie op de arbeidsmarkt, huisvesting en de mate waarin wordt deelgenomen aan de Nederlandse cultuur.

In dit onderzoek staan verschillen in schoolprestaties van autochtone en allochtone leerlingen centraal. Schoolprestaties zijn in aanzienlijke mate bepalend voor iemands verdere beroepsloopbaan (De Jong, 1991), aangezien opleiding (naast onder andere ervaring) een belangrijk criterium is bij het wel of niet in aanmerking komen voor een bepaalde maatschappelijke functie in de samenleving (De Grip, 1988). Uit onderzoek is gebleken, dat allochtone kinderen gemid-

* Adres: Postbus 1738, 3000 DR Rotterdam.

deld een lagere onderwijspositie behalen dan autochtone Nederlandse kinderen (De Jong, 1987; Van Langen & Jungbluth, 1990; Roelandt & Veenman, 1988; De Wit, Suhre & Mulder, 1993). Bovendien doen allochtone kinderen er langer over dan autochtone kinderen om een bepaalde onderwijspositie te bereiken. Daarnaast hebben zij bij het bereiken van die onderwijspositie meer te maken met schoolverzuim en schooluitval dan autochtone kinderen (Kerkhoff, 1989). Hoewel er steeds meer allochtone leerlingen zullen zijn die hun volledige schoolloopbaan in Nederland hebben doorlopen en de gemiddelde onderwijspositie van allochtone leerlingen zal stijgen, vanwege de in onderzoek geconstateerde samenhang tussen verblijfsduur en schoolprestaties (De Jong, 1985; Van Hooft, 1987), is het niet waarschijnlijk dat allochtone leerlingen op korte termijn hun onderwijsachterstand ten opzichte van autochtone leerlingen zullen inlopen (Van Hooft, 1987).

OORZAKEN VAN VERSCHILLEN IN SCHOOLPRESTATIES: THEORIE

Dat allochtone leerlingen gemiddeld genomen minder goede schoolprestaties behalen dan autochtone leerlingen staat buiten kijf. Het antwoord op de vraag waarom dat zo is, is minder eenduidig. Pas in de jaren tachtig is men begonnen met het doen van onderzoek naar de oorzaken van de onderwijsachterstanden van allochtone kinderen. Theorievorming ten aanzien van dit onderwerp heeft slechts incidenteel plaatsgevonden. Bij het onderzoek naar de oorzaken van de achterstandspositie van allochtone leerlingen wordt in het algemeen aangesloten bij het onderzoek naar oorzaken van verschillen in de schoolprestaties van leerlingen in het algemeen. Geconstateerd kan worden dat veel onderzoek op dit gebied is gebaseerd op één (of meerdere) van de volgende theoretische aannames:

- a. Verschillen in schoolprestaties hangen samen met verschillen in intelligentie;
- b. Verschillen in schoolprestaties hangen samen met verschillen in milieu van herkomst;
- c. Verschillen in schoolprestaties hangen samen met verschillen in schoolkenmerken.

Een verklaring voor het bestaan van verschillen in schoolprestaties van leerlingen is dat deze verschillen te wijten zijn aan verschillen in intelligentie (Dronkers, 1986). Uit onderzoek blijkt dat allochtone kinderen in het algemeen systematisch lagere scores behalen op intelligentietests dan autochtone kinderen (Hessels, 1993; Resing, Bleichrodt & Drenth, 1986; Vernon, 1979). De vraag is of dat te wijten is aan het minder intelligent zijn van allochtone kinderen, of dat deze lagere scores veroorzaakt worden door het feit dat de gehanteerde tests inadequate schattingen geven van de intelligentie van allochtone leerlingen (Jensen, 1961; Rohwer, 1971; Sewell, 1979), bijvoorbeeld omdat zij minder kans hebben gehad om de kennis en vaardigheden die in de test aan bod komen te verwerven (Guthke, 1993), of omdat zij minder testvaardig zijn dan autochtone leerlingen (Sarnacki, 1979; Röhr-Sendlmeier, 1987).

Een andere verklaring voor het bestaan van onderwijsachterstanden is, dat deze achterstanden te maken hebben met verschillen in milieu van herkomst (Dronkers, 1986). Uit onderzoek blijkt dat er sprake is van een samenhang tussen de leerprestaties van kinderen en het ouderlijk milieu waaruit zij afkomstig zijn (Van Heek, 1968). Dit effect zou worden veroorzaakt door milieuspecifieke socialisatie (Leseman, 1989; Tesser, 1986), vanwege een discrepantie tussen het geïnternaliseerde cultuurpatroon van een kind en de dominante cultuur van het onderwijsstelsel. Schoolsucces is derhalve afhankelijk van de mate van overlap tussen de beide (sub-)culturen.

De meest recente verklaring voor het bestaan van onderwijsongelijkheid is dat scholen ten aanzien van bepaalde kenmerken verschillen, en daardoor verschillen in de mate van effectiviteit van het door hun aangeboden onderwijs. Schoolkenmerken die relevant worden geacht, blijken te convergeren rondom sterk onderwijskundig leiderschap, accentuering van de verwerking van basisvaardigheden, een ordelijk en veilig klimaat, het hebben van hoge verwachtingen van de prestaties van leerlingen, en een frequente evaluatie van de vorderingen van de leerlingen (Scheerens, 1989). De onderscheiden factoren blijken in het Nederlandse effectiviteitsonderzoek echter niet duidelijk als relevante kenmerken van effectieve scholen naar voren te

komen (Scheerens, 1989). De invloed van onderwijskundig leiderschap bijvoorbeeld wordt niet eenduidig in Nederlands onderzoek teruggevonden (vgl. Van der Grift, 1990; Van der Grift & Akkermans, 1991). Dat enig effect van bepaalde schoolkenmerken op schoolprestaties zich voordoet lijkt echter zonder twijfel. De kwaliteit van de algemene effectieve instructiekenmerken (Van der Werf & Weide, 1991) en het door schoolbesturen gevoerde beleid (Hofman, 1992) blijken mede bepalend te zijn voor de effectiviteit van scholen. Overigens wijst onderzoek uit dat het effect van sommige schoolkenmerken op schoolprestaties sterker is bij allochtone dan bij autochtone leerlingen (Hofman, 1993; Van der Werf & Weide, 1991). In deze bijdrage zal aan de eventuele invloed van kenmerken van scholen op de schoolprestaties van leerlingen worden voorbij gegaan.

OORZAKEN VAN VERSCHILLEN IN SCHOOLPRESTATIES: ONDERZOEK

Bij het overzien van het onderzoek naar de oorzaken van de onderwijsachterstanden van allochtone leerlingen wordt allereerst geconstateerd, dat intelligentie veelal niet is opgenomen in het onderzoeksdesign (vgl. Van Langen & Jungbluth, 1989; Van 't Hof & Dronkers, 1992). Indien intelligentie daarin wel is opgenomen, is niet of nauwelijks gecontroleerd op de mate van cultuurredelijkheid ten aanzien van de verschillende etnische groepen (De Jong, 1987). Het is niet terecht dat intelligentie niet of nauwelijks aan bod komt in het onderzoek naar onderwijsachterstanden van allochtone leerlingen, daar de intelligentie van leerlingen een belangrijke voorspeller is van hun schoolprestaties (Hessels, 1993; Kaufman, Kamphaus & Kaufman, 1985; Resing, Bleichrodt & Drenth, 1986; Sewell, 1979). Een mogelijke reden voor het buiten beschouwing laten van intelligentie is de eventuele cultuurbepaaldheid van intelligentietests. Deze cultuurbepaaldheid is ongewenst, omdat tussen leden van verschillende populaties verschillen in testcores kunnen ontstaan die (gedeeltelijk) dienen te worden toegeschreven aan variatie in eigenschappen die men niet bedoelt te meten (Drenth & Sijtsma, 1990). Er zijn procedures waarmee intelligentietests (en ook schoolprestatietoetsen) op cultuurbepaaldheid onderzocht kunnen worden, en aan de hand waarvan – door het verwijderen van onzuivere (biased) items – cultuurredelijke testresultaten verkregen kunnen worden (Glas & Ouborg, 1993; Jensen, 1980; Osterlind, 1983; Van der Flier, 1980).

Daarnaast kan geconstateerd worden dat in onderzoek veelal wordt geconcludeerd dat etnisch-culturele verschillen tussen leerlingen, onder controle van sociaal-economische verschillen, niet of nauwelijks relevant zijn voor de verklaring van schoolprestaties (Driessen, 1993; De Jong, 1987; Van Langen & Jungbluth, 1990; Van 't Hof & Dronkers, 1992). De beleidsvoering ten aanzien van het bestrijden van onderwijsachterstanden is dan ook gebaseerd op de veronderstelling dat sociaal-economische status de belangrijkste determinant is van individuele schoolprestaties. Het onderwijsvoorrrangsbeleid ter bestrijding van onderwijsachterstanden van kansarme autochtonen wordt tevens geacht de achterstanden van allochtonen te verkleinen (Klopprogge, 1989). Het is derhalve aannemelijk dat in toekomstig onderzoek naar de oorzaken van onderwijsachterstanden van migranten zal worden uitgegaan van het standpunt dat sociaal-economische aspecten van het milieu van herkomst de voornaamste verklarende factoren zijn. Het onderscheid tussen autochtone en allochtone leerlingen betreft echter niet alleen sociaal-economische, maar ook etnisch-culturele verschillen. Het is de vraag of deze twee situaties zonder meer tot elkaar kunnen worden herleid (Tesser, 1989).

PROBLEEMSTELLING: INVLOED VAN ETNISCHE HERKOMST?

De idee dat slechtere schoolprestaties van allochtone leerlingen verklaard kunnen worden door alleen hun sociaal-economische herkomst, wordt in deze bijdrage ter discussie gesteld. De achtergrond hiervan is de veronderstelling dat de invloed van de etnische herkomst van leerlingen op hun schoolprestaties niet volledig verklaard wordt door hun lagere sociaal-economische

positie. Achterstandsbestrijdend beleid dat alleen een sociaal-economische basis kent, en voorbij gaat aan de mogelijke etnisch-culturele determinanten van onderwijsprestaties, is vanuit deze gedachte onvoldoende.

Omdat geconstateerd wordt dat de intelligentie van leerlingen een belangrijke voorspeller is van hun schoolprestaties, wordt het eventuele effect van etnische herkomst ook voor deze variabele gecontroleerd. Aldus zijn de volgende onderzoeksvragen geformuleerd:

- a. Is er sprake van een effect van de etnische herkomst van leerlingen op hun schoolprestaties?
- b. Is er sprake van een effect van de etnische herkomst van leerlingen op hun schoolprestaties na controle voor hun sociaal-economische herkomst?
- c. Is er sprake van een effect van de etnische herkomst van leerlingen op hun schoolprestaties na controle voor hun sociaal-economische herkomst en intelligentie?

ONDERZOEKSOPZET

Om de bovenstaande vragen te kunnen beantwoorden dienen verschillende gegevens van leerlingen verzameld te worden aan de hand waarvan de etnische herkomst, de schoolprestaties, de intelligentie en de sociaal-economische herkomst van leerlingen kunnen worden bepaald. De gegevens waarop de analyses zijn uitgevoerd, komen uit een bestand dat gecreëerd is in het kader van het SVO-project "Invloed Schoolspectifieke Factoren bij Allochtone Leerlingen (IS-FAL)" (Hofman, 1993). Het betreft een steekproef van tweedejaars leerlingen uit een steekproef van scholen voor LTO, LHNO, LEAO, MAVO en HAVO. Deze steekproef van leerlingen is representatief voor de totale leerlingpopulatie van de hiervoor genoemde schooltypen (Hofman, Latuheru & Van der Velden, 1989). De gegevens zijn verzameld in het voorjaar van 1989 en hebben betrekking op 3066 leerlingen.

Het antwoord op de onderzoeksvragen is verkregen door toetsing van de in figuur 1 weergegeven modellen.

In het eerste model wordt onderzocht of er een relatie bestaat tussen etnische herkomst en schoolprestaties. Als dat het geval is, wordt vervolgens onderzocht of deze relatie blijft bestaan na correctie voor sociaal-economische herkomst (model II). In het laatste model tenslotte wordt onderzocht of er sprake is van invloed van de etnische herkomst van leerlingen op hun schoolprestaties na controle voor hun sociaal-economische herkomst en intelligentie. Voordat de

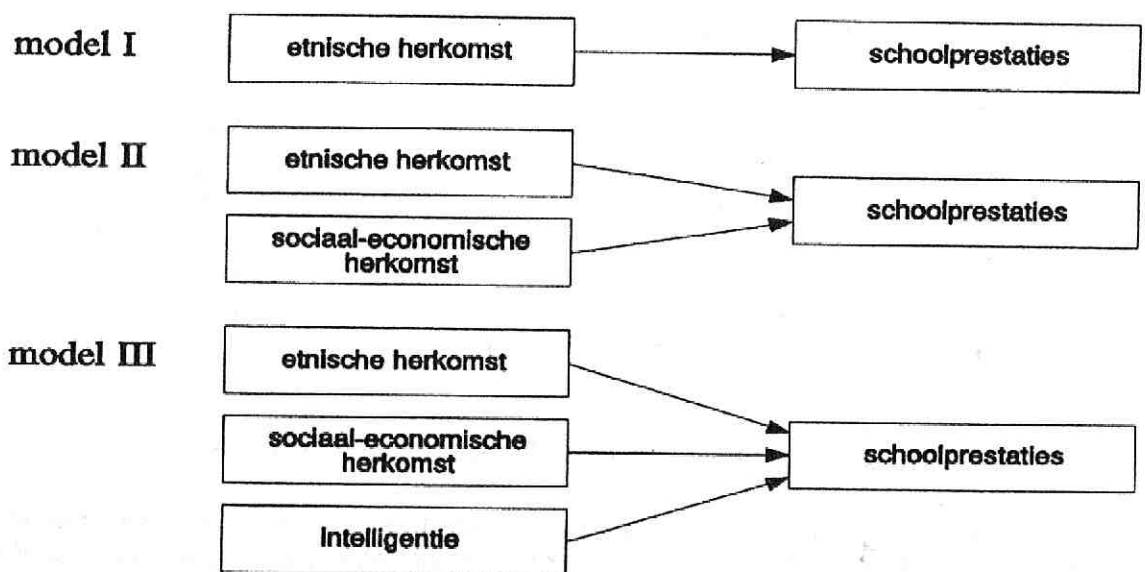


Fig. 1. Te toetsen modellen.

modellen getoetst zullen worden, zal eerst ingegaan worden op de variabelen in het model en hoe deze tot stand zijn gekomen.

CENTRALE ONDERZOEKSVARIABELEN

Etnische herkomst

Met iemands etnische herkomst wordt de etnische identiteit volgens de primordiale benadering bedoeld. In primordiale benaderingen wordt de traditionele cultuur van een volk bepalend geacht voor de etnische identiteit van de individuele cultuurdragers (Van Soest & Verdonk, 1984). De etnische herkomst wordt aldus opgevat als een statische eigenschap: als men als lid van een bepaald volk geboren wordt, verandert de etnische identiteit niet wezenlijk als men zich in een andere cultuur begeeft en beweegt. Ter vergelijking, in situationele benaderingen van etnische identiteit wordt identiteit gezien als een dynamisch gegeven (Van Soest & Verdonk, 1984).

In dit onderzoek is de etnische herkomst van een leerling vastgesteld aan de hand van het objectieve criterium geboorteland van de ouders. Een leerling wordt tot een bepaalde etnische groep gerekend als beide ouders in het corresponderende land zijn geboren. De etnische groepen die in dit onderzoek worden onderscheiden zijn Nederlanders, Surinamers, Turken en Marokkanen. De overige te onderscheiden etnische groepen zijn buiten de analyses gehouden enerzijds vanwege de geringe aantallen (bijvoorbeeld Antillianen en Molukkers) en anderzijds vanwege de grote etnische variatie binnen een categorie (met name de kinderen uit gemengde huwelijken).

Sociaal-economische herkomst

Gekozen is voor het vaststellen van de sociaal-economische herkomst van leerlingen op basis van het opleidingsniveau van de beide ouders. Gezien de hoge werkloosheid van de beroepsbevolking in Nederland (Roelandt, Roijen & Veenman, 1991), is het aan te bevelen om niet het beroepsniveau van beide ouders te hanteren als indicator voor de sociaal-economische herkomst. Als iemand geen beroep uitoefent, kan immers niet gesproken worden over het beroepsniveau van deze persoon. Het opleidingsniveau van een persoon is wat dat betreft een betrouwbaarder gegeven, omdat het een vaststaand gegeven betreft, terwijl het beroepsniveau als gevolg van werkloosheid of arbeidsongeschiktheid kan variëren.

Het opleidingsniveau van de beide ouders is vastgesteld op basis van het hoogst gevolgde en afgemaakte onderwijstype. Gezien de oververtegenwoordiging van allochtone ouders in de lagere onderwijsniveaus zijn de oorspronkelijk onderscheiden categorieën teruggebracht tot [a] geen onderwijs, [b] alleen lager onderwijs en [c] voortgezet onderwijs. Vervolgens is met behulp van een eendimensionale HOMALS-oplossing een samengestelde variabele geconstrueerd, waarin de opleidingsniveaus van de beide ouders zijn gekwantificeerd tot een variabele van interval meetniveau, te weten sociaal-economische herkomst. De eigenwaarde van de gevonden oplossing is hoog (.87), hetgeen inhoudt dat er sprake is van weinig informatieverlies (Van den Berg, 1986).

Intelligentie

Een indicatie voor de intelligentie van een leerling is in dit onderzoek verkregen aan de hand van een figurale subtest uit de Nederlandse versie van het 'Prüfsystem für Schul- und Bildungsberatung' (PSB), te weten de nonverbale speedtest PSB3, die geacht wordt het vermogen tot logisch redeneren te meten (Horn, 1969). De test bestaat uit veertig steeds moeilijker wordende opdrachten, die door de leerlingen in drie minuten dienen te worden beantwoord. Bij elke vraag wordt een leerling voor de taak gesteld om uit een verzameling van acht figuren aan te geven welke figuur niet in de betreffende verzameling thuishoort. Van de in het ISFAL-onderzoek afgenomen intelligentietests, is juist deze subtest gekozen, omdat dergelijke figurale tests het meest cultuurredelijk worden geacht (Jensen, 1980). Echter, het feit dat het om een speedtest

gaat, kan de cultuurredelijkheid ondermijnen. Immers, de tijdlimiet waaraan individuen gebonden zijn bij het maken van de test, kan nadelig zijn voor leerlingen die moeite hebben met het doseren van snelheid en nauwkeurigheid (Van de Vijver, 1993).

Bij de constructie van de schaal score die wordt geacht het vermogen tot logisch redeneren te representeren, zijn alleen die leerlingen betrokken die op minimaal een kwart van de vragen een antwoord hebben gegeven. Deze selectie is toegepast, omdat er bij minder beantwoorde vragen ons inziens geen verantwoorde schatting kan worden verkregen van iemands vermogen tot logisch redeneren. Bovendien zorgt het zeer consistente antwoordpatroon van leerlingen, die minder dan een kwart van de vragen hebben beantwoord, voor een slechts schijnbaar hoge betrouwbaarheid.

Om een zoveel mogelijk cultuurredelijke meting te bewerkstelligen is de test op item- en testniveau gecontroleerd op het bestaan van systematische (onbedoelde) verschillen tussen de scores van personen uit de diverse etnische groepen. Allereerst zijn, aan de hand van verschillende factoranalyses, die items geselecteerd op basis waarvan een voor elke etnische groep betrouwbare schaal kan worden geconstrueerd. Uiteindelijk zijn negen van de veertig items gebruikt voor de constructie van een intelligentieschaal. Cronbach's α van deze schaal bedraagt .75, .78, .79 en .74 voor respectievelijk de Nederlandse, Surinaamse, Turkse en Marokkaanse leerlingen. Deze waarden zijn voldoende hoog voor experimenteel onderzoek (Nunnally, 1978). Ook de itemrestcorrelaties zijn voldoende hoog ($\geq .24$). Ten tweede is gecontroleerd of de moeilijkheidsgraden van de items vergelijkbaar zijn tussen de groepen. Verschillen in moeilijkheidsgraden betekenen dat items niet alleen de bedoelde vaardigheid meten, maar ook onbedoelde vaardigheden die verschillen tussen de populaties (Drenth & Sijtsma, 1990; Osterlind, 1983). Een eenvoudige manier om vast te stellen of er sprake is van item-equivalentie is om via rangcorrelatiecoëfficiënten na te gaan in hoeverre de volgorden van moeilijkheidsgraad van de items variëren tussen de groepen (Van der Flier, 1980). Van der Flier (1980) noemt voor deze rangcorrelatiecoëfficiënt een kritische waarde van .80. Voor de intelligentieschaal variëren de rangcorrelatiecoëfficiënten van .97 tot 1.00. Er treden dus geen verschuivingen op in moeilijkheidsvolgorden. De conclusie luidt dat er voor de onderscheiden etnische groepen een intelligentieschaal is geconstrueerd die zowel betrouwbaar als cultuurredelijk is.

Schoolprestaties

De schoolprestaties van een leerling zijn vastgesteld aan de hand van een rekentoets. Het betreft een aangepaste versie van een door het Centraal Instituut voor Toetsontwikkeling (CITO) ontwikkelde instaptoets voor de middenschool. De bij de leerlingen afgenomen toets wordt geacht te voldoen aan de eis van een breed bereik, omdat de prestaties van zowel LBO-leerlingen als HAVO-leerlingen binnen de range van de toets behoren te vallen. De rekentoets bestond uit twintig meerkeuze-vragen die betrekking hadden op bewerkingen in formule- en contextvorm, die de leerlingen uit het hoofd in twaalf minuten dienden te beantwoorden.

Evenals bij de intelligentietest is als eis gesteld dat minimaal een kwart van de vragen beantwoord diende te zijn. Voor de constructie van een rekenschaal die voor alle etnische groepen voldoet, is dezelfde procedure gevolgd als bij de intelligentietest. Aan de hand van factoranalyses zijn twaalf items geselecteerd. Cronbach's α voor de Nederlandse, Surinaamse, Turkse en Marokkaanse leerlingen bedraagt .69, .80, .72 en .73. Ook hier zijn de itemrestcorrelaties bevredigend ($\geq .28$). De rangcorrelatiecoëfficiënten van de itemmoeilijkheidsgraden variëren van .78 tot .88. Er treden ook hier geen grote verschuivingen op in moeilijkheidsvolgorden. Geconcludeerd wordt dat er voor de onderscheiden etnische groepen een voldoende betrouwbare en cultuurredelijke rekenschaal is verkregen.

Bij de leerlingen is tevens een leestoets afgenomen. Ook hier betrof het een aangepaste versie van een door het CITO ontwikkelde instaptoets voor de middenschool. Het ging om een taak stillezen met beantwoording van tien vragen over een tekst, waarvoor de leerlingen tien minuten de tijd kregen. Het bleek echter niet mogelijk om op basis van deze toetsresultaten een voldoende betrouwbare en cultuurredelijke schaal te construeren die als indicator voor de leesprestaties van leerlingen kon fungeren. Dit betekent dat alleen de rekenprestaties van leerlingen

als indicator van schoolprestaties fungeren. Hoewel dit enigszins beperkt is, blijkt uit onderzoek dat rekenprestaties hoog correleren met schoolprestaties in het algemeen (Bleichrodt, Resing, Drenth & Zaal, 1987).

RESULTATEN

Voordat tot het testen van de onderzoeksmodellen is overgegaan, is de Nederlandse herkomstgroep teruggebracht tot een omvang die niet al te veel afwijkt van die van de overige etnische herkomstgroepen. Er bleek namelijk sprake te zijn van een grote oververtegenwoordiging van Nederlandse leerlingen ten opzichte van de Surinaamse, Turkse en Marokkaanse leerlingen. Gezien het feit dat gebruik is gemaakt van (co)variantie-analyses is deze reductie onvermijdelijk. De resultaten van variantie-analyses worden namelijk beïnvloed door een ongelijke verdeling van eenheden op de onafhankelijke variabele (Tacq, 1991). Ook HOMALS-analyses zijn gevoelig voor extreme ongelijke verdelingen over categorieën. De constructie van de variabele sociaal-economische herkomst heeft dan ook na de reductie van het aantal Nederlandse leerlingen plaatsgevonden. Er is een aselechte steekproef van 9% getrokken uit de groep Nederlandse leerlingen. De steekproef blijkt niet significant te verschillen van de totale Nederlandse groep op de relevante variabelen. In de analyses zijn vervolgens alleen die personen betrokken, waarvan alle gegevens beschikbaar waren. In tabel 1 wordt een overzicht gegeven van de scores van deze leerlingen op de in de analyse betrokken variabelen.

Uit tabel 1 blijkt dat de Nederlandse leerlingen het hoogst scoren op de gekwantificeerde variabele sociaal-economische herkomst. Zij worden gevolgd door respectievelijk de Surinaamse en Turkse leerlingen. De Marokkaanse leerlingen scoren gemiddeld het laagst op deze variabele. Met behulp van Scheffé-tests is onderzocht welke groepen van elkaar verschillen op een significantieniveau van 5%. Slechts het verschil tussen de Turkse en de Marokkaanse leerlingen en tussen de Nederlandse en Surinaamse leerlingen is niet significant. Wat betreft intelligentie scoren de Nederlandse leerlingen significant hoger dan de Surinaamse, Turkse en Marokkaanse leerlingen. De verschillen tussen de Surinaamse, Turkse en Marokkaanse leerlingen zijn niet significant. Ten aanzien van de rekenprestaties is de volgorde van gemiddelde scores van de onderscheiden etnische groepen hetzelfde als bij sociaal-economische herkomst. De Nederlandse leerlingen scoren significant hoger dan de Turkse en Marokkaanse leerlingen. De Surinaamse leerlingen scoren daarnaast significant hoger dan de Marokkaanse leerlingen. De overige verschillen zijn niet significant.

Aan de hand van de modelvergelijkingsprocedure (Tacq, 1991) is getoetst of een variabele uit het model een significante bijdrage levert aan de hoeveelheid verklaarde variantie in de afhankelijke variabele. Bij de modelvergelijkingsprocedure wordt getoetst of de door het volledige model (met de betreffende variabele) verklaarde variantie (R^2_{volledig}) significant groter is

Tabel 1. Gemiddelden (M) en standaarddeviaties (SD) naar etnische herkomst.

etnische herkomst	aantal		sociaal economische herkomst		intelligentie		rekenen	
	N	%	M	SD	M	SD	M	SD
Nederlands	112	35.1	.80	.43	4.55	2.50	8.04	2.79
Surinaams	54	16.9	.60	.63	2.98	2.24	6.72	3.55
Turks	86	27.0	-.69	.69	3.36	2.43	6.02	2.85
Marokkaans	67	21.0	-.90	.86	2.75	1.93	5.04	2.84
Totaal	319	100.0	.01	1.00	3.59	2.43	6.65	3.16

dan de door het beperkte model (zonder de betreffende variabele) verklaarde variantie (R^2_{beperkt}). Deze toetsing vindt plaats aan de hand van de toetsingsgrootte F die χ^2 -verdeeld is met als eerste vrijheidsgraad (v_1) het verschil tussen het aantal vrijheidsgraden van het volledige en het beperkte model ($d_{\text{beperkt}} - d_{\text{volledig}}$) en als tweede vrijheidsgraad (v_2) het aantal vrijheidsgraden van het volledige model (d_{volledig}). De toetsingsgrootte F kan worden berekend via de volgende formule.

$$F = \frac{(R^2_{\text{volledig}} - R^2_{\text{beperkt}}) / (d_{\text{beperkt}} - d_{\text{volledig}})}{(1 - R^2_{\text{volledig}}) / d_{\text{volledig}}}$$

Als de gevonden F -waarde significant is, is de verklaarde variantie van het volledige model significant groter dan die van het beperkte model. Er is dan sprake van een significant uniek effect van de betreffende variabele en deze is dus terecht in het model opgenomen. De proportie verklaarde variantie van de te vergelijken (co-)variantiemodellen is vastgesteld via de dummy-regressiebenadering (Green, 1987; Tacq, 1991).

Bij toetsing van model I wordt nagegaan of het (volledige) model met etnische herkomst significant meer variantie verklaart in rekenprestaties dan het (beperkte) model zonder verklaarende variabelen. Het blijkt dat de verklaarde variantie van het volledige model significant groter is dan die van het beperkte model ($F=15.03$, $v_1=3$, $v_2=315$, $p<.005$). Er is sprake van een significant effect van de etnische herkomst van leerlingen op hun rekenprestaties.

In model II is sprake van twee onafhankelijke variabelen. Voorwaarde bij het uitvoeren van (co-)variantieanalyses via de dummy-regressie benadering is dat er niet al te hoge samenhangen bestaan tussen de onafhankelijke variabelen (multicollineariteit). Indien dat wel het geval is, leveren de gestandaardiseerde regressiecoëfficiënten problemen op in inferentiële zin: de resultaten zijn niet generaliseerbaar van steekproef naar populatie (Tacq, 1991). In tabel 2 is de onderlinge samenhang tussen de onderzoeksvariabelen weergegeven.

De samenhang tussen de verschillende variabelen is over het algemeen zwak (cf. Nijdam & Van Buuren, 1982), met uitzondering van de samenhang tussen etnische en sociaal-economische herkomst (.77). Deze sterke samenhang is niet verwonderlijk, omdat de (groot-)ouders van Turkse en Marokkaanse kinderen over het algemeen geworven zijn voor het verrichten van ongeschoolde arbeid (Entzinger, 1984). Omdat bij de modelvergelijkingsprocedure alleen getoetst wordt op een verschil in verklaarde variantie, en geen gebruik wordt gemaakt van gestandaardiseerde regressiecoëfficiënten in inferentiële zin, levert eventuele multicollineariteit bij deze procedure geen probleem op. Als de verklaarde variantie van het model significant groter wordt na toevoeging van een onafhankelijke variabele, is deze stijging te wijten aan het unieke effect van de toegevoegde variabele. Immers, de gemeenschappelijke variantie die de toegevoegde variabele heeft met de overige in het model opgenomen onafhankelijke variabelen

Tabel 2. Samenhang tussen de variabelen¹.

	etnische herkomst ²	sociaal-economische herkomst ³	intelligentie ³
sociaal-economische herkomst	.77		
intelligentie	.31	.17	
rekenprestatie	.37	.30	.47

¹ samenhang significant op $p<.01$ niveau

² eta

³ pearson correlatiecoëfficiënt

is vanwege die gemeenschappelijkheid al tot uitdrukking gekomen in de verklaarde variantie van het model zonder de betreffende variabele.

Gezien de aanwezigheid van twee onafhankelijke variabelen (te weten etnische herkomst en sociaal-economische herkomst) in model II, is het model allereerst getoetst op additiviteit. Modelvergelijking laat zien dat er geen sprake is van een significant interactie-effect ($F=-.08$, $v_1=3$, $v_2=311$, $p>.05$). Vervolgens is getoetst of het model met sociaal-economische herkomst meer variantie in de rekenprestaties verklaart dan het model zonder sociaal-economische herkomst. Modelvergelijking toont aan dat dat niet het geval is ($F=.33$, $v_1=1$, $v_2=314$, $p>.05$). Er is dus geen sprake van een significant effect van de sociaal-economische herkomst van leerlingen op hun rekenprestaties, gecontroleerd voor etnische herkomst¹. Daarnaast is onderzocht of het model met etnische herkomst meer variantie in de rekenprestaties verklaart dan het model zonder etnische herkomst. Modelvergelijking toont aan dat dat inderdaad het geval is ($F=4.70$, $v_1=3$, $v_2=314$, $p<.005$): ook na controle voor sociaal-economische herkomst blijft er sprake van een significant effect van de etnische herkomst van leerlingen op hun rekenprestaties.

Ook model III is vanwege de aanwezigheid van meerdere onafhankelijke variabelen (te weten etnische herkomst, sociaal-economische herkomst en intelligentie) getoetst op additiviteit. Evenals bij model II blijkt er geen sprake te zijn van interactie-effecten ($F=-.18$, $v_1=7$, $v_2=306$, $p>.05$). Vervolgens is voor elk van de onafhankelijke variabelen getoetst of het volledige model meer variantie verklaart dan het model zonder de betreffende variabele. Voor intelligentie blijkt dat er sprake is van een significante bijdrage aan de verklaarde variantie van het model onder controle van de overige onafhankelijke variabelen ($F=64.34$, $v_1=1$, $v_2=313$, $p<.005$). Voor etnische herkomst apart blijkt dat er geen sprake is van een significante bijdrage aan de verklaarde variantie van het model ($F=1.19$, $v_1=3$, $v_2=313$, $p>.05$). Met andere woorden, na controle voor sociaal-economische herkomst en intelligentie is er geen significant effect meer van de etnische herkomst van leerlingen op hun rekenprestaties. Dat geldt echter ook voor sociaal-economische herkomst: na controle voor etnische herkomst en intelligentie is er geen significante bijdrage van sociaal-economische herkomst ($F=.71$, $v_1=1$, $v_2=313$, $p>.05$). Dat betekent echter niet, dat er geen sprake is van een effect van het milieu van herkomst van leerlingen op hun rekenprestaties. Het buiten het model laten van zowel etnische herkomst als sociaal-economische herkomst leidt namelijk wel tot een significante daling van de verklaarde variantie van het model ($F=6.08$, $v_1=4$, $v_2=313$, $p<.005$). De gemeenschappelijke variantie van etnische herkomst en sociaal-economische herkomst draagt blijkbaar bij tot het verklaren van verschillen in rekenprestaties van leerlingen².

De resultaten van bovenstaande analyses zijn nogmaals kort weergegeven in tabel 3.

Tabel 3. Getoetste modellen.

	effect	significant
Model I	etnische herkomst	ja
Model II	additiviteit	ja
	etnische herkomst	ja
	sociaal-economische herkomst	nee
Model III	additiviteit	ja
	etnische herkomst	nee
	sociaal-economische herkomst	nee
	etnische en sociaal-economische herkomst	ja
	intelligentie	ja

CONCLUSIES

In onderzoek naar de schoolprestaties van allochtone leerlingen wordt veelal uitgegaan van de veronderstelling dat de etnische herkomst van leerlingen niet van belang is, omdat etnisch-culturele verschillen in het milieu van herkomst zouden samenvallen met sociaal-economische verschillen. Het milieu van herkomst van leerlingen wordt daarom veelal geoperationaliseerd aan de hand van sociaal-economische indicatoren. De vraag in dit onderzoek was of deze omissie van etnisch-culturele indicatoren gerechtvaardigd is. Omdat daarnaast geconstateerd werd dat in onderzoek naar de oorzaken van verschillen in schoolprestaties van leerlingen afkomstig uit verschillende etnische groepen intelligentie zelden in het onderzoeksdesign is opgenomen, terwijl verondersteld wordt dat deze variabele een potentiële voorspeller is, is tevens onderzocht wat er met het effect van de etnische herkomst van leerlingen gebeurt als intelligentie aan het model wordt toegevoegd.

Uit de toetsing van model I blijkt dat de etnische herkomst van leerlingen van invloed is op hun schoolprestaties: er bestaan significante verschillen in de gemiddelde rekenprestaties van leerlingen afkomstig uit verschillende etnische groepen. Uit toetsing van model II blijkt dat er ook na controle voor de sociaal-economische herkomst van leerlingen sprake is van een significant effect van hun etnische herkomst. Hieruit kan geconcludeerd worden dat naast sociaal-economische indicatoren ook etnisch-culturele indicatoren in het onderzoeksdesign dienen te worden opgenomen. Bij toetsing van model III blijkt echter dat er na controle voor de sociaal-economische herkomst en de intelligentie van leerlingen geen sprake is van een significant effect van hun etnische herkomst. Hieruit mag echter niet worden afgeleid dat etnisch-culturele indicatoren, naast sociaal-economische indicatoren, niet in het onderzoeksdesign hoeven te worden opgenomen. Omgekeerd geldt namelijk dat er geen effect is van sociaal-economische herkomst na controle voor etnische herkomst (en intelligentie). De effecten van beide herkomstvariabelen blijken inwisselbaar, gezien de constatering dat omissie van beide herkomstvariabelen leidt tot een significante vermindering van de verklaarde variantie. De vraag of de etnische herkomst van leerlingen van invloed is op hun schoolprestaties kan derhalve niet eenduidig beantwoord worden. De keuze om het onderwijsvoorrangsbeleid te baseren op de veronderstelling dat achterstanden van (allochtone) leerlingen hoofdzakelijk een sociaal-economische achtergrond hebben, wordt niet bevestigd door dit onderzoek.

Gezien het feit dat etnische herkomst en sociaal-economische herkomst in relatief sterke mate met elkaar gecontamineerd zijn, verdient het aanbeveling om in toekomstig onderzoek meer aandacht te besteden aan de mogelijk vertekenende invloed hiervan op de onderzoeksresultaten. Omdat de resultaten voorts laten zien dat intelligentie een belangrijke bijdrage levert aan het verklaren van verschillen in schoolprestaties tussen leerlingen, verdient het tevens aanbeveling om in toekomstig onderzoek intelligentie in het verklaringsmodel op te nemen.

NOTEN

1. Uit onderzoek van Van Langen & Jungbluth (1992) blijkt dat na statistische correctie voor de etnische herkomst van leerlingen, grote verschillen in bereikt onderwijsniveau naar sociaal-economische herkomst blijven bestaan. Een mogelijke verklaring voor dit verschil in onderzoeksresultaat is dat in het onderhavige onderzoek de afhankelijke variabele de schoolprestaties (i.c. rekenprestaties) van leerlingen betreffen en niet het door hun gevolgde schooltype. Hierbij dient opgemerkt te worden dat de eerste van interval-meetniveau is, terwijl het meetniveau van de laatste hooguit ordinaal is. Strikt genomen is het bereikte onderwijsniveau daarom ongeschikt om als afhankelijke variabele te dienen in een (co-)variantie-analyse (Tacq, 1991). Een andere verklaring kan zijn dat, door de controle op bias bij de constructie van de rekenschaal en de daarmee gepaard gaande verwijdering van onzuivere items, verschillen in rekenprestaties naar sociaal-economische herkomst zijn gereduceerd (Drenth & Sijsma, 1990).
2. Bij een klassieke covariantie-analyse met een simultane schatting van de effecten van de onafhankelijke

variabelen, waarbij het effect van een onafhankelijke variabele gecorrigeerd wordt voor het effect van de overige onafhankelijke variabelen, luidt de conclusie dat er geen sprake is van een effect op rekenprestaties van zowel de etnische herkomst als de sociaal-economische herkomst (respectievelijk $F=2.21$, $p>.05$ en $F=1.71$, $p>.05$). Alleen het effect van intelligentie is significant ($F=65.55$, $p<.001$). De conclusie dat het milieu van herkomst van leerlingen niet van invloed is op hun rekenprestaties, na correctie voor intelligentie, is echter onjuist (zie tabel 3). Immers, het effect van elke onafhankelijke variabele wordt wel gecorrigeerd voor de gemeenschappelijke variantie van etnische herkomst en sociaal-economische herkomst, maar het effect van de gemeenschappelijke variantie zelf komt niet in de analyseresultaten tot uiting. Het is juist deze gemeenschappelijke variantie die verantwoordelijk is voor het effect van het milieu van herkomst van leerlingen op hun rekenprestaties.

LITERATUUR

- Berg, G.M. van den (1986). *Homals voor beginners*. Leiden: Rijksuniversiteit Leiden.
- Drenth, P.J.D., & Sijtsma, K. (1990). *Testtheorie. Inleiding in de psychologische test en zijn toepassingen*. Houten: Bohn Stafleu Van Loghum.
- Bleichrodt, N., Resing, W.C.M., Drenth, P.J.D., & Zaal, J.N. (1987). *Intelligentiemeting bij kinderen. Empirische en methodologische verantwoording van de gereviseerde Amsterdamse Kinder Intelligentie Test*. Lisse: Swets & Zeitlinger.
- Driessen, G. (1993). Binnen en buiten Onderwijsvoorrangsgebieden. Sociaal-etnische achtergronden en onderwijspositie van leerlingen in het voortgezet onderwijs. *Pedagogische studiën*, 70, 4, 252-262.
- Dronkers, J. (1986). Onderwijs en sociale ongelijkheid. In J.A. van Kemenade, N.A.J. Lagerweij, J.M.G. Leune & J.M.M. Ritzzen (Eds.), *Onderwijs, bestel en beleid. Deel 2a. Onderwijs en samenleving*. Groningen: Wolters Noordhoff.
- Entzinger, H.B. (1984). *Het minderhedenbeleid; dilemma's voor de overheid in Nederland en zes andere immigratielanden in Europa*. Meppel: Boom.
- Erf, R.F. van der (1989). Beleidsrelevante migrantengroepen demografisch gezien. *Migrantenstudies*, 3, 11-24.
- Flier, H. van der (1980). *Vergelijkbaarheid van individuele testprestaties*. Lisse: Swets & Zeitlinger.
- Glas, C.A.W., & Ouborg, M.J. (1993). Vraagonzuiverheid. In T.J.H.M. Eggen & P.F. Sanders (Eds.), *Psychometrie in de praktijk*. Arnhem: Cito Instituut voor Toetsontwikkeling.
- Green, P.E. (1978). *Analyzing multivariate data*. Hinsdale, IL: The Dryden Press.
- Grift, W. van de (1990). Educational leadership and academic achievement in elementary education. *School Effectiveness and School Improvement*, 1, 26-40.
- Grift, W. van de, & Akkermans, W. (1991). Onderwijskundig leiderschap en leerlingprestaties in het basisonderwijs. *Tijdschrift voor Onderwijsresearch*, 16, 4, 244-257.
- Grip, A. de (1988). *Scholing en werk. Een analyse van de scholingproblemen op de arbeidsmarkt en het arbeidsmarktbeleid*. Epe: Stichting Maatschappij en Onderneming.
- Guthke, J. (1993). Currents trends in theories and assessment of intelligence. In J.H.M. Hamers, K. Sijtsma & A.J.J.M. Ruijsseenaars (Eds.), *Learning potential assessment. Theoretical, methodological and practical issues*. Lisse: Swets & Zeitlinger.
- Heek, F. van (Ed.), (1968). *Het verborgen talent: milieu, schoolkeuze en schoolgeschiktheid*. Meppel: Boom.
- Hessels, M.G.P. (1993). *Leertest voor Etnische Minderheden: theoretisch en empirische verantwoording*. Proefschrift Universiteit Utrecht. Rotterdam: RISBO.
- Hof, L. van 't, & Dronkers, J. (1992). *Onderwijsachterstanden van allochtonen. Klasse, gezin of cultuur*. Paper voor de sessie 'Socialisatie en levensloop' van de Sociaalwetenschappelijke studiedagen 1992. Amsterdam: SCO.
- Hofman, R.H. (1992). *Effectief schoolbestuur. Een studie naar de bijdrage van schoolbesturen aan de effectiviteit van basisscholen*. Monografieën onderwijsonderzoek, 13. Proefschrift Universiteit Groningen. Groningen: RION.
- Hofman, W.H.A. (1993). *Effectief onderwijs aan allochtone leerlingen. Een empirische studie naar de invloed van school- en klasfactoren op de loopbaan van allochtone en autochtone leerlingen in het voortgezet onderwijs*. Proefschrift Erasmus Universiteit Rotterdam. Delft: Eburon.
- Hofman, W.H.A., Latuheru, E.J., & Velden, T.M.H. van der (1989). *ISFAL interimrapportage deel I*. Rotterdam: RISBO.

- Hooft, J. van (1987). *Het onderwijsniveau van migrantenkinderen*. Amsterdam: Gemeente Amsterdam.
- Horn, W. (1969). *Prüfsystem für Schul- und Bildungsberatung P-S-B. Handanweisung*. Göttingen: Verlag für Psychologie.
- Jensen, A.R. (1961). Learning abilities in Mexican-American and Anglo-American children. *California Journal of Educational Research*, 7, 4, 147-159.
- Jensen, A.R. (1980). *Bias in mental testing*. New York: Free Press.
- Jong, M.J. de (1985). Het onderwijsniveau van allochtone leerlingen in Rotterdam. In M.J. de Jong (Ed.), *Allochtone kinderen op Nederlandse scholen. Prestaties, problemen en houdingen*. Lisse: Swets & Zeitlinger.
- Jong, M.J. de (1987). *Herkomst, kennis en kansen. Allochtone en autochtone leerlingen tijdens de overgang van basis- naar voortgezet onderwijs*. Proefschrift Erasmus Universiteit Rotterdam. Lisse: Swets & Zeitlinger.
- Jong, S.C.N. de (1991). Prestatie, accumulatie zorg en geweld in het wereldkader. In L.J.M. d' Anjou & A.P.J. Bernts (Eds.), *Presteren en verdelen in Nederland. Over individuele belangen en maatschappelijke solidariteit*. Meppel: Boom.
- Kerkhoff, A. (1989). De geschiedenis herhaalt zich. Onderwijskansen van allochtone kinderen. *Migrantenstudies*, 2, 23-47.
- Kaufman, A.S., Kamphaus, R.W., & Kaufman, N.L. (1985). New directions in intelligence testing: the Kaufman Assessment Battery for Children (K-ABC). In B.B. Wolman (Ed.), *Handbook of intelligence. Theories, measurements and applications*. New York: Wiley & Sons.
- Kloprogge, J., m.m.v. Werf, G. van de, & Tesser, P. (1990). *Naar beter onderwijsvoorrangsbeleid*. Notitie over het onderwijsvoorrangsbeleid in 1989. 's-Gravenhage: SVO.
- Langen, A. van, & Jungbluth, P. (1990). *Onderwijskansen van migranten. De rol van sociaal-economische en culturele factoren*. Lisse: Swets & Zeitlinger.
- Leseman, P. (1989). *Structurele en pedagogische determinanten van schoolloopbanen. Berichten summatieve evaluatie project onderwijs en sociaal milieu 89/5*. Proefschrift Erasmus Universiteit Rotterdam. Rotterdam: Rotterdamse Schooladviesdienst (SAD).
- Nijdam, B. & Buuren, J. van (1982). *Statistiek voor de sociale wetenschappen*. Alphen a/d Rijn: Samson.
- Nunnally, J.C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- Osterlind, S.J. (1983). Test Item Bias. *Sage University Paper Series on Quantitative Applications in the Social Sciences, 07-030*. Newbury Park, CA: Sage.
- Penninx, R. (1988). *Minderheidsvorming en emancipatie. Balans van kennisverwerving ten aanzien van immigranten en woonwageneigenaren*. Alphen a/d Rijn: Samson.
- Resing, W.C.M., Bleichrodt, N., & Drenth, P.J.D. (1986). Het gebruik van de RAKIT bij allochtoon etnische groepen. *Nederlands Tijdschrift voor Psychologie*, 41, 179-188.
- Roelandt, Th., & Veenman, J. (1988). *Minderheden in Nederland. Positie in het onderwijs*. Rotterdam: EUR/ISEO.
- Roelandt, Th., Roijen, J.H.M., & Veenman, J. (1991). *Minderheden in Nederland. Statistisch vademecum 1991*. 's-Gravenhage: CBS/ISEO, SDU.
- Rohwer, W.D. (1971). Learning, Race, and School Success. *Review of Educational Research*, 41, 191-210.
- Röhr-Sendlmeier, U.M. (1987). Cross-cultural Intelligence Testing. In C. Schwarzer & B. Seipp (Eds.), *Trends in European educational research*. Braunschweiger Studien zur Erziehungs- und Sozialarbeitswissenschaft. Braunschweig: Rössner.
- Sarnacki, R.E. (1979). An examination of test-wiseness in the cognitive test domain. *Review of Educational Research*, 49, 252-279.
- Scheerens, J. (1989). *Wat maakt scholen effectief? Samenvatting en analyse van onderzoeksresultaten*. SVO Balans van onderwijsonderzoek. Den Haag: SVO.
- Sewell, T.E. (1979). Intelligence and learning tasks as predictors of scholastic achievement in black and white first-grade children. *Journal of School Psychology*, 17, 325-332.
- Soest, R. van, & B. Verdonk (1984). *Etnische identiteit, begrip en theorieën. Een literatuurstudie*. Instituut preventieve en sociale psychiatrie. Rotterdam: Erasmus Universiteit.
- Tacq, J. (1991). *Van probleem naar analyse. De keuze van een gepaste multivariate analysetechniek bij een sociaal-wetenschappelijke probleemstelling*. De Lier: Academisch Boeken Centrum.
- Tesser, P. (1986). *Sociale herkomst en schoolloopbanen in het voortgezet onderwijs*. Nijmegen: ITS.
- Tesser, P. (1989). Achterstand en cultureel verschil. Een dilemma in het onderwijsvoorrangsbeleid. *Jeugd en samenleving*, 1989, 2/3, 189-201.
- Vernon, P.E. (1979). *Intelligence, heredity and environment*. San Francisco: Freeman.
- Vijver, F.J.R. van de (1993). Learning potential assessment from a cross-cultural perspective. In J.H.M.

- Hamers, K. Sijtsma & A.J.J.M. Ruijsenaars (Eds.), *Learning potential assessment. Theoretical, methodological and practical issues*. Lisse: Swets & Zeitlinger.
- Werf, M.P.C. van der, & Weide, M.G. (1991). Effectief onderwijs aan allochtone leerlingen. *Tijdschrift voor Onderwijsresearch*, 16, 4, 231-243.
- Wit, W. de, Suhre, C., & Mulder, L. (1993). *De onderwijspositie van de OVB-doelgroepen na drie jaar voortgezet onderwijs*. Landelijke Evaluatie Onderwijsvoorrangsbeleid. Nijmegen: ITS.

Manuscript ontvangen 10-12-1993

Definitieve versie ontvangen 3-5-1994

Overadvisering en de schoolloopbanen van migrantenleerlingen en autochtone leerlingen in het voortgezet onderwijs*

M. Kocslag¹ en J. Dronkers²

¹ ROA, Rijksuniversiteit Limburg

² SCO-Kohnstamm Instituut, Universiteit van Amsterdam

ABSTRACT

Migrant pupils in the Netherlands get on the average higher recommendations for the most suited type of secondary education than Dutch pupils, given their scores on cognitive tests at the end of primary school. These higher recommendations might explain the less successful career in secondary education. This hypothesis is tested with a national sample of pupils who entered secondary education in 1989. This hypothesis is rejected. The most important factor to explain differences in educational attainment between Dutch and migrant pupils is the low parental educational and occupational position. This doesn't mean that the processes which effects the educational attainment of Dutch and migrant pupils are equal. It was found that these processes differ significantly. But these differences in processes do not explain the differences in educational attainment between Dutch and migrant pupils.

1. INLEIDING

De schoolloopbanen van migrantenleerlingen en autochtone leerlingen verschillen van elkaar. De prestaties van migrantenleerlingen op de basisschool verschillen van die van autochtonen, migrantenleerlingen komen vaker dan autochtonen in de lagere schooltypen van het voortgezet onderwijs terecht en de onderwijsloopbaan van de gemiddelde migrantenleerling in het voortgezet onderwijs verloopt minder gunstig dan die van de gemiddelde autochtone leerling: (Mulder & Tesser, 1992; De Jong, 1987; Van Langen & Jungbluth, 1990; Schouten, 1990; Van 't Hof & Dronkers, 1993.).

In de literatuur worden verschillende verklaringen genoemd voor de minder gunstige schoolloopbanen van migrantenleerlingen. Onderzoek (Tesser, Van der Werf, Mulder, 1990:156) wijst uit dat *de verblijfsduur* van migrantenleerlingen in Nederland, hun *beheersing van de Nederlandse taal* en hun *oriëntatie op Nederlandse waarden en normen* een verklaring voor onderwijsachterstanden kunnen zijn. Ook bij leerlingen die de basisschool niet in zijn geheel in Nederland hebben gevolgd, de zogenaamde *zij-instromers*, heeft dit tot slechtere schoolloopbanen geleid (Roelandt, Martens, Veenman, 1990: 120).

* Dit artikel is één van de producten van het project 'Secundaire analyses over de schoolloopbanen in het voortgezet onderwijs met de VOCL'89 data' van het SCO/Kohnstamm Instituut van de Universiteit van Amsterdam, uitgevoerd in opdracht van het Instituut voor Onderzoek van het Onderwijs (SVO) in Den Haag (project 92102). Het onderzoek is mede verricht door Mieke Kocslag, als onderdeel van de doctoraal scriptie voor de studie Sociologie (specialisatie cultuur en onderzoek) aan de Katholieke Universiteit Brabant te Tilburg, onder begeleiding van prof. dr. J. Dronkers (UvA). Correspondentie over dit artikel richten aan: J. Dronkers, SCO-Kohnstamm Instituut, Universiteit van Amsterdam, Grote Bickersstraat 72, 1013 KS Amsterdam.

Dit is het laatste artikel behorende bij het themanummer.

Deze invloeden veroorzaken dat migrantenleerlingen aan het einde van de basisschool gemiddeld lagere adviezen krijgen dan autochtone leerlingen (Tesser, Van der Werf, Mulder, 1990; Kohnstamm, 1993; Mulder & Tesser, 1992; Van 't Hof & Dronkers, 1993). "Zo krijgt ongeveer 60% van de Turkse en Marokkaanse leerlingen een LBO-advies, tweemaal zoveel als de Nederlandse kinderen, en vrijwel nooit een VWO-advies. Voor Surinaamse en Antilliaanse kinderen zijn de cijfers iets gunstiger, maar alleen de Molukse kinderen komen dicht in de buurt van de Nederlandse". (Kohnstamm, 1993). Ander onderzoek (De Jong, 1987; Tesser & Mulder, 1990; Kohnstamm, 1993; Van 't Hof & Dronkers, 1993) wijst uit dat migrantenleerlingen, bij dezelfde prestaties als autochtone leerlingen, aan het einde van de basisschool juist hogere adviezen krijgen dan de vergelijkbare autochtone leerlingen: migrantenleerlingen worden overgeadviseerd. "Globaal genomen krijgt een allochtone leerling een HAVO-advies bij prestatiescores waarbij een Nederlandse leerling een MAVO-advies krijgt. Analoge verhoudingen doen zich voor bij de MAVO- en LBO-adviezen" (Tesser & Mulder, 1990:37).

Overadvisering van migrantenleerlingen kan op verschillende manieren verklaard worden. In de eerste plaats is het mogelijk dat op de basisschool lagere eisen aan de prestaties van migrantenleerlingen worden gesteld dan aan autochtone leerlingen. Deze lagere eisen zijn een gevolg van het rekening houden door de leerkrachten met de negatieve effecten van het migratieverleden van deze leerlingen waardoor zij de feitelijke schoolprestaties van migrantenleerlingen positiever beoordelen. Ook is het mogelijk dat leerkrachten aan migrantenleerlingen lagere eisen stellen dan aan autochtone leerlingen omdat zij bang zijn te discrimineren. In beide gevallen krijgen migrantenleerlingen bij de advisering eerder het voordeel van de twijfel dan autochtone leerlingen en dus een relatief hoger advies.

Overadvisering van migrantenleerlingen kan in de tweede plaats verband houden met de sociale samenstelling van de basisschool die door migrantenleerlingen wordt bezocht. "Allochtone leerlingen zitten meestal op scholen met een hoog percentage leerlingen uit achterstandssituaties. Het gemiddelde prestatieniveau op die scholen is daardoor laag. Een goede leerling op zo'n school haalt ongeveer dezelfde prestaties als een middelmatige leerling op een school met een hoog gemiddeld prestatieniveau. Toch zullen beide scholen bij de advisering een rangorde aanbrengen waarbij de betere leerlingen een hoger advies krijgen dan de zwakkere leerlingen. Het gevolg daarvan is dat een goede leerling op de scholen met een laag gemiddeld prestatieniveau een hoger advies krijgt dan de middelmatige leerlingen op scholen waar het gemiddelde prestatieniveau hoger is, ook al zijn de taal en rekenprestaties van beide leerlingen gelijk." (Mulder & Tesser, 1992:59).

Overadvisering van migrantenleerlingen, hoe goed bedoeld ook, zou wel eens gedeeltelijk kunnen verklaren waarom deze leerlingen in het voortgezet onderwijs achterblijven bij autochtone leerlingen. Want het is mogelijk dat migrantenleerlingen door overadvisering vaker dan autochtone leerlingen in schooltypen terecht komen, waar van hen hogere schoolprestaties worden gevraagd dan zij feitelijk kunnen leveren. Op die hogere schooltypen zou bovendien minder dan op de basisschool rekening gehouden worden met de negatieve effecten van het migratieverleden. Migrantenleerlingen hebben daardoor een grotere kans om in het voortgezet onderwijs uit- of af te stromen dan autochtone leerlingen. Overadvisering kan echter bij migrantenleerlingen ook tot gunstige schoolloopbanen leiden, omdat de leerlingen in het voortgezet onderwijs uitgedaagd kunnen worden tot het leveren van prestaties die hun schoolkeuze ondersteunen.

Naast overadvisering zijn er andere verklaringen voor de ongunstige schoolloopbanen van migrantenleerlingen. De sociale herkomst van de leerlingen en de culturele afstand tot de Nederlandse samenleving kunnen ook de slechtere schoolloopbanen verklaren. Onderzoek (De Jong, 1987; Van Langen & Jungbluth, 1990; Roelandt, Martens, Veenman, 1990; Mulder & Tesser, 1992) wijst namelijk enerzijds uit dat, gegeven andere invloeden op de schoolloopbaan, leerlingen uit hogere milieus gunstiger schoolloopbanen hebben dan leerlingen uit lagere milieus. Anderzijds blijkt dat leerlingen wier afstand tot de dominante cultuur in het onderwijs klein is, betere schoolloopbanen hebben dan andere leerlingen (Sociaal en Cultureel Rapport 1992:200).

In dit artikel wordt geanalyseerd of overadvisering, naast sociale herkomst en culturele afstand tot de waarden en normen van de Nederlandse samenleving, een oorzaak is voor de minder gunstige schoolloopbanen van migrantenleerlingen vergeleken met autochtone leerlingen. Alvorens deze analyse te presenteren, geven we een overzicht van de categorieën migrantenleerlingen die we hebben onderscheiden en een beschrijving van hun schoolloopbanen.

2. DATA EN METHODE

2.1. Data

Wij maken gebruik van individuele gegevens van leerlingen uit het onderzoek 'Voortgezet Onderwijs Cohort Leerlingen 1989', (VOCL'89). Deze gegevens zijn verzameld door het CBS. Uit de scholen van het voortgezet onderwijs is door het CBS een steekproef getrokken van ongeveer 400 scholen. Vervolgens zijn uit deze steekproef aselect klassen met leerlingen getrokken. Het gaat hier om leerlingen die in 1989 in het voortgezet onderwijs zijn ingestroomd. Deze leerlingen zijn gedurende drie jaar gevolgd in het voortgezet onderwijs (tot en met 1991). Dit cohort leerlingen bestaat uit 19524 leerlingen. Voor dit onderzoek zijn leerlingen die voor 1989 al in een eerste klas van het voortgezet onderwijs zaten, uit het analyse bestand verwijderd omdat deze leerlingen de uitkomsten kunnen vertekenen. Ook zijn leerlingen uit éénoudergezinnen buiten de analyse gelaten. Leerlingen uit éénoudergezinnen hebben een slechtere schoolloopbaan dan leerlingen uit tweeoudergezinnen (Dronkers, 1992). In sommige categorieën migrantenleerlingen komen éénoudergezinnen vaker voor dan in andere categorieën. Deze categorieën leerlingen hebben daarom gemiddeld grotere onderwijsachterstanden. Omdat deze achterstanden een gevolg kunnen zijn van het opgroeien in een éénoudergezin en niet direct van de eigen cultuur, zijn leerlingen die afkomstig zijn uit éénoudergezinnen uit de analyses gelaten (Zie ook Van 't Hof & Dronkers, 1993:6). Alle overblijvende migrantenleerlingen zijn in het analyse bestand opgenomen. Uit de autochtone leerlingen is, met het oog op de hanteerbaarheid, een a-selecte steekproef getrokken van ruim 3700 leerlingen. Uiteindelijk zijn 5057 leerlingen in de analyse betrokken (zie tabel 1).

2.2. Verzameling van de gegevens:

Van de leerlingen zijn door middel van een aan de ouders/verzorgers gerichte vragenlijst achtergrondgegevens verzameld. Bij elke leerling is een door het CITO ontworpen schoolvorderings-toets afgenomen, om een indruk te krijgen van het prestatieniveau van de leerling aan het einde van de basisschool. Deze toets is in 1989, toen de leerlingen in de eerste klas van het voortgezet onderwijs zaten, afgenomen. De scholen van voortgezet onderwijs zijn door het CBS benaderd om informatie te geven over het schooltype, het leerjaar en het schooladvies dat door het hoofd van de basisschool is gegeven. In 1990 en 1991 zijn de scholen benaderd om gegevens te geven over het leerjaar en schooltype waarin een leerling op dat moment verbleef.

2.3. De variabelen

Van de volgende categorieën leerlingen wordt de schoolloopbaan onderzocht (zie ook bijlage 1):

- 1) Turkse leerlingen. Dit zijn leerlingen waarvan tenminste één van de ouders in Turkije is geboren (TUR).
- 2) Marokkaanse leerlingen. Dit zijn leerlingen waarvan tenminste één van de ouders in Marokko is geboren (MAR).
- 3) Surinaamse leerlingen. Dit zijn leerlingen waarvan tenminste één van de ouders in Suriname is geboren (SUR).
- 4) Antilliaanse leerlingen. Dit zijn leerlingen waarvan tenminste één van de ouders op de Nederlandse Antillen is geboren (ANT).
- 5) Molukse leerlingen. Dit zijn leerlingen waarvan tenminste één van de ouders op de Molukken is geboren (MOL).

- 6) Leerlingen waarvan tenminste één van de ouders de Griekse, Italiaanse, Portugese, Spaanse of Maltese nationaliteit heeft en niet in Nederland is geboren (ZUER).
- 7) Leerlingen waarvan tenminste één van de ouders de Chinese nationaliteit heeft en niet in Nederland is geboren (CHIN).
- 8) Leerlingen waarvan tenminste één van de ouders een nationaliteit van een Aziatisch land (met uitzondering van China en de Molukken) heeft en niet in Nederland is geboren (OVAZ).
- 9) Leerlingen met tenminste één van de ouders die een nationaliteit van een land uit Noord-, West- of Oost-Europa heeft en niet in Nederland is geboren. Ook leerlingen waarvan tenminste één van de ouders de Amerikaanse, Canadese of Australische nationaliteit heeft en niet in Nederland is geboren, zijn ingedeeld in deze categorie (OVIW).
- 10) Leerlingen waarvan tenminste één van de ouders een nationaliteit van een land uit Afrika of Zuid-Amerika heeft en niet in Nederland is geboren (OV3W).
- 11) Leerlingen die de Nederlandse nationaliteit hebben en ook in Nederland is geboren. Ook de ouders van deze leerlingen hebben de Nederlandse nationaliteit en zijn in Nederland geboren (AUT).

Tabel 1. Indeling categorieën leerlingen (in %).

	TOTAAL AANTAL	%
TUR	382	7.6
MAR	346	6.8
SUR	156	3.1
ANT	68	1.3
MOL	22	0.4
ZUER	80	1.6
CHIN	28	0.6
OVAZ	34	0.7
OVIW	218	4.3
OV3W	18	0.4
AUT	3705	73.3
TOTAAL	5057	100.0

De prestaties van de leerlingen zijn gemeten met behulp van een door het CITO ontworpen schoolvorderingstoets. Deze toets bestaat uit drie delen: een taaltoets, een rekentoets en een informatieverwerkingstoets.

Het schooladvies heeft 8 categorieën: individueel onderwijs, LBO, LBO-MAVO, MAVO, MAVO-HAVO, HAVO, HAVO-VWO, VWO. Wanneer er een advies voor drie schooltypen is gegeven, is het gewaardeerd naar het middelste schooltype, bijvoorbeeld een MAVO-HAVO-VWO-advies is gewaardeerd als een HAVO-advies. De schaling van deze 8 categorieën is gebeurd met behulp van de schaal van Van der Eeden e.a. (1993).

Om de hoogte van het gekozen type voortgezet onderwijs en het succes in het voortgezet onderwijs te bepalen is de onderwijspositie in 1989 tot en met 1991 gewaardeerd met een score. Voor de brugklassen is een gemiddelde waardering gegeven van de schooltypen waarop in die brugklas wordt voorbereid (b.v. brugklas MAVO-HAVO-VWO krijgt eenzelfde waardering als de categoriale HAVO). Voor deze procedure is gekozen, omdat zo het effect van brede brugklassen het best verdisconteerd wordt. Voor elk hoger leerjaar in een hogere klas op eenzelfde schooltype wordt de onderwijspositie 10 hoger. Leerlingen die zijn gestopt met het volgen van voortgezet onderwijs of het cohort hebben verlaten, hebben de onderwijspositie nul gekregen. De schaling van deze onderwijsposities is gebeurd met behulp van de schaal van Van der Eeden e.a. (1993).

De sociale herkomst is gemeten aan de hand van het opleidings- en beroepsniveau van beide

ouders. Het opleidingsniveau is gecodeerd van een niet voltooide lagere school (score 1) tot en met een voltooide opleiding in het hoger onderwijs (score 8). Het beroepsniveau van de vader loopt van arbeider (score 1) tot en met hogere employé (score 5).

Overadvisering is door ons gedefinieerd als het verschijnsel dat migrantenleerlingen, bij dezelfde prestaties als autochtone leerlingen, aan het einde van de basisschool hogere adviezen krijgen dan vergelijkbare autochtone leerlingen. De mate van over- (of onder)advisering is daarom door ons als volgt vastgesteld: wij hebben bij de autochtone leerlingen een gemiddeld advies berekend door de relatie tussen de prestaties op de schoolvorderingstoets en het advies aan het einde van de basisschool te schatten. Overadvisering wil zeggen: een hoger advies dan het op grond van de individuele prestaties te verwachten advies, gegeven de relatie tussen de prestaties op de schoolvorderingstoets en het advies bij autochtone leerlingen. Onderadvisering betekent een lager advies dan het op grond van eigen prestaties te verwachten advies, gegeven de relatie tussen de prestaties op de schoolvorderingstoets en het advies bij autochtone leerlingen¹. Wij hebben de relatie tussen prestaties en advies bij autochtone leerlingen gebruikt voor de vaststelling van de mate van over- of onderadvisering, omdat de over- of onderadvisering van migrantenleerlingen altijd vergeleken wordt met de advisering van autochtone leerlingen. De hier gevolgde procedure maakt de verschillen tussen de advisering van migrantenleerlingen en autochtone leerlingen maximaal zichtbaar.

3. DE RESULTATEN

3.1. Verschillen tussen de categorieën leerlingen

De gemiddelde prestaties van de categorieën leerlingen op deze onderdelen staan vermeld in Koeslag (1993) en in CBS (1991). Er zijn significante verschillen tussen de categorieën leerlingen wat betreft hun prestaties op de schoolvorderingstoets. Vergelijking van de gemiddelde prestaties per categorie leerlingen laat zien dat de Turkse en Marokkaanse leerlingen op alle onderdelen gemiddeld lager scoren dan andere categorieën leerlingen. Dit is, op rekenen na, ook het geval bij de categorie Chinese leerlingen. De categorie leerlingen uit Noord-, West-, en Oost-Europa scoren gemiddeld hoger op de onderdelen dan andere categorieën leerlingen. De categorie Molukse leerlingen wijkt niet veel af van de categorie autochtone leerlingen; deze categorie scoort het beste op rekenen.

De (gemiddeld) gekregen adviezen per categorie leerlingen staan ook vermeld in Koeslag (1993) en in CBS (1991). Ook deze blijken significant van elkaar te verschillen. Turkse en Marokkaanse leerlingen en leerlingen uit Zuid-Europa en de Derde Wereldlanden krijgen gemiddeld lagere adviezen dan andere categorieën leerlingen. Van de Turkse leerlingen krijgt 49% ten hoogste een LBO-advies; bij de Marokkaanse leerlingen is dat 45% en bij de autochtone leerlingen slechts 27%. De categorie Molukse leerlingen, Antilliaanse leerlingen en de categorieën leerlingen uit landen in Azië en West-, Oost- en Noord-Europa krijgen hogere adviezen dan de autochtone leerlingen: van de autochtone leerlingen krijgt 15% een HAVO-, VWO- of VWO-advies, van de Molukse leerlingen 27%, de Antilliaanse leerlingen 18% evenals de leerlingen uit landen in Azië. Van de leerlingen uit Noord-, West- of Oost-Europa krijgt 22% een HAVO-, VWO- of VWO-advies.

Uit Koeslag (1993) en CBS (1991) blijkt dat Turkse en Marokkaanse leerlingen en de categorie leerlingen uit landen in de Derde Wereld, vergeleken met andere categorieën leerlingen in het eerste leerjaar (1989) significant een lager schooltype kiezen: ongeveer 45-55% van deze leerlingen maakt een schoolkeuze van hooguit LBO. Bij de autochtone, Antilliaanse, Molukse en Surinaamse leerlingen is dit percentage respectievelijk 25, 15, 18, 21. Van de Antilliaanse, Molukse en Surinaamse categorie leerlingen kiest ongeveer 45% voor de HAVO of het VWO; bij de autochtonen is dit percentage 35,9 en bij de Turkse en Marokkaanse categorie leerlingen respectievelijk 23 en 21.

De onderwijspositie die de leerlingen in het derde leerjaar (1991) hebben staat vermeld in Koeslag (1993). Ook hierin verschillen de categorieën leerlingen significant van elkaar. Leer-

lingen die in 1989 voor het LBO of MAVO kiezen, blijken in de volgende twee jaar in de meeste gevallen gewoon door te stromen binnen hetzelfde schooltype. Opvallend is dat er in de categorie Chinese leerlingen relatief veel leerlingen uitvallen. Een andere categorie leerlingen die opvalt is de categorie leerlingen uit de Derde Wereld. Zij maken vaak (55,5%) een schoolkeuze in het schooltype IBO of LBO. Veel van deze leerlingen stromen op naar een hoger schooltype.

Categorieën leerlingen verschillen ook significant in sociaal milieu (Koeslag 1993; CBS, 1991). Vooral de Turkse, Marokkaanse, Surinaamse en Chinese leerlingen komen gemiddeld vaak uit een laag sociaal milieu, omdat het gemiddelde opleidingsniveau van de ouders en het gemiddelde beroepsniveau van de vaders van deze categorieën leerlingen lager is dan bij de andere leerlingen. Ouders van Antilliaanse, Molukse leerlingen en leerlingen uit de Eerste Wereld zijn gemiddeld hoger opgeleid dan de ouders van autochtone leerlingen. Alleen de vaders van de Molukse leerlingen hebben gemiddeld een hoger beroepsniveau dan de vaders van de autochtone leerlingen.

3.2. Over- en onderadvisering

In tabel 2 worden de percentages leerlingen die over- of ondergeadviseerd zijn, of conform de prestaties op de schoolvorderingstoets geadviseerd is, voor de verschillende categorieën leerlingen weergegeven. Bij deze beschrijving zijn alleen leerlingen, die een werkelijk advies hebben dat in belangrijke mate² boven het verwachte advies ligt, overgeadviseerd en zijn alleen de leerlingen die een werkelijk advies hebben dat in belangrijke mate onder het verwachte advies ligt, ondergeadviseerd. De andere leerlingen zijn conform geadviseerd.

Tabel 2. Advisering in % per categorie leerlingen.

ADVISERING	ONDER	CONFORM	OVER
Turkse leerlingen	6	83	11
Marokkaanse leerlingen	2	86	11
Surinaamse leerlingen	1	89	10
Antilliaanse leerlingen	9	77	15
Molukse leerlingen	9	82	9
Leerlingen uit Z.-Europa	5	88	8
Chinese leerlingen	-	86	14
Leerlingen uit Ov. Azië	-	94	6
Overige leerlingen uit 1e W.	7	84	9
Overige leerlingen uit 3e W.	11	89	-
Autochtone leerlingen	6	88	6

Antilliaanse (9%) en Molukse (9%) leerlingen, en leerlingen uit Noord-, West-, Oost-Europa (7%) en de Derde Wereldlanden (11%) worden vaker dan autochtone leerlingen (6%) ondergeadviseerd. Op de categorie leerlingen uit landen in Azië (behalve China (6%)) na, worden de leerlingen uit alle andere categorieën leerlingen vaker overgeadviseerd dan de autochtone (6%) leerlingen. Turkse (11%), Marokkaanse (11%), Antilliaanse (15%) en Chinese (14%) leerlingen worden vergeleken met de rest het vaakst overgeadviseerd. Bij alle migranten, behalve de Molukse categorie, blijkt overadvisering vaker voor te komen dan onderadvisering. Bij de Molukse categorie komt over- en onderadvisering even vaak voor. Autochtone leerlingen worden vaker ondergeadviseerd dan overgeadviseerd.

3.3. De onderwijsposities van leerlingen die een onder- of overadvies hebben gekregen

In de tabellen 3 en 4 zijn, per categorie leerlingen, de onderwijsposities gepresenteerd van leerlingen die over- of ondergeadviseerd zijn en van leerlingen die conform hun prestaties op de schoolvorderingstoets zijn geadviseerd.

Tabel 3. Het aantal afgesloten leerjaren per categorie leerlingen (in %).³

DE SCHOOLLOOPBAAN VAN	advies	0 JAAR	1 JAAR	2 JAAR	N
Turkse leerlingen	onder	-	-	100	22
	conform	-	20	74	318
	over	-	43	52	42
Marokkaanse leerlingen	onder	-	75	25	8
	conform	1	21	76	300
	over	-	32	63	38
Surinaamse leerlingen	onder	-	-	100	2
	conform	-	16	78	138
	over	-	25	75	16
Antilliaanse leerlingen	onder	-	-	100	6
	conform	-	19	73	52
	over	-	20	80	10
Molukse leerlingen	onder	-	-	100	2
	conform	-	22	78	18
	over	-	100	-	2
Leerlingen uit Z.-Europa	onder	-	-	100	4
	conform	-	20	74	70
	over	-	33	67	6
Chinese leerlingen	onder	-	-	-	-
	conform	-	-	67	24
	over	-	50	50	4
Leerlingen uit Ov.Azië	onder	-	-	-	-
	conform	-	13	75	32
	over	-	-	100	2
Overige leerlingen 1e W.	onder	-	13	75	16
	conform	-	17	80	182
	over	-	-	90	20
Overige leerlingen 3e W.	onder	-	-	100	2
	conform	-	-	100	16
	over	-	-	-	-
Autochtone leerlingen	onder	2	7	91	234
	conform	-	13	86	3253
	over	1	18	80	218

Uit tabel 3 blijkt dat leerlingen die een onderadvies krijgen nauwelijks doubleren in het voortgezet onderwijs dat zij tot en met 1991 gevolgd hebben. Alleen de categorie Marokkaanse leerlingen valt op. Slechts een vierde van deze categorie leerlingen doubleert niet: 75% doubleert één maal. In de andere categorieën ondergeadviseerde leerlingen komen alleen doubleurs voor bij de categorie autochtone leerlingen en bij de categorie migrantenleerlingen uit de rest van de Eerste Wereld. Ook wat betreft leerlingen die een advies conform aan de prestaties op de

Tabel 4. Uitstromen (zonder diploma het onderwijs verlaten), opstromen (in 1991 een hoger schooltype volgen dan in 1989 gekozen) en afstromen (in 1991 een lager schooltype volgen dan in 1989 gekozen) per categorie leerlingen (in %).

