

## VU Research Portal

### **Implementatie en Effecten van Interne Differentiatie. Een empirisch, vergelijkend onderzoek naar de realisering en effecten van interne differentiatie in heterogene groepen in de eerste fase voortgezet onderwijs bij wiskunde.**

Terwel, J.; Herfs, P.G.P.; Dekker, R.; Akkermans, W.

1988

[Link to publication in VU Research Portal](#)

#### ***citation for published version (APA)***

Terwel, J., Herfs, P. G. P., Dekker, R., & Akkermans, W. (1988). *Implementatie en Effecten van Interne Differentiatie. Een empirisch, vergelijkend onderzoek naar de realisering en effecten van interne differentiatie in heterogene groepen in de eerste fase voortgezet onderwijs bij wiskunde.* (Selectareeks). Stichting voor Onderzoek van het Onderwijs (SVO).

#### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

#### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

# IMPLEMENTATIE EN EFFECTEN VAN INTERNE DIFFERENTIATIE

---

*J. Terwel*

*P. Herfs*

*R. Dekker*

*W. Akkermans*

---







het Instituut voor  
Onderzoek van het Onderwijs

SVO

De SVO is het onafhankelijke, landelijke, instituut voor de bevordering van het wetenschappelijk onderzoek van het onderwijs waaronder ook de evaluatie van vernieuwingsprojecten. Middelen zijn o.a. de financiering, begeleiding en programmering van het onderzoekswerk. De onderzoeksactiviteiten worden in uitvoering gegeven bij gespecialiseerde researchinstituten in het land.

Een van de hoofdtaken van de SVO is ook het publiek maken van de onderzoeksresultaten. Dat doet de SVO door het verspreiden van persberichten, brochures en het maandblad Didaktief. En voorts door middel van de uitgave van een wetenschappelijke boekenserie, de SVO-reeks, en een semi-wetenschappelijke boekenserie, de Selectareeks.

Als u meer wilt weten over een bepaald onderzoek of meer informatie wenst, bel of schrijf dan naar de afdeling Voorlichting van de SVO, Sweelinckplein 14, 2517 GK Den Haag, telefoon 070-469679.

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

ISBN 90-6472-122-X

© het Instituut voor Onderzoek van het Onderwijs S.V.O. 's-Gravenhage 1988

No part of this book may be reproduced in any form: by print, photoprint, microfilm, or any other means without written permission of the Institute.

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

# IMPLEMENTATIE EN EFFECTEN VAN INTERNE DIFFERENTIATIE

---

*J. Terwel*

*P. Herfs*

*R. Dekker*

*W. Akkermans*

---

*J Terwel*





# **IMPLEMENTATIE EN EFFECTEN VAN INTERNE DIFFERENTIATIE**

J. Terwel  
P. Herfs  
R. Dekker  
W. Akkermans

Rijksuniversiteit Utrecht  
Vakgroep Onderwijskunde  
Project Interne Differentiatie  
SVO 0647  
april 1988





# **IMPLEMENTATIE EN EFFECTEN VAN INTERNE DIFFERENTIATIE**

Een empirisch, vergelijkend onderzoek naar de realisering en de effecten van interne differentiatie in heterogene groepen in de eerste fase voortgezet onderwijs bij wiskunde.

J. Terwel  
P. Herfs  
R. Dekker  
W. Akkermans

Rijksuniversiteit Utrecht  
Vakgroep Onderwijskunde  
Project Interne Differentiatie  
SVO 0647  
april 1988



"..... in een groep moeten de leerlingen samen, elk op het voor hem gepast niveau, hetzelfde onderwerp bewerken. De samenwerking moet de leerlingen op alle niveaus de toegang naar een hoger niveau openen, die van het lagere, omdat ze zich op die op het hogere kunnen oriënteren, die van het hogere, omdat de blik naar het lagere niveau hun nieuwe inzichten verschaft".

Prof. dr. H. Freudenthal

## Ten geleide

Lang niet altijd leidt onderwijsvernieuwing tot de beoogde verbetering van leerresultaten. Vaak meent men dan te moeten concluderen dat de vernieuwing mislukt is. In dit rapport is verslag gedaan van een onderzoek naar effecten van een nieuw programma voor wiskunde-onderwijs van 12 tot 16-jarigen. Het onderzoek laat zien dat het programma positieve effecten heeft op het leren in de klas. Over het geheel genomen zijn de leereffecten beter in vergelijking met een traditioneel lesprogramma. Toch blijven er nog wensen over.

Het ideaal 'Wiskunde voor iedereen' blijkt nog niet volledig gerealiseerd. Ook bij dit programma blijven sommige leerlingen achter. Meisjes blijken anders te reageren op wiskunde dan jongens. Zogezien gaat het om een gedeeltelijk geslaagde vernieuwing. Maar ook blijkt dat niet iedere leerkracht op dezelfde wijze omgaat met het nieuwe programma. Dat blijkt gevolgen te hebben voor de leerresultaten. Ook de individuele leerkracht doet er toe.

Zo blijkt dat er meer nodig is voor onderwijsverbetering dan alleen een nieuw programma. Vormen van gerichte nascholing en het klimaat in de school zijn voorbeelden van andere belangrijke bepalende factoren in veranderingsprocessen. Het is dan ook van groot belang, dat dit succesvolle programma kansen krijgt op een weloverwogen implementatie in het onderwijs.

Prof. dr. N.A.J. Lagerweij.



## Woord vooraf

Dit onderzoek is tot stand gekomen dankzij de inzet van vele personen en diverse instanties. Wij bedanken de leraren en directies van scholen voor hun medewerking en voor het in ons gestelde vertrouwen.

Het leerplanontwikkelingswerk van de sectie wiskunde van de Stichting voor de Leerplanontwikkeling (SLO) was een belangrijke basis voor het onderzoek. We zijn drs. P. Verstappen en zijn ontwikkelingsteam van de SLO erkentelijk voor hun bijdrage aan de totstandkoming van het onderzoek.

Zonder de subsidie van de Stichting voor Onderzoek van het Onderwijs (SVO) en van de SLO zou dit onderzoek niet ten uitvoer zijn gebracht. We danken mevrouw drs. E. Patijn - Stroink van de SVO voor haar aandacht voor de financiële aspecten, waardoor de continuïteit van het onderzoek werd gewaarborgd.

Prof. dr. H. Freudenthal inspireerde ons met zijn ideeën over heterogene groepen, niveaus in het leerproces en rijke contexten. (Vergelijk Freudenthal 1973, a en b, en 1980). Zijn steun was van grote betekenis voor het onderzoek. Ook personen uit de kring rondom Freudenthal zijn we erkentelijk voor hun belangstelling voor ons onderzoek. We denken daarbij in het bijzonder aan dr. A. Treffers van de vakgroep Onderzoek Wiskunde-onderwijs en Onderwijs Computercentrum (OW & OC) van de Rijksuniversiteit Utrecht en prof. dr. F. Goffree van de SLO en de Universiteit van Amsterdam. Dr. K.M. Stokking van de Rijksuniversiteit Utrecht, bedanken we voor zijn steun en deskundig advies in verschillende fasen van het onderzoek.

Van verschillende personen ontvingen wij constructieve reacties naar aanleiding van de concept-versie van dit rapport. Een deel van deze mensen is hierboven al genoemd. We bedanken de volgende mensen voor hun waardevolle opmerkingen: drs. J.C. van Bruggen, dr. H. Bonset, dr. C. Boonman, dr. H. Franssen, dr. P. van den Eeden, dr. P. de Koning, drs. J. Perrenet, drs. A. de Vries en drs. B. Versloot. We bedanken J. Stekelenburg en M. ten Brink voor de deskundige en vlotte wijze waarop de teksten zijn verwerkt.

De projectleden van ID 12-16 hebben zich verheugd over de belangstelling en steun van vele mensen in de afgelopen jaren. Als er tekortkomingen zijn ligt het niet aan hen, daarvoor zijn wij verantwoordelijk.

We zien terug op een boeiende en leerzame periode. We hopen dat de gegevens uit ons onderzoek een bijdrage kunnen leveren aan de verdieping van het inzicht in vraagstukken van onderwijs in (heterogene) groepen. We hopen ook dat de resultaten van het onderzoek een rol kunnen spelen in de discussie over de vormgeving van de eerste fase van het voortgezet onderwijs. We denken daarbij ondermeer aan het werk van de commissie F. van der Blij: de Commissie Ontwikkeling Wiskunde-

onderwijs (COW). Ons rapport kan echter ook voor mensen buiten de kring van het wiskunde-onderwijs relevante informatie opleveren. We denken in principe aan al die mensen die zich inzetten voor 'Basisvorming voor Iedereen'. In het wiskunde-onderwijs zijn de problemen van Basisvorming als het ware onder een vergrootglas te zien.

Misschien kan ons onderzoek ook een reflectie op gang brengen over methodologische vraagstukken bij het onderzoek in complexe veldsituaties (vgl. Terwel 1987a).

Inmiddels zijn we gestart met een nieuw onderzoek op het gebied van interne differentiatie: een onderzoek naar de effecten van Adaptief Groepsonderwijs voor 12-16 jarigen (AGO 12-16/ SVO 7031). Evenals ID 12-16 maakt AGO 12-16 onderdeel uit van het voorwaardelijk gefinancierde programma 'Innovatie van onderwijs en Curriculum' van de vakgroep Onderwijskunde aan de Rijksuniversiteit Utrecht.

## INLEIDING

Voor u ligt het eindrapport van het onderzoeksproject 'Interne Differentiatie Wiskunde-onderwijs voor 12-16 jarigen' (ID 12-16). Het betreft een empirisch onderzoek naar interne differentiatie in (heterogene) klassen in de eerste fase van het voortgezet onderwijs bij wiskunde.

Het ontstaan van dit project gaat terug tot 12 november 1980. Op deze dag organiseerde de SLO een landelijke manifestatie in Utrecht over leerplanontwikkeling en verzorgingsstructuur. Tijdens deze manifestatie was er een gesprek tussen P. Verstappen, coördinator van de sectie wiskunde van de SLO en J. Terwel van de Vakgroep Onderwijskunde aan de Rijksuniversiteit te Utrecht. In dit gesprek bleek dat de SLO was begonnen met een nieuw ontwikkelingsproject op het gebied van wiskunde voor 12-16 jarigen. Uitgangspunten waren ideeën van professor H. Freudenthal. De strijdroep was: Wiskunde voor Iedereen.

De SLO bleek geïnteresseerd in onderzoek dat in relatie tot dit ontwikkelingsproject zou worden opgezet. Na dit eerste contact zijn verkennende gesprekken gevoerd en zijn plannen gemaakt. Deze mondden uit in een aanvraag voor subsidie bij de Stichting voor Onderzoek van het Onderwijs (vergelijk Terwel 1981). Na honorering van deze aanvraag ging het onderzoek op 1 augustus 1982 officieel van start. De totale looptijd van het project bedroeg, inclusief de voorbereidingsfase, ruim zeven jaar. Hiervan zijn ruim vier jaar gesubsidieerd uit externe bronnen (SVO/SLO).

Het onderzoeksteam bestond uit R. Dekker, P. Herfs en J. Terwel. Voor de statistische verwerking van de gegevens werd in verschillende fasen, ondersteuning verleend door respectievelijk D. van der Ploeg en W. Akkermans. Voorts is een deelstudie verricht door M. Posthuma de Boer naar het gebruik van een handleiding voor leraren. Verschillende studenten van de Vakgroep Onderwijskunde hebben hun doctoraal-onderzoek in relatie tot dit project uitgevoerd en daarvan in scripties verslag gedaan.

Het onderzoek kent drie fasen. De eerste twee fasen zijn afgesloten. Deze hebben ondermeer twee interimrapporten en een handleiding voor het werken met heterogene groepen opgeleverd (Dekker e.a. 1983a en 1985, Posthuma de Boer 1986). Het proefschrift van Terwel (1984) bevat een beschrijving van de opzet en de theoretische achtergronden van het onderzoek. Deze vier publikaties vormen samen met de onderzoeksaanvragen Terwel (1981) en de continueringsaanvraag Dekker e.a. (1985d) de basisliteratuur van het project. Lezers die een indruk willen krijgen van het gehele project (dat wil zeggen onderzoeksopzet, theoretische achtergronden, tussentijdse resultaten en praktische aanwijzingen voor leraren) verwijzen we naar deze zes publikaties. In de literatuurlijst staan deze bovenaan.

Voorts zijn in diverse wetenschappelijke en praktijkgerichte tijdschriften publikaties verschenen. Ook door middel van lezingen en paperpresen-

taties op nationale en internationale conferenties van onderwijsonderzoekers zijn de resultaten van het onderzoek verspreid (zie literatuurlijst onder overige literatuur).

Dit eindrapport is een verslag van de derde en laatste onderzoeksfase. Na deze inleiding volgt in hoofdstuk 1 een beschrijving van de opzet van het onderzoek. Hoofdstuk 2 geeft een beeld van het onderwijs zoals dat in de klaspraktijk verloopt. Het gaat hier om gegevens die verkregen zijn door middel van directe observaties. Dit beeld wordt in hoofdstuk 3 aangevuld met gegevens over het onderwijsleerproces zoals de leerlingen dat zien: percepties van leerlingen. Hoofdstuk vier bevat de gegevens over de houding van de leerlingen vóór en na het doorlopen van het programma. Hoofdstuk 5 geeft een uitgebreid overzicht van de toetsresultaten van de leerlingen vóór en na het doorlopen van het wiskunde-programma. In hoofdstuk 6 komen de lijnen uit de voorgaande hoofdstukken bij elkaar. Het gaat om een integratie van de gegevens, vooral op basis van correlaties en regressie-analyses tussen gegevens uit de beginsituatie, het onderwijsleerproces en de effecten aan het eind.

In hoofdstuk 7 vatten we alles samen en trekken we hoofdconclusies uit het onderzoek. Het rapport besluit met een hoofdstuk (8) discussie en aanbevelingen. Tenslotte is een literatuurlijst opgenomen van onder meer alle publikaties die in het kader van dit project of in relatie tot dit project door projectmedewerkers zijn uitgebracht.

Dit rapport is in de eerste plaats geschreven voor onderwijsonderzoekers en vakdidactici. Wellicht zijn bepaalde hoofdstukken ook interessant voor beleidsmakers, leerplanontwikkelaars, opleiders van leraren en leraren in het voortgezet onderwijs.

Wie echter niet primair in de effectvraag en daarmee samenhangende methodologische vraagstukken is geïnteresseerd en vooral belangstelling heeft voor concrete suggesties betreffende groepswork en didactisch handelen verwijzen we naar het SLO materiaal en naar de handleiding van M. Posthuma de Boer, "Werken met heterogene groepen" (1986).