DE SCHOOLLOOPBAAN VAN	advisering	uitstr.	opstr.	afstr.	gelijkstr.	N
Turkse leerlingen	onder	-	-	-	100	22
	conform	6	13	8	73	318
	over	5	10	5	81	42
Marokkaanse leerlingen	onder	-	-	25	75	8
	conform	2	15	3	80	300
	over	5	32	5	58	38
Surinaamse leerlingen	onder	-	-	-	100	2
	conform	6	13	9	73	138
	over	-	13	-	88	16
Antilliaanse leerlingen	onder	-	-	-	100	6
	conform	8	19	8	65	52
	over	-	20	-	80	10
Molukse leerlingen	onder	-	-	-	100	2
	conform	-	22	-	78	18
	over	-	-	-	100	2
Leerlingen uit Z.-Europ :	onder	-	-	-	100	4
	conform	6	14	9	71	70
	over	-	67	-	33	6
Chinese leerlingen	onder	-	-	-	-	-
	conform	33	25	25	17	24
	over	-	-	-	100	4
Leerlingen uit Ov.Azië	onder	-	-	-	-	-
	conform	13	38	13	38	32
	over	-	100	-	-	2
Overige leerlingen 1e W.	onder	13	13	-	75	16
	conform	3	22	6	69	182
	over	10	70	10	10	20
Overige leerlingen 3e W.	onder	-	-	-	100	2
	conform	-	38	-	63	16
	over	-	-	-	-	-
Autochtone leerlingen	onder	2	9	3	86	234
	conform	1	19	2	77	3253
	over	2	23	5	70	218

schoolvorderingstoets hebben gekregen, zijn er geen grote verschillen tussen de categorieën leerlingen te constateren. In de meeste categorieën doubleert één-kwart van de leerlingen en dat meestal eenmaal. Alleen de categorie Chinese leerlingen doubleert meer: 33%. Ook is in deze categorie conform geadviseerde Chinese leerlingen de uitstroom hoog (33%). Er zijn verschillen tussen de categorieën overgeadviseerde leerlingen vast te stellen. Van de Turkse, Marok-

kaanse, Molukse, Chinese en Zuideuropese leerlingen die zijn overgeadviseerd, doubleren vaker leerlingen dan van de andere categorieën overgeadviseerde leerlingen. Van de categorie overgeadviseerde Turkse en Chinese leerlingen doubleert de helft van de leerlingen. Marokkaanse en Zuideuropese leerlingen doubleren in ruim 30% van de gevallen wanneer ze overgeadviseerd zijn. Bij de allochtone overgeadviseerde leerlingen ligt dit percentage doubleurs slechts op 18%.

Uit tabel 4 blijkt dat leerlingen die zijn overgeadviseerd, niet vaker uitstromen dan overige leerlingen. Alleen in de categorie Marokkaanse leerlingen stromen meer leerlingen uit die zijn overgeadviseerd. In de andere categorieën stroomt het hoogste percentage leerlingen uit die conform de schoolvorderingstoets zijn geadviseerd. In de meeste categorieën stromen leerlingen met een overadvies vaker op naar een hoger schooltype dan leerlingen die geen overadvies hebben gekregen. Dit is het geval bij de categorie Marokkaanse, Antilliaanse, Zuideuropese, autochtone leerlingen en leerlingen uit overig Azië en de rest van de Eerste Wereld.

Ondergeadviseerde leerlingen die uit- of afstromen, zijn alleen in de autochtone categorie en de categorieën leerlingen uit Marokko en de rest van de Eerste Wereld te vinden. Zij stromen minder vaak op vergeleken met leerlingen die conform of hoger dan de prestaties op de schoolvorderingstoets zijn geadviseerd. Bij de Marokkaanse leerlingen is dit percentage afstromers erg groot: 25%. In de categorie Turkse leerlingen stromen leerlingen die conform de prestaties op de schoolvorderingstoets zijn geadviseerd vaker af dan leerlingen die overgeadviseerd zijn. Alleen de overgeadviseerde leerlingen uit de Turkse, Marokkaanse, autochtone categorie leerlingen en de categorie leerlingen uit de rest van de Eerste Wereld stromen af. Leerlingen uit de rest van de Eerste Wereld stromen het vaakst (10%) af wanneer ze overgeadviseerd zijn; de autochtone leerlingen het minst vaak (5%). In de categorieën Turkse, Surinaamse, Antilliaanse, Molukse, Chinese leerlingen en in de categorie leerlingen uit Derde Wereldlanden, stromen leerlingen die conform de prestaties op de schoolvorderingstoets zijn geadviseerd, minder vaak door in het aanvangstype dan leerlingen die anders zijn geadviseerd. In alle categorieën stromen leerlingen die ondergeadviseerd zijn het vaakst door in het aanvangsschooltype.

4. INVLOEDEN OP DE SCHOOLLOOPBAAN

3/10/1990
Met behulp van een lineair structureel model onderzoeken we de relaties tussen de verschillende schoolloopbanen. Hiervoor hebben we LISREL gebruikt. Het doel van deze analyse is systematisch na te gaan in hoeverre de verschillen in de schoolloopbanen van migrantenleerlingen en autochtone leerlingen zijn toe te schrijven aan verschillen in ouderlijk milieu, schoolprestaties of overadvisering.

We beginnen de analyse met het basismodel uit figuur 1. In dit model veronderstellen we alleen rechtstreekse effecten: van ouderlijk milieu op de prestaties op de schoolvorderingstoets, van de prestaties op de schoolvorderingstoets op de mate van onder- en overadvisering, van de mate van onder- of overadvisering op de schoolkeuze, van de schoolkeuze op de onderwijspositie in 1990, van de onderwijspositie in 1990 op die van 1991. Het ouderlijk milieu is een latente variabele, geïndiceerd door de beroeps- en opleidingsniveaus van beide ouders. Prestaties op de schoolvorderingen is ook een latente variabele, geïndiceerd door de taal-, reken- en informatieverwerking. Er zijn geen effecten van categorie leerling op ouderlijk milieu, prestaties, overadvisering of schoolloopbaan.

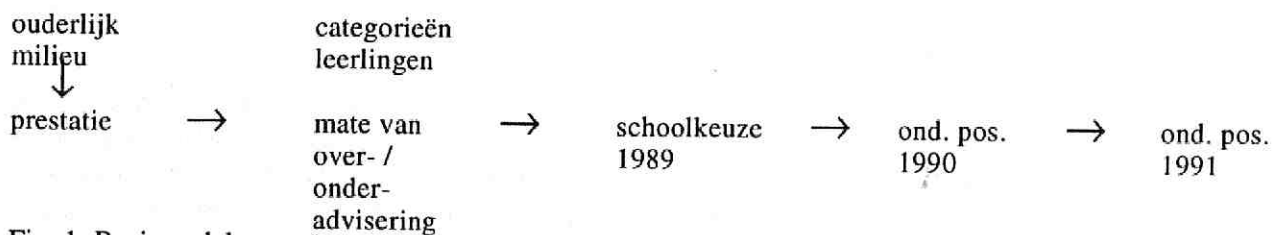


Fig. 1. Basismodel.

Model 1 blijkt met een χ^2 van 9860 en 188 vrijheidsgraden zeer slecht te passen (zie tabel 5, eerste rij). In model 2 worden de effecten van de prestaties op de schoolkeuze en op de onderwijspositie in 1990 en 1991 toegestaan. Hoge prestaties op de schoolvorderingstoets vergroten de kans op de keuze voor een hoog schooltype en een hoge onderwijspositie in 1990 en 1991. Tevens veronderstellen we effecten van de mate van onder- en overadvisering op de onderwijspositie in 1990 en 1991. Hiermee stellen we dat leerlingen die zijn overgeadviseerd enerzijds in het voortgezet onderwijs een grotere kans op mislukken hebben dan andere leerlingen. Anderzijds hebben leerlingen die zijn overgeadviseerd een grotere kans om door een hoger schooltype uitgedaagd te worden dan leerlingen die niet zijn overgeadviseerd. Ook veronderstellen we in model 2 dat de schoolkeuze van invloed kan zijn op de onderwijspositie in 1990 en 1991. Dit model 2 blijkt in vergelijking met model 1 een significant beter model (χ^2 verschil tussen model 1 en 2 van 6362 bij een verschil van 6 vrijheidsgraden; zie tabel 5), maar het past nog steeds zeer slecht bij de data.

Model 3 bouwt verder op model 2. Naast de reeds veronderstelde effecten van model 2 veronderstellen we dat in model 3 ook dat het ouderlijk milieu, naast de invloed op de prestaties op de basisschool, ook van invloed is op de verdere loopbaan van de leerlingen. Ook model 3 is vergeleken met model 2 een verbetering (χ^2 verbetering 232, vrijheidsgraden verschil 2), maar het past nog steeds slecht bij de data.

We verbeteren model 3 stapsgewijs (model 4 tot en met model 8) door te veronderstellen dat de schoolloopbaan van de tien categorieën migrantenleerlingen en de categorie autochtone leerlingen verschillend is. Elk model bouwt dus verder op de veronderstellingen van de eerdere modellen. Elk van deze modellen blijkt een significante verbetering op te leveren in vergelijking met het voorafgaande model. Vervolgens veronderstellen we in model 9 de effecten de verschillende categorieën leerlingen op het ouderlijk milieu gelijk. In model 10 voegen wij de significante gecorreleerde meetfouten tussen de indicatoren toe. Alle toegevoegde gecorreleerde meetfouten hebben betrekking op bepaalde indicatoren van het ouderlijk milieu, die voor bepaalde categorieën migrantenleerlingen significant afwijken van die van andere categorieën migrantenleerlingen en van autochtone leerlingen.

Tabel 5. Modelselectie.

TOEGEVOEGDE EFFECTEN:	χ^2	DF	FIT
1. Basismodel: model 1	9860	188	P= <.00
2. Onderlinge invloed schoolloopbaan-gebeurtenissen: model 2	3498	183	P= <.00
3. Sociaal milieu op de schoolloopbaan-gebeurtenissen: model 3	3266	179	P= <.00
4. Categorieën op prestaties: model 4	3246	174	P= <.00
5. Categorieën op de mate van overadvisering: model 5	3066	170	P= <.00
6. Categorieën op schoolkeuze: model 6	2973	165	P= <.00
7. Categorieën op onderwijspositie in 1990: model 7	2951	163	P= <.00
8. Categorieën op onderwijspositie in 1991: model 8	2943	160	P= <.00
9. Categorieën op het ouderlijk milieu: model 9	1416	151	P= <.00
10. Gecorreleerde meetfouten van de indicatoren: model 10	865	145	P= <.00
11. Niet-significante effecten eruit: model 11	887	158	P= <.00
12. Toevoeging van significant geworden effecten en verwijdering van niet-significante effecten: model 12	574	152	P= <.00

In de laatste twee modellen zijn parameters die in de vorige modellen niet significant zijn geworden alsnog op nul gesteld en zijn parameters die in de vorige modellen significant geworden zijn alsnog toegevoegd.

De hier geschetste modelselectie is samengevat in tabel 5. Het eindmodel past niet perfect, maar dat is gezien zijn complexiteit niet vreemd. De adjusted fit is zeer bevredigend: .99. Een verder toevoegen van meer parameters zou tot theoretisch onwaarschijnlijke effecten leiden.

De parameters van het eindmodel zijn samengevat in tabel 6a en 6b. Het eerste dat in tabel 6a opvalt is de grote betekenis die het ouderlijk milieu heeft bij de verklaring van verschillen in schoolloopbanen tussen migrantenleerlingen en autochtone leerlingen. In de eerste kolom van

Tabel 6a. Parameters van het eindmodel 12 (gestandaardiseerd voor causale parameters, ongestandaardiseerd voor meetmodellen).

	MILIEU	PRES	OVERADV	POS89	POS90	POS91
TUR	-.42	.00	.17	.03	.00	.00
MAR	-.41	.00	.19	.00	.00	.00
SUR	-.08	-.03	.07	.05	.00	.00
ANT	.05	-.06	.00	.02	.00	.00
MOL	.00	.00	.00	.00	.00	.00
ZUER	-.13	.00	.06	.00	.00	.00
CHIN	-.10	.00	.07	.00	-.04	-.03
OVAZ	-.05	.00	.03	.00	.00	-.02
OV1W	.00	.00	.03	.02	.00	-.02
OV3W	.00	.00	.00	-.03	.00	.00
MILIEU		.54	.44	.00	.00	.00
PRES			-.22	.79	.51	.30
OVERADV				.53	.27	.14
POS89					.32	-.24
POS90						.79
BERVA	1.00					
BERMA	.93					
OPLVA	1.38					
OPLMA	1.42					
TAAL		1.00				
REK		1.01				
INFO		.96				
R ²		.29	.10	.84	.75	.76

legenda: alle genoemde parameters die afwijken van .00 zijn significant (t-toets, +/- 1.96).

Tabel 6b. De gecorreleerde meetfouten bij de meetmodellen van het milieu van herkomst.

	BERVA	BERMA	OPLVA	OPLMA
TUR				
MAR		-.14		
SUR		-.07		
ANT				
MOL	.03			
ZUER				
CHIN		.10	.03	.06
OVAZ		.11	-.02	
OV1W		-.03		
OV3W	-.06	.08		

legenda: alle genoemde parameters die afwijken van .00 zijn significant (t-toets +/- 1.96)

tabel 6a is te zien dat met name Turkse en Marokkaanse leerlingen, vergeleken met de autochtone leerlingen, uit een ongunstig ouderlijk milieu komen. De effecten van het ouderlijk milieu op de schoolloopbaan zijn groot, voor zowel allochtone als autochtone leerlingen, terwijl de negatieve effecten van de categorieën migrantenleerlingen, gecontroleerd voor het effect van het ouderlijk milieu, gering zijn. Dit toont nogmaals aan dat de onderwijsachterstanden van migrantenleerlingen voor het belangrijkste deel voortvloeien uit de relatief lage maatschappelijke positie die de ouders van de leerlingen in de Nederlandse samenleving innemen. Als men voor dit effect van de sociale herkomst van de leerlingen controleert, zijn er nauwelijks directe effecten van één van de categorieën migrantenleerlingen op de prestaties op de schoolvoorderingstoets over. Alleen de Antilliaanse en de Surinaamse leerlingen presteren iets lager dan de autochtone leerlingen. De gecorrleerde meetfouten tussen de indicatoren van het ouderlijk milieu (tabel 6b) bij sommige categorieën migrantenleerlingen (Marokkaanse, Surinaamse, Antilliaanse, Aziatische (incl. Chinese) categorieën leerlingen en de categorie leerlingen uit de Derde Wereld) laten zien dat deze indicatoren niet in alle gevallen hetzelfde gewicht hebben als bij de autochtone leerlingen. Met name de betekenis van het niveau van het beroep van de moeder wijkt nogal eens af van dat van autochtone leerlingen. Dit is echter een nuance, vergeleken met de algemene betekenis van de verschillen in maatschappelijke positie van de ouders van migrantenleerlingen en autochtone leerlingen. De afwezigheid van belangrijke effecten van de categorieën migrantenleerlingen op de schoolprestaties onderstreept dat de veronderstelde cultuur van het land van herkomst de migrantenleerlingen niet in de weg hoeft te staan bij hun schoolprestaties.

De derde kolom van tabel 6a (overadvisering) laat zien dat, gecontroleerd voor ouderlijk milieu en voorafgaande schoolprestaties, bijna alle categorieën migrantenleerlingen in vergelijking met autochtone leerlingen overgeadviseerd worden (uitgezonderd de Antilliaanse, Molukse en Derde Wereld leerlingen). Overadvisering heeft geen negatieve effecten op de schoolloopbaan van migrantenleerlingen. De effecten van overadvisering op de verdere schoolloopbaan zijn vrij groot (.53, .27, .14), maar dit leidt, na controle voor de onderwijspositie in het voorgaande jaar, de schoolprestaties en het ouderlijk milieu, tot een slechtere schoolloopbaan dan die van de autochtone leerlingen. De overadvisering van migrantenleerlingen kan hun verschil in schoolloopbanen met autochtone leerlingen niet verklaren. De lage maatschappelijke positie van de ouders van de leerlingen en de daarmee samenhangende lage schoolprestaties op de basisschool, zijn wél een verklaring voor deze verschillen in schoolloopbanen. In de derde kolom van tabel 6a vallen nog twee zaken op. In de eerste plaats is te zien dat leerlingen die op de basisschool laag presteren, een grotere kans hebben dan de andere leerlingen om een overadvies te krijgen. Leerlingen die goed presteren op de schoolvoorderingstoets hebben, in vergelijking met andere leerlingen, een grotere kans op een onderadvies. Tegelijkertijd laat de derde kolom zien dat overadvisering een milieu-effect is: het effect van ouderlijk milieu op overadvisering is .44: overadvisering komt vooral voor bij leerlingen met ouders uit de hogere maatschappelijke lagen (dus meestal autochtone leerlingen).

Er is geen categorie migrantenleerlingen die significant betere schoolprestaties heeft dan autochtone leerlingen, na controle voor ouderlijk milieu (kolom 2 tabel 6a). Wel blijkt uit de vierde kolom dat sommige categorieën migrantenleerlingen in het eerste leerjaar voortgezet onderwijs een hogere onderwijspositie hebben dan autochtone leerlingen, na controle voor ouderlijk milieu, schoolprestaties en overadvisering (Turken, Surinamers, Antillianen, migranten uit de Eerste Wereld). Voor de onderwijsposities in het tweede en derde leerjaar voortgezet onderwijs (laatste twee kolommen tabel 6a) bestaan dergelijke positieve directe effecten niet. Dit resultaat betekent enerzijds dat migrantenleerlingen uit minder geëuropaniseerde samenlevingen geen slechtere schoolloopbanen hebben dan leerlingen uit meer geëuropaniseerde samenlevingen, gecontroleerd voor ouderlijk milieu en eerdere schoolprestaties. Anderzijds laat tabel 6a zien dat migrantenleerlingen betere schoolloopbanen kunnen hebben dan autochtone leerlingen, gecontroleerd voor ouderlijk milieu en eerdere schoolprestaties.

Tenslotte laat tabel 6a zien dat aan een hoge schoolkeuze voor voortgezet onderwijs, als die niet worden ondersteund door goede schoolprestaties, bepaalde risico's vastzitten: het kan in het

derde jaar leiden tot een lagere onderwijspositie. Het directe effect van de schoolkeuze in het eerste leerjaar op de onderwijspositie in het derde leerjaar is namelijk $-.24$, terwijl het directe effect van de prestaties op de basisschool op de onderwijspositie in het derde leerjaar $.30$ is. Dit negatieve effect van de schoolkeuze op de onderwijspositie in het derde leerjaar kan een gevolg zijn van een zogenaamd plafond-effect: een hoge schoolkeuze in het eerste leerjaar kan in het derde leerjaar niet meer leiden tot een nog hogere onderwijspositie maar alleen tot een lagere onderwijspositie. Een lage schoolkeuze in het eerste leerjaar daarentegen biedt nog wel ruimte voor een hoger onderwijspositie in het derde jaar.

5. VERSCHILLEN TUSSEN SCHOOLLOOPBANEN

Uit de resultaten van de vorige paragraaf mag echter niet geconcludeerd worden dat de schoolloopbanen van migrantenleerlingen en autochtone leerlingen niet verschillen. Een hoog advies kan voor de ene categorie migrantenleerlingen andere gevolgen voor de schoolloopbaan hebben dan voor een andere categorie, omdat een hoog advies in beide gevallen anders geïnterpreteerd wordt. Bij de vaststelling van goede schoolprestaties aan het eind van de basisschool kunnen taal-, reken- en informatietoets een andere effect of gewicht hebben bij autochtone leerlingen dan bij migrantenleerlingen. Deze gewichten kunnen tussen categorieën migrantenleerlingen, afhankelijk van hun migratiegeschiedenis, ook nog verschillen. Hetzelfde geldt voor de meting van het ouderlijk milieu. Een bepaald ouderlijk opleidingsniveau heeft in het land van herkomst vaak een andere betekenis dan hetzelfde opleidingsniveau van autochtone ouders. Om deze mogelijke verschillen nader te onderzoeken vergelijken wij de afzonderlijke schoolloopbanen van migrantenleerlingen en autochtone leerlingen met elkaar. In deze vergelijking staat de vraag centraal in hoeverre de schoolloopbanen (inclusief meting van ouderlijk milieu en schoolprestaties) tussen de verschillende categorieën migrantenleerlingen en autochtone leerlingen verschillen. Daarbij besteden wij extra aandacht aan de verschillen in effecten van overadvisering.

In tabel 7 zijn de verschillende ongestandaardiseerde regressie-coëfficiënten van het in de vorige paragraaf gebruikte LISREL-model⁴ voor vijf⁵ verschillende categorieën (autochtonen; Turken; Marokkanen; Surinamers; Antillianen) weergegeven. Deze tabel laat duidelijk zien dat er verschillen bestaan tussen de schoolloopbanen van deze vijf categorieën migrantenleerlingen en autochtone leerlingen: de structuur van het schoolloopbaanmodel is niet erg verschillend, het zijn vooral de parameters die verschillen. Een aantal parameters, dat nodig is voor een goede voorspelling van de schoolloopbaan van allochtone leerlingen, is niet significant⁶ bij migrantenleerlingen: met name beroep en opleiding van de migrantenmoeder hebben minder betekenis voor de bepaling van het ouderlijk milieu: op de opleiding van moeders van Surinaamse leerlingen na, zijn de parameters van het beroeps- en opleidingsniveau van de moeders van de andere categorieën migrantenleerlingen niet significant afwijkend van nul. Tabel 7 laat tevens zien dat het ouderlijk milieu van migrantenleerlingen minder invloed op de schoolprestaties en de mate van overadvisering heeft dan bij de autochtone leerlingen: ook hier zijn, op het effect van het ouderlijk milieu van Surinaamse leerlingen op hun prestaties na, alle invloeden van het ouderlijk milieu van migrantenleerlingen niet significant afwijkend van nul.

Gezien de migratiegeschiedenissen van deze categorieën leerlingen is dit resultaat niet vreemd: in de jaren zestig zijn door de Nederlandse overheid en het Nederlandse bedrijfsleven ongeschoolde Turkse en Marokkaanse werknemers geworven. Hun opleidings- en beroepsniveau was laag. Dit heeft tot gevolg gehad dat dit opleidings- en beroepsniveau weinig verklaart in de variantie van de prestaties bij de Turkse en Marokkaanse leerlingen. Het is namelijk goed mogelijk dat onder deze leerlingen nog potentieel "verborgen talent" aanwezig is. Turkse en Marokkaanse leerlingen kunnen, gezien hun migratiegeschiedenis, niet systematisch de gelegenheid hebben gekregen zich, onder de randvoorwaarden van het Nederlandse onderwijs, te ontwikkelen. Dit is wel het geval bij Antilliaanse en Surinaamse leerlingen: hun prestaties op de basisschool zijn reeds goed voorspelbaar.

Algemeen kan men uit tabel 7 concluderen dat er, ondanks alle overeenkomsten in de werking van de belangrijkste schoolloopbaanprocessen, toch verschillen bestaan tussen de categorieën migrantenleerlingen en de autochtone leerlingen.

Teneinde deze verschillen te toetsen hebben wij de parameters van de vijf categorieën leerlingen systematisch in LISREL analyses vergeleken op de mate van onderlinge overeenkomst. De procedure bij een dergelijke toetsing met LISREL is het gelijkstellen van de te analyseren parameters teneinde de χ^2 van deze analyse te vergelijken met de χ^2 van een LISREL analyse waarin deze parameters niet gelijk gesteld zijn. Indien het verschil in χ^2 en vrijheidsgraden van beide analyses geen significant verschil oplevert (het verschil in χ^2 is klein in vergelijking met het verschil in vrijheidsgraden), dan mag men aannemen dat beide parameters niet significant van elkaar verschillen. In tabel 8 zijn de uitkomsten van dergelijke analyses opgenomen en worden hun uitkomsten onderling vergeleken. De meest opvallende uitkomst van tabel 8 is dat

Tabel 7. De ongestandaardiseerde parameters van het schoolloopbaanmodel voor autochtonen, Turken, Marokkanen, Surinamers en Antillianen.

	Aut.	Tur.	Mar.	Sur.	Ant.
Prestaties op:					
- rekenen	1.22	1.20	1.14	1.22	1.21
- informatie	.96	.72	.79	.71	.94
ouderlijk milieu op:					
- beroep moeder	.33	.42	2.90	-.05	1.76
- opleiding vader	1.92	5.41	6.66	2.64	4.30
- opleiding moeder	1.44	2.34	4.34	3.09	5.46
prestaties op:					
- overadvies	-.42	-.89	-.42	-.36	-1.31
- onderwijspositie 1989	2.37	2.83	2.61	1.84	1.90
- onderwijspositie 1990	1.98	2.26	1.76	1.02	-1.06
- onderwijspositie 1991	1.42	.69	.98	1.14	.93
overadvies op:					
- onderwijspositie 1989	.84	.94	.69	.75	.83
- onderwijspositie 1990	.55	.55	.44	.25	-.69
- onderwijspositie 1991	.38	.05	.31	.30	.18
onderwijspositie 1989 op:					
- onderwijspositie 1990	.29	.29	.51	.57	1.47
- onderwijspositie 1991	-.37	-.11	-.16	-.46	-.57
Onderwijspositie 1990 op:					
- onderwijspositie 1991	.88	1.04	.78	1.10	1.27
Ouderlijk milieu op:					
- prestaties	2.03	2.15	-2.46	2.77	5.37
- overadvies	3.41	3.72	-2.50	1.95	17.54
Gecorreleerde meetfouten:					
Taal & rekenen	-1.08	-1.06	1.09	-2.97	-.84
beroep & opleiding vader	.38	-.04	.08	.52	1.40
ber. moeder & opl. vader	-.08	-.09	-.37	.27	-.22
beroep & opleiding moeder	.12	.20	.24	.70	.01
R ²					
prestaties	.25	.03	.02	.20	.23
overadvies	.13	.14	.03	.04	.40
onderwijspositie 1989	.87	.84	.72	.69	.81
onderwijspositie 1990	.79	.57	.71	.58	.48
onderwijspositie 1991	.79	.76	.70	.73	.73

legenda: onderstreepte parameters zijn niet-significant (t-toets < +/- 1.96)

bijna alle parameters significant verschillen tussen de vijf categorieën. Zelfs de parameters van de meetmodellen van ouderlijk milieu en prestaties aan het eind van de basisschool verschillen. Dat betekent dat de verschillen in ongestandaardiseerde parameters uit tabel 7 reële verschillen zijn, die niet verklaard kunnen worden uit toevallige verschillen of uit de verschillende omvang van de categorieën migrantenleerlingen en allochtone leerlingen. Opvallend is dat zelfs onderwijsprocessen die verondersteld worden "universalistisch", dat wil zeggen voor allen op dezelfde wijze te werken, dat blijkens deze analyse niet doen. Hun effect verschilt per categorie leerlingen.

Dit alles betekent niet dat de gevonden verschillen steeds of zelfs in hoofdzaak in het nadeel van migrantenleerlingen werken. Een goed voorbeeld is het effect van het ouderlijk milieu op de schoolloopbaan: de prestatie en het overadvies van de autochtone leerling wordt door zijn ouderlijk milieu beïnvloed; bij de migrantenleerlingen is dit nauwelijks het geval (Surinaamse leerlingen zijn hier de uitzondering). Door deze mengeling van voor- en nadelen van de verschillende schoolloopbanen van autochtone en migrantenleerlingen is het mogelijk dat uiteindelijk een groot deel van de onderwijsachterstanden van migrantenleerlingen verklaard kan worden door de normale "klassieke" oorzaken van onderwijsachterstand, zoals de vorige paragraaf al liet zien. Ook de effecten van overadvisering op de verdere schoolloopbaan verschillen tussen de vijf categorieën migranten- en autochtone leerlingen. Over het algemeen profiteren autochtone, Turkse en Marokkaanse leerlingen het langst van de overadvisering (met name voor onderwijspositie 1990), terwijl Surinaamse leerlingen er voor hun onderwijspositie 1990 minder voordeel uittrekken: Antilliaanse leerlingen worden door de overadvisering voor hun onderwijspositie 1990 benadeeld. Net als bij autochtone leerlingen heeft overadvisering ook bij migrantenleerlingen een positief effect op het eerste deel van hun loopbaan in het voortgezet onderwijs.

Tabel 8. Verschillende LISREL analyses over de vijf verschillende categorieën leerlingen uit tabel 7.

model	chi ²	df	model	chi ²	df
A. Alle parameters vrij	414	160			
B. prestatie model gelijk	445	168	B-A	30	8
C. milieumodel gelijk	508	172	C-A	94	12
D. effecten prestatie en onderwijsposities gelijk	535	200	D-A	121	40
E. effecten prestatie gelijk	467	176	E-A	53	16
F. effecten overadvies gelijk	448	172	F-A	34	12
G. overadvies-onderwijs '89	427	164	G-A	13	4
H. overadvies-onderwijs '90	430	164	H-A	16	4
I. overadvies-onderwijs '91	417	164	I-A	3	4
J. effecten milieu gelijk	471	168	J-A	57	8

legenda: onderstreepte chi² geeft een verschil tussen de modellen dat niet significant (p=.95) afwijkt van geen verschil tussen de modellen.

6. CONCLUSIE

De analyses van dit artikel laten duidelijk zien dat de onderwijsachterstand van migrantenleerlingen niet verklaard kan worden door overadvisering. Migrantenleerlingen worden meer overgeadviseerd dan autochtone leerlingen uit een vergelijkbaar sociaal milieu, maar de overadvisering wordt nog sterker beïnvloed door het ouderlijk milieu: hoe hoger dat is, hoe vaker er van overadvisering sprake is. Tot een vergelijkbare conclusie komen Mulder en Pijl (1992: 59) ook.

Allochtone leerlingen ondervinden van die overadvisering geen extra nadeel in hun verdere

loopbaan gedurende de eerste drie jaar van het voortgezet onderwijs. Het lijkt erop dat de leerkrachten op de basisschool het 'verborgen talent' van migrantenleerlingen kunnen inschatten: van het positief discrimineren van migrantenleerlingen zou men immers negatieve consequenties verwachten in de verdere schoolloopbaan. Uit de analyses in dit artikel blijkt dat dit niet het geval is. Dit resultaat stemt overeen met een analyse van Mulder en Pijl (1992: 68) van de onderwijspositie van OVB leerlingen in het tweede leerjaar van het voortgezet onderwijs. Zij concluderen daar dat de ongunstiger loopbaan in het voortgezet onderwijs van allochtone leerlingen veel meer het gevolg is van de hoge schoolkeuze in het eerste leerjaar dan van het relatief hoge advies. Een negatief effect van een hoge onderwijspositie in het eerste leerjaar op de onderwijspositie in het derde leerjaar vinden wij ook. Van overadvisering profiteren bovendien niet alle categorieën leerlingen in gelijke mate: autochtone, Turkse en Marokkaanse leerlingen profiteren het langste van overadvisering, terwijl Surinaamse leerlingen er minder voordeel uittrekken. Antilliaanse leerlingen worden door de overadvisering uiteindelijk benadeeld.

Het milieu van herkomst, gemeten in het opleidings- en beroepsniveau van de ouders van de leerlingen, is voor het grootste deel verantwoordelijk voor de onderwijsachterstanden van de leerlingen. Het grootste deel van de onderwijsachterstanden van migrantenleerlingen kan dus verklaard worden door de lage maatschappelijke positie die de ouders van deze leerlingen in de Nederlandse samenleving innemen. Dit betekent niet dat de schoolloopbanen van migrantenleerlingen hetzelfde verlopen als die van autochtone leerlingen. De vergelijking van de schoolloopbanen van de vier belangrijke categorieën migrantenleerlingen laat duidelijk verschillen zien (tabel 7). Bij de meeste categorieën migrantenleerlingen die voor deze analyse zijn opgenomen, heeft het ouderlijk milieu over het algemeen een geringere invloed dan bij autochtone leerlingen. Deze beide analyses geven dus steun aan twee schijnbaar tegengestelde stellingen: enerzijds zijn de normale "klassieke" milieu effecten in hoofdzaak verantwoordelijk voor de lagere schoolprestaties van migrantenleerlingen, anderzijds verschillen de patronen van schoolloopbanen van autochtone en migrantenleerlingen, zodat men op goede gronden kan volhouden dat een aparte aanpak van migrantenleerlingen gerechtvaardigd is. Deze schijnbare tegenstelling is goed op te lossen. Door de migratiegeschiedenis van de verschillende categorieën allochtone leerlingen (meestal emigreren alleen de laaggeschoolden naar een welvarender land waarin zij de laagst gewaardeerde banen krijgen) behoren hun ouders gemiddeld tot een veel lager sociaal milieu dan de ouders van autochtone leerlingen. Dit leidt ertoe dat een zeer laag ouderlijk milieu in hoge mate bijna uitsluitend bij migrantenleerlingen gevonden wordt. Het verschil in sociaal milieu is zo groot dat andere, kleinere verschillen tussen allochtone en autochtone leerlingen daarbij in een analyse wegvallen. In een analyse waarin allochtone en autochtone leerlingen afzonderlijk behandeld worden zijn deze andere, kleinere verschillen wel zichtbaar omdat men dan het dominante verschil in ouderlijk milieu buiten de analyse houdt. Het voordeel van een dergelijke afzonderlijke analyse is dat het zicht op andere verschillen toeneemt, maar tegelijkertijd verwaarloost men in een dergelijke afzonderlijke analyse de dominante verschillen in ouderlijk milieu. Deze kleinere verschillen in schoolloopbanen zijn vervolgens goed te verklaren met de migratiegeschiedenis en met culturele verschillen. Een ander schoolloopbaanpatroon behoeft niet altijd alleen maar nadelen met zich mee te brengen. Het kan juist ook een voordeel betekenen (bijvoorbeeld kansen bieden aan potentieel 'verborgen talent'). Door deze mengeling van voor- en nadelen van de verschillende schoolloopbanen van autochtone en migrantenleerlingen en door de grote verschillen in ouderlijk milieu van autochtone en migrantenleerlingen, is het mogelijk dat uiteindelijk een groot deel van de onderwijsachterstanden van migrantenleerlingen verklaard kan worden door de normale 'klassieke' oorzaken van onderwijsachterstand, terwijl de processen van schoolloopbanen van de verschillende categorieën migranten- en autochtone leerlingen toch verschillen als gevolg van de migratiegeschiedenissen en uiteenlopende culturen. Indien dit juist is, zou dit betekenen dat de tegenstelling tussen òf klasse òf etnische afkomst als de verklaring van ongelijke schoolloopbanen van migrantenleerlingen een valse tegenstelling is.

NOTEN

1. Overadvisering is niet alleen afhankelijk van de capaciteiten van de leerling maar ook van het aanbod van schooltypen in de regio van de betrokken school. Leerlingen die in een regio wonen met een groot aanbod van onderwijstypen of leerlingen die terecht komen op grote scholengemeenschappen hebben om die reden meer kans op overadvisering en hebben betere doorstroomkansen dan leerlingen in regio met weinig 'hogere' schooltypen of weinig 'brede' scholengemeenschappen. Terzijde willen wij opmerken dat dit mogelijke effect niet overdreven mag worden, de vele nodale gebieden in Nederland zijn ruim voorzien met alle mogelijke schooltypen en scholengemeenschappen, maar uit dit verschil in aanbod zijn wel bepaalde regio-effecten te verklaren. Belangrijker voor onze probleemstelling zou echter zijn, wanneer de verschillende migranten ongelijk verdeeld zouden zijn over de regio's of een grote voorkeur zouden hebben voor brede scholengemeenschappen. Hoewel die verdeling inderdaad niet gelijk is, blijkt uit de VOCL '89 data (hier niet weergegeven) dat migranten zeker niet geconcentreerd zijn in gebieden met het grootste aanbod en zeker niet alleen kiezen voor brede scholengemeenschappen. Ten slotte blijkt ook uit de VOCL '89 data (hier niet weergegeven) dat het aantal brede adviezen (meer dan twee schooltypen; in feite niet-adviezen) niet zo heel groot is en dat deze brede adviezen niet alleen of in hoofdzaak aan migrantenleerlingen gegeven worden. Dat betekent vervolgens dat het genoemde probleem niet tot een ernstige kunstmatige overschatting van de mate van over- en onderadvisering bij migranten geleid kan hebben.
2. Een belangrijke mate betekent dat het waargenomen advies tenminste 1 niveau (in schooltypen gemeten) afwijkt van het advies dat op grond van de prestaties verwacht mocht worden. Dus wanneer een leerling op basis van zijn prestaties een MAVO-advies zou moeten krijgen, maar in werkelijkheid tenminste een HAVO-advies of hoger krijgt.
3. Het resterende percentage leerlingen heeft het cohort verlaten.
4. Uiteraard zonder de tien variabelen die de migranten categorieën indiceerden.
5. Wegens gebrek aan computerwerkruimte konden niet alle elf categoriën leerlingen met een dergelijk complex schoolloopbaan binnen LISREL vergeleken worden. Daarom hebben wij ons in eerste instantie beperkt tot die categoriën migrantenleerlingen waarvoor maatschappelijk en wetenschappelijk in Nederland op dit moment de meeste aandacht bestaat.
6. Men moet zich echter niet blind staren op de significantie van de parameters van de verschillende groepen leerlingen. Uiteraard zijn de aantallen leerlingen bij de migrantenleerlingen kleiner dan bij de autochtone leerlingen, waardoor verschillen kunnen ontstaan bij het significant worden van parameters. De kleinste categorie was de Antilliaanse (68), de andere groepen waren veel groter.

LITERATUUR

- Batenburg, Th.A. Van & Lokman, A.H. (1992). *Leerlingen, scholen en schoolloopbanen in het MBO*. Tweede rapport SVO-project 9075. Groningen: RION.
- Benton, G., Vermeulen, H., (red.) (1987). *De Chinezen*. Migranten in de Nederlandse samenleving; 4. Dick Coutinho, Muiderberg, p. 6-11.
- CBS (1991). *Schoolloopbanen en achtergrond van leerlingen*. (VOCL '89, cohort 1989, peildatum september 1990; Mededelingen no. 7918.
- CBS (1991). *Schoolloopbanen en achtergrond van leerlingen*. Deel 1; Cohort 1989, instroom. CBS-publicaties, 's-Gravenhage: SDU.
- Cultureel Planbureau (1992). *Sociaal en Cultureel Rapport 1992*. Hoofdstuk 7: Onderwijs, Den Haag: VUGA, (p. 229-284).
- Dronkers, J. (1992). Zullen wij voor de kinderen bij elkaar blijven? De veranderende effecten van eenoudergezinnen op de schoolloopbanen van de kinderen. In *Mens en Maatschappij*, 67, 23-44.
- Eeden, P. Van den, Jong, U. De, Koopman, P., & Roeleveld, J. (1993). *Schoolloopbanen in Amsterdam*. Leerlingen op Amsterdamse scholen voor AVO in de jaren zeventig. Amsterdam/Lisse: Swets & Zeitlinger.
- Hof, L. Van 't & Dronkers, J. (1993). Onderwijsachterstanden van allochtonen: klasse, gezin of etnische cultuur? In *Migrantenstudies*, 9, 2-25.
- Jong, M.J. de (1987). *Herkomst, kennis en kansen*. Allochtone en autochtone leerlingen tijdens de overgang van basis- naar voortgezet onderwijs. Lisse: Swets & Zeitlinger.
- Koeslag, M. (1993). *Sociaal Milieu en Taal? Verklaringen voor het verschil in schoolloopbanen van mi-*

- grantenleerlingen en autochtone leerlingen uit twee-oudergezinnen*. Doctoraal scriptie sociologie Katholieke Universiteit Brabant, Tilburg.
- Kohnstamm, R. (1993). Hogerop. In *NRC-Handelsblad*, 25 februari.
- Langen, A. Van & Jungbluth, P. (1990). *Onderwijskansen van migranten. De rol van sociaal-economische en culturele factoren*. Amsterdam/Lisse: Swets & Zeitlinger.
- Mulder, L., & Tesser, P. (1992). *De schoolkeuzen van allochtone leerlingen*. Nijmegen: ITS.
- Mulder, L., & Pijl, B. (1992). *De onderwijspositie van leerlingen uit de OVB-doelgroepen na twee jaar voortgezet onderwijs*. Nijmegen: ITS.
- Pels, T. (1993). Het belang van een cultuurbeleid in het onderwijs. In *Comenius*, 13, 42-55.
- Roelandt, T., Martens, E., & Veenman, J. (1990). Achterstanden van allochtonen in het onderwijs: sociaal milieu en migratie-achtergronden. *Mens en Maatschappij*, 65, 103-125.
- Schouten, S.P. (1990). Schoolloopbanen van Turken en Marokkanen in het voortgezet onderwijs. In: M. du Bois-Reymond en L. Eldering (red), *Nieuwe oriëntaties op school en beroep: de rol van sexe en etniciteit* (p. 163-176). Amsterdam/Lisse: Swets & Zeitlinger.
- Tesser, P., Werf, G. van der & Mulder, L. (1990). Het onderwijsvoorrangsbeleid en de onderwijskansen van leerlingen uit etnische minderheidsgroepen. In: M. du Bois-Reymond en L. Eldering (red), *Nieuwe oriëntaties op school en beroep: de rol van sexe en etniciteit* (p. 151-162). Amsterdam/Lisse: Swets & Zeitlinger.
- Tesser, P., & Mulder, L. (1990). Concentratie en prestatie van allochtone leerlingen in het basisonderwijs. In *Migrantenstudies*, 6, 31-45.
- Veenman, J. (1993). Voorzichtigheid geboden! Commentaar bij het artikel van Van 't Hof en Dronkers. In: *Migrantenstudies*, 9, 26-30.

Manuscript ontvangen 10-12-1993

Definitieve versie ontvangen 3-5-1994

BIJLAGE 1

Het onderscheid tussen de (meeste) categorieën is gemaakt op basis van de komst van migranten naar Nederland. De categorieën verschillen op basis van redenen van migratie: ex-kolonialen en arbeidsmigranten.

Sommige categorieën zijn bovendien van elkaar onderscheiden op basis van herkomstgebied (d.w.z. Azië, Noord-, West-, Oost-Europa, Canada en Australië en 'de Derde Wereld'). Het indelingscriterium voor deze categorieën is nationaliteit van de ouders geweest, omdat gegevens over het geboorteland ontbreken. Voor de categorieën die zijn gemaakt op basis van hun migratieverleden, geldt als indelingscriterium het geboorteland van de ouders, tenzij anders is vermeld.

De eerste categorie migranten zijn mensen uit de voormalige Nederlandse koloniën. Rond het onafhankelijk worden van Indonesië (1949) kwamen mensen van Nederlands-Indonesische afkomst (± 150.000) naar Nederland. Deze categorie migranten is geheel door de Nederlandse samenleving opgenomen, omdat ze al voor hun komst op de Nederlandse samenleving waren georiënteerd en omdat het overheidsbeleid gericht is geweest op integratie. Omdat het hier gaat om 12 moeders en 1 vader, is deze categorie migranten niet als aparte categorie onderscheiden. Onder de categorie migranten uit de voormalige koloniën vallen ook de Surinamers en Antillianen. Vooral in het begin van de jaren 70 kwamen er veel Antillianen (er werd actief door Nederlandse bedrijven in de periode 1964-1966 geworven) naar Nederland. Tot 1975 (onafhankelijkheid van Suriname) kwamen er veel (Creoolse) Surinamers uit de midden en hogere klasse naar Nederland om te studeren of te werken (in de hogere beroepen). Na 1975 kwamen ook ongeschoolde Surinamers (Hindoestaanse en Javaanse Surinamers, uit vrees voor de politieke overheersing van de Creoolse Surinamers) naar Nederland. Zij waren minder goed georiënteerd op de Nederlandse samenleving en beheersten de Nederlandse taal minder dan hun voorgangers.

Een andere categorie migranten die ook in de jaren '50 naar Nederland is gekomen, zijn de migranten van de Molukken. De Molukken kunnen, strikt genomen geen land van herkomst zijn

en vallen in principe onder Indonesië. Toch is er voor gekozen om leerlingen waarvan ten minste één van de ouders op de Molukken is geboren als een aparte categorie aan te merken. Ten eerste vallen de Molukken om politieke redenen niet samen met Indonesië. Toen Indonesië in 1949 onafhankelijk werd, riep dit verzet op op het eiland Ambon. Ambon riep de zelfstandige staat Maluku Selatan uit. Indonesië, met een uit oud-KNIL-militairen geformeerd nieuw leger, bezette dit gebied. De voormalig KNIL-militairen die van de Zuid-Molukken afkomstig waren, konden er moeilijk voor kiezen om het nieuwe leger van Indonesië te dienen. In Nederland werd besloten om deze KNIL-militairen en hun gezinnen tijdelijk op te vangen. Ze werden gehuisvest in kampen en het was voor hen verboden om aan de arbeidsmarkt deel te nemen (ze kregen wachtgeld). Het gevolg was dat ze (zeker in het begin) moeilijk in de Nederlandse samenleving integreerden. Ook de schoolprestaties van Molukse leerlingen bleven lange tijd achter bij die van autochtone leerlingen; velen stroomden dan ook naar de lagere schooltypen. Dit is een andere reden om de Molukse leerlingen als een eigen categorie aan te merken.

De derde categorie migranten is de arbeidsmigranten. Dit zijn buitenlandse werknemers die op verzoek van de Nederlandse regering of op verzoek van het Nederlandse bedrijfsleven, naar Nederland zijn gekomen.

Er is een onderscheid te maken in twee soorten arbeidsmigranten:

- 1) werknemers uit lidstaten van de E.G. Dit zijn Italiaanse (zij kwamen vooral in de jaren '50 in de mijnbouw werken), Spaanse, Portugese, en Griekse werknemers. Velen zijn getrouwd met een Nederlandse partner. Ook wat betreft religie en cultuur staan zij niet ver af van de Nederlanders. Deze categorie is ingedeeld op basis van de nationaliteit van de ouders.
- 2) werknemers uit landen ten zuiden van de Middellandse Zee. Pas in de periode 1960-1973 kwamen Turkse en Marokkaanse werknemers naar Nederland (arbeidsmigratie). Na 1973 werd er door de overheid niet meer actief geworven. Toch is vooral deze categorie migranten na 1973 sterk gegroeid als gevolg van de komst van de gezinnen van de werknemers.

Ook de Chinezen zijn als een eigen categorie onderscheiden. Eigenlijk kan deze categorie moeilijk als een homogene categorie worden gezien: China zelf is een land met grote verschillen in taal en cultuur per provincie en er zijn veel Chinezen in Nederland die eerder al naar andere landen zijn geëmigreerd (Nederlands-Indië, Hongkong, Suriname, Taiwan, Vietnam, Maleisië, Singapore). Dit heeft zijn weerslag in de samenstelling van de Chinese gemeenschap in Nederland, maar omdat ik geen informatie heb over het land van herkomst, is er geen onderscheid gemaakt in de diverse categorieën Chinese migranten. De Chinese migranten zijn als een eigen categorie onderscheiden omdat ze in Nederland een vrij gesloten gemeenschap vormen. Als indelingscriterium gold de nationaliteit van de ouders.

Ontwikkeling van beginnende lees- en spellingvaardigheid

Ludo Verhoeven¹ en Paul Gillijns²

¹Letterenfaculteit K.U.B., Tilburg

²Instituut voor Toetsontwikkeling (Cito), Arnhem*

ABSTRACT

In this paper the acquisition of early reading and spelling in Dutch as a first and second language will be dealt with. In a longitudinal study, the reading and spelling processes of native Dutch children and ethnic minority children were documented during grades 3 and 4 of primary school. The children were administered a number of tasks to test the efficiency of word identification and word spelling and comprehension processes. The results show that minority children keep up with native Dutch children on word identification tasks. However, on word spelling and reading comprehension they lag behind their Dutch peers. From linear structural analysis it is shown that both the native Dutch and the minority children's word identification and word spelling results can be predicted from metalinguistic word skills and grapheme-phoneme matching. Reading comprehension can be predicted from both word decoding and vocabulary knowledge. The structural components of first and second language reading and spelling processes turn out to be highly comparable.

1 INLEIDING

De uitkomsten van onderzoeken waarin het leesniveau van kinderen op bepaalde momenten in de basisvorming is gepeild, hebben de nodige discussie losgemaakt. Een internationale vergelijking door de International Association for the Evaluation of Educational Achievement (Elley, 1992) van de leesprestaties van 9- en 14-jarige leerlingen in 32 landen heeft voor Nederland tegenvallende prestaties laten zien. In Nederland zouden de 9-jarigen ongeveer een half tot een heel jaar achterlopen op leerlingen uit bijvoorbeeld de Scandinavische landen. Bij de 14-jarigen is de gemeten achterstand ietwat geringer, maar nog steeds significant.

Uit onderzoek blijkt dat lees- en schrijfproblemen reeds in een vroege fase van de basisvorming ontstaan (vgl. Van Dongen, 1984) en bovendien in hoge mate persistent zijn (Stanovich, 1986). Enkele recente onderzoeken laten zien dat reeds in de eerste fase van de basisvorming sprake is van grote verschillen in prestaties op (deel)vaardigheden tussen leerlingen op het terrein van lezen en schrijven. Bij de onlangs door het Instituut voor Toetsontwikkeling gehouden taalpeilingen medio en einde basisonderwijs is geprobeerd op basis van een expliciet criterium een profiel van de (schriftelijke) taalvaardigheid van 9-jarige leerlingen in kaart te brengen (Sijtsma, 1993). Voor *technisch lezen* en spelling wordt van een redelijke mate van beheersing gesproken. Wel wordt gewezen op grote verschillen tussen kinderen. Zo wordt bij technisch lezen het niveauverschil tussen de zwakste en de beste leerlingen geschat op tenminste drie leerjaren. Het algemeen niveau van *begrijpend lezen* wordt aangeduid als laag, vooral waar het gaat om het begrijpen van zakelijke teksten. Bovendien wordt ook bij begrijpend lezen gewezen op grote individuele verschillen tussen leerlingen. Daarnaast blijkt uit onderzoek van Verhoeven en Vermeer (1992) dat er ook in de *woordenschatontwikkeling* reeds bij de aanvang van het basisonderwijs sprake is van grote individuele verschillen. Met name bij kinderen die van huis uit een andere taal spreken dan het Nederlands kunnen we spreken van een substantiële achter-

*Adres: Cito, Postbus 1034, 6801 MG Arnhem

stand in de woordenschatomvang. Naarmate kinderen meer leeservaring opdoen, blijken de verschillen in lexicale kennis ook eerder groter dan kleiner te worden.

Bovenstaande onderzoeken wijzen erop dat de effectiviteit van de eerste fase van de basisvorming in schriftelijke vaardigheden te wensen overlaat. De vraag dringt zich op hoe het onderwijs zodanig kan worden verbeterd dat nagenoeg alle leerlingen tot een aanvaardbaar niveau van geletterdheid worden gebracht. Daarbij verdient de positie van allochtone leerlingen die voor de taak staan te leren lezen en schrijven in het Nederlands als tweede taal speciale aandacht. In het onderzoek dat hier wordt gepresenteerd, laten we vanuit een longitudinaal perspectief zien na hoe de woordenschat en de lees- en spellingvaardigheid van autochtone en allochtone kinderen zich ontwikkelt in de jaargroepen 3 en 4 van de basisschool. Bovendien gaan we binnen zowel de autochtone als de allochtone groep na welke relaties bestaan tussen lees- en spellingvaardigheden en belangrijke deelvaardigheden worden onderzocht.

2 THEORETISCH KADER

2.1 Leren lezen en schrijven in de moedertaal

Het leren lezen en schrijven van tekst is een complexe activiteit (voor een overzicht zie Verhoeven 1992a). Het veronderstelt het leren interpreteren of produceren van een betekenisrepresentatie op basis van informatie uit de tekst en uit het geheugen. Daarbij is het van groot belang dat de woorden uit de tekst snel en correct beschikbaar komen en dat er voldoende taalkennis en kennis van de wereld aanwezig is om betekenis te kunnen toekennen aan die woorden.

In het beginnend lees- en schrijfonderwijs krijgt het leren (de)coderen van woorden veel aandacht. Vanaf groep 3 volgen kinderen in de regel een gestructureerd curriculum, waarin stap-voor-stap het inzicht in het alfabetisch principe wordt bijgebracht (vgl. Verhoeven & Mommers, 1992; Adams, 1990). Zo ontwikkelt het kind een fonemisch bewustzijn en leert het via de relatie tussen letters en klanken woorden lezen en spellen. Wat het lezen betreft wordt de elementaire leeshandeling aangeleerd, waarbij de deelvaardigheden grafeem-foneem-koppeling en auditieve synthese centraal staan. Min of meer tegelijkertijd leren kinderen woorden spellen. Auditieve analyse en foneem-grafeem-koppeling zijn daarbij belangrijke deelvaardigheden.

Nadat het inzicht in de alfabetische structuur van ons schriftsysteem is bijgebracht, leren kinderen de grafo-fonemische processen efficiënter te laten verlopen, zodat geschreven woorden sneller kunnen worden herkend en opgeroepen. Bij de identificatie van woorden hoeft het kind daardoor niet meer 'spellenderwijs' te werk te gaan. In langere woorden worden deelstructuren herkend, waardoor het decoderen aanzienlijk wordt versneld en het korte-termijn-geheugen minder wordt belast. Naarmate woorden sneller worden herkend en opgeroepen, komt in het werkgeheugen meer capaciteit beschikbaar voor het verwerken van zinnen en het integreren van tekstuele informatie (vgl. Perfetti, 1985; Stanovich, 1986; Mommers, et al, 1986; Van den Bos & Lutje Spelberg, 1993).

Ten aanzien van begrijpend lezen geldt dat kinderen niet alleen informatie leren begrijpen die letterlijk in de tekst wordt verwoord, maar ook informatie leren afleiden en relaties leren leggen tussen verwijswaarden en hun antecedenten in de tekst. Onderzoek laat verder zien dat woordenschat een belangrijke determinant is van begrijpend lezen. Naarmate de woordenschat van kinderen groter is, blijken zij in staat om een tekst sneller en efficiënter te lezen (Aarnoutse & Van Leeuwe, 1988; Anderson & Freebody, 1981). Ook blijkt uit studies dat het begrijpend lezen kan worden verbeterd door onderwijs in lexicale kennis (McKeown, Omanson & Perfetti, 1983; Beck, Perfetti & McKeown, 1982).

Lees- en schrijfproblemen kennen uiteenlopende verschijningsvormen (Van der Leij, 1992; Spear & Sternberg, 1987). Zij kunnen het gevolg zijn van beperkingen in 'top down'-processen die worden gevoed door conceptuele kennis en geleid door strategieën, van tekorten in 'bottom up'-processen die de actuele tekst als uitgangspunt nemen, of van een combinatie van beide typen processen. Stanovich (1986) heeft er bovendien nog op gewezen dat top-down beperkin-

gen kunnen voortkomen uit bottom-up tekorten, in die zin dat een gebrekkige decodeervaardigheid de ontwikkeling van lexicale kennis kan belemmeren.

2.2 Leren lezen en schrijven in een tweede taal

Onderzoek laat zien dat de lees- en schrijffresultaten van allochtone leerlingen sterk achterblijven bij die van autochtone leeftijdsgenoten (Teunissen, 1986; Verhoeven, 1987, 1990a/b, 1992b; Hacquebord, 1989; Damhuis, et al., 1989). Wat betreft het begrijpen van functionele leestaken lopen allochtone leerlingen reeds in een vroeg stadium een achterstand op die veelal niet wordt ingelopen. Peilingsonderzoeken laten zien dat op het einde van de basisschool naar schatting 15 procent van de allochtone leerlingen als functioneel analfabeet kan worden aangemerkt (Triescheijn, Van den Bergh & Hoeksma, 1985; Zwarts, 1990). Naar verhouding blijken allochtone leerlingen vooral problemen te hebben met het begrijpen en produceren van informatieve teksten.

Vanuit pedagogisch-didactisch oogpunt geldt als belangrijke vraag vanuit welke variabelen de verschillen in schriftelijke taalvaardigheid tussen allochtone en autochtone leerlingen kunnen worden verklaard. Allochtone kinderen hebben een linguïstische achtergrond die in meer of mindere mate van die van Nederlandstalige kinderen verschilt. Welke invloed deze verschillen hebben op een mogelijk succes, of falen met betrekking tot het leren lezen is onduidelijk. In principe kunnen leesproblemen van allochtone kinderen worden teruggevoerd op problemen met decoderen of begrijpend lezen, of beide.

Decodeerproblemen kunnen op uiteenlopende wijzen verklaard worden. Ten eerste zouden allochtone leerlingen, als gevolg van een beperkt vermogen tot auditieve discriminatie in de tweede taal, problemen kunnen hebben met het vormen van een fonologische representatie van geschreven woorden. Het kan ook zijn dat allochtone leerlingen bij woordherkenning problemen ervaren bij het gebruik maken van orthografische 'constraints'. Ten slotte zou er bij decodeerproblemen sprake kunnen zijn van problemen met de directe woordherkenning als gevolg van een beperkte woordenschatomvang.

Met betrekking tot begrijpend lezen kunnen allochtone leerlingen als gevolg van een beperkte lexicale kennis van de tweede taal problemen hebben met het toepassen van syntactische en semantische beperkingen die top down-processen bij het lezen reguleren. Onderzoek in de Verenigde Staten heeft aan het licht gebracht dat allochtone leerlingen als gevolg van hun beperkte woordenschat vaak problemen hebben met begrijpend lezen in een tweede taal (vgl. Lucas & Singer, 1975; Rosales, 1981; Hauptman, 1981; Tregar & Fun Wong, 1984; Parish & Perkins, 1985).

3 VRAAGSTELLING

In tot dusver uitgevoerde studies naar de ontwikkeling van beginnende lees- en spellingvaardigheid is de vaardigheid in begrijpend lezen steeds gerelateerd aan één bepaald deelaspect van de linguïstische competentie van kinderen, zoals de woordenschat, de grammaticale kennis, of de kennis van tekststructuren. In de meeste gevallen was sprake van cross-sectionele studies, zoals de peilingsonderzoeken, waarbij de leerresultaten van leerlingen op één moment in de tijd werden onderzocht. Uitzonderingen vormen studies als die van Mommers et al. (1986) en Aarnoutse et al. (1986) naar de samenhang tussen decoderen, begrijpend lezen en spellen, en studies als die van Voeten (1991) en Vooijs et al. (1992) waarin met behulp van de multiniveaubenadering determinanten van de ontwikkeling van leesvaardigheid zijn onderzocht. In de eerdere studies is doorgaans verzuimd begrijpend lezen te definiëren als resultaat van een interactief proces waarin naast de decodeervaardigheid van kinderen hun woordenschat een minstens zo belangrijke variabele vormt. Bovendien geven de tot dusver uitgevoerde studies geen antwoord op de vraag in welk opzicht (deel)vaardigheden op het terrein van lezen en spellen van allochtone leerlingen verschillen van die van autochtone leerlingen. Eerder kleinschalig onderzoek van Verhoeven (1987, 1990a, 1992) heeft aangetoond dat de structuur van lees-leerproces-

sen in het Nederlands als eerste en tweede taal in hoge mate vergelijkbaar is en dat er alleen verschillen zijn te constateren van temporele aard. Voor de vaardigheden decoderen, spellen en begrijpend lezen kon met behulp van foutenanalyses worden vastgesteld dat tweede-taalleerders lager scoren dan moedertaalleerders, maar dat de structuur van leerprocessen in beide groepen leerders in hoge mate vergelijkbaar is. Genoemd onderzoek was niet alleen beperkt in de omvang en het aantal meetmomenten (2), het had bovendien grotendeels betrekking op allochtone leerlingen die in het land van herkomst waren geboren en op een onderwijscontext die nauwelijks was voorbereid op de grote toeloop van anderstaligen. Verder was het onderzoek naar structurele aspecten van lees- en spellingvaardigheden beperkt tot analyses van door leerlingen gemaakte fouten; een vergelijking van lees- en spellinggedrag in termen van onderliggende deelvaardigheden bij eerste- en tweede-taalleerders is niet eerder gemaakt.

In het onderhavige onderzoek zal het lees- en spellingproces van een steekproef van autochtone en in Nederland geboren allochtone leerlingen in de jaargroepen 3 en 4 van het basisonderwijs worden gevolgd en geanalyseerd. Daarbij proberen we na te gaan in hoeverre de leerprocessen van tweede-taalleerders naar tempo en structuur verschillend zijn ten opzichte van moedertaalleerders. De volgende vragen staan centraal in dit onderzoek:

1. Is er tussen autochtone en allochtone leerlingen in groep 3 en 4 van het basisonderwijs sprake van significante verschillen in tempo van ontwikkeling van (deel)vaardigheden van lezen en spellen?
2. In welke mate kunnen bij autochtone en allochtone leerlingen in groep 3 van het basisonderwijs individuele verschillen in decodeervaardigheid worden verklaard vanuit de deelvaardigheden grafeem-foneem-koppeling en auditieve synthese?
3. In welke mate kunnen bij autochtone en allochtone leerlingen in groep 3 van het basisonderwijs individuele verschillen in spellingvaardigheid worden verklaard vanuit de deelvaardigheden auditieve analyse en foneem-grafeem-koppeling?
4. In welke mate kunnen bij autochtone en allochtone leerlingen in groep 3 en 4 van het basisonderwijs individuele verschillen in de vaardigheid in begrijpend lezen worden verklaard vanuit de variabelen decodeervaardigheid en woordenschat?

4 OPZET VAN HET ONDERZOEK

4.1 Informanten

In dit onderzoek wordt gerapporteerd over gegevens die zijn verzameld in een longitudinaal onderzoek ten behoeve van de constructie en normering van het Leerlingvolgsysteem van het Cito (vgl. Gillijns & Verhoeven, 1991; 1992). Dit onderzoek is in 1989 bij een landelijk cohort van circa 3600 leerlingen gestart. Voor de samenstelling van het cohort werd eenzelfde procedure toegepast als bij het onderzoek in het kader van de Periodieke Peiling van het Onderwijsniveau (Wijnstra, 1988). De totale populatie van Nederlandse scholen werd, op basis van de schoolscore in het schooljaar 1989/1990, opgedeeld in drie strata van ongelijke omvang. In stratum 1 (schoolscore < 1.06) zijn vooral scholen vertegenwoordigd met relatief weinig arbeiderskinderen en relatief veel kinderen uit gezinnen met een hogere sociaal-economische status. Scholen in stratum 3 (schoolscore > 1.15) kenmerken zich door een relatief hoog aantal allochtone kinderen. De scholen in stratum 2 (schoolscore 1.06-1.15) vormen een middenpositie tussen beide andere strata. Per stratum werd een gelijk aantal scholen (N₆₀) getrokken. Het totaal aantal scholen per stratum werd vervolgens gelijkelijk verdeeld over zes zogenaamde afnamegroepen, elk met een omvang van circa 600 kinderen. Deze afnamegroep zijn zodanig samengesteld dat ze als representatief kunnen worden beschouwd voor de gehele steekproef. Aan iedere afnamegroep werd een deel van de toetsen voorgelegd. Het totale meetinstrumentarium was namelijk dermate omvangrijk dat het niet in zijn geheel door alle informanten gemaakt kon worden. In Tabel 1 is de verdeling van de lees- en spellinginstrumenten over de zes afnamegroepen af te lezen.

Tabel 1. Verdeling van de lees- en spellinginstrumenten over de zes afnamegroepen.

afnamegroep	1	2	3	4	5	6
decoderen	X			X	X	X
woordenschat	X			X	X	X
begrijpend lezen	X	X	X			
auditieve synthese	X					
grafemetoets	X					
spelling		X		X	X	
auditieve analyse					X	
fonemendictee					X	

Met het oog op een zo groot mogelijke betrouwbaarheid van gegevens is voor de beschrijving van de ontwikkeling van kennis en (deel-) vaardigheden steeds het maximale aantal leerlingen in de analyse betrokken. Dat wil zeggen alle leerlingen uit alle groepen waarin de betreffende toetsen werden afgenomen.

Voor de analyse van de interactie tussen kennis en (deel-) vaardigheden zijn de gegevens gebruikt van de afnamegroepen 1 en 5. In afnamegroep 1 zijn alle leestoetsen in combinatie met elkaar afgenomen, in afnamegroep 5 de taken die betrekking op spelling (-deel-) vaardigheden. Voor deze analyse konden alleen complete cases gebruikt worden, een criterium waaraan een groot aantal van de cases niet voldeed, omdat nogal wat leerlingen bij één van de afnames vanwege ziekte of om anderszins niet op school waren. We hebben geen redenen om aan te nemen dat hier sprake is van systematische uitval.

Om een dataset te krijgen met louter complete cases zijn de cases waarbij de eerste of de laatste afname op het gebied van decoderen, begrijpend lezen, spellen of woordenschat ontbraken uit die dataset verwijderd. Bij de cases die daarop resteerden zijn de ontbrekende resultaten met behulp van het door Little en Rubin (1987) ontwikkelde EM-algoritme geschat. Hierdoor bleven er in afnamegroep 1 317 cases (165 jongens en 152 meisjes) en in afnamegroep 5 363 complete cases (178 jongens en 185 meisjes) over.

In beide afnamegroepen zijn scholen uit stratum 2 en in mindere mate scholen uit stratum 3 enigszins oververtegenwoordigd. In oktober 1991 werd stratum 2 gevormd door 29.2% van de basisscholen. In afnamegroep 1 was 42% en in afnamegroep 5 37% van de deelnemende leerlingen afkomstig van een stratum 2-school. Van alle Nederlandse basisscholen behoorde eind 1991 20.4% tot stratum 3. In onze afnamegroepen bedroegen die percentages respectievelijk 23% en 26%. De groep allochtone leerlingen is in afnamegroep 1 enigszins ondervertegenwoordigd (10%) en in afnamegroep 5 enigszins oververtegenwoordigd (15%) in vergelijking met het landelijke beeld van oktober 1991. Toen bestond de populatie leerlingen in het Nederlandse onderwijs voor 11.4% uit allochtone kinderen.

Aan de deelnemende scholen is gevraagd welke onderwijsleermethode men ten behoeve van het leesonderwijs hanteert. Circa 90% van de leerlingen in beide afnamegroepen heeft aanvankelijk leesonderwijs genoten aan de hand van de methode *Veilig Leren Lezen*.

4.2 Instrumenten

In het onderzoek is gebruik gemaakt van meetinstrumenten uit het Leerlingvolgsysteem van het Cito.

Voor het meten van *decodeervaardigheid* werd de Drie-Minuten-Toets (Verhoeven, 1992c) ingezet. Deze toets bestaat uit drie leeskaarten, elk met specifieke woordtypen. Op Leeskaart 1 staan woorden van het type medeklinker-klinker, klinker-medeklinker en medeklinker-klinker-medeklinker; op Leeskaart 2 staan eenlettergrepige woorden met voor- of achteraan in het woord een cluster van medeklinkers; op Leeskaart 3 staan woorden met twee, drie of vier

lettergrepen. Elke leeskaart wordt door de leerling gedurende één minuut hardop gelezen. De toetsscore is het aantal correct gelezen woorden.

Om vast te stellen hoe goed en hoe snel kinderen *grafemen aan fonemen* kunnen koppelen is de Grafementoets gebruikt (Verhoeven, 1992c). Kinderen krijgen een kaart waarop alle grafemen staan afgedrukt die verwijzen naar de 34 fonemen die in het Nederlands voorkomen, met de opdracht deze te verklanken. De toetsscore bestaat uit twee delen: het aantal correct verklankte grafemen en de gebruikte leestijd in seconden. Bij de uitgevoerde variantie en LISREL-analyses is gewerkt met het quotiënt van het aantal correct verklankte grafemen en de leestijd.

De vaardigheid in *auditieve synthese* is gemeten met behulp van de Toets voor Auditieve Synthese (Verhoeven, 1992d). Aan de hand van 20 opgaven is nagegaan in hoeverre kinderen losse klanken kunnen samenvoegen tot woorden. De toetsscore is het aantal correct gereconstrueerde woorden.

De *spellingvaardigheid* is gemeten met behulp van de Schaal voor Spellingvaardigheid (Van den Bosch et al, 1991). De toets bestaat uit meerdere modules die ieder bestaan uit een reeks onveranderlijke één- en tweelettergrepige woorden. Deze zijn met behulp van een Item-Response model op één unidimensionele schaal zijn gebracht (Kamphuis & Moelands, 1990; Gillijns & Verhoeven, 1991; Gillijns & Moelands, 1992). De woorden worden in de context van een plaatje of een zin aangeboden, waarna de kinderen ze opschrijven. De toets wordt groepsgewijs afgenomen. De toetsscore, het aantal correct gespelde woorden, is omgezet in een vaardigheidsschatting.

Voor het bepalen van de vaardigheid in *auditieve analyse* is gebruik gemaakt van Toets voor Auditieve Analyse (Verhoeven, 1992d). In deze toets krijgen kinderen 20 woorden aangeboden met de opdracht deze foneem-voor-foneem te segmenteren. Het aantal correct gesegmenteerde woorden vormt de toetsscore.

Hoe goed kinderen *fonemen aan grafemen* kunnen koppelen, is gemeten met behulp van het Fonemendictee (Verhoeven, 1992d). De toets wordt groepsgewijs afgenomen, waarbij de leerkracht telkens een foneem dicteert nadat deze eerst als eerste klank in een woordcontext is aangeboden. In het opgavenboekje staat bovendien per item een plaatje afgedrukt dat naar dat woord verwijst. Het aantal correct geschreven fonemen is hier de toetsscore.

Voor het meten van de vaardigheid in *begrijpend lezen* zijn een tweetal toetsen gebruikt: de Schaal Betekenisrelaties en de Schaal Verwijsrelaties beide opgenomen in het pakket Lezen met Begrip (Verhoeven, 1992c). Beide worden groepsgewijs afgenomen en in beide gevallen zijn de opgaven met behulp van een Item-Response model op één unidimensionele schaal geplaatst. De Schaal Betekenisrelaties bevat korte tekstfragmenten met meerkeuzevragen naar zinsbetekenis en betekenisrelaties tussen zinnen. De Schaal Verwijsrelaties bevat teksten met meerkeuzevragen naar de relatie tussen de verwijswaarden in de tekst en hun antecedenten.

De *woordenschat* van kinderen is ten slotte is gepeild met een experimentele versie van

Tabel 2. Afnamemomenten van de meetinstrumenten.

	groep 3			groep 4		
	begin	medio	eind	begin	medio	eind
Drie-Minuten-Toets	x	x	x	x	x	x
Toets voor Auditieve synthese		x	x			
Grafementoets		x	x			
Spellingschaal		x	x		x	x
Toets voor Auditieve analyse		x	x			
Fonemendictee		x	x			
Toetsen Begrijpend Lezen			x		x	x
Woordenschattoets 1	x		x	x		x

Woordenschattoets 1 (Verhoeven, 1992f). De toets bevat 60 opgaven die elk bestaan uit vier tekeningen. Bij iedere opgave leest de proefleider een woord op en vraagt de kinderen de bijbehorende tekening te omcirkelen. De toetsscore is het totaal aantal correct aangegeven woordbetekenissen.

Tabel 2 laat zien op welke tijdstippen de toetsen zijn afgenomen.

4.3 Procedure

Ter beantwoording van de eerste onderzoeksvraag worden de scores van autochtone en allochtone leerlingen op (deel)vaardigheden decoderen, spelling, woordenschat en begrijpend lezen in de tijd vergeleken. Met behulp van variantie-analyse (MANOVA) met herhaalde metingen is nagegaan in hoeverre de hoofdeffecten 'onderwijsperiode' en 'herkomst' en hun interactie significant zijn. Op deze wijze is nagegaan of er sprake is van vooruitgang in de tijd, of er verschillen zijn tussen autochtone en allochtone leerlingen en of de verschillen tussen autochtonen en allochtonen in de loop van de tijd convergeren, divergeren, dan wel constant blijven.

Ter beantwoording van de onderzoeksvragen 2, 3 en 4 is met behulp van LISREL nagegaan hoe bij autochtone en allochtone leerlingen de ontwikkeling van decoderen, spelling en begrijpend lezen kan worden beschreven in termen van deelvaardigheden (vgl. Jöreskog & Sörbom, 1981; Van Leeuwe, 1984). In deze analyse zijn alleen de leerlingen betrokken van afnamegroep 1 en 5. De leerlingen van afnamegroep 1 maakten alle aan het leesproces gerelateerde toetsen, de kinderen uit afnamegroep 5 maakten alle aan spelling gerelateerde toetsen. Er is gekozen voor een LISREL-analyse, omdat bij die procedure wordt uitgegaan van een meetmodel waarbinnen zowel geobserveerde als latente variabelen kunnen worden gehanteerd, dat zowel cross-sectionele als longitudinale verbanden kunnen worden blootgelegd en dat de passendheid van structurele modellen met behulp van maximum-likelihood-analyse kan worden bepaald. Allereerst is nagegaan in hoeverre in groep 3 de ontwikkeling van decodeervaardigheid bij autochtone leerlingen kan worden beschreven vanuit de deelvaardigheden grafeem/foneem-koppeling en auditieve synthese. Daartoe is het best passende structurele model gezocht. In het bijbehorend meetmodel is de decodeervaardigheid gerepresenteerd als de factorscore van leerlingen op de drie leeskaarten van de Drie-Minuten-Toets. Vervolgens is de spellingvaardigheid van autochtone leerlingen gerelateerd aan de vaardigheden auditieve analyse en foneem/grafeem-koppeling, eveneens in jaargroep 3. Daarna is de vaardigheid in begrijpend lezen gerelateerd aan de decodeervaardigheid en de woordenschat van autochtone leerlingen. In het bijbehorend meetmodel is decodeervaardigheid wederom gerepresenteerd als de factorscore op de drie leeskaarten, terwijl voor begrijpend lezen is uitgegaan van de factorscore op de twee schaalscores. Ten slotte zijn bovenstaande LISREL-procedures in een multi-sample analyse met autochtone en allochtone leerlingen herhaald. Daarbij is in eerste instantie gepoogd het model dat bij de autochtone leerlingen werd geïdentificeerd ook te fitten op de populatie allochtone leerlingen. Indien dit niet lukte is gezocht naar afzonderlijke modellen voor beide populaties die tezamen een goede fit laten zien.

5 RESULTATEN

5.1 Ontwikkeling van decoderen

In Tabel 3 staan de gemiddelden en standaarddeviaties op de leeskaarten van de Drie-Minuten-Toets en op de onderdelen Grafementoets en de Toets voor Auditieve Synthese van alle autochtone en allochtone leerlingen die aan deze afnames hebben deelgenomen. Daarbij moet worden aangetekend dat aan de afname van de Drie-Minuten-Toets eind groep 3 om praktische redenen alleen aan de leerlingen in afnamegroep 1 konden deelnemen. Bij alle afnames van de Drie-Minuten-Toets gold verder een afbreekcriterium. Bij leerlingen die de eerste vijf woorden zeer langzaam en fout lezen is de afname gestaakt. Sommige leerlingen hebben derhalve niet op alle drie leeskaarten een toetsscore. Het zal duidelijk zijn dat dit met name bij de eerste afnames speelt.

Tabel 3. Gemiddelden, standaarddeviaties en aantal proefpersonen op de leeskaarten van de Drie-Minuten-Toets (DMT 1, DMT 2, DMT 3), de Grafemetoets (Toetscore=GRA, leestijd=GRtd) en de Toets voor Auditieve Synthese (TAS), uitgesplit naar autochtone (aut) en allochtone (all) leerlingen.

	begin 3			medio 3			eind 3		
	X	sd	N	X	sd	N	X	sd	N
DMT 1 aut	13.1	11.4	1812	29.7	13.5	1802	39.1	19.3	499
all	12.8	10.6	314	28.1	16.3	331	41.5	21.6	82
DMT 2 aut	6.5	8.8	1655	16.9	13.5	1802	24.0	17.0	499
all	7.3	7.8	286	16.2	11.8	306	25.0	16.1	79
DMT 3 aut	4.2	6.6	1582	11.3	10.2	1685	15.5	13.6	488
all	4.5	5.5	264	11.1	8.6	278	16.4	11.0	75
GRA aut				31.5	2.7	486	33.3	1.1	482
all				30.1	4.1	77	32.6	2.4	76
GRtd aut				48.1	20.9	224	35.1	11.5	223
all				47.6	18.2	49	34.2	8.9	49
TAS aut				18.0	2.9	481	19.4	1.1	477
all				17.2	4.0	77	19.2	1.1	68

	begin 4			medio 4			eind 4		
	X	sd	N	X	sd	N	X	sd	N
DMT 1 aut	50.9	20.6	1671	65.7	19.6	1617	71.8	19.3	1625
all	47.7	20.9	278	65.7	20.8	275	71.0	21.1	269
DMT 2 aut	37.5	20.9	1671	53.2	21.2	1617	59.9	21.3	1625
all	33.2	19.3	278	51.7	22.0	274	57.9	23.8	268
DMT 3 aut	25.9	16.3	1644	40.0	18.0	1610	46.0	18.7	1615
all	23.0	14.0	274	37.5	16.9	274	43.5	18.2	268

In Figuur 1 zijn de gemiddelde scores van autochtone en allochtone leerlingen op de leeskaarten van de Drie-Minuten-Toets grafisch weergegeven.

Er valt af te lezen dat de gemiddelde scores op Leeskaart 1 nagenoeg identiek zijn. Op de twee leeskaarten met meer complexe orthografische structuren is er een tendens dat de gemiddelde scores van allochtone leerlingen in groep 4 enigszins achterblijven bij die van autochtone leerlingen. Variantie-analyse (MANOVA) met periode en herkomst en leeskaart als hoofdeffecten laat zien dat de factor 'periode' significant is ($F(5,1455)=671.6$, $p<.001$) evenals de factor 'leeskaart' ($F(2,582)=1482.81$, $p<.001$), terwijl de factor 'herkomst' dat niet is. Verder is de interactie tussen de factoren 'herkomst' en 'periode' en tussen de factoren 'herkomst', 'periode' en 'leeskaart' niet significant. De interactie daarentegen tussen 'leeskaart' en 'periode' ($F(10,2910)=71.56$, $p<.01$), alsmede die tussen 'leeskaart' en 'herkomst' ($F(2,582)=5.59$, $p<.001$) is dat wel. Eerstgenoemde interactie wijst erop dat de scores op de drie leeskaarten niet in dezelfde mate toenemen: kinderen gaan het snelst vooruit op de eerste leeskaart en het traagst op de derde leeskaart. De interactie tussen leeskaart en herkomst geeft aan dat toch sprake is van een zeker verschil in prestaties op de drie leeskaarten. Naarmate woorden langer worden blijken allochtone leerlingen ietwat minder te presteren in vergelijking met hun autochtone leeftijdgenoten.

Uit tabel 3 is af te lezen dat de gemiddelde scores van allochtone leerlingen op de Grafemetoets ietwat achterblijven bij die van autochtone leerlingen, hoewel in de loop van de tijd de verschillen kleiner worden. Variantie-analyse op het quotiënt van het aantal correct verklankte

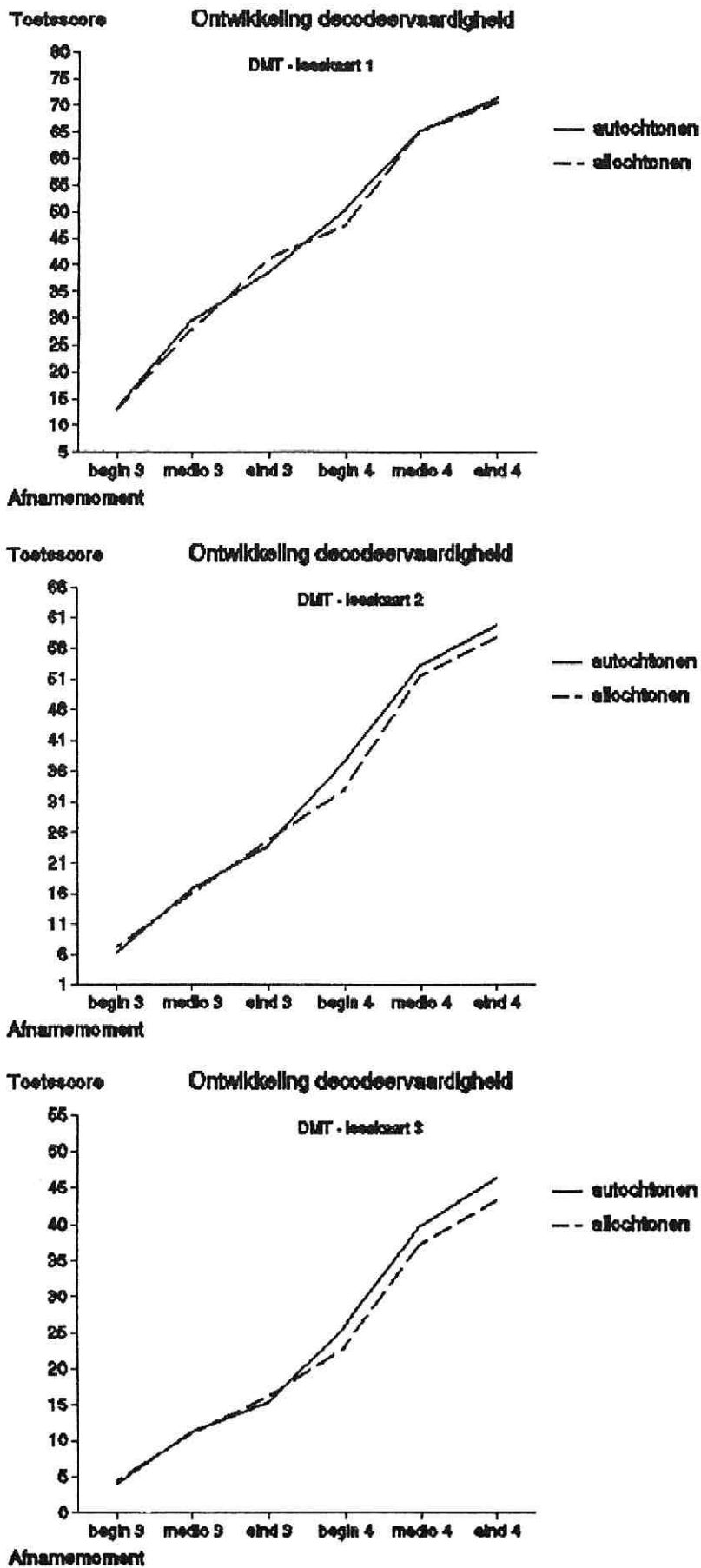


Fig. 1. Gemiddelde scores op de Drie-Minuten-Toets bij allochtone en autochtone leerlingen.

grafemen en de leestijd geeft hiervoor evidentie: De hoofdeffecten 'periode' ($F(1,290)=191.60$, $p<.001$) en 'herkomst' ($F(1,290)=1979.81$, $p<.01$) zijn beide significant, evenals de interactie tussen beide effecten ($F(1,290)=20.63$, $p<.001$).

Bij de Toets voor Auditieve Synthese blijkt sprake van geringe verschillen in groepsgemiddelden tussen autochtone en allochtone leerlingen, zeker bij de afname eind groep 3. Variantie-analyse laat zien dat de factor 'periode' significant ($F(1,524)=80.57$, $p<.001$) is, terwijl de factor 'herkomst' tendeeft naar significantie ($p<.10$). De interactie tussen beide factoren blijkt niet significant

5.2 Ontwikkeling van spelling

Tabel 4 geeft de gemiddelden en standaarddeviaties op de Spellingtaak en op de onderdelen Fonemendictee en Toets voor Auditieve Analyse van de totale steekproef van leerlingen.

De gemiddelde schaalscores voor spelling van autochtone en allochtone leerlingen worden in Figuur 2 grafisch weergegeven.

Uit Figuur 2 valt af te lezen dat allochtone leerlingen aanzienlijk meer problemen hebben met spelling dan hun autochtone leeftijdsgenoten. Het lijkt erop dat de verschillen medio groep 4 nivelleren, terwijl de verschillen eind groep 4 weer neigen te divergeren. MANOVA laat zien dat de factoren 'periode' ($F(3,2409)=609.68$, $p<.001$) en 'herkomst' ($F(1,803)=47.57$, $p<.001$) significant zijn, maar hun interactie niet.

Allochtone leerlingen blijken minder vaardig te zijn in auditieve analyse dan hun autochtone leeftijdsgenoten, zo blijkt uit tabel 4. De verschillen in de loop van groep 3 worden niet veel kleiner. Variantie-analyse ondersteunt dit beeld. De hoofdeffecten 'periode' ($F(1,501)=139.52$, $p<.001$) en 'herkomst' ($F(1,501)=14.94$, $p<.001$) zijn beide significant, terwijl hun interactie dat niet is.

Ook voor wat de foneemkennis betreft blijken allochtonen lager uit te komen dan autochtonen. MANOVA laat zien dat de hoofdeffecten 'periode' ($F(1,498)=286.84$, $p<.001$) en 'herkomst' ($F(1,498)=24.17$, $p<.001$) allebei significant zijn, evenals hun interactie ($F(1,498)=73.78$, $p<.001$). De significante interactie tussen beide effecten laat zien dat allochtone leerlingen wat betreft foneemkennis in de loop van groep 3 inlopen op hun autochtone leeftijdsgenoten.

Tabel 4. Gemiddelden, standaarddeviaties en aantal proefpersonen op de Schaal Vorderingen in Spellingvaardigheid (SVS), de Toets voor Auditieve Analyse (TAA) en het Fonemendictee (FD), uitgesplit naar autochtone (aut) en allochtone (all) leerlingen.

		medio 3			eind 3		
		X	sd	N	X	sd	N
SVS	aut	104.5	11.4	959	110.4	7.9	934
	all	100.2	12.5	195	106.1	7.6	145
TAA	aut	15.0	5.1	433	18.4	3.0	440
	all	13.6	5.1	93	16.6	3.6	92
FD	aut	29.5	4.0	431	31.9	2.4	440
	all	27.3	4.9	91	30.8	3.0	93
		medio 4			eind 4		
		X	sd	N	X	sd	N
SVS	aut	117.6	7.4	868	122.5	8.2	848
	all	114.9	8.0	187	117.5	7.9	162

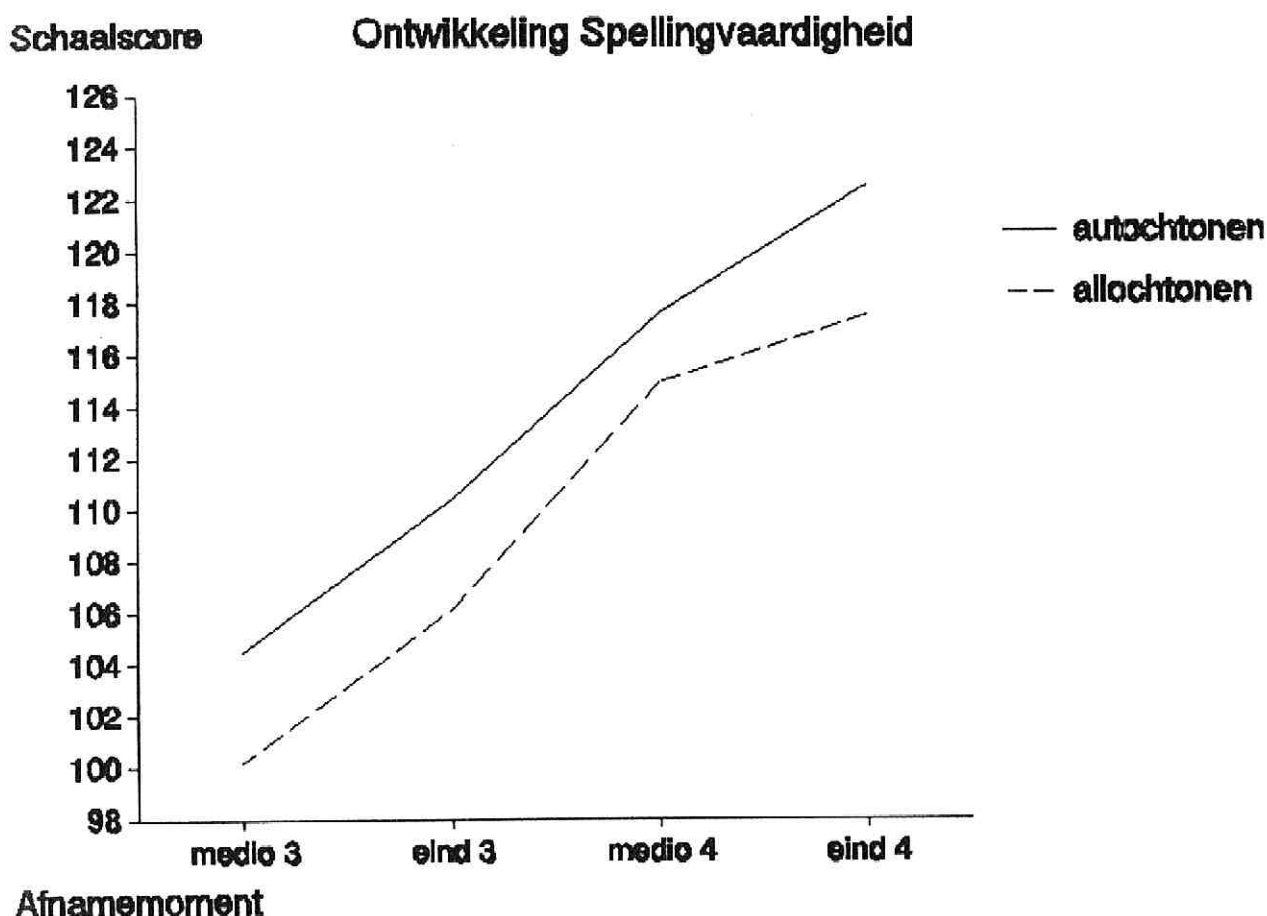


Fig. 2. Gemiddelde scores op de Schaal Vorderingen in Spellingvaardigheid bij allochtone en autochtone leerlingen.

5.3 Ontwikkeling van woordenschat

In Tabel 5 staan de gemiddelden en standaarddeviaties op de Woordenschattoets van de totale steekproef van leerlingen vermeld. In Figuur 3 worden de gemiddelde scores van autochtone en allochtone leerlingen grafisch weergegeven.

Er blijkt sprake van grote verschillen tussen allochtone en autochtone leerlingen. De factor 'herkomst' blijkt significant ($F(1,1519)=940,68$, $p<.001$), evenals de factor 'periode'

Tabel 5. Gemiddelden, standaarddeviaties en aantal proefpersonen op de Woordenschattoets, uitgesplit naar autochtone (aut) en allochtone (all) leerlingen.

		Woordenschattoets		
		X	sd	N
Medio 3	aut	45.1	6.5	1811
	all	32.1	9.3	320
Eind 3	aut	49.3	5.8	1828
	all	36.4	9.1	309
Medio 4	aut	52.3	4.9	1634
	all	41.1	8.4	286
Eind 4	aut	54.1	4.6	1594
	all	43.3	7.5	278

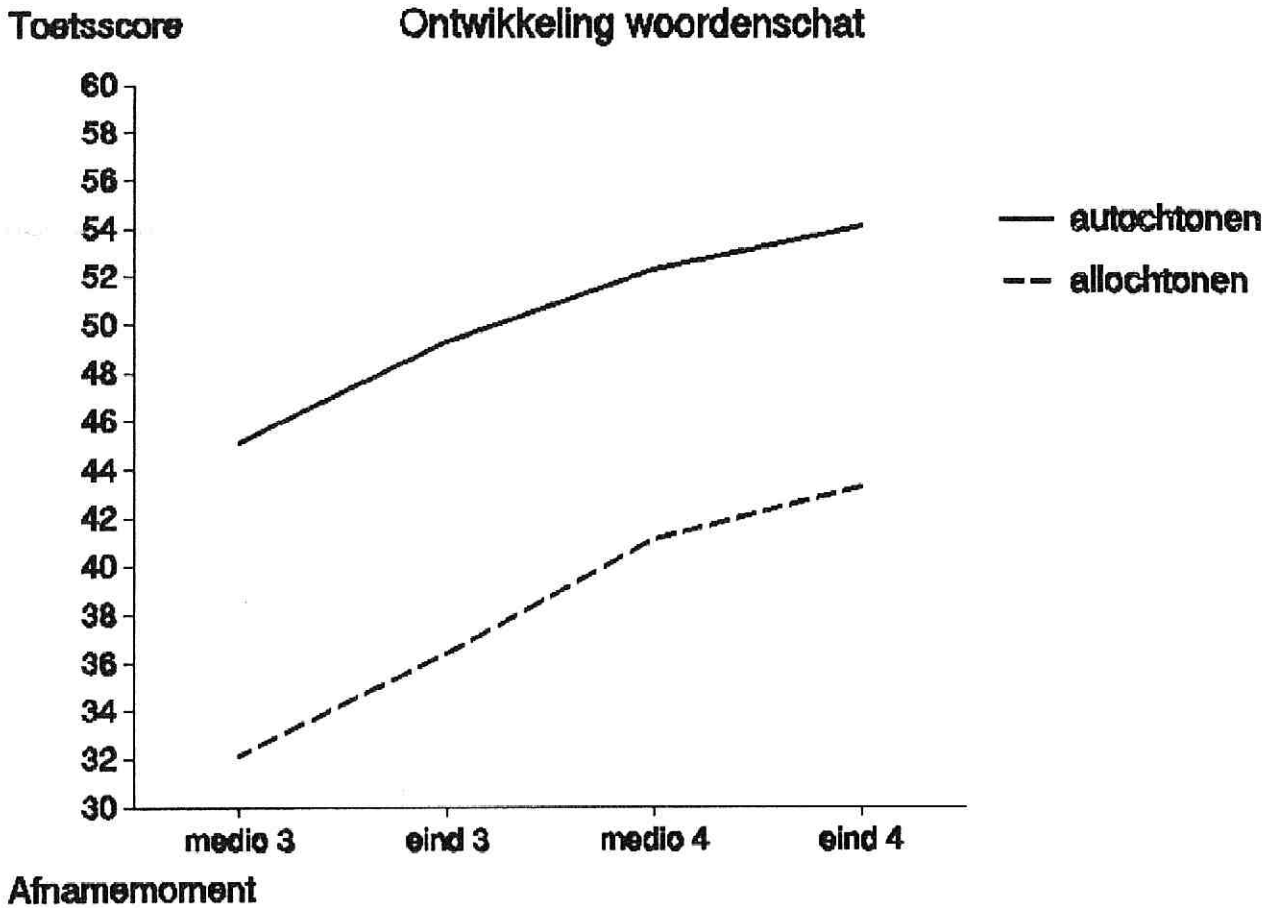


Fig. 3. Gemiddelde scores op de Woordenschattoets bij allochtone en autochtone leerlingen

($F(3,4557)=1229.07$, $p<.001$). Hetzelfde geldt voor de interactie tussen beide factoren ($F(3,4557)=21.00$, $p<.001$), hetgeen erop lijkt te wijzen dat de allochtone leerlingen wat hun Nederlandse woordenschat betreft in de loop van de tijd ietwat op hun autochtone leeftijdsgenoten inlopen. Het licht nivellerende effect kan echter ook worden verklaard vanuit een plafondefect voor de autochtone groep.

5.4 Ontwikkeling van begrijpend lezen

Tabel 6 geeft de gemiddelden en standaarddeviaties op de Schaal Betekenisrelaties en de Schaal Verwijsrelaties van de totale steekproef van leerlingen. De gemiddelde scores van autochtone en allochtone leerlingen worden vervolgens in Figuur 4 weergegeven.

Tabel 6. Gemiddelden, standaarddeviaties en proefpersonen op de Schaal Betekenisrelaties (SBR) en de Schaal Verwijsrelaties (SVR), uitgesplitst naar autochtone (aut) en allochtone (all) leerlingen.

		SBR			SVR		
		X	sd	N	X	sd	N
Eind 3	aut	93.6	8.3	1359	99.7	14.1	469
	all	89.4	7.2	202	89.7	10.1	70
Medio 4	aut	99.9	6.2	1277	111.9	11.3	413
	all	96.0	5.8	177	105.2	10.2	54
Eind 4	aut	102.5	4.7	1301	117.0	12.9	413
	all	99.8	3.9	169	109.5	10.0	52

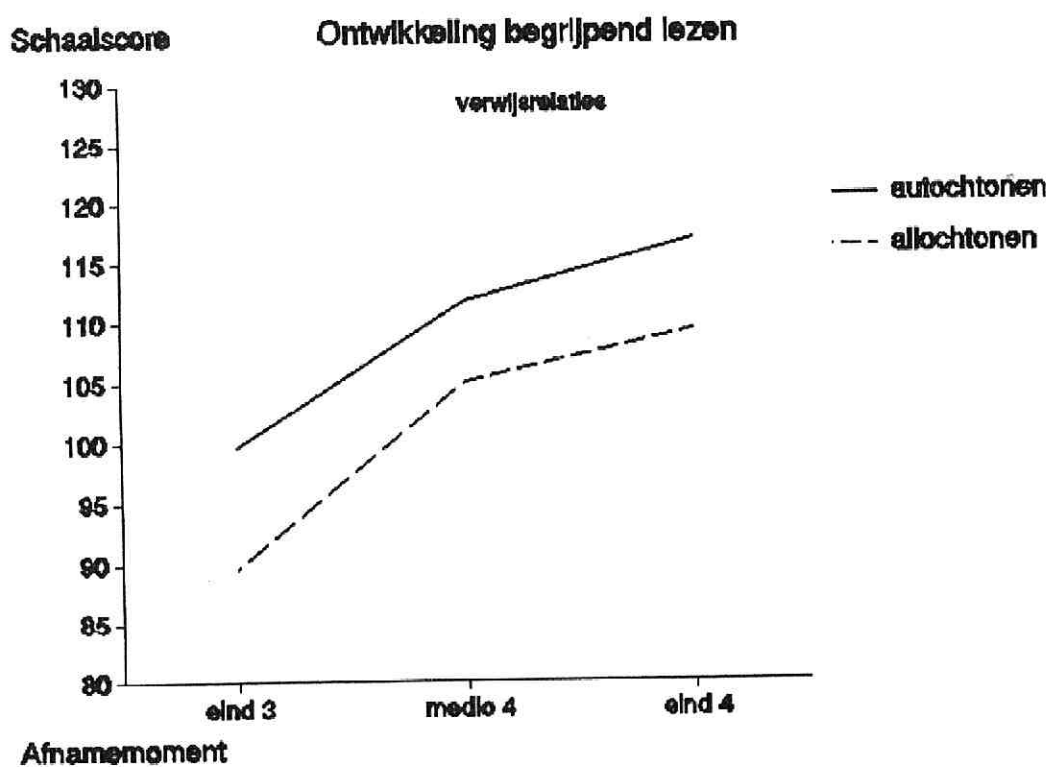
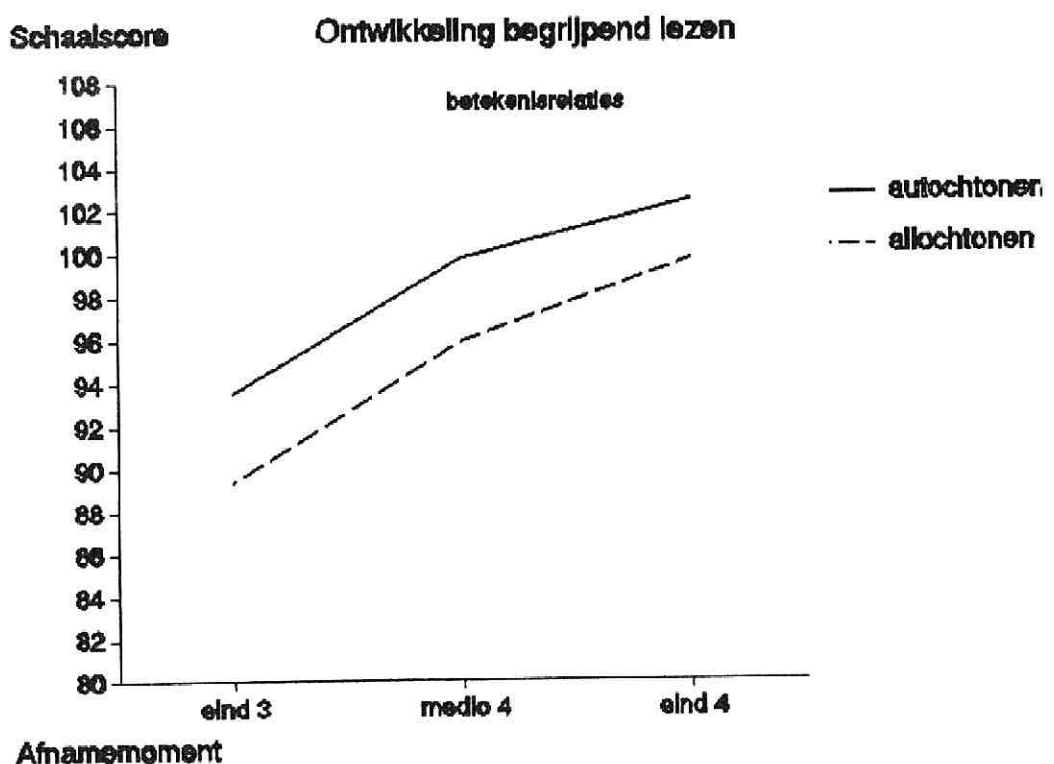


Fig. 4. Gemiddelde scores op de Schaal Betekenisrelaties en de Schaal Verwijsrelaties bij allochtone en autochtone leerlingen.

Op het onderdeel Betekenisrelaties blijkt tot medio groep 4 sprake van grote verschillen tussen allochtone en autochtone leerlingen. Eind groep 4 lijkt sprake te zijn van enige nivellering. Deze nivellering kan echter worden toegeschreven aan plafondeffecten op deze taak. MANOVA laat zien dat de hoofdeffecten 'herkomst' ($F(1,224)=9.74$, $p<.01$) en 'periode'

($F(2,448)=21.63$, $p<.001$) significant zijn, evenals de interactie tussen beide effecten ($F(2,448)=7.28$, $p<.01$).

Wat het onderdeel Verwijsrelaties betreft valt te constateren dat de substantiële verschillen in de prestaties van allochtone en autochtone leerlingen constant blijven in de tijd. MANOVA ondersteunt dit beeld: De hoofdeffecten 'herkomst' ($F(1,394)=23.81$, $p<.001$) en 'periode' ($F(2,788)=200.33$, $p<.001$) zijn beide significant, terwijl hun interactie dat niet is.

5.5 Interactie tussen kennis en (deel)vaardigheden

In Figuur 5 is voor de autochtone leerlingen in afnamegroep 5 het best passende structurele LISREL-model weergegeven voor decodeervaardigheid. Daarbij is de decodeervaardigheid in groep 3 gerelateerd aan hun vaardigheid in grafeem-foneem-koppeling en auditieve synthese.

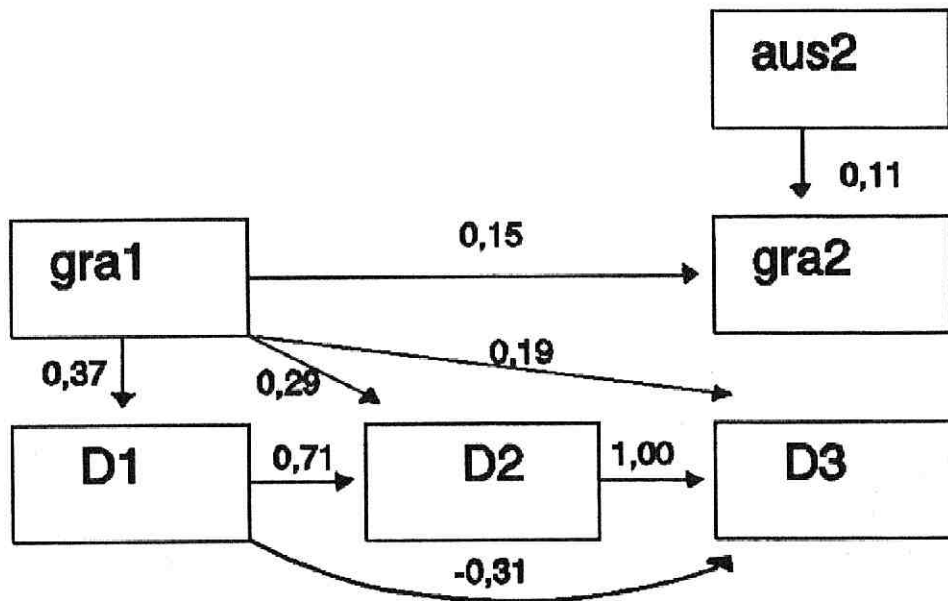
De modelfit is goed. De goodness-of-fit (GOF) bleek .98, terwijl toetsing van het model leidde tot $X^2(9)=19.83$, $p=.02$. In een LISREL multi-sample analyse met autochtone en allochtone leerlingen bleek bovenstaand model ook de beste fit te laten zien: $GOF=.83$; $X^2(37)=50.21$; $p=.07$. Een exclusief LISREL-model voor allochtone leerlingen resulteerde in $GOF=.82$; $X^2(24)=48.24$, $p<.001$.

Uit het model in figuur 5 valt af te lezen dat de grafeemkennis een belangrijke predictor vormt voor decodeervaardigheid. De vaardigheid in auditieve synthese blijkt in deze fase voor decodeervaardigheid geen predictor van betekenis, wat vermoedelijk te wijten is aan het plafondeffect op de Toets voor Auditieve Synthese.

Figuur 6 beschrijft het best passende structurele LISREL-model gegeven voor spellingvaardigheid bij autochtone leerlingen. Daarbij is in groep 3 de spellingvaardigheid gerelateerd aan de vaardigheid in auditieve analyse en foneem/grafeem-koppeling.

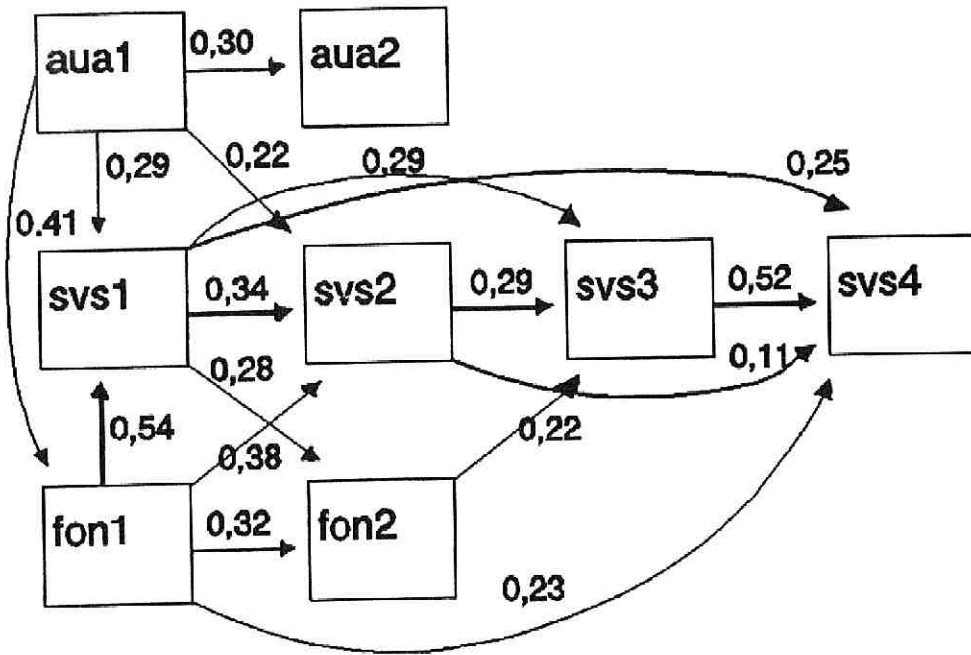
De fit van het model lijkt op basis van de verhouding van het Chi-kwadraat en het aantal vrijheidsgraden redelijk te noemen: $GOF=.99$; $X^2(13)=17.92$, $p=.16$.

In een multi-sample analyse met autochtone en allochtone leerlingen heeft bovenstaand model ook een acceptabele fit: $GOF=.82$; $X^2(49)=66.0$; $p=.06$. Aanpassing van de indices in het



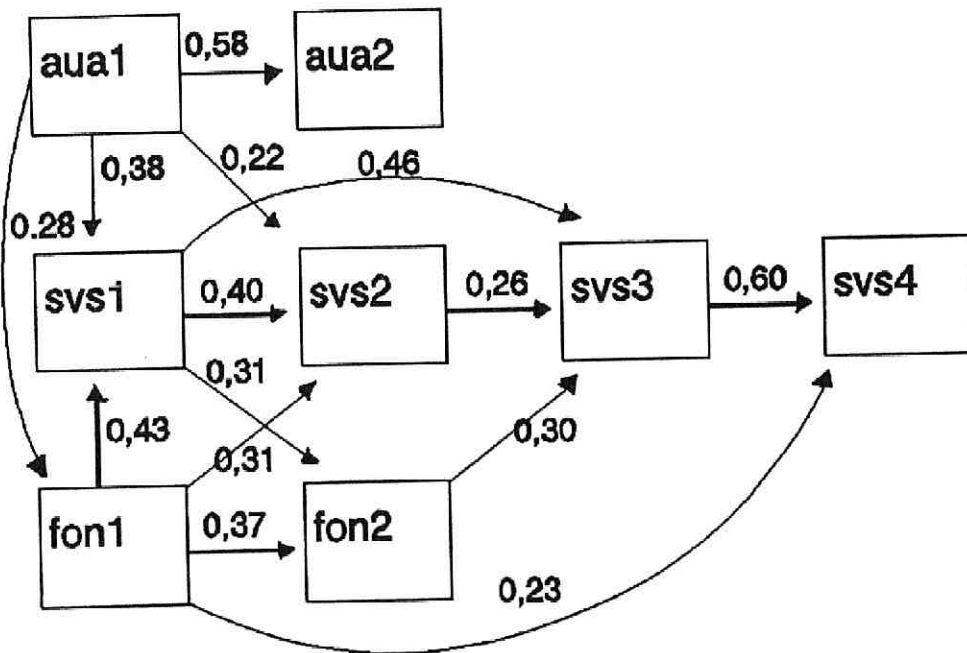
D = decoderen
 aus = auditieve synthese
 gra = grafeem-foneem-koppeling

Fig. 5. LISREL-model voor decodeervaardigheid bij autochtone en allochtone leerlingen in afnamegroep 1.



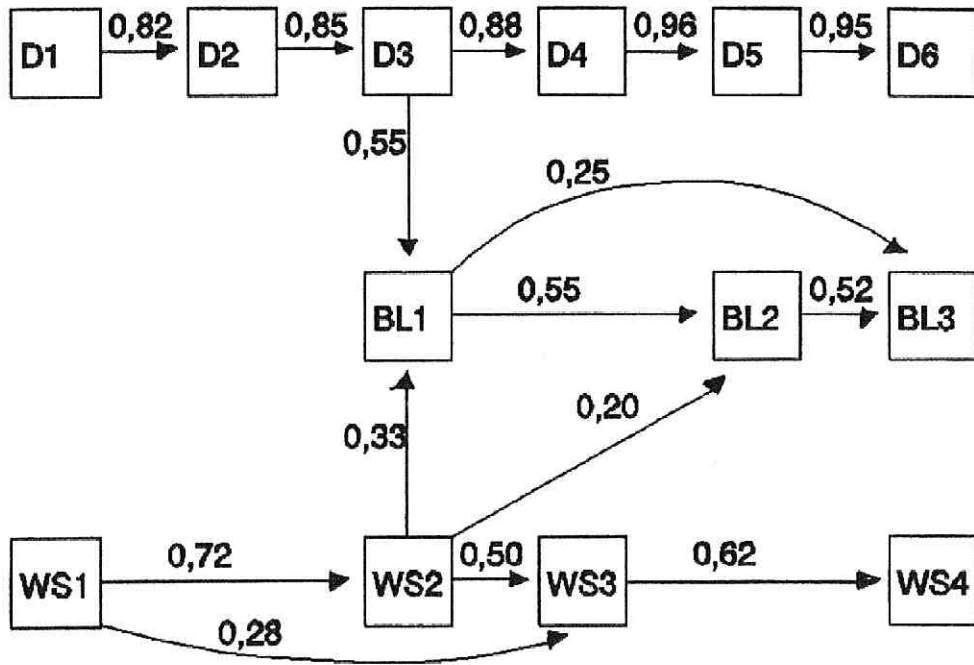
svs = spelling
 aua = auditieve analyse
 fon = foneem-grafeem-koppeling

Fig. 6. LISREL-model voor spellingvaardigheid bij de autochtone leerlingen in afnamegroep 5



svs = spelling
 aua = auditieve analyse
 fon = foneem-grafeem-koppeling

Fig. 7. LISREL-model voor spellingvaardigheid bij de allochtone leerlingen in afnamegroep 5



BL = begrijpend lezen
 D = decoderen
 WS = woordenschat

Fig. 8. LISREL-model voor begrijpend lezen bij de autochtone leerlingen in afnamegroep 1.

model voor allochtone leerlingen resulteert echter in een aanzienlijke verbetering van de fit: $GOF=.97$; $X^2(27)=26.03$, $p=.52$. Figuur 7 toont het aangepaste structurele LISREL-model voor de spellingvaardigheid van allochtone leerlingen.

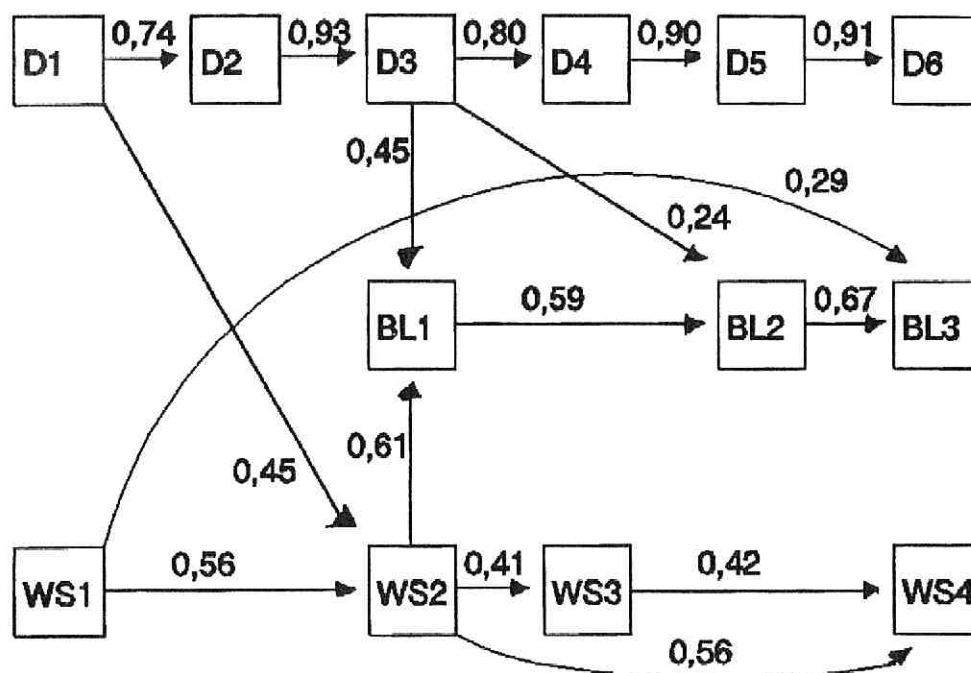
Beide modellen laten zien dat sprake is van een interactie binnen het op gang komen van foneemkennis en spellingvaardigheid: De relatie tussen de ontwikkeling van spellingvaardigheid en foneemkennis in de loop van groep 3 en 4 is wederzijds te noemen. Verder blijkt de vaardigheid in auditieve analyse de initiële ontwikkeling van spellingvaardigheid te beïnvloeden en tevens het op gang komen van foneemkennis te voorspellen. Wel lijkt de directe voorspelling van spellingvaardigheid vanuit de vaardigheid in auditieve analyse bij allochtone leerlingen sterker dan bij autochtone leerlingen, terwijl de invloed van foneemkennis bij allochtone leerlingen iets geringer lijkt in vergelijking met hun autochtone leeftijdsgenoten.

In Figuur 8 wordt het best passende structurele LISREL-model gepresenteerd voor autochtone leerlingen bij begrijpend lezen. In dit geval is de vaardigheid in groep 3 en 4 in begrijpend lezen gerelateerd aan de decodeervaardigheid en de woordenschat.

Hoewel de p -waarde laag is, valt de verhouding tussen het aantal vrijheidsgraden en het Chi-kwadraat gunstig uit. De fit van het model mag redelijk genoemd worden: $GOF=.94$; $X^2(54)=117.36$, $p<.001$.

Het model laat zien dat de onderdelen decodeervaardigheid en woordenschat in de loop van de tijd gezien als uiterst stabiele vaardigheden kunnen worden aangemerkt. Verder blijkt de vaardigheid in begrijpend lezen in iets hogere mate te worden voorspeld door de decodeervaardigheid dan door de woordenschat van leerlingen.

Een multi-sample analyse met hetzelfde model, maar met autochtone én allochtone leerlingen levert een minder goed resultaat op: $GOF=.62$; $X^2(145)=302.13$; $p<.001$. Aanpassing in de indices in het model voor allochtone leerlingen leidt in ieder geval tot een betere fit-maat: $GOF=.75$; $X^2(114)=228.69$, $p<.001$. Figuur 9 geeft het aangepaste model weer. In vergelijking



BL = begrijpend lezen
 D = decoderen
 WS = woordenschat

Fig. 9. LISREL-model voor begrijpend lezen bij de allochtone leerlingen in afnamegroep 1

met de autochtone leerlingen valt de grotere rol op van woordenschat als voorspeller van de vaardigheid in begrijpend lezen.

6 CONCLUSIES EN DISCUSSIE

Uit het onderhavige onderzoek kan een aantal conclusies worden getrokken. In de eerste plaats blijkt in de jaargroepen 3 en 4 sprake van een sterke groei in (deel)vaardigheden op het terrein van woordenschat, decoderen, begrijpend lezen en spellen. Daarbij blijkt het beeld voor autochtone en allochtone leerlingen op een aantal aspecten gedifferentieerd. Opvallend is dat dit niet geldt voor de ontwikkeling van decodeervaardigheid.

Wat decodeervaardigheid betreft is het prestatieniveau van beide groepen nagenoeg gelijk. Alleen bij het lezen van gelede woorden (Leeskaart 3 van de Drie-Minuten-Toets) blijken allochtonen in de loop van de tijd ietwat lager uit te komen dan hun autochtone leeftijdsgenoten. De verschillen zijn echter uitermate gering te noemen. Ook bij de deelvaardigheden Auditieve Synthese en Grafeemkennis blijken de verschillen tussen autochtone en allochtone leerlingen gering. Op het eind van groep 3 is zelfs sprake van volledige nivellering. Blijkbaar voldoet een relatief geringe beheersing van de Nederlandse taal voor het op gang brengen van metalinguïstisch bewustzijn in die taal en voor het totstandkomen van visuele woordrepresentaties. Dit resultaat is sterk afwijkend van de bevindingen uit eerder onderzoek van Verhoeven (1987; 1990a/b) waar geconcludeerd werd dat de decodeervaardigheid van allochtone leerlingen in de jaargroepen 3 en 4 achterblijft bij die van autochtone leeftijdsgenoten. Als mogelijke verklaring voor deze discrepantie geldt dat de beheersing van de Nederlandse taal bij de kinderen in het onderhavige onderzoek als gevolg van een langere verblijfsduur in Nederland waarschijnlijk

beduidend hoger is dan die van de leerlingen uit het eerder genoemde onderzoek. Gezien het feit dat in beide onderzoeken verschillende meetinstrumenten zijn gehanteerd valt deze veronderstelling niet te toetsen. Een tweede mogelijke verklaring vormt het feit dat de onderwijspraktijk inmiddels beter is ingespeeld op de didactiek van het aanvankelijk lezen in het Nederlands als tweede taal.

In spellingvaardigheid blijkt er tussen autochtone en allochtone leerlingen sprake te zijn van verschillen die constant blijven in de tijd. Allochtone kinderen blijken ten opzichte van hun autochtone leeftijdsgenoten relatief veel moeite te hebben met de deelvaardigheden Auditieve Analyse en Foneemkennis. Te concluderen valt dat voor tweede-taal-leerders het *coderen* van woorden aanmerkelijk lastiger is dan het *decoderen* van woorden.

Allochtone leerlingen blijken ook wat hun woordenschat en begrip- leesvaardigheid betreft beduidend lager uit te komen dan autochtone leerlingen. In de loop van de jaargroepen 3 en 4 blijven de verschillen bij zowel *woordenschat* als *begrip- lezen* nagenoeg constant. Dit resultaat komt overeen met uitkomsten uit eerdere studies van Verhoeven (1987, 1990a/b, 1992b), Verhoeven en Vermeer (1992), Teunissen (1986) en Hacquebord (1989).

Een tweede belangrijk resultaat is dat de structuur van de taalcompetentie van eerste- en tweede-taalleerders opmerkelijke overeenkomsten vertoont. LISREL-componenten-analyse laat zien dat sprake is van een organisatie van decodeervaardigheid in twee deelcomponenten: auditieve synthese en grafeem-foneem-koppeling. Opmerkelijk is dat auditieve synthese bij allochtone en autochtone leerlingen mede de (deel)vaardigheid in grafeem-foneem-koppeling voorspelt. Dit resultaat valt te verklaren vanuit het feit dat de vaardigheid in auditieve synthese als onderdeel van metalinguïstisch bewustzijn kan worden gezien. Het lijkt redelijk te veronderstellen dat een zekere mate van metalinguïstisch bewustzijn nodig is voor het op gang brengen van koppelingen tussen grafemen en fonemen.

Voor spellingvaardigheid blijken auditieve analyse en foneem-grafeem-koppeling de belangrijke deelprocessen te vormen. Ook hier blijkt dat de vaardigheid in auditieve analyse, als onderdeel van metalinguïstisch bewustzijn, de deelvaardigheid in foneem-grafeem-koppeling voorspelt. Bij autochtone leerlingen blijkt dit in sterkere mate het geval dan bij allochtone leerlingen. Wellicht dat de vaardigheid in foneem-grafeem-koppeling bij allochtonen in grotere mate worden bepaald door de mate van fonologische kennis van het Nederlands als tweede taal, waarover zij beschikken.

LISREL-analyse laat verder zien dat het leersucces in begrip- lezen bij zowel allochtone als autochtone leerlingen wordt bepaald door zowel decodeersnelheid als woordenschat. Bij allochtone leerlingen blijkt de factor woordenschat echter relatief van groter belang.

In het licht van de tot dusverre gevonden resultaten is de vraag van belang hoe het lees- en spellinggedrag van met name de allochtone leerlingen in de onderbouw van het basisonderwijs kan worden geoptimaliseerd. Het onderhavige onderzoek laat zien dat allochtone leerlingen wat betreft decoderen vergelijkbare resultaten boeken als autochtone leerlingen. Voor spelling, woordenschat en begrip- lezen is dit echter niet het geval. Een specifieke didactische aanpak van allochtone leerlingen op deze onderdelen is daarom dringend gewenst. Ten aanzien van spelling is er behoefte aan aanvullende oefenstof, gericht op het analyseren van auditieve informatie en het inslijpen van letters en spellingpatronen.

Met het oog op woordenschatontwikkeling zijn er twee mogelijkheden voor stimulering: het aanbieden van woorden in context en het geven van directe instructie (vgl. Verhoeven & Van Kuyk, 1991). Algemeen wordt inmiddels aangenomen (cf. Schouten-Van Parreren 1985; McKeown & Curtis 1987) dat de meeste woorden worden geleerd in context, maar dat woordbetekenissen het beste worden onthouden via directe instructie (cf. Nagy & Herman 1987; Pressley, Levin & McDaniel 1987). Een gecombineerde aanpak lijkt de beste resultaten te bieden (Sternberg, 1987).

Ten slotte kan de vaardigheid in begrip- lezen worden gestimuleerd door het verbeteren van strategieën met betrekking tot het analyseren van de tekststructuur (vgl. Aarnoutse, 1982, 1992). Daarbij gaat het om strategieën als het afleiden van de hoofdgedachte uit een tekst (Rinehart, Stahl & Erickson, 1986; Baumann, 1984), het doen van inferenties (Hansen, 1981;

Paris & Oka, 1986) en het oplossen van anaforische relaties in de tekst (Baumann, 1986). Enkele studies laten zien dat dergelijke strategieën voor het begrijpend lezen in een tweede taal uiterst relevant zijn (Barnitz, 1979; Barnitz, 1986, Carrell, 1984; Weber, 1991). Systematische effectstudies met betrekking tot het aanleren van dit soort leesstrategieën bij tweede-taalleerders zijn echter nog niet eerder uitgevoerd. Gegeven het feit dat allochtone leerlingen veel minder gelegenheid hebben gehad zich de linguïstische kennis eigen te maken van de doeltaal die in teksten wordt verondersteld, mag worden verwacht dat het aanleren van adequate leesstrategieën in die taal daardoor voor hen extra moeilijk is (vgl. Weber, 1991).

Naast gerichte instructie in een gestructureerd curriculum zijn goede en systematische evaluatieprocedures onontbeerlijk om de ontwikkeling van lezen en spellen die kinderen in deze leeftijdsfase doormaken op de voet te volgen (Janssens, 1986). Op die manier kan de leerkracht de informatie verzamelen die noodzakelijk is voor het optimaliseren van het didactische handelen ten behoeve van zowel autochtone als allochtone leerlingen. Het is bekend dat dergelijke evaluatiepraktijken een positieve invloed hebben op de leervorderingen (vgl. Brandsma & Knuver, 1991; Scheerens, 1990). Het Leerlingvolgsysteem van het Cito is bedoeld als aanzet voor zo'n systematische aanpak van de onderwijsevaluatie. Op het gebied van het beginnend lezen en spellen bevat het meetinstrumenten die de ontwikkeling van kinderen goed in kaart kunnen brengen. In die zin is het een geschikt hulpmiddel voor het begeleiden van leerlingen naar functionele geletterdheid.

LITERATUUR

- Aarnoutse, C. (1982). *Aspecten van Begrijpend Lezen in het Vierde Leerjaar van het GLO*. Dissertatie. Nijmegen: KUN.
- Aarnoutse, C., Mommers, M., Smits, B. & Leeuwe, J. van (1986). De ontwikkeling en samenhang van technisch lezen, begrijpend lezen en spellen. *Pedagogische Studiën*, 63, 97-110.
- Aarnoutse, C. & Leeuwe, J. van (1988). Het belang van technisch lezen, woordenschat en ruimtelijke intelligentie voor begrijpend lezen. *Pedagogische Studiën*, 65, 2, 49-59.
- Aarnoutse, C. (1992). Tekstgericht onderwijs in begrijpend lezen. In L. Verhoeven (red.), *Handboek lees- en schrijfdidactiek*. Lisse: Swets & Zeitlinger.
- Adams, M.J. *Beginning to read*. Cambridge, MA: MIT Press.
- Anderson, R.C. & Freebody, P. (1981). Vocabulary knowledge. In J.T. Guthrie (ed.), *Comprehension and teaching: Research reviews*. Newark: International Reading Association.
- Barnitz, J. G. (1979). *Reading comprehension of pronoun-referent structures in grades two, four and six*. Illinois: Urbana Campaign.
- Barnitz (1986). Towards understanding the effects of cross-cultural schemata and discourse structure on second language comprehension. *Journal of Reading Behavior*, 18, 95-116.
- Baumann, J. F. (1984). The effectiveness of a direct instruction paradigm for teaching main idea comprehension. *Reading Research Quarterly*, 20, 93-115.
- Baumann, J. F. (1986). Teaching third grade students to comprehend anaphoric relationships. *Reading Research Quarterly*, 21, 70-90.
- Beck, I.L., Perfetti, C.A. & McKeown, M.G. (1982). The effects of longterm vocabulary instruction on lexical access and reading comprehension. *Journal of Educational Psychology*, 74, 506-521.
- Bos, K. van den & Lutje Spelberg, H. (1993). Reading comprehension and related skills in nine-year-old normal and poor readers. In R.M. Joshi & C.K. Leong (Eds.), *Reading disabilities: Diagnosis and component processes*. Dordrecht: Kluwer.
- Bosch, L. van den, Gillijns, P., Krom, R. & Moelands F. (1991). *Schaal Vorderingen in Spellingvaardigheid 1*. Arnhem: Cito.
- Brandsma, H. & Knuver, A. (1991). De contextgebondenheid van effectiviteitsbevorderende schoolkenmerken. *Tijdschrift voor Onderwijsresearch*, 16 (4), 219-230.
- Carrell, P. L. (1984). The effects of rhetorical organization on ESL readers. *TESOL Quarterly*, 18, 441-469.
- Damhuis, R., Gloppe, K. de & Schooten, E. van (1989). Leesvaardigheid in het Nederlands van allochtone en autochtone leerlingen in groep 3 van het basisonderwijs. *Pedagogische Studiën*, 66, 4, 158-171.
- Elley, W. (1992). *How in the world do children read?* Hamburg: IEA.
- Dongen, A. van (1984). *Leesmoeilijkheden*. Tilburg: Zwijzen.

- Gillijns, P. & Moelands, F. (1992). Toepassing van item-response-theorie ten behoeve van een leerlingvolgsysteem. *Tijdschrift voor Orthopedagogiek*, 31 (4), 162-176.
- Gillijns, P. & Verhoeven, L. (1991). Naar een leerlingvolgsysteem voor het basisonderwijs. *Pedagogische Studiën*, 68, 1-15.
- Gillijns, P. & Verhoeven, L. (1992). Het Cito-Leerlingvolgsysteem: met het oog op de praktijk. *Pedagogische Studiën*, 69, 291-296.
- Hansen, J. (1981). The effects of inference training and practice on young children's reading comprehension. *Reading Research Quarterly*, 16, 391-417.
- Hacquebord, H. (1989). *Tekstbegrip van Turkse en Nederlandse Leerlingen in het Voortgezet Onderwijs*. Dissertatie. Dordrecht: Foris.
- Hauptman, P. C. (1981). A comparison of first and second reading strategies. *ITL*, 51, 37-57.
- Janssens, F.J.G. (1986). *De evaluatiepraktijken van leerkrachten. Een beschrijvend onderzoek naar het evalueren tijdens het rekenen*. Dissertatie. Arnhem: Cito.
- Jöreskog, K.G. & Sörbom, D. (1981) *LISREL V: Analysis of linear structural relationships by maximum likelihood and least square methods*. Chicago.
- Kamphuis, F.H. & Moelands, A.H.J. (1990). Longitudinaal meten van individuele leervorderingen met het Cito-Leerlingvolgsysteem. In P.R.L. Simons & J.G.L.C. Lodewijks (red), *Onderwijs Research Dagen 1990: Technologie/Methodologie*, Nijmegen: I.T.S.
- Leeuwe, J.F.J. van (1984). *Lisrel notities*. Nijmegen: RTD-PAW.
- Leij, A. van der (1992). Risico op functionele ongeletterdheid in de basisschool. *Pedagogische Studiën*, 69, 5, 352-370.
- Little, R.J.A. & Rubin, D.B. (1987). *Statistical analysis with missing data*. New York: John Wiley & Sons.
- Lucas, M.H. & Singer H. (1975). Dialect in relation to oral reading achievement. Recoding, encoding or merely a code? *Journal of Reading Behavior*, 7, 137-148.
- McKeown, M.G., Beck, I.L., Omanson, R.C. & Perfetti C.A. (1983). The effects of longterm vocabulary instruction on reading comprehension. *Journal of Reading Behavior*, 15, 3-18.
- McKeown, M. & Curtis, M. (1987) (eds), *The nature of vocabulary acquisition*, Hillsdale, New Jersey/London: Lawrence Erlbaum.
- Mommers, M.J.C., Van Leeuwe, J.F.J., Oud, J.H.L. & Janssens, J.M.A.M. (1986). Decoding skills, reading comprehension and spelling: a longitudinal investigation. *Tijdschrift voor Onderwijsresearch*, 11, 2, 97-113.
- Nagy, W. & Herman, P. (1987), Breadth and depth of vocabulary knowledge: Implications for acquisition and instruction. In: M.G. McKeown & M. Curtis (eds.), *The nature of vocabulary acquisition*. (19-36). Hillsdale, New Jersey/London: Lawrence Erlbaum.
- Paris, S. G. & Oka, E. R. (1986). Children's reading strategies, metacognition and motivation. *Developmental Review*, 6, 25-56.
- Parish, C. & Perkins K. (1985). Factors influencing anaphoric processing in second language reading comprehension. *Journal of Research in Reading*, 8, 106-115.
- Perfetti, C. (1985). *Reading Ability*. New York: Oxford University Press.
- Pressley, M, Levin, L. & McDaniel, M. (1987), Remembering versus inferring what a word means: mnemonic and contextual approaches. In: M. McKeown & M. Curtis (eds), *The nature of vocabulary acquisition* (107-128). Hillsdale, New Jersey/London: Lawrence Erlbaum.
- Rayner, K. & Pollatsek, A. (1989). *The psychology of reading*. NJ: Prentice Hall.
- Rinehart, S. D., Stahl, S. A. & Erickson L. G. (1986). Some effects of summarization training on reading and studying. *Reading Research Quarterly*, 21, 422-438.
- Rosales, S. M. (1981). *A study of the relationship between oral language proficiency and reading achievement of Mexican American children*. Dissertation. Texas: A. & I. University.
- Scheerens, J. (1990). School effectiveness research and the development of process indicators of school functioning. *School effectiveness and school improvement*, 1 (1), 61-80.
- Schouten-Van Parreren, C. (1985), *Woorden leren in het vreemde-talenonderwijs*. Apeldoorn: Van Walraven.
- Sijstra, J. (1993). Leesprestaties medio basisonderwijs. *School en Begeleiding*, 10 (8), 8-10.
- Spear, L.C. & Sternberg, R.J. (1987). An information-processing framework for understanding reading disability. In: S.J. Ceci (ed.), *Handbook of cognitive, social and neuropsychological aspects of learning disabilities*. Hillsdale, NJ: LEA.
- Stanovich, K.E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360-407.
- Sternberg, R.J. (1987). Most vocabulary is learned from context. In: M. McKeown & M. Curtis (eds), *The nature of vocabulary acquisition* (89-106). Hillsdale, New Jersey/London: Lawrence Erlbaum.

- Teunissen, F. (1986). *Een School, Twee Talen*. Dissertatie. Utrecht: VOU.
- Tregar, B. & Fun Wong, B. (1984). The relationship between native and second language reading comprehension and second language oral language ability. In C. Rivera (Ed.), *Placement procedures in bilingual education*. Clevedon: Multilingual Matters.
- Triesscheijn, T., Van de Berg, H. & Hoeksma, J. B. (1985). *Voorstudie periodieke peiling van het onderwijsniveau, Deel V: Lees- en schrijffprestaties van allochtone leerlingen in de zesde klas*. Amsterdam: SCO.
- Verhoeven, L. (1987). *Ethnic Minority Children Acquiring Literacy*. Dissertatie. Dordrecht: Foris.
- Verhoeven, L. (1990a). Acquisition of reading in a second language. *Reading Research Quarterly*, 15, 90-114.
- Verhoeven, L. (1990b). Language variation and learning to read. In P. Reitsma & L. Verhoeven (eds), *Acquisition of reading in Dutch*. Dordrecht: Foris.
- Verhoeven, L. (1991). Begrijpend lezen in de aanvangsfase: Rol van coherentie, inferentie en anafora. In P. Reitsma & M. Walraven (red.), *Instructie in begrijpend lezen*. Delft: Eburon.
- Verhoeven, L. (1992a). Inleiding functionele geletterdheid. In L. Verhoeven (red.), *Handboek lees- en schrijfdidactiek*. Lisse: Swets & Zeitlinger.
- Verhoeven, L. (1992b). Mondelinge en schriftelijke vaardigheid in het Nederlands als eerste en tweede taal. *Toegepaste Taalwetenschap in artikelen*, 43, 66-77.
- Verhoeven, L. (1992c). *Drie-Minuten-Toets, Grafemendictee en Toets voor Auditieve Synthese*. Arnhem: Cito.
- Verhoeven, L. (1992d). *Fonemendictee en Toets voor auditieve analyse*. Arnhem: Cito.
- Verhoeven, L. (1992e). *Lezen met Begrip 1*. Arnhem: Cito.
- Verhoeven, L. (1992f). *Woordenschattoets 1 (experimentele versie)*. Arnhem: Cito.
- Verhoeven, L. & Mommers, C. (1989). De didactiek van het leren lezen: Van 'reading readiness' naar 'emergent literacy'. *School en Begeleiding*, 6, 21, 60-65.
- Verhoeven, L. & Mommers, C. (1992). Ontwikkelings- en programmagericht leesonderwijs. In L. Verhoeven (red.), *Handboek lees- en schrijfdidactiek*. Lisse: Swets & Zeitlinger.
- Verhoeven, L. & Van Kuijk, J. (1991). Peiling van conceptuele en metalinguïstische kennis bij de aanvang van het basisonderwijs. *Pedagogische Studiën*, 68, 9, 415-425.
- Verhoeven, L. & Vermeer, A. (1992). Woordenschat van leerlingen in het basis- en MLK-onderwijs. *Pedagogische Studiën*, 69, 3, 218-234.
- Voeten, M.J.M. (1991) Beschrijving van de individuele ontwikkeling van leesvorderingen. In J. Hoogstraten & W.J. van der Linden (red.), *Methodologie; Onderwijsresearchdagen 1991*. Amsterdam: SCO.
- Vooijs, M.W., Kamp, L.J.Th. van der, Kooistra, C.M. & Voort T.H.A. van der (1992). De multiniveaubenadering toegepast op leesvorderingen. *Tijdschrift voor Onderwijsresearch*, 17, 6, 329-338.
- Weber, R. (1991). Linguistic diversity and reading in American society. In R. Barr, et al. (eds.), *Handbook of Reading Research*, New York: Longman.
- Wijnstra, J.M. (red) (1988). *Balans van het rekenonderwijs in de basisschool; uitkomsten van de eerste rekenpeiling medio en einde basisonderwijs*; PPOON-reeks, nr. 1. Arnhem: Cito.
- Zwarts, M. (1990). *Balans van het taalonderwijs aan het einde van de basisschool*. Arnhem: Cito.

Manuscript ontvangen 14-7-1993

Definitieve versie ontvangen 3-5-1994

Boekbespreking

B. Vreeburg, *Identiteit en het verschil. Levensbeschouwelijke vorming en het Nederlands voortgezet onderwijs*. Zoetermeer: Uitgeverij De Horstink, Boekencentrum, 1993. ISBN 90-6184-374-x, 351 p. (f 59,-).

Wat gebeurt er in het Nederlandse voortgezet onderwijs aan levensbeschouwelijke vorming? Dat is de brede vraag, waarop Bruno Vreeburg een antwoord zoekt in de dissertatie "Identiteit en het verschil". Op 20 september 1993 promoveerde hij aan de Universiteit van Amsterdam op dit proefschrift. Hier bespreken we de handelseditie daarvan. Het betreft een vergelijking tussen 8 onderscheiden onderwijsrichtingen, n.l. openbaar, algemeen-bijzonder, samenwerkingscholen, rooms-katholiek, interconfessioneel, protestants-christelijk, reformatorisch en gereformeerd onderwijs.

Vreeburg ontwikkelt in het eerste hoofdstuk zijn centrale pedagogische en onderwijskundige uitgangspunten. Volgens hem introduceren opvoeding en onderwijs mensen altijd in de symbolische orde van een samenleving en daarbinnen van dominante levensbeschouwelijke subculturen. Daardoor eigenen zij zich vrij oproepbare handelingsstructuren toe. Zo vormen opvoeding en onderwijs mensen tot vrije subjecten binnen de grenzen van die symbolische orde.

In het tweede hoofdstuk werkt hij nader uit, wat hij onder levensbeschouwing verstaat. Volgens hem staan in levensbeschouwelijke tradities rituelen centraal, d.w.z. herhaalde symbooltaal en symboolhandelingen, waarin verwezen wordt naar de goddelijke en spirituele werkelijkheid. Deze levensbeschouwingen functioneren enerzijds als een cultureel kapitaal met ruilwaarde, maar bevatten anderzijds altijd een gratis element. Omdat aan deze verwijzingen telkens nieuwe betekenissen gehecht worden vraagt levensbeschouwing een cyclisch leerproces, waarin niet zozeer een cumulatie alswel een voortdurende verschuiving en vernieuwing van betekenissen plaatsvindt. De vraag rijst dan wel in hoeverre en op welke wijze aan deze rituele aspecten van levensbeschouwing nog recht gedaan kan worden in ons huidig onderwijsbestel.

Daarom gaat de auteur in het derde hoofdstuk in op de historische ontwikkeling van het verzuilde onderwijsbestel in Nederland en op de vragen, die dat oproept, met name voor de positie van de levensbeschouwelijke vorming in het onderwijs. Dat bestel vormt op zich slechts de formele context. Inhoudelijk is vooral de invloed van levensbeschouwelijke instituties, van kerkelijkheid en godsdienstigheid van belang. Daarom bespreekt hij in hoofdstuk 4 belangrijke onderwijsstandpunten van de verschillende kerken en het Humanistisch Verbond, gaat hij in op een aantal kwantitatieve gegevens rond kerkelijkheid, met name van jongeren en op de secularisatie. Daaruit blijkt, dat het geestelijke leven in Nederland en dus ook op de scholen gedurende de laatste decennia sterk veranderd is. Dat roept de vraag op, hoe de verschillende richtingen in het onderwijs zelf op deze veranderingen reageren en zelf hun relatie tot onderwijs formuleren.

Op die vraag gaat Vreeburg in hoofdstuk 5 – het centrale hoofdstuk – als volgt in. Eerst bespreekt hij van de verschillende richtingen de bestuurlijk-organisatorische structuur, de identiteitsdiscussie en eventuele empirische onderzoeksgegevens. Daarna ontwikkelt hij een theoretisch kader, waarbinnen hij de identiteitsvormgevingen in de onderscheiden richtingen vergelijkt. Dit leidt tot een aantal verwachtingen ten aanzien van verschillen tussen onderscheiden soorten scholen.

In hoofdstuk 6 wordt verslag gedaan van een empirisch onderzoek op scholen ter toetsing van deze verwachtingen. In feite gaat het om een uitgebreide schriftelijke enquête onder schooldirecties (N=550). Opvallende, maar niet onverwachte uitkomsten zijn, dat er bij het openbaar onderwijs wel een sterk identiteitsbesef bestaat met als uitgangspunten pluriformiteit en tolerantie, maar dat dit weinig tot uitdrukking komt in concrete activiteiten. Slechts 18% van de openbare scholen heeft een vak als geestelijke stromingen of een ander soort van levensbeschouwelijke vorming. Alleen als het gaat om aandacht voor religieuze verplichtingen van allochtone leerlingen scoren zij beter dan confessionele scholen. De samenwerkingschool heeft veelal dezelfde uitgangspunten als het openbaar en het algemeen bijzonder onderwijs, maar is

op het terrein van levensbeschouwelijke vorming en de vormgeving van de identiteit actiever. De inspanningen op dit terrein zijn zelfs groter dan bij het protestants-christelijk en rooms-katholiek onderwijs. Bij met name het katholiek onderwijs blijkt de levensbeschouwelijke identiteit het meest diffuus te zijn geworden. Op deze scholen zijn wel een groot aantal levensbeschouwelijke activiteiten aangetroffen, maar 44% van de scholen geeft aan dat het vak voor levensbeschouwing geen binding meer heeft met de eigen katholieke traditie. Gebed wordt nog op 11% van de r.k. scholen wekelijks of dagelijks en op 40% van de scholen soms gepraktiseerd. Er is sprake van een kloof tussen kerkelijke leiding en het katholiek onderwijs. Voorts komt uit het onderzoek naar voren dat, hoewel men in katholieke bestuurskringen zeer afwijzend staat ten opzichte van de samenwerkingschool, samenwerkingscholen en katholieke scholen elkaar qua levensbeschouwelijke vorming niet veel ontlopen. Het protestants-christelijk onderwijs kent een sterkere traditiebinding dan het katholieke onderwijs. Dat blijkt onder meer uit de stringentere eisen bij de benoeming van leerkrachten. Op 86% van de protestants-christelijke scholen is sprake van wekelijks of dagelijks gebed en op 8% is dat soms. Toch rapporteert ook 35% van de p.c. scholen dat het levensbeschouwelijk vak geen binding meer heeft met de eigen traditie. Het interconfessioneel onderwijs zit tussen het r.k. en p.c. onderwijs in, maar ligt dicht bij het p.c. onderwijs. De reformatorische en gereformeerde scholen richten zich op verschillende orthodox protestantse groepen. Met in totaal ca. 35 scholen voor voortgezet onderwijs en een volledig uitgebouwde verzorgingsstructuur bestaan er twee orthodoxe minizuilen. Hier is sprake van een sterke binding met de verschillende orthodox protestantse tradities. In het algemeen geldt trouwens, dat de binding aan een levensbeschouwelijke traditie een belangrijke latente variabele blijkt te zijn, die met name veel van de rituele en algemeen christelijke activiteiten verklaart (niet de omvang van het vak levensbeschouwelijke vorming). Deze traditiebinding wordt zelf vooral verklaard door de visie op de gebondenheid aan een levensbeschouwing en door de algemene identiteitsvisie (niet door kerkelijkheid en urbanisatiegraad).

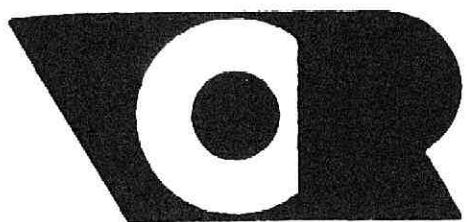
Een belangrijke activiteit op levensbeschouwelijk gebied is natuurlijk het vak, dat al of niet onder de naam levensbeschouwelijke vorming wordt gegeven. Daaraan besteedt Vreeburg het zevende hoofdstuk. Hij onderscheidt daarbij 3 hoofdsoorten, n.l. geestelijke stromingen of kennis van het geestelijk leven, de agogisch gerichte levensbeschouwelijke vorming (levo) en de richtinggebonden levensbeschouwelijke vorming (p.c. godsdienstonderwijs, r.k. catechese, humanistisch vormingsonderwijs en islamitisch godsdienstonderwijs). Van elk soort bespreekt hij recente studies, welke hij er karakteristiek voor acht. En naar aanleiding van die studies formuleert hij de specifieke problemen en dilemma's en de voor- en nadelen van elk soort. Daaruit concludeert hij, dat de discussie over dit vak nog onbeslist is.

In hoofdstuk 8 ten slotte bespreekt hij nog eens kort de opbrengst van deze studie en doet hij een aantal suggesties voor verder onderzoek op dit gebied. Een interessant punt daarbij is de hypothese, dat het voortbestaan van de verzuiling in het onderwijsbestel minder verklaarbaar is door machtsverhoudingen en institutionele waarborgen van de zuilen (institutionaliseringshypothese), maar meer door de effectiviteit van of behoefte aan bijzondere scholen en doordat de overheid zich afzijdig houdt van levensbeschouwelijke strijd (de pacificatiehypothese) en vooral doordat de onderwijsrichtingen kunnen voorzien in de nauwe relatie tussen onderwijskundig en pedagogisch spreken enerzijds en levensbeschouwelijk spreken anderzijds (pedagogische hypothese). Vreeburg sluit zijn studie af met een relativering van de mogelijkheden voor levensbeschouwelijke vorming in het huidige onderwijs.

Al bij al is het een omvangrijke en belangrijke studie geworden met interessante resultaten. Met name in de politieke discussies rond de levensbeschouwelijke identiteit van verschillende soorten scholen zal men er zeker in de eerstkomende jaren moeilijk omheen kunnen. Het brengt die discussies hopelijk ook een stap verder, omdat we nu voor het eerst beschikken over een vergelijking van relevante actuele feiten in de onderscheiden richtingen. Een aantal op louter vermoedens gebaseerde argumenten kunnen voortaan gelukkig vervangen worden door meer feitelijke gegevens. Juist de inventarisatie van die feiten is de grootste verdienste van dit proefschrift.

Anderzijds ligt in dit inventariserend karakter tegelijk ook de zwakte. Mij is althans niet duidelijk geworden, waarom Vreeburg zelf het van belang heeft gevonden om zo breed uit te zoeken, wat er in Nederland gedacht en gedaan wordt op het gebied van levensbeschouwelijke vorming in het voortgezet onderwijs. Wat was zijn "erkenntnisleidende Interesse"? Waarin bestaat de dynamiek van zijn onderzoek? In zijn inleiding schrijft Vreeburg: "Soms is het goed om nauwgezet de problemen en de (denk)structuren die deze situatie beheersen in kaart te brengen, zonder dat men zich direct tot één van de onderwijsrichtingen bekent." Dat lijkt me op zich wel juist. De vraag is echter, waarvoor dat dan precies goed is. Vreeburg formuleert dat aldus: "Wanneer deze studie er toe kan bijdragen, dat men beter toegerust deze problematiek tegemoet treedt en een aansporing geeft om gezamenlijk over de vraagstukken van onderwijs en levensbeschouwing na te denken, dan acht ik haar doel bereikt". Was het Vreeburg er dus alleen om te doen, dat de discussie voortgaat en heeft hij zelf verder geen enkele inhoudelijke bijdrage aan die discussie willen leveren? Dat lijkt me onwaarschijnlijk en is ook in tegenspraak met zijn kritische weergave van wat anderen denken en willen. Ik kan alleen niet zo'n duidelijke lijn in zijn eigen standpunten ontdekken. Dat vind ik jammer, want dat discussieert zo moeilijk. Mij lijkt, dat werkelijke voortgang op dit gebied alleen geboekt kan worden als we niet alleen zo nauwkeurig mogelijk weten hoe het feitelijk zit, maar ook wat betrokkenen denken en willen. En met dit proefschrift is Vreeburg in elk geval een belangrijke betrokkene geworden.

A. de Jong
Katholieke Hogeschool Heerlen

The logo consists of a stylized 'O' inside a white circle, which is set against a black background that forms a shape resembling the letter 'R'.