Uitgebreide kwalitatieve beschrijvingen van het werken in heterogene klassen en van het probleemoplossen in kleine groepen zijn onder meer te vinden in onze twee interimrapporten Dekker e.a. (1983a en 1985). Het is ook mogelijk 'Wiskunde voor Iedereen' in actie te zien op een video-film van de RUU. Daarin presenteren we beelden van classesituaties. We gaan ook in op theoretische achtergronden en onderzoeksresultaten.

Tenslotte geven we nog een leessuggestie voor mensen die te weinig tijd hebben om het rapport van A tot Z te lezen. Wellicht is het dan het beste om eerst de samenvatting en eindconclusies (hoofdstuk 7) te lezen. Daarna zou men hoofdstuk 6 kunnen lezen. In dat hoofdstuk brengen we alle gegevens bij elkaar. Zo heeft men in korte tijd een vrij compleet beeld van de inhoud van dit rapport.

<b>INHOUDSOPGAVE</b>	<b>pagina</b>
Titelblad	I
Ten Geleide	III
Woord Vooraf	IV
Inleiding	VI
Inhoudsopgave	VIII
<b>Hoofdstuk 1: ACHTERGROND EN OPZET VAN HET ONDERZOEK</b>	<b>1</b>
1.1 Inhoudelijk-theoretische achtergrond	1
1.2 Methodologische achtergrond	6
1.3 Vraagstelling en hypothese	8
1.4 Methoden van onderzoek en fasering van het project	12
1.5 Onderzoeksdesign	14
1.6 Onderzoekssituatie	15
1.7 Wiskundig-didactische inhoud	16
<b>Hoofdstuk 2: TWEE KLASSEN IN BEELD</b>	<b>19</b>
2.1 Werkwijze	19
2.2 Résumé	19
2.3 Klas A in beeld	20
2.4 Klas B in beeld	26
2.5 Reflectie	32
<b>Hoofdstuk 3: PERCEPTIES VAN HET ONDERWIJSLEERPROCES</b>	<b>35</b>
3.1 Vraagstelling	35
3.2 Operationalisering	35
3.3 Betrouwbaarheid	39
3.4 Procedure en analyse	40
3.5 Resultaten	40
3.5.1 Gemiddelden en correlaties	40
3.5.2 Verschillen tussen condities, scholen, klassen	46
3.6 Conclusies	49
<b>Hoofdstuk 4: DE HOUDING VAN DE LEERLINGEN</b>	<b>51</b>
4.1 Vraagstelling	51
4.2 Operationalisering	52
4.3 Betrouwbaarheid	52
4.4 Procedure en analyse	53
4.5 Resultaten	53
4.6 Conclusies	59



	pagina
Hoofdstuk 5: LEERRESULTATEN: VOORTOETS EN NATOETS	61
5.1 Vraagstelling	61
5.2 Operationalisering	61
5.3 Betrouwbaarheid	65
5.4 Procedure en analyse	66
5.5 Resultaten	66
5.5.1 De beginkennis van de leerlingen: toets 1	66
5.5.2 De toetsresultaten op de drie toetsen	73
5.6 Conclusies en reflectie	87
Hoofdstuk 6: INTEGRATIE VAN ONDERZOEKSGEGEVENS: RELATIES TUSSEN VARIABLEN	93
6.1 Vraagstelling en hypothese	93
6.2 Operationalisering en procedures	94
6.3 Correlaties op klassenniveau	97
6.4 Verschillen tussen leerlingen op individueel niveau	106
6.5 Verschillen in effecten tussen projectklassen (hypothese A)	112
6.6 Verschillen in leereffecten tussen condities (hypothese B)	114
6.7 Conclusies	117
Hoofdstuk 7: SAMENVATTING EN EINDCONCLUSIES	119
7.1 Opzet en vraagstelling	119
7.2 Wat is de beginsituatie?	120
7.3 Worden de onderwijsleerprocessen gerealiseerd?	121
7.4 Zijn de beoogde leereffecten opgetreden?	123
7.5 Wat zijn de relaties tussen beginsituatie, processen en effecten?	125
Hoofdstuk 8: DISCUSSIE EN AANBEVELINGEN	127
LITERATUUR	139
Bijlage 1 : Items percia-vragenlijst	145
Bijlage 2 : Items vragenlijst houding	151
Bijlage 3 : Items toets 3 (natoets)	155
Bijlage 3a: Relatieve scores voortoets	183
Bijlage 3b: Relatieve scores natoets	186
Bijlage 4 : Observatiegegevens projectklassen	189





# 1 ACHTERGROND EN OPZET VAN HET ONDERZOEK

## 1.1 Inhoudelijk-theoretische achtergrond

Dit onderzoek betreft het effect van een programma voor intern gedifferentieerd wiskunde-onderwijs in de eerste fase van het voortgezet onderwijs. Het vraagstuk van interne differentiatie staat daarbij centraal. Hoe wordt op verschillen tussen leerlingen ingespeeld en wat zijn de effecten? Het gehanteerde 'differentiatie-model' is in zekere zin uniek en tamelijk gecompliceerd. Het vormt een integraal onderdeel van een bepaald onderwijsconcept. Daarom beschrijven we het 'differentiatie-model' in de context van dat concept en het daarop gebaseerde vernieuwingsprogramma.

Door de SLO is een innovatieprogramma ontwikkeld voor nieuw wiskunde-onderwijs voor 12-16 jarigen. Dit programma bestaat uit lesmateriaal voor leerlingen en leraren en achtergrondmateriaal (visiestukken). Dit programma bevat 'modellen' voor nieuwe doelen, inhouden en werkvormen. Het betreft dus niet alleen curriculum materiaal in de smalle betekenis van leerstofvernieuwing. Er worden vernieuwingen voorgesteld die het didactisch handelen van leerkrachten betreffen. Ook de schoolorganisatie en de klasse-organisatie is in het geding bij het werken in heterogene groepen van 12-16 jaar. Daarbij heeft de SLO zich laten inspireren door ideeën van prof. H. Freudenthal die stelde dat het mogelijk is kinderen na de basisschool nog jaren bijeen te houden in heterogene klassen. Freudenthal keerde zich tegen het opsplitsen van leerlingen als gevolg van het vasthouden aan verouderde leerplannen. Dit geheel aan ideeën en materialen heeft de SLO uitgedragen onder de vlag van 'Wiskunde voor Iedereen'.

De ideeën en uitgangspunten die ten grondslag liggen aan het nieuwe wiskundeprogramma zijn omschreven in het concept: 'Wiskunde voor Iedereen'. In dit concept zijn ideeën terug te vinden uit het vroegere 'Instituut voor Ontwikkeling van het Wiskunde Onderwijs' (IOWO). In het bijzonder verwijzen we naar het werk van Freudenthal (1973a en b en 1980). Dit geheel aan ideeën hebben we elders uitgebreid beschreven en geanalyseerd (Terwel 1984, blz. 179 e.v. en Dekker e.a. 1985, blz. 254, e.v.).

Het onderwijsleerproces zoals de SLO dat voor ogen staat is te typeren met vier vernieuwende kenmerken:

- a. klassikale introducties en reflecties in heterogene klassen;
- b. samenwerking in heterogene subgroepjes van bijvoorbeeld vier leerlingen;
- c. niveaus in het leerproces (verschillen in oplossingen en oplossingswegen en de reflectie daarop);
- d. wiskunde in (rijke) contexten.

Het onderwijsleerproces verloopt in de tijd gezien als volgt. De leraar introduceert klassikaal nieuwe onderwerpen. Daarna werken leerlingen in heterogene subgroepjes van bijvoorbeeld vier leerlingen aan opgaven in 'rijke' contexten. Deze opgaven kunnen op verschillende manieren worden opgelost. Soms is het nodig een klassikaal intermezzo te organiseren, bijvoorbeeld wanneer meerdere groepjes op hetzelfde punt vastlopen. Tenslotte is er een klassikale reflectie. Deze sequentie van onderwijsleeractiviteiten kan men in één les of in een lessenreeks doorlopen.

Omdat het niveauprincipe (kenmerk c) zo'n centrale rol vervult in de differentiatie zoals de leerplanontwikkelaars van de SLO dat zien, gaan we daar nader op in. Het niveauprincipe is door Freudenthal in verschillende publikaties beschreven. In zijn inleiding op de 'Gesamtschule Conferentie 1973' te Beekbergen waarschuwde Freudenthal voor het gevaar van separatie van leerlingen. Het is volgens Freudenthal evident dat dan alleen de 'goeden' er 'beter' van worden. "Wie heeft die wordt gegeven. We moeten echter luisteren naar de dringende vraag: hoe reageren we op de natuurlijke differentiatie in het leerproces?" (Freudenthal 1973a, blz. 91). Freudenthal's pleidooi voor toepassing van het niveauprincipe ging vergezeld van een afwijzing van andere differentiatie-maatregelen. Hij formuleerde het als volgt:

"De niveaus in de wiskunde, die ik bedoel, zijn niveaus van begrip. Wat op het ene niveau activiteit is, wordt op het volgende onderwerp van beschouwing. Zodoende presenteert elk niveau zich als verdieping van het voorafgaande. In 't algemeen bevinden leerlingen zich naast elkaar op verschillende niveaus van het leerproces, ook als ze dezelfde materie bewerken. Dit is een ervaring die men in elk klassikaal onderwijs kan opdoen. Men beschouwt het als een noodzakelijk kwaad - ik wil er een deugd van maken, maar genuanceerd in dier voege, dat de leerlingen niet naast, maar met elkaar met dezelfde stof op verschillende niveaus werkzaam zijn, en wel in bewust heteroogeen samengestelde kleine groepen van bijvoorbeeld vier leerlingen.

Dit verschilt hemelsbreed van de ABC-methode en van de methode van fundamentum en additum, of welke termen men anders zou gebruiken. Een verrijksstof voor de één, een repetitiestof voor de ander, terwijl een middengroep ondertussen de pas moet markeren - dat betekent separatie. Maar wat erger is, het betekent, dat men aan de niveaus van de wiskunde

in het leerproces geen recht laat wedervaren.

Ik herhaal het: in een groep moeten de leerlingen samen, elk op het voor hem gepaste niveau, hetzelfde onderwerp bewerken. De samenwerking moet de leerlingen op alle niveaus de toegang naar een hoger niveau openen, die van het lagere, omdat ze zich op die op het hogere kunnen oriënteren, die van het hogere, omdat de blik naar het lagere niveau hun nieuwe inzichten verschaft". (Freudenthal 1973a, blz 93).

Freudenthal vroeg zich in zijn inleiding overigens af, of het niveauprincipe in de praktijk gerealiseerd kan worden. "Of deze niveaus ook werkelijk didactisch in de door mij bedoelde zin zich laten toepassen, is de grote vraag, die praktisch moet worden beantwoord. In de praktijk moet blijken of men de leraar in staat kan stellen, de leerprocessen dusdanig te programmeren en te sturen, dat werkelijk leerlingen van verschillend niveau met succes samen hetzelfde onderwerp kunnen bewerken. Een vanzelfsprekende vereiste is hierbij dat de leraar zelf de niveauctuur van het leerproces heeft begrepen". (1973a, blz. 94).

Toen de eerste onderzoeken met betrekking tot deze vraag van Freudenthal in het project ID 12-16 waren uitgevoerd werd duidelijk hoe complex het vraagstuk van de niveaus in het leerproces is. De niveaus in het leerproces werden aanvankelijk nauwelijks gerealiseerd in de concrete klaspraktijk. Het niveauprincipe was ook nog te weinig geëxpliciteerd in het schriftelijke materiaal voor leerlingen en leraren. Herhaaldelijk stond in discussies tussen ontwikkelaars en onderzoekers het niveauprincipe centraal. Is het niveauprincipe niet te moeilijk? Weten leraren, leerplanontwikkelaars en onderzoekers eigenlijk wel waar ze het over hebben als ze spreken over niveaus en over reflectie?

Dat waren enkele vragen in de beginperiode van het onderzoek.

In de dissertatie van Terwel (1984, blz. 212) wordt het probleem als volgt omschreven. "De gedachte van de niveaus in het leerproces als benaderingswijze van het vraagstuk van differentiatie in heterogene groepen stelt hoge eisen aan leraren, leerplanontwikkelaars en onderzoekers. Het blijkt moeilijk tot een goede operationalisatie van het niveaubegrip te komen. Soms lijkt het er zelfs op dat hier sprake is van een nimbusbegrip, namelijk een begrip met een waas van heiligheid waarvan de betekenis meestal vaag is".

Sindsdien is er echter vooruitgang geboekt. In latere fasen van het onderzoek vonden we geslaagde voorbeelden van de realisering van het niveauprincipe in de concrete klaspraktijk (vgl. Dekker e.a. 1985 en Dekker e.a. 1986). Toch zijn de problemen, zoals de operationalisering van het begrip, de uitwerking in concreet materiaal en de toepassing ervan in het onderwijsleerproces, niet volledig opgelost. Om nog maar te zwijgen

van de problemen om via onderzoek tot de niveaus door te dringen. Nu na zo'n zeven jaar ontwikkelingswerk en onderzoek is de conclusie dat de ene leraar er wél in slaagt het niveau principe te realiseren terwijl de andere de niveaus niet of nauwelijks in praktijk brengt (zie hoofdstuk 2 van dit rapport en ook de discussie in hoofdstuk 8).

Het lijkt nog steeds buitengewoon zinvol om verder te werken aan de ontwikkeling en het onderzoek betreffende het niveau principe.

Gedurende de gehele onderzoeksperiode was er overeenstemming tussen Prof. H. Freudenthal, de leerplanontwikkelaars en de onderzoekers op het punt van de pedagogische motieven achter de 'niveaus' als differentiatie principe bij het werken in heterogene groepen. Deze motieven zijn (a) het voorkomen van (permanente) separatie tussen leerlingen in de eerste fase van het voortgezet onderwijs en (b) het creëren van een menswaardige, stimulerende leeromgeving voor alle leerlingen. Dat was de basis waarop de gesprekspartners in het vernieuwingsproces elkaar steeds konden vinden en aanspreken.

Er waren ook verschillen in inzichten en verschillen in verwachtingen. Door de onderzoekers is het niveau principe steeds als een belangrijk middel bij interne differentiatie gezien. Maar er bestond bij de onderzoeker(s) vanaf het begin de verwachting dat naast het niveau principe óók gewerkt zou worden aan de ontwikkeling en het onderzoek van andere differentiatie maatregelen. Dit blijkt bijvoorbeeld uit het onderzoeksplan (Terwel, 1981) dat in nauwe samenwerking met de leerplanontwikkelaars tot stand is gekomen. In dat onderzoeksplan worden naast het kenmerk 'niveaus in het leerproces' twee andere kenmerken genoemd. Het ging om een *combinatie* van differentiatie maatregelen.