ORD BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie-adres

Dr. P.J.J. Stijnen, Open Universiteit,
Postbus 2960, 6401 DL Heerlen,
tel. 045-762293

Jrg. 18, nr. 2

1994

Inhoudsopgave

- Internationale balans van het onderwijsonderzoek in Nederland
(Openingstoespraak Onderwijs Research Dagen 1994, 24 mei 1994)
Dr. L.S.J.M. Henskens 284
- Het Maatschappelijk belang van onderwijskunde
(Inleiding bij de ORD 1994)
N.A.J. Lagerweij 289

Internationale balans van het onderwijsonderzoek in Nederland

Openingstoespraak Onderwijs Research Dagen 1994, 24 mei 1994

Dr. L.S.J.M. Henskens
Ministerie van O. en W.

Dames en heren,

Deze week is het kabinet het wetsvoorstel tot herziening van de WOV aan de orde: de wet op de onderwijsverzorging. Binnenkort zult u kunnen vernemen welke gedachten daarin zijn ontwikkeld over de toekomst van het onderwijsonderzoek, in het bijzonder over de SVO. Het nieuwe kabinet en het parlement zullen verder beslissen hoe in de toekomst de verzorgingsstructuur gaat functioneren. De huidige politieke windstille periode leent zich op dit moment goed voor het opmaken van een balans. Hoe staat het onderwijsonderzoek ervoor? Deze vraag wordt niet alleen in Nederland gesteld. De laatste jaren blijkt in verschillende landen, vooraanstaand op het gebied van onderwijsonderzoek, deze vraag aan de orde: de Verenigde Staten, Zweden, Australië, Duitsland. Ook internationale organisaties spelen hier een rol.

In deze korte bijdrage behandel ik eerst een overzicht van wat er nationaal en internationaal gaande is aan evaluaties en inventarisaties. In de tweede plaats geef ik u een financiële balans en een inhoudelijke balans vanuit internationaal perspectief. Ik sluit af met twee trends die uw werk zullen beïnvloeden.

Schema 1: Evaluaties/inventarisaties onderwijsonderzoek.

	object	optiek
WRR: WOV	vraag scholen aanbod verzorging	marktwerking doelmatigheid
OCV	onderzoeksprojecten, witte plekken	kwaliteit, doelgerichtheid
EG	programma's, onderzoeksinstellingen	inventarisatie Europees netwerk
OECD	country case studies	structure and functioning educational R&D

In schema 1 ziet u de verschillende instanties die een analyse gemaakt hebben of doende zijn een analyse te maken van het onderwijsonderzoek. De afgelopen jaren heeft een evaluatie van de WOV plaatsgevonden door de WRR en heeft de WRR een advies opgesteld. Het Leitmotief van dit advies was enerzijds het versterken van de marktwerking uitgaande van autonome scholen die hun eigen vraag articuleren en anderzijds het vergroten van de doelmatigheid. In het kader van de herziening van de WOV heeft SVO niet centraal gestaan.

Van geheel andere aard zijn de activiteiten van de OCV, de overleg Commissie Verkenningen. De OCV is een instrument van het wetenschapsbeleid. Een aantal wetenschapsgebieden wordt onder auspiciën van de OCV geanalyseerd aan de hand van de vraag of deze wetenschappen voldoende doelgericht zijn. Het gaat hier om de inhoudelijke aansluiting tussen maatschappelijke vraag en wetenschappelijk aanbod én om wetenschappelijke kwaliteit. De OCV heeft

zojuist een analyse laten uitvoeren door het ITS van het projectenbestand in het onderwijsonderzoek. Mede naar aanleiding van een discussiemiddag hierover zal de OCV met het voorstel komen voor een verkenning van gewenste toekomstige ontwikkelingen in het onderwijsonderzoek.

Internationaal zijn er twee belangrijke organisaties actief. Na het verdrag van Maastricht heeft de *Europese Unie* een meer nadrukkelijke taak voor het gehele onderwijs. In Brussel is een Directoraat Generaal voor het onderwijs in de maak en in dat kader is men bezig met een inventarisatie van onderzoeksinstituten en onderzoeksprogramma's in de landen van de EU. Voor zover wij dat kunnen overzien gaat het Brussel om de opbouw van een netwerk om kennis over het onderwijs te genereren. Niet onbelangrijk is het ook te vermelden dat een nieuw sociaaleconomisch actieprogramma voor het eerst een substantieel aandeel voor onderwijsonderzoek heeft, zo'n 25 miljoen ECU.

De *OECD* tenslotte is, op basis van een uitspraak van ministers van onderwijs van OECD-landen, een project gestart naar de 'structure and functioning of educational research and development'. Het project bestaat uit 'country case studies'. Landen maken een evaluatie van het functioneren van hun eigen 'educational R&D' en beleggen daarover een internationale conferentie. Nederland is gevraagd om in 1995 de slotconferentie te organiseren. Momenteel stellen SVO en O. en W. een beschrijvend basisrapport op, met behulp waarvan een internationaal samengestelde reviewcommissie een evaluatie zal uitvoeren, vergelijkbaar met de onderwijsreview. In de verschillende betrokken landen en bij de internationale organisaties is men ervan overtuigd dat kennis van het onderwijs een onmisbare basis is voor het verbeteren van de onderwijspraktijk en het onderbouwen van het onderwijsbeleid. Hoe staat het Nederlandse onderwijsonderzoek er internationaal voor? Ik ga eerst in op het volume of de omvang van de middelen.

Schema 2: Middelen voor onderwijs R&D in miljoenen gulden (1992).

Universiteiten		30
- vakgroepen onderwijskunde e.d.	15	
- didaxologisch onderzoek	5	
- ULO's	4	
- RWO bureau's + OU	6	
SVO		20 (- > 15)
O&W		18
- hoger onderwijs	3.7	
- beroepsonderw. en volw. educatie	3.4	
- arbeidsvoorwaarden leraren	3.9	
- PPON	3	
- overig	4	
'Onderwijsverzorging'		26
R&D SLO + CITO	20	
CIBB	5	
SVE	1	
R&D LPC + SBD	?	
Overig (gemeenten, HBO Raad, NWO, andere departementen)	5	
		100

De omvang van het onderzoek van het onderwijs wordt nogal eens verbonden aan het budget van SVO. Tengevolge van de maatregelen bij de Tussenbalans loopt het SVO-budget terug van zo'n 20 naar 15 miljoen. Er is echter meer. In totaal wordt er in Nederland nu zo'n 100 miljoen besteed. Het grootste aandeel neemt de eerste en tweede geldstroom van de universiteiten voor zijn rekening, zo'n 30 miljoen. O. en W. is goed voor 18 miljoen beleidsonderzoek. Volgens de internationaal erkende definitie van 'research and development' (het Frascati manual) valt een deel van de onderwijsverzorging ook onder 'educational R&D': 26 miljoen is de schatting

(SLO/CITO/CIBB/SVE) met daarnaast een onbekend deel van de activiteiten van LPC's en SBD's die ook onder de definitie van 'research and development' vallen. Tenslotte is er een aantal overige opdrachtgevers met kleinere budgetten. Ondanks afgenomen budgetten voor SVO en voor het eerste geld-stroomonderzoek op de universiteiten in de tachtiger jaren is onze indruk dat het totaalbudget voor onderwijsonderzoek is gestegen.

Is die 100 miljoen of – zonder de development functie – 75 miljoen nu veel of weinig? Internationaal beschikken wij over twee vergelijkingen.

Schema 3: Uitgaven voor onderwijsonderzoek als % van alle wetenschapsuitgaven.

	1e geldstroom	2e en 3e geldstroom
Verenigde Staten	6.2 %	0.4 %
Nederland	3.1 %	4.8 %
Groot Brittannie	3.0 %	2.1 %
West Duitsland	2.6 %	0.8 %
Frankrijk	0.4 %	0.0 %

Bron: Irvine, Martin en Isard, 1990.

De eerste vergelijking is gebaseerd op een verdeling van uitgaven over wetenschapsgebieden. Uit deze studie van Irvine, Martin, Spru en Isard, is afgeleid welk percentage van alle wetenschapsuitgaven door vijf landen wordt besteed aan onderwijsonderzoek. Wat betreft de eerste geldstroom geeft de Verenigde Staten het meeste uit, gevolgd door Nederland en Engeland. Bij het 2e en 3e geldstroom-onderzoek is Nederland koploper. In dit overzicht gaat het alleen om onderzoek, niet om ontwikkeling.

Schema 4: Uitgaven voor onderwijsonderzoek en ontwikkeling als % van de totale onderwijsuitgaven.

Australië	0.37 %
Verenigde Staten	0.25 % (< - schatting)
Zweden	0.24 %
Canada	0.22 %
Nederland	0.20 % - 0.25%
Zwitserland	0.06 %

Bron: McKenzie and McGaw, 1993

Een andere benaderingswijze is het bekijken welk percentage of promillage van de onderwijsbegroting wordt uitgegeven aan onderwijsonderzoek en ontwikkeling (R&D). Deze exercitie is ondernomen in een pilot study voor een nieuwe indicator bij de uitgave van de OECD over onderwijsindicatoren 'Education at a Glance'. De CBS wetenschapsstatistieken vormen de basis voor het Nederlandse percentage.

Australië is hier het mekka van de onderwijsonderzoekers. Wanneer we in Nederland de 'development' meerekenen is Nederland met de Verenigde Staten tweede in deze groep landen. Als alleen het percentage research wordt meegerekend leidt dat voor Nederland tot een vijfde plaats.

De verschillen tussen de landen op de positie 2 t/m 5 zijn klein en de percentages zijn, zoals we in het Nederlandse overzicht hebben kunnen zien, gevoelig voor definiëring wat wel en niet onder R&D wordt verstaan.

Wanneer we een poging doen om op meer *kwalitatieve* dimensies de Nederlandse R&D in internationaal opzicht te karakteriseren zijn er vergelijkbare problemen.

Schema 5: Internationale positie Nederlands onderzoek op 7 dimensies.

	laag 1	2	3	4	hoog 5
omvang	(FR)				NL
institutionalisering	(UK)				NL
marktgerichtheid	(DL)				NL
multidisciplinariteit		(DL)	(UK)		NL
wetenschapsoriëntatie			NL		(DL)
praktijkgerichtheid		NL			(UK)
internationale oriëntatie		(DL)		NL	(UK)

Schema 5 bevat een aantal dimensies waarmee het onderzoek in een land kan worden getypeerd. Om aan te geven dat deze dimensies verschillen tussen landen laten zien is ook de positie van enkele grote Europese buurlanden aangegeven. Het gaat hier om inschattingen gemaakt in het eerdergenoemde OECD-project. Over de *omvang* is het nodige gezegd. Wat betreft *institutionalisering* is Nederland ver ontwikkeld. Enerzijds komt dit door de professionele verzuiling van onderzoek, ontwikkeling en toetsing; SLO en CITO behoren in hun soort tot de grootste instituten ter wereld. Anderzijds kent Nederland in vergelijking tot andere landen professionele onderzoeksinstituten, de meeste verankerd binnen de universiteiten. Vanwege een omvangrijke derde geldstroom en het functioneren van SVO als opdrachtgever die bemiddelt voor aanvragers uit overheid en veld, is het Nederlandse onderzoek sterk *marktgericht*. Het ontstaan van de onderwijskunde uit de pedagogie, de psychologie en de sociologie geeft de *multidisciplinariteit* al aan. De laatste jaren laten een toename zien van de inbreng van 'nieuwe' disciplines als de bestuurskunde, de onderwijsconomie, het onderwijsrecht en de taalkunde.

Alhoewel er in Nederland een behoorlijke universitaire eerste geldstroom en daarmee een *wetenschappelijke oriëntatie* is, is per saldo het totaal van het onderzoek vooral gericht op algemene onderwijsproblemen, gedefinieerd door het beleid of door koepelorganisaties. In tegenstelling tot Engeland en Duitsland is de opleiding van leraren, zeker die voor het basisonderwijs, bij ons niet of minder ingebed in universiteiten. De Nederlandse opleidingen voor onderwijskunde bereiden in de eerste plaats voor op functies in de onderwijsverzorging of in het onderwijsbeleid op verschillende niveaus. In combinatie met het gegeven dat scholen hier nauwelijks invloed hebben op de opstelling van de onderzoeksagenda is de directe *praktijkgerichtheid* van het onderzoek gering. Deze functie wordt nu ingevuld door schoolbegeleidingsdiensten en landelijke pedagogische centra. Het meeste didaxologische onderzoek vindt niet plaats bij de onderwijskundige vakgroepen, maar bijvoorbeeld bij faculteiten der wiskunde- en natuurwetenschappen of bij algemene taalwetenschappen.

Recent is een nieuwe versie van Nederlandse wetenschapsindicatoren uitgekomen. Tot onze verrassing stond bij de indicator, *internationale oriëntatie* gemeten naar co-publicaties met buitenlandse auteurs, de onderwijswetenschap op de tweede plaats van een lijst van alle Nederlandse wetenschappelijke disciplines. De omvang van de internationale publicatie-activiteit van Nederlandse onderwijswetenschappers was gemiddeld en bijvoorbeeld lager dan Engelse wetenschappers maar duidelijk hoger dan een aantal andere Europese landen waaronder Zweden of Duitsland.

Zoals ik in het begin heb gezegd ligt een herijking van het onderzoekbeleid nog voor ons. Twee algemene trends zullen dit beleid zeker beïnvloeden. Ten eerste gaat de *inter-*

nationalisering verder. De EU is een belangrijke nieuwe beleidsmatige en wellicht in de toekomst ook belangrijke financiële factor in het onderwijsonderzoek. In OECD-kaders wordt in toenemende mate geparticipeerd in multilaterale activiteiten. Dicht bij huis wordt de samenwerking met grenslanden, Vlaanderen en Duitse deelstaten, geïntensiveerd. De tweede trend betreft de toenemende *concurrentie op de markt van kennis* over het onderwijs. In de wetenschapsdynamica wordt dit fenomeen het Agora-model van wetenschap genoemd. Dat staat tegenover het traditionele Olympus-model waar de wetenschap in relatief isolement kennis produceert die om toepassing vraagt. In het Agora-model is de wetenschap slechts een van de partijen op de markt van kennis. Een opkomende concurrent op de Nederlandse markt van kennis over het onderwijs is bijvoorbeeld de inspectie die zich profileert met een meer systematische evaluatietaak. Het Proces Management Basisvorming houdt zijn eigen implementatie-enquêtes. Deze concurrentie om de rol van kennis is gepaard gegaan met een concurrentie om middelen voor onderzoek. Voor zover het volume aan werk in de grote instituten voor onderwijsonderzoek is afgenomen, een fenomeen dat ons niet is ontgaan, is de oorzaak eerder gelegen in meer concurrentie op een meer open markt dan in minder middelen. Tegelijkertijd zal de koopkracht van scholen en lagere overheden of organisaties van scholen toenemen, leidend tot vragen om lokale kennis. In de nabije toekomst komen daar internationale concurrentie, maar ook een internationale vraag, bij.

Naar onze mening is het Nederlandse onderwijsonderzoek van voldoende kwaliteit en efficiëntie om deze concurrentie aan te kunnen.

Ik wens u een succesvolle conferentie toe.

Het maatschappelijk belang van Onderwijskunde (Inleiding bij de ORD 1994)

N.A.J. Lagerweij

De titel van deze inleiding bij de ORD 94 is voor mijn opvatting over onderwijskunde van grote waarde: het maatschappelijk belang van onderwijskunde. Ik ben van mening dat wetenschapsbeoefening op het gebied van onderwijs in nauwe relatie staat en moet staan met maatschappelijke taken. Ik heb mij de vraag gesteld, hoe de onderwijskunde zich tegenwoordig in dit opzicht manifesteert. Niet zo zeer omdat het naar mijn waarneming de verkeerde kant opgaat. Integendeel, de onderwijskunde blijkt als vanouds een nauwe relatie te hebben met de actuele ontwikkelingen in het studieobject: het onderwijsveld. Toch zijn er enkele signalen te bespeuren die om enige waakzaamheid vragen.

Laat ik, om duidelijk te maken wat ik bedoel, beginnen met het weergeven van een gesprek, dat ik onlangs had met een kennis. Deze kennis is adjunct-directeur van een basisschool. We spreken met elkaar nogal vaak over allerlei zaken rondom het onderwijs en het onderwijsbeleid. Vorige week vertelde ze me met enige opwinding het volgende. De inspectrice had de school bezocht, maar niet op de gewone wijze. Zij had in de helft van alle klassen zitten observeren. De bedoeling was, aldus deze kennis, om de effectiviteit van het rekenonderwijs te bekijken.

“Ze lette er op hoeveel leerlingen niet goed bij de les waren. Ze deed dat elke vijf minuten. “ In de middagpauze heeft de inspectrice de gegevens in de computer ingevoerd. ’s Middags om drie uur was de inspectrice klaar met haar observaties. Daarna volgde een teambespreking.

“En wat was de uitslag?” vroeg ik nieuwsgierig.

“Nou, we hadden het goed gedaan. Ze vergeleek de uitslag van onze school met de cijfers van andere scholen. Zij noemde onze school vrij goed. Na enig aandringen gaf ze ons een acht.”

“Taakgerichte leertijd”, precies wat u zegt. De inspectie is blijkbaar bezig expliciet aandacht te besteden aan rendement en effectiviteit van basisscholen. De “effectieve school beweging” is in Amerika inmiddels over zijn populariteit heen. De voors en tegens zijn daar uitvoerig aan het licht gekomen. In ons land ligt dat kennelijk nog wat anders. Onlangs verscheen het geruchtmakende rapport “Zicht op kwaliteit”. Aan de opstelling daarvan werkten enkele hoogleraren onderwijskunde mee. Ik waardeer het ten zeerste dat zij bereid waren zich te mengen in vraagstukken van onderwijspolitieke aard. Zij gaan in dat rapport zelfs zo ver dat zij pleiten voor een cultuurwijziging in de basisschool. Er moet in basisscholen meer sprake zijn van een “opbrengstgerichte cultuur”. Die boodschap blijkt aan te slaan. Die opbrengstgerichte cultuur heeft de inspectie al in de greep. Of de mensen voor de klas zich in zo’n cultuur willen begeven is een geheel andere kwestie. Persoonlijk ben ik niet erg gelukkig met de aanduiding van “opbrengstgerichte cultuur”. Zeker niet als dat leidt tot een eenzijdig accent op output meting en zeker niet als ik het voorbeeld van de inspectrice zie. Ik kom daar aan het eind van mijn bijdrage nog op terug.

Als voorzitter van de VOR wil ik bij de opening van de ORD 1994 eerst enige aandacht besteden aan de situatie waarin de VOR verkeert. Daarna maak ik enkele opmerkingen over de onderwijskunde.

1. DE VOR

Over de VOR kan ik kort zijn: het gaat uitstekend met onze vereniging. Als ik in de geest van opbrengstgerichte cultuur dat kwantitatief uitdruk kan het volgende gezegd worden. Het afgelopen jaar groeide het aantal leden spectaculair: van 432 naar 609 per 1 maart 1994. De onafhan-

kelijke variabele, die deze groei bewerkstelligde, is nog niet goed in kaart gebracht. Er is wel een plausibele verklaring. In kringen van het bestuur ondersteunt iedereen de hypothese, dat de nieuwe folder aanzienlijk heeft bevorderd dat het ledenaantal is gestegen. Het verspreiden van de folder onder alle afgestudeerde onderwijskundigen in ons land heeft ongetwijfeld bijgedragen aan het genoemde effect. Een ander eveneens positief kwantitatief gegeven heeft betrekking op de begroting. De begroting over het afgelopen jaar is sluitend en bevat voldoende reserves voor een gezonde toekomstige ontwikkeling.

Minder kwantitatief, maar kwalitatief uiterst belangrijk zijn de activiteiten in de divisies. De activiteiten die kenmerkend voor deze beroepsvereniging zijn, krijgen voornamelijk vorm in de 8 divisies. De VOR-leden hebben zich keurig verspreid over de divisies, waarbij enkele honderden ook van meerdere divisies lid zijn. Het hart van de vereniging zit in de divisies. Sommige divisies verspreiden relevante informatie in een eigen vorm. Op deze ORD is nu voor de tweede keer een indeling van presentaties gemaakt conform de divisies. Als extra thema koos de vakgroep onderwijskunde Utrecht voor onderzoek rondom de basisvorming. Het betreft hier, zoals bekend, een grootschalige vernieuwing die het gehele voortgezet onderwijs vooral inhoudelijk raakt. Over het succes van deze onderwijsvernieuwing verkeren we nog in het ongewisse. De Utrechtse onderwijsonderzoekers hebben hun interesse in actuele beleidsontwikkelingen getoond door dit thema voor hun onderzoek te kiezen. Het blijkt dat ook op andere plaatsen deze onderwijsvernieuwing voorwerp van onderzoek is. Daarmede toont de onderwijsresearch op eigen wijze betrokkenheid bij actuele maatschappelijke ontwikkelingen.

Voor de VOR leden is het Tijdschrift voor onderwijs Research (TOR) het orgaan waarin het verenigingsnieuws is opgenomen. Er zijn nieuwe afspraken met de uitgever gemaakt. Dat blad verschijnt in het vervolg 4 keer per jaar overigens met hetzelfde aantal pagina's als tot nu toe gebruikelijk. Op meer opvallende wijze dan tot nu toe gebruikelijk zal het verenigingsnieuws daarin blijven voorkomen.

Inmiddels zijn de voorbereidingen voor een Europese vereniging voor onderwijsonderzoek afgesloten. In juni vindt in Straatsburg de oprichting plaats van de European Educational Research Association (EERA). Onze vereniging neemt daaraan op actieve wijze deel. In oktober 1993 is in Aarau in Zwitserland door zo'n 14 landen het principebesluit genomen tot stichting van een Europese vereniging. Het doel van deze vereniging is het bevorderen van onderzoek en ontwikkeling (research and development) van onderwijs in Europa en zodoende het bevorderen van onderwijskunde (educational science) en het verbeteren van onderwijs.

In het bijzonder streeft men naar het bevorderen van samenwerking tussen onderwijsresearch en onderwijsbeleid en -bestuur, de lerarenopleiding en nascholing. Vanzelfsprekend wil men samenwerking bevorderen tussen de verschillende nationale organisaties op het gebied van onderwijsonderzoek, ook in relatie tot de internationale organisaties als de Europese Unie, de OECD en Unesco. Men streeft naar overleg tussen andere gespecialiseerde verenigingen op het gebied van onderwijs zoals EARLI, CESE, ATEE en CIDREE. De nationale organisaties kunnen als lid toetreden, maar ook is een individueel lidmaatschap mogelijk.

2. ONDERWIJSKUNDE

De tweede vraag die ik hier aan de orde wil stellen luidt: Hoe staat het met de onderwijskunde in Nederland? Die vraag kan vanuit verschillende invalshoeken beantwoord worden. Ik kies er drie, te weten

1. de ontwikkelingen op het gebied van het onderwijsonderzoek;
2. de situatie bij de opleidingen;
3. de bemoeienis van onderwijskunde met het onderwijsbeleid.

Onderwijsonderzoek

Op het terrein van onderwijsonderzoek doen zich twee belangwekkende ontwikkelingen voor. In de eerste plaats dient vermeld te worden dat we er in geslaagd zijn een onderzoekschool voor onderwijskunde te construeren. Alle universiteiten met een vakgroep onderwijskunde nemen er

aan deel. De betreffende aanvraag is bij de KNAW ingediend en we wachten af of erkenning ons ten deel zal vallen.

Op dit moment hebben in totaal 65 personen de criteria gehaald, die voor toetreding zijn opgesteld. Zij behoren tot de groep excellente onderzoekers, omdat zij tenminste drie artikelen in een internationaal tijdschrift gedurende de laatste vijf jaren gepubliceerd hebben. In mijn ORD lezing van vorig jaar heb ik er al op gewezen, dat een eenzijdig accent op internationaal publiceren niet zonder risico's is. Ik hoop niet dat na de erkenning van deze onderzoekerschool de leden alleen nog in een vreemde taal gaan publiceren. Hoe begrijpelijk dat ook moge zijn, de nationale en culturele aspecten van onderwijsvraagstukken vragen ook een verantwoording aan de Nederlandse belastingbetalers.

Een tweede belangrijke ontwikkeling heeft te maken met de wijze waarop de verdeling van de geldstromen in de toekomst zal gaan verlopen. In de eerste plaats krijgen we een andere situatie als gevolg van het expireren van de Wet op de Verzorgingsstructuur in 1995. De kabinetsvoorstellen betekenen een verandering in de positie van SVO en zullen leiden tot een andere verdeling van de geldstromen over groeperingen in het onderwijsveld. In de tweede plaats is er bij NWO een reorganisatie toegepast. Het onderwijsonderzoek dat op succesvolle wijze geregeld werd door de stichting PEDON is nu ondergebracht in een nieuwe, vooral grotere Stichting voor de Gedragwetenschappen (SGW). Er is daar nog wel sprake van een onderzoeksgroep "Onderwijsonderzoek en Cognitief leren & probleemoplossen". Deze onderzoeksgroep heeft echter niet veel functie meer, omdat zij geen aanvragen meer mag beoordelen.

Beide veranderingen hebben de positie van onderwijsonderzoek en van onderwijsonderzoekers minder duidelijk gemaakt. We zullen als beroepsvereniging de nieuwe lijnen kritisch moeten volgen om de belangen van onderwijsonderzoek niet onnodig te laten beschadigen.

Het onderwijsonderzoek in Nederland is onlangs door enkele personen werkzaam bij het ITS in kaart gebracht. Dit op verzoek van de overlegcommissie Verkenningen. Deze commissie heeft tot taak om in Nederland een breed gedragen proces van verkenningen op te zetten en oplossingen aan te dragen voor vraagstukken in het wetenschapsbeleid. Als voorstudie met het oog op een eventuele verkenning inzake onderwijsonderzoek verscheen van de hand van Van Gennip, Pelkmans (1994) het rapport "Zwaartepunten in het Nederlandse onderwijsonderzoek". Het rapport geeft een helder zicht op de typen onderzoek en de gerichtheid daarvan. Uitgaande van het kennisdoel blijkt dat zo'n 16 procent van het onderzoek gericht is op het vergroten van wetenschappelijke kennis. Ruim veertig procent is gericht op kennis inzake het bovenscholings onderwijsbeleid en eveneens ruim veertig procent beoogt inzichten op te leveren voor de onderwijspraktijk. Wie geen vreemdeling in dit Jeruzalem is, herkent daarin de formule die al vele jaren geldt voor de verdeling van SVO subsidie. Sinds de jaren zeventig blijkt dat ongeveer een kwart van het onderwijsonderzoek conclusiegericht is en driekwart beslissingsgericht. Eveneens een kwart is fundamenteel-strategisch van aard en driekwart is te typeren als beleidsonderzoek of als bestemd voor de onderwijspraktijk. Onderwijsonderzoek is nauwelijks fundamenteel, zuiver wetenschappelijk, aldus de auteurs (a.w. pag. 76). "Onderwijsonderzoek gaat immers per definitie over een maatschappelijk verschijnsel."

Over de kwaliteit van het onderzoek door beleidsinstanties of door anderen bevat het rapport geen gegevens. Over de benutting van onderzoeksgegevens doen de samenstellers helaas geen uitspraken. Daarover zal een Verkenningscommissie hopelijk nog eens haar licht laten schijnen. Toch is de relevantie van het onderwijsonderzoek afhankelijk van het gebruik dat belanghebbenden ervan maken. In een workshop over onderwijsonderzoek in maart 1994 is veel beharrendwaardigs opgemerkt over de verschillende aspecten aan de kwaliteit van het Nederlandse onderwijsonderzoek. Ik wijs op drie problematische kanten, die naar mijn mening voor de komende jaren ook vanuit de VOR extra aandacht verdienen.

Dat is in de eerste plaats de problematiek van de overbevoering en de daarmee gepaard gaande grote percentages non-respons. Scholen worden bestookt met vragenlijsten, verzoeken om interviews en dergelijke. Steeds vaker weigeren scholen aan dergelijke afstandelijke benaderingen mee te werken. Het wordt daardoor bijna onmogelijk nog representatieve steekproeven te verkrijgen. Waarom motiveert dergelijk onderwijsonderzoek nog zo weinig de respondenten?

In schril contrast met de lage responspercentages bij onderzoek is de jaarlijkse enquête van het Procesmanagement Basisvorming. Die vragenlijst krijgt een respons van bijna 80%. Kunnen onderzoekers hiervan wellicht wat leren?

De tweede kwestie is al vanaf het ontstaan van onderzoek gesignaleerd. Het betreft de versnippering in veel kleingoed aan onderzoekjes en het tekort aan langlopend onderzoek op dezelfde thematieken. De programmering van een samenhangend onderzoekprogramma kent inmiddels een lange en weinig tot tevredenheid stemmende traditie.

Ten derde verdient het vraagstuk van de gebrekkige verspreiding en de achterblijvende toepassing of implementatie van onderzoeksgegevens naar de onderwijspraktijk onze aandacht. We zullen ons als onderzoekers meer inspanning moeten getroosten dan tot nu toe gebruikelijk is. Er is inmiddels op allerlei terreinen kennis voor handen die erom vraagt toegepast te worden in de dagelijkse praktijk. Als in de medische sector zo nonchalant met nieuwe wetenschappelijke ontwikkelingen werd omgesprongen als in het onderwijs was half Nederland nu ziek, zwak of misselijk. In het onderwijs valt een dergelijke ongezonde situatie niet op, en doen we of er niets aan de hand is. Vooral in de opleiding van nieuwe leraren, maar ook in de nascholing zou de doorwerking van onderwijskundige informatie een vaste plaats moeten hebben. Dat is thans, zoals we ook uit visitatierapporten weten, aanzienlijk minder het geval dan wenselijk is. Naar mijn mening zouden deze drie genoemde kwesties in kringen van de VOR nader geanalyseerd moeten worden. We hebben hier een voor alle onderwijskundigen uiterst relevante problematiek aan de orde.

De opleidingen

Er is voor zover ik kan overzien niet veel bijzonders te vermelden over de wijze waarop de opleidingen tot onderwijskundigen momenteel functioneren. Na de bevredigend verlopen visitatie is er in de kamer onderwijskunde, een overlegorgaan in het kader van de VSNU, enig beraad geweest omtrent de conclusies en aanbevelingen van de visitatiecommissie (zie VSNU, 1993). Vooralsnog zoekt iedere instelling naar verbetering in eigen huis. Gemeenschappelijke vraagstukken, die een soort nationale aanpak vereisen, zijn nog nauwelijks uitgewerkt. Naast het nauwkeurig inspecteren van de eigen navels zou het de vakgroepen passen ook te streven naar een gezamenlijke opstelling in kwesties rondom de opleiding van onderwijskundigen. Zo ontbreekt het bijvoorbeeld nog aan een beroepsprofiel, een beroepscode en aan een inventaris van een minimum pakket dat in de opleiding van alle onderwijskundigen aan de orde zou moeten komen.

Nieuw is dat onder de WHW het mogelijk is een propedeuse onderwijskunde te volgen. Dat was tot dan toe alleen mogelijk in Twente. Nu bieden de Universiteit van Amsterdam en de Universiteit Utrecht een dergelijke mogelijkheid met ingang van het studiejaar 1993-1994 aan. Ik vind dat een buitengewone gunstige ontwikkeling. Het is voor mij een volgende stap in de ontwikkeling van de onderwijskunde tot een volwaardige discipline.

De toekomst van de onderwijskunde aan de universiteiten is niet geheel rooskleurig. Binnen enkele jaren kan zich een voor alle instellingen geldend probleem manifesteren. De laatste jaren blijft het aantal studenten dat voor een studie onderwijskunde kiest redelijk constant. Op korte termijn dreigt echter het gevaar van een aanzienlijke terugloop als het anti-stapel beleid van de regering succesvol blijkt. Veel HBO studenten kiezen momenteel voor een universitaire vervolgstudie. Vooral PABO-ers kiezen dan voor onderwijskunde. Dat is, zo leert onze ervaring, een vaak zeer gelukkige en verstandige keuze. Juist vanuit de directe ervaring opgedaan in de onderwijspraktijk kan de studie onderwijskunde een waardevolle aanvulling en verdieping betekenen. Om als onderwijskundige in het onderwijsveld goed te kunnen functioneren is het hebben van praktijkervaring in veel gevallen een aanbeveling. Ik vind het dan ook voor deze categorie studenten en voor onze opleidingen een bijzonder verlies als de beperkte duur van de studiefinanciering ertoe leidt dat PABO studenten gaan afzien van een universitair vervolg.

Onderwijskunde en onderwijsbeleid

Er doet zich de laatste jaren een verheugend verschijnsel voor als het gaat om de bemoeienis van

onderwijskundigen met de onderwijspolitiek. Enkele malen per jaar ontmoeten ambtenaren van het Departement van O en W een netwerk van hoogleraren onderwijskunde. De gedachtenwisselingen daar hebben vooral ten doel vraagstukken van onderwijsbeleid te confronteren met de stand van zaken op het vakgebied. Daarnaast zie je steeds vaker dat verschillende wetenschappers maatschappelijke rollen vervullen door bijvoorbeeld zitting te nemen in commissies en adviesorganen. In de afgelopen regeringsperiode hebben bewindslieden vaak daartoe het initiatief genomen. De doorwerking van onderwijskundige kennis naar maatschappelijke ontwikkelingen in de sector onderwijs krijgt zo een aantrekkelijke kans. Het zitting nemen in dergelijke gremia betekent vaak dat er ook veel kritiek over de deelnemers uitgestort wordt. Het rapport "Zicht op kwaliteit" is daar een treffend voorbeeld van'.

'Kundes' blijken te laveren tussen wetenschappelijke pretenties en maatschappelijke dienstbaarheid. Het theorie-praktijk vraagstuk vormt een centraal probleem voor 'kundes' (zie Boon, e.a. 1989). Maatschappelijke rollen van wetenschappelijke onderzoekers roepen nogal eens felle discussies op. Ook onder onderwijskundigen onderling. Het rapport "Schaal en kwaliteit" (Projectgroep schaalvergroting basisonderwijs, 1990) uit het begin van de jaren negentig was daarvan een treffend voorbeeld (zie Ax & Van Wieringen, 1990). Ik hoop niet dat dat een reden is om minder bereidwillig te zijn deel te nemen aan dergelijk advieswerk. De ervaring leert dat het succes van dergelijke participatie op wat langere termijn wel degelijk aanwezig kan zijn. Als bijvoorbeeld de bijdrage van de onderwijskundigen aan "Zicht op kwaliteit" even veel effect oplevert als de bijdrage van onderwijskundigen aan schaalvergroting, dan zijn er binnen drie jaren zo'n 70 % van de scholen bezig effectief te worden. Immers dat percentage scholen is, drie jaar na het verschijnen van de nota over schaalvergroting, thans betrokken bij fusies.

Het kritisch volgen van maatschappelijke ontwikkelingen is van groot belang voor het bestaansrecht van de onderwijskunde. Een voorbeeld van zo'n maatschappelijke trend is de onstuitbare tendens tot vergroting van autonomie van scholen. Leune (1993) analyseerde dit vraagstuk op voorbeeldige wijze en maakte duidelijk welke gevaren aan deze ontwikkeling kleven.

Reeds eerder noemde ik mijn bezorgdheid omtrent het propageren van een opbrengstgerichte cultuur (zie ook Lagerweij & Haak, 1994). Het gevaar dat ik daarbij vrees is, dat kwaliteit van het onderwijs beperkt wordt tot het vaststellen van leereffecten. De verwachting is dat dan de cognitieve kanten centraal zullen staan, waardoor de aandacht kan verslappen voor pedagogische, normatieve en affectieve aspecten van het onderwijs. Kieviet geeft in zijn afscheidsrede aan dat de onderwijskunde als discipline zich moet richten op de gehele breedte van opvoeding in en door de school. "Bij outsiders bestaat nogal eens het beeld van de onderwijskunde als een tamelijk technologisch georiënteerde discipline, die zich met name bezighoudt met evaluatie en effectmeting, en dan nog vooral met betrekking tot cognitieve aspecten" (p. 11). Dit beeld hoeft niet correct te zijn, "maar de onderwijskunde als wetenschapsbedrijf heeft wel zelf schuld aan het ontstaan van dit beeld" (pag. 12).

De positie van wetenschappers in de onderwijsarena is niet onomstreden. Debatteren over de functies van het onderwijs moet mijns inziens onderdeel blijven van het wetenschapsbedrijf (zie voorgangers als Idenburg, Van Gelder die hun verantwoordelijkheid voor het onderwijs als sector in de samenleving op vele wijzen vorm gaven). Kritische distantie ten opzichte van beleid is noodzakelijk ook al kan daarbij niet altijd empirische onderbouwing voorhanden zijn.

Actieve bemoeienis met het actuele onderwijs is kenmerkend in het dertigjarig bestaan van de onderwijskunde. Van Gelder, de vader van de onderwijskunde, is daarvan voor mij de verpersoonlijking. Bij zijn afscheid (Van Gelder, 1981) in 1981 als hoogleraar onderwijskunde in Groningen verkondigde hij de opvatting dat de wetenschappelijke taken en de maatschappelijke taken een nauwe relatie met elkaar hebben. Hij stelt de vraag aan de orde naar de grondslag van de onderwijskunde: is dat "slechts een geperfectioneerd hulpmiddel om bestaande onderwijssystemen effectiever te maken of is die onderwijskunde in staat tot zelfkritiek te komen, respectievelijk tot kritisch bezien van de maatschappelijke context van haar activiteiten". Hij verwerpt een technocratische onderwijskunde "omdat daarmee de doelstellingen van het onderwijs buiten spel gezet worden" (pag. 16). Hij eindigt zijn rede onder meer met de vaststelling "dat een maatschappelijke betrokkenheid voorwaarde is om vruchtbaar voor het onderwijs te

kunnen werken (sociale context)" (pag. 19) en "dat wij ons in dit werk gesteund weten door een visie op mens en samenleving (antropologische context)".

Ik rond hiermee mijn betoog af. Ik heb de vraag gesteld naar de wijze waarop onderwijskunde zich maatschappelijk manifesteert. Een aantal van die verschijningsvormen heb ik geschetst. De balans is niet alleen maar positief. Als levensecht voorbeeld startte ik met het bezoek van de inspectie aan een basisschool. De praktijkmensen waren enigszins onthutst door het nieuwe karakter van dat bezoek. De observatie activiteiten van die inspectrice lijken verdacht veel op het werk van een onderzoeksassistent (de salarissen zullen wel niet zijn aangepast). Ik vermoed dat deze ontwikkeling sterk beïnvloed is door onderwijskundige overwegingen. De inspectie heeft wettelijk gezien vier taken: controleren, informeren, stimuleren en adviseren. Ik hoop dat het effectiviteitsdenken er niet toe leidt dat een of meer van deze taken in de knel komen.

Ik wens de onderwijskunde toe dat zij in haar maatschappelijke taakstelling nog vele jaren inspirerend en vruchtbaar kan bijdragen aan het onderwijs en de daar levende onderwijsvraagstukken.²

LITERATUUR

- Ax, J., Wieringen, A.M.L. van, (red.) (1990). *De maat van de school*. Amsterdam: Universiteit van Amsterdam, Vakgroep Onderwijskunde; De Lier: ABC.
- Boon, L., Gottschal, P., Harbers, H., Otten, R., & Vries, G.H. de (1989). *Disciplines, kundes en de kwaliteit van wetenschap*. Maastricht/Groningen: z.u.
- Commissie Evaluatie Basisonderwijs (1994). *Zicht op kwaliteit*. Inspectie van het onderwijs; Leiden: Distributiecentrum.
- Gelder, L. van (1981). *Ervaring en Opdracht*. Groningen: Wolters-Noordhoff.
- Gennip, H. van, & Pelkmans, A. (1994). *Zwaartepunten in het Nederlandse onderwijsonderzoek*. Nijmegen: ITS.
- Kieviet, F.K. (1993). *Onderwijs en Opvoeding*. Leiden: Rijksuniversiteit Leiden.
- Lagerweij, N.A.J., & Haak, E.M. (1994). *Eerst goed kijken...* Leuven/Apeldoorn: Garant.
- Leune, H. (1993). *Onderwijskwaliteit en de autonomie van scholen*. Rotterdam: RISBO.
- Projectgroep schaalvergroting basisonderwijs (1990). *Schaal en kwaliteit in het basisonderwijs*. Zoetermeer: Ministerie van onderwijs en Wetenschappen.
- VSNU-Visitatiecommissie Pedagogiek en onderwijskunde (1993). *Onderwijsvisitatie. Pedagogiek en Onderwijskunde 1992*. Utrecht: VSNU.

Aantekeningen

1. De vijf hoogleraren die naast inspecteur-generaal Stekelenburg in de commissie zaten, werden getypeerd als "algemene onderwijskundigen, die ingehuurd zijn om voor de inspectie vooraf vaststaande conclusies van wetenschappelijke status te voorzien" (in Het Schoolblad 29, nr. 4, pag. 7: Onno Bosma, Ondeskundig, ondeskundig, ondeskundig.) Treffers stelt dat de hele wereld niet blijkt te kunnen voldoen aan de normen van de commissie. "In 1980 werd internationaal onderzoek gedaan naar rekenen, Japan scoorde gemiddeld zestig procent, Nederland was tweede met 59 procent. Volgens de norm van de commissie zit de hele wereld fout. Dat kan natuurlijk niet."
2. De Werkgroep Opleiding Onderwijskunde gaf in de jaren zeventig de volgende omschrijving: "Onderwijskunde is gericht op wetenschappelijk verantwoorde bestudering en analyse van het onderwijs in al zijn facetten om, mede op basis daarvan, bij te dragen tot de oplossing van onderwijsvraagstukken".

Affectieve variabelen: Hun effect op leermotivatie en leerprestatie

Monique Boekaerts en Gerard Seegers
*Vakgroep Onderwijsstudies, Rijksuniversiteit Leiden**

AFFECTIEVE VARIABELEN

Het is belangrijk een onderscheid te maken tussen objectieve en subjectieve competentie. Eerstgenoemd begrip verwijst naar domein-specifieke kennis en vaardigheden, naar cognitieve strategieën en naar taakgerichte gedachten die het leerproces sturen, bewaken en evalueren (metacognitieve vaardigheden). Subjectieve competentie refereert daarentegen aan persoonlijke opvattingen en overtuigingen die leerlingen hebben opgebouwd ten overstaan van hun competentie in dat domein. Meer specifiek, subjectieve competentie verwijst naar affectieve zelfgerelateerde variabelen zoals o.a. het zelfbeeld, het beeld van eigen-bekwaamheid, intrinsieke motivatie en interesse, prestatiemotivatie, doeloriëntatie, attitude, toekomstperspectief, angst en stress.

Zelfbeeld en Beeld van Eigen-Bekwaamheid

Het zelfbeeld verwijst naar de theorie die leerlingen over zichzelf hebben en over hoe anderen over hen denken. Het is eigenlijk een verzamelbegrip dat bestaat uit diverse beelden die leerlingen van zichzelf hebben op bepaalde terreinen: zoals o.a. de zelfperceptie van hun uiterlijk, aantrekkingskracht, sociale status, en lichamelijke, sociale en schoolse bekwaamheid. Het beeld dat leerlingen hebben van hun schoolse bekwaamheid is een doorslaggevende component van hun subjectieve competentie t.a.v. een bepaald vakgebied.

Uit onderzoek is bekend (vgl. Marsh, Walker & Debus, 1991) dat leerlingen reeds vanaf jonge leeftijd een gedifferentieerd beeld van eigen-bekwaamheid hebben ten aanzien van taal en rekenen. Wat echter opvalt is dat bij het eind van de basisschool het beeld van eigen-bekwaamheid voor rekenen niet samenhangt met het beeld van eigen-bekwaamheid voor taal. Dit is vreemd omdat de prestaties voor de beide vakken wel vrij sterk samenhangen. De resultaten van longitudinaal onderzoek (Helmke, 1989) wezen erop dat tot het eind van groep zeven het beeld van eigen-bekwaamheid niet kan worden afgeleid van de leerprestaties. Pas in de loop van groep acht is er sprake van wederzijdse beïnvloeding. Dit betekent in concreto, dat het actueel of toekomstig beeld van eigen-bekwaamheid met betrekking tot een vakgebied vanaf dat moment als een soort motivator gaat werken om inzet te leveren voor taken binnen dat vakgebied.

Motivatie, vakattitude, en angst

Een ander aspect van de subjectieve competentie is de motivatie van de leerling. Diverse motivatietheorieën geven een specifieke inkleuring aan dit begrip. Nicholls (1984) en Dweck (1986) deden een poging ideeën uit de prestatie-motivatietheorie en de intrinsieke motivatietheorie te integreren in een nieuw begrip: doeloriëntatie. Zij maakten een onderscheid tussen twee soorten doeloriëntatie: nl. actuele doelen gericht op het bereiken van beheersing en actuele doelen gericht op het leveren van prestaties. Leerlingen die taaksituaties waarnemen als een uitnodiging of gelegenheid om een belangrijk leerdoel te bereiken streven leergeoriënteerde doelen na. Zij richten zich op de leertaak zelf en worden gemotiveerd door plezier in eigen-kunnen en de perceptie van vooruitgang. Leerlingen die overwegend prestatiegericht zijn willen daarentegen vooral hun competentie bewijzen, en vermijden dat anderen hun onzekerheid, fouten en inzet zien. Nicholls (1984) liet zien dat leerlingen die ego-georiënteerd zijn fundamenteel anders reageren t.a.v. fouten, feedback, hulp zoeken en inzet dan leerlingen die leergeoriënteerd zijn.

* Adres: Postbus 9555, 2300 RB Leiden.

De vakattitude van leerlingen kan worden gezien als een specifieke vorm van motivatie. Vakattitudes, of de tendens om positief, neutraal of negatief te reageren t.a.v. informatie in een bepaald vakgebied ontstaan vrij vroeg. Helmke (1992) toonde in een longitudinaal onderzoek aan dat in het eerste leerjaar van het basisonderwijs de meeste leerlingen een positieve attitude hebben t.a.v. taal en rekenen. Plezier in het vak bleef op peil in de eerste klas, maar kreeg een dip in de tweede klas, vooral bij meisjes. Het is echter interessant daarbij te vermelden dat de leerlingen die aangaven weinig plezier in rekenen te hebben niet lager scoorden op intelligentietesten. Helmke liet eveneens zien dat de vakattitude t.a.v. rekenen bij het eind van de tweede klas een goede voorspeller was voor de rekenprestaties in de derde klas. Omgekeerd hadden leerlingen die bij het eind van de tweede klas hoge cijfers hadden voor rekenen ook meer plezier in rekenen bij het eind van de derde klas. Plezier beleven aan schoolse taken blijkt ook voor oudere leerlingen een belangrijke motivator te zijn, terwijl negatieve gedachten en gevoelens, vooral angst, remmend werken. Wigfield en Meece (1988) stelden in dit verband vast dat leerlingen die een laag beeld van eigen-bekwaamheid t.a.v. rekenen hebben maar het vak als nuttig en relevant definiëren, meer last hebben van remmende angst dan leerlingen die een laag beeld van eigen-bekwaamheid hebben maar het vak niet zinvol achten. Het is dus vooral het twijfelen aan eigen-kunnen en zich daarover zorgen maken dat het verlamme effect produceert.

Seegers en Boekaerts (1993) toonden aan dat positieve gedachten en gevoelens met betrekking tot leertaken de leerintentie bevorderen. Negatieve gedachten en gevoelens, vooral twijfel aan eigen kunnen, maken leerlingen minder inzet bereid en sorteren een rechtstreeks negatief effect op de leerprestaties.

ZELFGERELATEERDE VARIABELEN GEMETEN OP VERSCHILLENDE NIVEAUS

Tegenwoordig gaan de meeste onderzoekers ervan uit dat zelfgerichte gedachten en gevoelens het doelgericht handelen van leerlingen sterk kunnen beïnvloeden, en dat deze variabelen op verschillende niveaus kunnen worden gemeten. Cantor (1981) maakte een onderscheid tussen drie niveaus: nl. (1) het algemeen of trekniveau, (2) het midden niveau en (3) het situatie-specifieke niveau. Het eerste meetniveau verwijst naar het meten van persoonlijkheidskenmerken, zoals bijvoorbeeld de motivatie die leerlingen hebben t.a.v. allerlei activiteiten met inbegrip van leren op school. Naast motivatie zijn er tal van andere trekken die bij leerlingen op het algemeen niveau kunnen worden gemeten, zoals bijvoorbeeld het beeld wat leerlingen hebben t.a.v. hun bekwaamheid, hun faalangst, en hun sociale angst.

Het tweede meetniveau refereert aan het meten van dezelfde soort eigenschappen maar nu t.a.v. een specifiek deelgebied, bijvoorbeeld het vakgebied wiskunde of taal. Voorbeelden van variabelen gemeten op het midden niveau zijn interesse in natuurkunde, het beeld van eigenbekwaamheid t.a.v. sport, het toekomstperspectief t.a.v. vrije tijd, en faalangst voor wiskunde. Het derde meetniveau valt samen met de actuele situatie. Zelfgerelateerde variabelen gemeten in een leersituatie geven meer gedetailleerde informatie over de gedachten en gevoelens die bij leerlingen opkomen voor, tijdens en na het leerproces. Zo wordt bijvoorbeeld de leerbereidheid van leerlingen gemeten t.a.v. een specifieke leertaak, maar ook hun competentieoordeel t.a.v. die taak, en hun twijfel of het wel zal lukken.

Sommige onderzoeksgroepen maken een keuze voor één meetniveau, terwijl anderen variabelen meten op verschillende meetniveaus. Zo zijn er studies die de algemene motivatie of prestaties van leerlingen proberen te verklaren vanuit hun algemene bekwaamheid, toekomstperspectief, of faalangst. Andere studies situeren zich op het middenniveau en concentreren zich op één vakgebied, bijvoorbeeld wiskunde. Nog andere studies meten zowel de motivatie als de prestatie in concrete leersituaties en bestuderen de invloed van trekmaten en van variabelen gemeten op het midden niveau op de leerintentie en -prestaties van leerlingen (zie ook Boekaerts, 1994).

VERKLAREN VAN LEERMOTIVATIE EN -PRESTATIES

Dit themanummer gaat over de invloed van affectieve variabelen op leermotivatie en -prestaties. Het bestaat uit vijf artikelen die een thematische eenheid vormen. De keuze voor deze artikelen werd gemaakt op inhoudelijke en methodologische gronden. Ze vormen een eenheid in die zin dat ze zelfgerelateerde cognities en affect meten. Elke onderzoeksgroep bestudeert specifieke theoretische constructen en heeft de uiteindelijke bedoeling de onderlinge relaties tussen deze nauw verwante constructen bloot te leggen en hun effect op de leermotivatie en -prestaties te beschrijven en te verklaren. De artikelen vormen ook een methodologische eenheid omdat ze gebruik maken van dezelfde geavanceerde technieken, nl. lineaire structurele modellen (LISREL) om de onderliggende samenhang tussen de variabelen te analyseren. Het is de bedoeling van de diverse onderzoeksgroepen om mee te werken aan de ontwikkeling van krachtige verklaringsmodellen die de impact en de relatieve bijdrage meten van cognitieve en affectieve variabelen op leermotivatie en -prestaties.

Met deze vijf artikelen willen we de oningewijde lezer een beeld geven van de verscheidenheid van affectieve variabelen die momenteel worden bestudeerd. De ingewijde lezer willen wij verslag doen van de stand van zaken binnen de diverse theoretische stromingen voor zover die zich ten onzent richten op affectieve variabelen én LISREL modellen hanteren.

In het eerste artikel, gaan Crombach, Voeten en Boekaerts uit van een structureel model dat door Eccles en haar medewerkers werd geconstrueerd om het keuzegedrag voor wiskundevakken bij leerlingen van de middelbare school te verklaren met behulp van affectieve variabelen gemeten op het midden niveau. Nagegaan wordt of gelijksoortige affectieve variabelen, gemeten op het situatie-specifieke niveau, de leerintentie van basisschoolleerlingen beïnvloedt voor verschillende curriculaire taken. Met andere woorden, de auteurs onderzoeken of het structureel model van Eccles ook opgaat wanneer zelfgerelateerde variabelen in actuele leersituaties worden bestudeerd en leerintentie voor concrete reken-, teken- en leestaken de afhankelijke variabelen zijn. In een reeks van opeenvolgende stappen passen ze het uitgangsmodel aan en toetsen ze of er verschillen moeten worden aangenomen voor diverse vakgebieden voor zowel jongens als meisjes. De conclusie wordt getrokken dat het model, zoals voorgesteld door Eccles voldoende robuust is om de onderliggende mechanismen op het situatie-specifieke niveau te beschrijven.

Moreas en Lens meten de attitude die leerlingen hebben t.a.v. van het heden en t.a.v. de toekomst en hun faalangst op het algemene of trekniveau. Zij bestuderen de invloed van deze trekmaten op de studiemotivatie en de -resultaten. De doelgroep bestaat uit leerlingen van het secundair onderwijs in Vlaanderen. Met behulp van structurele modellen wordt onderzocht of deze variabelen rechtstreeks de studiemotivatie en de -resultaten beïnvloeden. De auteurs rapporteren dat hoe positiever de attitude van leerlingen is t.a.v. het heden, hoe meer ze gemotiveerd zijn om te studeren en hoe beter hun prestaties zullen zijn. Deze attitude is vooral gebaseerd op het persoonlijk belang dat ze toekennen aan het heden, aan hun engagement, en aan hun oordeel over de moeilijkheidsgraad. Verder blijkt dat er geen verschillen zijn qua tijd die faalangstige en niet faalangstige leerlingen aan hun schoolwerk besteden. De studie-efficiëntie van de faalangstige leerlingen is echter inferieur aan die van de niet-faalangstigen. Het verband tussen faalangst en studiemotivatie verloopt echter niet rechtstreeks, maar indirect via de attitude t.a.v. het heden. Hier geldt dat hoe faalangstiger de leerlingen zijn des te negatiever hun attitude is t.a.v. het heden en een negatieve attitude t.a.v. het heden leidt tot lagere studiemotivatie.

In het artikel van Peetsma wordt motivatie gemeten op het midden niveau. Er wordt uitgegaan van het toekomstperspectief dat leerlingen hebben t.a.v. vier deelterreinen: nl. school en beroep, sociale relaties, eigen ontwikkeling, en vrije tijd. De vraag die wordt gesteld is of leerlingen die een positief toekomstperspectief hebben t.a.v. deze vier deelterreinen zich meer inzetten voor school dan leerlingen met een negatief toekomstperspectief. Vastgesteld wordt dat het toekomstperspectief object gebonden is: het toekomstperspectief t.a.v. vrije tijd werkt averechts op de inzet voor schooltaken. De drie andere toekomstperspectieven hebben tezamen een

positief effect op de schoolse inzet. Dit geldt ongeacht schooltype, geslacht van de leerlingen, hun faalangstscore en de schoolfase waarin ze zich bevinden. Alleen de mate waarin leerlingen denken controle te hebben over hun prestaties beïnvloedt het verband tussen toekomstperspectief en inzet. Peetsma geeft hier echter geen verklaring voor.

In het artikel van Otten, Boekaerts en Seegers wordt de handelingscontrole van brugklassers nader onder de loep genomen. Deze auteurs meten de controle die leerlingen uitoefenen op hun alledaagse handelingen op het algemene niveau. Meer specifiek meten ze of leerlingen in de startfase van dagelijkse activiteiten (zoals bijv. boodschappen doen) snel een gedragsintentie kunnen koppelen aan een handelingsplan (initiatief), of ze in de uitvoeringsfase deze koppeling doorzetten (persistentie); en of ze in de evaluatiefase de gedragsintentie kunnen loskoppelen van het handelingsplan zonder dat negatieve emoties gaan overheersen (ontkoppeling). Verder gaan ze na of de drie aspecten van handelingscontrole een invloed uitoefenen op de wijze waarop leerlingen concrete leertaken beoordelen (appraisals), op hun leerintentie en op de gerapporteerde inzet na afloop van specifieke leertaken. De auteurs stellen verschillen vast in de gemiddelde scores op de diverse aspecten van handelingscontrole tussen brugklassers in de HAVO/VWO stroom en leerlingen in de andere brugklastypen. Er werden echter geen verschillen geconstateerd in het onderliggende mechanisme: de invloed die van handelingscontrole uitgaat op de appraisals, de leerintentie en op de gerapporteerde inzet verloopt op dezelfde wijze voor alle brugklassers.

In de laatste bijdrage, wordt de stress die leerlingen op school ondervinden gemeten. Boekaerts en Seegers vroegen leerlingen de Stress en Copingvragenlijst in te vullen. Aan de hand hiervan kon de frequentie van meemaken en de intensiteit van de ervaren stress met betrekking tot zeven dagelijkse ergernissen worden vastgesteld. De wijze waarop brugklassers met deze zeven stressoren omgaan werd eveneens in kaart gebracht. Met behulp van structurele modellen werd onderzocht of dagelijkse stressoren (maar ook trauma's meegemaakt in het afgelopen jaar) de gerapporteerde copingstrategieën beïnvloeden en of ervaren sociale steun hierbij een moderator effect heeft. Vastgesteld werd dat de frequentie van meemaken van stressvolle levensgebeurtenissen en dagelijkse beslommeringen agressieve copingvormen bevordert. De intensiteit van ervaren stress hangt samen met de beide andere copingmodi (toenadering en vermijding). Waargenomen sociale steun heeft een beperkt modererend effect. Prestaties en prestatiemotivatie worden echter niet of nauwelijks beïnvloed door stress en coping. De onderliggende structuur van de variabelen vastgesteld bij brugklassers bleek na twee jaar ongewijzigd te zijn.

LITERATUUR

- Boekaerts, M. (1994, in druk). The interface between intelligence and personality as determinants of classroom learning. In D.H. Saklofske & M. Zeidner (Eds.). *Handbook of personality and intelligence*. New York: Plenum Press.
- Cantor, N. (1981). Perceptions of situations. In D. Magnusson (ed.). *Toward a psychology of situations. An interactional perspective* (pp. 229-244). Hillsdale, NJ: Erlbaum.
- Dweck, C.S. (1986). Motivational processes affecting learning. *American Psychologist*, 41, 1040-1048.
- Helmke, A. (1989). Affective student characteristics and cognitive development: Problems, pitfalls, perspectives. *International Journal of Educational Research*, 13(8), 915-932.
- Helmke, A. (1992). *The Development of children's attitude towards learning in elementary school: A longitudinal study*. Paper presented at the XXV International Congress of Psychology, July 19-24. Brussels.
- Marsh, H.W., Walker, R., & Debus, R. (1991). Subject-Specific Components of Academic Self-Concept and Self-Efficacy. *Contemporary Educational Psychology*, 16, 331-345.
- Nicholls, J.G. (1984). Achievement motivation: Conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91, 328-346.
- Seegers, G., & Boekaerts, M. (1993). Task motivation and mathematics achievement in actual task situations. *Learning and Instruction*, 3, 133-150.
- Wigfield, A., & Meece, J. (1988). Math anxiety in elementary and secondary school students. *Journal of Educational Psychology*, 80, 210-216.

A model for explaining individual differences between students in intended effort on curricular tasks

Marjo J. Crombach and Marinus J.M. Voeten
*Department of Educational Sciences, University of Nijmegen**

Monique Boekaerts
Centre for the Study of Education and Instruction, University of Leiden

ABSTRACT

In this study we specify and test a model about the way in which appraisals, in this case task-specific cognitions, determine the student's intended effort for a curricular task. The model is based on the work of Eccles with respect to mathematics achievement and on Boekaerts' model for adaptable learning. The model was successfully fitted to data on three curricular tasks: a reading, an arithmetic and a drawing task. Approximately 1,000, mainly sixth grade Flemish students completed these tasks and a questionnaire (the OMQ) which assessed the students' appraisals at the onset of the task as well as their intention to spend effort on the task. The appraisal model appeared to be adequate for each task. There were only minor differences between the tasks. The models seemed to be structurally invariant for boys and for girls. Only a modest amount of the variance of intended effort could be explained (varying from 21% to 32%). Results were largely similar to the results obtained by Eccles and her co-workers in the area of mathematics achievement.

INTRODUCTION

In this study we specify and test a model about the way in which task-specific cognitions determine the student's intended effort for a curricular task. As a point of departure in developing such a model, we took the relevant part of a model for achievement behaviours constructed by Eccles (Parsons) and her co-workers (Eccles et al., 1983, 1985; Eccles, Adler, & Meece, 1984; Meece, Eccles, Kaczala, Goff, & Futterman, 1982). We will first present the Eccles' model. Then, we will outline our own theoretical framework and present a working model that will serve as a guide for formulating a more final model. Finally, we will fit this working model to our data, allowing for certain modifications when necessary, and we will test whether the model is adequate for boys as well as for girls.

The Eccles' Model

Eccles et al. (1983, 1984, 1985) presented the *General Model for Achievement Behaviors* in order to explain individual differences in achievement behaviours, especially differences between male and female students. This model consists of two components. The first is a psychological component in which the effects of various cognitive factors at one point in time are specified. The second component is a developmental one; it specifies the development of individual differences in these cognitive factors. The model, a cognitive mediation model, is built on the assumption that 'it is not reality itself (i.e., past successes or failures) that most directly determines children's expectancies, values, and behavior, but rather the interpretation of that reality' (Eccles et al., 1983, p. 79-81). The origin of the Eccles' model lies in the expectancy x value models (Atkinson, 1957, 1964; Heckhausen, 1977, 1980; for a review see Spence &

*Adres: Postbus 9104, 6500 HE Nijmegen.

Helmreich, 1983). This implies that expectancy and value are the central constructs in the model, and in fact they are the only constructs directly influencing achievement behaviours. Other constructs can only indirectly influence achievement behaviours via expectancy or value. For example, students' expectancies and their perceptions of task value are supposed to be determined by their goals, their self-concept of ability, and their perception of task difficulty. In addition, several exogenous variables were included in the model: the students' perceptions of parents' and teachers' attitudes and expectations, the students' beliefs about appropriate role characteristics, and the students' causal attributions of past performance. The focus of the Eccles' study was on achievement behaviours with respect to a particular school subject, namely mathematics. More specifically, the students' intention to enroll in mathematics courses was investigated. The majority of the sample in Eccles' study consisted of junior high school students.

Eccles et al. (1983) concluded that their psychological model provides a good representation of the data, especially for high school students. Ethington (1991) carried out a replication study and simultaneously extended the work of Eccles et al. She carried out path analyses separately for male and female eighth grade students and paid attention to direct as well as to indirect effects on intention to enroll in mathematics courses. She concluded that her results strongly support the importance of the constructs identified by Eccles et al. as critical determinants of the intention to take more mathematics in their study program.

The main goal for Eccles et al. to start their extensive research was to explain the unequal enrollment of male and female students in more advanced courses in high school mathematics, while mathematics is crucial for many career and job opportunities in today's increasingly technological society (Eccles et al, 1985; Meece et al., 1982; Meece, Wigfield, & Eccles, 1990). With respect to sex differences, Eccles et al. (1983, p. 115) concluded that no single construct emerged as the crucial mediator of sex-differentiated math achievement behaviours. However, there were small but consistent sex differences on several constructs (e.g., self-concept of ability, perception of task difficulty, expectancy). Possibly, these constructs could mediate differential achievement behaviours in combination. Since the model was not estimated separately for males and females, it is not known whether the relations between the constructs are the same for both sexes. Ethington's (1991) results suggested that the model works differently for males and females. Meece et al. (1990) took the work of Eccles et al. (1983, 1985) as a starting point, and investigated models with less constructs. They applied the LISREL program to test the invariance of the covariance matrices of male and female junior high school students. Their results suggested that the pattern of relations is similar for males and females. Consequently, they concluded that sex differences in course enrollment and performance patterns are due more to the mean differences in males' and females' achievement-related perceptions than to differences in the pattern of relationships among the constructs.

In the next section we will outline our theoretical framework in order to end up with a working model to explain the students' intention to invest effort for accomplishing their curricular assignments. The model will be tested using data concerning curricular assignments within three different school subjects.

Theoretical Framework

As compared with Eccles and her co-workers, Boekaerts and her co-workers have narrowed down the domain of achievement behaviours to learning intention for concrete curricular tasks. Following Cantor (1981), who made a distinction between perceiving situations at the superordinate, the middle, and the subordinate level, Boekaerts (1987a) introduced the terms superordinate, middle, and momentary level of motivation. At the superordinate level general motivation for school learning is measured, at the middle level the students' motivation for a specific school subject is assessed, and at the momentary level motivation with respect to concrete tasks is measured. From this point of view, Eccles' model is a model developed and tested at the middle level (intention to take more math courses), whereas our model is situated at the momentary level (intention to invest effort in a specific task).

The model for adaptable learning formulated by Boekaerts (e.g., Boekaerts, 1985, 1988, 1992) has continuously been tested, revised and elaborated over recent years. She based this model on existing theories of achievement motivation (Atkinson, 1964; Atkinson, & Feather, 1966; Heckhausen, 1977, 1980) and intrinsic motivation (Deci, 1975; deCharms, 1984; Harter, 1978; Lepper and Greene, 1978). In addition, she gave a central role to appraisal processes. In doing so, Boekaerts relied heavily on the ideas of Lazarus and his associates (Lazarus, 1991a, 1991b; Lazarus & Folkman, 1984). Lazarus made it clear that appraisal processes arise when persons are confronted with new or changing situations. Boekaerts applied this idea explicitly to learning situations in the classroom. She reasoned that confronting students with a task means that cognitions and affects related to that task and the situation within which it is embedded are activated. Following Lazarus, she called those task-specific cognitions and affects 'appraisals'. The self-report instrument, that she constructed, viz. the On-line Motivation Questionnaire (OMQ) (Boekaerts, 1986, 1987b, 1987c), measures these appraisals and intended effort, as well as result assessment and attributions. Four cognitive appraisals identified by Boekaerts, namely self-efficacy (SEff), perceived difficulty (PDiff), success expectancy (SE), and perceived utility (PU), are the focus of this study. These appraisals allow for a comparison with variables in the Eccles' model.

In order to elaborate the appraisal part of Boekaerts' model for adaptable learning, we started by developing a model for the relations between the four appraisals mentioned above and for the effects of these appraisals on intended effort (IEffo), also measured by the OMQ. We specified the relations between these constructs in much the same way as Eccles did. In this study, we tested whether this model, henceforth referred to as the appraisal model, fits the data obtained by the 1985-version of the OMQ (Boekaerts, 1987b). In subsequent studies we will further elaborate the appraisal model by adding additional constructs and ideas as emphasized in Boekaerts' model of adaptable learning.

Model Specification

Eccles used, among others, the following constructs: (1) self-concept of ability, (2) perception of task difficulty, (3) success expectancy, (4) perception of task value, and (5) intention to take further courses in mathematics (see the first column of Table 1). All these constructs are defined at the level of the individual student. In the second column of Table 1 the comparable terms as used in the appraisal model, are mentioned. The relations between the four appraisals and their presumed effects on intended effort are outlined in Figure 1.

In line with the expectancy x value model, success expectancy and perceived utility have a direct influence on intended effort; self-efficacy and perceived difficulty, which may be correlated as indicated by the curved arrow in Figure 1, only indirectly influence intended effort. Expectancy and value are hypothesized to have the same determinants. The relationship of success expectancy with self-efficacy and perceived difficulty has been well established in the literature, but less research has been done on task value (Eccles et al., 1983; Wigfield & Eccles, 1992). With respect to task value Eccles et al. (1983) suggested that "the value of a particular task to a particular person is a function of both the perceived qualities of the task and the individual's needs, goals, and self-perceptions" (o.c., p. 90). The variables included in our model only partly cover the antecedents of task value implied by this wide-ranging quotation from Eccles et al. In this study, we depart from a more narrow concept of task value, called perceived utility (see below).

In agreement with Eccles et al. (1983) we did not specify a relationship between expectancy and value in our model. From the work of Atkinson (1964) one would expect an inverse relationship between expectancy and value. According to Wigfield and Eccles (1992), however, expectancy and value are positively correlated. In our first version of the model we suppose that this correlation can be explained by common causes, i.e. self-efficacy and perceived difficulty of the task.

When comparing the two columns of Table 1, it is evident that the Eccles' constructs are not identical with the appraisals as measured with the OMQ. This is due to the different focus of

Table 1. Variables in the Appraisal Model Compared with Variables of Eccles' Model.

Psychological component of Eccles' model for mathematics (cf. Eccles et al., 1983, fig. 2-2(b) p. 101)	Appraisal Model
Student's self-concept of math ability	Self-efficacy with respect to the present task (Seff)
Student's perception of difficulty of math as a school subject	Perceived Difficulty of the present task (PDiff)
Student's expectancies 1. Current 2. Future	Success Expectancy for the present task (SE): <i>current</i>
Student's perception of task value 1. Liking of math 2. Perceived usefulness of math 3. Importance of doing well in math 4. Worth of amount of effort needed to do well in math	Perceived Utility of the present task (PU): <i>perceived usefulness of the task</i>
Intention to take a further course in mathematics	Intended Effort on the present task (IEffo)

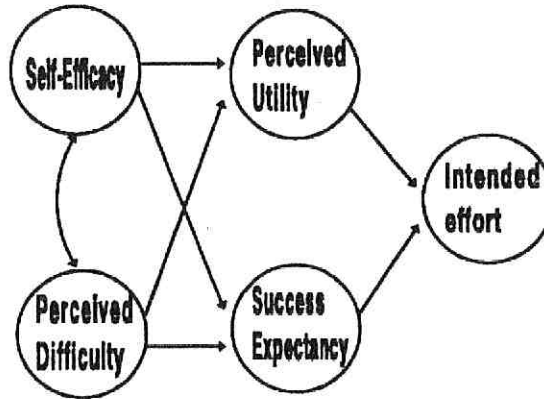


Fig. 1. The Appraisal Model.

each study. In general, the Eccles' constructs are broader than our constructs. For the constructs self-concept of ability and task difficulty the different operationalizations are merely a matter of perspective: middle level versus momentary level. In Eccles' model the dependent variable is defined as a long-term intention to enroll in math courses, while our model tries to explain the short-term intention for putting effort into a curricular task. A more important difference is the following. Eccles et al. distinguished between current and future success expectancies for math. We use success expectancy for a particular task, which approximately corresponds to Eccles' aspect of current success expectancy. Therefore we placed 'current' in italics below the appraisal of success expectancy in Table 1.

As stated before our concept of perceived utility is more narrowly conceived than the concept of task value as used by Eccles. She distinguished four major components of task value: the attainment value, the intrinsic or interest value, the utility value, and the costs (Eccles et al., 1983; Wigfield and Eccles, 1992). The only aspect of task value represented in our appraisal

model at this moment is the student's appraisal of the utility value of the task. In a forthcoming study (Boekaerts, Crombach & Voeten, in preparation) we will extend the appraisal model with an intrinsic value component, task attractivity. The intrinsic value of a task is the immediate pleasure experienced in performing the task, while the utility value is determined by future goals. From analyses not reported here, it was concluded that perceived utility and attractiveness cannot be considered as indicators of the same construct, viz., 'task value'. Wigfield & Eccles also (1992) referred to empirical work that showed utility, attainment and interest to be separate components. Boekaerts (1987b, 1987c) argued that the intrinsic value, the liking of a task, plays a crucial role in becoming involved in a task. It is our view that the intrinsic-value component deserves a separate position in a motivation model as a determinant of intended effort, especially in a model focused at task-specific motivation. However, in this study, we did not include task attractivity in the appraisal model. To disentangle the intrinsic and extrinsic motivation aspects of a task, we first developed an appraisal model working within the achievement motivation framework. Compare Meece et al. (1990), who also used utility (importance) as the only component of task value in their models. In Boekaerts et al. (in preparation) we study the intrinsic value component.

Because it was the intention to measure motivation in a task-specific way, the appraisal model was applied to data from separate tasks. OMQ-data with respect to three different curricular tasks were available, viz., a reading task, an arithmetic task, and a drawing task. It was tested whether the same model would fit data stemming from these three tasks. As an alternative the interrelations between the appraisals, or the effects of the appraisals on intended effort, could differ for the three tasks, presumably because of the respondents' attitude toward different school subjects.

The exogenous variables available in this study are different from those in the Eccles' study. We included only the student's sex and the teacher's judgement of the student's task-specific ability (TJSA) as exogenous variables. It is assumed that the teacher's judgement of ability is communicated to the students through his or her behaviour in the classroom and through grades students receive, and that it forms a frame of reference for the students' perception of their ability.

Our main goal is to gain more insight into appraisal processes. Hence, successive models will be formulated in which the interrelations between specific cognitions, affects, and intended effort are specified. The appraisal model, presented in Figure 1, is the first step in developing such a model. A subsequent goal is to investigate whether appraisal processes are sensitive to sex differences. In this article we will answer the following questions: (1) Is the appraisal model (Figure 1) adequate to describe the data for each specific task, or are task-specific modifications necessary? and (2) Is the appraisal model (or a modified version of it for a particular task) adequate for boys as well as for girls?

METHOD

Subjects

Subjects were approximately 1,000 Flemish fifth and sixth grade students. The age of these students ranged from 10 to 12 years. Nineteen percent of the students was from the fifth grade. The students filled in the OMQ three times, namely for each of the three tasks: reading (547 boys and 413 girls), arithmetic (548 boys and 440 girls), and drawing (539 boys and 440 girls). The data were collected in 1985 as a part of a larger study carried out by Boekaerts (1986, 1987c).

Materials

The On Line Motivation Questionnaire

The questionnaire consists of two parts (Boekaerts, 1987b). The variables included in the ap-

praisal model were measured by the first part of the questionnaire, which was administered just before starting a task. Self-efficacy was measured by two items, the other variables (Pdiff, SE, PU, and IEffo) by one item. The text of these items is given in the Appendix. In addition to the items used for this study, the first part of the questionnaire also contained questions on current affects, not analysed here. The second part was filled in after task completion; questions were asked about perceived result assessment, causal attributions for success or failure, and affects after the task. Students were trained to complete the first part of the questionnaire just after the introduction of a task at the teacher's signal. On finishing the task, students knew they had to proceed immediately with the second part of the questionnaire. But, the data from the second part were not used in this study. All items were in the format of four-point Likert scales. In the 1985-version of the OMQ all items had an additional fifth answer opportunity: 'I do not know'.

Description of the three tasks

Teachers introduced the tasks as part of the curriculum during normal classroom interaction. The *reading* task was a reading comprehension task. The students read a text and answered 20 questions. The *arithmetic* assignment involved making a shopping list. The students had to do a number of calculations (+, -, x, :) in order to get subtotals and totals. The *drawing* task requested students to illustrate a story they were told. For a further description of the tasks see Boekaerts (1987c, p. 173).

Teacher judgement

The classroom teacher's judgements of student ability in the three respective domains were obtained on a five-point scale. These teacher ratings were collected together with the teacher's judgement of student overall motivation. In this study, only the teacher judgements of student ability (TJSA) were used.

Procedures for Analysis

Preliminary analyses

The items of the OMQ were scored from 1 to 4 (see Appendix). Students who did not provide an answer to two or more of the items, were excluded. We did the same for students who gave the answer 'I do not know' two or more times. In order to investigate the value of the do-not-knows and the missings relative to the four Likert-type categories, we carried out Homals analyses (Crombach, Boekaerts, & Voeten, in preparation). From these analyses it was concluded that do-not-know answers as well as missing answers can be placed in the middle of the four-point scale. Therefore, we randomly assigned the values 2 or 3 to the remaining do-not-knows and missings.