Er wordt in dat plan gesproken over een kenmerk 'gemeenschappelijke doelen'. Dit kenmerk heeft betrekking op de formulering en het nastreven van communale doelen. Het impliceert het nemen van differentiatie maatregelen om te voorkomen dat bepaalde leerlingen achterblijven (zie Terwel 1981, blz. 32).

Direct hieraan gekoppeld werd nóg een kenmerk voorgesteld namelijk het kenmerk "Toets- en registratiesysteem". In het plan staat: "Gedifferentieerd onderwijs staat of valt met een goed systeem van toetsing en registratie" (zie Terwel 1981, blz. 33). De veronderstelling was dat een diagnostische toetsing tot ontwikkeling zou worden gebracht als een onderdeel van het nieuwe programma. Deze twee kenmerken zijn door de SLO echter niet tot ontwikkeling gebracht.

Toen in het empirisch onderzoek bleek dat er leerlingen achterop raakten, hebben we deze voorstellen herhaald in de vorm van een reeks ontwerpregels ten behoeve van de leerplanontwikkelaars (Dekker e.a. 1985, blz. 258 en Dekker e.a. 1985d). Deze ontwerpregels zijn door de SLO echter niet in concreet lesmateriaal uitgewerkt. Dit heeft ook gevolgen



gehad voor het onderzoek. In het onderzoeksplan (Terwel 1981, blz. 32) formuleerden we twee onderzoeksvragen die betrekking hebben op deze twee kenmerken:

1. Bereiken alle leerlingen de minimumdoelstellingen?  
(Voldoen de prestaties aan de eisen? Daarbij zou speciaal worden gelet op bepaalde categorieën leerlingen: meisjes en zwakke of kansarme leerlingen).
2. Hoe functioneert het toets- en registratiesysteem?  
(Voldoet het systeem aan de eisen?).

Deze twee onderzoeksvragen waren niet te beantwoorden omdat de twee kenmerken waarop deze vragen betrekking hebben niet door de SLO tot ontwikkeling zijn gebracht. Vooral het ontbreken van minimumdoelstellingen en criteria raakte onmiddellijk aan het toetsende karakter van het onderzoek. Hier was de functie van het onderzoek in het geding. Dit aspect is door de onderzoekers regelmatig in vergaderingen aan de orde gesteld. Daarbij ging het niet om invloed op de inhoud van het programma uit te oefenen. De verantwoordelijkheid voor de programma-ontwikkeling ligt bij de SLO. Dat is hun opdracht. Het betreft hun professionaliteit. Maar omdat twee essentiële onderzoeksvragen, niet te beantwoorden waren, wegens het ontbreken van het onderzoeksobject i.c. de twee kenmerken, raakte het direct aan de opdracht en professionaliteit van de onderzoekers. Hoe kan men bijvoorbeeld vaststellen of alle leerlingen bepaalde minimumdoelstellingen hebben bereikt als deze niet zijn geformuleerd in toetsbare termen?

De verwachting bij de onderzoekers was dat elke lessenreeks zou worden afgesloten met een diagnostische toetsing waarbij de leerlingen individueel opgaven maken. Op basis van het resultaat van deze toetsing treft de leraar aanvullende didactische maatregelen voor specifieke groepen leerlingen met het oog op het bereiken van zo hoog mogelijk niveau voor alle leerlingen. Dit impliceert het formuleren van communale (en differentiële) doelen. De onderzoekers verwachtten, op basis van het overeengekomen onderzoeksplan en de latere continueringsaanvragen, dat de uitwerking hiervan door de leerplanontwikkelaars als normaal onderdeel van het programma zou worden geconstrueerd (vergelijk Dekker e.a. 1985d).

In de loop van het onderzoek werd duidelijk dat de leerplanontwikkelaars dit werk niet op zich wilden nemen omdat het niet in hun visie op onderwijs past (zie hoofdstuk 8). Toen werd duidelijk dat er tussen onderzoekers en leerplanontwikkelaars niet alleen verschillende verwachtingen maar ook verschillende theorieën over effectief onderwijs bestaan. Dat is een unieke kans vanuit een oogpunt van onderzoek. Wij als



onderzoekers hadden die verschillende theorieën als uitgangspunt willen nemen voor het onderzoek in de laatste fase van het project. Uit deze verschillende theorieën zijn concurrerende hypothesen af te leiden die in een experiment kunnen worden getoetst.

## 1.2 Methodologische achtergrond

Met dit laatste punt raken we aan een tweede punt van verschil van inzicht tussen leerplanontwikkelaars en de onderzoekers.

De leerplanontwikkelaars wensten een (kwalitatief) beschrijvend onderzoek. De onderzoekers wensten (in de laatste fase) een *toetsend onderzoek* aan de hand van *criteria vooraf* zoals in het onderzoeksplan was vastgelegd. Bovendien was er de financieringsvoorwaarde van de SVO dat er een *vergelijking* moest worden uitgevoerd van de SLO werkwijze met een meer klassikale aanpak.

In de toekenningsbrief van de SVO van 29 maart 1985, is deze voorwaarde als volgt omschreven. "Daarbij stelt het bestuur als voorwaarde dat door u wordt nagegaan in hoeverre de SLO werkwijze een versnellend dan wel een vertragend effect heeft op de benodigde lestijd/leertijd voor verschillend begaafde leerlingen, in vergelijking met een meer klassikale aanpak. Ook dient inzicht te worden verkregen in de consequenties van de SLO werkwijze voor het behaalde eindniveau zowel qua eindniveau als qua veelsoortigheid van onderwerpen".

De onderzoekers hebben deze voorwaarde geaccepteerd. Dat impliceerde dat het onderzoek een *toetsend* en een *vergelijkend* karakter moest hebben. De onderzoekers hebben het toetsende en vergelijkende karakter van het onderzoek doorgezet, ook toen bleek dat de leerplanontwikkelaars geen criteria konden leveren en bovendien lieten blijken weinig fiducia te hebben in een vergelijkende aanpak.

We stelden voor *varianten* van het nieuwe programma te maken (met en zonder individuele toetsing en aanvullende didactische maatregelen) en deze varianten te vergelijken op effectiviteit. De vraag was: bereiken de leerlingen in de ene variant de doelen sneller of beter dan in de andere variant? Daarbij zouden we ook kijken naar de dimensie meer klassikaal versus meer groepswork. Met deze laatste vergelijking zou tevens worden voldaan aan de financieringsvoorwaarde van de SVO waarin een vergelijking met een meer klassikale aanpak werd voorgesteld. In feite ging het dus om een effectiviteitstoetsing van vier varianten van het ene SLO programma. Het design was een quasi-experimenteel design met a-selectie toewijzing van varianten aan klassen (en leraren).

Als dit design ten uitvoer zou zijn gebracht zouden veel methodologische problemen zijn opgelost. Het design voorziet namelijk in een

vergelijking tussen varianten gemeten aan dezelfde (in alle varianten nagestreefde) SLO-doelstellingen. Het design betekent ook een ont koppeling van klas en variant (door a-selectie toewijzing van varianten aan klassen). Dat sluit een aantal alternatieve interpretaties uit, waardoor men beter conclusies kan trekken. Het plan was echter praktisch niet uitvoerbaar. We presenteerden dit design te laat. Het design was in strijd met de visie van de leerplanontwikkelaars op onderwijs en onderzoek. De pogingen om de leraren in deze opzet mee te krijgen mislukten gedeeltelijk; een deel van de leraren was er tegen. Zo vielen er teveel klassen uit om nog door te gaan op die weg. Bovendien waren er géén concrete materialen en procedures beschikbaar voor de uitvoering van bepaalde varianten, omdat de leerplanontwikkelaars de twee genoemde kenmerken niet als onderdeel van hun ontwikkelingstaak hebben uitgevoerd. Evenmin waren zij bereid materialen te ontwikkelen voor de vergelijking van een meer klassikale aanpak tegenover een meer op groepswerk gerichte benadering. Zo dreigden we niet alleen in conflict te komen met de financieringsvoorwaarde van de SVO, we zagen ook de opbrengst van het onderzoek de mist in gaan.

Mede om die reden besloten we vergelijkingsscholen in het onderzoek te betrekken (zie ook hoofdstuk 5). De vergelijking was nu niet langer gericht op een vergelijking van varianten van het SLO-programma, maar een vergelijking van het SLO-programma met een meer traditioneel, klassikaal programma.

We hebben nadrukkelijk vastgehouden aan een toetsend/vergelijkend aspect in de laatste fase van het onderzoek. Niet alleen omdat de vergelijkingen een financieringsvoorwaarde was van de SVO, maar ook omdat we ons niet wilden beperken tot het puur beschrijven. Voor een methodologische verantwoording van dit standpunt zie Terwel 1987a. We hebben alles in het werk gesteld om criteria op tafel te krijgen: absolute criteria en/of vergelijkende criteria. De weg van de absolute criteria was geblokkeerd: de SLO kon geen absolute criteria leveren (zie hoofdstuk 5). Een vergelijking van varianten van het SLO-programma kon niet worden gerealiseerd. Zo kwamen we uit bij het design dat voorligt: (a) enerzijds een beschrijving en vergelijking van spontane, ongeplande variaties in de implementatie van het SLO-programma (b) anderzijds een vergelijking van het SLO-programma met een meer traditioneel, klassikaal programma op controlescholen.

De vergelijking van het nieuwe en het traditionele programma impliceert een criterium: een nieuw programma moet beter zijn dan iets wat al bestaat. Daarmee was de oplossing van het criterium vraagstuk binnen bereik. We hadden eindelijk een criterium waaraan de effectiviteit van het nieuwe SLO-programma kon worden afgemeten. In beginsel was daarmee ook een hypothese gegeven. Dat dit design methodologisch

minder sterk is als het design met de vooraf geplande, a-select toegewezen varianten, is duidelijk. We komen daar ondermeer in hoofdstuk 8 uitgebreid op terug.

Achteraf gezien is de accentverschuiving van een absolute naar een relatieve (vergelijkende) evaluatie een goede ontwikkeling geweest. We hebben niet alleen vergelijkingen gemaakt tussen klassen binnen de projectconditie, maar ook tussen condities: het SLO-programma in vergelijking met een traditionele, klassikale aanpak. Die laatste vergelijking leverde enkele interessante conclusies op:

1. Over het geheel genomen bereiken de leerlingen, die het nieuwe programma hebben gevolgd, de doelen sneller dan de leerlingen in de traditionele, klassikale situatie. Voor de zwakke leerlingen is er echter geen verschil van betekenis.
2. De meisjes behalen ook na het doorlopen van het SLO-programma lagere resultaten dan de jongens. Dit is echter geen specifiek effect van het SLO-programma. In de traditionele, klassikale situatie ziet men hetzelfde verschijnsel.

Nu de inhoudelijk-theoretische achtergrond en de methodologische keuzen zijn geschetst gaan we verder met de vraagstelling, de hypothesen en de opzet van het onderzoek.

### 1.3 Vraagstelling en hypothesen

De algemene onderzoeksvraag die in alle drie fasen van het onderzoek centraal heeft gestaan is:

Hoe functioneert een op een bepaald onderwijsconcept gebaseerd programma voor gedifferentieerd wiskunde-onderwijs voor heterogene groepen in het voortgezet onderwijs?

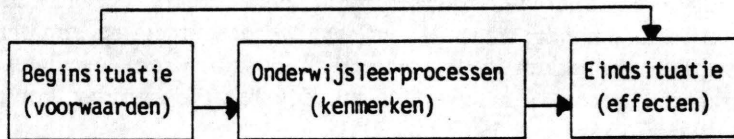
In deze derde en laatste fase van het onderzoek hebben we tegen de achtergrond van deze algemene onderzoeksvraag vier hoofdvragen geformuleerd. Bovendien is een algemene hypothese opgesteld die voor twee vergelijkingen is uitgewerkt in twee specifiekere hypothesen. De vier hoofdvragen zijn:

1. Wat is de beginsituatie van de leerlingen (houding en voorkennis)?
2. Worden de onderwijsleerprocessen volgens de bedoelingen gerealiseerd?
3. Zijn de beoogde leereffecten opgetreden?
4. Wat zijn de relaties tussen (1) beginsituatie, (2) onderwijsleerprocessen en (3) leereffecten?

De algemene werkhypothese die betrekking heeft op de vierde vraag is: Hoe meer de onderwijsleerprocessen volgens de 'bedoelingen' zijn gerealiseerd, des te beter zijn de leereffecten.

Zoals gezegd draait het onderzoek om de vier kenmerken van het onderwijsprogramma van de SLO. In feite zijn het globaal geformuleerde proceskenmerken. In het onderzoek is daarnaast de beginsituatie en de eindsituatie betrokken. Het volgende model vormt de basis voor het onderzoek.

*Figuur 1.1: Model voor het onderzoek*



In hoofdstuk 6 (figuur 6.1) is dit model gespecificeerd naar variabelen binnen de drie componenten. Tevens is aangegeven met welke instrumenten de variabelen zijn gemeten. Bij de bespreking van de methoden van onderzoek geven we ook al in dit eerste hoofdstuk, paragraaf 1.4, een nadere invulling voor wat betreft de gebruikte methoden en instrumenten.

In de algemene werkhypothese is al een indicatie gegeven over de relatie tussen onderwijsleerprocessen en effecten. We geven nu een nadere omschrijving van de drie componenten in het model en hun onderlinge relatie. Tegelijk is het een poging om de globale proceskenmerken in operationele (meetbare) termen te omschrijven. Dit resulteert in de formulering van twee specifieke hypothesen: één hypothese voor een vergelijking tussen klassen binnen de projectsituatie en één hypothese voor de vergelijking tussen twee condities (projectscholen en vergelijkingscholen). In deze twee specifieke hypothesen zijn de SLO kenmerken opgenomen. Bij de formulering van deze hypothesen is uitgegaan van bepaalde verwachtingen omtrent relaties tussen Beginsituatie, Onderwijsleerprocessen en Eindsituatie. In deze verwachtingen is ook de frequentie en de kwaliteit van bepaalde onderwijsleerprocessen, in relatie tot de Eindsituatie, aangegeven. Deze verwachtingen zijn gebaseerd op gegevens uit de literatuur en uit eigen onderzoek in eerdere fasen van het project. We vatten deze verwachtingen kort samen.