The Lisrel analyses

To answer the research questions, we used the Lisrel program incorporated in SPSSX. Each construct in the appraisal model was measured by just one item, except self-efficacy for which two items were used as indicators. The model was applied to the data and was modified stepwise when necessary. Such model modification makes cross-validation of the resulting model desirable (MacCallum, Roznowski, & Necowitz, 1992). Therefore, we divided the total group of students into two equivalent groups: an exploration sample and a validation sample. This was done separately for each task. We assigned the students, who were sorted by school and class number, alternately to these two samples.

We carried out Lisrel analyses for each task separately, starting with the reading task because this task had been analyzed most extensively. First, we modified a model without interaction effects for sex on the exploration sample. Next, we fitted four models simultaneously for both sexes, increasingly allowing for sex differences. In the cross-validation procedure these four models were fitted to the validation sample. Finally these models were fitted to the complete sample. For each task the same series of Lisrel analyses was carried out on the covariance matrices. The following steps were taken:

1. *Model modification.* We fitted the appraisal model (see Figure 1) to the exploration sample for the data on the reading task. A model was always modified in a stepwise way by introducing extra free parameters (one at a time). These parameters were selected on the basis of their modification indices, with the restriction never to free a parameter that would violate the theoretical presumptions. Next, the exogenous variables sex and teacher judgement of student ability were included. These variables were expected to influence the first two appraisals only (viz., self-efficacy and perceived difficulty). From this starting point we again modified the model stepwise until it fitted.

2. *Interaction effects with sex.* The model found in Step 1 was tested for differences between boys and girls by simultaneously fitting the model to the covariance matrices for boys and for girls. A sequence of models was formulated in such a way that each model allowed increasingly for sex differences. In the first model (Model a) it was assumed that all parameter matrices were equal for boys and for girls. In the second model (Model b) sex differences were allowed in the residual variances of the endogenous variables. In the third model (Model c) sex differences in the covariances between the residuals of the endogenous variables were also allowed for. Note that the structural part of the model is still equal for both sexes in Models a, b, and c. Finally, in the fourth model (Model d) the values of the parameters in the structural model were allowed to vary for boys and girls. Model d is the only model that allows interaction effects for sex. It should be noted that the measurement model remained invariant for boys and for girls in all analyses.

3. *Cross-validation.* The Models a, b, c, and d (as defined in Step 2) were applied to the validation sample, fixing all parameters at the values estimated from the exploration sample in Step 2. The likelihood ratio chi-square was used as a cross-validation index. This index is a kind of distance measure between two variance-covariance matrices: the reproduced matrix obtained from the model developed on the exploration sample and the matrix observed in the validation sample (Cudeck & Browne, 1983).

4. *Selection of the final model, testing sex differences, and parameter estimation for the complete sample.* Models a, b, c, and d were fitted to the complete sample. The differences in chi-squares and degrees of freedom for these four models were compared. In order to choose a final model, differences in chi-squares for the four models as well as the cross-validation index were taken into account. In addition, the Bayesian Information Criterion (BIC) as proposed by Raftery (1993), based on the complete sample, was applied. The model with the lowest value of BIC is to be preferred; the value of BIC should in any case be less than zero.

The procedure described above was repeated for the three tasks separately, using the model found in Step 1 for the reading task as a starting model.

RESULTS

Table 2 gives the means and standard deviations of the variables used in this study for boys and girls separately. In most cases, the means and standard deviations indicated only slight differences between boys and girls. There were, however, some significant differences. Girls scored higher on intended effort than boys for each task. Girls reported a higher perceived utility for the reading and the arithmetic task than boys. With regard to the arithmetic task, boys scored higher on self-efficacy and perceived the task as easier than girls. Table 3 shows the correlations among the variables for each task. Correlations are given separately for boys and for girls.

In this section only final results of the process of Lisrel model modification and estimation are presented, starting with the description of the results for the reading task. Next, the drawing task, and finally the arithmetic task are described. Figures 2, 3, and 4 show the final Lisrel models with standardized parameter estimates and goodness of fit measures for the reading, the drawing, and the arithmetic task, respectively. Only the structural part of the model is represented in the figures, omitting the measurement model. Note that except for self-efficacy there is only one indicator for each latent variable.

Table 2. Means and Standard Deviations of the Variables for Boys and for Girls per Task.

	Reading Task				Arithmetic Task				Drawing Task			
	Boys (n=547)		Girls (n=413)		Boys (n=548)		Girls (n=440)		Boys (n=539)		Girls (n=440)	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
TJSA	3.40	1.06	3.37	1.03	3.38	1.17	3.33	1.10	3.48	0.89	3.71*	0.84
Seff1	2.83	0.73	2.77	0.68	2.95	0.69	2.79*	0.71	3.00	0.81	2.96	0.76
Seff2	2.77	0.73	2.67*	0.66	2.91	0.69	2.68*	0.68	2.97	0.79	2.87	0.74
PDiff	2.15	0.84	2.11	0.79	2.01	0.77	2.28*	0.81	1.78	0.81	1.84	0.80
SE	3.19	0.67	3.25	0.66	3.26	0.68	3.22	0.64	3.26	0.69	3.28	0.66
PU	2.76	0.79	2.92*	0.68	3.05	0.80	3.15*	0.70	2.64	0.90	2.68	0.81
IEffo	3.20	0.74	3.39*	0.64	3.23	0.74	3.37*	0.66	3.30	0.76	3.41*	0.65

* Significant difference between boys and girls ($p < .05$).

Table 3. Correlations Between the Variables for Boys and for Girls per Task.

READING TASK							
	TJSA	Seff1	Seff2	PDiff	SE	PU	IEffo
TJSA	-	.24	.25	-.09	.14	.07	.14
Seff1	.22	-	.78	-.46	.48	.32	.24
Seff2	.30	.73	-	-.47	.50	.31	.22
PDiff	-.12	-.54	-.46	-	-.31	-.09	-.04
SE	.19	.40	.40	-.29	-	.37	.34
PU	.08	.27	.21	-.15	.26	-	.45
IEffo	.18	.17	.13	-.06	.35	.31	-

Correlations above the main diagonal are for boys (N=547), below the main diagonal for girls (N=413).

ARITHMETIC TASK							
	TJSA	Seff1	Seff2	PDiff	SE	PU	IEffo
TJSA	-	.38	.39	-.20	.25	.10	.17
Seff1	.28	-	.71	-.47	.48	.23	.23
Seff2	.35	.74	-	-.45	.51	.24	.28
PDiff	-.23	-.58	-.59	-	-.30	-.08	-.02
SE	.12	.36	.39	-.26	-	.33	.44
PU	.06	.11	.07	-.06	.24	-	.46
IEffo	.12	.15	.16	-.03	.37	.45	-

Correlations above the main diagonal are for boys (N=548), below the main diagonal for girls (N=440).

DRAWING TASK							
	TJSA	Seff1	Seff2	PDiff	SE	PU	IEffo
TJSA	-	.25	.26	-.04	.18	.12	.13
Seff1	.25	-	.84	-.43	.55	.42	.29
Seff2	.28	.79	-	-.46	.55	.42	.29
PDiff	-.20	-.58	-.55	-	-.36	-.18	-.15
SE	.13	.53	.55	-.41	-	.35	.36
PU	.11	.32	.35	-.26	.37	-	.38
IEffo	.13	.42	.41	-.27	.47	.45	-

Correlations above the main diagonal are for boys (N=539), below the main diagonal for girls (N=440).

Reading Task

From the first step of the model modification process it appeared that only one modification of the appraisal model (Figure 1) was needed to obtain a good fit, namely permitting a correlation between the residuals of success expectancy and perceived utility.¹ The teacher judgment of the student's reading ability appeared to have direct effects on self-efficacy and on perceived difficulty, but also on intended effort. There were no significant effects for sex. The model with the teacher judgement as the only exogenous variable provided a good fit. The next step was to test for interaction effects with sex (Step 2). The set of four nested models was tested, derived from the model obtained in Step 1, with Model a being the most parsimonious one. We compared the models by looking at the differences in chi-square. Models a, b, c, and d did all fit reasonably well (see Table 4, exploration sample). Model b (i.e., the model according to which only the residual variances differ for boys and for girls) differed significantly from Model a but not from Models c and d. In the cross-validation procedure (Step 3) the four models were fitted to the validation sample. The resulting chi-squares were all very satisfactory (see Table 4, validation sample). Model a showed the lowest chi-square but the chi-squares did not differ greatly, especially those for Models a, b, and c. In Step 4, the four models were fitted again, but now to the complete sample. Table 4 (see column complete sample) provides the relevant chi-squares. As can be seen, only Model a did not fit well. Models c and d did not differ significantly from Model b. Hence, in this comparison, Model b appears to be the best model. The BIC measures calculated on the complete sample are also shown in Table 4. Judging from this measure, Model a was slightly better than Model b. Considering all criteria, Model b was accepted as the final model for the reading data. Figure 2 shows this model. This model is only slightly different from the model depicted in Figure 1, with sex differences residing in the residual variances only.

It can be concluded that the appraisal model is adequate to describe the reading data. However, a few modifications were necessary. A correlation had to be included between the residuals of success expectancy and perceived utility. This correlation was small (.15) though. A substantial residual correlation (-.52) remained between self-efficacy and perceived difficulty, meaning that teacher judgement of ability is not very effective in explaining the correlation between these two appraisals. Not surprisingly, other exogenous variables will be needed. Finally, the appraisal processes could not fully explain the correlation between teacher judgement and intended effort. A direct effect of teacher judgement on intended effort had to be included. This effect, though, is only marginally significant. The variance of intended effort accounted for by the model is only about twenty percent (21% for boys and 22% for girls). Inspection of the parameter estimates in Figure 2 shows one nonsignificant coefficient, namely the effect of perceived difficulty on success expectancy. Other effects involving perceived difficulty also show low coefficients.

Finally, the model implies that there are no sex differences with respect to the structural relations between the variables. The only sex differences found had to do with the residual variances of the endogenous variables.

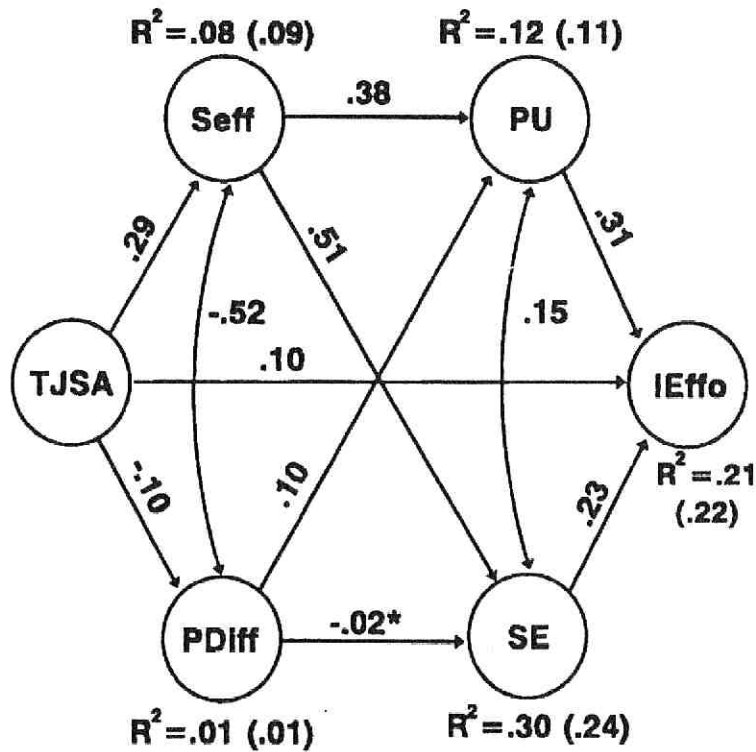
Drawing Task

The model resulting from Step 1 of the analyses for the reading task fitted the data for the drawing task well. The effect of teacher judgement of ability on intended effort was not significant at the 5% level, so the effect was removed from the model. As with the reading data, adding the student's sex to the model did not improve the fit. Thus, again, there were no main effects for sex. Next, the two-group analyses of sex differences were conducted (Step 2). Model a did not fit well in the exploration sample, while Models b, c, and d did (see Table 4). Model b did not differ significantly from Models c and d at the 1% level. In Step 3, the cross-validation, Model b showed the lowest chi-square (see Table 4). In Step 4 the four models were fitted to the complete sample (see Table 4). Only Model a did not fit well. Model b did not differ significantly from Models c and d. Thus, once again, Model b is the best choice. Model a showed the best BIC measure but the values for Models b, and c were nearly as good. Considering all criteria, Model b was accepted as the final model for the drawing data (see Figure 3).

Table 4. Goodness of Fit Statistics (Likelihood-Ratio Chi-Square and BIC) of the Various Models per Task, Simultaneously Fitted for Boys and Girls.

Task	Model	df	X ² Exploration Sample	X ² Validation Sample (df=56)	X ² Complete Sample	BIC
Reading Task	a	36	48.36	59.72	61.12	-256
	b	30	29.18	63.13	40.21	-224
	c	28	27.86	61.57	37.51	-209
	d	19	20.26	72.15	27.87	-140
Drawing Task	a	37	62.89	96.88	92.00	-235
	b	31	38.34	80.78	46.72	-227
	c	29	37.65	81.17	45.79	-210
	d	21	16.37	101.71	30.49	-155
Arithmetic Task	a	35	59.51	80.95	56.59	-253
	b	29	45.15	85.90	41.71	-215
	c	27	41.60	84.50	36.16	-203
	d	17	21.90	103.15	17.35	-133

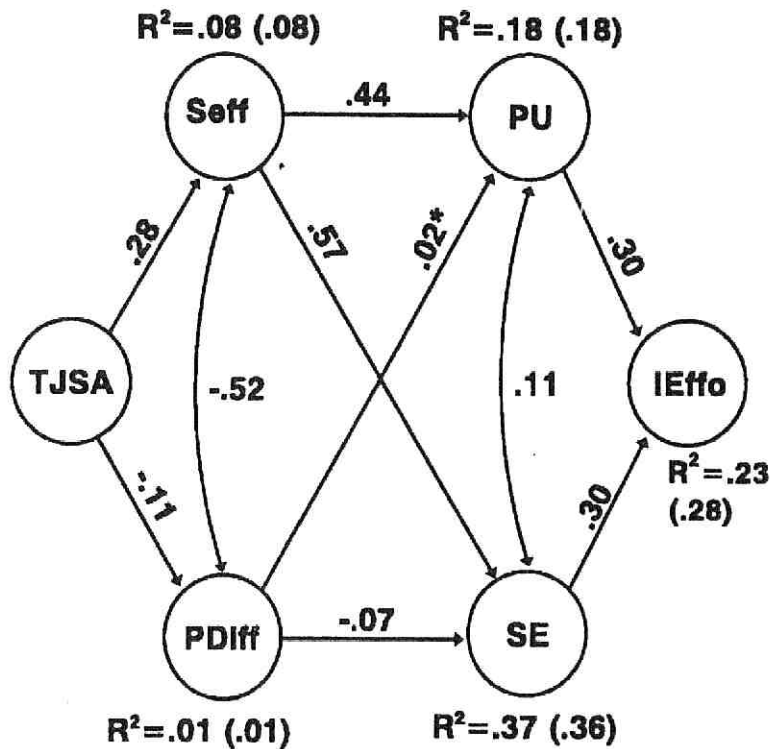
For the meaning of Models a, b, c and d see the Method section of the text.



Chi-Square: 40.21 (df=30, p=0.101)
 Goodness of Fit Index: Boys: 0.994 Girls: 0.981

Fig. 2. Final Lisrel Model for the Reading Task.

Parameter estimates (common metric standardized solution) are placed along the arrows. Coefficients marked with '*' are nonsignificant at the 5% level. Explained variances are denoted as R². The values for girls are placed between parentheses.



Chi-Square: 46.72 ($df=31$, $p=0.035$)

Goodness of Fit Index: Boys: 0.993 Girls: 0.979

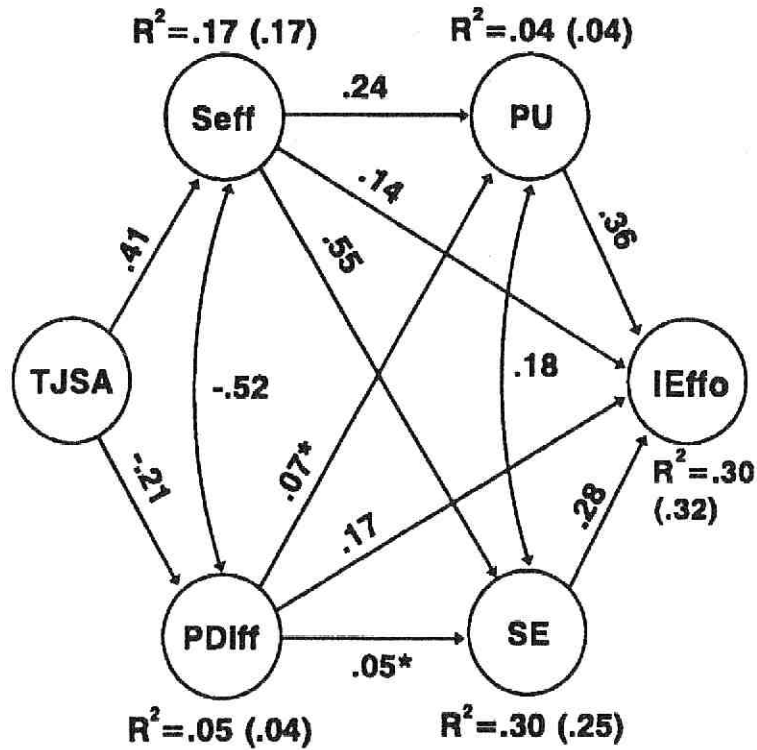
Fig. 3. Final Lisrel Model for the Drawing Task.

Parameter estimates (common metric standardized solution) are placed along the arrows. Coefficients marked with '*' are nonsignificant at the 5% level. Explained variances are denoted as R^2 . The values for girls are placed between parentheses.

We may conclude that the appraisal model is adequate to describe the data on the drawing task. Compared with the reading model, the direct effect of teacher judgement on intended effort could be removed. In other respects the model was highly similar to the one obtained for the reading task. The variance of intended effort accounted for by this model is 23% for boys and 28% for girls. Inspection of the parameter estimates in Figure 3 shows that the lowest coefficients are those for the effects involving perceived difficulty. The differences between boys and girls for the drawing task are limited to differences in the variances of the residuals from the structural model.

Arithmetic Task

The model for the reading task, resulting from Step 1, had a poor fit when applied to the arithmetic data. Two modifications were needed to improve the fit, namely inclusion of direct effects of both perceived difficulty and self-efficacy on intended effort. The direct effect of teacher judgement on intended effort was no longer statistically significant. In contrast with the reading and the drawing tasks, direct effects of sex on self-efficacy and on perceived difficulty had to be included, as well as on intended effort and on perceived utility. Boys perceive the arithmetic task as easier and score higher on self-efficacy than girls. Girls perceive the utility of the arithmetic task as higher, and they also score higher on intended effort than boys. In the two-group analyses on the exploration sample (Step 2) only Model d fitted well at the 5%-level. However this result did not survive the cross-validation. From the results of the cross-validation



Chi-Square: 41.71 ($df=29$, $p=0.060$)
 Goodness of Fit Index: Boys: 0.991 Girls: 0.984

Fig. 4. Final Lisrel Model for the Arithmetic Task.

Parameter estimates (common metric standardized solution) are placed along the arrows. Coefficients marked with '*' are nonsignificant at the 5% level. Explained variances are denoted as R^2 . The values for girls are placed between parentheses.

step, one could choose Model a as the best model. When estimated on the complete sample all four models adequately fitted the data. Comparing the four models by conditional tests, Model a appeared to be significantly worse than models b and c, which in turn were significantly worse than model d. At the 1%-level, only Model a deviated significantly from the other three models. The BIC measure identified Model a as the best choice, but the values for Model b and c were rather close. Although the results on the tests for sex differences are less clear-cut for the arithmetic task than for the other two tasks, Model b was accepted as the final model for the arithmetic data (see Figure 4).

It may be inferred from these findings that the appraisal model is adequate to describe the data on the arithmetic task. However, compared with the other tasks, two extra effects were needed, viz., the effects of self-efficacy and perceived difficulty on intended effort. This means that the correlations between self-efficacy and intended effort and between perceived difficulty and intended effort cannot be fully explained by the mediating variables, success expectancy and perceived utility. As with the drawing task, the direct effect from teacher judgement to intended effort could be removed. Another difference with the other tasks is the existence of main effects for sex: girls score higher than boys on perceived utility and intended effort. However, as with the other tasks, the relations between the variables are the same for both sexes, allowing for differences in the residual variances. Inspection of the parameter estimates in Figure 4 shows nonsignificant coefficients for the effect of perceived difficulty on both success

expectancy and perceived utility. The position of perceived difficulty in the model again only weakly confirms its presumed theoretical role. The relationships added to the model show low, though statistically significant, coefficients. The variance of intended effort accounted for by this model is about thirty percent (boys 30% and girls 32%).

CONCLUSIONS AND DISCUSSION

The results show that with few modifications the appraisal model is adequate to describe the structure of the data. A slightly different model was found for each task. The model found for the drawing task resembles best the original appraisal model specified in Figure 1. The only deviation from this original model is a small correlation between the residuals of success expectancy and perceived utility. So, the correlation between these two variables cannot be fully explained by the preceding variables in the model. Meece et al. (1990) also found a residual correlation between expectancies and 'importance' in their models. For the reading task, a direct effect of teacher judgment on intended effort was needed. For the arithmetic task, direct effects of both self-efficacy and perceived difficulty on intended effort had to be included. Main effects for sex were found only for the arithmetic task. Boys perceive their own mathematics ability as higher than girls. Boys also perceive the arithmetic task as easier than girls. Girls perceive the utility of the task as higher than boys and they express higher intended effort than boys. For all three tasks, Model b was accepted as the final model. This implies that there are no differences between the sexes in the structural part of the model. Only the residual variances slightly differ for the sexes, reflecting differences in the variances of the observed variables between boys and girls. This finding warrants the conclusion that the models found are equally suited for boys and for girls. Meece et al. (1990) also reported that the pattern of relations did not differ for boys and for girls. They concluded that sex differences in course enrollment and performance patterns are caused by sex differences in the means of the predictor variables, rather than by differences in the pattern of relationships among the predictors. This conclusion is supported by our data on intended effort for curricular tasks. Thus, we conclude that sex differences in intended effort are not caused by different patterns of relations in the appraisal model.

The mean sex differences we found at the momentary level are comparable with findings at the middle level (Eccles et al., 1983; Meece et al., 1982, 1990; Ethington, 1991). However, we found that girls perceive the utility of the arithmetic task as higher than boys, whereas in the literature it is generally reported that boys perceive the utility of math as higher than girls. Our findings suggest that at the end of elementary school boys and girls seem to have already somewhat different achievement-related perceptions at the momentary level, reflected in their appraisals measured.

Explained variances

The variance of intended effort accounted for by the three models is rather low. With respect to the arithmetic model, the percentages we found lie in between the percentages found by Eccles et al. (1983, 1985) and those found by Ethington (1991). These values are remarkably similar, considering the fact that there are many differences, such as different perspective (middle level versus momentary level), less and other exogenous variables, extra direct effects on intended effort in our arithmetic model, and the use of somewhat younger students.

With regard to the explained variances of other variables, it is noticed that success expectancy shows the highest percentage of variance explained. This is in accordance with the findings reported by Eccles et al. (1983, 1985) and Ethington (1991). For perceived difficulty and perceived utility the variance explained is rather low.

Total effects

We found the largest total effects on intended effort stemming from the constructs self-efficacy, success expectancy, and perceived utility. All these effects are moderate and of about the same

size (about .30, unstandardized). One exception is the considerably higher total effect of self-efficacy on intended effort with respect to the arithmetic task (.46, unstandardized). Ethington (1991) also reported that self-concept had the largest total influence on intentions to enroll in mathematics courses.

Parameter estimates

The coefficient for the effect of self-efficacy on success expectancy is the highest one in our models. Perceived utility has a stronger direct effect on intended effort than success expectancy, except for the drawing task where these direct effects are equal. All coefficients involving perceived difficulty are low or nonsignificant. Only with respect to the arithmetic task a substantial total effect of perceived difficulty on intended effort was found. Hence, the position of perceived difficulty is questionable. Eccles et al. (1983, 1985) and Ethington (1991) also found nonsignificant effects for task difficulty. A curvilinear relation between perceived task difficulty and intended effort could be an explanation for the absence of strong paths. We found a direct effect of perceived difficulty on intended effort but only for the arithmetic task. This effect was small and positive: perceiving the task as difficult leads to more intended effort. Another possibility is that the influence of perceived difficulty works via self-efficacy. Eccles et al. (1985), using cross-lagged panel analyses on longitudinal data, found a negative effect from difficulty of current math course on self-concept of math ability. The high, negative correlation between the residuals of self-efficacy and perceived difficulty could point in such direction. With respect to other parameter estimates we also noticed many resemblances among the tasks as well as between our findings and the findings of Ethington (1991) and Eccles et al. (1983), especially those from their second wave of measurements (Eccles et al., 1985).

Teacher judgement

The power of the appraisal processes in explaining the correlation between teacher judgement and intended effort is highest for the drawing task. However, for all tasks the observed correlation between teacher judgement of ability and intended effort is small. The presumption that the teacher judgement of student ability has only direct effects on self-efficacy and perceived difficulty is correct. One exception is the statistically significant, but small effect of teacher judgement on intended effort with the reading task. It is noteworthy that the direct effects of teacher judgement of math ability on students' self-efficacy and on perceived difficulty expressed just before working on a math task, are nearly twice as large as those displayed in the other two tasks.

The pattern of relations in Eccles' model, based on expectancy x value models and specified at the middle level has proven to be an adequate basis for explaining appraisal processes on different school subjects at the momentary level. There are many resemblances with the final models we found in this study. This suggests that the appraisals that students of this age group make with regard to academic tasks are, to a large extent, general in nature. They are only to a small extent task specific or school-subject specific. This is probably due to the fact that the appraisals studied here are still quite global in nature. The students did not exactly know the content of the tasks. We want to explore task differences in subsequent studies, and to include task attraction as an important appraisal in the model (Boekaerts, et al., in preparation). We also want to extend the model with affects, experienced by students when faced with curricular tasks, and to use a revised version of the On-Line Motivation Questionnaire (cf. Seegers & Boekaerts, 1993), in which the constructs were operationalized by more than one item.

NOTE

1. For the sake of convenience we represented correlations between residuals in the figures by curved arrows between the latent variables, without explicitly drawing the residuals.

REFERENCES

- Atkinson, J.W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, 64, 359-372.
- Atkinson, J.W. (1964). *An introduction to motivation*. Princeton, NJ: Van Nostrand.
- Atkinson, J.W. & Feather, N.T. (1966). *A theory of achievement motivation*. New York: Wiley.
- Boekaerts, M. (1985). Some new developments in the study of motivational processes in a classroom context. In G. D'Ydewalle (ed.), *Cognition, Information Processing, and Motivation*. Vol. 3, XXIII International Congress of Psychology (pp. 569-586). Amsterdam: North Holland.
- Boekaerts, M. (1986). The measurement of state and trait motivational orientation: Refining our measures. In E.E.J. de Bruyn, J.H.L. van den Bercken, & Th.C.M. Bergen (eds.), *Achievement and task motivation* (pp. 229-245). Berwyn: Swets North-America.
- Boekaerts, M. (1987a). Die Effekte von state- und trait-motivationaler Orientierung auf das Lernergebnis. *Zeitschrift für Pädagogische Psychologie*, 1, 29-43.
- Boekaerts, M. (1987b). Individual differences in the appraisal of learning tasks: An integrative view on emotion and cognition. *Communication and Cognition*, 20, 207-224.
- Boekaerts, M. (1987c). Situation specific judgements of a learning task versus overall measures of motivational orientation. In E. de Corte, H. Lodewijks, R. Parmentier & P. Span (eds.), *Learning and instruction* (pp. 169-179). Oxford/Leuven: Pergamon Press/Leuven University Press.
- Boekaerts, M. (1988). Motivated learning: bias in appraisals. *International Journal of Educational Research*, 12, 267-280.
- Boekaerts, M. (1992). The adaptable learning process: Initiating and maintaining behavioural change. *Applied Psychology: An International Review*, 41, 377-397.
- Boekaerts, M., Crombach, M.J., & Voeten, M.J.M. (in preparation). Task attraction as a determinant of intended effort on curricular tasks.
- Cantor, N. (1981). Perceptions of situations. In D. Magnusson (ed.), *Toward a psychology of situations. An interactional perspective* (pp. 229-244). Hillsdale, NJ: Erlbaum.
- Crombach, M.J., Boekaerts, M., & Voeten, M.J.M. (in preparation). On-line measurement of appraisals of students faced with curricular tasks.
- Cudeck, R., & Browne, M.W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research*, 18, 147-167.
- deCharms, R. (1984). Motivation enhancement in educational settings. In R. Ames & C. Ames (eds.), *Research on motivation in education. Vol 1: Student motivation* (pp. 275-310). Orlando, FL: Academic Press.
- Deci, E.L. (1975). *Intrinsic motivation*. New York: Plenum.
- Eccles (Parsons), J., Adler, T.F., Futterman, R., Goff, S.B., Kaczala, C.M., Meece, J.L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J.T. Spence (ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 75-146). San Francisco: W.H. Freeman and Co.
- Eccles (Parsons), J., Adler, T.F., Futterman, R., Goff, S.B., Kaczala, C.M., Meece, J.L., & Midgley, C. (1985). Self-perceptions, task perceptions, socializing influences, and the decision to enroll in mathematics. In S.F. Chipman, L.R. Brush, & D.M. Wilson (eds.), *Women and mathematics: Balancing the equation* (pp. 95-121). Hillsdale, NJ: Erlbaum.
- Eccles (Parsons), J., Adler, T., & Meece, J.L. (1984). Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology*, 46, 26-43.
- Ethington, C.A. (1991). A test of a model of achievement behaviors. *American Educational Research Journal*, 28, 155-172.
- Harter, S. (1978). Effectance motivation reconsidered: Towards a developmental model. *Human Development*, 21, 34-64.
- Heckhausen, H. (1977). Achievement motivation and its constructs: A cognitive model. *Motivation and Emotion*, 1, 283-329.
- Heckhausen, H. (1980). *Motivation und Handeln: Lehrbuch der Motivationspsychologie*. Berlin: Springer.
- Lazarus, R.S. (1991a). Cognition and motivation in emotion. *American Psychologist*, 46, 352-367.
- Lazarus, R.S. (1991b). Progress on a Cognitive-Motivational-Relational Theory of Emotion. *American Psychologist*, 46, 819-834.
- Lazarus, R.S., & Folkman, S. (1984). *Stress, appraisal, and coping*. New York: Springer.
- Lepper, M.R., & Greene, D. (1978). *The hidden costs of rewards: New perspectives on the psychology of human motivation*. Hillsdale, NJ: Lawrence Erlbaum.

- MacCallum, R.C., Roznowski, M., & Necowitz, L.B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490-504.
- Meece, J.L., Eccles (Parsons), J., Kaczala, C.M., Goff, S.B., & Futterman, R. (1982). Sex differences in math achievement: Toward a model of academic choice. *Psychological Bulletin*, *91*, 324-348.
- Meece, J.L., Wigfield, A., & Eccles, J.S. (1990). Predictors of math anxiety and its influence on young adolescents' course enrolment intentions and performance in mathematics. *Journal of Educational Psychology*, *82*, 60-70.
- Raftery, A.E. (1993). Bayesian model selection in structural equation models. In K.A. Bollen & J.S. Long (eds.), *Testing Structural Equation Models* (pp. 163-180). Newbury Park: Sage.
- Seegers, G., & Boekaerts, M. (1993). Task motivation and mathematics achievement in actual task situations. *Learning and Instruction* *3*, 133-150.
- Spence, J.T., & Helmreich, R.L. (1983). Achievement-related motives and behaviors. In J.T. Spence (ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 7-74). San Francisco: W.H. Freeman and Co.
- Wigfield, A., & Eccles, J.S. (1992). The development of achievement task values: a theoretical analysis. *Developmental Review*, *12*, 265-310.

Manuscript ontvangen 25-1-1994

Definitieve versie ontvangen 25-5-1994

APPENDIX

The Items of the OMQ Used in this Study

Self-efficacy (Seff1)

I am able to do the following task

very well. (score 4)

well.

not so well.

not well at all. (score 1)

I do not know.

Self-efficacy (Seff2)

How good do you think you can do this task?

very good (score 4)

good

not so good

not good at all (score 1)

I do not know.

Perceived Difficulty (PDiff)

I think this task is

very difficult. (score 4)

difficult.

not so difficult.

not difficult at all. (score 1)

I do not know.

Success Expectancy (SE)

I expect to do at this task.

very well (score 4)

well

not so well

not well at all (score 1)

I do not know.

Perceived Utility (PU)

I think the following task is

very important. (score 4)

important.

not so important.

not important at all. (score 1)

I do not know.

Intended Effort (IEffo)

I shall try to do this task

very well. (score 4)

well.

not so well.

not well at all. (score 1)

I do not know.

Tijdsattitudes, faalangst en studiemotivatie

Marie-Anne Moreas en Willy Lens

*Departement Psychologie, Universiteit Leuven**

ABSTRACT

In this study the effects of positive attitudes towards the present and towards the future on student motivation and achievement are studied. The results confirm the existence of these effects. In addition, it is investigated whether fear of failure has a direct effect on motivation and achievement, or whether the effect of fear of failure on motivation and achievement is mediated through the affective attitude towards the present and the future. The latter relation is confirmed by using linear structural modelling (LISREL).

INLEIDING

Het belang van de affectieve attitude ten aanzien van de toekomst wordt vaak in verband gebracht met de huidige economische recessie, waarbij de banen niet meer voor het grijpen liggen. Een negatieve, pessimistische kijk op de toekomst zou een belangrijke oorzaak zijn van demotivatie bij leerlingen in het secundair onderwijs. Uit eerder onderzoek blijkt echter dat de meeste leerlingen in het Vlaamse secundair onderwijs positief staan ten aanzien van hun toekomst (Rosseel, 1989; Van Calster, Lens & Nuttin, 1987).

Het effect dat de affectieve attitude die leerlingen hebben ten aanzien van hun individuele toekomst heeft op hun studiemotivatie werd onderzocht met de tijdsattitudeschaal (Lens & Decruyenaere, 1989, 1991; Rosseel, 1989; Van Calster, 1979; Van Calster, Lens & Nuttin, 1987). Van Calster (1979) vond een significant positief verband tussen de score op een subschaal van de tijdsattitudeschaal met vijf zuiver evaluatieve items (bijv. aangenaam-onaangenaam) en de studiepersistentie. Van de overige vijf subschalen voor de affectieve tijdsattitude ten aanzien van de toekomst correleerde slechts de schaal *gepercipieerde structuur in de toekomst* significant met de studiepersistentie. Lens en Decruyenaere (1989) maten de affectieve attitude met 10 items. Zij vonden in een steekproef van meer dan 1000 scholieren een zwak positief verband tussen de affectieve attitude ten aanzien van de toekomst enerzijds en de studiepersistentie en de studie-efficiëntie anderzijds. Wanneer zij deze relatie echter per studierichting onderzochten, vonden zij alleen in het algemeen secundair onderwijs de verwachte verschillen voor de affectieve attitude in functie van het niveau van de studiepersistentie of in functie van het niveau van de studie-efficiëntie. Bij de leerlingen uit het technisch secundair onderwijs had de toekomst-attitude blijkbaar geen motivationeel effect. In het secundair beroepsonderwijs bleek er wel een verband te bestaan met de studie-inzet. Studiepersistentie en -efficiëntie konden niet worden gemeten in het beroepsonderwijs. In tegenstelling tot Van Calster vond Rosseel (1989) geen significant verband tussen de affectieve attitude ten aanzien van de toekomst en studiemotivatie.

Het beeld dat deze studies over het motivationeel effect van de tijdsattitude ten aanzien van de toekomst (TAAT) geven is dus niet éénduidig. De veronderstelde motivationele effecten blijven soms uit en de gevonden positieve samenhangen zijn over het algemeen zwak. Vanuit de theoretische definitie van de affectieve attitude wordt echter een motivationeel effect van de toekomst-attitude verwacht. Een positieve kijk op de toekomst, of de toekomst als waardevol zien en het gevoel controle te hebben over de toekomst, maakt het zinvol om te werken aan de toekomst. De motivationele impact die uitgaat van een positief toekomstbeeld wordt in deze

* Adres: Tiensestraat 102, B-3000 Leuven.

studie meer diepgaand onderzocht door naast de totaalscore voor de affectieve attitude ten aanzien van de toekomst ook de afzonderlijke factoren of componenten van de affectieve attitude in beschouwing te nemen. Uit een studie van Lens en Decruyenaere (1989) bleek dat het zinvol is de relatie tussen de toekomst-attitude en de verschillende aspecten van de studiemotivatie (de studiepersistentie en de studie-efficiëntie) apart te onderzoeken.

Het motivationeel effect van de affectieve tijdsattitude ten aanzien van het *heden* werd ook gemeten met de tijdsattitudeschaal. Indien het theoretisch zinvol is het motivationeel effect van de affectieve attitude ten aanzien van de toekomst te veronderstellen, loont het ook de moeite om te onderzoeken of er een motivationeel effect uitgaat van de affectieve attitude ten aanzien van het heden. Een negatieve kijk op het heden en een gevoel van gebrek aan controle over het heden, zou de leerlingen kunnen demotiveren. Daarom zal ook het motivationele effect van de affectieve attitude ten aanzien van het heden op een meer diepgaande wijze worden onderzocht.

Vanuit de hypothesen over de positieve effecten van de beide vormen van affectieve attitude op de studiemotivatie wordt tevens een positief effect van de affectieve attitude ten aanzien van elke tijdsperiode afzonderlijk op de studieresultaten verwacht.

De resultaten zullen in twee delen worden besproken. In een eerste stap worden de a priori hypothesen met betrekking tot de effecten van de affectieve attitude ten aanzien van het heden en de toekomst op de studiemotivatie en -resultaten getoetst. In een tweede stap wordt een post-hoc model omtrent de mediërende rol van de affectieve attitudes bij het effect van faalangst op studiemotivatie en -prestatie geformuleerd en getoetst.

METHODE

Steekproef

De subjecten zijn leerlingen van het derde jaar (leeftijd 14-15 jaar) in het algemeen secundair onderwijs in één meisjesschool en één jongensschool. De gegevens werden verzameld in twee sessies van elk twee uur, waarin ook andere variabelen werden gemeten dan de nu gerapporteerde (Moreas & Lens, 1991). In de eerste sessie participeerden 129 meisjes en 137 jongens. In de tweede sessie waren 127 meisjes en 128 jongens aanwezig. De tijdsattitude-schalen werden slechts aangeboden aan de leerlingen die alle overige vragenlijsten uit de tweede sessie tijdig ingevuld hadden. Verschillende leerlingen kregen de tijdsattitude-schalen niet volledig ingevuld. Verder schreven sommige respondenten hun code niet op één of meerdere formulieren en lieten anderen één of meerdere items onbeantwoord. Daardoor zijn hun antwoorden voor de betreffende vragenlijst niet meer bruikbaar. Dit impliceert dat het aantal respondenten niet hetzelfde is voor alle analyses.

Instrumenten

Faalangst

De State-Trait Anxiety Inventory (Van der Ploeg, Defares & Spielberger, 1980) meet de algemene (Trait) en de specifieke (State) angst. Algemene angst betreft een relatief stabiele dispositie tot het zich angstig voelen in het algemeen. Het is deze dispositionele angst die situationeel opgewekt wordt in testsituaties. In testsituaties opgewekte angst wordt als faalangst aangeduid. Daarom wordt de Trait-versie gebruikt voor het meten van de mate van faalangst.

De vragenlijst bevat 20 items (bijv. 'Ik voel me prettig'; 'Ik voel me nerveus en onrustig'). De leerlingen geven aan hoe vaak ze de aangeduide gemoedsgestelden ervaren door één van vier antwoordalternatieven aan te duiden (1=bijna nooit; 4=bijna altijd). Hoe hoger de score, des te hoger de faalangst.

Affectieve attitude ten aanzien van heden en toekomst

Het gebruikte instrument is een aangepaste versie van Nuttins tijdsattitude-schaal of T.A.S.

(Nuttin & Lens, 1985; Van Calster, 1979). Deze schaal is gebaseerd op Osgoods semantische differentiaal techniek. Elk item bestaat uit een paar adjectieven met tegengestelde betekenis (bijv. 'aangenaam-onaangenaam'; 'koud-warm'). Op een zevenpuntsschaal geven respondenten aan in welke mate zij hun persoonlijk heden (toekomst) als (bijvoorbeeld) aangenaam of onaangenaam (koud of warm) ervaren. Van de T.A.S. zijn verschillende vormen beschikbaar. De hier gebruikte vorm bestaat uit 30 items, die afgeleid werden uit vroegere versies van de T.A.S.. De items uit de oorspronkelijke versie van de T.A.S. werden integraal overgenomen omdat deze versie al op haar waarde getoetst is (Nuttin & Lens, 1985) om de attitude ten aanzien van het heden én om de attitude ten aanzien van de toekomst te meten. Van Van Calsters versie om de affectieve attitude ten aanzien van de toekomst te meten (Van Calster, 1979, tweede factoranalytische studie), namen we voor elke factor een aantal items over die specifiek op één factor laadden. Zo resulteerde een gewijzigde versie van de T.A.S.. De items werden tweemaal aangeboden: een eerste keer voor het meten van de affectieve (tijds)attitude ten aanzien van het heden (TAAH), en een tweede keer voor het bepalen van de affectieve (tijds)attitude ten aanzien van de toekomst (TAAT).

Op beide datasets werden principale componenten analyses (met varimax-rotatie) uitgevoerd. In de oplossing voor het heden werden drie factoren met een eigenwaarde groter dan 1 geëxtraheerd. De drie factoren verklaren respectievelijk 25.6%, 5.5% en 4.0% van de variantie. Van de 30 items zijn er 27 met een lading groter of gelijk aan .30 op de eerste factor, en van de 14 items met een lading hoger dan .50 zijn er vier met een zuiver evaluatieve betekenis en vijf met de betekenis 'waarde'. Hieruit blijkt dat de eerste factor de evaluatieve betekenis van het heden weergeeft. De totaalscore voor de TAAH werd berekend op basis van de 27 items die laden op de eerste factor. De principale componenten analyse voor de TAAT geeft eenzelfde beeld. De drie factoren verklaren respectievelijk 30.03%, 5.5% en 3.87% van de totale variantie. Dezelfde 27 items als voor het heden hebben een lading hoger dan .30. Al de zuiver evaluatieve items en vijf items met de betekenis 'waarde' hebben een lading hoger dan .50. Analoog aan de berekeningen voor het heden, werd de totaalscore voor de affectieve attitude ten aanzien van de toekomst berekend op basis van de 27 items die laden op de eerste factor. De Pearson correlatie tussen de affectieve attitude ten aanzien van het heden en de affectieve attitude ten aanzien van de toekomst is .76 ($n=187$, $p<.001$).

Vervolgens werd apart voor het heden en de toekomst de oplossing geroteerd (Varimax criterium). Voor de TAAH resulteerde dit in de volgende factoren: (1) gepercipieerde waarde van het heden (HWAARDE), (2) persoonlijk belang en engagement (HBELANG), en (3) gepercipieerde gemakkelijkerheid (HMAK). Voor de toekomst resulteerde dit in drie parallele factoren: (1) de verwachte waarde van de toekomst (TWAARDE), (2) persoonlijk belang en engagement (TBELANG), en (3) geanticipieerde gemakkelijkerheid (TMAK). In de Bijlage worden voor de factoren voor het heden en voor de toekomst afzonderlijk de hoog ladende items opgesomd. Voor het bepalen van de scores op deze factoren worden de factorscores gebruikt.

Studiemotivatie

Studiemotivatie werd gemeten met een vragenlijst van 19 items, die gebaseerd is op de studiemotivatie-vragenlijst van De Volder (1980). Respondenten geven aan of ze al dan niet akkoord zijn met de inhoud van de items (bijv. 'Ik vind dat ik harder werk dan de meeste leerlingen van mijn klas'; 'Ik studeer regelmatig'). De totaalscore wordt berekend over alle items: een hoge score betekent een hoge studiemotivatie. Op de data van deze vragenlijst werd een principale componenten analyse uitgevoerd. De oplossing werd geroteerd (Varimax criterium). De resulterende factoren zijn gelijk aan die van De Volder (1980; De Volder & Lens, 1982): nl. studie-efficiëntie (bijv. 'Ik moet me tijdens het studeren dwingen om bepaalde gedachten uit mijn hoofd te zetten'; 'Dikwijls kom ik moeilijk tot studeren') en studiepersistentie (bijv. 'Ik studeer regelmatig'; 'Zelfs voor saaie vakken span ik me voldoende in'). De factorscores werden gebruikt voor verdere analyses.

Studieresultaten

Als studieresultaten werden de uitslagen op de examens voor het eerste trimester van het schooljaar gebruikt. De verschillende klassen komen echter uit twee verschillende scholen, hebben andere leerkrachten en een ander vakkenpakket. Hierdoor houdt eenzelfde uitslag in verschillende klassen niet noodzakelijk eenzelfde studieprestatie in. Om die reden werden de examenuitslagen per klas gestandaardiseerd. Een hoge score betekent een goede prestatie in vergelijking met de andere leerlingen van dezelfde klas, en niet noodzakelijk een goede prestatie in de totale steekproef. Door de gegevens te standaardiseren per klas wordt niet alleen de foutvariantie uitgezuiverd (andere leerkrachten, andere vakken en dergelijke), maar wordt ook de relevante variantie tussen de klassen gereduceerd. De variantie tussen de klassen die veroorzaakt wordt door werkelijke prestatieverschillen verdwijnt. Als men de studieresultaten in verband wil brengen met een andere variabele, moet men ook die variabele standaardiseren per klas.

RESULTATEN

Een score van 108 op de TAAH en de TAAT komt overeen met een neutrale attitude. Vooreerst wordt opgemerkt dat in deze steekproef de overgrote meerderheid van de leerlingen een positieve attitude heeft (score hoger dan 108) ten aanzien van het heden (91%) en de toekomst (93%). Dit is in overeenstemming met de resultaten van Rosseel (1989) en van Van Calster, Lens en Nuttin (1987).

De hypothesen over de effecten van de TAAH en de TAAT op de studiemotivatie en op de studieresultaten worden onderzocht via ANOVA's. De totaalscores voor de TAAH en de TAAT worden daarbij als onafhankelijke variabelen genomen, terwijl studiemotivatie en studieresultaten afhankelijke variabelen zijn. Het motivationele effect van de TAAH en van de TAAT wordt verder onderzocht door de Pearson correlaties tussen de TAAH- of de TAAT-variabelen (totaalscore en factorscores) en alle variabelen voor de studiemotivatie (totaalscore en de factorscores) te bestuderen.

Tot slot worden de hypothesen omtrent het positieve effect van de studiemotivatie-variabelen (totaalscore en factoren) op de studieresultaten onderzocht. Hiervoor wordt voor elke studiemotivatie-variabele apart een ANOVA uitgevoerd met de studiemotivatie-variabele (mediaan-splitsing) als onafhankelijke variabele en de studieresultaten als afhankelijke variabele.

De affectieve attitude ten aanzien van het heden (TAAH)*Studiemotivatie*

De ANOVA met TAAH (mediaan-splitsing) als onafhankelijke variabele en de score op studiemotivatie als afhankelijke variabele toont een significant effect van TAAH ($F(1, 186)=7.71$, $p<.007$). De studenten met de meest positieve attitude ($\bar{X}=9.81$, $n=96$) zijn meer gemotiveerd dan de studenten met een minder positieve attitude ($\bar{X}=8.11$, $n=92$). Om dit motivationeel effect verder te interpreteren zijn Pearson correlaties tussen alle variabelen voor TAAH en voor de studiemotivatie berekend. De resultaten zijn weergegeven in Tabel 1.

De TAAH (totaalscore) heeft een significant positief verband met alle studiemotivatie-variabelen. Dit betekent dat studenten die een hogere score op de TAAH hebben ook meer studiemotivatie en studiepersistentie vertonen, en dat hun studie-efficiëntie groter is.

Van de factoren van de affectieve attitude ten aanzien van het heden heeft alleen de factor *persoonlijk belang en engagement* een significant positief verband met alle motivatievariabelen. Naarmate het actueel persoonlijk belang en persoonlijk engagement groter is, is ook de gerapporteerde studiepersistentie en de studie-efficiëntie, en dus ook de totale studiemotivatie, groter. De factor *gepercipieerde waarde van het heden* heeft enkel met de studie-efficiëntie een significant positief verband. De factor *gepercipieerde gemakkelijkeheid* heeft een significant positief verband met de totale studiemotivatie en met de studie-efficiëntie

Tabel 1. Pearson correlaties tussen de TAAH-variabelen en de studiemotivatie-variabelen.

		STUDIEMOTIVATIE		
		totaalscore	persistentie	efficiëntie
HTOT	r	.27***	.19*	.33***
	n	188	166	166
HWAARDE	r	.06	.06	.16*
	n	183	162	162
HBELANG	r	.30***	.26***	.22**
	n	183	162	162
HMAK	r	.19*	.05	.32***
	n	183	162	162

*: $p < .05$, **: $p < .01$, ***: $p < .001$

HTOT = de totaalscore voor TAAH; HWAARDE = subscore *gepercipieerde waarde van het heden*; HBE-LANG = subscore *persoonlijk belang en engagement*; HMAK = subscore *gepercipieerde gemakelijkheid*.

Studieresultaten

Een ANOVA met de score op de TAAH (mediaansplitsing) als onafhankelijke variabele en de studieresultaten als afhankelijke variabele laat een significant effect zien ($F(1, 190)=5.08$, $p<.05$). De leerlingen met de meest positieve attitude presteren significant beter ($\bar{X}=27$, $n=96$) dan degenen met de minst positieve attitude ($\bar{X}=-.04$, $n=96$). De Pearson correlatie tussen de TAAH (totaalscore) en de studieresultaten is positief en significant ($r=.18$, $n=192$, $p<.05$). De factor *gepercipieerde gemakelijkheid* correleert eveneens significant met de studieresultaten ($r=.15$, $n=187$, $p<.05$). De correlaties tussen de studieresultaten en de overige factoren zijn, zoals verwacht, ook positief, maar ze bereiken het 5%-significantieniveau niet: *persoonlijk belang en engagement* ($r=.13$, $n=187$, n.s.), *gepercipieerde waarde van het heden* ($r=.11$, $n=187$, n.s.).

De affectieve attitude ten aanzien van de toekomst (TAAT)

Studiemotivatie

Een ANOVA met TAAT (mediaan-splitsing) als onafhankelijke variabele en de totale studiemotivatie als afhankelijke variabele laat een significant effect voor TAAT op studiemotivatie zien ($F(1, 191)=12.91$, $p<.001$). De leerlingen met de meest positieve houding ten aanzien van de toekomst zijn significant meer gemotiveerd ($\bar{X}=9.99$, $n=94$) dan de andere leerlingen ($\bar{X}=7.87$, $n=99$).

De Pearson correlaties tussen de TAAT-variabelen enerzijds en de studiemotivatie (totaalscore en de scores op de twee afzonderlijke factoren) zijn in Tabel 2 weergegeven. De score op de TAAT (totaalscore) heeft een significant positief verband met de studiemotivatie-variabelen. Hoe positiever de affectieve attitude ten aanzien van de toekomst, hoe groter de studiepersistentie en de studie-efficiëntie, en dus ook de totale studiemotivatie. Van de TAAT-factoren heeft alleen de factor *persoonlijk belang en engagement* een significant positief verband met al de studiemotivatie-variabelen. De factor *verwachte waarde van de toekomst* correleert alleen met de studiepersistentie: de studiepersistentie is groter naarmate men een hogere score voor deze factor heeft. De factor *geanticipieerde gemakelijkheid* heeft een significant positief verband met de totale studiemotivatie en met de studie-efficiëntie.

Studieresultaten

De leerlingen met de meest positieve houding ten aanzien van hun toekomst presteren beter

Tabel 2. Pearson correlaties tussen de TAAT-variabelen en de studiemotivatie-variabelen.

		STUDIEMOTIVATIE		
		totaalscore	persistentie	efficiëntie
TKTOT	r	.26***	.24**	.29***
	n	193	167	167
TWAARDE	r	.09	.18*	.04
	n	187	162	162
TBELANG	r	.22**	.22**	.20*
	n	187	162	162
TMAK	r	.17*	.10	.26***
	n	187	162	162

*: $p < .05$, **: $p < .01$, ***: $p < .001$

TKTOT = de totaalscore voor TAAT; TWAARDE = subscore *verwachte waarde van de toekomst*; TBELANG = subscore *persoonlijk belang en engagement*; TMAK = subscore *geanticiperde gemakkelijheid*.

($\bar{X} = .17$, $n = 99$) dan degenen met de minst positieve houding ($\bar{X} = -.06$, $n = 100$), maar het verschil is niet significant ($F(1, 197) = 2.82$, $p > .05$). De Pearson correlatie tussen de TAAT en de studieresultaten is positief en laag, maar significant ($r = .16$, $n = 199$, $p < .05$). Het niet-significante effect volgens de variantie-analyse kan een gevolg zijn van de dichotomisering van de onafhankelijke variabele, wat een zeker informatieverlies impliceert. Bij de correlatie wordt er maximaal gebruik gemaakt van alle informatie, wat in dit geval tot de conclusie leidt dat er een significant verband is tussen de TAAT en de studieresultaten. De factoren van de TAAT correleren niet significant met de studieresultaten: geanticiperde waarde ($r = .05$, $n = 193$, n.s.); persoonlijk belang en engagement ($r = .14$, $n = 193$, n.s.); geanticiperde gemakkelijheid ($r = .13$, $n = 193$, n.s.).

Motivatie en resultaten

Tot slot werd het effect van de studiemotivatie-variabelen op de studieresultaten getoetst. Voor elke studiemotivatie-variabele werd apart een ANOVA uitgevoerd met de studiemotivatie-variabele (mediaan-splitsing) als onafhankelijke variabele en de studieresultaten als afhankelijke variabele. Elke studiemotivatie-variabele heeft een significant effect op de studieresultaten: totale studiemotivatie ($F(1, 242) = 17.01$, $p < .001$); studiepersistentie ($F(1, 206) = 7.61$, $p < .01$); studie-efficiëntie ($F(1, 206) = 4.42$, $p < .05$). De leerlingen met een sterke totale studiemotivatie ($\bar{X} = .26$, $n = 123$), met een sterke studiepersistentie ($\bar{X} = .15$, $n = 106$) of met een sterke studie-efficiëntie ($\bar{X} = .11$, $n = 105$) halen betere resultaten dan de leerlingen met respectievelijk een zwakkere totale studiemotivatie ($\bar{X} = -.24$, $n = 121$), een zwakkere studiepersistentie ($\bar{X} = -.22$, $n = 102$) of een zwakkere studie-efficiëntie ($\bar{X} = -.17$, $n = 103$).

CONCLUSIE EN DISCUSSIE

Affectieve attitude ten aanzien van het heden

Hoe positiever de affectieve attitude ten aanzien van het heden, des te sterker studenten van 14-15 jaar gemotiveerd zijn om te studeren en hoe beter ze presteren. De motivationeel meest relevante aspecten van de TAAH zijn het persoonlijk belang van en het engagement in het heden en de gepercipieerde gemakkelijheid van het heden.

Moreas en Lens (1992) vonden dit motivationeel effect van de TAAH ook in een steekproef met studenten van 20 tot 40 jaar (55 licentiaatsstudenten en 85 mensen die post-universitaire studies volgden in combinatie met een volledige baan). In deze steekproef werd een significant effect in de verwachte richting van TAAH op de studiemotivatie ($F(1, 135)=14.28, p<.001$), op de studiepersistentie ($F(1, 135)=4.02, p<.05$) en op de studie-efficiëntie ($F(1, 135)=17.95, p<.001$) gevonden. De Pearson correlaties tussen de TAAH-variabelen en de studiemotivatie-variabelen verduidelijken dit beeld. Persoonlijk belang en engagement heeft een significant positief verband met alle motivatievariabelen. Gepercipieerde waarde van het heden heeft enkel met de studie-efficiëntie een significant positief verband, terwijl de waargenomen gemakkelijkerheid een significant positief verband heeft met de studie-efficiëntie, maar niet met de totale studiemotivatie. De sterke gelijkenis tussen de data uit de twee steekproeven wijst op de generaliseerbaarheid van het motivationele effect van de TAAH.

Affectieve attitude ten aanzien van de toekomst

Naarmate de affectieve attitude ten aanzien van de toekomst positiever is, zijn studenten van 14-15 jaar meer gemotiveerd om te studeren en presteren ze beter. De motivationeel meest relevante componenten van de TAAT zijn het persoonlijk belang en engagement en de geanticiperde gemakkelijkerheid.

Het gevonden motivationele effect van de TAAT komt overeen met de resultaten van Van Calster (1979) en Rosseel (1989). Zij meten de TAAT aan de hand van een beperkte schaal met vijf zuiver evaluatieve items. De items laden hoog op de factor *geanticiperde waarde van de toekomst*. Deze factor correleert in onze gegevens niet significant met studiemotivatie (zoals ook bij Rosseel, 1989) noch met de studieresultaten (zoals ook bij Van Calster, 1979), maar wel met de studiepersistentie (zoals bij Van Calster, 1979). Deze niet éénduidige resultaten kunnen dus een gevolg zijn van verschillen in de operationalisering van de TAAT en van de studiemotivatie.

HET MOTIVATIONEEL EFFECT VAN FAALANGST

Tot nog toe werden a priori hypothesen omtrent het motivationeel effect van de TAAH en de TAAT getoetst. Deze affectieve attitudes kunnen bij studenten samenhangen met een sterke of minder sterke faalangst. In het volgende deel van deze bijdrage wordt daarom onderzocht welke relatie er is tussen de TAAH en de TAAT en faalangst. Vervolgens wordt getoetst of het effect van faalangst op studiemotivatie en op studieresultaten kan worden verklaard via de TAAT- en de TAAH-variabelen.

Het is aannemelijk dat er voor schoolgaande jongeren een negatief effect is van faalangst op de attitude zowel ten aanzien van het heden als van de toekomst. Faalangst heeft een effect op TAAH en TAAT omdat faalangst de kans op succeservaringen negatief beïnvloedt. De Pearson correlatie tussen de TAAH en faalangst is hoog negatief ($r=-.67, n=192, p<.001$). Ook de correlatie tussen de TAAT en faalangst is hoog negatief ($r=-.47, n=200, p<.001$). Hoe faalangstiger leerlingen zijn, des te negatiever hun attitudes ten aanzien van hun persoonlijk heden én ten aanzien van hun persoonlijke toekomst.

Op grond van deze correlaties kan een aantal post-hoc hypothesen geformuleerd worden. Deze worden geïntegreerd in een post-hoc model. Hierin zal worden getoetst of het effect van faalangst op de studiemotivatie en op de studieresultaten volledig verklaard kan worden via de TAAH en de TAAT. Gebruik zal worden gemaakt van lineaire structurele modelanalyse (LISREL).

Faalangst en affectieve attitude ten aanzien van het heden

De eerste post-hoc hypothese betreft de samenhang tussen faalangst en de TAAH. Een leerling uit het secundair onderwijs kan dagelijks geconfronteerd worden met zowel onaangekondigde, als aangekondigde toetsen. Bij een student met hoge angst om te falen zal dit waarschijnlijk tot

een constante angst leiden met een negatief effect op de TAAH. Verwacht kan worden dat hoog faalangstigen een minder hoge verwachting hebben van het heden, minder vertrouwen hebben in de realisatie van waardevolle doelen (éénduidig affectieve items, items die de waarde van het heden aanduiden en items die verwachtingen van succes aanduiden) en dat ze het heden als moeilijker ervaren dan laag faalangstigen. Of de hoog faalangstigen verschillen van de laag faalangstigen op de factor *persoonlijk belang en engagement* is theoretisch niet voorspelbaar: het persoonlijk engagement kan afnemen ten gevolge van de faalangst, maar anderzijds veronderstelt de angst om te falen dat men de toetsen en het huidige schoolse leven wel belangrijk vindt. Vanuit de theoretische veronderstellingen volgt dat het sterk negatieve verband tussen faalangst en de TAAH het best verklaard wordt via de relaties tussen faalangst en de factoren *gepercipieerde waarde van het heden* en *gepercipieerde gemakkelijkerheid*. In het post-hoc model is de TAAH dan ook een weerspiegeling van deze twee relevante factoren: de affectieve attitude ten aanzien van het heden in het post-hoc model is de som van de z-waarden (voor de hele groep) van de factorscores voor de factoren *gepercipieerde gemakkelijkerheid van het heden* en *gepercipieerde waarde van het heden*.

Faalangst en de affectieve attitude ten aanzien van de toekomst

De tweede hypothese betreft de samenhang tussen faalangst en de TAAT. De angst voor mislukking op een prestatietoets neemt af naarmate de prestatietoets nog ver is (Gjesme, 1975). De angst voor prestatietoetsen in de toekomst is bijgevolg een weinig waarschijnlijke verklaring voor het sterke verband tussen faalangst en de TAAT. Verwacht wordt daarentegen dat het effect van faalangst op de TAAT verklaard kan worden via de TAAH. Studenten die het heden als moeilijk ervaren zullen dat ook doen ten aanzien van de toekomst. Verder verwachten we dat de geanticipeerde waarde van de toekomst zal worden beïnvloed door de gepercipieerde waarde van het heden, en het persoonlijk belang van en engagement in de toekomst zal worden beïnvloed door het persoonlijk belang van en engagement in het heden. Vanuit de verwachte effecten van faalangst op de TAAH-factoren verwachten we een verband tussen faalangst en de *geanticipeerde waarde van de toekomst* en de *geanticipeerde gemakkelijkerheid van de toekomst*. In het post-hoc model is de TAAT-score een weerspiegeling van deze twee relevante factoren: het is de som van de z-waarden (voor de hele groep) van de factorscores voor deze twee factoren.

Faalangst en studiemotivatie

Correlaties tussen faalangst en de factoren voor studiemotivatie zijn berekend. Faalangst correleert significant negatief met studie-efficiëntie ($r = -.34$, $n = 193$, $p < .001$), maar niet met studie-persistentie ($r = -.11$, $n = 193$, n.s.). Dit betekent dat faalangstige jongeren niet minder (vaak) studeren dan anderen, maar ze studeren op een minder efficiënte wijze (Lens, 1993, p. 58). Analyse van de items die hoog laden op de factor *studie-efficiëntie* suggereert dat faalangstigen zich tijdens het studeren moeten dwingen om bepaalde gedachten uit hun hoofd te zetten, dat ze vlug afgeleid zijn, dat het studeren niet vlot en dat ze, hoewel ze een hele tijd achter hun boeken hebben gezeten, weinig gestudeerd hebben. Ze komen dikwijls ook moeilijk tot studeren. Faalangstigen komen minder gemakkelijk op dreef en ze kunnen hun concentratie moeilijk behouden omdat ze zich zorgen maken. De redenen van hun zorgen zijn te vinden in de TAAH- en TAAT-factoren: hun gepercipieerde waarde van het heden is lager, ze hebben minder vertrouwen in de realisatie van hun waardevolle doelen en ze percipiëren het heden en de toekomst als moeilijker. Vandaar dat verwacht wordt dat het effect van faalangst op de studie-efficiëntie verklaard kan worden via de factoren voor de affectieve attitude ten aanzien van een periode.

Faalangst en studieresultaten

Mensen die minder efficiënt studeren, zullen ook slechter presteren. Dit verklaart de samenhang tussen faalangst en studieresultaten; het effect van faalangst op de studieresultaten wordt dan volledig verklaard via de studie-efficiëntie.

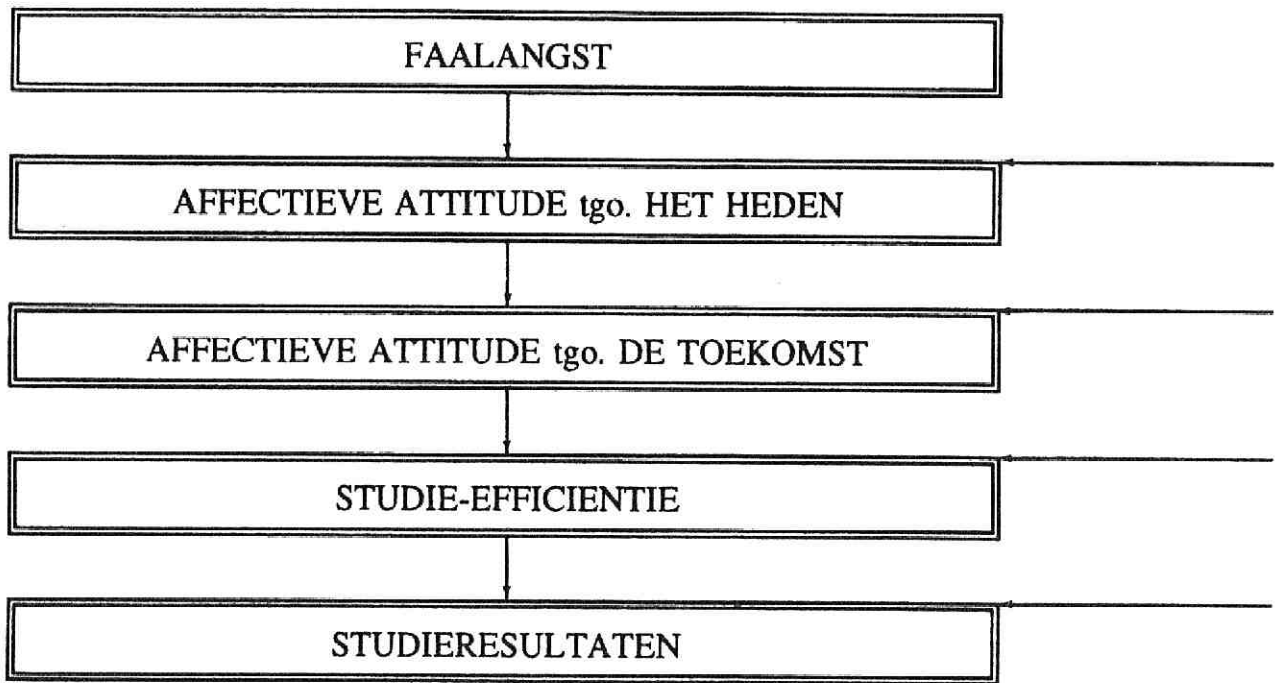


Fig. 1. Post hoc model ter verklaring van het motivationeel effect van faalangst.

Het post-hoc model zoals dat volgt uit de relaties die zijn gespecificeerd in de post-hoc hypothesen, wordt schematisch weergegeven in Figuur 1.

Faalangst heeft alleen een direct effect op de TAAH. Het effect van faalangst op de overige factoren uit het model is indirect van aard. Concreet betekent dit dat faalangst geen substantieel deel van de variantie in de TAAT, in de studiemotivatie en in de studieresultaten verklaart, wanneer rekening wordt gehouden met de variantie in deze variabelen die verklaard wordt door de TAAH-score. De affectieve attitude ten aanzien van het heden heeft geen motivationeel effect onafhankelijk van een effect via de affectieve attitude ten aanzien van de toekomst. En de affectieve attitude ten aanzien van de toekomst heeft alleen via de studie-efficiëntie een effect op de studieresultaten.

RESULTATEN

Dit model wordt getoetst met het LISREL-programma (Jöreskog & Sörbom, 1981). De correlatiematrix wordt als input gebruikt. De parameters worden geschat met de Maximum Likelihood-methode.

Omdat het standaardiseren van de data per klas de variantie in de data reduceert, is het model tweemaal getoetst. Eerst wordt het model zonder de studieresultaten getoetst. In dat geval wordt de variantie maximaal gehouden: een hoge score betekent een hoge score ten opzichte van de totale steekproef. Een tweede maal wordt het model met de studieresultaten getoetst. In dat geval worden alle variabelen per klas gestandaardiseerd: een hoge score betekent dan een hoge score voor de betreffende variabele tegenover de eigen klasgenoten.

Het model zonder de studieresultaten

Het volledige post hoc model in Figuur 1, zonder de studieresultaten maar met de variabelen faalangst tot en met de studie-efficiëntie, past niet goed bij de data ($\chi^2=17.67$, Df=3, $p<.001$). Op grond van de modificatie-indices lijkt het aangewezen om het directe effect van de affectieve attitude ten aanzien van het heden op de studiemotivatie vrij te laten. Het gewijzigde model

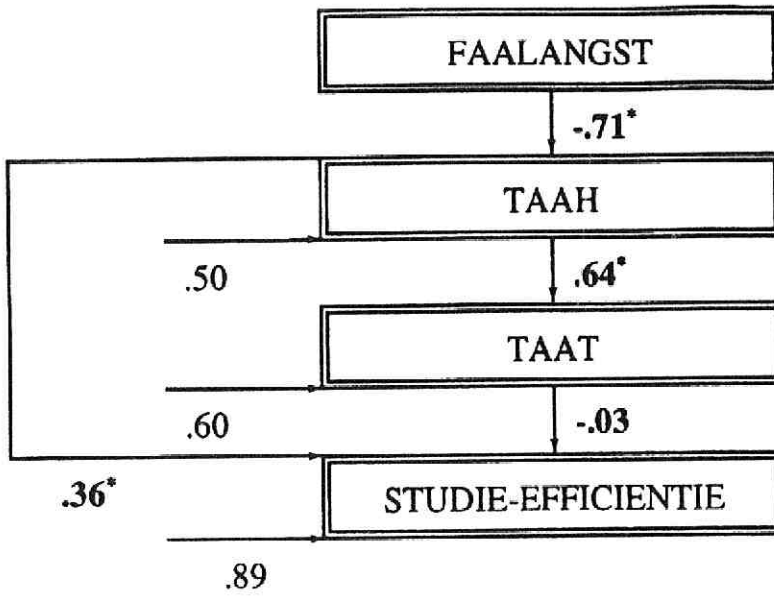


Fig. 2. De Maximum Likelihood-oplossing van het structurele vergelijkingsmodel.

*: $p < .05$

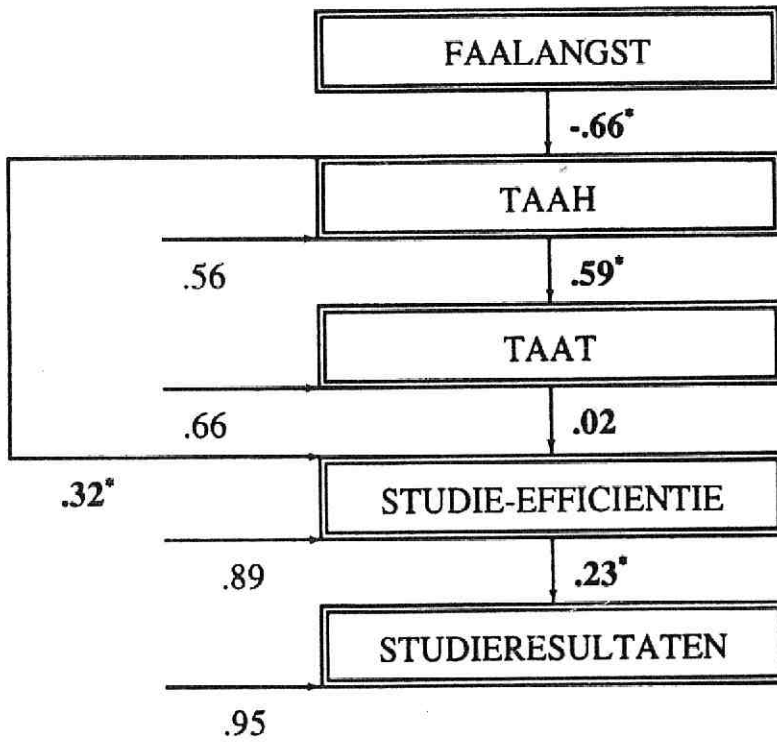


Fig. 3. De Maximum-Likelihood oplossing van het structurele vergelijkingsmodel (met studieresultaten).

*: $p < .05$

komt goed overeen met de data (zie Figuur 2). De χ^2 heeft een waarde van 4.59. Met 2 vrijheidsgraden is dit niet significant ($p > .10$). De Goodness-of-Fit Index is .986 (Adjusted Goodness-of-Fit Index .931). De Root Mean Square Residual is .034, terwijl de totale determinatiecoëfficiënt .505 is. Omdat het nu volgende model met de studieresultaten dit model als een component heeft en gelijkaardige resultaten geeft, wordt de effect-analyse daar besproken.

Het model met de studieresultaten

Het model is vervolgens uitgebreid door het directe effect van de studie-efficiëntie op de studieresultaten toe te voegen. Dit model past goed bij de data (zie Figuur 3). De χ^2 bedraagt 7.41 met $df=5$ ($p>.10$). De Goodness-of-Fit Index is .982 (Adjusted GFI .946). De Root Mean Square Residual= .048, terwijl de totale determinatiecoëfficiënt .439 bedraagt.

Faalangst verklaart 44% van de variantie in de TAAH. Verder wordt door het model 34% van de variantie in de TAAT, 11% van de variantie in de studie-efficiëntie en 5% van de variantie in de studieresultaten verklaard. De TAAH heeft een significant direct effect op de studie-efficiëntie. Als men rekening houdt met dit directe effect verklaart de TAAT geen significant deel van de variantie in de studie-efficiëntie meer.

Het effect van faalangst op de studie-efficiëntie en op de studieresultaten wordt volledig verklaard via de variabelen voor de affectieve attitude. Faalangst heeft alleen een significant direct effect op de TAAH (-.66). Deze attitude is een combinatie van de factoren *gepercipieerde gemakkelijkerheid van het heden* en *gepercipieerde waarde van het heden*. De TAAH heeft een significant direct effect op de studie-efficiëntie (.32) en op de TAAT (.59). Deze attitude is een combinatie van de factoren *geanticiperde gemakkelijkerheid van de toekomst* en *geanticiperde waarde van de toekomst*. De TAAT heeft geen significant direct effect op een andere endogene variabele in dit model. Van alle variabelen heeft alleen de studie-efficiëntie een significant direct effect op de studieresultaten (.23).

CONCLUSIE EN DISCUSSIE

Voor wat betreft het motivationeel effect van faalangst op de studie-efficiëntie en de studieresultaten laten de resultaten zien dat dit volledig te verklaren is via de variabele voor de TAAH. De TAAT blijkt in het geformuleerde model geen relevante variabele te zijn om het effect van faalangst op de studie-efficiëntie te verklaren. De vraag die zich opdringt is hoe het effect van faalangst op de studieresultaten vanuit deze gegevens verklaard kan worden. Er is geen verband met studie-persistentie: faalangstige leerlingen werken even veel uren als niet-faalangstige leerlingen. Maar de efficiëntie waarmee ze studeren neemt af naarmate ze faalangstiger zijn. De reden hiervoor moet worden gezocht in de relevante factoren van de TAAH. We kunnen dit effect verklaren in termen van Sarason's stelling dat '... test anxiety is conceptualized in terms of worrisome, self-preoccupying thoughts that interfere with task performance, ...' (Sarason, 1984, p. 933). En: '... self-preoccupation about the task on which subjects performed was negatively related to their actual level of performance but off-task or wandering thoughts were not' (Sarason, Sarason, Keefe, Hayes & Shearin, 1986, p. 219).