We verwachten dat leraren die een 'banende' rol vervullen relatief goede leerresultaten met hun leerlingen boeken. Met 'banen' is onder meer



bedoeld (a) kernbegrippen verhelderen aan de hand van aansprekende exempels (b) doelstellingen en taken expliciteren (c) aansluiten bij reeds aanwezige begrippen en ervaringen en het aanbrenge van een heldere structuur van begrippen (d) heldere samenvattingen en generalisaties geven. Het gaat om het 'banen' van de weg naar het abstracte, het doorstoten naar de kern, het leren hanteren van ideeën, begrippen en oplossingsprocedures in nieuwe probleemsituaties.

We verwachten ook dat in klassen met een duidelijke leiding en coördinatie van de klas als geheel en van de subgroepjes, de effecten beter zijn dan in klassen waarin minder 'sturing' vanuit de leraar is, dat wil zeggen waarbij de leerlingen meer in autonome, los van elkaar opererende subgroepjes werken en waarbij zij over lange perioden zijn aangewezen op eigen initiatief en/of leeftijdsgenoten. Anders gezegd als er géén adequaat klasmanagement is en er worden niet regelmatig klassikale onderwijsleergesprekken georganiseerd dan zullen de leerlingen in de groepjes minder goed samenwerken en hun leertijd minder goed (taakgericht) besteden. (Dekker, e.a. 1985, blz. 206).

Uiteraard speelt in dit geheel ook de wiskundig-didactische inhoud een belangrijke rol. Daarbij letten we vooral op het gebruik van wiskunde in zogenaamde contexten. Het gaat om een evenwichtig didactisch arrangement van leraar- en leerlingactiviteiten en inhoud.

Op grond van gegevens uit eerdere fasen van ons onderzoek en van gegevens uit literatuur verwachten we dat de beginsituatie van de leerlingen een belangrijke factor is. Te denken valt aan de beginkennis van de leerlingen op individueel niveau, maar ook aan de beginkennis van een klas; bijvoorbeeld de proportionele verdeling van sterke en zwakke leerlingen in een klas (vergelijk Terwel, 1986a). Daarnaast heeft het vraagstuk van meisjes en wiskunde steeds bijzondere aandacht gehad in het onderzoek (vergelijk Dekker 1984 en 1985 en vergelijk Meeder en anderen 1984). Daarom besteden we in dit rapport ook aandacht aan verschillen in prestaties, percepties en houdingen tussen jongens en meisjes.

Tegen de achtergrond van bovenstaande verwachtingen, waarin de SLO kenmerken uitdrukkelijk zijn opgenomen, hebben we de hypothesen geformuleerd. De algemene werkhypothese is primair ontwikkeld met het oog op klassen waarin met het nieuwe programma van de SLO is gewerkt. De verwachting was dat *binnen* deze 'experimentele' conditie verschillen in onderwijsleerprocessen zouden optreden. Analyse van die verschillen in het licht van deze hypothese kan inzicht verschaffen over meer of minder effectieve varianten bij de implementatie van het nieuwe programma. De specifieke hypothese voor de projectsituatie omschrijven we als volgt:

- A. In klassen waarin de leraar een banende, taakstellende, (bege)leidende rol vervult en waarin de leerlingen taakgericht samenwerken aan de oplossing van opgaven in contexten, zijn de leerresultaten beter dan in klassen waarin dat minder (vaak) het geval is.

Behalve vergelijking van varianten *binnen* de 'experimentele' conditie is in dit onderzoek ook een vergelijking gemaakt *tussen* de 'experimentele' conditie en een 'controle' conditie. Deze controle conditie bestaat uit klassen waarin met een meer traditioneel, klassikaal wiskundeprogramma is gewerkt. De specifieke hypothese voor deze vergelijking is:

- B. In onderwijs waarin met het nieuwe wiskundeprogramma van de SLO wordt gewerkt verlopen de onderwijsleerprocessen beter en zijn de leerresultaten beter dan in onderwijs met een meer traditionele aanpak.

Een beter verloop van onderwijsleerprocessen slaat op de frequentie en kwaliteit van bepaalde activiteiten van leraren en leerlingen en de mate waarin de wiskunde in voor de leerlingen herkenbare, motiverende contexten is geplaatst. (Zie ook de algemene werkhypothese en hypothese A).

In deze paragraaf moeten we ook iets zeggen over onze pogingen om de "niveaus in het leerproces" te meten. De hypothesen zijn geformuleerd met het oog op alle klassen en alle leerlingen. Dat stelt bepaalde eisen aan de methoden van onderzoek. Deze moeten geschikt zijn voor gegevensverzameling op grote schaal. Dat is de reden voor het overwegend kwantitatieve karakter van het onderzoek in deze laatste fase.

Het bleek echter niet gemakkelijk het 'niveauprincipe' met die relatief grootschalige, kwantificerende aanpak adequaat te beschrijven. Mede om die redenen is in dit onderzoek óók gebruik gemaakt van diepgaande, kwalitatieve observaties van de onderwijsleerprocessen in twee klassen. Daardoor zijn we toch in staat uitspraken te doen over het feitelijk functioneren van het niveauprincipe.

In de kwantitatieve analyses kunnen we de variabele 'niveaus in het leerproces' als afzonderlijk gemeten variabele niet volledig meenemen. Toch speelt dit niveauprincipe in enkele kwantitatieve analyses mee. Aspecten van het niveauprincipe komen voor als onderdeel van de gehanteerde perceptie-vragenlijst (zie hoofdstuk 3). Ook speelt het niveauprincipe een rol in analyses waarin we dummy variabelen gebruiken om het gehele complex van proceskenmerken (treatment) in de twee verschillende condities aan te duiden (zie hoofdstuk 6). In die analyses zijn alle proceskenmerken dus ook de niveaus in het leerproces als ongemeten



variabelen in de 'black box' van de projectconditie opgenomen. Met andere woorden voor zover het niveauprincipe in de klassen van de projectscholen is gerealiseerd kan het samen met de andere kenmerken de leereffecten hebben veroorzaakt. Het eventuele effect van de niveaus in het leerproces is dus niet buiten beschouwing gebleven. We hebben de toepassing van het niveauprincipe alleen niet in alle 22 klassen kunnen observeren en analyseren. Dat zou te veel onderzoekstijd hebben gevegd en dat zou betekenen dat we aan de beantwoording van de effectvraag niet zouden zijn toegekomen.

#### 1.4 Methoden van onderzoek en fasering van het project

Ter beantwoording van de onderzoeksvragen is gebruik gemaakt van verschillende methoden en instrumenten: toetsen, observatiemethoden, en vragenlijsten voor het meten van de houding en de percepties van de leerlingen.

De beginsituatie van de leerlingen is bepaald aan de hand van een voortoets en een vragenlijst naar de houding van de leerlingen.

Het onderwijsleerproces in de klas is op drie manieren in kaart gebracht (a) door kwalitatieve observaties van het onderwijsleerproces in twee klassen en verschillende groepjes binnen die klassen (b) door kwantitatieve observaties van de tijd aan groepswerk besteed en van de taakgerichtheid van de leerlingen (c) door het meten van de percepties van de leerlingen, waarbij de leerlingen als het ware als observators zijn opgetreden.

De bepaling van de eindsituatie vond plaats door middel van een eindtoets en een vragenlijst naar de houding van de leerlingen.

Omdat de methodekeuze moet worden gezien tegen de achtergrond van het gehele project dient iets gezegd te worden over de fasering. De keuze van methoden is in dit project niet gebaseerd op een vooringenomen standpunt ten faveure van één bepaalde onderzoekstraditie. Ons onderzoek als geheel is gebaseerd op een breed-spectrum-methodologie. Wij hebben daar bewust voor gekozen.

In de eerste fase van het onderzoek toen het concept 'Wiskunde voor Iedereen' nog erg open leek en de SLO-materialen nog volop in ontwikkeling waren lag het accent op kleinschalig, kwalitatief onderzoek op één school. Het onderzoek ondersteunde het ontwikkelingswerk, vooral door empirische gegevens aan te dragen over het functioneren van de 'vernieuwende' kenmerken van de SLO, maar ook door theoretische studies waarin de uitgangspunten in breder verband werden gezien.

In de tweede fase lag het accent nog sterk op het kwalitatief-

beschrijvende aspect. Maar daarnaast werd ook aandacht besteed aan instrumentontwikkeling ten behoeve van een grootschaligere opzet. Het aantal SLO-pakketten dat onderzocht werd nam toe. Er werden kwantitatieve gegevens verzameld op twee projectscholen en in totaal 18 klassen. Ook in deze fase leverde het onderzoek gegevens op ter ondersteuning van het ontwikkelingswerk. We formuleerden een reeks ontwerpregels voor de leerplanontwikkelaars (Dekker e.a. 1985) en we stelden een handleiding samen voor leraren die (gaan) werken met heterogene groepen in het voortgezet onderwijs (Posthuma de Boer 1986).

In deze (derde) laatste fase ligt het accent op een empirisch-analytische benadering, maar er is ook veel aandacht besteed aan het maken van portretten van classesituaties, gebaseerd op participerende observatie in twee klassen. Wederom nam het aantal onderzochte SLO-pakketten in vergelijking tot de vorige ronde toe. Er werden nog meer scholen in het onderzoek betrokken.

Deze accentverschuiving is te typeren in termen van polariteiten:

- van klein naar groot
- van formatief naar summatief
- van kwalitatief naar kwantitatief
- van interpreterend naar analyserend
- van beschrijvend naar toetsend/vergelijkend
- van deelname naar distantie.

Voor een adequaat beeld van het hele project is het aan te bevelen tenminste de zes basispublicaties (zie literatuurlijst bovenaan) en dit rapport te lezen. Wie vooral is geïnteresseerd is in kwalitatieve beschrijvingen en analyses verwijzen we naar fase 1 en 2 van dit onderzoek. Wie vooral belang stelt in de vraag naar de effecten en naar methodologische vraagstukken rondom onderzoek in complexe veldsituaties, verwijzen we naar fase 2 en 3 van het onderzoek.

Geheel in overeenstemming met onze keuze voor een breed-spectrum-methodologie hebben we de keuze van methoden afgeleid uit de doelstelling, vraagstelling en situatie van het onderzoek en niet omgekeerd. Tegelijkertijd is geprobeerd om in elke fase verbindingen te leggen tussen kwalitatieve en kwantitatieve gegevens.

Het vervolg van dit hoofdstuk gaat uitsluitend over de laatste (derde) fase van het onderzoek.

## 1.5 Onderzoeksdesign

Het design is te typeren als een voortoets - natoets design. Dit geldt zowel voor hypothese A als voor hypothese B. Bij hypothese A gaat het om een vergelijking tussen verschillende implementaties van het experimentele SLO programma in 22 klassen.

Bij hypothese B gaat het om een vergelijking tussen een experimentele en een controlegroep. In de experimentele conditie is het SLO programma uitgevoerd (22 klassen). In de controle conditie is gewerkt met een traditionele klassikale aanpak (11 klassen). Omdat is uitgegaan van bestaande groepen is het design voor hypothese B nader te omschrijven als een non-equivalent group design.

Hoewel deze typering van dit laatste design als eerste aanduiding voldoet, is nadere informatie nodig. Er is een vergelijking gemaakt tussen een experimenteel, innovatief programma en een meer traditioneel, klassikaal programma. Het begrip 'controlegroep' is echter methodologisch te pretentius voor de groep die het traditionele programma heeft gevolgd. Voor de 'controlegroep' zijn géén criteria geformuleerd waaraan het klassegebeuren zou moeten voldoen. Voor de experimentele groep heeft de SLO materiaal voor leerlingen en docenten ontwikkeld. Bovendien beschikten de docenten over een handleiding voor het werken met heterogene groepen die door de onderzoekers was samengesteld. (Posthuma de Boer, 1986). Zowel bij de samenstelling van de klassen en groepjes als bij de vormgeving van de onderwijsleerprocessen in de klassen waren de scholen en docenten autonoom. De termen experimentele- en controlegroep suggereren criteria, afspraken en controles op de naleving daarvan, die er in feite niet waren.

Daarom spreken we van een voortoets - natoets design met een projectgroep en een vergelijkingsgroep. Tussen deze twee groepen bestonden van meet af aan verschillen in beginkennis zoals gemeten met de voortoets. Tegenover het relatieve gebrek aan voorstructurering en controle van de onderzoekssituatie staat een zorgvuldige beschrijving van de beginsituatie, de onderwijsleerprocessen in de klas en de eindsituatie. We beschrijven en toetsen verschillen tussen beide condities.

Binnen de project-conditie toetsen we de hypothese over de relatie tussen onderwijsleerprocessen en effecten. Daarbij maken we gebruik van verschillen in implementatie van het vernieuwingsprogramma binnen de projectconditie. Op de achtergronden en beperkingen van dit design komen we terug in hoofdstuk 8: Discussie en aanbevelingen.

## 1.6 Onderzoekssituatie

In het onderzoek zijn drie project scholen, (twee brede scholengemeenschappen en een middenschool) betrokken die werken met een nieuw programma voor intern gedifferentieerd onderwijs. Daarnaast zijn gegevens verzameld over het onderwijs in twee vergelijkings scholen (mavo, havo, vwo scholengemeenschappen) die op een meer traditionele wijze werken. In de project scholen werd het nieuwe programma uitgevoerd in 22 klassen. Er waren ruim 500 leerlingen bij betrokken. In de vergelijkings scholen deden 11 klassen en ongeveer 300 leerlingen mee aan het onderzoek. In totaal waren er dus 5 scholen, 33 klassen en ruim 800 leerlingen in het onderzoek betrokken. Nu volgt nadere informatie over de drie project scholen en de twee vergelijkings scholen.

School 1 is een brede scholengemeenschap in een middelgrote stad in het noord-westen van het land. De adviezen van de leerlingen variëren van LBO t/m VWO. Er zijn relatief veel leerlingen met LBO-adviezen. De school is een stimulerings school en heeft relatief veel allochtone leerlingen. De school heeft veel 'zwakke' leerlingen en weinig 'sterke' leerlingen, dat wil zeggen leerlingen met respectievelijk veel lage en weinig hoge scores op de door ons afgenomen voortoets. Hoewel de verhouding 'zwakke' en 'sterke' leerlingen over de verschillende klassen schommelt, weerspiegelt de schoolpopulatie zich ook in de classesamenstelling. Dat wil zeggen er zijn veel klassen met relatief\* veel 'zwakke' leerlingen. Gelet op de voortoets scores zijn de klassen niet heterogeen van samenstelling. De klassegrootte ligt lager dan in de andere, door ons onderzochte, scholen namelijk, rond 22 leerlingen per klas. De school heeft acht tweede klassen. Niet alle leraren hebben ervaring met het nieuwe wiskunde onderwijs.