Er is echter een verschil tussen Sarason's inzichten en ons model. Sarason beschrijft de processen die zich voordoen wanneer een faalangstig persoon aan een specifieke taak werkt. In het huidige model wordt met meer algemene constructen gewerkt, namelijk met meer algemene angst ('trait anxiety') en met de factoren van de affectieve attitude ten aanzien van het heden. Hiervan wordt verondersteld dat het persoonlijkheidstrekken zijn die operationeel worden wanneer men aan een prestatietaak (inclusief studeren) werkt. Het zou interessant zijn om in verder onderzoek na te gaan in hoeverre deze meer algemene trekken in specifieke taken aanleiding geven tot interfererende gedachten over de taak en in hoeverre het motivationeel effect via deze interfererende gedachten verklaard kan worden.

LITERATUUR

- De Volder, M. (1980). *Motivatie en toekomstperspectief als mogelijke determinanten van studiegedrag en studieprestaties*. Leuven: Katholieke Universiteit, Niet gepubliceerde doctoraatsverhandeling.
- De Volder, M., & Lens, W. (1982). Academic achievement and future time perspective as a cognitive-motivational concept. *Journal of Personality and Social Psychology*, 42, 566-571.

- Gjesme, T. (1975). Slope of gradients for performance as a function of achievement motive, goal distance in time, and future time perspective. *The Journal of Psychology*, 91, 143-160.
- Jöreskog, K.G., & Sörbom, D. (1981). *Analysis of linear structural relationships by maximum likelihood and least squares methods – V*. Chicago: National Educational Resources.
- Lens, W. (1993). *Studiemotivatie: Theorie voor de praktijk op school en thuis*. Leuven: Universitaire Pers Leuven.
- Lens, W., & Decruyenaere, M. (1989). *Motivatie en demotivatie in het secundair onderwijs: een theoretische en een empirische studie*. Leuven: Katholieke Universiteit, Departement Psychologie.
- Lens, W., & Decruyenaere, M. (1991). Motivation and demotivation in secondary education: student characteristics. *Learning and Instruction*, 1, 145-159.
- Moreas, M.-A., & Lens, W. (1992). *Test anxiety, affective attitude towards life periods, student motivation, and school performance*. Leuven: Katholieke Universiteit, Departement Psychologie.
- Nuttin, J., & Lens, W. (1985). *Future time perspective and motivation: theory and research method*. Leuven & Hillsdale, NJ: Leuven University Press & Erlbaum.
- Rosseel, E. (1989). The impact of attitudes toward the personal future on study motivation and work orientations of nonworking adolescents. *Adolescence*, 24, 73-93.
- Sarason, I.G. (1984). Stress, anxiety, and cognitive interference: reactions to tests. *Journal of Personality and Social Psychology*, 46, 929-938.
- Sarason, I.G., Sarason, B.R., Keefe, D.E., Hayes, B.E., & Shearin, E.N. (1986). Cognitive interference: situational determinants and traitlike characteristics. *Journal of Personality and Social Psychology*, 51, 215-226.
- Van Calster, K. (1979). *De affectieve attitudes tegenover verleden, heden en toekomst: analyse en empirisch onderzoek*. Leuven: Katholieke Universiteit, Niet gepubliceerde doctoraatsverhandeling.
- Van Calster, K., Lens, W., & Nuttin, J. (1987). Affective attitude toward the personal future: impact on motivation in high school boys. *American Journal of Psychology*, 100, 1-13.
- Van der Ploeg, H.M., Defares, P.B., & Spielberger, C.D. (1980). *Handleiding bij de zelfbeoordelingsvragenlijst, ZBV. Een nederlandstalige bewerking van de Spielberger State-Trait Anxiety Inventory, STAI-DY*. Lisse: Swets & Zeitlinger.

Manuscript ontvangen 25-1-1994

Definitieve versie ontvangen 25-5-1994

BIJLAGE

De hoog ladende items (lading > .40) voor de factoren voor het heden en voor de toekomst afzonderlijk

DE AFFECTIEVE ATTITUDE TEN AANZIEN VAN HET HEDEN

Gepercipieerde waarde van het heden (HWAARDE)

Op deze factor laden vooral éénduidig affectieve items. Dit zijn items die de waarde van het heden en items die de verwachting van succes aanduiden.

1. éénduidig affectieve items: afschuwelijk – prachtig (.66), aangenaam – onaangenaam (-.60), donker – licht (.59), warm – koud (-.51), aantrekkelijk – bedreigend (-.52)
2. waarde van het heden: boeiend – saai (-.68), leeg – vol (.51), belangrijk – onbelangrijk (-.46), nuttig – nutteloos (-.42)
3. verwachting van succes: hoopvol – hopeloos (-.67), teleurstelling (mislukt) – vervulling (geslaagd) (.57), zeker – onzeker (-.55), gemakkelijk – moeilijk (-.46), gekend – ongekend (-.42)
4. overige: gesloten – open (.53)

Persoonlijk belang en engagement (HBELANG)

De items die verwijzen naar het persoonlijk belang van het heden zijn: nuttig – nutteloos (-.50), persoonlijk – onpersoonlijk (-.49) en belangrijk – onbelangrijk (-.43). De items die verwijzen naar het persoonlijk engagement zijn: zelf gepland – door anderen bepaald (-.50), waar ik passief afwachtend tegenover sta – dat voor mij een actieve opgave is (.46), van buiten uit bepaald – iets van mijzelf (.41), ongestructureerd – gestructureerd (.47). Eén item met een hoge lading past niet goed binnen onze interpretatie: uitgesteld – direct (.60).

Gepercipieerde gemakkelijkerheid (HMAK)

De hoogst ladende items zijn: vol conflicten – zonder conflicten (.62), problematisch – probleemloos (.56), gemakkelijk – moeilijk (-.52), eenvoudig – gecompliceerd (-.48), zich steeds wijzigend – onveranderd blijvend (.42) en zeker – onzeker (-.41).

DE AFFECTIEVE ATTITUDE TEN AANZIEN VAN DE TOEKOMST

Verwachte waarde van de toekomst (TWAARDE)

Op deze factor laden vooral zuiver evaluatieve items en items die de waarde van de toekomst aanduiden hoog.

1. éénduidig affectieve items: afschuwelijk – prachtig (.65), warm – koud (-.60), donker – licht (.60), aangenaam – onaangenaam (-.57), aantrekkelijk – bedreigend (-.57)
2. waarde van de toekomst: leeg – vol (.62), boeiend – saai (-.52)
3. overige: gesloten – open (.54), teleurstelling – vervulling (.50)

Persoonlijk belang en engagement (TBELANG)

De items die verwijzen naar het persoonlijk belang van de toekomst zijn: nuttig – nutteloos (-.63), boeiend – saai (-.57), belangrijk – onbelangrijk (-.50), teleurstelling – vervulling (.52), persoonlijk – onpersoonlijk (-.47), hoopvol – hopeloos (-.44). De items die verwijzen naar het persoonlijk engagement zijn: waar ik passief afwachtend tegenover sta – dat voor mij een actieve opgave is (.51), van buiten uit bepaald – iets van mezelf (.45), afhankelijk van eigen inspanning en/of bekwaamheid – afhankelijk van toeval of uitwendige omstandigheden (-.41). Verder laden ook de items die verwijzen naar structuur in de toekomst op deze factor: ongestructureerd – gestructureerd (.55), wanordelijk – overzichtelijk (.47), zeker – onzeker (-.45) en gekend – ongekend (-.41). We interpreteren dit patroon in de factorladingen in de zin dat een groter persoonlijk belang van de toekomst en een groter persoonlijk engagement in de toekomst leidt tot / samengaat met een grotere mate van structuur in de toekomst. Twee items met een hoge lading passen niet goed binnen deze interpretatie van de factor: uitgesteld – direct (.61) en gesloten – open (.43)

Geanticiperde gemakkelijkerheid (TMAK)

De hoogst ladende items zijn: problematisch – probleemloos (.62), gemakkelijk – moeilijk (-.67), vol conflicten – zonder conflicten (.49) en eenvoudig – gecompliceerd (-.64), zeker – onzeker (-.50), hoopvol – hopeloos (-.41).

Toekomstperspectief als voorspeller van inzet voor school

Thea Peetsma

*SCO-Kohnstamm Instituut, Universiteit van Amsterdam**

ABSTRACT

This article is concerned with the relation between pupils' future time perspectives and their school investment. Future time perspective is regarded as a predictor, and school investment as the behaviour to be explained. In this study, involving over 600 pupils of all grades of secondary education, it was found that future time perspectives, particularly perspectives related to school- and professional career, are good predictors of school investment. It was also investigated whether pupils' gender, school type, grade, social-economic status, fear of failure, and locus of control influence the relation between future time perspectives and school investment. This proved to be the case for locus of control.

INLEIDING

Toekomst als motor? Reeds in één van de eerste publicaties over het begrip toekomstperspectief (Frank, 1939) ging het om beweegredenen voor gedrag. Wat motiveert iemand? Wat maakt dat de een zich wel inzet voor iets en de ander nauwelijks? Motivatie van leerlingen is een voortdurende bron van zorg in het onderwijs en heeft uiteraard een belangrijke plaats in onderwijsonderzoek. Daarbij is onderzoek van het begrip toekomstperspectief onder meer van belang omdat een negatief toekomstperspectief van leerlingen vaak wordt aangevoerd als verklaring voor ongemotiveerdheid. In de onderzoeksliteratuur over motivatie treft men een onderscheid aan tussen 'inzet' (gemotiveerdheid) en 'motivatie'. Met het begrip inzet wordt datgene aangegeven wat zich van de 'achterliggende' motivatie manifesteert in concreet gedrag dat moet worden verklaard, terwijl 'motivatie' één van de verklarende variabelen is. Bij inzet kan men bijvoorbeeld denken aan de intensiteit waarmee leerlingen zich inzetten voor een vak in de klas of de volharding waarmee zij zich inzetten voor huiswerk. In dat licht hoort toekomstperspectief tot de verzameling verklarende variabelen voor bijvoorbeeld de mate waarin leerlingen zich inzetten bij het maken van huiswerk.

Uitwerking van het begrip toekomstperspectief

Het begrip toekomstperspectief wordt doorgaans omschreven als een voorstelling of conceptualisering van een levensterrein of object, bijvoorbeeld van een beroepsloopbaan of van sociale relaties. Het wordt gekenmerkt door de aspecten 'extensie' en 'valentie' (zie o.m. Winnubst, 1975). Bij perspectief gaat het om een voorstelling van een voorval of van een object in de verre of nabije toekomst. De extensie geeft daarbij aan hoe ver in de toekomst de voorstelling gaat. Wat een zinvolle afbakening van de tijd (extensie) is, kan van onderwerp tot onderwerp verschillen. Voor leerlingen vormen 'de tijd na deze school' of 'dit lopende schooljaar' zinvolle afbakeningen in de tijd. De valentie of relevantie van de variabele 'toekomstperspectief' geeft de waarde aan van een object in de toekomst. In de omschrijving van het begrip toekomstperspectief als een motivationele variabele speelt iemands waardering voor een object in de toekomst een essentiële rol.

Het begrip toekomstperspectief wordt in dit onderzoek opgevat als een variant van het begrip 'attitude' met de drie componenten: affectie, cognitie en gedragsintentie (Rosenberg & Hov-

* Adres: Grote Bickersstraat 72, 1013 KS Amsterdam.

land, 1960). Toekomstperspectief wordt gezien als een attitude ten aanzien van een object op een zekere termijn in de tijd. Kennis (cognitie) wordt hier opgevat enerzijds als de aanwezigheid van ideeën en/of verwachtingen over de toekomst, anderzijds als kennis van de maatschappelijke realiteit. Gevoel (affectie) wordt opgevat als een optimistische danwel een pessimistische benadering van een object in de toekomst en ook als een positieve danwel negatieve waardering. De attitude-component 'gedrag' wordt in de context van het begrip toekomstperspectief in de eerste plaats gezien als een 'gedragsintentie'.

Probleemstelling

Zoals aangegeven ligt er in de literatuur een sterke nadruk op het motivationele karakter van toekomstperspectief (zie o.a. Nuttin, 1980 en Lens, 1987). De verwachting luidt dat toekomstperspectief als een goede voorspeller van inzet voor school kan dienen. Verondersteld wordt dat inzet voor school deel uitmaakt van gedrag (de uitvoering van een reeks taken) gericht op een positief resultaat voor school en mogelijk op een beroepsloopbaan in de toekomst. Het gedrag bestaat dan uit opeenvolgende stappen die implicaties hebben voor een doel in de toekomst. Roede (1986) operationaliseert inzet voor school als de aanzet, intensiteit en volharding (duur) van drie verschillende soorten gedrag, namelijk: naar school gaan, voor een vak werken, en huiswerk maken. In dit onderzoek zijn als objecten van toekomstperspectief opgenomen: school- en beroepsloopbaan, sociale relaties, persoonlijke ontwikkeling en vrije tijd. In onderzoeken van Peetsma (1985a) en Sannen (1990) bleek het belang van deze onderwerpen voor leerlingen.

De concrete verwachting in dit onderzoek luidt dat leerlingen met positievere toekomstperspectieven op een beroepsloopbaan op korte en lange termijn (reikend tot na de huidige school) zich meer inzetten voor school dan leerlingen met negatievere perspectieven. Bovendien is de verwachting, onder meer gebaseerd op de onderzoeken van McClelland (1961) en Hermans (1967) naar kenmerken van zeer prestatiegemotiveerde mensen, dat leerlingen met positieve toekomstperspectieven in het algemeen (dus op al de vier objecten: school- en beroepsloopbaan, sociale relaties, persoonlijke ontwikkeling en vrije tijd) zich meer inzetten voor school dan leerlingen met negatievere perspectieven. Andere variabelen waarbij vaak samenhangen met toekomstperspectief of motivatie zijn gevonden (zie o. m. Peetsma, 1985a, 1985b & 1992) zijn: schooltype, leerjaar, sekse, sociaal-economische status (ses), locus of control en faalangst. Deze zijn opgenomen in het onderzoek als achtergrondvariabelen.

De uitwerking van de probleemstelling leidt tot de volgende specifieke vraagstellingen:

1. Zetten leerlingen met positieve toekomstperspectieven voor een school- en beroepsloopbaan op korte en lange termijn zich meer in voor school dan leerlingen bij wie deze toekomstperspectieven negatiever zijn?
2. Zetten leerlingen met positieve toekomstperspectieven voor alle vier objecten zich meer in voor school dan leerlingen waarbij deze toekomstperspectieven negatiever zijn?
3. In hoeverre wordt de samenhang tussen de toekomstperspectieven en gemotiveerdheid voor school beïnvloed door de variabelen schooltype, leerjaar, sekse, sociaal-economische status, locus of control en faalangst?

METHODE

Subjecten

Er nemen 606 leerlingen deel aan het onderzoek. Dit betreft 270 jongens en 327 meisjes. Van 9 leerlingen is de sekse niet bekend. De leeftijd van de leerlingen varieert van 11 tot 20 jaar. De leerlingen komen uit 30 afdelingen van 21 scholen(gemeenschappen). Van de 606 leerlingen volgen er 138 een lbo opleiding, 199 mavo, 145 havo of brugklas havo/vwo en 124 vwo. De representativiteit van de onderzoeksgroep wat betreft de aantallen leerlingen per schooltype is achteraf nagegaan door vergelijking met de gegevens hierover in de CBS-statistieken van het schooljaar 1986-1987. Doordat veel scholen(gemeenschappen) een brugklas-systeem kennen, is bij een groot aantal leerlingen nog geen 'schooltype' bekend. Er is een grove verdeling van deze

leerlingen over de schooltypen uitgevoerd op basis van de verdeling van leerlingen waarvan wel een schooltype bekend is. Zo kunnen de percentages leerlingen per schooltype in de onderzoeksgroep vergeleken worden met de percentages in de populatie. De indeling van de leerlingen in de onderzoeksgroep levert de volgende percentages op: lbo 23%, mavo 33%, havo 24% en vwo 20%. Voor de populatie zijn de percentages als volgt: lbo 30%, mavo 30%, havo 19% en vwo 21%. Opvallende verschillen tussen de percentages van een schooltype in de onderzoeksgroep en in de populatie doen zich niet voor. Het verschil van 7% bij het lbo is het grootst. De conclusie is gewettigd dat de representativiteit van de steekproef goed is.

Instrumenten

Alle meetinstrumenten in dit onderzoek zijn vragenlijsten. De antwoordmogelijkheden bij de vragen liggen op een vijfpunts Likert-type schaal, lopend van 'past helemaal bij mij' tot 'past helemaal niet bij mij'. Alleen bij de Inzet vragenlijst lopen de antwoordmogelijkheden van '(bijna) nooit' tot '(bijna) altijd'.

Toekomstperspectief

De Toekomstperspectief Vragenlijst bevat 48 items. Als aspecten van het begrip toekomstperspectief zijn object, extensie en valentie opgenomen. Deze drie onderdelen van het begrip zijn op de volgende wijze in de vragenlijst systematisch gevarieerd via een facet-benadering: namelijk

- a. Onderwerpen/objecten op vier niveau's: school- en beroepsloopbaan; sociale relaties; persoonlijke ontwikkeling en vrije tijd;
- b. Termijn/extensie op twee niveau's: de tijd nu, dit lopende schooljaar, op school en later na deze school;
- c. Waardering/valentie op twee niveau's: positieve en negatieve waarde.

De vragenlijst bestaat dus uit $4 \times 2 \times 2$ combinaties. Bij de analyses voor deze studie zijn de positieve en de negatieve valentie, na omschaling, samengenomen, waardoor acht subschalen ontstaan die elk uit zes items bestaan. Drie items van deze subschalen zijn positief gericht en drie negatief. Uitzondering daarop vormen de subschalen 'sociale relaties op lange termijn' en 'persoonlijke ontwikkeling op lange termijn'. Die bevatten vier positieve en twee negatieve items. Als voorbeeld volgen hier enige items van verschillende subschalen.

School en beroepsloopbaan op lange termijn: 'Het kan me weinig schelen wat ik later, na deze school, voor werk of studie kan doen.'

School en beroepsloopbaan op korte termijn: 'Ik vind het fijn dat ik er dit schooljaar weer allerlei dingen bij leer.'

Sociale relaties op lange termijn: 'Het lijkt me goed om later na deze school veel contact te hebben met vrienden en vriendinnen.'

De schalen van de Toekomstperspectief Vragenlijst hebben een voldoende tot goede psychometrische kwaliteit. De schaalhomogeniteit (Cronbach's alpha's) loopt van .70 tot .86. Alle schalen zijn scheef verdeeld naar de positieve kant (zie Bollen, 1989, blz. 206-218). De 'skewness' varieert tussen -.165 en -.994. Voor nadere gegevens over de validiteit van de toekomstperspectief vragenlijst wordt verwezen naar Peetsma (1985a), Roede (1986) en Peetsma (1992).

Inzet voor school

De Inzet Vragenlijst is overgenomen uit onderzoek van Roede (1986). Op basis van wat de leerlingen zelf aangeven over aspecten van hun gedrag wordt een maat vastgesteld voor (1) hun inzet bij het werken voor een vak in de klas ('Ik vind het leuk om in de klas iets te gaan doen voor Engels'), (2) bij het maken van huiswerk voor een vak ('Ik raffel mijn huiswerk voor Engels af'), (3) inzet voor school in het algemeen ('Op school denk ik over dingen buiten school'), en (4) inzet bij het maken van huiswerk in het algemeen ('Ik kan goed mijn aandacht bij het huiswerk houden'). Elke schaal bestaat uit 14 items. Per schaal wordt met evenveel positieve als negatieve items nagegaan (1) de aanzet/richting van het gedrag; (2) de intensiteit

waarmee iets wordt gedaan; en (3) de volharding/persistentie waarmee iets wordt gedaan. Per schaal wordt een totaalscore van aanzet, intensiteit en volharding berekend.

De psychometrische kwaliteit van de vragenlijst is goed: de homogeniteit van de vier subschalen varieert van .85 tot .92. Drie van de vier schalen zijn scheef verdeeld naar de positieve kant (skewness varieert van -.335 tot -.615). De schaal voor inzet voor school in het algemeen is enigszins scheef verdeeld naar de negatieve kant. De skewness bedraagt .479. Resultaten van verschillende onderzoeken naar validiteit en betrouwbaarheid van het instrument staan beschreven in Roede (1986).

Faalangst

Voor het meten van de variabele 'negatieve faalangst' is van de Achievement Motivation Scale (AMS) (Nygard & Gjesme, 1973) de schaal opgenomen die de vermijdingstendens (MF) meet. Deze bestaat uit 15 uitspraken. De uitspraken slaan niet op concrete schoolse situaties, maar gelden meer algemeen (bijvoorbeeld 'Ik vind situaties vervelend waarin ik moet laten zien wat ik kan'). De homogeniteit van de schaal is goed (Cronbach's alpha is .90).

Locus of control

Voor het meten van de mate van interne locus of control is een Nederlandse vertaling en bewerking gebruikt van één van de drie schalen van de vragenlijst 'Spheres of Control' (Paulhus & Christie, 1981): persoonlijke doeltreffendheid in prestaties. Daarvan zijn de zes item gebruikt die de interne locus of control meten. Een voorbeelditem is: 'Als ik maar echt wil, kan ik alles leren.' De homogeniteit van de schaal is voldoende (Cronbach's alpha is .69).

Biografische gegevens van de leerlingen

Voor de indeling van de biografische gegevens over de sociaal-economische status van de leerlingen is gebruik gemaakt van de Beroepenklapper (ITS) (Van Westerlaak, Kropman & Collaris, 1975).

Procedure

Het onderzoek is uitgevoerd bij leerlingen van alle leerjaren van het voortgezet onderwijs. Een proefleider heeft de vragenlijsten in schooltijd afgenomen. Bij de selectie van de scholen is gezorgd voor een evenredige vertegenwoordiging van de deelnemende schooltypen, te weten lbo, mavo, havo en vwo. Bovendien is gezorgd voor een spreiding over urbanisatiegraad (stads- en streekscholen), denominatie en regio. De scholen werd verzocht om per klas een klein aantal (2 tot 4) leerlingen deel te laten nemen. Daarmee is een klas-effect op de resultaten voorkomen. De leerlingen zouden ofwel de betere en de slechtere moeten zijn of de meer gemiddelde leerlingen van de klas. Afhankelijk van de grootte van de school is om een totaal aantal leerlingen verzocht van 20 tot 40. Dit laatste moet een school-effect voorkomen. Klas- en school-effecten kunnen optreden doordat leerlingen van eenzelfde klas en school een aantal kenmerken gemeen hebben. Voorbeelden daarvan zijn zitten in een 'slechte' klas of op een 'goede' school, de sociale structuur en het onderwijsklimaat van de klas en de school: kenmerken die van invloed kunnen zijn op de inzet en de toekomstperspectieven van de leerlingen en op de relatie tussen toekomstperspectieven en inzet. Door de selectieprocedure willen we klas- en school-effecten vermijden en daardoor voorkomen dat een systematische vertekening van de resultaten optreedt. Het verzamelen van de gegevens op school vond plaats in de winter van november 1986 tot februari 1987.

RESULTATEN

Toekomstperspectieven en inzet voor school

Er zijn product-moment correlatiecoëfficiënten berekend tussen de schalen van de Toekomstperspectief vragenlijst en de Inzetvragenlijst (Tabel 1) voor de totale groep leerlingen (n=606).

Tabel 1. Product-moment correlaties (*) tussen facetten van toekomstperspectief en inzet voor school, berekend voor de totale groep (n = 606).

Inzet voor:	Engels in de klas	huiswerk Engels	school algemeen	huiswerk algemeen
toekomst perspectief: korte termijn				
vrije tijd	.03	.03	-.07	.01
eigen ontwikkeling	.21	.20	.20	.23
sociale relaties	.06	.09	.04	.07
S- en B-loopbaan	.50	.48	.50	.55
toekomst perspectief: lange termijn				
vrije tijd	-.15	-.24	-.24	-.23
eigen ontwikkeling	.16	.10	.11	.15
sociale relaties	.09	.11	.09	.13
S- en B-loopbaan	.35	.34	.32	.39

* Bij n = 500 $p(r) = 0.01$ $r = .10$

Uit de resultaten blijkt duidelijk een verband tussen alle schalen voor inzet en de toekomstperspectieven op een school- en beroepsloopbaan (variërend van .32 tot .55). Het verband met het perspectief op korte termijn is over de hele linie sterker (.48 en .55) dan met het perspectief op lange termijn (.32 en .39), maar ook bij deze laatste is de positieve samenhang duidelijk. Ook na berekening van de voor attenuatie gecorrigeerde correlaties blijft dit verschil bestaan. Deze resultaten ondersteunen dus de eerste verwachting dat leerlingen met positieve toekomstperspectieven op school- en beroepsloopbaan zich meer inzetten voor school en huiswerk in het algemeen en voor een bepaald vak (Engels) dan leerlingen met negatievere perspectieven.

Bij het perspectief op vrije tijd op korte termijn vinden we geen samenhang met inzet voor school en bij dit perspectief op lange termijn zien we een negatieve samenhang (variërend van -.15 tot -.24). De perspectieven op persoonlijke ontwikkeling op korte en lange termijn hangen positief samen met de inzet- metingen, maar bij de perspectieven op sociale relaties is alleen de samenhang met het perspectief op lange termijn zwak significant. We moeten dus concluderen dat er onvoldoende ondersteuning is voor de tweede verwachting dat leerlingen met positieve toekomstperspectieven op alle vier objecten zich meer inzetten voor school. Vooral de perspectieven op vrije tijd zijn daaraan debet. De negatieve correlaties suggereren dat perspectieven op vrije tijd samenhangen met motivationele tendensen die competitief zijn met studiemotivatie (Atkinson & Birch, 1970; Roede, 1989). De relatie tussen toekomstperspectief en inzet voor school blijkt duidelijk niet voor elk willekeurig object van het perspectief op te gaan.

Samenhang met achtergrondvariabelen

Het verband tussen toekomstperspectief en inzet voor school is op een andere wijze nogmaals geanalyseerd. Exploratief, maar voortbouwend op de onderzoeksliteratuur en bevindingen van de bivariate analyses (zie Peetsma, 1992), is daarbij nagegaan in hoeverre de achtergrondvariabelen de gevonden relatie tussen toekomstperspectief en inzet voor school beïnvloeden.

Voor de analyse van de verbanden tussen toekomstperspectief en inzet wordt gebruik gemaakt van het LISREL VI programma¹ (Jöreskog en Sörbom, 1986). Een leesbare introductie in dit programma en de techniek is te vinden in Bollen (1989). Om het meetmodel te definiëren worden de vier schalen voor inzet opgevat als vier metingen van één latente variabele: 'ware' inzet of gemotiveerdheid voor school. De toekomstperspectieven op korte en lange termijn op éénzelfde object zijn opgenomen als twee metingen van één latente variabele: het toekomstperspectief op dat ene object. Zo ontstaan vier latente toekomstperspectieven: toekomstperspectief op vrije tijd, toekomstperspectief op persoonlijke ontwikkeling, toekomstperspectief op sociale relaties en toekomstperspectief op school- en beroepsloopbaan.

Om na te gaan of achtergrondvariabelen invloed hebben op het verband tussen toekomstperspectieven en inzet voor school zijn subgroepen gevormd voor verschillende niveau's van de achtergrondvariabelen: bijvoorbeeld meisjes en jongens of hoog- en laag-faalangstigen. Het perspectief-inzet model wordt gelijktijdig voor de vergeleken groepen geanalyseerd. Gebruik is hierbij gemaakt van de multi-groep optie uit het LISREL programma (Bollen, 1989, blz 355-369). De hypothese die steeds werd getoetst was of de regressie-coëfficiënten constant zijn over groepen. Hiertoe werd de aannemelijkheid van een model met invariante regressie-coëfficiënten vergeleken met de aannemelijkheid van een model waarbij de regressie-coëfficiënten variëren. Het meetmodel wordt in de groepen constant gehouden zodat de latente variabelen en de regressie-coëfficiënten vergelijkbaar zijn. Onder aanname van multivariaat-normaal verdeelde scores en een voldoende grote steekproef, kan gebruik worden gemaakt van de log-likelihood test: het verschil van de chi-kwadrat waarden van de vergeleken modellen volgt bij benadering weer een chi-kwadrat verdeling, met als vrijheidsgraden het verschil in graden van de vergeleken modellen. De waarschijnlijkheid ervan kan dus worden bepaald (verg. Bollen, 1989, hoofdstuk 4). Om de interpretatie van de resultaten te vergemakkelijken zijn alle variabelen gestandaardiseerd.²

Daar bij de analyse van de psychometrische kwaliteit van de acht metingen van toekomstperspectief en vier metingen voor inzet is gebleken dat de variabelen enigszins scheef verdeeld zijn moeten we vertrouwen op de gerapporteerde robuustheid van de maximum likelihood (ML) methode bij (milde) schending van de aanname van normaliteit (zie Bollen, 1989, hoofdstuk 9). Om minder afhankelijk te zijn van de normaliteitsaanname kunnen de modelparameters geschat worden volgens de Unweighted Least Squares (ULS) methode. Helaas krijgt men bij ULS-schattingen geen chi-kwadrat grootte die getoetst kan worden. Daarom zijn voor de analyses ML-schattingen gebruikt. Ter controle is echter ook gebruik gemaakt van de Unweighted Least Squares methode om modelparameters te schatten.

Het perspectief-inzet model

Een overzicht van geschatte parameters in het model staat in Figuur 1.

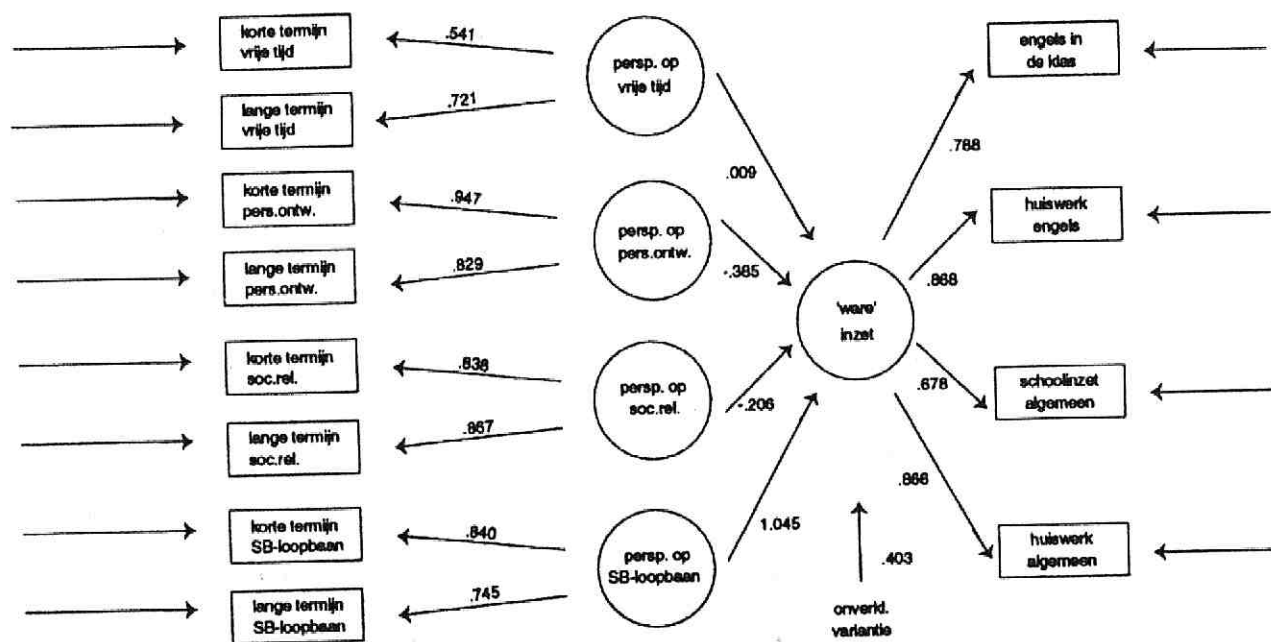


Fig. 1. Perspectief-Inzet model met vier perspectieven. Gestandaardiseerde oplossing.

* De vier latente perspectief variabelen zijn onderling gecorreleerd.

Het model past redelijk (χ^2 -kwadraat = 509.35, $df = 278$, $p = .00$). De Goodness-of-Fit Index is .85. Bij de presentatie van het model is uitgegaan van een gestandaardiseerde oplossing, waarin de varianties van de latente variabelen op 1 zijn gesteld. In de figuur zien we dat het perspectief op vrije tijd niet bijdraagt aan de verklaring in variantie van de 'ware' inzet. Het perspectief op vrije tijd wordt echter zwak gemeten. Vooral het perspectief op korte termijn representeert deze variabele matig. Het is met name deze korte termijn schaal die hoge restcorrelaties veroorzaakt en daarmee de matige modelfit. We vinden negatieve bijdragen van de perspectieven op eigen ontwikkeling en sociale relaties. Het perspectief op school- en beroepsloopbaan draagt zeer sterk (1.045) bij aan de verklaarde variantie in 'ware' inzet voor school. Aan onverklaarde variantie van de 'ware' inzet-variantie blijft over 40%. Dit houdt in dat ongeveer 60% van de 'ware' inzet-variantie wel wordt verklaard door de perspectieven. Bij de vier metingen van inzet voor de variabele 'ware' inzet voor school zien we een iets lagere bijdrage van de schaal 'school-inzet in het algemeen' dan van de drie andere schalen. In de ongestandaardiseerde oplossing zijn ook standaardfouten verkregen. Alleen voor vrije tijd ligt nul in het 95% betrouwbaarheidsinterval.

De geschatte correlatie-matrix van latente perspectief-variabelen staat in Tabel 2. De correlatie van het perspectief op vrije tijd met die op school- en beroepsloopbaan is negatief.

Bij de LISREL-analyse wordt bevestigd wat uit eerdere analyses bleek: Het toekomstperspectief op vrije tijd past niet goed in de structuur van de andere drie toekomstperspectieven. Aan de verklaring van inzet voor school draagt het niet bij. In de volgende LISREL-analyses is het toekomstperspectief op vrije tijd niet meer in het perspectief-inzet model opgenomen, vooral ook omdat het een ongunstige invloed heeft op de modelfit.

Perspectief-Inzet model met drie perspectieven

We zullen nu een model met drie toekomstperspectieven (toekomstperspectief op persoonlijke ontwikkeling, toekomstperspectief op sociale relaties en toekomstperspectief op school- en beroepsloopbaan) schatten. Een overzicht van geschatte parameters in dit model met invariante parameters staat in Figuur 2.

In principe is hetzelfde resultaat verkregen als eerder, maar nu zijn twee variabelen weggelaten. In het geheel past het model redelijk (χ^2 -kwadraat = 261.16, $df = 84$, $p < .01$). De goodness of fit index is .92. Het gemiddelde residu is .068. Ook bij de presentatie van dit model is uitgegaan van een gestandaardiseerde oplossing, waarin de varianties van de latente variabelen op 1 zijn gesteld. Het model is zeer waarschijnlijk geïdentificeerd: de standaard-fouten konden worden berekend. In Figuur 2 zien we geen perspectief meer dat niet bijdraagt aan de verklaring in variantie van de 'ware' inzet. De bijdragen van de drie resterende perspectieven komen overeen met hun bijdragen in het eerste model met vier perspectieven. We vinden ook in dit model negatieve bijdragen van de perspectieven op eigen ontwikkeling en sociale relaties. Het perspectief op school- en beroepsloopbaan draagt ook hier zeer sterk (1.063) bij aan de verklaarde variantie in 'ware' inzet voor school.

Tabel 2. Geschatte correlatie-matrix van latente variabelen perspectief op vrije tijd, perspectief op persoonlijke ontwikkeling, perspectief op sociale relaties en perspectief op school- en beroepsloopbaan bij model met vier perspectieven.

	perspectief op vrije tijd	perspectief op pers. ontw.	perspectief op soc. rel.	perspectief op SB- loopbaan
perspectief op vrije tijd	1.0			
perspectief op pers. ontw.	.252	1.0		
perspectief op soc. rel.	.453	.427	1.0	
perspectief op SB-loopbaan	-.107	.694	.441	1.0

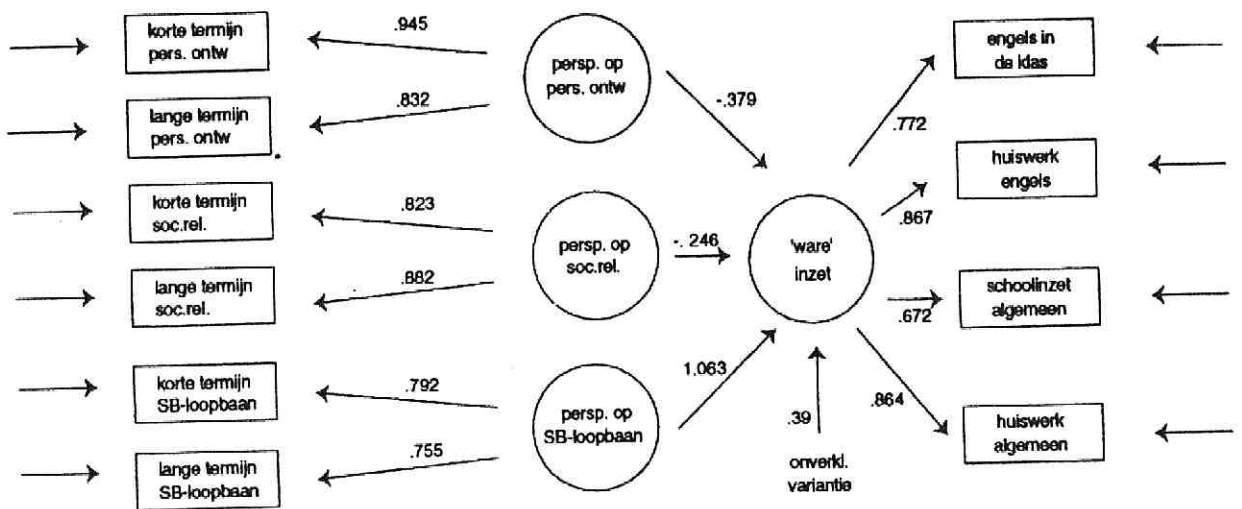


Fig. 2. Perspectief-Inzet model met drie perspectieven. Gestandaardiseerde oplossing.
* De drie latente perspectief variabelen zijn onderling gecorreleerd.

De variabelen zijn enigszins scheef verdeeld. Daarom is het perspectief-inzet model ook geanalyseerd volgens een methode waarin de maximum-likelihood-schatting is vervangen door schattingen met de ULS methode. Daaruit blijkt dat de regressie-schattingen van inzet op perspectieven nauwelijks veranderen. De Goodness-of-Fit Index blijkt .983, wat op een redelijk model duidt.

Bijna 40 procent van variantie van de 'ware' inzet blijft in het perspectief-inzet model onverklaard. Dit houdt in dat ruim 60 procent van de 'ware' inzet-variantie wordt verklaard door de perspectieven. De geschatte correlatie-matrix van latente perspectief-variabelen staat in Tabel 3. De correlaties komen nauw overeen met de correlaties bij het model met vier variabelen.

Invloed van achtergrondvariabelen

Als volgende stap is een aantal analyses uitgevoerd met het perspectief-inzet model met drie perspectieven. Onderzocht is of de relatie tussen toekomstperspectief en motivatie afhankelijk is van schooltype, leerjaar, sekse, sociaal-economische achtergrond (SES), locus of control, en faalangst. Bij elke analyse werd de steekproef in subgroepen verdeeld. Getoetst wordt of een restrictief model met identieke parameterwaarden voor beide groepen voldoet, of dat een minder restrictief model met verschillende parameterwaarden moet worden aangenomen.

Tabel 3. Geschatte correlatie-matrix van latente variabelen perspectief op persoonlijke ontwikkeling, perspectief op sociale relaties en perspectief op school- en beroepsloopbaan bij model met drie perspectieven.

	perspectief op pers. ontw.	perspectief op soc. rel.	perspectief op SB-loopbaan
perspectief op pers. ontw.	1.0		
perspectief op soc. rel.	.443	1.0	
perspectief op SB-loopbaan	.704	.457	1.0

Voor de invloed van het schooltype (lbo, mavo, havo of vwo) werd het perspectief-inzet model twee keer geanalyseerd met LISREL VI. Een model met verschillende parameterwaarden voor de vergeleken groepen voldeed niet significant beter dan een model met identieke parameterwaarde. Dit resultaat geeft aan dat schooltype geen invloed heeft op het verband tussen toekomstperspectief en inzet voor school.

Voor wat betreft de invloed van leerjaar op de samenhang tussen toekomstperspectief en inzet, op basis van de resultaten van bivariate analyses (Peetsma, 1992) is besloten de leerjaren niet allemaal afzonderlijk te nemen maar in drie groepen in te delen: alle eerste leerjaren in groep 1, alle laatste leerjaren (eindexamenklassen) in groep 3 en de tussenliggende leerjaren in groep 2. Vergelijking van het restrictieve en minder restrictieve model liet zien dat het toestaan van verschillende parameterwaarden in de vergeleken groepen geen verbetering in de fit opleverde ten opzichte van het model zonder effecten van leerjaar. Daaruit kunnen we concluderen dat de regressies van inzet op perspectieven niet afhangen van de leerjaren. Leerjaar (schoolfase) heeft geen invloed op het gevonden verband tussen toekomstperspectief en inzet voor school.

Ook voor wat betreft de invloed van sekse is een model waarbij verschillende parameterwaarden zijn toegestaan voor jongens en meisjes nauwelijks een verbetering van de fit ten opzichte van het model zonder effecten van sekse. Daaruit vloeit de conclusie voort dat de regressies van inzet op perspectieven niet afhangen van sekse.

Om de invloed van sociaal-economische status (SES) te onderzoeken zijn drie groepen gevormd op basis van het beroep van de vader: laagste, middelste en hoogste sociale status. Opnieuw betekent een minder restrictief model nauwelijks een verbetering van de fit van het model. De sociaal-economische status van leerlingen heeft geen invloed op het gevonden verband tussen toekomstperspectief en inzet voor school.

Voor wat betreft de invloed van faalangst is er een drie-deling gemaakt in de mate waarin de leerlingen negatief faalangstig zijn: een groep met weinig, een groep met gemiddeld, en een groep met veel negatieve faalangst. Ook nu kan worden geconcludeerd dat de regressies van inzet op perspectieven niet afhangen van de mate van negatieve faalangst bij de leerlingen. Faalangst heeft geen invloed op het gevonden verband tussen toekomstperspectief en inzet voor school.

Voor wat betreft toetsing van de samenhang met locus of control is er een drie-deling gemaakt in de mate waarin de leerlingen de controle over hun effectiviteit in prestaties bij zichzelf leggen (interne locus of control): een groep met weinig, een groep met gemiddeld, en een groep met veel interne controle. Bij vergelijking van de beide modellen vinden we een chi-kwadraat van 19.56 met een verschil-df van 8 ($p < .01$). Dit betekent dat het toestaan van verschillende waarden in de vergeleken groepen voor de regressies van inzet op perspectief een statistisch significante verbetering van de modelfit betekent. Daaruit vloeit de conclusie voort dat de mate waarin de leerlingen de controle over hun effectiviteit in prestaties bij zichzelf leggen invloed heeft op het gevonden verband tussen toekomstperspectief en inzet voor school.

Uit Tabel 4 volgt dat hoe groter de waargenomen interne controle is, des te negatiever is de regressie van inzet op het perspectief op persoonlijke ontwikkeling. Hoe minder de ervaren interne controle is, des te negatiever is de regressie van inzet op het perspectief op sociale

Tabel 4. Regressies van inzet op de drie perspectieven en de hoeveelheid onverklaarde variantie, voor de drie locus of control-groeperingen.

	regressie van inzet op persp. pers. ontw.	regressie van inzet op persp. soc. rel.	regressie van inzet op persp. SB-loopb.	percentage onverklaarde variantie
locus weinig intern	-.123	-.307	.926	.328
locus gemiddeld intern	-.326	-.234	1.079	.269
locus sterk intern	-.493	-.048	.609	.609

relaties. Bij leerlingen die de controle over prestaties vooral bij zichzelf leggen, vinden we een negatief verband tussen hun perspectief op persoonlijke ontwikkeling en hun inzet voor school. Als ze een heel positief perspectief hebben op hun persoonlijke ontwikkeling zetten ze zich minder in voor school en, omgekeerd, als ze een nogal negatief perspectief hebben op hun persoonlijke ontwikkeling zetten ze zich meer in voor school. Bij leerlingen die de controle weinig bij zichzelf leggen vinden we echter een negatief verband tussen hun perspectief op sociale relaties en hun inzet voor school. Als ze een heel positief perspectief hebben op hun sociale relaties zetten ze zich minder in voor school en omgekeerd als ze een nogal negatief perspectief hebben op hun sociale relaties zetten ze zich meer in voor school.

Voor de groep leerlingen die de controle het meest bij zichzelf legt, blijkt de regressie van inzet op het perspectief op school-en beroepsloopbaan lager dan bij de beide andere groepen. Bij deze groep leerlingen heeft het perspectief op school- en beroepsloopbaan minder verband met de mate waarin ze zich inzetten voor school. Tegelijk blijkt dat voor de groep leerlingen met veel interne controle de meeste inzet-variantie (60%) onverklaard blijft, terwijl de hoeveelheid onverklaarde variantie voor de groepen met gemiddeld en weinig interne controle ongeveer 30% bedraagt.

DISCUSSIE

Toekomstperspectief is gedefinieerd als een attitude die iemand heeft ten aanzien van een object op een zekere termijn in de toekomst. Verondersteld is dat toekomstperspectief een motivationeel effect heeft. Ten behoeve van het motivationele karakter van het begrip toekomstperspectief zijn alle drie aspecten van een attitude (affectie, cognitie en intentie/gedrag) opgenomen in de omschrijving van het begrip (Peetsma, 1992).

In eerder onderzoek (Peetsma, 1985a) bleek het toekomstperspectief vrij sterk object-gebonden te zijn. De perspectieven op korte en lange termijn hingen sterk samen. Dezelfde object-gebondenheid van de perspectieven komt terug in deze resultaten. De termijn-gebondenheid is ook hier minder sterk dan de object-gebondenheid. We moeten concluderen dat het object een belangrijker rol heeft binnen de toekomstperspectieven van leerlingen dan de termijn.

De conclusie over de hoofdvragen van het onderzoek moet luiden dat de verwachtingen voor een deel zijn bevestigd: er is een verband gevonden tussen inzet en de toekomstperspectieven op school-en beroepsloopbaan, maar niet tussen inzet en toekomstperspectief in het algemeen (toekomstperspectieven op alle vier objecten). Het toekomstperspectief op vrije tijd draagt niet bij aan de verklaring van variantie in inzet. Met name het perspectief op vrije tijd op lange termijn lijkt zelfs averechts te werken. Indachtig de betekenis van vrije tijd voor veel leerlingen is dit resultaat niet zo heel verwonderlijk. Toch is het toetsen van de hypothese over een motiverende werking van de perspectieven op de vier objecten de moeite waard geweest. In verschillende theoriën (o.a. McClelland, 1961 en Hermans, 1967) gaat men uit van een algemeen toekomstperspectief dat motiverend zou zijn. Strikt genomen zal het toekomstperspectief op vrije tijd daaronder vallen. Deze redenering blijkt echter meer theoretisch dan levensvatbaar en moet dus worden verworpen. Toekomstperspectief blijkt bij leerlingen wel degelijk object-gebonden, ook als voorspeller van inzet voor school. De toekomstperspectieven ten aanzien van school- en beroepsloopbaan, sociale relaties en persoonlijke ontwikkeling, in geval korte en lange termijn perspectieven worden samengenomen, dragen samen ruim 60% bij aan verklaarde variantie in inzet. Dit is een hoog percentage.

Bij analyse van de invloed van een aantal variabelen op het verband tussen toekomstperspectieven van leerlingen en hun inzet voor school blijken schooltype, sekse, leerjaar (schoolfase), sociaal-economische achtergrond noch de mate van faalangst er toe te doen. Dat betekent bijvoorbeeld dat er noch tussen jongens en meisjes verschil is in de sterkte van het verband tussen toekomstperspectief en inzet, noch tussen zeer faalangstige en weinig faalangstige leerlingen. Alleen de mate waarin de leerlingen de controle over prestaties bij zichzelf leggen lijkt

het verband tussen toekomstperspectief en inzet te beïnvloeden. Voor de groep leerlingen die de controle het meest bij zichzelf legt, heeft het perspectief op school- en beroepsloopbaan minder verband met de mate van inzet voor school dan bij de beide andere groepen leerlingen. Tevens blijkt dat voor de groep leerlingen met veel interne controle de meeste inzet-variantie (60%) onverklaard blijft, terwijl de hoeveelheid onverklaarde variantie voor de groepen met gemiddeld tot weinig interne controle ongeveer 30% bedraagt. Het hoge percentage verklaarde variantie in inzet voor school maakt nieuwsgierig naar verdere bestudering van de perspectief-inzet relatie, maar ook naar mogelijke toepassingen van de bevindingen.

NOTEN

1. Ik dank Dr. M. Eiting en drs. T.M. Bechger voor hun adviezen bij de toepassing van dit programma.
2. Deze wijze van analyseren kan worden opgevat als een uitbreiding van covariantie-analyse. Technische informatie hierover kan de lezer vinden in enkele artikelen van Dag Sörbom (1978 en 1982).

LITERATUUR

- Atkinson, J.W., & Birch, D. (1970). *The dynamics of action*. New York: Wiley.
- Bollen, K.A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Frank, L.K. (1939). Time Perspectives. *Journal of Social Philosophy*, 4, 293-312.
- Hermans, H. (1967). *Motivatatie en Prestatie* (proefschrift). Amsterdam: Swets & Zeitlinger.
- Jöreskog, K.G. and Sörbom, D. (1986). *LISREL VI; Analysis of Linear Structural Relationships by Maximum Likelihood, Instrumental Variables and Least Squares Methods*. Uppsala: University of Uppsala.
- Lens, W. (1987). Future time perspective, motivation, and school performance. In E. de Corte, J. Lodewijks, R. Parmentier, & P. Span (eds.). *Learning and Instruction: European research in an international context* (blz. 181-189). Leuven/Oxford: Leuven University Press/Pergamon.
- McClelland, D. (1961). *The achieving society*. Princeton: Van Nostrand.
- Nuttin, J.R. (1980). *Theorie de la motivation humaine*. Paris: Presses Universitaire de France.
- Nygaard, R. & Gjesme, T. (1973). Assessment of achievement motives: Comments and suggestions. *Scandinavian Journal of Educational Research*, 17, 39-46.
- Paulhus, D.L., & Christie, R. (1981). Spheres of control: An interactionist approach to assessment of perceived control. In H.M. Lefcourt (ed.). *Research with the locus of control construct* (Vol. 1. Assessment Methods, blz. 161-188). New York: Academic Press.
- Peetsma, T.T.D. (1985a). *Meting van toekomstperspectief bij leerlingen van mavo, havo en vwo*. Amsterdam: Stichting Centrum voor Onderwijsonderzoek.
- Peetsma, T.T.D. (1985b). *Toekomstperspectief*, literatuurstudie. (SCO-cahier, no.34) Amsterdam: Stichting Centrum voor Onderwijsonderzoek.
- Peetsma, T.T.D. (1992). *Toekomst als motor? toekomstperspectieven van leerlingen in het voortgezet onderwijs en hun inzet voor school*. Amsterdam: Stichting Centrum voor Onderwijsonderzoek.
- Roede, E. (1986). *Gemotiveerdheid en ongemotiveerdheid van mavo-, havo- en vwo-leerlingen*. Amsterdam: Stichting Centrum voor Onderwijsonderzoek.
- Roede E. (1989). *Explaining Student Investment; an investigation of high school students' retrospective causal accounts of their investment in school*. Amsterdam: Stichting Centrum voor Onderwijsonderzoek.
- Rosenberg, M.J. & Hovland, C.I. (1960). Cognitive, affective, and behavioral components of attitudes. In: C.I. Hovland & M.J. Rosenberg (eds.). *Attitude organization and change* (blz. 1-14). New Haven, Conn.: Yale University Press.
- Sannen, H. (1990). *'Mijn toekomst tot 2030': Verwachtingen van jeugdigen*. Tilburg: Katholieke Universiteit Brabant.
- Sörbom, D. (1978). An alternative to the methodology for analysis of covariance. *Psychometrika*, 43, 381-396.
- Sörbom, D. (1982). Structural equation models with structured means. In K.G. Jöreskog & H. Wold (eds.). *Systems under indirect observation*. Amsterdam: Elsevier.

- Westerlaak, J.M. van, Kropman, J.A., & Collaris, J.W.M. (1975). *Beroepenklapper*. Nijmegen: Instituut voor Toegepaste Sociologie.
- Winnubst, J.A.M. (1975). *Het westerse tijdssyndroom; conceptuele integratie en eerste aanzet tot constructvalidatie van een reeks molaire tijdsvariabelen in de psychologie*. Nijmegen: Stichting Studentenpers Nijmegen.

Manuscript ontvangen 25-1-1994

Definitieve versie ontvangen 25-5-1994

Handelingscontrolevaardigheden in het onderwijs

Roelie Otten, Monique Boekaerts en Gerard Seegers
*Vakgroep Onderwijsstudies, Universiteit Leiden**

ABSTRACT

Self-regulatory behaviour does not only refer to metacognitive skills to steer and guide the learning process, but also to behavioural control mechanisms. Recently, various authors have pointed out that metamotivational skills are an important aspect of independent learning. In this study metamotivational skills are operationalized in terms of Kuhl's (1984) action control theory (i.e., the abilities to take initiative, to persist, and to disengage in relation to everyday activities). Differences in action control in 7th grade students from different educational streams were investigated. The relationships were examined between action control, on the one hand, and the appraisals and effort investment vis-a-vis concrete learning tasks, on the other. Results revealed that students of the highest educational stream score lower on initiative, and disengagement than expected, but that they score higher on persistence than the other students. Despite these mean differences there were no differences between the educational streams as regards the effect of the three aspects of action control on the appraisal of learning situations, on intended effort, and on reported effort, which suggests that the same mechanism is underlying the effect of action control on cognitions and effort expenditure.

INLEIDING

In de literatuur over instructie en leren wordt veelal een onderscheid gemaakt tussen cognitieve kennis en vaardigheden aan de ene kant en vaardigheden om het eigen leerproces te sturen aan de andere kant. Bij deze laatste categorie van vaardigheden gaat het er om dat leerlingen in staat zijn het eigen leren op een zodanige wijze te reguleren dat cognitieve kennis en vaardigheden op een adequate wijze in het leerproces worden ingezet. Deze sturingsactiviteiten werden in het verleden veelal aangeduid met de term meta-cognitieve vaardigheden. Recentelijk wordt meer en meer de term (zelf)regulatievaardigheden gebruikt om het onderscheid met de externe sturing van het leerproces, door bijvoorbeeld de leerkracht, te benadrukken (Simons & Vermunt, 1986).

Onderzoek naar zelfregulatie heeft zich tot nu toe vooral gericht op meta-cognitieve regulatievaardigheden. Activiteiten die de informatiestroom in het leerproces sturen, zoals plannen, bewaken, controleren en reflecteren, stonden centraal. Succesvolle leerlingen blijken deze meta-cognitieve regulatieactiviteiten vaker te hanteren dan minder succesvolle leerlingen. Zo is bijvoorbeeld gevonden dat LBO/MAVO-leerlingen niet alleen achterblijven bij leerlingen van andere schooltypen wat betreft cognitieve kennis en vaardigheden, maar dat zij ook minder goed in staat zijn om het eigen leerproces te reguleren (De Jong, 1992).

Recentelijk hebben verschillende auteurs er op gewezen (Boekaerts, 1992; Kuhl & Kraska, 1989) dat zelfregulatie op het motivationele vlak eveneens een belangrijke rol speelt bij het zelfstandig uitvoeren van leeractiviteiten. Deze auteurs benadrukken dat zelfregulerend leren niet alleen betekent dat leerlingen over vaardigheden beschikken om het leren in engere zin te reguleren, maar ook dat zij hun inzet of motivatie om te leren weten te mobiliseren en te reguleren. Naar deze motivationele zelfregulatievaardigheden is nog weinig onderzoek verricht.

Handelingscontrolevaardigheden

Zelfregulatie op het motivationele vlak bestaat uit twee componenten: (1) zichzelf kunnen 'motiveren' en (2) beschikken over een intern gedragscontrolesysteem om voorgenomen gedrag

* Adres: Postbus 9555, 2300 RB Leiden.

ook ten uitvoer te brengen (Kuhl & Gotschke, in press). Het eerste aspect is het onderwerp waarmee motivatie-onderzoek zich vanouds bezighoudt. Belangrijk is dat de bereidheid van leerlingen om zich voor schoolse taken in te zetten niet alleen wordt beïnvloed door traditionele motivatie-aspecten, zoals succesverwachting en de persoonlijke waarde die een leerling aan het resultaat hecht, maar ook door stress en sociale aspecten (Boekaerts, 1988). Het tweede aspect, het interne (metamotivationale) gedragscontrolesysteem, is vooral door Kuhl (1984) in zijn handelingscontrole theorie naar voren gebracht. Ook de vaardigheden van een leerling om op een constructieve wijze om te gaan met eventuele negatieve emoties die de leersituatie oproept ('coping') en de vaardigheid van leerlingen om de sociale omgeving te interpreteren en die kennis te benutten om persoonlijke en sociale doelen te bereiken, kunnen genoemd worden (Boekaerts & Simons, 1992).

Voor wat betreft het meta-motivationale controlesysteem bevat Kuhl's werk een gedetailleerde beschrijving van de mentale processen die persoonlijke doelen of intenties koppelen aan zelfregulerende activiteiten. Volgens Kuhl gaat het bij handelingscontrole niet zozeer om de subjectieve opvattingen en verwachtingen van mensen betreffende hun persoonlijke mogelijkheden om controle uit te oefenen, zoals bijvoorbeeld in Bandura's (1986) self-efficacy concept, maar gaat het om feitelijke gedragscontrolemechanismen en hun efficiëntie. Kuhl & Eisenbeiser (1986) omschrijven handelingscontrole als de vaardigheid om een handelingsintentie in stand te houden en ten uitvoer te brengen ondanks concurrerende afleidende activiteiten. Individuen verschillen in de mate waarin ze in staat zijn om hun (leer)intenties voldoende af te schermen van concurrerende activiteiten. Goede motivationele zelfregulatie (handelingscontrole)-vaardigheden worden aangeduid met de term actie-oriëntatie. Deze oriëntatie is kenmerkend voor personen die hun gedachten richten op de activiteiten die nodig zijn om hun doel te bereiken. Geringe handelingscontrolevaardigheden worden omschreven als toestand-oriëntatie. Deze oriëntatie is kenmerkend voor personen die in zodanige mate gericht zijn op het analyseren van de situationele condities en op eigen emoties (bijv. zich zorgen maken over de taak en het voortdurend op zoek zijn naar de oorzaken van hun problemen) dat het handelen voortdurend uitgesteld wordt. Kuhl veronderstelt dat toestand-oriëntatie interfereert met optimaal cognitief functioneren omdat concurrerende motivationele tendensen niet op een afstand worden gehouden. Dit is in overeenstemming met de aandachtshypothese dat slechte prestaties van personen met testangst veroorzaakt worden door taakirrelevante gedachten die de aandacht van de taak afleiden (Sarason, 1975; Wine, 1982).

Kuhl onderscheidt drie aspecten van handelingscontrole. Hij stelt dat het voor het succesvol volbrengen van taken nodig is dat leerlingen initiatief ontplooiën om een gedragsintentie aan een handelingsplan te koppelen, persistentie tonen in de uitvoering van het plan, en, wanneer nodig, de intentie van het handelingsplan weten te ontkoppelen. *Initiatief* betekent dat selectief de aandacht wordt gericht op stimuli die relevant zijn voor het versterken van een bepaalde intentie. Hierdoor treedt in de planningsfase van een activiteit geen onnodig tijdverlies op. *Persistentie* heeft betrekking op de uitvoeringsfase en komt tot uiting in het doorzettingsvermogen tijdens de taakuitvoering. *Ontkoppeling* verwijst naar de vaardigheid zich los te maken van doelen die in de huidige situatie niet haalbaar zijn. Dit is van belang in de evaluatiefase en uit zich bijvoorbeeld in het op een constructieve manier omgaan met frustratie, falen, en het behaalde resultaat. Op grond van de theorie wordt verondersteld dat individuen die laag scoren op initiatief, persistentie en/of ontkoppeling voorgenomen handelingen minder snel of minder vaak ten uitvoer brengen dan individuen die hoog scoren op handelingscontrolevaardigheden.

Doel van het onderzoek

Verwacht wordt dat tekorten in dergelijke zelfregulatievaardigheden vooral naar voren zullen komen bij de overgang van het ene naar het andere schooltype. De reden is dat in de nieuwe situatie 'oude' studietechnieken en -strategieën vaak niet meer voldoen. Leertaken worden complexer, moeten in toenemende mate zelfstandig door leerlingen worden uitgevoerd, en leerlingen krijgen in het onderwijsleerproces meer verantwoordelijkheden. Dit geldt zeker voor de overgang van het basisonderwijs naar het voortgezet onderwijs. In het basisonderwijs be-

perkt zelfstandig werken zich tot het individueel uitvoeren van opdrachten van verschillende aard. In het voortgezet onderwijs wordt een organisatorische zelfstandigheid gevraagd in de zin van: grotere stukken leerstof overzien, huiswerk plannen, maatregelen nemen voor langere perioden, alsmede de vaardigheid om gedurende langere tijd zelfstandig te kunnen werken aan een verscheidenheid aan opdrachten (Gieles, 1985). Dit maakt dat met name de brugklas een periode is waarin zelfregulatievaardigheden verder ontwikkeld moeten worden.

De volgende onderzoeksvraag staat in deze bijdrage centraal: In welke mate beschikken leerlingen van verschillende typen brugklassen over handelingscontrolevaardigheden ten aanzien van alledaagse activiteiten? Concreter gesteld, in welke mate beschikken zij over initiatief, persistentie en de mogelijkheid om zonodig het voorgenomen handelingsplan van een intentie te ontkoppelen? Zijn er verschillen tussen de klastypen in de wijze waarop handelingscontrolevaardigheden, naast traditionele motivatievariabelen (prestatie-motivatie) en vakspecifieke competentie, de beleving van leertaken alsmede de voorgenomen- en daadwerkelijke inzet voor concrete taken beïnvloeden? Wanneer gesproken wordt over de "beleving van leertaken" wordt verwezen naar drie aspecten die in de motivatieliteratuur veelvuldig als belangrijke predictoren voor voorgenomen inzet worden aangemerkt: het zelfvertrouwen van de leerling voorafgaand aan het maken van de taak, de waargenomen attractiviteit van de taak en het belang dat de leerling aan de taak hecht. In de literatuur wordt aan het zelfvertrouwen van de leerling gerefereerd met diverse begrippen, bijv. Atkinson's (1957) succesverwachting, Bandura's (1986) doelmatigheidsbeleving (self-efficacy), Marsh and Shavelson's (Marsh, 1990; Shavelson & Marsh, 1986) beeld van eigen bekwaamheid en White's (1959) competentieconcept. De motiveerende invloed van de attractiviteit van de leertaak is in de intrinsieke motivatie traditie veelvuldig aangetoond (Deci, 1975; Harter, 1974). Het belang dat een leerling aan de leertaak hecht is recentelijk in de doeloriëntatie theorie (Ames & Archer, 1988; Dweck, 1986; Nicholls, 1984) opnieuw als een belangrijke predictor voor gedrag naar voren geschoven. Aangenomen wordt dat persoonlijke doelen die leerlingen nastreven doorslaggevend zijn voor de bereidheid om zich voor taken in te zetten. Boekaerts ontwikkelde de On-line Motivatie Vragenlijst om deze begrippen situatie-specifiek te meten. Deze vragenlijst is inmiddels in verschillende onderzoeken gebruikt en is op psychometrische kwaliteiten onderzocht (Boekaerts, 1992; Boekaerts, 1994; Otten & Boekaerts, 1992; Seegers & Boekaerts, 1993).

METHODE

Steekproef

In het onderzoek participeerden 821 brugklassers, 399 jongens en 422 meisjes (leerjaar 1) van acht scholen uit voornamelijk de Randstad. De groep van acht scholen werd gevormd door zes

Tabel 1. Verdeling van leerlingen over typen brugklassen van de verschillende scholen (N=821).

SCHOOL	LBO/MAVO	MAVO/HAVO	HETEROGEEN	HAVO/VWO	VWO
1	79	50		16	
2		94			
3	26	55	25		
4			58		
5					28
6			164		
7			169		
8					57
Tot	105	199	391	41	85

scholengemeenschappen en twee Lycea. De volgende vijf klastypen zijn vertegenwoordigd: LBO/MAVO, MAVO/HAVO, heterofoon, HAVO/VWO en VWO klassen. In Tabel 1 wordt de samenstelling van de steekproef weergegeven.

Instrumenten

Kuhl's handelingscontrole vragenlijst voor volwassenen (de HAKEMP90) werd vertaald en bewerkt. Het belangrijkste verschil met de oorspronkelijke vragenlijst is dat in de *Leidse Handelings Controle Schaal* (LHCS) in plaats van dichotome items voor vierpuntsschalen werd gekozen. De LHCS bevat evenals de HAKEMP90 36 items, verdeeld in drie schalen met elk 12 items. Bevraagd wordt actie- versus toestand-oriëntatie in de planningsfase (initiatief), tijdens de taakuitvoering (persistentie), en na confrontatie met falen (ontkoppeling). Elk item geeft een schets van een alledaagse situatie, zoals t.v. kijken, boodschappen doen, gevolgd door een vierpuntsschaal met aan beide uiteinden een antwoord-alternatief, één verwijzend naar actie-oriëntatie het andere naar toestand-oriëntatie. De schaalcores worden verkregen door aan de vier antwoorden waarden van 1 tot 4 toe te kennen, waarbij een hogere score een toenemende mate van actie-oriëntatie uitdrukt. De betrouwbaarheid (interne consistentie) van de verschillende schalen is bepaald door Cronbach's alpha te berekenen. Deze bedraagt .76 voor *initiatief*, .67 voor *persistentie* en .73 voor *ontkoppeling* (zie Otten & Boekaerts, 1992). Een hoge score op initiatief betekent dat leerlingen zonder tijdverlies aan een activiteit beginnen. Een hoge score op persistentie betekent dat leerlingen een activiteit zonder onderbrekingen uitvoeren, en een hoge score op ontkoppeling houdt in dat leerlingen afstand kunnen nemen van niet-bereikte doelen. Ter illustratie, wordt van elke subschaal één voorbeelditem gegeven.

Initiatief

- | | | | |
|--|--|---------|--|
| 2. Als ik weet dat iets snel af moet zijn: | moet ik me er vaak toe brengen om er aan te beginnen | O O O O | vind ik het makkelijk om het snel af te werken |
|--|--|---------|--|

Persistentie

- | | | | |
|--|--|---------|------------------------------------|
| 3. Als ik een leuk nieuw spel heb geleerd: | heb ik daar ook weer snel genoeg van en doe ik graag iets anders | O O O O | blijf ik lang in het spel verdiept |
|--|--|---------|------------------------------------|

Ontkoppeling

- | | | | |
|--|--|---------|---|
| 1. Als ik iets waardevols kwijt ben en het nergens kan vinden: | kan ik me moeilijk op iets anders concentreren | O O O O | zet ik het na korte tijd uit mijn hoofd |
|--|--|---------|---|

Prestatiemotivatie werd gemeten met behulp van Gjesme and Nygard's *Achievement Motive Scale* (AMS) in een vertaling van Van Hal (1980). De AMS bestaat uit een prestatie-motief- en een faalangstmotiefschaal met elk 15 uitspraken (vierpuntsschaal). Een voorbeeld van een prestatie-motiefitem is: 'Ik doe graag dingen waarvan ik van tevoren niet helemaal zeker weet dat ik ze kan'. Een voorbeeld van een faalangstmotiefitem is: 'Ik houd er niet van als ik moet laten zien wat ik kan'. Cronbach's alpha als maat voor interne consistentie is respectievelijk .84 voor de prestatie-motiefschaal en .87 voor de faalangstmotiefschaal.

De Situatiebeleving bij concrete leertaken alsmede de voorgenomen- en daadwerkelijke inzet voor deze taken werd gemeten met de 1991 versie van Boekaerts' *On-line Motivatie Vragenlijst* (OMV). Deze vragenlijst bestaat uit twee gedeelten. Het eerste deel bevraagt de gedachten en gevoelens van leerlingen voorafgaand aan de taak. Het tweede deel brengt de gedachten en gevoelens na afloop van de taak in kaart. In dit onderzoek werden de aspecten *waargenomen attractiviteit van de taak*, *zelfvertrouwen voorafgaand aan de taak*, *het waargenomen belang van de taak*, *de intentionele inzet voorafgaand aan de taak*, en *de gerapporteerde*

inzet na afloop van de taak, in de analyses betrokken. Betrouwbaarheid (Cronbach's alpha) van de subschalen is in de verschillende onderzoeken tenminste rond de .70 of hoger (zie Crombach, Boekaerts & Voeten, in voorbereiding; Otten & Boekaerts, 1992; Seegers & Boekaerts, 1993).

De OMV wordt bij concrete leertaken ingevuld. In dit onderzoek betrof dit taken voor vier verschillende vakken, namelijk Frans, geschiedenis, Nederlands en wiskunde. Voor Frans kregen leerlingen opdrachten waarin zowel werkwoordvervoegingen als communicatieve aspecten waren opgenomen. De geschiedenisopdrachten hadden de Romeinen en de ontwikkeling van het schrift als onderwerp. De leerlingen kregen een aantal opdrachten van verschillende aard (zoals de informatie uit een tabel in eigen woorden weergeven en er conclusies uit trekken, een historische kaart lezen, een tijdbalk maken, en een historische tekst in eigen woorden weergeven). De beide taken voor het vak Nederlands waren opdrachten tekstverklaren met meerkeuzevragen. De wiskundetaken tenslotte bestonden voor een deel uit herhaling van basisonderwerpen (namelijk procenten, verhoudingen, breuken en redactieopgaven), en voor een deel uit brugklasonderwerpen (namelijk wiskundige lichamen benoemen, werken met draadmodellen, grafieken tekenen, rekenen met negatieve getallen en vergelijkingen met één onbekende oplossen). Tot slot werden de rapportcijfers van de leerlingen verzameld als maat voor de competentie voor specifieke schoolvakken.

Procedure

Er werd op twee tijdstippen gemeten. De eerste afnames vonden plaats in december '91 en januari '92, de tweede serie afnames in mei/juni '92. Elke school deed op beide meetmomenten voor (dezelfde) twee schoolvakken aan het onderzoek mee. De scholen werden gelijkelijk verdeeld over de vakkencombinaties Geschiedenis/Frans, Frans/Nederlands, Nederlands/Wiskunde en Wiskunde/Geschiedenis. Deze procedure was bedoeld om overbelasting van de leerlingen en leerkrachten te voorkomen. De werkwijze tijdens beide afnames was gelijk. In het eerste lesuur werden de Prestatiemotivatie Vragenlijst en de Handelingscontrole Vragenlijst afgenomen. Met minstens een week tussenperiode werd vervolgens in twee aparte lessen (één lesuur per vak) de taak gemaakt en de OMV ingevuld. Voor het maken van de taak was ongeveer een half uur beschikbaar.

RESULTATEN

De steekproef bestaat uit vijf verschillende klastypen, geordend van laag naar hoog: LBO/MAVO, MAVO/HAVO, MAVO/HAVO/VWO (heterogeen), HAVO/VWO, en alleen VWO. Het heterogene klastype is wat betreft het aantal leerlingen in de steekproef oververtegenwoordigd en de HAVO/VWO klassen zijn sterk ondervertegenwoordigd. Nadat de resultaten van t-tests hadden uitgewezen dat er geen significante verschillen tussen HAVO/VWO- en VWO-klassen waren wat betreft de drie aspecten van handelingscontrole werd besloten om voor verdere analyses deze beide klastypen samen te voegen.

Verschillen tussen klastypen in handelingscontrolevaardigheden

Om de verschillen in handelingscontrolevaardigheden tussen de vier klastypen te toetsen werd een MANOVA met herhaalde metingen op de twee tijdstippen uitgevoerd. In deze analyse is klastype (4 niveaus) de onafhankelijke variabele, terwijl de drie vormen van handelingscontrole genest onder de factor tijd (2 meetmomenten) de afhankelijke variabelen zijn. Van de 821 leerlingen werden 184 leerlingen vanwege ontbrekende gegevens niet in de uiteindelijk analyses betrokken. De gemiddelde scores, standaarddeviaties en het aantal leerlingen voor de verschillende klastypen op beide meetmomenten staan weergegeven in Tabel 2.

De variantie analyse resulteerde in een significant klastype-effect ($F(3,633)=2.89, p<.05$) een significant tijdseffect ($F(1,633)=120.89, p<.001$); het interactie-effect voor klastype en tijd was niet significant ($F(3,633)=2.60, p>.05$). Wel significant zijn alle interacties met handelingscontrole, te weten het interactie-effect "tijd * handelingscontrole" ($F(2,1266)=109.43, p<.001$)

Tabel 2. Gemiddelde scores en standaarddeviaties voor de aspecten van handelingscontrole per klastype op de twee meetmomenten.

	December						Mei					
	Initiatief		Persistentie		Ontkoppeling		Initiatief		Persistentie		Ontkoppeling	
	M	sd	M	sd	M	sd	M	sd	M	sd	M	sd
LBO/MAVO (N=70)	33.78	6.5	34.88	5.7	29.76	6.1	33.73	5.8	34.28	6.7	29.52	6.3
MAVO/HAVO (N=148)	33.54	6.3	35.36	5.6	29.08	6.4	33.24	6.1	33.89	5.6	30.39	5.8
HETEROGEEN (N=312)	33.71	6.2	35.77	5.6	30.52	6.2	33.83	6.6	35.81	5.6	31.49	6.8
HAVO/VWO (N=107)	32.22	6.8	37.18	4.9	29.21	6.2	31.75	7.7	37.13	5.3	29.36	6.4
Totaal (N=637)	33.43	6.4	35.82	5.6	29.88	6.2	33.33	6.6	35.42	5.8	30.66	6.5

en het interactie-effect "klastype * tijd * handelingscontrole" ($F(6,1266)=5.36, p < .001$). Met andere woorden, de klastypen verschillen in handelingscontrolevaardigheden t.a.v. alledaagse activiteiten en de handelingscontrolevaardigheden veranderen gedurende de tijdspanne van vijf maanden, waarbij de grootte van de verschillen varieert over de drie aspecten van handelingscontrole. In Tabel 3 zijn de effecten voor klastype voor de verschillende aspecten van handelingscontrole op de twee meetmomenten weergegeven.

Bekijken we het klastype effect voor elk van de handelingscontrolevariabelen op de twee meetmomenten, dan zien we dat er op het eerste meetmoment alleen een significant klastype effect is voor persistentie, terwijl op het tweede meetmoment het klastype effect voor alle drie de aspecten van handelingscontrole significant is. De gevonden klastype-effecten zijn vooral toe te schrijven aan verschillen tussen de HAVO/VWO klassen en de andere klastypen (zie Tabel 2). HAVO/VWO klassen scoren duidelijk hoger op persistentie, terwijl zij op initiatief tijdens het tweede meetmoment een duidelijk lager gemiddelde hebben dan alle andere klastypen. Bovendien scoren de HAVO/VWO klassen tijdens het tweede meetmoment (tezamen met de LBO/MAVO klassen) laag op ontkoppeling.