X School 2 is een middenschool, in een middelgrote stad in het midden van het land. Er is een instroom van leerlingen met LBO t/m VWO adviezen. Deze school voert een actief beleid met het oog op een evenwichtige instroom. Deze school heeft volgens onze voortoets gegevens een evenwichtige samenstelling. De klassen zijn heterogeen gelet op voorkennis bij wiskunde. De school heeft zes tweede klassen. De sectie heeft ervaring met het nieuwe wiskunde onderwijs.

---

\*Relatief betekent hier in relatie tot het totaal van scholen, klassen of leerlingen die in het onderzoek waren betrokken. 'Sterk' of 'zwak' slaat uitsluitend op voorkennis bij wiskunde zoals bepaald met de voortoets.



X School 3 is een brede scholengemeenschap in een middelgrote stad in het oosten des lands. De adviezen van de leerlingen variëren van LBO t/m VWO. Gelet op de voortoetsgegevens heeft deze school een evenwichtige samenstelling. De school werkt niet met volledig heterogene tweede klassen, maar past een vorm van streaming toe. Er zijn in totaal 8 tweede klassen: 3 LBO/MAVO klassen, 3 MAVO/HAVO klassen en 2 HAVO/VWO klassen. Deze school onderhoudt al jaren contacten met de SLO als proefschool en participeert voor de tweede keer in ons onderzoek. De sectie heeft dus ervaring met het nieuwe wiskundeonderwijs.

School 4 is een MAVO-HAVO-VWO scholengemeenschap in een middelgrote stad in het midden des lands. Omdat leerlingen met LBO-adviezen ontbreken heeft deze school relatief weinig 'zwakke' leerlingen in termen van onze voortoets. Deze school werkt niet met heterogene klassen maar past een vorm van streaming toe. Er zijn in totaal 8 tweede klassen: 3 MAVO klassen, 3 HAVO/VWO-determinatieklassen en 2 VWO klassen. Voor de MAVO klassen gebruikt deze school als wiskundemethode "Getal en Ruimte". De overige klassen werken met "Moderne Wiskunde".

School 5 is een MAVO-HAVO-VWO scholengemeenschap in een grote stad in het midden van het land. Ook hier ontbreken dus leerlingen met LBO-adviezen en voor het overige past ook deze school een vorm van streaming toe. Voor de MAVO klassen gebruikt men 'Denken, doen, begrijpen' en in de HAVO/VWO klassen hanteert men 'Getal en Ruimte'. Deze school participeerde slechts met een deel van haar tweede klassen in het onderzoek.

Op beide vergelijkingsscholen ontbreekt het LBO-type. Dat is geen opzettelijke keuze van de onderzoekers geweest. Integendeel, graag hadden wij onze projectscholen willen vergelijken met even brede (LBO t/m VWO) vergelijkingsscholen. Helaas konden we geen brede scholengemeenschap bereid vinden te participeren in het onderzoek. Een potentiële vergelijkingsschool (LBO t/m VWO) werd door brand getroffen. Daarmee viel die school af. School 4 en 5 werden na dit incident aangezocht door de onderzoekers. Een potentiële LBO school haakte op het laatste moment af.

### 1.7 Wiskundig-didactische inhoud

We geven nu een korte typering van het wiskundeprogramma zoals dat op de projectscholen is uitgevoerd. Het programma zoals dat in de periode van oktober 1985 t/m april 1986 is uitgevoerd in de projectscholen bestaat

uit de volgende onderdelen.

- Regelrecht A** : Lineaire verbanden in de vorm van grafieken, tabellen en regels.  
Contexten bij bijna alle opgaven.
- Pythagoras B** : Stelling van Pythagoras, wortels.  
Contexten bij 25% van de opgaven.
- Regelrecht B** : Lineaire verbanden in de vorm van grafieken, tabellen en regels.  
Contexten bij bijna alle opgaven.
- Vergelijkingen/ formules** : Eerstegraads vergelijkingen en lineaire verbanden in formulevorm. Variaties per school in gebruik van contexten en wiskundige inhoud.
- Uitstippelen** : Lineaire, kwadratische en hyperbolische verbanden in de vorm van grafieken, tabellen en regels.  
Contexten bij bijna alle opgaven.
- Vlak Voorbij** : Plattegrond, perspectief, kubus, vlakken, aanzichten, bouwplaten.  
Contexten bij bijna alle opgaven.

Op school 1 is wegens tijdgebrek 'Vlak Voorbij' niet behandeld terwijl per school en per klas soms eigen accenten zijn gelegd. Het pakket 'Uitstippelen' in school 1 is in enkele klassen slechts gedeeltelijk behandeld. Deze school heeft met ziekte van docenten te kampen gehad. In enkele klassen van school 2 is 'Uitstippelen' niet behandeld.

Voor uitgebreide informatie verwijzen we naar de betreffende SLO publicaties. Voor beschrijving van het onderwijs in de concrete klaspraktijk zie hoofdstuk 2 van dit rapport.

Aan het begin van de onderzoeksperiode is door de onderzoekers een studiedag voor docenten georganiseerd waarin de vernieuwingsideeën, materialen en didactische procedures van de SLO zijn geïntroduceerd.

Als we tot slot het vernieuwingsprogramma vergelijken met de programma's op de vergelijkingsscholen dan valt op dat de programma's wat wiskunde-inhoud betreft grotendeels samenvallen. Het overgrote deel van de inhoud van alle programma's heeft betrekking op lineaire verbanden in de vorm van grafieken, tabellen, regels en formules. Een klein deel dat



ook in alle programma's terugkomt wordt gevormd door de stelling van Pythagoras en het rekenen met wortels. Alleen de wiskunde-inhoud van 'Vlak Voorbij' zit niet in de programma's van de vergelijkingsscholen, maar ook op de projectscholen is dit pakket niet in alle klassen aan de orde gekomen.

Een van de meest in het oog springende verschillen tussen het vernieuwingsprogramma en de programma's op de vergelijkingsscholen is het vrijwel volledig ontbreken van contexten in de programma's op de vergelijkingsscholen. Een uitzondering vormt de methode 'Moderne Wiskunde' met contexten in 25 tot 30% van de opdrachten. Vijf van de elf vergelijkingsklassen (school 4) werken met deze methode. In het vernieuwingsprogramma zijn daarentegen contexten bij bijna alle opdrachten.

Nu gaan we in hoofdstuk 2 een begin maken met de beschrijving van het onderwijsleerproces. Deze beschrijving beoogt een beeld te schetsen van de praktijk in de klas. Dit hoofdstuk heeft een kwalitatief-beschrijvend karakter. Daarmee verschilt dit hoofdstuk van de rest van dit rapport dat overwegend kwantitatief-vergelijkend van aard is.

## 2 TWEE KLASSEN IN BEELD

### 2.1 Werkwijze

Als onderdeel van de procesmeting in ons onderzoek hebben we op 2 van onze onderzoeksscholen van 2 klassen en hun docenten portretten gemaakt om een beeld te geven van Wiskunde voor Iedereen in de klassepraktijk. Daarvoor hebben we in elke klas 300 minuten geobserveerd. In de ene klas, die we klas A\* zullen noemen, waren dat 6 lessen à 50 minuten, in de andere, klas B, 4 lessen à 75 minuten. De observaties vonden plaats met tussenpozen van ruim een week in de periode december '85 tot maart '86.

In deze ronde van ons onderzoek gaat het dus om een portret van de klas en zijn docent als geheel. We hebben per les telkens van een ander groepje bandopnames gemaakt terwijl we zelf, als observator, van het betreffende groepje, de docent en de klas als geheel aantekeningen hebben gemaakt. Deze aantekeningen en bandopnames hebben we verwerkt in uitgebreide lesverslagen. Uit deze lesverslagen hebben we fragmenten gekozen die wij typerend vonden voor de klas en de docent. Daarbij is in het bijzonder gelet op de vernieuwingskenmerken zoals geformuleerd in hoofdstuk 1. Deze fragmenten hebben we per klas tot een portret bewerkt waarbij we observatie en interpretatie typografisch van elkaar gescheiden hebben (zie 2.3 en 2.4). We voegen daar een reflectie aan toe (zie 2.5). Allereerst geven we een korte samenvatting van de analyses van de observatieverslagen.

### 2.2 Résumé

In beide klassen worden de lessen klassikaal begonnen. In klas A wordt het huiswerk besproken of geeft de docent een inleiding op een nieuw onderwerp waarbij hij de *contexten* verlevendigt. De inbreng van de leerlingen is echter gering en aan *verschillen in oplossingen* wordt weinig aandacht besteed. In klas B geeft de docent zijn leerlingen veel inbreng en worden *verschillen in oplossingen* besproken. De klassikale start beperkt zich echter tot het nabespreken van opdrachten, contexten etc. *Contexten* verlevendigt de docent van klas B zelden.

---

\*In de andere hoofdstukken van dit rapport heeft klas A een nummer gekregen namelijk 38.

Voor klas B is dat nummer 22. Klas A is een klas van school 3, een brede scholengemeenschap, klas B is een klas van school 2 (een middenschool).

Na de klassikale start begint in beide klassen het groepswerk. In klas A is er duidelijk sprake van *samenwerking binnen kleine heterogene groepen*. De docent stimuleert vooral uitleg aan elkaar. In klas B is dat veel minder het geval. De hulp van de docent is voornamelijk individueel gericht. In beide klassen komen *verschillen in oplossingen* tijdens het groepswerk weinig naar voren.

Een klassikale nabespreking wordt in klas A regelmatig gehouden en in klas B zelden.

Beide docenten treffen geen *extra didactische maatregelen* voor leerlingen die achterblijven.

### 2.3 Klas A in beeld

Klas A bestaat uit 29 leerlingen, 14 meisjes en 15 jongens. De leerlingen hebben een MAVO- of een HAVO-advies. Docent D. begint de les altijd met een klassikale inleiding op een nieuw onderwerp of een klassikale nabespreking van het huiswerk. Maar allereerst begint hij met orde houden, want klas A is een drukke klas.

D.: "Dat is de tweede keer dat ik probeer jullie stil te krijgen. En Mira snapt dat helemaal niet".

Het wordt rustig en D. gaat blz. 9 van *Regelrecht B* klassikaal bespreken.

D.: "Als ik die tekst lees van die speedboot... wat houdt dat in? Want dat snapt niet iedereen direct heb ik gemerkt".

Kennelijk heeft D. dit de vorige les tijdens de begeleiding van de groepjes opgemerkt.

Jongen: "Je kunt ook een half uur huren...".

D.: "En dan kost het ook?"

D. stuurt met 'ook' op 22 gulden af.

Jongen: "22 gulden".

"Nee, nee", zeggen enkele kinderen.

Er leven kennelijk verschillende interpretaties van de tekst op het bord. Een mogelijkheid tot discussie.

"Jawel", zegt D. "er staat juist 'of een gedeelte daarvan'. Je betaalt altijd 22 gulden. Je kunt hem voor een heel uur huren, maar heb je geen zin meer na een half uur, dat moet je zelf weten, maar dan moet je toch voor een uur betalen".

Maar D. geeft zelf al aan hoe de tekst geïnterpreteerd moet worden.

D.: "Dat doet die man natuurlijk, anders verdient die niet genoeg vindt die zelf. Hij denkt een uur is een mooie tijd. Willen ze korter, dan moeten ze dat zelf maar weten, maar dan kan ik 'm meestal niet meer verhuren. Dus al kom je na een kwartier terug, moet je toch die 22

gulden betalen. Daarom staat er op dat bordje '22 gulden per uur of gedeelte daarvan'.

D. maakt zijn interpretatie van de tekst aannemelijk door de context naar voren te halen en zich in de botenverhuurder te verplaatsen. Dat doet hij op een heel overtuigende manier. Maar nu is hij het die de context en het antwoord invult en niet de klas als geheel.

D. gaat verder: "Dan die gordijnstof. Eerste kwaliteit gordijnstof, 22 gulden per meter. Het kan net zo goed een prachtige lap zijn waarvan je moeder iets maakt, wat de meisjes betreft. Want zo gaat het bij ons thuis tenminste".

TWEË-EN-TWINTIG PEEK

AANBIEDING

**3 P's**

22 piek per stuk

SPEEDBOOT  
TE HUUR

22 GULDEN  
PER UUR OF  
GEDEELTE DAARVAN

EERSTE KWALITEIT  
GORDIJN STOF

f 22,- per meter

a. Welke grafiek past het beste bij welk bordje?

b. Waarom?

D.: "Maar dan gaan ze, de dochters, het liefst zelf mee. Waarom? Omdat ze dan zeggen, dan kan ik tenminste de kleur uitzoeken. Hebben ze dan altijd precies een meter of twee meter nodig? Wie is al eens mee geweest met z'n moeder?"

Weer is het D. zelf die de context verlevendigt, zelfs verandert om deze voor zijn meisjesleerlingen aantrekkelijk te maken. Hij schuwt niet er ook iets persoonlijks in te leggen. Nu daagt hij zijn leerlingen ook uit om eigen kennis en ervaringen in te brengen.

Margreet: "Je meet eerst uit wat je nodig hebt..."

D.: "Ja, wat je nodig hebt... waar haal je dat dan uit?"

Margreet: "Uit het patroon".

D.: "En wat dan?"

Margreet: "Nou, als je bijvoorbeeld 1 meter 60 bent, dan koop je dat".

D. wijst op de tekst: "Hier, 1 meter 60, waar gaat het hier om? Moet je dan 44 gulden betalen voor die lap?"

D. gaat in gesprek met Margreet en 'duwt' haar naar 1 meter 60 waar hij dan op door kan om de tekst op het bordje te interpreteren. Daarbij gaat hij wel voorbij aan het feit dat Margreet wel een erg eenvoudige stelregel hanteert om de hoeveelheid stof te berekenen.

"Nee, nee", zeggen enkele meisjes.

"Nee, natuurlijk niet", zegt D., "dan rekenen ze dat uit. Zouden we dat kunnen ...rekenen?"

D. legt verder niet uit waarom dat 'natuurlijk niet' gebeurt. Hij stuurt aan op de berekening en doet al een stap in het oplossingsproces.

Jongen: "Een zesde deel van 22".

D.: "Nee, ik zou niet een zesde deel nemen. Ik zou een ....nemen?"

Enkele leerlingen: "Tiende".

D.: "Ja, een tiende deel. Noem eens een tiende deel van 22 gulden?"

Meisje: "2 gulden en 20 cent".

D.: "Nou en dat moet je ....?"

Meisje: "13 gulden 20".

D.: "Uitstekend, dan komt er nog 13 gulden 20 bij. En dan betaal je voor die lap geen 44 gulden net zoals bij dat bootje, maar 35 gulden 20".