Het interactie-effect klastype * tijd * handelingscontrole werd nader onderzocht door de verschillende klastypen per meetmoment en voor de verschillende aspecten van handelingscontrole te vergelijken (Scheffé tests). Op het eerste meetmoment (in december) geeft de Scheffé-test (getoetst op .05 niveau) voor geen van de drie aspecten van handelingscontrole significante verschillen tussen de klastypen. Met andere woorden, in december is er in de MANOVA wel

Tabel 3. Effecten van klastype voor de verschillende aspecten van handelingscontrole per meetmoment.

		SS	F (3,633)	sig
December	Initiatief	191.19	1.57	n.s.
	Persistentie	291.93	3.19	.05
	Ontkoppeling	268.98	2.33	n.s.
Mei	Initiatief	357.82	2.75	.05
	Persistentie	795.60	8.21	.001
	Ontkoppeling	496.32	3.96	.01

een significant klastype effect voor persistentie, maar bij paarsgewijze vergelijking verschillen de klastypen niet statistisch significant van elkaar. Op het tweede meetmoment is wel een aantal verschillen significant. Wat betreft de handelingscontrolevaardigheden ten aanzien van alledaagse activiteiten scoren de HAVO/VWO-klassen significant lager op initiatief dan de heterogene klassen. De heterogene klassen scoren significant hoger op persistentie dan de MAVO/HAVO klassen, terwijl de HAVO/VWO klassen significant hoger scoren dan de LBO/MAVO en de MAVO/HAVO klassen. Tot slot scoren de HAVO/VWO klassen wat betreft de handelingscontrolevaardigheden in de evaluatiefase (ontkoppeling) significant lager dan de heterogene klassen.

De resultaten kunnen als volgt worden geïnterpreteerd. Persistentie bij alledaagse activiteiten hangt positief samen met schoolcarrière-aspecten: meer specifiek met het brugklasniveau dat de leerlingen bij het begin van het voortgezet onderwijs volgen. Alle gevonden verschillen wijzen in de richting van een toename van de persistentiescore met dit niveau. Voor initiatief en ontkoppeling waren de uitkomsten minder eenduidig. Bij deze beide aspecten van handelingscontrolevaardigheden hebben de HAVO/VWO klassen een onverwacht lage score en werd het theoretisch veronderstelde positieve verband met schoolsucces niet gevonden.

Verschillen tussen klastypen in de invloed van handelingscontrole op de beleving van leertaken

Verbanden tussen de variabelen voor de totale groep werden met behulp van pad-analyse onderzocht. Gebruik werd gemaakt van het LISREL VII-programma (Analysis of Linear Structural Relationships by Maximum Likelihood and Least Squares Methods, Jöreskog & Sörbom, 1990). Meer specifiek werd onderzocht of er veranderingen in het model noodzakelijk zijn om tegemoet te komen aan eventuele verschillen tussen de klastypen. Het LISREL programma maakt het mogelijk de hypothese te toetsen dat eenzelfde model adequaat is om de data-structuur van meerdere groepen te verklaren (multi-sample analysis). Van deze mogelijkheid wordt in de analyses gebruik gemaakt.

Effecten op leertaken

Eerst is nagegaan of de klastypen van elkaar verschillen in de wijze waarop handelingscontrolevaardigheden, naast prestatie-motivatie en de competentie voor specifieke schoolvakken (gerefleeteerd in rapportcijfers), de beleving en inzet bij leertaken beïnvloeden. Het model van het affectieve leerproces beschreven in Boekaerts (1992), werd door Seegers en Boekaerts (1993) vertaald in een structureel model en getoetst. Dit model diende als uitgangspunt voor deze studie. Het werd uitgebreid met de drie aspecten van handelingscontrole. Prestatie-motivatie (Prestatie-motief en Faalangstmotief) en de vakspecifieke competentie (gerepresenteerd in het rapportcijfer voor een vak) worden verondersteld de situatiebeleving, dat wil zeggen het zelfvertrouwen, de waargenomen attractiviteit van de taak en het waargenomen belang van de taak te beïnvloeden. Zowel de objectieve competentie als het prestatie- en faalangstmotief worden verondersteld het vertrouwen dat leerlingen hebben in hun eigen competentie om aan de taakeisen te kunnen voldoen te beïnvloeden. Competentie en prestatie-motief oefenen een positieve invloed uit op het zelfvertrouwen, terwijl faalangst deze in negatieve zin beïnvloedt. Het prestatie-motief heeft tevens een positieve invloed op het belang dat leerlingen aan leertaken hechten. Leerlingen met een hoge score op het prestatie-motief vinden "presteren" belangrijk. Dit maakt dat zij meer belang zullen hechten aan toetsituaties dan leerlingen met een minder uitgesproken prestatie-motief. Een direct verband tussen het prestatie-motief en de waargenomen attractiviteit van de taak wordt, gezien het veelvuldig gemaakte onderscheid in extern- en intern gemotiveerd gedrag, niet verondersteld. Daarentegen wordt een indirect effect van het prestatie-motief en het faalangstmotief op de attractiviteit verondersteld, via het beïnvloeden van het zelfvertrouwen van de leerlingen. Het zijn de situatiespecifieke belevingen die vervolgens de intentionele inzet en via deze intentionele inzet de daadwerkelijke geleverde inzet beïnvloeden. Kuhl stelt dat handelingscontrole zich richt op de feitelijke gedragscontrolemechanismen en hun efficiëntie, dat wil zeggen op het omzetten van een intentie in daadwerkelijk gedrag. Voor handelingscontrolevaardigheden (Initiatief, Persistentie en Ontkoppeling) wordt dan ook verondersteld dat zij

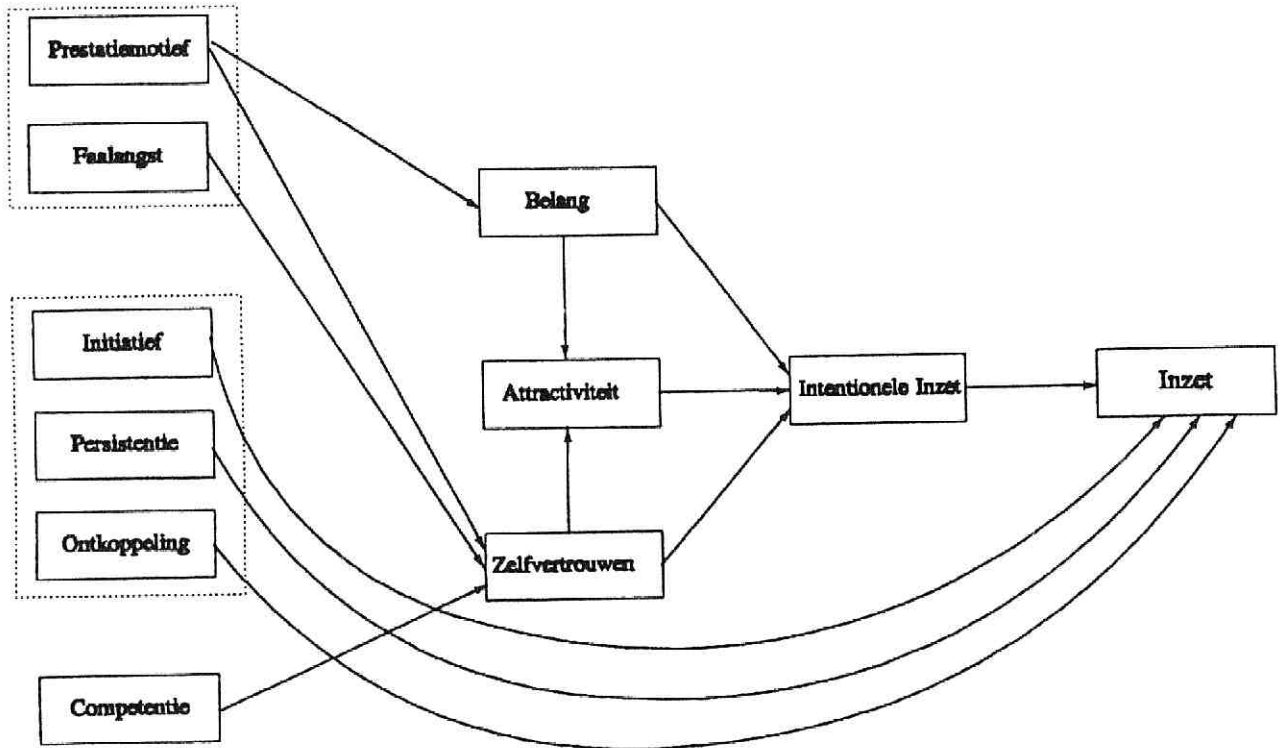


Fig. 1. Theoretisch uitgangsmodel.

dan ook verondersteld dat zij direct van invloed zijn op de daadwerkelijke inzet (Geleverde Inzet) die leerlingen bereid zijn voor een specifieke leertaak te leveren.

Het geformuleerde model zoals weergegeven in Figuur 1 werd getoetst voor het vak Frans. Wanneer we het criterium hanteren dat de Chi-kwadraat waarde niet hoger mag zijn dan twee tot drie keer het aantal vrijheidsgraden, geldt dat het model ($\chi^2=105.44$, $df=27$, $GFI=.95$, $AGFI=.89$, $RMSR=.08$) onvoldoende bij de data past. Inspectie van de modificatie indices leidde tot aanpassingen van het model. Een extra invloed werd aangenomen van zelfvertrouwen op het belang van de taak, en een directe invloed van belang op de geleverde inzet. Geen van de aspecten van handelingscontrole bleek direct van invloed te zijn op de geleverde inzet. Initiatief is van invloed op zelfvertrouwen; persistentie is van invloed op attractiviteit van de taak en op intentionele inzet. Ontkoppeling heeft geen significante relaties met één van de andere variabelen in het model.

Het aangepaste model heeft bij het vak Frans een Chi-kwadraat waarde die iets beneden twee keer het aantal vrijheidsgraden ligt. Dit geldt als een redelijke (globale) fit. Inspectie van het model op een meer gedetailleerd niveau gaf geen aanwijzingen tot serieuze misspecificaties.

Tabel 4. Goodness of fit indices LISREL analyses.

	χ^2	df	GFI	AGFI	RMS
Frans (N=392)	48.70	25	.978	.943	.046
Geschiedenis (N=381)	47.61	25	.978	.943	.044
Nederlands (N=346)	35.36	25	.982	.952	.041
Wiskunde (N=345)	71.86 (46.34)	25 (23)	.963 (.977)	.903 (.933)	.057 (.047)

voor wiskunde staan de resultaten van het aangepaste model tussen haakjes.

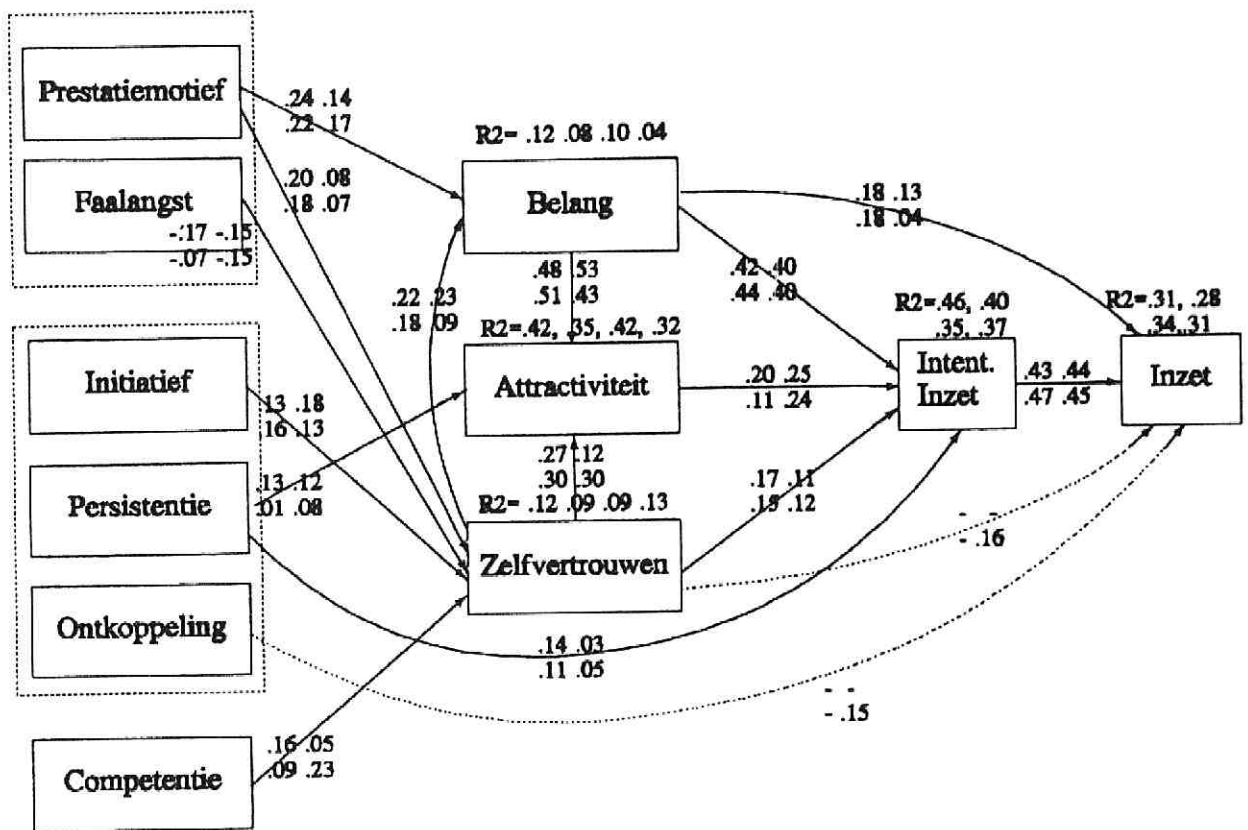


Fig. 2. Parameterschattingen en verklaarde variantie voor de vier vakken.

Ditzelfde model leverde bij de vakken geschiedenis en Nederlands eveneens een goede fit op. Voor wiskunde moesten twee extra relaties worden opgenomen om een goede modelfit te krijgen: zelfvertrouwen en ontkoppeling zijn van invloed op de geleverde inzet. In Tabel 4 wordt een overzicht gegeven van de (globale) fitmaten bij toetsing van het model voor de vier verschillende vakken. In Figuur 2 wordt het model gepresenteerd met de parameter-schattingen voor de vier vakken. Hierbij is uitgegaan van de gestandaardiseerde oplossing. Bovendien zijn de percentages verklaarde variantie van de afhankelijke variabelen in de figuur weergegeven.

Het percentage verklaarde variantie voor zelfvertrouwen ligt tussen de 9% voor geschiedenis en Nederlands en 13% voor wiskunde. Prestatiemotief, faalangstmotief, initiatief en competentie zijn hierbij de verklarende variabelen. De sterkte van de invloed van deze vier variabelen is globaal genomen vergelijkbaar, maar er zijn wel verschillen te zien bij de vier schoolvakken. Het percentage verklaarde variantie voor belang van de taak ligt tussen de 4% voor wiskunde en 12% voor Frans. Prestatiemotief en zelfvertrouwen zijn de verklarende variabelen. De percentages verklaarde variantie voor attractiviteit, intentionele inzet en geleverde inzet, zijn beduidend hoger. Zelfvertrouwen, belang en persistentie verklaren gezamenlijk tussen de 32% voor wiskunde en 42% voor Frans en Nederlands van attractiviteit, waarbij belang de belangrijkste invloed uitoefent. De door het model verklaarde variantie voor intentionele inzet ligt tussen de 35% en 46%. Intentionele inzet wordt beïnvloed door de drie situatie-specifieke appraisals, zelfvertrouwen, attractiviteit en belang, en het handelingscontrole aspect persistentie. Uit de sterkte van de verbanden is af te lezen dat het belang van de taak de belangrijkste verklarende variabele is. Het percentage verklaarde variantie voor geleverde inzet is ongeveer 30%. Geleverde inzet wordt bij Frans, geschiedenis en Nederlands bepaald door intentionele inzet en belang. Bij het vak wiskunde beïnvloeden ook zelfvertrouwen en ontkoppeling de geleverde inzet. Wel geldt ook bij dit laatste vak dat intentionele inzet en belang van de taak de belangrijkste verklarende variabelen zijn.

Samenvattend kan gesteld worden dat prestatie-motivatie en vakspecifieke competentie zoals weergegeven in rapportcijfers vooral van invloed zijn op het zelfvertrouwen dat leerlingen hebben voor zij met een taak beginnen. Gezien de lage percentages verklaarde variantie moeten we aannemen dat zelfvertrouwen en het belang dat leerlingen aan een taak hechten voor een belangrijk deel met andere dan in het model opgenomen variabelen samenhangen. Handelingscontrolevaardigheden beïnvloeden de situatiebeleving (initiatief en persistentie) en de intentionele inzet (persistentie). Alleen bij het vak wiskunde is handelingscontrole (ontkoppeling) van invloed op de geleverde inzet. Dit suggereert dat leerlingen die geen tijd verliezen in de plan-ningsfase van alledaagse activiteiten meer zelfvertrouwen hebben bij het uitvoeren van concrete leertaken. Leerlingen die een hoge score op persistentie hebben bij alledaagse activiteiten vinden de leertaken attractiever en scoren hoger op intentionele inzet. Leerlingen die op een constructieve manier met falen in alledaagse situaties weten om te gaan (ontkoppeling) hebben bij wiskunde een hogere gerapporteerde inzet.

De afhankelijke variabelen intentionele en geleverde inzet worden in belangrijke mate (30% tot 40%) door het model verklaard. Vooral de situatiebelevingsaspecten zijn van invloed op de intentionele inzet, waarbij belang van de taak de belangrijkste verklarende variabele is. Intentionele inzet op zijn beurt is de belangrijkste verklarende variabele voor geleverde inzet. Dit betekent dat wat leerlingen zich voornemen te gaan doen een belangrijke voorspeller is voor wat ze daadwerkelijk aan inzet leveren.

Verschillen tussen klastypen

Vervolgens is nagegaan of het hierboven beschreven model geldig is voor leerlingen van de verschillende klastypen. Omdat bij het vak geschiedenis het merendeel van de leerlingen tot het heterogene klastype behoort is toetsing bij dit vak niet zinvol. Getoetst is in hoeverre eenzelfde model adequaat is om de onderliggende structuur van de data van verschillende klastypen te verklaren. Vergeleken zijn modellen die verschillen in restricties. Het programma geeft een Chi-kwadraat waarde die aangeeft in hoeverre het veronderstelde model de structuur van de data van verschillende groepen adequaat verklaart. De meest restrictieve aanname is dat een identiek model adequaat is om de onderliggende structuur van de data van alle klastypen te verklaren. Een minder restrictief model neemt eenzelfde onderliggende structuur voor de verschillende klastypen aan, maar de sterkte van de verbanden kan verschillen. Het verschil in chi-kwadraat waarde van de vergeleken modellen heeft bij benadering opnieuw een chi-kwadraat verdeling met als vrijheidsgraden het verschil in vrijheidsgraden van de vergeleken modellen. Een statistisch significant effect betekent dat een minder restrictief model moet worden geaccepteerd. In Tabel 5 zijn de fit maten voor de vergeleken modellen weergegeven.

Uit de resultaten in Tabel 5 blijkt dat in geen enkel geval een minder restrictief model een significante verbetering van de model fit oplevert. Dit houdt in dat er geen aanwijzingen zijn om aan te nemen dat de sterkte van de samenhang tussen de variabelen in het uitgangsmodel voor

Tabel 5. Fit maten van de vergeleken modellen voor specifieke klastypen.

Vak	χ^2	df	GFI	RMS
Frans	141.98	115	.971	.051
	103.33	75	.974	.049
Nederlands	156.71	115	.918	.083
	99.93	75	.956	.056
Wiskunde	135.33	113	.840	.012
	93.23	69	.908	.074

leerlingen uit de verschillende klastypen verschilt. Met andere woorden het onderliggend mechanisme dat de invloed van handelingscontrolevaardigheden op situatiespecifieke motivatie en inzet verklaart, is voor alle brugklassen hetzelfde. Wel werden er verschillen in gemiddelden voor een aantal variabelen gevonden.

DISCUSSIE

De verschillen in de handelingscontrolevaardigheden (initiatief, persistentie en ontkoppeling) tussen leerlingen van verschillende klastypen zijn gering. Aan het begin van het eerste brugklasjaar (december) zijn er geen significante verschillen tussen de klastypen, aan het eind van het leerjaar zijn het vooral de HAVO/VWO klassen die in het oog springen. HAVO/VWO leerlingen scoren theoretisch gezien onverwacht laag op initiatief en ontkoppeling. Dat wil zeggen dat ze meer tijd dan verwacht nodig hebben bij het beginnen van dagelijkse activiteiten en dat ze meer moeite hebben om niet bereikte doelen los te laten. Kuhl interpreteerde een lage score op initiatief als 'hesitation' (aarzelen) en een lage score op ontkoppeling als 'preoccupation' (doormalen). Het zou echter ook kunnen betekenen dat leerlingen die meer tijd besteden aan het keuzeprocess bij het begin van een activiteit zich beter oriënteren voordat ze aan een activiteit of taak beginnen en ook langer reflecteren over het resultaat na afloop van een activiteit of taak. HAVO/VWO leerlingen laten zich het minst afleiden wanneer ze eenmaal met een taak bezig zijn. Globaal genomen geldt, dat hoe hoger het niveau van het klastype is des te hoger de score voor persistentie. Deze resultaten lijken er op te wijzen dat persistentie bij alledaagse activiteiten op een meer eenduidige wijze samengaat met succes in het onderwijs, afgemeten aan het niveau van het klastype in het begin van het voortgezet onderwijs, dan de beide andere aspecten van handelingscontrole. Verder onderzoek zal moeten uitwijzen of deze resultaten in volgende leerjaren worden bevestigd. Het feit dat er aan het begin van de brugklas nog geen significante verschillen tussen de klastypen zijn maar aan het eind van het leerjaar wel, wijst in de richting van divergerende handelingscontrolevaardigheden bij leerlingen tussen de 12 en 13 jaar.

Dat er geen verschillen gevonden werden tussen de klastypen wat betreft de invloed van handelingscontrolevaardigheden op de situatiebeleving en op de inzet voor concrete leertaken kan worden geïnterpreteerd als een aanwijzing voor eenzelfde onderliggend mechanisme. Voor alle leerlingen geldt namelijk dat het kunnen nemen van initiatief in de oriëntatie fase van een handeling van invloed is op het zelfvertrouwen ten aanzien van concrete leersituaties. Persistentie bij het uitvoeren van alledaagse activiteiten heeft een positief effect op de waargenomen attractiviteit en op de intentionele inzet voor leertaken. Alleen voor het vak wiskunde blijkt ontkoppeling van invloed te zijn op de geleverde inzet. Met andere woorden, verschillende aspecten van handelingscontrole beïnvloeden de situatiebeleving en de intentionele inzet bij concrete leertaken. De theoretisch veronderstelde invloed van handelingscontrole op de daadwerkelijk geleverde inzet werd echter alleen bij wiskunde voor ontkoppeling terug gevonden. Dit beschouwen we echter niet zondermeer als een falsificatie van de theorie van Kuhl. Het is mogelijk dat handelingscontrolevaardigheden hun effect op de omzetting van de intentionele inzet in geleverde inzet alleen uitoefenen wanneer er aanleiding toe is, namelijk wanneer leerlingen laag scoren op diverse aspecten van de situatiebeleving (c.f. Boekaerts, 1994). Ook Heckhausen, Schmalt en Schneider (1985) stelden dat de effecten van handelingscontrolevaardigheden voornamelijk waarneembaar zijn wanneer er in de situatie hindernissen zijn die het ten uitvoer brengen van een intentie bemoeilijken.

Het feit dat actie-oriëntatie of met andere woorden handelingscontrolevaardigheden van invloed zijn op de situatiebeleving en minder op de intentionele- en geleverde inzet voor leertaken maakt het daarom niet een minder belangrijk aspect voor verder motivatie-onderzoek. Immers, wanneer handelingscontrolevaardigheden, naast meer traditionele motiatiematen, de motivatie voor concrete leertaken beïnvloeden betekent dit dat ze een belangrijke variabele zijn voor verder motivatie onderzoek. Er dient nader onderzoek te worden verricht naar de houdbaarheid van Kuhl's handelingscontrole theorie in de classesituatie. In ons onderzoek bleken

handelingscontrolevaardigheden eerder van invloed te zijn op de beleving van leertaken en de voorgenomen inzet dan op de daadwerkelijk geleverde inzet.

LITERATUUR

- Ames, C. & Archer, J. (1988). Achievement goals in the classroom. Student's learning strategies and motivation processes. *Journal of Educational Research*, 80(3), 260-268.
- Atkinson's (1957). Motivational determinants of risk-taking. *Psychological Review*, 64, 359-372.
- Bandura, A. (1986). *Social foundations of thought and action: A social-cognitive theory*. Englewood Cliffs, N.J.: Prentice Hall.
- Boekaerts, M. (1987). Individual differences in the appraisal of learning tasks: An integrative view on emotion and cognition. *Communication and Cognition*, 20, 207-224.
- Boekaerts, M. (1988). Motivated learning: Bias in appraisals. *International Journal of Educational Research*, 12, 267-280.
- Boekaerts, M. (1992). The adaptable learning process: Initiating and maintaining behavioural change. *Applied Psychology: An International Review*.
- Boekaerts, M. (1994, in druk). The interface between intelligence and personality as determinants of classroom learning. In D.H. Saklofske & M. Zeidner (eds.). *Handbook of personality and intelligence*. New York: Plenum Press.
- Boekaerts, M. & Simons, P.R.J. (1992). *Instructie en leren: Psychologie van de leerling en het leerproces*. Assen: Dekker & Van de Vegt.
- Cito (1987). *Belevingsschaal voor Wiskunde: Handleiding*. Arnhem: Centraal Instituut voor Toetsontwikkeling.
- Crombach, M.J., Boekaerts, M., & Voeten, M.J.M. (in voorbereiding). On-Line Measurement of Appraisals of Students Faced with Curricular Tasks.
- Deci, E.L. (1975). *Intrinsic motivation*. New York: Plenum.
- Dweck, C.S. (1986). Motivational process affecting learning. *American Psychologist*, 41(10), 1040-1048.
- Gieles, P. (1985). *Begeleiding in de brugperiode VWO-AVO* (Deel 1. De Klassementor). Den Bosch: Katholiek Pedagogisch Centrum; Nijmegen: Berkhout.
- Hal, G. van (1980). *Een vergelijking tussen de PrestatieMotivatie Test PMT-K voor kinderen en de Achievement Motivation Scale (AMS)*. Nijmegen: Katholieke Universiteit.
- Harter, S. (1974). Pleasure derived from cognitively challenge and mastery. *Child development*, 45, 661-669.
- Heckhausen, H., Schmalt, H.D., & Schneider, K. (1985). *Achievement motivation in perspective*. Orlando: Academic Press.
- De Jong, F.P.C.M. (1992). *Zelfstandig leren. Regulatie van het leerproces en leren reguleren: een procesbenadering*. Academisch proefschrift, KUB Tilburg (pp. 219).
- Kuhl, J. (1984). Volitional aspects of achievement motivation and learned helplessness: Toward a comprehensive theory of action control. In B.A. Maher & W.B. Maher (eds.). *Progress in Experimental personality research* (Vol. 13, pp. 99-171). New York: Academic Press.
- Kuhl, J. & Eisenbeiser, T. (1986). Mediating versus meditating cognitions in human motivation action-control, inertial motivation, and the alienation effect. In J. Kuhl & J.W. Atkinson (eds.). *Motivation thought and action* (pp. 288-306). New York: Praeger.
- Kuhl, J. & Gotschke, T. (in press). Formation and maintenance of intentions in memory. In J. Kuhl & J. Beckman (eds.). *Volition and personality. Action versus state orientation*. Toronto: Hogrefe.
- Kuhl, J. & Kraska, K. (1989). Self-regulation and metamotivation: Computational mechanisms, development, and assessment. In R. Kanfer, P.L. Ackerman, & R. Cudeck (eds.). *Abilities, motivation and methodology* (pp. 343-369). Hillsdale, NJ: Erlbaum.
- Marsh, H.W. (1990). The structure of academic self-concept: The Marsh/Shavelson Model. *Journal of Educational Psychology*, 82(4), 623-636.
- Nicholls, J.G. (1984). Achievement motivation: conceptions of ability, subjective experience, task choice, and performance. *Psychological Review*, 91, 328-346.
- Otten R. & Boekaerts, M. (1992). *Metamotivatie en inzet. Motivationale oriëntatie en metamotivatie als determinanten van leerintentie, inzet en prestatie* (Eindrapport SVO-project 354). Leiden: Centre for the Study of Education and Instruction (101 p.).
- Sarason, I.G. (1975). Test anxiety, attention, and the general problem of anxiety. In C.D. Spielberger & I.G. Sarason (eds.). *Stress and Anxiety* (Vol. 1). Washington, DC: Hemisphere.

- Seegers, G. & Boekaerts, M. (1993). Task motivation and mathematics achievement in actual task situations. *Learning and Instruction, Vol. 3*, 133-150.
- Shavelson, R.J. & Marsh, H.W. (1986). On the structure of self-concept. In R. Schwarzer (Ed.). *Anxiety and cognitions*. Hillsdale, NJ: Erlbaum.
- Simons, P.R.J. & Vermunt, J.D.H.M. (1986). Self-regulation in knowledge acquisition: a selection of Dutch research. In G. Beukhof & P.R.J. Simons (Eds.). *German and Dutch research on learning and instruction: general topics and self-regulation in knowledge acquisition* (pp. 101-136). 's-Gravenhage: SVO Selecta Reeks.
- White, R.W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review, 66*, 297-333.
- Wine, J.D. (1982). Evaluation anxiety: A cognitive-attentional construct. In H.W. Krohne & L. Laux (Eds.). *Achievement, stress, and anxiety* (pp. 207-219). Washington DC: Hemisphere.

Manuscript ontvangen 25-1-1994

Definitieve versie ontvangen 25-5-1994

Stress, coping, en prestatie-(motivatie)

Monique Boekaerts en Gerard Seegers
*Vakgroep Onderwijsstudies, Universiteit Leiden**

ABSTRACT

In this study, it is investigated whether students' perceived stress and the coping strategies they use influence achievement and achievement motivation during the first years of secondary education. A moderating position was hypothesized for social support from parents, (math) teacher, and classmates. Data were collected in a longitudinal set-up with two times of measurement. A linear structural model was hypothesized to underly the structure of the data. The LISREL programme was applied for testing. It is concluded that the same model adequately explained the data structure for both measurements. However, the model was not very successful in explaining achievement and achievement-motivation as outcome variables.

INLEIDING

Er zijn verschillende definities van het begrip 'stress'. Binnen de transactionele stressbenadering wordt stress gezien als een verstoorde relatie tussen de persoon en zijn omgeving. Meer concreet wordt gesteld dat mensen stress hebben als zij denken niet meer in staat te zijn te voldoen aan de eisen die aan hen gesteld worden. Lazarus en Folkman (1984) definiëren stress als de ervaren discrepantie tussen de waargenomen omgevingseisen en de ingeschatte eigen mogelijkheden om daaraan te voldoen. De eerste stressonderzoeken richtten zich vooral op het effect van traumatische levensgebeurtenissen op het dagelijks functioneren van volwassenen. Toen echter bleek dat alledaagse ergernissen ('daily hassles') evenveel, en misschien zelfs meer, psychisch dysfunctioneren, ziekte en depressie kunnen veroorzaken dan traumatische levensgebeurtenissen, heeft het stressonderzoek een belangrijke verandering ondergaan. Momenteel wordt algemeen aanvaard dat de mate van stress voor een belangrijk deel afhangt van de manier waarop mensen hebben leren reageren op specifieke stressoren, of concreter gezegd, van de wijze waarop zij hun persoonlijke mogelijkheden om stress te reduceren hebben leren inzetten. Voor de schoolsituatie betekent dit dat leerlingen over vaardigheden moeten beschikken om op een constructieve wijze negatieve emoties (angst, irritatie, teleurstelling) die tijdens de leersituatie worden opgeroepen, te reduceren. Bovendien moeten ze over zelf-regulerende vaardigheden beschikken om met dagelijkse ergernissen om te gaan.

Stress op Nederlandse scholen

Onderzoek naar stress op Nederlandse en Vlaamse scholen richtte zich in het begin vooral op de schoolbeleving van leerlingen (De Corte, Van Buggenhout & Pyck, 1985; Compennolle, 1987; Moos, 1979; Stoel, 1980). Uit deze onderzoeken bleek dat verschillen in schoolbeleving niet toe te schrijven zijn aan materiële factoren (school-, klasgrootte, infrastructuur), maar eerder aan aspecten van de school als sociale organisatie. Uit het nationaal survey onderzoek dat het Centraal Bureau voor de Statistiek in 1979 bij ongeveer 4000 jongeren uitvoerde kwam naar voren dat onvrede van leerlingen met het onderwijs meevalt en dat de overgrote meerderheid van de leerlingen geen hekel heeft aan school. Van der Linden en Roeders (1982) brachten de leefwereld en de zelfbeleving van 540 LBO, MAVO, HAVO, en VWO-leerlingen uit het derde leerjaar in kaart. Zij concludeerden dat 21% van de leerlingen het nut van de school niet inziet, dat 17% van de leerlingen schoolmisnoegen rapporteert, en dat 17% de school als een gevangenis ervaart. Stoel (1980) van zijn kant rapporteert op grond van onderzoek bij 7000 leerlingen

* Adres: Postbus 9555, 2300 RB Leiden.

Dit is het laatste artikel behorende bij het themanummer.

uit de brugklas en het tweede en derde jaar van het voortgezet onderwijs dat de meeste leerlingen aangeven het contact met klasgenoten als het meest waardevolle aspect van het schoolgebeuren te ervaren. Maar lang niet alle leerlingen ervaren de school als een prettig sociaal gebeuren. Mooij (1992) geeft aan dat in het basisonderwijs ruim 57.000 leerlingen verschillende keren per week worden gepest. In het voortgezet onderwijs zijn ruim 18.000 leerlingen wekelijks het slachtoffer van sociale isolaties en pesterijen. Een vrij groot aantal leerlingen beleeft frequent stress op school. Welk effect dit heeft op hun functioneren op school, op hun leerprestaties, motivatie en ambitie, en uiteindelijk op hun schoolloopbaan en toekomstmogelijkheden is nauwelijks onderzocht. Om deze verbanden te kunnen onderzoeken is het nodig dat leerlingen worden bevraagd en geobserveerd in dagelijkse schoolsituaties zodat in kaart kan worden gebracht welke dagelijkse ergernissen veel voorkomen in bepaalde school- en klastypen en hoe leerlingen van diverse leeftijden en schoolniveaus met deze dagelijkse ergernissen omgaan. Bovendien moet worden onderzocht of het frequent meemaken van bepaalde stressvolle situaties, en het intens ervaren van negatieve emoties in dergelijke situaties, leiden tot het hanteren van typische wijzen van omgaan met stress, die op hun beurt de leerprestaties en de motivatie beïnvloeden.

Het is uiterst ingewikkeld om de invloed van stress en stressreacties op het leerproces te bestuderen. Ten eerste omdat stress niet objectief kan worden vastgesteld maar retrospectief moet worden bevraagd. Ten tweede omdat stress een dynamisch proces is dat fluctueert over de tijd. Iemand kan bij confrontatie met een stressvolle situatie (bijvoorbeeld gepest worden) veel stress ervaren. Wanneer hij echter op een bepaalde wijze reageert en hierdoor de pester afschrikt kan de spanning snel weer afnemen en omslaan in een gevoel van voldoening. Dit betekent in concreto dat een gehanteerde strategie zo effectief kan zijn dat leerlingen haast vergeten dat zij stress in dergelijke situaties hebben meegemaakt, en daardoor op een vragenlijst zullen aangeven weinig negatieve emoties te ervaren in peestsituaties. Daarom zal aan leerlingen een aantal stressvolle schoolsituaties worden voorgelegd waarbij zal worden onderzocht welke situaties door leerlingen van diverse leeftijden en klastypen frequent worden meegemaakt, hoe intens ze worden beleefd, en hoe ze meestal omgaan met de negatieve emoties die ze in die situaties ervaren.

Omgaan met stress

De typische wijze waarop mensen omgaan met ervaren stress wordt in de literatuur 'coping' genoemd. Het effect dat ervaren stress heeft hangt af van de mate waarin individuen menen dat ze beschikken over effectieve copingstrategieën. Er is een grote verscheidenheid aan copingstrategieën geconstateerd die leerlingen hanteren in problematische schoolsituaties. Bijvoorbeeld bij stress t.a.v. slechte cijfers past de ene leerling zich aan door zich voor te nemen harder te gaan werken, een tweede leerling reageert door hulp te gaan zoeken bij de leraar of bij medeleerlingen, terwijl een derde leerling er genoeg van heeft en helemaal niet geholpen wil worden.

Er is nog weinig bekend over de effectiviteit van diverse copingstrategieën en over de impact ervan op het leer- en motivatieproces. Hendriksen (1990) formuleerde de hypothese dat leerlingen die voornamelijk toenadering hanteren bij stressvolle situaties, beter zouden presteren op school dan leerlingen die vermijding hanteren. Deze hypothese werd echter niet bevestigd. De meeste auteurs stelden overigens vast (zie bijvoorbeeld, Seiffge-Krenke, 1990) dat, hoewel leerlingen verschillende strategieën ontwikkelen om problemen op te lossen en sommige van die strategieën door leerlingen zelf en door hun omgeving worden gezien als 'succesvol' en andere als 'minder adequaat', dit niets zegt over de effectiviteit van een bepaalde copingstrategie voor een bepaalde leerling in een bepaalde situatie. Het is belangrijk in te zien dat een specifieke copingstrategie effectief kan zijn vanuit het standpunt van herstel van het welbevinden (stressreductie), maar niet vanuit het standpunt van leerstofbeheersing (controle over leeruitkomsten). Ook is het relevant te constateren dat leerlingen op school vaak geen vluchtgedrag kunnen ontwikkelen vanwege het verplichte karakter van het onderwijs. Zij moeten accepteren dat zij bepaalde taken minder goed kunnen, of bepaalde situaties problematisch vinden. Dat kan tot gevolg hebben dat zij bepaalde leersituaties als irrelevant voor hun persoonlijke ontwikkeling

gaan bestempelen en daardoor niet meer bereid zijn er veel inzet voor te leveren. Anderen kunnen zich actief tegen de schoolcultuur gaan verzetten door bijvoorbeeld agressief of lesondergravend gedrag, of door te gaan spijbelen. Wanneer leerlingen gewend raken zich op een bepaalde wijze op school staande te houden, spreken we van een copingstijl. Deze raakt verankerd in de persoonlijkheidsstructuur en is daardoor nog moeilijk te beïnvloeden.

Doel van het onderzoek

In dit onderzoek zijn in een longitudinale opzet gegevens verzameld over stress, coping, en sociale steun, naast schoolprestaties en prestatie-motivatie. De belangrijkste onderzoeksvragen betreffen enerzijds de relaties tussen stress en coping, en de invloed van sociale steun hierbij, anderzijds de effecten van copingstrategieën op de gemeten schoolse variabelen. Aangenomen wordt dat het effect van stress en sociale steun op prestatie en prestatie-motivatie wordt gemedieerd door de copingmodi. Deze aanname volgt uit het werk van Lazarus et al. (cf. Lazarus & Folkman, 1984).

Globaal hebben onze verwachtingen betrekking op twee aspecten. Allereerst wordt verwacht dat sociale ondersteuning een positief effect heeft op de copingmodus. In eerder onderzoek (Boekaerts, Wijdemans, & Seegers, 1991) werd gevonden dat wanneer leerlingen aangaven dat ze konden rekenen op sociale ondersteuning van hun omgeving ze frequenter toenaderingsgedrag als copingmodus hanteerden. In het hier gerapporteerde onderzoek verwachten we bovendien dat frequent hanteren van vermijdingsgedrag als copingmodus leidt tot een afnemende prestatie-motivatie, terwijl toenaderingsgedrag eerder samengaat met een hoger motivatieniveau. Deze verwachting volgt uit resultaten van onderzoek met betrekking tot de relatie tussen test-angst en prestatie-motivatie (bijv. Wieland-Eckelman, Bösel, & Badorrek, 1989). In stressvolle situaties zal een repressieve en defensieve copingmodus gepaard gaan met een afnemend aspiratieniveau (Schönplflug, 1986). Een verminderde prestatie-motivatie kan worden gezien als een defensieve strategie.

METHODE

Steekproef

Het onderzoek kende twee meetmomenten. Gegevens van brugklasleerlingen werden verzameld in de maanden september en oktober. Er werkten 10 scholengemeenschappen mee in de regio Doetinchem-Grave-Venlo. De onderzoeksgroep bestond uit 34 brugklassen met samen 861 leerlingen. De tweede afname vond plaats aan het begin van het derde leerjaar. Van de oorspronkelijke 10 scholengemeenschappen vielen er twee af omdat ze betrokken waren in een reorganisatie. In totaal werden van 561 leerlingen (286 jongens en 275 meisjes) op beide tijdstippen gegevens verkregen. Alleen de gegevens van deze leerlingen zijn in de analyses opgenomen.

Instrumenten

In dit onderzoek werden 4 vragenlijsten afgenomen: de *Stress- en Copingvragenlijst voor Brugklassers*, dit is een aanpassing van een eerdere vragenlijst (Boekaerts, Hendriksen, Maes, 1987), de *Vragenlijst Sociale Ondersteuning* (Boekaerts, 1987), de *Vragenlijst Stressvolle Levensgebeurtenissen* (Maes & Van Asten, 1988), en de *Prestatie Motivatie Vragenlijst*. Het laatste instrument bestaat uit één subschaal van de *Achievement Motivation Scale* van Gjesme & Nygard (1970). Verder zijn rapportcijfers voor wiskunde gebruikt als maat voor prestatie.

De *Stress- en Copingvragenlijst voor Brugklassers* is een aangepaste versie van de *Stress- en Copingvragenlijst voor Basisschoolleerlingen*. Acht typen situaties werden door leerlingen als de meest voorkomende dagelijkse ergernissen genoemd (gepest worden door klasgenoten, de leraar gaat te snel/is onduidelijk, de klas is niet leuk, op school komen en merken dat je dingen vergeten bent, de leraar is niet leuk/erg streng, een slecht punt krijgen, een klasgenoot wordt gepest of uitgescholden, en onterecht strafwerk krijgen). Per situatie wordt gevraagd hoe vaak

leerlingen deze meemaken, hoe intens de ervaren stress is, en wat ze in deze situaties (meestal) doen. Bij de analyse van de data op het eerste meetmoment zijn de copingstrategieën die leerlingen hanteren herleid tot een drietal typen, nl. agressief gedrag ('gaan schelden'), toenaderingsgedrag ('beter gaan leren'), en vermijdingsgedrag ('niks doen of zeggen, want dat helpt toch niet'). Voor een uitgebreide beschrijving zie Boekaerts & Van de Goor (1991). De schaal *agressief gedrag* bestaat uit 31 items. De interne consistentie (Cronbach's alpha) is .87. Voor de tweede meting is de interne consistentie .90. De subschaal *Toenaderingsgedrag* bestaat uit 30 items. Cronbach's alpha is respectievelijk .83 en .84 voor de twee meetmomenten. De subschaal *vermijding* bestaat uit 27 items. Cronbach's alpha voor de twee meetmomenten bedraagt respectievelijk .71 en .78. Voor alle leerlingen is een score bepaald op elk van de drie copingmodi. Deze score is berekend als de gemiddelde score over de verschillende situaties. De correlaties tussen deze scores voor de twee meetmomenten zijn .48 voor agressie, .49 voor toenadering, en .41 voor vermijding. Agressie en toenadering hangen matig positief samen. De correlatie-coëfficiënten zijn respectievelijk .22 en .30 voor de eerste en de tweede meting. De correlatie van agressie met toenadering en vermijding is niet statistisch significant.

De frequentie en intensiteit van ervaren stress werden bepaald door de gemiddelde score over de verschillende situaties te berekenen. De betrouwbaarheid van de schaal voor stressintensiteit was goed. Cronbach's alpha was .88 op beide meetmomenten (aantal items is 21). Dit betekent dat de 7 situaties die werden gebruikt tamelijk homogeen zijn in de mate waarin ze negatieve emoties bij leerlingen oproepen: leerlingen die aangeven veel negatieve emoties te ervaren in een bepaalde stressvolle situatie, geven dit over het algemeen ook aan voor een andere stressvolle situatie.

Voor de schaal stressfrequentie werd een lage betrouwbaarheid gevonden. Dit resultaat is niet verrassend omdat een oordeel geven over de frequentie van meemaken van stress in onderling niet duidelijk samenhangende situaties sterk kan verschillen. Een hogere score op de variabele stressfrequentie drukt uit dat een leerling meerdere van de door ons gehanteerde stressoren herkent en bovendien aangeeft dat ze in het dagelijkse schoolgebeuren frequent voorkomen. Frequentie van meemaken van stressvolle situaties en intensiteit van ervaren stress zijn verschillende aspecten. Dit wordt weerspiegeld in de lage correlatie tussen beide. Voor de twee meetmomenten bedraagt deze respectievelijk .22 en .21. De correlatie tussen de scores voor frequentie van ervaren stress op de twee meetmomenten is .38. Voor intensiteit van ervaren stress is deze samenhang .50. Dit impliceert dat er slechts een matige samenhang is tussen de mate waarin de 7 situaties negatieve emoties uitlokken bij brugklassers en bij dezelfde leerlingen 2 jaar later.

De *Vragenlijst Sociale Ondersteuning* bestaat uit twee delen. Het eerste deel is gebaseerd op de Social Support Questionnaire van Sarason (1984) en be vraagt de ervaren kwantiteit en kwaliteit van sociale steun. Het tweede deel werd door Boekaerts (1987) aan deze algemene lijst toegevoegd. Hier wordt gevraagd hoe leerlingen de didactische en emotionele ondersteuning van de wiskundeleraar en van klasgenoten waarnemen. In dit verslag gaan we alleen in op het toegevoegde deel. Vier subschalen kunnen worden onderscheiden: (1) sociale ondersteuning door klasgenoten (bijv. 'zijn leerlingen blij voor je als je goede resultaten haalt?'), bestaat uit 4 items. Cronbach's alpha is respectievelijk .75 en .71 voor de twee meetmomenten; (2) sociale ondersteuning van de wiskundeleraar (bijv. 'geeft de wiskundeleraar extra hulp als je daarom vraagt?'), bestaat uit 12 items. Cronbach's alpha is hier .80 en .87; (3) waardering van ouders en leerkracht (bijv. 'waarderen je ouders je inzet op school?') bestaat uit 3 items. Cronbach's alpha voor de twee meetmomenten bedraagt respectievelijk .62 en .63; (4) waargenomen sfeer in de klas (bijv. 'leven je klasgenoten met elkaar mee?'). Deze subschaal bevat 2 items. De correlatiecoëfficiënten bedragen .67 en .71 voor de twee meetmomenten (zie Boekaerts & Seegers, in voorbereiding).

De *Vragenlijst Stressvolle Levensgebeurtenissen* be vraagt 11 levensgebeurtenissen. Aan leerlingen wordt gevraagd of zij de gebeurtenis hebben meegemaakt. Als dat het geval is wordt

gevraagd op een vier-puntsschaal aan te geven hoe erg ze deze gebeurtenis vonden. Voorbeelden van situaties zijn een ernstige ziekte thuis, een overleden familielid, en ruzie met klasgenoten. De correlatie tussen deze frequentie van meemaken op de eerste en de tweede meting bedraagt .39.

Om de prestatie-motivatie te meten is de schaal *toenadering* uit de Achievement Motivation Scale van Gjesme & Nygard (1970) gebruikt. Deze schaal bestaat uit 15 items. Een voorbeeld is 'Ik voel me tevreden wanneer ik gebruik kan maken van wat ik kan'. Cronbach's alpha is .87 en .88 voor de twee meetmomenten. De correlatie tussen de scores voor eerste en tweede meting is .47.

RESULTATEN

Een model is verondersteld om de onderliggende verbanden tussen de variabelen te beschrijven en een samenvattend beeld te geven. De structuur van het model in hoofdlijnen is dat de zelf-regulerende vaardigheden die de leerlingen aangeven te hanteren in de 7 situaties worden beïnvloed door de 3 stressvariabelen, namelijk frequentie en intensiteit van ervaren stress in de 7 situaties, en frequentie van meemaken van stressvolle levensgebeurtenissen, en door aspecten van sociale ondersteuning. Op hun beurt beïnvloeden de copingstrategieën die leerlingen hantieren hun leerprestaties en prestatie-motivatie. Het is daarnaast mogelijk dat sociale ondersteuning en stressvariabelen een direct effect hebben op prestatie en motivatie.

Voor toetsing is gebruik gemaakt van het programma LISREL. De gekozen benadering komt overeen met wat Jöreskog (1993) 'model generating' noemt. Een onderliggend model is in hoofdlijnen gespecificeerd, terwijl er relatief veel ruimte is voor modelaanpassingen. Een situatie waarbij een model wordt ontwikkeld en getoetst op eenzelfde dataset, leidt gemakkelijk tot kanskapitalisatie. Een nodige, maar niet voldoende, voorwaarde is dat mogelijke aanpassingen a priori worden vastgelegd en theoretisch adequaat worden onderbouwd. Wanneer van 'model generating' sprake is, is een vorm van cross-validering vereist. In dit onderzoek is de steekproef daarom in twee willekeurige groepen verdeeld, een exploratie- en een validatiesteekproef. Eerst werd een model ontwikkeld voor de exploratiesteekproef. Vervolgens is getoetst of dit model voor beide delen van de steekproef de onderliggende structuur van de data adequaat weergeeft. Hierbij is gebruik gemaakt van de multi-sample optie uit het LISREL programma, waarbij waarden voor de modelparameters voor verschillende groepen gelijktijdig worden geschat. Het programma geeft een chi-kwadraat waarde die uitdrukt in hoeverre het gespecificeerde model voldoet om de structuur van de data van beide groepen te verklaren. Bij toetsing zijn verschillende geneste modellen voor de beide groepen vergeleken op houdbaarheid. Elk model verschilt van een voorgaand model in het aantal restricties dat aan de modelparameters wordt opgelegd. Hierdoor kan gebruik worden gemaakt van de log-likelihood methode om modellen te vergelijken. Het verschil in chi-kwadraat waarden voor twee modellen volgt bij benadering weer een chi-kwadraat verdeling met als aantal vrijheidsgraden het verschil in vrijheidsgraden van het restrictieve en het minder restrictieve model. Is deze chi-kwadraat statistisch significant, dan vormt het minder restrictieve model een meer adequate representatie van de data.

Bij toetsing werd als uitgangspunt een model met maximale restricties gekozen. Hierbij wordt niet alleen aangenomen dat eenzelfde model voor beide groepen adequaat is om de onderliggende structuur van de data te beschrijven, maar worden ook alle relevante parameterwaarden in beide groepen identiek verondersteld. In de daarop volgende stappen wordt het aantal restricties verminderd om na te gaan of de fit van het model hierdoor verbetert. In een eerste reductie van het aantal restricties werden de covarianties tussen de error-termen vrij gelaten. Vervolgens werd een model met minimale restricties aangenomen waarin het onderliggende model voor de beide groepen hetzelfde is, maar waarin de parameters per groep verschillende waarden kunnen aannemen.

Het eerste meetmoment

Het uitgangsmodel in deze studie was een vereenvoudigd model waarin enkel de stressvariabelen, de sociale ondersteuning, en de copingmodi zijn opgenomen. Een model verklaart de onderliggende structuur van de data in termen van directe en indirecte effecten. In de gebruikelijke LISREL opzet worden interactie-effecten niet expliciet uitgedrukt. Daarom is met behulp van multiële lineaire regressie-analyses de invloed van interactie-effecten apart onderzocht. In een eerste stap zijn de interactie-effecten van de 4 aspecten van sociale ondersteuning en de 3 stressvariabelen op de copingmodi (toenadering, agressie, vermijding) onderzocht. In de toetsing werden deze effecten in één stap aan de vergelijking toegevoegd nadat de hoofdeffecten in de vergelijking waren opgenomen. In geen enkel geval resulteerde een statistisch significante toename van de verklaarde variantie. Dit betekent dat we in het model kunnen uitgaan van een direct effect van social support op de copingmodi. Ook in andere onderzoeken werden geen of slechts marginale effecten van de interactie tussen stress en social support gevonden, terwijl wel directe effecten werden gevonden (Dubow & Tisak, 1989; Compas, 1987). In een tweede stap werden ook nog prestatie en prestatie-motivatie als afhankelijke variabelen genomen. De interactie-effecten van respectievelijk stress en copingmodus, en sociale ondersteuning en copingmodus werden in de vergelijking opgenomen. Toevoeging van deze interactie-effecten leidde in geen enkel geval tot een statistisch significante toename van de verklaarde variantie. Besloten is daarom om een model zonder interactie-effecten te toetsen.

De respondenten werden random in twee groepen gesplitst. Ten aanzien van geen enkele variabele was er een statistisch significant verschil tussen de delen van de steekproef. Wanneer een model met maximale restricties wordt aangenomen voor de twee random delen van de steekproef resulteren Goodness-of-Fit Indices van respectievelijk .96 en .95. De χ^2 bedraagt 143.56 met 81 vrijheidsgraden. Wanneer de covarianties tussen de error termen van de afhankelijke variabelen voor de vergeleken groepen werden vrijgelaten, resulteerde dit in een betere fit. De χ^2 waarde bedraagt dan 122.16 met 74 vrijheidsgraden. Het verschil van 21.4 heeft bij benadering een χ^2 verdeling met 7 vrijheidsgraden. Dit is statistisch significant, en bevestigt dat een minder restrictief model een betere representatie van de structuur van de data vormt. Wanneer wordt toegestaan dat de modelparameters voor de verbanden tussen de variabelen in het model verschillende waarden kunnen aannemen voor de vergeleken groepen resulteert bij toetsing een χ^2 waarde van 112.56 met 62 vrijheidsgraden. Ten opzichte van het voorgaande model is deze reductie van de χ^2 waarde niet statistisch significant (χ^2 is 9.6 met $df=12$). Deze reductie in restricties is bijgevolg overbodig. Parameterschattingen voor de exploratie- en de validatiesteekproef (per groep gestandaardiseerde oplossing) zijn in figuur 1 weergegeven, gescheiden door een schuine streep. Squared Multiple Correlations for Structural Equations (SMC) representeren de proporties verklaarde variantie (R^2) voor de afhankelijke variabelen. Voor de copingmodi bedragen deze .06/.10, .26/.25 en .06/.06 voor respectievelijk agressie, toenadering, en vermijding, terwijl voor prestatie-motivatie een SMC resulteert van .06 voor beide steekproefdelen.

In de brugklas zien we een duidelijk effect van frequentie van meemaken van stressvolle situaties op agressie als copingmodus, maar niet op toenadering en vermijding. Verder heeft frequentie van meemaken van stressvolle levensgebeurtenissen een (beperkte) invloed op agressie, maar niet op de beide andere copingmodi. Intensiteit van ervaren stress beïnvloedt zowel toenadering als vermijding, maar beïnvloedt agressie niet. Alle vier de aspecten van sociale ondersteuning hebben een gunstige invloed op toenadering als copingmodus. Waardering van ouders en wiskundeleraar bevordert vermijdingsgedrag. Er is geen effect van sociale ondersteuning op agressie. De resultaten laten ook een samenhang zien tussen agressie en toenadering enerzijds en prestatie-motivatie anderzijds. Beide vormen van coping beïnvloeden de prestatie-motivatie. Er is geen verband tussen vermijding en prestatie-motivatie. De steun die van de wiskundeleraar wordt ervaren draagt direct bij aan de prestatie-motivatie. Geen van de variabelen in het model vertoont een samenhang met het rapportcijfer voor wiskunde.

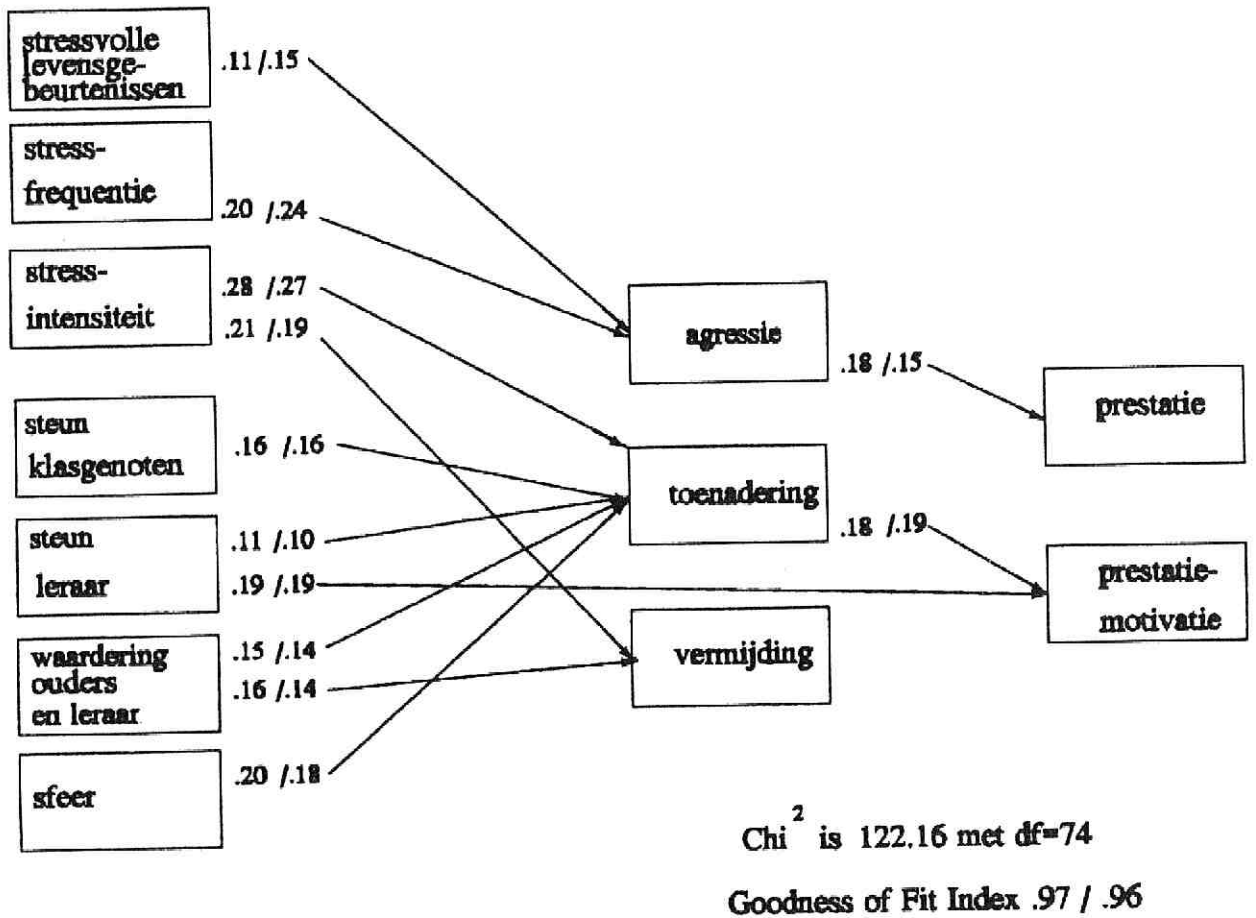


Fig. 1. Grafische weergave van het basismodel voor de eerste meting (brugklas) bij random verdeling van de steekproef. Parameterschattingen voor de twee delen van de steekproef zijn in de figuur weergegeven.

Het tweede meetmoment

Voor de tweede meting (begin derde leerjaar) zijn we uitgegaan van een verdeling van de steekproef in dezelfde twee delen. Ook nu werd t.a.v. geen enkele variabele een statistisch significant verschil gevonden tussen de delen van de steekproef. Om de onderliggende structuur van de data van het tweede meetmoment te toetsen werd verondersteld dat het basismodel een adequate representatie vormt van de onderliggende structuur van de data. De houdbaarheid van dit model werd op dezelfde manier onderzocht als in de vorige analyses. Uitgangspunt was een model met maximale restricties, waarbij alle parameterschattingen voor beide delen als identiek werden verondersteld. Toetsing van dit model onder de meest restrictieve assumpties geeft Goodness-of-Fit Indices van respectievelijk .95 en .96 voor de twee groepen. De χ^2 waarde bedraagt 157.28 met 81 vrijheidsgraden. Wanneer de assumptie van gelijke covarianties tussen de error termen van de afhankelijke variabelen wordt losgelaten resulteert een waarde van 143.73 met 74 vrijheidsgraden. Het verschil van 13.55 heeft weer bij benadering een χ^2 verdeling met 7 vrijheidsgraden. Deze waarde is echter niet statistisch significant. Wanneer slechts minimale restricties worden opgelegd aan het model resulteert een model met een χ^2 waarde van 129.88 met 62 vrijheidsgraden. Ten opzichte van het meest restrictieve model is de afname van de χ^2 niet statistisch significant (χ^2 is 27.4; $df=19$; $p>.05$).

Op dit punt moet echter beoordeeld worden of de aanname dat de onderliggende structuur van de data op beide meetmomenten dezelfde is, wel juist is. Met andere woorden, is de aanname dat eenzelfde model een adequate representatie vormt voor de structuur van de data op de

twee meetmomenten correct, of zijn verbanden tussen variabelen in de tussenliggende tijd structureel veranderd? In dat geval zal reductie van opgelegde restricties uiteindelijk geen bevredigend model opleveren omdat verbanden tussen variabelen aan verandering onderhevig zijn. De resultaten van de parameterschatting laten zien dat modificaties de fit van het model aanzienlijk kunnen verbeteren. Deze resultaten worden hier niet grafisch weergegeven omdat het gevonden model nog onvoldoende adequaat is. We bespreken de meest opmerkelijke verschillen. Het effect van stressvolle levensgebeurtenissen is niet langer statistisch significant in beide groepen. Het toevoegen van een direct verband tussen waardering door ouders en wiskundeleraar enerzijds en agressieve copingstrategieën anderzijds, verbetert de fit van het model voor beide vergeleken groepen. Het gemodificeerde model wordt weergegeven in Figuur 2.

Wanneer deze veranderingen ten opzichte van het voor het eerste meetmoment ontwikkelde model worden geaccepteerd vinden we bij toetsing van dit model met maximale restricties is de Goodness of Fit Index .96 voor beide steekproefdelen. De χ^2 waarde bedraagt 138.83 met 81 vrijheidsgraden. Wanneer de assumptie van gelijke covarianties tussen de error termen van de afhankelijke variabelen wordt losgelaten resulteert een χ^2 waarde van 125.83 met $df=74$. Deze verbetering is niet statistisch significant (χ^2 is 13.01, $df=7$; $p>.05$). Ook bij verdere vermindering van de restricties is geen sprake van een statistisch significante verbetering. Het minst restrictieve model heeft een χ^2 waarde van 113.07 met 62 vrijheidsgraden. Ten opzichte van het restrictieve beginmodel is deze afname niet statistisch significant (χ^2 is 25.76 met $df=19$; $p>.05$).

In het eindmodel zijn voor de copingmodi agressie, toenadering, en vermijding de proporties verklaarde variantie respectievelijk .19/.24, .20/.28, en .14/.16. Voor prestatiemotivatie is dit respectievelijk .08 en .06 voor de twee delen van de steekproef.

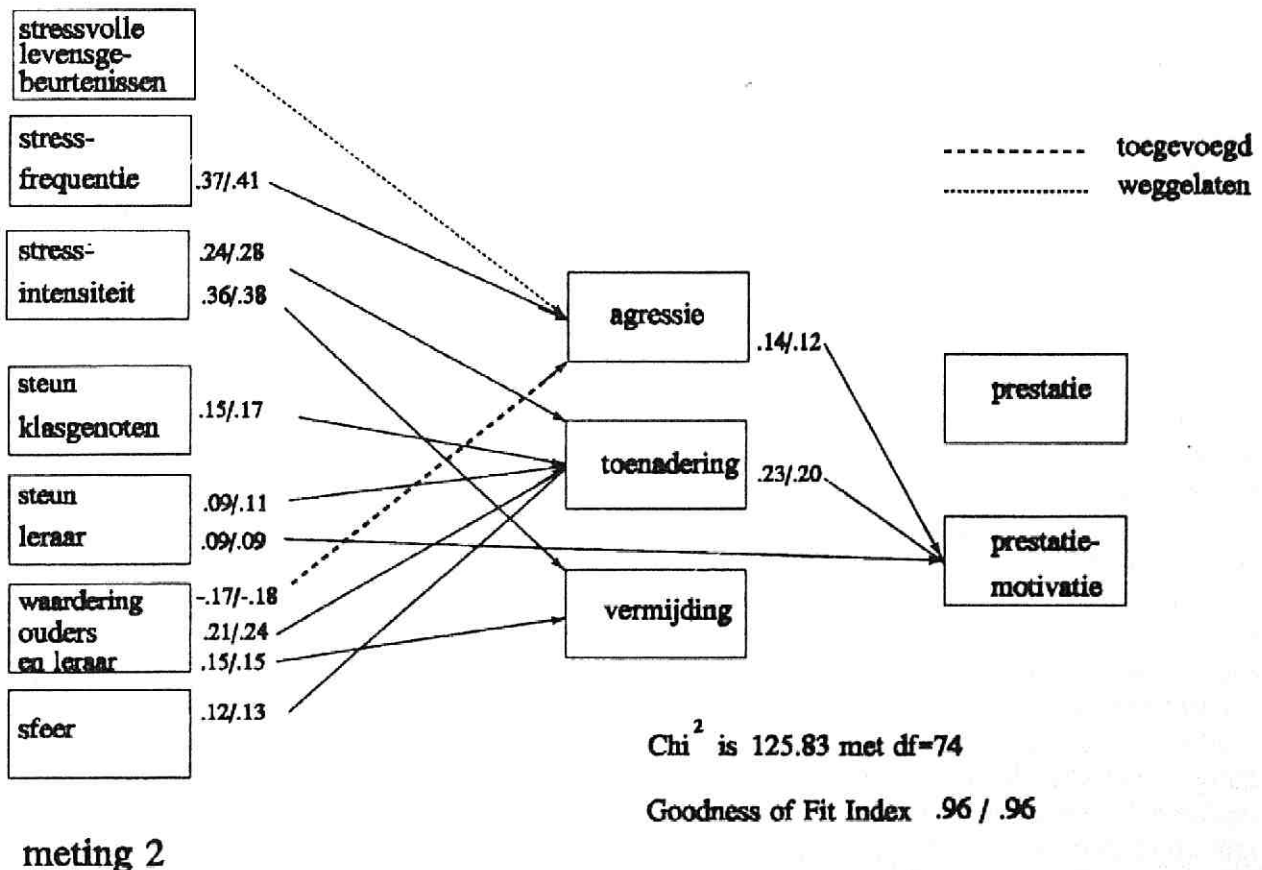


Fig. 2. Grafische weergave van het model voor het tweede meetmoment (begin derde leerjaar). Parameterschattingen voor de twee delen van de steekproef zijn in de figuur weergegeven.

Het is belangrijk op te merken dat er correlaties bestaan tussen aspecten van sociale ondersteuning en stressvariabelen. Deze correlaties wijzen mogelijk op substantiële indirecte effecten van sociale ondersteuning op copingmodi. We zullen de meest belangrijke samenhangen bespreken. Op het eerste meetmoment is de correlatie tussen ervaren sfeer in de klas en de frequentie waarmee stressvolle situaties worden meegemaakt respectievelijk $-.22$ en $-.28$ voor de twee steekproefdelen. Een positief ervaren sfeer gaat dus samen met een lagere frequentie van meemaken van de 7 situaties. Dit betekent dat sfeer in de klas een indirect effect heeft op de copingmodus agressie via waargenomen frequentie van stressvolle situaties. Op het tweede meetmoment is deze samenhang echter afgenomen tot respectievelijk $-.11$ en $-.10$. Wel is de samenhang tussen waardering van ouders en wiskundeleraar enerzijds, en waargenomen frequentie anderzijds, toegenomen ($-.24$ en $-.12$), evenals de samenhang tussen ervaren steun van de wiskundeleraar en frequentie ($-.18$ en $-.25$). Ook hier is sprake van een indirect effect op agressie, maar de grootte van de gevonden verbanden dwingt niet tot het toekennen van een belangrijke rol aan deze indirecte effecten.

CONCLUSIES EN DISCUSSIE

De resultaten van dit onderzoek bevestigen de samenhang tussen stressvariabelen en copingmodi. De frequentie waarmee de 7 stressvolle situaties worden ervaren beïnvloedt vooral het hanteren van agressief copinggedrag. De intensiteit waarmee stress wordt ervaren in de 7 situaties die we de leerlingen voorlegden hangt samen met toenaderings- en vermijdingsgedrag. Sociale steun heeft daarbij een beperkt moderend effect. Dit wil zeggen dat sociale steun vooral toenadering als copingmodus bevordert. Ook is er een samenhang tussen de waardering die leerlingen van ouders en wiskundeleraar ontvangen en agressief gedrag als copingmodus. Het verband is negatief: naarmate leerlingen meer waardering waarnemen neemt de voorkeur voor agressief gedrag als copingmodus af.

Vergelijking van Figuren 1 en 2 laat zien dat de onderliggende structuur van de resulterende modellen op beide meetmomenten grotendeels gelijk is. Er zijn echter ook verschillen. Zo neemt het effect van stressvolle levensgebeurtenissen op agressie als copingmodus af, en oefent de waardering die van ouders en wiskundeleraar wordt ondervonden in toenemende mate een modererend effect uit op agressie. Verder neemt de invloed van de frequentie waarmee stressvolle situaties worden ervaren op agressief copinggedrag toe.

In het algemeen worden prestatie en prestatiemotivatie niet of nauwelijks beïnvloed door stress en coping. De resultaten laten zien dat leerlingen die in stressvolle situaties bij voorkeur agressie of toenadering hanteren meer prestatiegemotiveerd zijn dan leerlingen die vooral reageren met vermijdingsgedrag. De verwachting dat vermindering gepaard zou gaan met een afnemende prestatiemotivatie werd echter niet bevestigd. Verder wijzen de resultaten erop dat de ervaren steun van de wiskundeleraar een directe invloed heeft op de prestatiemotivatie. Bij vergelijking van de situatie in de brugklas met die aan het begin van het derde jaar neemt de sterkte van dit verband echter af.

Het geringe effect dat stress- en copingvariabelen hebben op de in het onderzoek betrokken schoolse variabelen hangt wellicht samen met het feit dat stress in het algemeen geen probleem vormt voor de meerderheid van de onderzochte leerlingen. Dat meer leerlingen uitvallen die een relatief hoge stressbelasting en weinig social support ervaren, is inherent aan dit type onderzoek ('subject attrition'). Het gevolg daarvan is dat de kans afneemt om effecten te vinden die samenhangen met stressbelasting. Het identificeren van risicogroepen was echter niet het doel van dit onderzoek. Het belangrijkste doel was de samenhang tussen de onderzochte variabelen vaststellen, en nagaan of deze structuur relatief stabiel blijft van brugklas naar het begin van het derde schooljaar. De resultaten kunnen gebruikt worden als uitgangspunt voor verder onderzoek.

Een ander punt waarop gewezen moet worden is dat coping in dit onderzoek als algemene strategie is beschouwd die leerlingen in stressvolle maar ongelijksoortige situaties hanteren. Er zijn echter aanwijzingen dat coping deels domeingebonden is (Seiffge-Krenke, 1990). Om de

rol van coping nauwkeuriger in te schatten is het daarom van belang dat rekening wordt gehouden met deze domein-afhankelijkheid.

In het onderzoek is het gebruik van lineaire modellen exploratief geweest ('model generatig'). Daarom is replicatie van het onderzoek essentieel om de validiteit van gevonden resultaten op de proef te stellen.

LITERATUUR

- Boekaerts, M. (1987). *De Social Support vragenlijst* (ongepubliceerde versie). Nijmegen: Vakgroep Onderwijspsychologie.
- Boekaerts, M., & Goor, J. van de (1991). *Stress in de brugklas* (Eindrapport SVO 9404). Nijmegen/Leiden: Vakgroep Interdisciplinaire Onderwijskunde/Vakgroep Onderwijsstudies.
- Boekaerts, M., Hendriksen, J., & Maes, S. (1988). *Stress- en Copingvragenlijst voor leerlingen*. Kath. Universiteit Nijmegen: Vakgroep Interdisciplinaire Onderwijskunde.
- Boekaerts, M. & Seegers, G. (in voorbereiding). *Stress and social support*.
- Boekaerts, M., Wijdemans, M., & Seegers, G. (1991). Stress op de basisschool: Reken maar op je vriendjes! In B.H.A.M. van Hout-Wolters & L.F.W. de Klerk (eds.). *Onderwijsleerprocessen* (pp. 9-19). Amsterdam: SCO.
- Compas, B.E. (1987). Coping with stress during childhood and adolescence. *Psychological Bulletin*, 101, 393-403.
- Compennolle, T.H.L. (1987). *Stress op de middelbare school*. Leuven: ACCO.
- Corte, E. de, Van Buggenhout, J., & Pyck, K. (1985). *Houding tegenover de leervakken van het middelbaar onderwijs. Een exakt onderzoek bij zittenblijvers en normaal vorderende leerlingen*. Amsterdam: Swets & Zeitlinger.
- Dubow, E.F., & Tisak, J. (1989). The relation between stressful life events and adjustment in elementary school children: The role of social support and social problem-solving. *Child Development*, 60, 1412-1423.
- Gjesme, T., & Nygard, R. (1970). *Achievement-related motives: Theoretical considerations and construction of a measuring instrument*. Unpublished Report, University of Oslo.
- Hendriksen, J. (1990). *Stress en copinggedrag bij leerlingen uit groep zeven en acht van de basisschool*. Acad. proefschrift, Kath. Universiteit Nijmegen: vakgroep Interdisciplinaire Onderwijskunde.
- Jöreskog, K.G. (1993). Testing structural equation models. In K.A. Bollen & J. Scott Long (eds.). *Testing structural equation models* (pp.294-316). Newbury Park: Sage.
- Lazarus, R.S., & Folkman, S. (1984). *Stress, appraisal and coping*. New York: Springer.
- Linden, F.J. van der, & Roeders, P.J.B. (1982). *Schoolgaande jongeren, hun leefwereld en zelfbeleving*. Nijmegen: Hoogveld Instituut.
- Maes, S., & Asten, J. van (1988). *Vragenlijst Stressvolle Levensgebeurtenissen* (ongepubl.). Tilburg: Kath. Universiteit Brabant.
- Moos, R.H. (1979). Educational climates. In H. Walberg (ed.). *Educational environments and effects*. Berkely: McCuthan.
- Mooij, T. (1992). *Pesten in het onderwijs*. Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.
- Sarason, I.G. (1984). Stress, anxiety, and cognitive interference: Reaction to tests. *Journal of Personality and Social Psychology*, 46, 929-938.
- Schunk, D.H. (1987). Peer model and children's behavioral change. *Review of Educational Research*, 57, 149-174.
- Schönpflug, W. (1986). Behavior economics as an approach to stress theory. In M.H. Appley & R. Trumbull (eds.). *Dynamics of stress* (pp.81-98). New York: Plenum.
- Seiffge-Krenke, I. (1990). Developmental processes in self-concept and coping behaviour. In H. Bosma & S. Jackson (eds.). *Coping and self-concept in adolescence* (pp. 51-68). Berlin: Springer.
- Stoel, W.G. (1980). *De relatie tussen grootte voor voortgezet onderwijs en het welbevinden van de leerling* (Deel 1 en 2). Haren RION.
- Wieland-Eckelmann, R., Bösel, R., & Badorrek, W. (1989). Anxiety-related coping styles and the achievement process: State control and task performance. In F. Halisch & J.H.L. van den Bercken (eds.). *International perspectives on achievement and task motivation* (pp. 131-152). Amsterdam-Lisse: Swets & Zeitlinger.