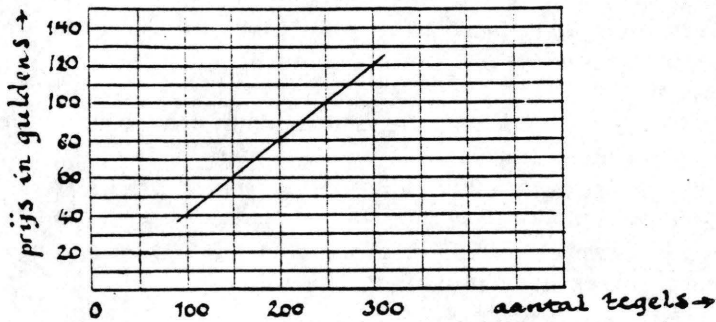
Want de rest van die 2 meter kan de stoffenverkoper wel weer kwijt, anders dan bij de botenman. Als D. consequent was geweest had hij er een dergelijke redenering aan toegevoegd. Nu gebruikt hij alleen ervaringen van de kopers. De berekening doet hij weer in gesprek met één leerling waarbij hij het is die de stappen zet.

D.: "Dan krijg je een duur jurkje deze keer, of valt dat nogal mee?"

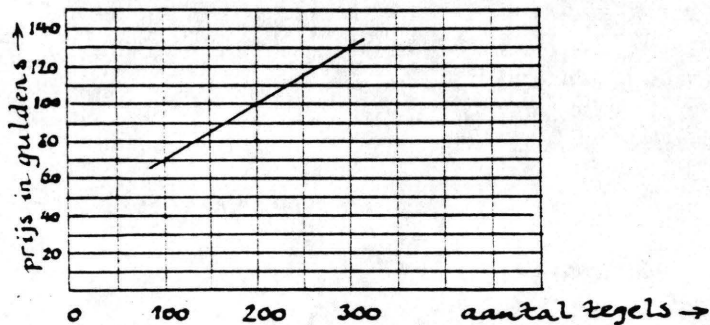
"Nou", zegt een meisje.

Toch nog even terug naar de context. D. besteedt daar duidelijk aandacht aan.





Deze grafiek laat zien hoeveel je moet betalen voor een bepaald soort tegels als je ze door firma Bouwhuis thuis laat bezorgen, afhankelijk van het aantal tegels dat je koopt. Hieronder zie je net zo'n grafiek, maar dan die van firma van Tegelen.



Als je de grafieken met elkaar vergelijkt, dan zie je dat ze niet hetzelfde zijn. Dat verschil komt doordat de ene firma géén bezorgkosten laat betalen, maar iets duurder per tegel is, terwijl de andere firma wél bezorgkosten rekent, maar een lagere tegelprijs in rekening brengt.

Een goede kennis van jullie is van plan om tegels te gaan kopen en die thuis te laten bezorgen.

Ze heeft gehoord dat jullie er iets van af weten.

Wat zouden jullie haar aanraden?

Na de klassikale start van de les wordt er bij D. in groepjes gewerkt. Er zijn 7 groepjes, 3 meisjes-, 3 jongens- en 1 gemengd groepje. D. helpt de groepjes en stimuleert uitleg aan elkaar. Ferry, Jos, Stan en Cor werken samen in een groepje. Ze zijn bezig met Bouwhuis/Van Tegelen, maar blijven steken bij de vraag waar je het beste 1000 tegels kunt kopen.

*Ferry roept D. en hij komt.*

*Jos: "Hoe moeten we dit uitrekenen?"*

*D.: "Wat denk je? Dat moet je maar gaan onderzoeken. Als je 1000 tegels gaat kopen, dan zoek je uit wat een tegel bij Bouwhuis kost en je zoekt uit wat een tegel bij Van Tegelen kost. Dus hoeveel kosten ze per stuk. Dat kun je toch wel vinden?"*

*D. loopt weg.*

*D. denkt dit groepje adequaat te helpen door een begin van een oplossingsweg voor ze uit te stippelen.*

*Jos: "Wacht effe, hierzo ... 200 stuks is 100 gulden precies, dus 1000 voor 500 precies, dank u, dank u".*

*Jos, die zijn afwezige leraar bedankt, gaat de oplossingsweg op, maar wel verkeerd. Hij rekent de bezorgkosten mee.*

*Ferry: "500, dat is niet erg veel".*

*Stan: "En dan per tegel bij Bouwhuis".*

*Jos: "Dat is die bovenste, hè?"*

*Stan: "Ja".*

*Ferry: "Wat is het nou per tegel bij de firma Bouwhuis?"*

*Stan: "Nou, dan moet je even die lijn doortrekken. Dan komt die precies bij nul uit. Dat heb ik met de geodriehoek gedaan".*

*Jos: "100 zijn 40 guldens".*

*Stan: "Als 200 nou 80 is, dan is 50 20, dan is 25 een tientje en zo kan je terugtellen als je slim bent".*

*Jos: "12 en een halve gulden".*

*Stan: "Per tegel?"*

*Jos: "Ja".*

*Stan: "Jongen, we zijn nog geen eens bij één tegel. Doe niet zo achterlijk".*

*Jos: "Ik doe niet achterlijk tegen jou!"*

*Stan: "12 en een halve voor 5 gulden".*

*Jos: "12 en een halve gulden voor 5 tegels".*

*Stan: "Nee, 5 gulden voor 12 en een halve tegel".*

*Jos: "12 en een halve tegel voor 5 gulden! Daar klopt ook geen zak van".*

*Stan en Jos stappen beiden op de oplossingsweg van D. Jos leest af wat 100 tegels kosten en maakt vervolgens een berekeningsfout. Stan leest af wat 200 tegels kosten en rekent dan terug door eerst door 4 te delen en dan door 2. Misschien dat Jos op dit rijtje van Stan is ingehaakt en een verwisselfout maakt. Bij Bouwhuis is het in ieder geval wel mogelijk zo terug te werken tot 1 tegel aangezien daar geen bezorgkosten zijn. Maar Jos en Stan raken verstrikt in onderling gekibbel en halen uiteindelijk D. er maar weer bij.*

*D. komt.*

*D.: "Ik wil jullie op weg helpen. Dan moet je hier bij Bouwhuis eens*

*kijken hoeveel 100 tegels kosten en dat doe je hier ook".*

Dit had Jos al, maar D. stelt zich niet eerst op de hoogte van de stand van zaken in het groepje.

*Ferry: "Dat is 40 gulden".*

*Jos: "Maar Van Tegelen begint ook al met 40 gulden bezorgkosten. Dat moet je toch eerst aftrekken".*

*D.: "Ja, daar moet je rekening mee houden".*

*Ferry: "Ik heb een briljant idee. Dan moet het bij Van Tegelen 70 zijn. Dat is 700 per duizend".*

*Jos: "Dat is 30 cent per tegel".*

*D.: "Ja, en maak dat nou maar eens duidelijk aan hun. En dan kom je eruit".*

D. merkt dat Jos het snapt en deze doet het nu opeens wel goed met de bezorgkosten. D. rekent er nu op dat Jos het de anderen zal uitleggen en met een kleine bemoediging gaat hij weg. Jos legt inderdaad zijn groepsgenoten de berekening uit. Ook Ferry weet hij te overtuigen dat bij Van Tegelen eerst de bezorgkosten eraf moeten. D. stimuleert uitleg aan elkaar binnen de groepjes sterk en dit komt dan ook veel voor.

Een klassikale nabespreking houdt D. regelmatig. Ook de vraag waar het groepje van Jos, Stan en Ferry mee bezig was komt in de klassikale nabespreking aan de orde.

*D.: "Ik ben van plan om 1000 tegels te gaan kopen. Waar kan ik dan volgens jullie het beste terecht? 1000 tegels".*

*Marcel: "Ik dacht dat 't niet uitmaakt, want als je voorbij de 160 bent, dan loopt 't toch hetzelfde op. In het begin is Bouwhuis goedkoper".*

Marcel bedoelt vast 160 op de y-as, dat is de y-coördinaat van het snijpunt. De lineairiteit van het verband tussen aantal en kosten doorziet hij kennelijk niet.

*D.: "Ik begrijp je probleem al. Je bent goed op weg. Jij zegt 't maakt niks uit".*

*Jos fluistert: "Maar ze kruisen elkaar, stommerd".*

*D.: "Luister nog even naar Marcel, want ik vind een stukje van zijn antwoord best wel goed. Hij zegt, in 't begin maakt 't wel wat uit. Waar zou je dan kopen, als je heel weinig tegels nodig hebt?"*

*Marcel: "Bij Bouwhuis, bij die, die geen vervoerskosten rekent".*

*D.: "Ja natuurlijk, dus bij Bouwhuis, maar nu heb ik er 1000?"*

*Marcel: "Bij Van Tegelen".*

*D.: "En kun je dat duidelijk maken, Marcel? Of weet je het nog niet? Jij zegt Van Tegelen?"*

*Marcel zwijgt.*

*D.: "Hoe kun je dat laten zien? Neem dan eens even bladzijde 10 voor je".*

*Marcel: "Die ene lijn loopt steiler dan die andere".*

Dat is een goed argument van Marcel en het is D. die door zijn bemoedigen en doorvragen Marcel zover gekregen heeft. Dat steiler lopen kan D. koppelen aan de lineairiteit, want dat lijkt het probleem bij Marcel.

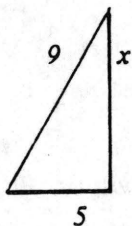
*D.: "Jazeker, ik vind dat een heel mooi antwoord. Alleen ... dat is een beetje je gevoel, hè? Neem allemaal bladzijde 10 eens voor je en dan doen we het straks precies. Wat Marcel zegt is dat die eerste lijn veel steiler omhoog loopt en die andere veel minder steil. Dus die andere is goedkoper: Van Tegelen. Op den duur is Van Tegelen goedkoper en in het begin is Bouwhuis goedkoper. Je had het bijna goed, Marcel. En je hebt al een getal genoemd waar het belangrijk verandert. Let op".*

Maar D. kiest niet voor een oplossing via de lineairiteit en het steiler lopen van de grafiek. Uit het vervolg blijkt dat hij aanstuurt op berekening van de prijs per tegel en vandaar per 1000 tegels, zoals hij ook in het groepje stimuleerde. Dit sturen naar een bepaalde oplossing bij voorkeur via rekenen doet D. vaak. Hij keurt andere oplossingswegen niet af, maar vindt ze niet voldoende, terwijl een oplossing hier via de grafiek misschien nog wel geschikter is.

#### 2.4 Klas B in beeld

Klas B bestaat uit 28 leerlingen, 17 meisjes en 11 jongens. de leerlingen hebben een LBO-, MAVO-, HAVO- of VWO-advies. Docent L., die mentor van de klas is, begint de les altijd met het bespreken van het huiswerk of van opdrachten die de vorige les gemaakt zijn. Een inleiding op een nieuw onderwerp geeft hij zelden. In het volgende fragment is L. bezig met de stelling van Pythagoras. Samen met de klas is L. geruime tijd bezig geweest met het herontdekken van de stelling. Nu zijn ze bezig de stelling toe te passen in contextrijke, maar ook contextloze situaties, zoals in het volgende fragment.

*L. zet op het bord:*



L.: "Deze is 9, die is 5, die is x".

"Ja, die wil ik wel doen", zegt Diane.

"Diane", zegt L.

Diane: "9 is  $5^2$  plus  $x^2$ ".

L. herhaalt en schrijft op:  $9 = 5^2 + x^2$

Diane: "9 is 25 ...".

L.: "9 is ...".

Diane: "25 plus x".

"25 plus x", zegt L.

Diane wil de beurt en die krijgt ze ook. Maar ze maakt fouten. L. herhaalt haar fouten neutraal om reacties uit te lokken.

" $x^2$  hoor", zegt iemand.

Diane: "x is 56".

L.: "x is 56".

"Hoe weet je dat nou?", zegt iemand.

"Dat had L. me uitgelegd, ja?", zegt Diane.

Diane: "Erne x is 7,48. Je moet er een kwadraat boven zetten".

L.: "Waar?"

"Bij 56", zegt Diane.

L. heeft op het bord gezet:

$$9 = 5^2 + x^2$$

$$9 = 25 + x$$

$$x^2 = 56$$

$$x = 7,48$$

"Nou, wat vinden jullie ervan?", vraagt L.

"Fout", wordt er gezegd.

"Oh, L. heeft die me uitgelegd", zegt Diane.

"Ja, ik heb het gedaan, hè?", zegt L.

Mieke krijgt de beurt.

"Even kijken hoor", zegt Mieke, "de eerste is goed, alleen bij 9 hoort een kwadraat te staan".

"Ja, daar hoort nog een kwadraat te staan", zegt L.

"Oh, dat heb ik hier ook", zegt Diane.

"Dat heb ik zeker uitgelegd zo", zegt L.

Gelach.

"Bij de tweede 9 hoort 81 te staan", zegt Mieke.

"Juist, heel goed", zegt L., "want daar moet je dat kwadraat uitrekenen. Dan komt het als een wonder bijna goed. Maar er staat nog een foutje in die regel".

Mieke: "Hoort nog bij die x een kwadraat te staan?"

L.: "Hoort ook nog daar een kwadraat. En dan is het goed, dan is x inderdaad 56".

"Wortel 56", zegt iemand.



L.: "En  $x$  is dus de wortel van 56. Dat moet er eigenlijk nog even tussen".

In tweegesprek met Mieke wordt Dianas oplossing verbeterd.

L. stimuleert voortdurend inbreng van zijn leerlingen waardoor deze zich vrij voelen om te reageren en L. de kans krijgt achter hun oplossingen te komen.

"Hè?", Arne snapt het niet.

"Ik heb het simpeler eigenlijk", zegt Patrien.

"Het is toch 81?", zeggen een paar.

L.: "En dan?"

"En dat is dan 7,48", zegt een jongen.

L.: "Wie heeft het anders gedaan, Patrien?"

Patrien: "Ik had gewoon die  $9^2$ , dat is de schuine zijde, dat is gewoon 81, heb ik 81 is 25, dus die 5 is 25,  $5^2$ , en dan doe ik gewoon 25 van de 81 aftrekken en dan weet je wat  $x$  is".

L.: "Dat doet zij in feite ook, maar hoe schrijf je het op? Die 56 krijgen Diane en Mieke ook door 81 min 25 te berekenen, maar hoe schrijf je het op?"

Patrien: "Ik zou opgeschreven hebben  $9^2$  is 81...".

L.: "Maar wat heb je opgeschreven?"

Patrien: " $x, x^2$  is eh ...".

L.: "Kijk, nu wordt het iets heel anders, zeggen hoe je iets uitrekent, zeggen hoe iets moet is heel anders dan het goed opschrijven en daar trainen we ons nu een beetje in dit blok op, want ik heb wel het idee dat de meeste van jullie wel weten hoe die stelling van Pythagoras werkt en ook uiteindelijk tot een goed antwoord kunnen komen. Maar het goed opschrijven zodat er geen fouten in staan, is toch wel een moeilijke zaak".

En daarmee expliciteert L. het doel van de les, het komen tot een goede neerslag van een oplossingsweg. Dit gebeurt in een open sfeer waarin onder L's leiding de leerlingen elkaar verbeteren zonder elkaar af te kraken. Diane is in ieder geval niet afgeschrokken, want even later geeft ze zelf de stoot tot meer algemene principes om de hypotenusa of een van de rechthoekszijden te berekenen.

Diane: "Hoe weet je nou dat je op moet tellen of af moet trekken?"

L. herhaalt Dianas vraag.

"Staat erboven", zegt Patrien.

"Ja, is toch makkelijk", zegt Arne.

Patrien: "Hypotenusa is de schuine zijde".

"Ja", zegt Arne, "dit plus dit is altijd die, snap je?"

"Ja, weet ik ook wel", zegt Diane.

"Arne?", vraagt L.

Arne: "Rechthoekszijde plus rechthoekszijde is altijd eh, bij een rechte

*dan, hypothenusa".*

*"Ja, de kwadraten van de rechthoekszijden", zegt L.*

*"Ja", zegt Arne.*

*Diane: "Dus als je de hypothenusa hebt en een rechthoekszijde, dan moet je aftrekken".*

*"Precies", zegt L.*

En zo heeft Diane uiteindelijk zelf haar eigen vraag beantwoord.

Na de klassikale start wordt er in groepjes gewerkt. Er zijn 7 groepjes, 6 gemengd en 1 met 5 meisjes. L. is geen voorstander van het afdwingen van samenwerking en dat weerspiegelt zich in de hulp die hij zijn leerlingen geeft. Deze is voornamelijk individueel gericht. Uitleg aan elkaar stimuleert hij nauwelijks. Binnen de groepjes werken de leerlingen individueel of meestal 2 aan 2. Er is ook overleg van individuele leerlingen tussen verschillende groepjes. Dus er is wel veel overleg in de klas maar niet binnen de groepjes als geheel.

Donald, Michiel, Cecile en Estelle zitten bij elkaar. Cecile en Estelle zijn al bijna klaar met Regelrecht B. Donald en Michiel hebben er meer moeite mee. Zij zijn bij bladzijde 14 en roepen zoals gebruikelijk niet de hulp van Estelle en Cecile in maar van L.

*Cecile is bezig met de rekenmachine. L. komt en kijkt bij Donald en Michiel.*

*"Wat kun je d'r van zeggen?", vraagt Michiel aan L. "Je kunt alleen zeggen dat je extra hoeft te betalen".*

Michiel bedoelt dit als antwoord op 14b en ziet kennelijk dat alle grafieken boven de oorsprong beginnen en trekt de goede conclusie dat er dus extra kosten zijn. Alleen hier ging het niet om de interpretatie van dit kenmerk van de grafiek maar om het onder woorden brengen van het kenmerk zelf.

*L.: "Als er geen extra kosten zijn, hoe loopt die grafiek dan, wordt er gevraagd".*

L. wil dat ze 14a beantwoorden, daar hadden ze nog geen antwoord staan.

*"Moet je dan gaan tekenen", vraagt Michiel.*

*L.: "Nou, je kunt het ook omschrijven, je kunt ook zeggen hoe die lijn loopt".*

*Cecile: "Je weet toch, als er geen extra kosten zijn, dan begint die lijn toch bij nul?"*

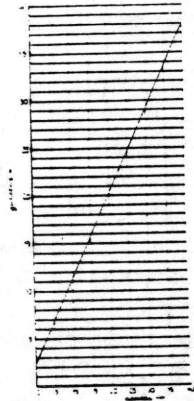
Helpt Cecile nu omdat L. erbij is? Ze verwoordt het in ieder geval uitstekend.

*"Oh ja ... ja, dat is heel goed!", zegt Michiel.*

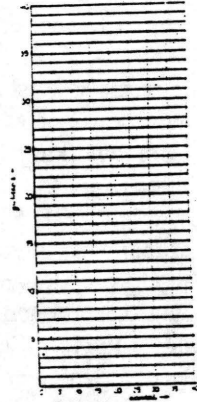
Haast een Aha-erlebnis. Precies wat je wilt bij groepswork. De een vindt woorden voor datgene waar de andere al dichtbij is.

De grafieken zagen er zo uit:

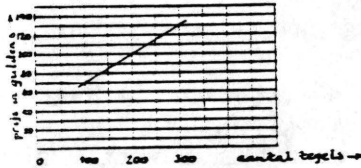
fotonuis Polak:



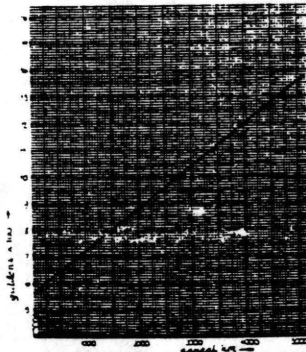
of



firma van Tegelen:



'Toyota':



Het hangt van de hoogte van de extra kosten af waar de grafiek 'begint'.

- a. Wat kun je zeggen over de grafiek als er *geen* extra kosten zijn?
- b. Kun je iets 'in 't algemeen' zeggen over dit soort kostengrafieken?

*L. is weer weg.*

Verwacht hij dat Cecile's hulp voldoende is? Of verwacht hij dat het overleg nu verder wel loopt?

*"Weet je hoe het gaat?", vraagt Cecile, "70,70 cent".*

Maar Cecile is alweer bij blz. 15 en richt zich op Estelle.

*"Waarom zeg je dat nou", zegt Estelle, "kan ik het zelf niet uitrekenen".*

Ook de samenwerking tussen die twee is niet optimaal.

*"Ze moeten allemaal extra kosten betalen", zegt Michiel.*

Michiel ziet wel een gemeenschappelijk kenmerk van de grafieken, maar interpreteert deze weer. In feite is hij bij zijn oude antwoord terug.

*"B.T.W. heet zoiets", zegt Donald.*

*"Oh zeggen", leest Donald bij 14b, "dus we schrijven niks op, je moet het zeggen".*

*Michiel lacht.*

Even later hebben Estelle en Cecile het boekje uit en krijgen extra werk van L. L. verplicht de 'betere' kinderen tot het maken van extra werk. Het SLO-materiaal komt toch al zo tegemoet aan de zwakkeren, vindt hij. Hij verwacht juist problemen voor de kinderen die naar HAVO of VWO gaan. Die zouden met dit materiaal tekort komen. Cecile en Estelle gaan dus met ander materiaal verder en daarmee wordt overleg binnen dit groepje als geheel onmogelijk.

Donald en Michiel blijven bij 15a hangen.

L. houdt zelden een klassikale nabespreking. Alleen aan het eind van bovenstaande les houdt hij een korte bespreking.

*"Goed dames en heren", zegt L., "over ruim 5 minuten is het tijd. Het huiswerk voor de volgende keer is het blok af. Ik wil graag even weten ..."*

*"Wat krijgen we hierna?", wordt er gevraagd.*

*"Vergelijkingen", zegt L.*

*"Oh Jezus!", zegt Estelle.*

*"Dat vind ik leuk", zegt Cecile, "met x en y is dat".*

*"Luister nog eens eventjes", zegt L., "want ik was nog niet uitgesproken. Hoe is dit blok jullie bevallen?"*

*"Erg moeilijk", "ja", "ja", "nee".*

*"Nee hoor!", roept Cecile.*

*"Wie vond het erg moeilijk?", vraagt L.*

*Donald en Michiel steken hun vinger op.*

*Anderen ook.*

*L.: "Kun je ook uitleggen waarom?"*

*"Sommige dingen", wordt er gezegd.*

*"Moeilijk om het uit te leggen", zegt iemand.*

*"De vragen waren niet moeilijk, maar de antwoorden wel", zegt een jongen achterin.*

"De vragen waren niet moeilijk maar de antwoorden wel", herhaalt L.

## 2.5 Reflectie

Als we kijken hoe docent D. met klas A en docent L. met klas B 'Wiskunde voor Iedereen' in de praktijk brengen dan zien we dat de klas het gedrag van de docent beïnvloedt, maar omgekeerd heeft de docent ook duidelijk invloed op het gedrag van de klas.

Klas A bijvoorbeeld is een drukke klas. Dat beïnvloedt D. Hij begint de les met orde houden en beperkt de inbreng van zijn leerlingen sterk. Aan het eind van de les als de orde veel minder een probleem is krijgen zijn leerlingen, zoals Marcel, veel meer ruimte om hun oplossingen te vertellen. Maar ongeacht de drukte, aan de *contexten* besteedt D. veel aandacht, dat vindt hij kennelijk zelf een belangrijk aspect van 'Wiskunde voor Iedereen'. *Verschillen in oplossingen* daarentegen vindt D. kennelijk minder belangrijk. Sterker nog, het lijkt erop dat hij ervan overtuigd is dat uiteindelijk één oplossing de goede is. Want drukte of niet, hij stuurt naar één bepaalde oplossing. Dit is natuurlijk ook van invloed op het leerproces binnen de kleine heterogene groepen. Hij stimuleert *samenwerking binnen kleine heterogene groepen*. Dat vindt hij belangrijk en er wordt ook veel samengewerkt. Maar doordat *verschillen in oplossingen* te weinig aandacht krijgen beperkt D. die samenwerking tot uitleggen en daarmee wordt de in het SLO-curriculum gehanteerde differentiatievorm onvoldoende gerealiseerd.

Docent L. is mentor van klas B. Hij heeft een goede band met zijn leerlingen en geeft ze veel inbreng. En zoals Diane laat zien verwachten de leerlingen die inbreng ook. Aan het verlevendigen van de *contexten* besteedt L. weinig aandacht. Alleen als hij merkt dat de geschreven tekst tot misverstanden leidt grijpt hij in met een klassikaal moment. Aan *verschillen in oplossingen* van zijn leerlingen en het elkaar becommentariëren besteedt L. tijdens klassikale momenten veel aandacht en daarmee komt dit kenmerk sterk naar voren. Maar binnen de kleine heterogene groepen wordt dit kenmerk onvoldoende gerealiseerd doordat L. de samenwerking binnen kleine heterogene groepen niet organiseert. Ook bij hem wordt dus de in het SLO-curriculum gehanteerde differentiatievorm onvoldoende gerealiseerd.

Beide docenten leggen sterk hun eigen accenten in het realiseren van 'Wiskunde voor Iedereen'. Ze laten zich daarbij ten dele leiden door de klas waarmee ze te maken hebben. Veel sterker lijkt hun eigen visie op wiskunde een rol in die keuze te spelen. Docent D. probeert zeer bewust een aantal kenmerken van 'Wiskunde voor Iedereen' te realiseren, maar de kern van de differentiatie, de niveaus in het leerproces, realiseert hij



niet. Misschien is dit wel in strijd met zijn visie op wiskunde (Thompson, 1984). Docent L. geeft de kern tijdens klassikale momenten de volle aandacht, maar verwaarloost de andere kenmerken van 'Wiskunde voor Iedereen'. Het lijkt erop of hij het met een ander curriculum in de hand niet veel anders zou doen. Kortom, zijn visie op wiskunde lijkt in hoge mate zijn lesgeven te bepalen, veel meer dan de kenmerken van 'Wiskunde voor Iedereen'.

Opvallend is dat beide leerkrachten geen *extra didactische maatregelen* treffen voor leerlingen die niet goed meekomen, die bijvoorbeeld een toets slecht gemaakt hebben. Dit is voor beiden kennelijk ongebruikelijk of te ingewikkeld om te organiseren. Het SLO-materiaal zoals het er ligt biedt ook niet de mogelijkheid om met deze verschillen tussen kinderen anders dan via oplossingswegen rekening te houden.

Dit hoofdstuk is aan docent D. van klas A en docent L. van klas B voorgelegd. Beiden waren accoord met de door ons gemaakte portretten en analyse.



### 3. PERCEPTIES VAN HET ONDERWIJSLEERPROCES

#### 3.1 Vraagstelling

Eén van de methoden om het onderwijsleerproces te beschrijven is door gebruik te maken van percepties van leerlingen. Door leerlingen te vragen wat zij gezien en gehoord hebben, krijgt men een beeld van het klassegebeuren. Het is een beschrijving op basis van waarnemingen van de betrokkenen i.c. de leerlingen. De vraagstelling is: hoe percipiëren de leerlingen het onderwijsleerproces? De leerlingpercepties geven een indicatie voor de mate waarin de kenmerken van het nieuwe curriculum zijn uitgevoerd zoals bedoeld. Door vergelijkingen te maken tussen gemiddelde perceptiescores, van bijvoorbeeld klassen of scholen, krijgen we zicht op verschillen in de implementatie van bepaalde kenmerken. Vergelijkingen tussen project scholen en vergelijkingsscholen kunnen gegevens opleveren over verschillen in de onderwijsleerprocessen in de klas tussen beide condities. Een tweede vraag is dus: zijn er verschillen in perceptie tussen de leerlingen in de verschillende klassen, scholen en condities?

#### 3.2 Operationalisering

Als methode van onderzoek is gekozen voor de schriftelijke vragenlijst. De methode maakt het mogelijk alle leerlingen in het onderzoek te betrekken. We construeerden de zogenaamde PERCIA vragenlijst (PERceptie van het Curriculum-In-Actie).

Deze vragenlijst vertoont overeenkomsten met elders gebruikte vragenlijsten. Het betreft perceptieschalen voor het meten van de leeromgeving. (vgl. Fraser 1981, Fischer en Fraser 1983). Er is een toenemende, internationale belangstelling voor het meten van de 'leeromgeving' door middel van leerlingpercepties. Dit blijkt bijvoorbeeld uit de oprichting van een Special Interest Group van de AERA onder leiding van B. Fraser. In verschillende Nederlandse onderzoeksprojecten zijn schalen ontwikkeld voor het meten van leerlingpercepties (vgl. Wierstra 1986 en Van der Grift en Mulder 1986, Wubbels e.a. 1987).

De PERCIA vragenlijst is qua vorm geïnspireerd op deze leeromgevingsschalen. Inhoudelijk is echter een invulling gemaakt die aansluit bij de vernieuwende kenmerken van het onderzochte curriculum. Het betreft vooral de kenmerken die betrekking hebben op de werkvormen of interactievormen in de klas. Er is echter ook een kenmerk betreffende de wiskundig didactische inhoud opgenomen: het gebruik van wiskunde in

contexten. De hieronder gepresenteerde subschalen worden onderscheiden. Zie voor de volledige lijst met items bijlage 1.