Boekbesprekingen

R.J. Bijleveld, *Numeriek rendement en studiestaking. Een theoretische analyse van factoren die samenhangen met rendement en studiestaking in het wetenschappelijk onderwijs*. Proefschrift Universiteit Twente 1993. Ook verschenen als handelsuitgave: Utrecht (Lemma), 1993.

M.J.F. Hulshof, J.A.A.M. Bevers, en J.M.H.M. Willems (red.), *Universiteit & Diversiteit. Liber amicorum voor Hans van Hout bij gelegenheid van zijn afscheid van het IOWO*. Nijmegen (KU Nijmegen / IOWO), 1993.

Inleiding

Ten aanzien van wetenschappelijk onderzoek zijn een theoretische en een praktische doelstelling te onderscheiden. Een theoretische doelstelling is in de regel gericht op een bijdrage aan de theorievorming ten aanzien van een aspect van het object van onderzoek. Een praktische doelstelling beoogt veelal een bijdrage te leveren aan de verbetering van een bestaande praktijk-situatie waarin het object van onderzoek is ingebed. Gegeven deze doelstellingen is het ook mogelijk de theoretische of praktische doelstelling te accentueren. Het proefschrift van Bijleveld gaat uit van een theoretische positie; het boek van Hulshof, Bevers & Willems van een praktische positie. Beide boeken gaan uit van geconstateerde verschillen in het hoger onderwijs. Gelet op hun gekozen positie reageert elk op een eigen wijze op deze verschillen: Bijleveld probeert de verschillen te verklaren; Hulshof c.s. proberen de verschillen te reguleren. Door deze boeken na elkaar te bespreken hoop ik het contrast tussen beide posities duidelijk uit te laten komen.

RIEKELE BIJLEVELD'S NUMERIEK RENDEMENT EN STUDIESTAKING

Samenvatting van de inhoud

Het proefschrift omvat negen hoofdstukken. De studie heeft tot doel een bijdrage te leveren aan het proces van theorievorming over de relatie tussen onderwijsprocesvariabelen en effectiviteit van het wetenschappelijk onderwijs. De centrale probleemstelling bestaat uit twee delen:

1. Welke samenhangen bestaan er volgens de nationale en internationale onderzoeksliteratuur tussen kenmerken van het onderwijsproces en het numeriek rendement van een studierichting in het wetenschappelijk onderwijs?
2. Hoe ziet een theoretisch model eruit dat met behulp van studierichtingkenmerken een verklaring kan geven voor verschillen in rendement tussen studierichtingen in het wetenschappelijk onderwijs in Nederland?

Bijleveld gaat na een inleidend, eerste hoofdstuk en een drietal hoofdstukken over de plaats, definitie en ontwikkeling van het numeriek rendement (hoofdstukken 2 t/m 4) op de twee genoemde probleemstellingen in met respectievelijk vier hoofdstukken (hoofdstuk 5 t/m 8) en één hoofdstuk (hoofdstuk 9).

1. Plaats, definitie en ontwikkeling van het numeriek rendement

Het numeriek rendement wordt in navolging van het CBS gedefinieerd als het percentage studenten gemeten aan het aantal studenten bij eerste inschrijving, dat na een gegeven aantal jaren een bepaald examen met succes heeft afgelegd. Pas met het Plaatsen-Geld Model (1982) komt er meer aandacht voor het rendement. De kwantitatieve ontwikkeling van het numeriek rendement van het wetenschappelijk onderwijs in Nederland kenmerkt zich in het algemeen sinds de dertiger jaren door een dalende tendens; pas met de invoering van de twee-fasenstructuur lijkt deze over te gaan in een stijging. De hoogte van het numeriek studierendement loopt op verschillende aspecten uiteen.

2. *Samenhangen tussen kenmerken van het onderwijsproces en het numeriek rendement van een studierichting*

Voor wat betreft de kenmerken richt Bijleveld zich eerst op de institutionele factoren: kenmerken die verschillen in "learning" en "educational attainment" (bereikt onderwijsniveau) tussen instellingen kunnen verklaren, en kenmerken die verschillen in intellectuele ontwikkeling ("learning") binnen instellingen kunnen verklaren. Bestaand onderzoek blijkt geen aanwijzingen te geven dat verschillen tussen instellingen van belang zijn voor het niveau van "learning". Twee variabelen, individualisering van instructietechnieken en het oordeel van studenten over docer-kwaliteiten van de wetenschappelijke staf, blijken verschillen binnen instellingen t.a.v. "learning" te kunnen verklaren. De beste voorspeller van "educational attainment" zijn de studieresultaten van de afzonderlijke onderdelen.

Vervolgens onderneemt Bijleveld een grondige analyse van vier theoretische invalshoeken die zijn toegepast ten aanzien van studiestaking: de psychologische, maatschappelijke, economische en organisationele invalshoek. De uiteindelijke conclusie is dat de vier benaderingen niet echt duidelijk maken waarom bepaalde individuele kenmerken of organisatie-kenmerken uiteindelijk tot studiestaking leiden.

Ook de interactionele invalshoek is volgens Bijleveld niet echt bruikbaar, omdat er sprake is van een te eenzijdige sociologische invalshoek waardoor onderwijskundige en institutionele factoren naar de achtergrond zijn gedrongen (b.v. het model van Tinto). Bijleveld concludeert dat juist variabelen die te maken hebben met het onderwijsproces in Nederland verschillen in studiestaking tussen Nederlandse studierichtingen kunnen verklaren. Bijleveld onderscheidt vervolgens student- en studierichtingfactoren in het geïnventariseerde Nederlandse onderzoek. Ten aanzien van de studentfactoren komt hij tot de conclusie dat geen van de onderzochte variabelen een overtuigende samenhang vertoont met studiestaking. Dit betreft de variabelen: diploma dat recht geeft op toelating tot het wetenschappelijk onderwijs, de samenstelling van het eindexamenpakket, de hoogte van de eindexamencijfers, intelligentie, algemene studievaardigheden, studieplanningsvaardigheden, studiestrategie, verwachtingen m.b.t. de studie, sociale integratie in het studentenleven, persoonlijkheidskenmerken en achtergrondvariabelen als leeftijd, geslacht en sociaal-economische status van de ouders. Voorts blijkt dat onderzoek naar het verband tussen studentkenmerken, opgevat als instroomkenmerken van studierichtingen, en numeriek rendement zeer schaars is. Ten aanzien van studierichtingskenmerken analyseert Bijleveld een groot aantal variabelen waarvan de meeste niet onderzocht blijken te zijn op studierichtingsniveau. Tentamenspreiding is een uitzondering: spreiding van tentamens leidt tot een meer gelijkmatige studie-inspanning en tot betere studieresultaten.

3. *Constructie van een eigen theoretisch model*

Het "mastery learning model" (b.v. Carroll, 1989) is het uitgangspunt voor het theoretisch model waarmee met name de leertijd die studenten kunnen en willen besteden van belang is voor het behalen van voldoende studieresultaat. Het theoretisch model richt zich dan ook in het bijzonder op de middengroep van studenten die in principe wel over de gevraagde capaciteiten beschikken, maar waarvan de inzet te gering is om een voldoende studieresultaat te boeken. Het theoretisch model spitst zich toe op het verband tussen studierichtingsfactoren en de inzet van studenten. De theorie stelt dat de invloed van de studierichting verloopt via het concept "motivatie" dat wordt gedefinieerd als de drang van het individu om prestaties te leveren. Deze drang kan zowel tot stand komen via interne prikkels, met name eigen interesse, als via externe prikkels, met name druk vanuit de onderwijsorganisatie om prestaties te leveren. De aanwezigheid van beide soorten prikkels is noodzakelijk om de student tot de noodzakelijke minimum-inspanning te brengen. De centrale veronderstelling is dat naarmate het onderwijsproces beter inspeelt op de interne prikkels tot studeren van de student ("activatie van de interesse") en er meer externe prikkels aanwezig zijn om te studeren ("studiedruk"), de kans groter is dat studenten de noodzakelijke studie-inspanning zullen leveren.

Vervolgens formuleert Bijleveld twee blokken van hypothesen aangaande het verband tussen studierichtingskenmerken en studieresultaat (het percentage studiestakers) gekoppeld aan de

concepten "activatie van de interesse" en "studiedruk". Bijleveld formuleert ook ten aanzien van de beschikbaarheid van minimale intellectuele capaciteiten twee blokken van hypothesen. Deze beschikbaarheid ziet Bijleveld afhankelijk van de moeilijkheidsgraad van het gegeven onderwijs (het gevraagde wiskundige inzicht) en van de samenstelling van de instroom. Bijleveld veronderstelt dat de instroomkenmerken de effecten van de hiervoor genoemde studierichtingsvariabelen conditioneren.

Commentaar

Bijleveld heeft geprobeerd met een brede benadering zijn vraagstellingen te beantwoorden en zijn bijdrage aan de theorievorming te leveren. Ik heb waardering voor een benadering waarin vanuit verschillende theoretische perspectieven confronterend naar resultaten van onderwijs en determinanten daarvan wordt gekeken. Ook kan ik waardering opbrengen voor een benadering waarin nu eens niet het individuele gedrag van de student als uitgangspunt wordt gekozen, maar de output van de onderwijsorganisatie: numeriek rendement en studiestaking. Een dergelijke studie past in de huidige ontwikkelingen aangaande kwaliteitszorg. Zowel vanuit theoretisch als vanuit praktisch gezichtspunt is het uiterst relevant over een goed beargumenteerd model te beschikken dat enerzijds een kader biedt voor hypothesevorming en -toetsing en anderzijds een basis kan zijn voor het selecteren van praktische maatregelen ter verbetering van de kwaliteit van het wetenschappelijk onderwijs. Aansluitend bij de laatste, praktische positie, komt het voorgestelde model in hoofdstuk 9 – even los gezien van de voorafgaande hoofdstukken – op zich als bruikbaar en waard om te toetsen over. Echter, als de voorafgaande hoofdstukken erbij worden betrokken, dan schort het aan de onderbouwing. Dit is te wijten aan de opzet van de studie, aan de wijze waarop met de aard van de vraagstelling is omgegaan en aan de methodologie. Een toelichting:

ad. 1. de opzet van de studie

Mijn kritiek betreft het tekort aan scherpte in de beantwoording van de onderzoeksvragen en de onduidelijke relatie tussen de antwoorden op de eerste en tweede onderzoeksvraag.

De tekortschietende scherpte komt bijvoorbeeld tot uitdrukking in de opname van hoofdstuk 2 waarin de plaats van het numeriek rendement in het naoorlogse hoger onderwijsbeleid wordt samengevat. Het hoofdstuk geeft nauwelijks een bijdrage aan het antwoord op de gestelde onderzoeksvragen (blz. 5). Wat mij betreft had Bijleveld zich voorts in hoofdstuk 4 over de kwantitatieve ontwikkeling kunnen beperken tot verschillen tussen studierichtingen.

De relaties tussen de antwoorden op de eerste en tweede onderzoeksvraag zijn onduidelijk of worden onduidelijk gerapporteerd. Op gezette tijden wordt gesteld dat de uitkomsten van de eerste onderzoeksvraag bouwstenen leveren voor de formulering van het theoretisch model. Een expliciete argumentatie voor het al of niet opnemen van bepaalde variabelen ontbreekt evenwel. Ook het feit dat na drie hoofdstukken meta-analyses van resultaten en theoretische invalshoeken van de internationale (vooral Amerikaanse) literatuur (de hoofdstukken 5 t/m 7) de auteur meent dat "juist variabelen die te maken hebben met het onderwijsproces in Nederland verschillen in studiestaking tussen studierichtingen kunnen verklaren" en vervolgens in hoofdstuk 8 aangaande de Nederlandse literatuur wordt geconcludeerd, dat er weinig of geen bouwstenen zijn (blz. 186), doet nog eens met nadruk de vraag rijzen wat nu de argumentatie is om bepaalde variabelen, voor zover aanwezig, al of niet op te nemen.

ad. 2 De aard van de vraagstelling

Er is sprake van een multi-niveau vraagstelling. Mijn kritiek is dat Bijleveld niet goed met deze vraagstelling omgaat, omdat hij de verschillende niveaus te veel impliciet laat en daarmee het gevaar loopt dat niveaus door elkaar gaan lopen. Op zich is, zoals gezegd, de invalshoek zeer interessant: het verklaren van verschillen in rendement op het niveau van de studierichting, de onderwijsorganisatie (meso-niveau; zie tabel 1.1., blz. 2). Dit is afwijkend van wat veelal in onderwijskundige toepassingen gebeurt waarin de leerresultaten van de student (micro-niveau) de afhankelijke variabelen zijn (zie b.v. onderzoek van schoolloopbanen: Voeten & Van der

Kamp, 1990; Raudenbush & Bryk, 1988). Door de te verklaren verschillen in resultaten op het meso-niveau als uitgangspunt te nemen, kunnen verklarende factoren zowel in de inrichting van het micro-niveau (directe interacties studenten – docenten), het meso-niveau (organisatie van de studierichting), als het macro-niveau (b.v. hoger onderwijsbeleid, arbeidsmarktontwikkelingen) nader worden onderzocht.

Met name het micro- en meso-niveau blijven te impliciet zodat het voor de lezer soms een puzzel is op welk niveau de auteur nu bezig is. Voorbeelden van een dergelijke implicietheid van niveaus zijn het conceptuele schema voor de analyse van factoren die “educational attainment” beïnvloeden (figuur 5.1., blz. 76), de vijf theoretische benaderingen in hoofdstuk 6 en 7 en de conclusies over het effect van studentfactoren (blz. 184 en 185).

ad. 3 De methodologie

Mijn kritiek betreft het definiëren van de gebruikte begrippen en de gevolgde werkwijze. Het definiëren van de kernbegrippen uit de probleemstelling is zeer verschillend. Het begrip “numeriek rendement” wordt zeer precies en zorgvuldig gedefinieerd, gebaseerd op een met anderen uitgevoerde aparte studie (Jurgens, Prins & Bijleveld, 1993). Het begrip “onderwijsproces” wordt evenwel nergens gedefinieerd.

Ook is het niet helder welke studierichtingen Bijleveld nu op het oog heeft: vergelijkbare studierichtingen met rendementsverschillen (blz. 7 en 8), of alle studierichtingen vanaf 1930 (hoofdstuk 4)? Voorts wordt er naar mijn idee te snel naar verwante begrippen gesprongen. Voorbeelden zijn het begrip “studiestaking” dat wordt uitgebreid tot “andere vormen van studievoortgang en studiestaking” (blz. 149) en het begrip “educational attainment” dat in vele verschillende betekenissen wordt gebruikt (zie blz. 75, 82 en 83). Bovendien leidt het complementaire gebruik van het begrip studiestaking (blz. 75) tot een m.i. problematische sprong van feiten naar een conclusie (zie voor de “logical flaw”: Campbell, 1974, 112).

Ten slotte vind ik de definiëring van het begrip “learning” als de “ontwikkelingen in taalvaardigheden, kwantitatieve vaardigheden en kennis van het onderwerp” (blz. 75; overigens, op blz. 78 maakt de ontwikkelingsdimensie weer geen deel uit van de definitie) veel te mager voor het type leerresultaten dat in het hoger onderwijs wordt beoogd. Een definitie waarin het begrip “expertise” met al haar verschillende soorten kennis centraal staat, lijkt hier meer op z'n plaats.

Voor wat betreft de gebruikte methode stelt Bijleveld de methode van meta-analyse wel zeer summier aan de orde (alleen op blz. 77). Bovendien is het niet duidelijk of de op deze bladzijde verwoorde werkwijze aangaande meta-analyse ook zal worden toegepast in de andere hoofdstukken. Ten slotte mis ik een nadere argumentatie dat de conclusies, gebaseerd op de Amerikaanse schoolcultuur, ook van betekenis zijn voor het Nederlands wetenschappelijk onderwijs (blz. 91). Naar mijn oordeel ware het beter geweest als deze methodische aspecten grondig waren besproken en deel hadden uitgemaakt van de eerder genoemde argumentatie (zie ad. 1) om tot een theoretisch model te komen.

HULSHOF C.S.' UNIVERSITEIT EN DIVERSITEIT

Samenvatting van de inhoud

Het boek, uitgebracht ter gelegenheid van het afscheid van dr. Hans van Hout als directeur van het IOWO, bevat vijf delen waarin verschillende auteurs op verschillende niveaus van het hoger onderwijssysteem een bepaalde manier van omgaan met diversiteit (en daarmee met ‘universiteit’ als contrast van ‘diversiteit’) aan de orde stellen: (a) macro-niveau: landelijk beleid; (b) meso-niveau: onderwijsorganisatie en -management, kwaliteitszorg, en onderwijsontwikkeling; en (c) micro-niveau: onderwijsontwikkeling, en computerondersteund onderwijs.

Mede gelet op de inleiding zou ik de wat impliciete centrale vraag van het boek als volgt willen omschrijven: op welke manier moet er een goede verhouding worden gevonden tussen enerzijds allerlei vormen van variatie in het hoger onderwijs en anderzijds een centrale tendentie?

Macro-niveau

Drie thema's uit het landelijk beleid komen aan de orde: (a) het systeem van kwaliteitsbewaking (Mertens & Paardekoper, en Vroeijensteijn); (b) de invulling van de bovenbouw, de tweede fase van het voortgezet onderwijs (Van der Weiden); en (c) het AIO-systeem (Hulshof).

De drie thema's staan tegelijkertijd voor drie tendensen. In de eerste plaats de tendens stelling te nemen tegen centralisatie in het onderwijsbeleid. Zowel Mertens & Paardekoper als Vroeijensteijn benadrukken een stelsel van decentrale kwaliteitszorg (en daarmee diversiteit) waarin de verbetering van de onderwijskwaliteit van een bepaalde faculteit voorop staat.

In de tweede plaats de tendens aandacht te schenken aan meisjes in relatie tot het kiezen van een bèta-studie. Van der Weiden pleit voor een grotere diversiteit in de keuzepatronen in het VWO en vermoedt dat er vooral bij meisjes nog reserve aan capaciteit aanwezig is ten aanzien van bèta-studies.

In de derde plaats de tendens in het beleid rekening te houden met de disciplinaire achtergrond van wetenschappers. Hulshof laat in dit verband in een empirische studie (multipel regressie) zien dat verschillen in houdingen van WO-docenten ten aanzien van het AIO-systeem en de onderzoeksschool verklaard kunnen worden uit discipline-kenmerken en daarmee samenhangende opleidingsconcepties.

Meso-niveau

Ten aanzien van onderwijsorganisatie en management komen drie thema's aan de orde: (a) samenwerking in het hoger onderwijs (Kleinbergen & Spruit); (b) strategisch beleid van scholen (Giesbers); en (c) de structuur van de universitaire onderwijsorganisatie (Korsten en Aarts).

De kwaliteitszorg betreft twee thema's: (a) de definitie van het begrip numeriek rendement (Prins); en (b) de relatie hoger onderwijs – maatschappelijke effectiviteit, in het bijzonder op de arbeidsmarkt (Verrijt, Nedermeijer).

De meeste auteurs gaan uit van de tendens een onderwijsinstelling te zien als een min of meer open systeem in wisselwerking met de omgeving. Een onderwijsinstelling wordt daarbij gezien als een levende organisatie. Voor wat betreft de externe relaties van het systeem "onderwijsinstelling" besteden Kleinbergen & Spruit aandacht aan de samenwerking in de vorm van interorganisatiele netwerken (strategisch beleid) en gaan Verrijt en Nedermeijer in op de relatie met de arbeidsmarkt. Kleinbergen & Spruit laten enerzijds allerlei verschillen zien in de motieven tot en vormgeving van samenwerkingsverbanden, en komen anderzijds tot de conclusie dat er ook een gemeenschappelijk strategisch beleid noodzakelijk is voor samenwerking. Verrijt analyseert de problemen van een onderwijsinstelling in het HO als deze het onderwijs wil afstemmen op de zeer diverse arbeidsmarkt. Deze analyse resulteert in een aantal voorwaarden waaraan een arbeidsmarktgeoriënteerd beleid moet voldoen. Nedermeijer sluit concreet bij één voorwaarde van Verrijt aan door zijn bespreking van een instrument om de relatie tussen de beroepspraktijk en het curriculum te optimaliseren: een beroepsopleidingsprofiel (BOP). De diversiteit van de arbeidsmarkt wordt gereguleerd met activiteiten als het opstellen van beroepsbeelden, het formuleren van kritische beroepsvaardigheden, het formuleren en legitimeren van eindtermen, en daaruit volgende onderwijsontwikkeling.

Voor wat betreft de interne relaties van het systeem "onderwijsinstelling" bekritisieren Korsten en Aarts de huidige, wettelijk verankerde facultaire onderwijsorganisatie die vooral wordt gekenmerkt door de preoccupatie met de belangen van de eigen discipline. Gelet op de huidige onderwijstaken en de benodigde samenhang in het onderwijsprogramma pleiten zij voor het Angelsaksische model van de "schools" die expliciet een opleidingstaak en ruime bevoegdheden krijgen.

Giesbers wijst er op dat het feit dat er ruimte is om beleid te bepalen en uit te voeren ten aanzien van interne en externe relaties, nog niet wil zeggen dat de ruimte ook zinvol benut wordt. Er zijn hier grote verschillen tussen onderwijsinstellingen.

Het begrip "numeriek rendement" speelt een belangrijke rol in dit kader, omdat het één van de maten is aan de hand waarvan een onderwijsinstelling extern en/of intern beleid kan gaan

voeren. Prins constateert hier een ongewenste diversiteit in de berekening van de maat en stelt derhalve een uniformering voor.

Micro-niveau

Met name onderwijsontwikkeling staat centraal. De bijdragen van de auteurs sluiten aan bij de tendens relatief nieuwe onderwijsvormen in te voeren in het hoger onderwijs: thematisch onderwijs (Mulders), probleemgestuurd onderwijs (Willems) of leren (Wijnen), en computerondersteund onderwijs (Mirande, Pilot). Deze nieuwe onderwijsvormen spelen door hun kenmerken in op de tendensen om motivatie, interesse en tempo alsmede verschillen daartussen van studenten serieus te nemen, de hiervoor al gekritiseerde disciplinaire versplintering tegen te gaan, en het zelfstandig leren te bevorderen. Met name Mulders pleit als oplossing voor de disciplinaire versplintering voor een thematische ordening van de vakinhouden aan de hand van een raamwerk van centrale thema's die met toepassingsgebieden corresponderen; specialisatie vindt later plaats. Op zichzelf is zelfstandig leren (zelfsturing) nastrevenswaardig, echter het kan tot een zodanige diversiteit leiden dat de leerdoelen van de makers van probleemgestuurd onderwijs in het gedrang komen. Willems gaat in een empirisch onderzoek deze (mis)fit na en komt tot geruststellende conclusies. Wijnen betreft echter juist de stelling dat het leren ècht centraal moet staan, vandaar probleemgestuurd leren (PGL) i.p.v. probleemgestuurd onderwijs (PGO). Hij bepleit een combinatie van PGL die het mogelijk maakt te individualiseren en te differentiëren, en voortgangstoetsen, die de mate van vooruitgang van de student aangaande de expertise (functionele, langer beklijvende kennis) bewaken.

De toepassing van computertechnologie is bij uitstek een middel om in te spelen op diversiteit bij studenten. Hoewel het een tendens is, maakt Mirande duidelijk dat het HO nog slechts een "beginnend gebruiker" is; een systematische aanpak en daarbij horende beleidsontwikkeling is nauwelijks aanwezig in het HO. Wel is er een behoorlijke (succesvolle) ervaring opgedaan met toepassingsvormen die vrij nauw aansluiten bij het vigerende systeem van lesgeven (bijvoorbeeld vervanging van werkcolleges). Wat dat betreft bepleit Pilot een ontwikkeling die meer aansluit bij de hiervoor genoemde opvatting van Wijnen waarin het zelfstandig leren centraal staat, de student als kennisnavigator, en waarin met name aangepaste software uit de beroeps- en onderzoekspraktijk wordt gebruikt in het leerproces.

Commentaar

Het is misschien wat ongebruikelijk om een liber amicorum kritisch te bekijken, omdat het doel toch is een vertrekende persoon ter herinnering aan het achtergelaten werk en de daarbij horende mensen iets mee te geven en hem of haar toch ook een beetje in het zonnetje te zetten. Tegelijkertijd brengt het boek een aantal onderwerpen naar voren waarvan het waard is deze nader te beschouwen.

In de eerste plaats geeft het boek een goed beeld van de werkzaamheden van een instituut dat onderwijskundige diensten verleent aan het hoger onderwijs en aan aanpalende onderwijssystemen. De werkzaamheden bestrijken terecht alle niveaus (micro, meso en macro) van het hoger onderwijssysteem. Wel heb ik de indruk dat het micro-niveau – met name Willems, Mirande en Pilot gaan daar op in – in het boek wat ondervetegenwoordigd is.

In de tweede plaats vind ik dat het thema goed gekozen is. Het omgaan met één of andere combinatie van eenheid en verscheidenheid is naar mijn oordeel een kernprobleem in het huidige hoger onderwijs. Het inspelen op de aanwezige verschillen bij studenten is altijd al een onderwijskundig thema geweest. De deregulering en autonomie hebben daar een organisatorische en beleidsmatige dimensie aan toegevoegd. Het boek had m.i. aan waarde kunnen winnen als de redacteuren in de inleiding of de afsluiting hun visie hadden gegeven hoe een instituut van onderwijskundige dienstverlening nu om zou moeten gaan met eenheid en verscheidenheid in het hoger onderwijs. Een zelfde behoefte aan een visie kreeg ik ook bij het zien van al die niveaus en werkterreinen waarop een instituut voor onderwijskundige dienstverlening zich zoal beweegt. Ook hier dreigt versplintering en het meelopen met modieuze trends; het ene jaar gaat het om blokonderwijs, dan staat probleemgestuurd onderwijs weer centraal, en ten slotte laten

we ons weer meeslepen met het begrip "studeerbaarheid". Bovendien lijkt een zekere consistentie tussen de niveaus van belang.

In de derde plaats wil ik enige kanttekeningen plaatsen bij de verschillende bijdragen. Ik wil dat doen vanuit het perspectief van een praktiserende onderwijskundige die voor de taak staat onderwijskundige diensten te verlenen. Ik zal daarbij gebruik maken van begrippen als probleemformulering, diagnose, ontwerp, invoering en evaluatie die ik als kenmerkend zie voor de onderwijskundige dienstverlener; onderwijskundig onderzoek van welke aard dan ook staat in dienst van deze activiteiten.

ad. 1 macro-niveau

Bij de meeste auteurs vind ik de overgang van de analyse en het formuleren van het probleem naar het zoeken, bepalen en ontwerpen van mogelijke oplossingen aan de zwakke kant.

Mertens & Paardekoper beperken zich tot het signaleren van een mogelijk probleem waarbij ik me afvraag of hun analyse van de visitatie-situatie wel grondig genoeg is om tot maatregelen in deze richting te besluiten. Van der Weiden springt wel erg snel naar de conclusie (en daarmee in de richting van maatregelen) dat er nog reservecapaciteit voor bèta-studies bij meisjes is, gelet op het feit dat de uitkomsten van haar onderzoek toch vooral mogelijke verklaringen en waarschijnlijkheden betreffen (blz. 44/45). Het op zichzelf degelijke diagnostische onderzoek van Hulshof loopt het gevaar dat te nemen maatregelen te beperkt blijven. In de lijn van het proefschrift van Koppen (1991) zou ik hier pleiten voor maatregelen op het niveau van de directe begeleiding van de AIO's, het systeem als zodanig, en de regulerende principes (b.v. de in het onderzoek genoemde disciplinaire achtergrond) die de interactie tussen de voorgaande twee reguleren.

Bij de bijdrage van Vroeijssteijn heb ik toch de behoefte aan een koppeling met een wat bredere implementatie-theorie om ook andere relevante factoren in het oog te krijgen; de nu geformuleerde 15 principes (blz. 33/34) krijgen dan ook een zekere ordening.

ad. 2 meso-niveau

Een algemeen zwak punt is wellicht het ontbreken van een theoretisch kader aan de hand waarvan probleemanalyse en -diagnose en het ontwerp scherper kan worden aangestuurd.

Terecht wijzen Kleinbergen & Spruit op centrale begrippen als motieven, kosten en baten ten behoeve van de diagnose van problemen rond samenwerking tussen instellingen, een aspect van strategisch management. Hun theoretisch diagnostische kader blijft helaas echter beperkt tot noties die een meer coherent geheel behoeven. Giesbers heeft zich beperkt tot een schetsmatige aanduiding van het probleem. Hij geeft bijvoorbeeld niet aan hoe nu de mate van gewenst "eigen meesterschap" kan worden vastgesteld en hoe er op moet worden gereageerd. Hoewel zowel Korsten als Aarts een goede analyse en diagnose van de problemen met de disciplinaire vakgroepen-structuur geven, geven zij niet aan hoe nu verder te gaan. Wie neemt het initiatief tot dit beleid? Verrijt beperkt zich in feite tot een schets van het probleem en een globale diagnose. De voorwaarden voor een arbeidsmarktgeoriënteerd beleid zijn nog op een te abstract niveau geformuleerd om daaruit aanwijzingen te destilleren voor het ontwerp van instrumenten om het beoogde beleid uit te voeren. Wat dat betreft geeft Nedermeijer wèl een concreet voorbeeld van een instrument, een beroepsopleidingsprofiel (BOP), voor een arbeidsmarktgeoriënteerd beleid. De aanpak loopt echter het gevaar dat tegen de tijd dat men aan de concrete onderwijsvormgeving toe is, men weer opnieuw kan beginnen, omdat de beroepssituatie gewijzigd is.

De bijdrage van Prins (samen met anderen; zie o.a. hiervoor Bijleveld) neemt een aparte positie in: hij verschaft eenvoudigweg een uitstekend instrument dat is gebaseerd op een nauwkeurige analyse en diagnose van de berekeningsproblemen aangaande numeriek rendement. Niet meer, maar ook niet minder. Het is natuurlijk pas een echt instrument als het op correcte en uniforme wijze wordt gebruikt in een professioneel kwaliteitsbewakingsstelsel.

ad. 3 micro-niveau

Een algemeen zwak punt is wellicht dat geen van de bijdragen een leertheorie expliciteert die

kan worden gebruikt als richtinggevend kader voor onderwijsontwikkeling. Het gaat hierbij om een leertheorie die aandacht schenkt aan de ontwikkeling van extern naar intern sturen.

Op zich geeft Mulders een goede probleemanalyse en diagnose van de versplintering van het onderwijs, gevolgd door een doordacht ontwerp dat goed ingevoerd en geëvalueerd kan worden in de onderwijspraktijk, echter in de voorgestelde sequentiëring is de ontwikkeling naar interne sturing niet verdisconteerd. De empirische studie van Willems over de beïnvloeding van de leerverwachting in PGO geeft een goede terugkoppeling aan onderwijsontwikkelaars over de effecten van hun inspanningen. Ik vraag me echter wel af of zijn suggesties over de noodzaak discussies te stimuleren niet voorbij gaat aan de mogelijkheid dat een leerhandeling zich kan ontwikkelen van een verbaal niveau (b.v. in discussie) naar een mentaal niveau en daarmee naar een toenemende interne sturing (Gal'Perin; zie Haenen, 1993). Wijnen is het meest expliciet in zijn keuze voor het leren en niet voor het onderwijzen. Echter, tot nu toe is de vormgeving van de toetsen voor de meting van expertise-ontwikkeling beperkt gebleven tot stelling-toetsen waarin declaratieve kennis de belangrijkste rol speelt. Wijnen geeft helaas geen aanwijzingen voor toetsconstructie aangaande de complexe vaardigheden die in het HO aan de orde zijn.

Mirande geeft in kort bestek een goed overzicht van de stand van zaken aangaande het gebruik van COO, zij het, dat ik de financiële kant van het ontwikkelen van COO mis. Voorts vind ik dat Mirande's vooruitblik te veel uitgaat van een klassieke opvatting van leren en onderwijzen waarin de laatste domineert. Aan de andere kant is Mirande's extrapolatie ook heel realistisch; de onderwijspraktijk in het hoger onderwijs ligt zeker niet altijd wakker van een nieuwe onderwijskundige visie op leren en onderwijzen. Dit laatste is ook het zwakke punt in de bevlogen bijdrage van Pilot. Overigens ziet hij dat zelf ook wel, blijkens zijn conclusies waarin hij een minder schools hoger onderwijs en een onderwijsprogramma waarin echt mogelijkheden zijn opgenomen voor zelfstandig leren als voorwaarden ziet voor een volwaardig gebruik van computertechnologie.

Conclusie

De beide boeken kruisen elkaar letterlijk op de definitie van het rendementsbegrip. De percentages die bij toepassing van de formules worden berekend, kunnen pas worden geïnterpreteerd in het kader van een theoretisch model dat verklaringen voor de uitkomsten kan genereren. Wat dat betreft heeft Bijleveld gelijk: er ontbreekt een theoretisch kader. Dat zijn poging naar mijn oordeel voor wat betreft de onderbouwing niet helemaal gelukt is, laat onverlet dat de noodzaak blijft bestaan. Bovendien heb ik van Bijleveld's studie geleerd dat een aantal dimensies van belang is: (a) het multi-niveau karakter van de problematiek; (b) de noodzaak uit te gaan van de Nederlandse situatie; en (c) de noodzaak een werkwijze te bedenken waarin wordt geëxpliciteerd op welke gronden variabelen in een model worden opgenomen.

Deze drie punten brengen mij ook bij het boek van Hulshof c.s. Bepaalde bijdragen uit dit boek illustreren nog eens de noodzaak van een theoretisch kader. Aan de andere kant laat dit boek ook heel goed zien dat als het hoger onderwijs een zo groot mogelijk aantal studenten met kwaliteit wil afleveren, er voor wat betreft de te ontwikkelen en in te zetten middelen in verschillende niveaus moet worden gedacht. Bovendien moet er sprake zijn van consistentie in en tussen de niveaus. Een theoretisch model kan hierbij een richtinggevend kader zijn als de op te nemen factoren niet alleen een verklarende, maar vooral ook een regulerende waarde hebben. Dit laatste is alleen te bepalen als er concreet wordt geëxperimenteerd met middelen in een praktijksituatie. Het boek van Hulshof c.s. geeft daarvan enkele voorbeelden en aanzetten. Op deze wijze kunnen een theoretische en een praktische doelstelling op elkaar betrokken worden en daarmee beiden worden nagestreefd (zie o.a. Terlouw, 1993). Naar mijn oordeel moet toekomstig onderzoek en ontwikkeling in het hoger onderwijs dergelijke ondernemingen gaan opzetten. Theoretisch-praktisch geïnteresseerde onderzoekers en praktisch-theoretisch geïnteresseerde instituten voor onderwijskundige dienstverlening zijn daartoe op vele universiteiten aanwezig.

Literatuur

- Campbell, S.K. (1974). *Flaws and fallacies in statistical thinking*. Englewood Cliffs NJ: Prentice Hall.
- Carroll, J.B. (1989). The Carroll model: A 25-year retrospective and prospective view. *Educational Researcher*, 18, 26-31.
- Haenen, J.P.P. (1993). *Piotr Gal'Perin. His lifelong quest for the content of psychology*. Proefschrift Vrije Universiteit van Amsterdam.
- Jurgens, B., Prins, J. & Bijleveld, R. (1993). *Haalbaarheidsstudie jaarlijkse rendementsrapportage Hoger Onderwijs*. (Achtergrondstudies Hoger Onderwijs en Wetenschappelijk onderzoek 15). Zoetermeer: Ministerie van Onderwijs en Wetenschappen.
- Koppen, J.K. (1991). *Een kwestie van discipline. Over de externe democratisering van het wetenschappelijk onderwijs*. Proefschrift Universiteit van Amsterdam.
- Raudenbush, S.W. & Bryk, A.S. (1988). Methodological advances in analyzing the effects of schools and classrooms on student learning. In E.Z. Rothkopf (Ed.), *Review of Research in Education*, 15 (pp. 423-475). Washington: AERA.
- Terlouw, C. (1993). A model for instructional development: integration of theory and practice. In C. Terlouw (ed.), *Instructional development in higher education: theory and practice* (pp.11-30). Amsterdam: Thesis Publishers.
- Voeten, M.J.M. & Kamp, L.J.Th. van der (red.) (1990). Themanummer Longitudinaal onderzoek naar schoolloopbanen. *Tijdschrift voor Onderwijs Research*, 15, 3-88.

C. Terlouw
Universiteit Twente
Onderwijskundig Centrum
Postbus 217
7500 AE Enschede

T. Mooij (1994). *Kenmerken en effecten van methoden begrijpend/studerend lezen*. Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen (ITS).

1. Inleiding

In dit onderzoeksrapport doet Mooij verslag van een vergelijkend onderzoek naar methoden voor begrijpend lezen in het basisonderwijs. Er zijn enkele redenen om aan het rapport bijzondere aandacht te besteden. In de eerste plaats is begrijpend lezen een belangrijk vak op de basisschool. Er wordt veel tijd aan besteed en de effecten bepalen in sterke mate hoe goed leerlingen in het voorgezet onderwijs in staat zijn om zelfstandig kennis uit studieboeken te verwerven. Voorts is van belang dat het onderwijs in begrijpend lezen al enige tijd van diverse kanten kritisch gezien wordt. Zo betogen Weterings en Aarnoutse (1986) dat er veel te weinig instructie gegeven wordt. En het internationale IEA-onderzoek laat zien, dat de Nederlandse negenjarige leerlingen vergeleken met leerlingen in andere Westerse landen slecht presteren (Elley, 1992; De Glopper & Otter, 1993). Een derde reden voor de extra aandacht is dat het door Mooij gerapporteerde onderzoek min of meer model staat voor ander methode-vergelijkend onderzoek dat door het Instituut voor Onderzoek van het Onderwijs (SVO) in de afgelopen jaren geëntameerd is. Gewezen kan worden op onderzoek van Van Batenburg (1988), Edelenbos (1990), Harskamp (1988) en Hol, De Haan en Kok (te verschijnen), waarin methoden voor taalonderwijs, respectievelijk Engels, reken/wiskunde-onderwijs en aanvankelijk lezen vergeleken zijn.

2. Korte weergave van de inhoud

Voor een korte weergave van de inhoud van het rapport maak ik gebruik van de documentaire samenvatting die in het rapport te vinden is (p. 145-146):

“Het onderzoek moest antwoord geven op vragen naar de inrichting van het onderwijs in begrijpend/studerend lezen volgens bepaalde methoden, de verschillen tussen methoden in de praktijk, de mate waarin het leerkrachten lukt om het leesonderwijs gestalte te geven overeenkomstig de bedoelingen van de methode(makers), welke van de methoden het meest gebruikersvriendelijk is voor leerkrachten en leerlingen, en of leesonderwijs dat vorm gegeven wordt met behulp van een bepaalde methode tot betere leesprestaties leidt dan leesonderwijs dat met andere middelen gestalte wordt gegeven.

Het onderzoek vond plaats op 38 scholen (in een straal van 70 km rondom Nijmegen), bij in totaal 44 groepen ofwel 1042 leerlingen. Er werden vijf methoden begrijpend/studerend lezen onderzocht: a. Leeswerk; b. Lees je wijzer; c. Taal kabaal; d. Begrijpend lezen (Malmberg); e. overige methoden. Gebruik is gemaakt van een herhaald survey bij leerkrachten in groep 7 en 8 van de basisschool, observatie in de klas, analyse van de handleiding van de onderzochte methoden, analyse van leerlingenboeken, en een longitudinaal leesprestatie-onderzoek bij leerlingen in groep 7-8 van de basisschool.

Uit de surveys blijkt dat er tussen methoden een aantal verschillen wat betreft uitwerking in de onderwijspraktijk bestaat. Tussen groep 7 en groep 8 worden (bij dezelfde leerlingen) nauwelijks verschillen aangetroffen. Het is niet haalbaar gebleken om via bestudering en scoring van de schriftelijke documentatie (handleidingen, leerlingenboeken) de kenmerken van de methoden op een eenduidige wijze vast te stellen. Nodig is een meer gesystematiseerd en eenduidig leerstof-begrippenkader. De leerkrachten ervaren ten aanzien van elke methode voor- en nadelen. Uit het longitudinaal effectenonderzoek blijkt dat Leeswerk, Lees je wijzer, Taal kabaal en Begrijpend lezen in vergelijking met ‘overige methoden’ géén effect hebben op de ontwikkeling in taalscore (Cito-score) in de groepen 7 en 8 van het basisonderwijs. Vergelijkenderwijs heeft Lees je wijzer een positief effect op de ontwikkeling van de Cito-rekenscore en daarnaast negatieve effecten op de ontwikkeling in metacognitief gedrag voorafgaand aan het lezen en na het lezen. Leeswerk heeft in dit opzicht geen enkel effect. Taal kabaal heeft positieve effecten op de ontwikkeling in rekenen, informatieverwerking en de Cito-totaalscore, maar negatieve effecten op de ontwikkeling van de metacognitieve gedragsscores vóór, tijdens

en na het lezen. Begrijpend lezen heeft géén effect op de ontwikkeling in taal, rekenen, informatieverwerking, of metacognitief leesgedrag. Er is relatief gezien géén effect van een begrijpend/studerend leesmethode op de leesontwikkeling van slechte of van goede lezers. In een enkel geval (Leeswerk) wordt de ontwikkeling van goede lezers geremd. Naarmate een groep hoger scoort op de begin-leesprestatie, maakt de groep een méér positieve ontwikkeling door wat betreft het lezen, rekenen en informatieverwerking." (einde citaat)

3. Bespreking

Ik verdeel mijn opmerkingen over drie hoofdonderdelen: de vraagstelling, de onderzoeksopzet en de uitvoering.

Met de gekozen *vraagstelling* ben ik niet helemaal gelukkig. Van de vier gestelde onderzoeksvragen zijn er maar twee die bij nadere beschouwing als relevante vragen overeind blijven. Relevant is de eerste vraag, naar mogelijke verschillen tussen scholen wat betreft de vormgeving van het leesonderwijs, samenhangend met de methode waarvan gebruik gemaakt wordt. De tweede vraag – of die vormgeving in overeenstemming is met de bedoelingen van de methodenschrijvers – is in wezen niet zo belangrijk, omdat die bedoelingen op zichzelf geen relevante toetssteen vormen. Methodenschrijvers laten leerkrachten namelijk een grote vrijheid in de manier waarop ze hun methode willen gebruiken. Dat bevordert de verkoopcijfers. Belangrijker was de vraag geweest of de methoden een consistente opbouw hebben: is de gekozen vakdidactische en onderwijskundige uitwerking consistent met de doelstellingen die men bij de leerlingen wil bereiken? Ook van de derde onderzoeksvraag – naar de gebruiksvriendelijkheid voor leerkrachten en voor leerlingen – acht ik het belang ondergeschikt. Gebruiksvriendelijkheid moge misschien voor computerprogramma's een belangrijk criterium zijn, bij onderwijsmethoden gaat het allereerst toch om de effectiviteit.

De vierde vraag – naar verschillen in effectiviteit ten aanzien van de vorderingen van de leerlingen – is de hamvraag. De mogelijkheid dat leerlingen met de ene methode sneller vorderingen maken dan met de andere, vormt de rationale van methodevergelijkend onderzoek. Overigens maak ik me over de grootte van eventuele verschillen in effectiviteit weinig illusies. In de eerste plaats kennen we inmiddels de resultaten van PPO, die suggereren dat er tussen methoden nauwelijks verschillen in effectiviteit bestaan. Dat er hoogstens kleine verschillen te verwachten zijn, valt ten tweede te baseren op de resultaten van het schooleffectiviteitsonderzoek, waaruit blijkt dat verschillen tussen scholen in de prestaties van hun leerlingen overwegend klein zijn en sterk samenhangen met verschillen in de leerlingenpopulatie, met name met de proportie leerlingen in achterstandssituaties (Blok & Hoeksma, 1993).

Bij de vierde vraag moet overigens wel worden opgemerkt, dat de vraag naar eventuele verschillen in effectiviteit slechts zinvol te onderzoeken is in combinatie met een voorafgaande analyse van de doelstellingen van de methoden en van de door de methoden gevolgde aanpak. Een analyse van de doelstellingen is nodig om de criteria te kiezen waarop de methoden vergeleken worden. Een vergelijking van methoden is namelijk slechts zinvol voor alleen dié doelstellingen die de methoden met elkaar gemeen hebben. Een analyse van de aanpak (waaronder de vakdidactische en leerpsychologische uitgangspunten, de uitgezette leerlijnen en de consistentie daarvan, de manier waarop gedifferentieerd wordt, de voorstellen voor het bijhouden van de vorderingen en de gekozen organisatieprincipes) is nodig om de uitkomst van de effectvergelijking te kunnen interpreteren. Helaas moet worden vastgesteld dat in Mooij's vraagstelling de vraag naar de overeenkomsten en verschillen in doelstellingen en werkwijzen ontbreekt.

Met dit laatste punt belanden we bij de door Mooij gekozen *onderzoeksopzet*. Aan die opzet zitten twee problemen vast. De eerste is het ontbreken van een analyse van wat de methoden qua doelstellingen en aanpak te bieden hebben. Dit gemis is vanzelfsprekend de consequentie van de incomplete vraagstelling. Maar het heeft er toch de schijn van, dat de onderzoeker zich wel degelijk gerealiseerd heeft dat een methodenanalyse niet gemist kan worden. De manier waarop deze evenwel is uitgevoerd, is erg oppervlakkig. Op dit punt kom ik bij de bespreking van de onderzoeksuitvoering nog terug.

Het tweede probleem met de onderzoeksopzet heeft te maken met de duur van de periode waarin de vorderingen van de leerlingen bepaald zijn. Ten behoeve van het effectonderzoek is gebruik gemaakt van een voormeting in september/oktober in groep 7 en een nameting in februari in groep 8. Dit deel van het onderzoek heeft derhalve betrekking op slechts 15 maanden lesonderwijs. Mooij geeft aan dat de condities van zijn opdrachtgever (SVO) hem geen andere keuze lieten. Maar er moet toch worden opgemerkt dat onder zulke condities onderzoek naar effecten van methoden geen recht doet aan de inspanningen van methodemakers. De onderzochte leesmethoden hebben een bereik vanaf groep 4 tot en met groep 8, een periode derhalve van 50 maanden lesonderwijs. De onderzochte periode betreft daarvan ongeveer 30 procent. Zodoende scheidt Mooij een selectief beeld van de verdiensten en tekortkomingen van de methoden. Van belang is verder dat vorderingen wat betreft begrijpend lezen tamelijk traag zijn, zoals bijvoorbeeld blijkt uit de Clib-toetsen die onderdeel vormen van het leerlingvolgsysteem van het Cito (Staphorsius, 1992). Om op een termijn van 15 maanden verschillen te kunnen aantonen, dient men dus over bijzonder gevoelige toetsen te beschikken.

Aangaande de *uitvoering* van het onderzoek beperk ik me tot vijf punten van kritiek. Het eerste punt is de manier waarop de methodenanalyse is uitgevoerd. De analyse had zowel betrekking op de handleidingen voor de leerkracht als op de leerlingenboeken, in beide gevallen overigens alleen op het materiaal dat voor groep 7 en 8 bedoeld is. Voor de analyse van de handleidingen is een coderingsschema ontworpen, bestaande uit negen aandachtspunten, dat bij de daadwerkelijke scoring onbruikbaar bleek. Dat brengt de onderzoeker tot de volgende conclusie: 'Op basis van de methode-handleidingen is het niet goed mogelijk om precies te achterhalen wat de kenmerken van een methode zijn' (p. 47). Bij de scoring van de leerlingenboeken is dezelfde werkwijze gevolgd, met de even teleurstellende conclusie dat de scoring binnen het beschikbare onderzoeksbudget niet uitvoerbaar is gebleken. De leesmethoden zijn voor de onderzoeker derhalve een ondoorgrondelijke 'black box' gebleven. Daarmee is in feite de bodem uit het onderzoek geslagen. Wat heeft het immers voor zin methoden op hun effecten te vergelijken, als men niet kan beschrijven waarin die methoden van elkaar verschillen? Gelukkig weten we uit andere bronnen dat er tussen de methoden wel degelijk verschillen zijn (Laarhoven, Van Rossum & Naber, 1987; Van Rossum, 1990). Het verbaast mij dat Mooij deze bronnen wel noemt, maar ze niet ten goede laat komen aan zijn onderzoek.

Het tweede te bespreken punt van uitvoering is de manier waarop de steekproef van deelnemende scholen tot stand gekomen is. De onderzoeker heeft 700 basisscholen in de omgeving van Nijmegen benaderd met de vraag of zij bereid waren aan het onderzoek deel te nemen. Van deze scholen hebben er zich daartoe 114 in eerste instantie bereid verklaard. Om redenen die in het rapport niet worden toegelicht, hebben er uiteindelijk slechts 38 scholen aan het onderzoek deelgenomen. Dat is vijf procent van de oorspronkelijk benaderde groep. Omdat er geen non-responsonderzoek verricht is en omdat de onderzoeker geen poging doet de representativiteit van zijn deelnemersgroep te toetsen, kan men niet anders dan een vraagteken zetten bij de generaliseerbaarheid van de uitkomsten. Toch had het niet zo ver hoeven komen. In de loop van het onderzoek zijn voldoende gegevens verzameld (scores op de Eindtoets Basisonderwijs van het Cito; scores op de Raven, de achterstandscategorieën waaruit de leerlingen afkomstig zijn) om althans post hoc een uitspraak te kunnen doen over de representativiteit. Mooij laat hier dus een kans liggen. Maar afgezien daarvan is een uiteindelijke respons van vijf procent natuurlijk iets wat een nadere analyse verdient. Hoe komt het dat maar zo weinig scholen tot deelname over te halen waren? Zou dat te maken kunnen hebben met het gepercipieerde nut van het onderzoek?

De verrichte observaties – terwijl de leerlingen in groep 7 zaten, is in elk van de 44 groepen één maal een leesles geobserveerd door één observator – vormen het volgende punt van bespreking. Gebruik is gemaakt van een semi-gestructureerde observatie aan de hand van acht aandachtspunten. Uit de in het rapport gegeven informatie valt niet te reconstrueren hoe de observaties zijn uitgevoerd. Als voorbeeld noem ik de als eerste vermelde variabele 'afstemming leesprocessen op leesniveau'. Deze variabele is geoperationaliseerd als een vraag met twee antwoorden: 'Afstemming van de leesprocessen op het leesniveau van elke leerling

(0 = nee; 1 = ja)'. Ik kan me eenvoudigweg geen voorstelling maken van de betekenis van de op deze manier verkregen ja/nee antwoorden. Uit de resultaten blijkt dat de gemiddelde score op deze variabele 0,2 bedraagt. Betekent dit dat in zeven à tien van de 44 geobserveerde leeslessen de leesprocessen afgestemd zijn op het niveau van elke leerling? Het is een tekort dat Mooij aan de betrouwbaarheid en de validiteit van de via observaties verkregen gegevens zo weinig aandacht schenkt. Het maakt deze gegevens ontoegankelijk voor interpretatie. Misschien is de enige winst van dit gedeelte van het onderzoek wel dat gebleken is, dat in twee groepen met een andere methode gewerkt werd dan de school had opgegeven. Deze groepen zijn alsnog aan een andere onderzoeksconditie toegevoegd.

De gekozen operationalisaties op het niveau van de leerlingen vormen het vierde aandachtspunt. Wat betreft het leesniveau is gebruik gemaakt van drie instrumenten. Als beginmeting is een toets van het Cito gekozen (Begrijpend lezen, eind leerjaar 4) en een nieuw ontwikkelde vragenlijst 'metacognitieve leesvaardigheden'. Bij de eindmeting is dezelfde metacognitieve vragenlijst gebruikt, naast de score op de toets Informatieverwerking die onderdeel is van de Eindtoets Basisonderwijs van het Cito. De metacognitieve vragenlijst is een belangwekkend instrument, waarmee leerlingen van 56 leesstrategieën (verdeeld over strategieën voor, tijdens en na het lezen) moeten aangeven of zij deze toepassen. Omdat alle items positief geformuleerd zijn, kunnen zich antwoordtendenties voordoen, bijvoorbeeld onder invloed van sociale wenselijkheid. Het is jammer dat over de validiteit van de metacognitieve vragenlijst geen nadere gegevens verstrekt worden. De toets Informatieverwerking is een 'omnibustoets' voor het kunnen omgaan met schriftelijke informatiebronnen. De toets bestrijkt duidelijk een breder domein dan begrijpend/studerend lezen. Daardoor is de toets onvoldoende gevoelig om de eventuele kleine verschillen te detecteren. Kritische vragen zijn er ook bij de twee intelligentietests die aan het begin zijn afgenomen en bij de analyse als covariaat zijn gebruikt. De validiteit van de 'Raven progressive matrices' is blijkens de 'Documentatie van tests en test-research in Nederland' (Evers, Van Vliet-Mulder & Ter Laak, 1992) onvoldoende. De andere gekozen test is de Duitse intelligentietoets uit 1968 (Prüfsystem für Schul- und Bildungsberatung), die naar ik veronderstel vertaald is, maar waarover geen psychometrische gegevens bekend zijn noch door Mooij worden verstrekt. Zelfs een eenvoudige correlatie tussen beide tests wordt de lezer onthouden. Onverklaard blijft waarom de keuze gevallen is op niet-verbale intelligentietests, terwijl algemeen bekend is dat verbale intelligentietests een betere voorspeller voor schoolprestaties vormen dan niet-verbale tests. Al met al maken de gekozen operationalisaties geen overtuigende indruk.

Tenslotte wil ik twee opmerkingen maken over de gegevensanalyses en de manier waarop de uitkomsten vermeld zijn. Hoofdmoot van de gegevensanalyses vormen variantie-, regressie- en meerniveau-analyses. Het valt te betreuren dat de rapportering van de uitkomsten niet zo volledig is. Zo ontbreekt bij de variantie-analyses een beschrijving van de gemiddelden en standaarddeviaties op de eindtoetsen. Daardoor is het niet mogelijk een indruk te krijgen van de omvang van de verschillen. Jammer is ook dat het accent gelegd wordt op bilaterale vergelijkingen van de vier afzonderlijke methoden met de restgroep 'overige methoden', waarin negen methoden vertegenwoordigd zijn. Die restgroep is bepaald niet het eenduidige criterium dat men zich bij een methodevergelijking zou wensen. Deze bilaterale vergelijkingen zijn daarom niet goed te interpreteren.

4. Conclusies

Mijn eendoordeel over het onderzoek en over de verslaggeving daarvan is negatief. De vraagstelling is deels weinig relevant, deels incompleet. De ontwikkelde opzet is niet geschikt voor een valide vergelijking van methoden. En in de uitvoering zijn op kernpunten tekortkomingen aan te wijzen.

Ter verdediging van Mooij zij erop gewezen dat de SVO in haar rol als opdrachtgever heeft verklaard, dat het onderhavige rapport aan haar minimum kwaliteitseisen voldoet. Ook de aanvrager van het onderzoek, het Christelijk Pedagogisch Studiecentrum (CPS), is naar ik vermoed tevreden met het verrichte werk. Weliswaar zijn de door het CPS ingediende onder-

zoeksvragen door de SVO slechts gedeeltelijk overgenomen, maar dat valt de onderzoeker niet te verwijten. Die heeft op zijn beurt tenminste toch een deel van de oorspronkelijk aangedragen onderzoeksvragen naar beste vermogen beantwoord. En de wijze waarop hij dat gedaan heeft, heeft van SVO het stempel van goedkeuring verkregen. Of de leerkrachten die op zoek zijn naar een nieuwe leesmethode – namens wie het CPS zich als aanvrager opgeworpen heeft – met het onderzoek tevreden zijn, waag ik te betwijfelen. Het grootste deel van de onderzochte methoden is inmiddels vervangen door de volgende generatie. En een belangwekkende nieuwkomer als 'Wie dit leest' is buiten beschouwing gebleven.

Ter verdediging van de onderzoeker zij er tevens op gewezen dat de condities waaronder deze zijn onderzoek moest verrichten – en die hem door de SVO zijn toebedeeld – een behoorlijke methodenvergelijking bij voorbaat verhinderd hebben. De opgedragen onderzoeksvragen zijn tevoren niet goed genoeg doordacht. Het is maar de vraag of men van een onderzoeker mag verwachten dat deze met de opdrachtgever onderhandelt om de vraagstelling te verbeteren. A.D. de Groot liet in de bundel ter gelegenheid van het afscheid van professor Geensen over zulk onderhandelen het volgende optekenen: 'Onderzoekers durven en kunnen dat niet. Ze zijn, en dat is begrijpelijk, dolblij als ze een opdracht krijgen. Ze hebben niet de guts en de status om tegen de opdrachtgever te zeggen: "Je moet dat anders doen, zo leveren wij ondermaats werk, we moeten samen de vraagstelling herformuleren."' (Haenen, e.a., 1984, p. 58). Niet alleen de vraagstelling, ook de korte termijn waarop SVO de onderzoeksresultaten op tafel wenste te hebben, stond kwalitatief hoogwaardig onderzoek in de weg. In twee jaar tijd een behoorlijk onderzoek uitvoeren naar de effecten van methoden die een looptijd van vijf jaar beslaan, is nu eenmaal niemand gegeven.

Methodenvergelijkend onderzoek is niet bij voorbaat zonder zin. Maar er moeten wel enkele condities aan verbonden worden. Ik noem er drie. De voornaamste is dat er serieuze aanwijzingen moeten bestaan dat methoden in hun effecten op leerlingen verschillen. Methodevergelijkend onderzoek hoort hypothese-toetsend onderzoek te zijn, geen exploratief onderzoek. Die hypothesen – dat is de tweede conditie – mogen niet uit de lucht komen vallen. Ze moeten verankerd zijn in een degelijke analyse van de vakdidactische en onderwijskundige kwaliteiten van de methoden. Zo'n analyse – derde conditie – moet voorafgaan aan de keuze van variabelen en operationalisaties. Met het noemen van deze drie condities heb ik tevens samengevat, waarom het onderzoek van Mooij zo weinig heeft opgeleverd en – dat moet hem worden toegegeven – dankzij SVO ook zo weinig kon opleveren.

5. Literatuur

- Batenburg, T.A. van (1988). *Een evaluatie van taalmethoden*. Groningen: RION.
- Blok, H. & Hoeksma, J.B. (1993). De stabiliteit van het schooleffect in de tijd: een analyse op basis van vijf jaar Eindtoets Basisonderwijs van het Cito. *Tijdschrift voor Onderwijsresearch*, 18, 331-342.
- Edelenbos, P. (1990). *Leergangen voor Engels in het basisonderwijs vergeleken*. Groningen: RION.
- Elley, W.B. (1992). *How in the world do students read?* Hamburg: The International Association for the Evaluation of Educational Achievement.
- Evers, A., Vliet-Mulder, J.C. van & Laak, T. ter (1992). *Documentatie van tests en testresearch in Nederland*. Assen/Maastricht: Van Gorcum.
- Glopper, K. de & Otter, M.E. (1993). *De Nederlandse leesprestaties in internationaal perspectief*. Amsterdam: SCO.
- Haenen, J.P.P., Hermans, J.J., Westhoff, G.J. & Wubbels, Th. (red.) (1984). *Onderwijswetenschap & onderwijspraktijk: een LAT-relatie?* Purmerend: Muusses.
- Harskamp, E.G. (1988). *Rekenmethoden op de proef gesteld*. Groningen: RION.
- Hol, G.G.J.M., Haan, M. de & Kok, W.A.M. (te verschijnen). *De effectiviteit van methodes voor aanvankelijk leesonderwijs*. Utrecht: ISOR.
- Laarhoven, J., Rossum, C. van & Naber, I. (1987). *Analyse van methoden voor begrijpend en studerend lezen*. 's-Hertogenbosch: KPC.
- Rossum, C. van (1990). *Analyse van methoden voor begrijpend en studerend lezen 2*. 's-Hertogenbosch: KPC.
- Staphorsius, G. (1992). *Clib-toetsen; handleiding*. Cito: Arnhem.

Weterings, A.C.E.M. & Aarnoutse, C.A.J. (1986). De praktijk van het onderwijs in begrijpend lezen.
Pedagogische Studiën, 63, 387-400.

H. Blok

SCO-Kohnstamm Instituut voor Onderzoek van Opvoeding en Onderwijs
van de Universiteit van Amsterdam



BULLETIN

ORGAAN VAN DE VERENIGING VOOR ONDERWIJSRESEARCH

Redactie-adres

Dr. P.J.J. Stijnen, Open Universiteit,
Postbus 2960, 6401 DL Heerlen,
tel. 045-762293

Jrg. 18, nr. 3

1994

Inhoudsopgave

Verslag VOR-ledenvergadering	382
Een nieuwe Europese vereniging voor onderwijsonderzoek	384

Verslag VOR-ledenvergadering 25 mei 1994: 12.45 - 13.15 uur te Utrecht (Jaarbeursgebouw tijdens ORD'94)

Aanwezig: Leden: Th. Bergen, M. Berger, H.J.M. van Berkel, M. Lunenberg, P. Meijer, H. Peters, P. Slegers, R. Voeten, A. van Wieringen
Bestuursleden: J. Ax, N. Lagerweij (voorzitter), F.J.J. van Bussel (penningmeester), W.Th.J.G. Hoeben, M. Mulder, J. Scheerens, P.J.J. Stijnen (secretaris)

1. Opening

De voorzitter opent de vergadering om 12.45 uur. Bericht van verhindering is ontvangen van: J.J. Beishuizen, P. van den Broek, J. Claessen, J. Dronkers, H. van Hout, A.J. Kallenberg, L. de Klerk, dhr. Klomp, J. Pieters, J. Slavenburg, J. Vastenburger, H. de Wolf.

2. Mededelingen

- De voorzitter deelt mede dat er het laatste jaar een aanzienlijke uitbreiding van het ledenbestand heeft plaatsgevonden
- De voorzitter deelt mede dat de twee ere-leden, prof. dr. A.D. de Groot en prof. dr. E. Warries, hun benoeming gisteravond met genoegen hebben aanvaard.
- De heer Hoeben deelt mede dat tijdens deze ORD een vergadering plaatsvindt van de divisie Curriculum. Er zal een nieuwe voorzitter worden gekozen: dr. J van den Akker van de UT. De heer J van den Akker zal daarna in zijn plaats toetreden tot het VOR-bestuur. De voorzitter dank de heer Hoeben voor zijn actieve bijdrage aan de activiteiten van het VOR-bestuur.

3. Verslag van de VOR-ledenvergadering van 16 december 1993

Het verslag wordt onder dankzegging aan de secretaris ongewijzigd vastgesteld. Naar aanleiding van het verslag wordt opgemerkt dat het tijdstip van de algemene ledenvergadering tijdens de ORD'95 in Groningen zo moet worden gekozen dat zo veel mogelijk leden de bijeenkomst kunnen bijwonen.

De voorzitter verzoekt de heer Hoeben - die betrokken zal zijn bij de organisatie van de ORD in Groningen - hier volgend jaar op toe te zien. Ook de secretaris zal dit punt bewaken.

4. Vaststelling jaarrekening 1993

De jaarrekening wordt ongewijzigd vastgesteld.

5. Rapport kascommissie en decharge kascommissie

De heer Voeten rapporteert namens de commissie dat de penningmeester uitstekend werk heeft verricht en dat de financiële situatie van de vereniging goed is. De financiële gegevens zijn door de commissie in orde bevonden. Onder dankzegging van de voorzitter wordt de kascommissie (de heren Voeten, Berger en Claessen) gedechargeerd. Vanuit de vergadering wordt gesuggered een deel van het bedrag op de girorekening tegen een hogere rentepercentage uit te zetten.

6. Instelling nieuwe kascommissie

De heren Voeten, Berger en Claessen worden bereid gevonden de nieuwe kascommissie te vormen. Voor de heer Voeten zal dit het laatste jaar zijn.

7. Begroting 1995

De begroting wordt ongewijzigd vastgesteld. Naar aanleiding van een opmerking van de heer Van Wieringen wordt afgesproken in het VOR-bestuur nog eens na te gaan of het mogelijk is

om voor VOR-leden een reductie te verkrijgen op Nederlandstalige tijdschriften. Hierdoor zou het toch mogelijk blijven het subsidie aan TOR te handhaven terwijl voor leden die mogelijk meer interesse hebben in anderssoortige tijdschriften een interessante kortingsmogelijkheid ontstaat.

8. Communicatie met de leden en divisies

De vergadering kan accoord gaan met een vorm van communicatie door middel van een los ingeschoven inlegvel in TOR. Daarnaast blijven de nu gehanteerde communicatievormen binnen de divisies vanzelfsprekend gehandhaafd.

9. Wat verder ter tafel komt

Er komt verder niets ter tafel.

10. Rondvraag

Van de rondvraag wordt geen gebruik gemaakt.

11. Sluiting

Niets meer aan de orde zijnde sluit de voorzitter onder dankzegging om 13 . 1 5 uur de vergadering.

N.B.: De volgende Algemene Ledenvergadering vindt plaats tijdens de ORD 1995 in Groningen.

Een nieuwe Europese vereniging voor onderwijsonderzoek

Op 18 juni 1994 is in Straatsburg de EERA opgericht. Het betreft de 'European Educational Research Association', een vereniging van nationale verenigingen en/of instituten op het gebied van onderwijsonderzoek. De bedoeling is het stimuleren van kwalitatief goed onderwijsonderzoek in Europa en het bevorderen van een vruchtbare afstemming tussen onderwijsonderzoek, onderwijsbeleid en -praktijk.

De nieuwe vereniging brengt een groot aantal nationale onderzoeksverenigingen en instituten bijeen uit Oost- en West-Europa en zal gaan functioneren als het belangrijkste orgaan voor Europese onderwijsonderzoekers. De Nederlandse Vereniging voor Onderwijsresearch (VOR) heeft zich van meet af aan ingespannen om een dergelijke Europese organisatie in het leven te roepen.

Tijdens de openingsceremonie zette Dr. Armin Gretler uiteen hoe EERA ontstaan is na uitvoerige discussies tussen onderzoekers en andere geïnteresseerde groepen uit Europa. Hij sprak van een uniek voorbeeld van internationale samenwerking. Prof. James Calderhead, de eerste voorzitter van de vereniging, verwees naar de rijke en gevarieerde Europese geschiedenis van opvoedkundige en onderwijskundige idealen, theorieën en onderwijsexperimenten. Die hebben belangrijk bijgedragen aan de snelle sociale veranderingen binnen Europa. Onderzoek heeft een rol gespeeld bij het ontwikkelen van alternatieve opvattingen die inspirerend gewerkt hebben op onderwijsbeleid en -praktijk. Bovendien is onderzoek van belang voor het volgen en evalueren van veranderingen in het onderwijs.

De komst van EERA werd van harte verwelkomd door Michael Vorbeck van de Raad van Europa en door Pierre Laderriere van de OECD. De laatste benadrukte het belang van een Europese organisatie op het gebied van onderwijsonderzoek die zich richt op uitbreiding van internationale samenwerking en op het bevorderen van onderzoek van hoge kwaliteit dat onafhankelijk is van politieke invloed.

De EERA heeft voor de komende jaren een aantal activiteiten in voorbereiding gericht op het in brede kring bespreken van de relevantie van onderwijsonderzoek en -evaluatie. In september 1995 wordt een grote conferentie georganiseerd in Bath, Engeland.

De leden van huidige bestuur zijn:

Prof. Martine Abdallah-Pretceille (Frankrijk)
Prof. James Calderhead (Engeland)
Prof. Arturo de la Orden (Spanje)
Dr. Peter Döbrich (Duitsland)
Prof. Mats Ekholm (Zweden)

Prof. Armin Gretler (Zwitserland)
Prof. Nijs Lagerwei; (Nederland)
Prof. Bartolo Paiva Campos (Portugal)
Prof. Jan Prucha (Tsjechie)
Dr. Marian Setinc (Slovenie)

Voor informatie is in Nederland de contactpersoon:

Prof.dr N.A.J.Lagerweij, Vakgroep onderwijskunde, Universiteit Utrecht, Postbus 80140, 3508 TC Utrecht. tel. 030-534940.

Artikelen

Gelijke kansen in het onderwijs? Een vergelijking van vier cohorten leerlingen in hun overgang naar het voortgezet onderwijs <i>B.F.M. Bakker en P.G. J. Cremers</i>	191
Affectieve variabelen: Hun effecten op leermotivatie en leerprestatie <i>M. Boekaerts en G. Seegers</i>	297
Stress, coping, en prestatie- (motivatie) <i>M. Boekaerts en G. Seegers</i>	356
School- en sekseverschillen in vakkenkeuzen in het VWO <i>R.J. Bosker en H.P.J.M Dekkers</i>	214
School Effects on Drop Out. A multi-level logistic approach to assessing school-level correlates of drop out of ethnic minorities <i>R.K. Bosker and W.H. A. Hofman</i>	50
Effective Instruction as a Basis for Effective Education in Schools <i>B.P.M Creemers</i>	3
A Model for Explaining Individual Differences between Students in Intended Effort on Curricular Tasks <i>M.J. Crombach, Marinus Voeten and M. Boekaerts</i>	301
Inleiding: Ongelijkheid in onderwijskansen <i>J. Dronkers, JA.M. Heijke en R.K.W. van der Velden</i>	189
Taakkenmerken die bijdragen aan de moeilijkheid van functionele leestaken <i>W.C de Geus en P. Reitsma</i>	85
Effecten van streefniveaus in het basisonderwijs op schoolloopbanen in het voortgezet onderwijs <i>A.A. van der Hoeven-van Doornum</i>	204
Overadvisering en de schoolloopbanen van migrantenleerlingen en autochtone leerlingen in het voortgezet onderwijs <i>M. Koeslag en J. Dronkers</i>	240
Effecten van schoolkenmerken en steldidactiek op functionele stelvaardigheid Nederlands in het MAVO <i>H. Kuhlemeier</i>	161
Schoolprestaties van allochtone leerlingen: de invloed van etnische herkomst <i>E.J. Lataheru en M.G.P. Hessels</i>	227
Teaching Effectiveness: An Overview of Studies <i>J. Lowyck</i>	17
Tijdsattitudes, faalangst en studiemotivatie <i>M.-A. Moreas en W. Lens</i>	318

IV

Handelingscontrolevaardigheden in het onderwijs <i>R. Otten, M. Boekaerts en G. Seegers</i>	343
Toekomstperspectief als voorspeller van inzet voor school <i>T. Peetsma</i>	331
Leereffecten van stages in sociaal-wetenschappelijke studierichtingen <i>R.F. Poell en M.J.M. Voeten</i>	105
The School-level Context of Instructional Effectiveness: A Comparison between School Effectiveness and Restructuring Models <i>J. Scheerens</i>	26
Diagnose van cognities als interventie in de training van leraren <i>H.H. Tillema</i>	118
Van de redactie	1
Schoolloopbanen van Molukse leerlingen in het voortgezet onderwijs <i>A. Veen en M. Robijns</i>	143
Ontwikkeling van beginnende lees- en spellingvaardigheid <i>L. Verhoeven en P. Gillijns</i>	259
Eindexamens in de tijd vergeleken: het einde van de wet Posthumus <i>H.D. Webbink</i>	135
School and Instruction Characteristics and Pupils' Arithmetic Achievements at the End of Primary Education <i>G. van der Werf</i>	39
A Coupled Process of Conceptualizing a Model of School Effectiveness and Developing an Indicator System for Monitoring Effectiveness <i>R. Zuzovsky and M. Aitkin</i>	65
Boekbesprekingen	
M. Boekaerts en P.R.J. Simons: <i>Leren en instructie: psychologie van de leerling en het leerproces</i> (P.C. van der Sijde)	184
R.J. Bijleveld: <i>Numeriek rendement en studiestaking. Een theoretische analyse van factoren die samenhangen met rendement en studiestaking in het wetenschappelijke onderwijs</i> (C. Terlouw)	366
J.W. Holleman: <i>Over studielast en studeerbaarheid</i> (J. Wolfhagen)	182
M.J.F. Hulshof, J.A.A.M. Bevers en J.M.H.M. Willems (red) <i>Universiteit & Diversiteit</i> . (C. Terlouw)	366
T. Mooij: <i>Kenmerken en effecten van methoden begrijpend/studerend lezen</i> (H. Blok)	375

W. Tomic en P. Span (red): <i>Onderwijspsychologie: beïnvloeding, verloop en resultaten van leerprocessen</i> (P.C. van der Sijde)	184
B. Vreeburg: <i>Identiteit en het verschil. Levensbeschouwelijke vorming en het Nederlands voortgezet onderwijs</i> (A. de Jong)	280

Redactionele medewerkers

In 1994 werd redactionele medewerking verleend door:

C.A.J. Aarnoutse	Katholieke Universiteit Nijmegen
H.J.M. van Berkel	Rijksuniversiteit Limburg
R.J. Bosker	Universiteit Twente
H.P. Brandsma	OCTO Enschede
H.P.J.M. Dekkers	ITS Nijmegen
S. Dijkstra	Universiteit Twente
P. Edelenbos	GION Groningen
W.J.M. van Esch	ITS Nijmegen
K. de Glopper	SCO-Kohnstamm Instituut Amsterdam
D.N.M. de Gruijter	Bureau Onderzoek van Onderwijs Leiden
W.H.A. Hofman	RISBO Rotterdam
J.F.M.J. van Hout	Universiteit van Amsterdam
G.G.H. Jansen	Rijksuniversiteit Groningen
U. de Jong	SCO-Kohnstamm Instituut Amsterdam
G. Kanselaar	Universiteit Utrecht
G. Lang	Rijksuniversiteit Groningen
D.A.V. van der Leij	Vrije Universiteit Amsterdam
G.J. Mellenbergh	Universiteit van Amsterdam
T. Mooij	ITS Nijmegen
J.M. Pieters	Universiteit Twente
G.J. Reezigt	GION Groningen
P. Reitsma	Paedologisch Instituut Duivendrecht
K. Reynders	Rijksuniversiteit Groningen
E. Roede	SCO-Kohnstamm Instituut Amsterdam
J. Roeleveld	SCO-Kohnstamm Instituut Amsterdam
P.F. Sanders	Cito Arnhem
T.A.B. Snijders	Rijksuniversiteit Groningen
J.D.H.M. Vermunt	Katholieke Universiteit Brabant
L. Verschaffel	Katholieke Universiteit Leuven
M.C. van der Wende	Nuffic Den Haag
M.P.C. van der Werf	GION Groningen