#### *Leraar, instructie, klasseklimaat (LIK)*

Deze subschaal betreft de leerkracht en de klas als geheel. De items betreffen aspecten van leraarsgedrag, instructie en klasseklimaat die in vele onderzoeken effectief zijn gebleken. Kort samengevat gaat deze subschaal over de mate waarin bepaalde activiteiten van de leraar worden gerealiseerd en over kwaliteitsaspecten van die activiteiten. Daarbij gaat het om zaken als: duidelijke (klassikale) uitleg, goede leiding, overzicht op de voortgang van de leerlingen, feedback. Daarnaast zijn er items over het klimaat: vriendelijke, begrijpende leraar, prettige sfeer in de klas.

#### *Begeleiding van individuele leerlingen en groepen (BIG)*

Deze subschaal bevat items over de mate waarin individuele leerlingen en groepjes worden begeleid door de leraar. Daartoe behoort controle op de individuele voortgang, aandacht voor zwakke leerlingen en begeleiding van groepjes.

#### *Samenwerking tussen leerlingen (SAM)*

In dit cluster wordt gevraagd naar de samenwerking in de groepjes en in de klas. Het gaat om zaken als: hulp geven, hulp vragen, elkaar begrijpen, aandacht voor verschillen in oplossingswijze.

#### *Taakgerichtheid en orde (TAKO)*

In deze subschaal gaat het om aspecten als orde, rumoer, verveling, systematiek bij het werken en tijdsbesteding (time on task). Het betreft een dimensie die heel belangrijk is, juist bij onderwijs waarbij in heterogene klassen en heterogene groepjes wordt gewerkt.

#### *Inhoud (IN)*

Deze subschaal betreft de wiskundig-didactische inhoud zoals deze in de opdrachten en de lessen naar voren komt. Een toespitsing is gemaakt op het gebruik van 'contexten'. De items gaan over aansluiting van de contexten bij de leefwereld en de voorkennis van de leerlingen, de aantrekkelijkheid en de duidelijkheid van de contexten.

De PERCIA - vragenlijst bestaat uit 54 items verdeeld over vijf subscha-

len. Twee subschalen betreffen de leraar, twee subschalen gaan over de leerlingen en één betreft de leerinhoud. De subschalen hebben de volgende inhoud (zie ook bijlage 1)

1. Leraar, instructie en klasseklimaat (LIK)
2. Begeleiding van individuele leerlingen en groepjes (BIG)
3. Samenwerking tussen leerlingen (SAM)
4. Taakgerichtheid en orde (TAKO)
5. Inhoud (IN)

De items hebben de volgende vorm:

	bijna niet	niet zo vaak	soms	vaak	zeer vaak		
Wij helpen elkaar bij het zoeken van een oplossing.	Dat komt	A	B	C	D	E	voor

De indeling in subschalen is primair op inhoudelijke gronden tot stand gekomen. Daarbij zijn twee hoofdbronnen gebruikt:

1. De vier SLO kenmerken:
  - a. klassikale introducties en reflecties in de (heterogene) klas als geheel (zie LIK)
  - b. samenwerking tussen de leerlingen in (heterogene) groepjes (zie SAM)
  - c. niveaus in het leerproces
  - d. rijke contexten (zie IN)
2. Researchgegevens uit literatuur. De volgende aspecten hebben we bij de verbreding en operationalisering van de SLO kenmerken bij de itemconstructie betrokken:
  - e. het klimaat in de klas als geheel (zie LIK)
  - f. de begeleiding door de leraar (zie BIG)
  - g. de taakgerichtheid (zie TAKO).

Een combinatie van de vier kenmerken van de SLO a, b, c en d met de andere drie aspecten e, f en g leverde de indeling op voor de PERCIA-schaal. De schaal bevat dus uitdrukkelijk de SLO-kenmerken. Bij de uitwerking tot een onderzoeksinstrument zijn daarnaast enkele andere



effectief gebleken principes van goed onderwijs betrokken. Deze kunnen als aanvulling op, respectievelijk als uitwerking van, de SLO-kenmerken worden beschouwd. Zoals gezegd in hoofdstuk 1, was het niet eenvoudig kenmerk c (niveaus) in de Percia adequaat te operationaliseren. Het is misschien wel een te moeilijk begrip om aan de leerlingen te vragen. We zijn er niet in geslaagd een volledige subschaal NIVEAUS samen te stellen. Wel komen binnen LIK, BIG, SAM en IN enkele items voor die betrekking hebben op het niveauprincipe. We hebben via diepgaande kwalitatieve observaties het niveauprincipe beschreven (zie hoofdstuk 2).

Deze indeling in vijf subschalen gaat terug op een eerdere indeling in vier subschalen (Terwel, 1986c). Aan deze vier subschalen is een vijfde subschaal Inhoud toegevoegd, terwijl het aantal items binnen de vier subschalen is uitgebreid. Bij de analyses zijn de schaaleinden van de negatief geformuleerde items omgedraaid. Na uitvoering van betrouwbaarheidsanalyses en factoranalyses zijn enkele items verwijderd, bijvoorbeeld items die onvoldoende betrouwbaar waren en/of onvoldoende laadden. Enkele andere items zijn verplaatst van de ene naar de andere subschaal. Zo ontstond de uiteindelijke schaal met 5 subschalen en 54 items (zie bijlage 1).

Bij het vergelijken van de perceptie van projectleerlingen en vergelijkingsleerlingen doet zich een moeilijkheid voor. De PERCIA-schaal is inhoudelijk toegespitst op de kenmerken van het experimentele curriculum. Is het wel terecht om de vergelijkingsklassen en scholen te meten met een meetlat die op een ander concept is gebaseerd?

Het antwoord is dat dit inderdaad niet geheel terecht is, in het bijzonder daar waar het de specifieke wiskundig-didactische inhoud betreft, dat wil zeggen bij subschaal 5, betreffende wiskunde in zogenaamde contexten. In deze subschaal komen enkele specifieke items voor die niet passen bij het curriculum van de vergelijkings scholen. Bij de interpretatie van de gegevens moet daar rekening mee worden gehouden. De items in de overige subschalen zijn in principe ook van toepassing op meer 'traditioneel' onderwijs, zij het dat de mate waarin een kenmerk voorkomt in beide condities zal verschillen. Bijvoorbeeld de frequentie van samenwerking tussen de leerlingen (subschaal 3).

Het overgrote deel van de items in de subschalen 1, 2, 3, en 4 betreft het voorkomen van bepaalde werk- en interactievormen en de kwaliteit daarvan. De scores op deze subschalen geven dus vooral een indicatie van de mate van implementatie van effectief gebleken (of effectief veronderstelde) kenmerken of interactievormen. Wel is het zo dat in de projectklassen bepaalde 'kenmerken' nadrukkelijker worden nagestreefd

en met curriculummaterialen worden ondersteund, zoals opdrachten voor leerlingen en een handleiding voor leerkrachten. Het project 'wiskunde voor iedereen' beoogt onder meer de frequentie en de kwaliteit van bepaalde onderwijsleerprocessen of interactievormen te verhogen. De mate waarin dat is gerealiseerd proberen we te meten met de PERCIA. Vergelijking met scholen die volgens een ander concept te werk gaan is zinvol, omdat algemene kwaliteitscriteria aan de items ten grondslag liggen. Een vergelijking van condities levert informatie op over de mate waarin de bedoelingen of kenmerken van de leerplanontwikkelaars zijn gerealiseerd mede in vergelijking met de realisatie van deze bedoelingen of kenmerken in vergelijkingsscholen. Voorzichtigheid bij de interpretatie is echter geboden.

### 3.3 Betrouwbaarheid

De betrouwbaarheid van de totale schaal en de subschalen is berekend met behulp van alpha-coëfficiënten zowel op leerlingniveau als op klasseniveau. Bovendien is de coëfficiënt van Horst (1949) bepaald. De Horst-coëfficiënt geeft een indicatie van de mate waarin de subschalen discrimineren tussen klassen in relatie tot verschillen binnen klassen.

*Tabel 3.1: Betrouwbaarheid Percia-schalen*

Subschaal	aantal items	alpha (II) N = 800	alpha (klas) N = 33	Horst N = 33
LIK	15	.85	.94	.93
BIG	11	.64	.74	.84
SAM	12	.65	.84	.85
TAKO	11	.70	.81	.86
IN	5	.63	.83	.81
Totaalschaal	54	.85	.94	.92

Uit de gegevens in tabel 3.1 is te concluderen dat de interne consistentie van de PERCIA op leerlingniveau over het geheel gezien voldoende is. De betrouwbaarheid op leerlingniveau van LIK en TAKO is redelijk tot goed. De interne consistentie van overige subschalen (BIG, SAM en IN) is aan de lage kant maar toch nog voldoende.

De interne consistentie (alpha) op klasseniveau, gebaseerd op klassegemiddelden, is hoog. Ook de Horst-coëfficiënten zijn hoog. De klassegemid-

delden geven dus een betrouwbaar beeld van het gebeuren in de klas. Het aggregeren van individuele scores naar het klasseniveau is echter een ingewikkeld vraagstuk (vergelijk Koehler en Van den Eeden 1987).

### **3.4 Procedure en analyse**

De vragenlijst werd afgenomen tegen het einde van de periode waarin het curriculum werd uitgevoerd. Voor alle subschalen zijn de gemiddelden en standaarddeviaties bepaald voor verschillende niveaus zoals klas, school en conditie. Tevens zijn one way-variantie-analyses uitgevoerd ter bepaling van het al of niet significant zijn van gevonden verschillen.

### **3.5 Resultaten**

#### **3.5.1 Gemiddelden en correlaties**

We presenteren in deze paragraaf eerst gemiddelde scores van leerlingpercepties. Een globale vergelijking van verschillen in gemiddelde leerlingpercepties geeft een indicatie van verschillen in onderwijsleerprocessen tussen klassen, scholen en condities.

In tabel 3.2 staan de gemiddelde scores op de subschalen per klas (klasnummer 11-18, 21-26, 31-39, 41-48 en 51-53).

Tabel 3.2: *Klassegemiddelden en standaardafwijkingen per subschaal*

		PERCIA totaal		LIK		BIG		SAM		TAKO		IN	
N	Klas	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
835	Totaal	3.26	.36	3.64	.62	2.39	.50	3.54	.47	3.40	.55	3.04	.64
521	Tot.	3.33	.34	3.66	.59	2.47	.48	3.67	.45	3.43	.50	3.14	.64
projectklassen													
20	11	3.26	.14	3.60	.34	2.45	.32	3.64	.35	3.25	.48	3.14	.46
16	12	3.35	.18	3.75	.24	2.60	.40	3.68	.46	3.41	.44	2.84	.60
20	13	3.24	.27	3.53	.51	2.61	.44	3.55	.42	3.14	.33	3.26	.56
20	14	2.99	.47	3.09	.93	2.48	.59	3.35	.36	3.02	.57	2.92	.54
20	15	3.32	.34	3.82	.41	2.50	.54	3.54	.65	3.43	.64	2.89	.60
19	16	3.06	.30	2.74	.49	2.46	.50	3.78	.50	3.44	.28	2.79	.69
13	17	2.98	.31	3.09	.57	2.57	.62	3.31	.36	2.82	.40	3.05	.53
15	18	2.96	.34	3.14	.57	2.60	.35	3.23	.47	3.01	.48	2.49	.67
28	21	3.46	.29	4.04	.36	2.58	.39	3.64	.46	3.50	.42	3.18	.51
27	22	3.41	.26	3.78	.43	2.62	.54	3.71	.37	3.47	.40	3.25	.62
26	23	3.36	.24	3.81	.41	2.23	.41	3.61	.30	3.81	.36	2.95	.59
28	24	3.55	.28	3.84	.43	2.99	.42	3.72	.45	3.59	.43	3.45	.56
23	25	3.36	.21	3.88	.37	2.33	.36	3.52	.46	3.63	.33	3.10	.71
26	26	3.57	.19	4.15	.34	2.63	.43	3.56	.41	3.75	.36	3.45	.49
23	31	3.26	.25	3.50	.44	2.42	.46	3.69	.31	3.51	.47	2.82	.56
31	32	3.35	.24	3.76	.40	2.27	.43	3.82	.53	3.42	.37	3.20	.53
32	33	3.31	.34	3.48	.52	2.35	.40	3.85	.42	3.39	.56	3.43	.67
20	34	2.88	.37	2.98	.76	2.10	.41	3.56	.31	2.79	.61	2.84	.74
26	35	3.12	.30	3.42	.45	2.17	.38	3.52	.39	3.26	.54	2.99	.65
32	36	3.47	.26	3.88	.35	2.33	.35	3.91	.40	3.59	.34	3.46	.67
29	38	3.56	.25	3.99	.35	2.55	.52	3.94	.35	3.64	.30	3.36	.55
27	39	3.63	.28	4.18	.42	2.60	.42	3.92	.42	3.77	.37	3.26	.58
314	Tot.	3.14	.37	3.59	.67	2.25	.52	3.32	.42	3.34	.62	2.86	.60
vergelijkingsklassen													
26	41	3.13	.25	3.71	.33	2.42	.65	3.22	.40	3.06	.63	2.92	.57
23	42	3.11	.24	3.66	.38	2.21	.34	3.19	.25	3.26	.59	2.91	.52
26	43	3.07	.55	3.57	.87	2.32	.61	3.20	.49	3.06	.72	2.90	.54
29	44	2.79	.38	2.88	.61	1.99	.49	3.28	.47	2.91	.69	2.87	.66
31	45	3.22	.36	3.68	.54	2.16	.53	3.33	.42	3.61	.49	3.12	.65
30	46	3.30	.28	3.91	.52	1.99	.41	3.44	.40	3.75	.46	3.05	.53
33	47	3.22	.28	3.62	.55	2.14	.37	3.50	.49	3.48	.61	3.14	.51
31	48	2.81	.33	2.77	.62	1.94	.47	3.38	.35	3.11	.52	2.86	.54
28	51	3.28	.31	3.96	.45	2.49	.42	3.31	.51	3.47	.59	2.49	.52
29	52	3.17	.26	3.70	.47	2.40	.41	3.30	.33	3.48	.50	2.29	.49
28	53	3.40	.22	4.13	.47	2.78	.28	3.25	.30	3.43	.53	2.89	.53

Bij een eerste, globale inspectie van de klassegemiddelden in tabel 3.2 zien we dat de klassegemiddelden in de projectscholen hoger liggen dan in de vergelijkingsscholen. Dit wordt het duidelijkst zichtbaar in de gemiddelden tussen de condities die ook in deze tabel zijn opgenomen. Voorts valt op dat school 2 (in de projectconditie) relatief hoog scoort op de meeste subschalen. Alvorens deze verschillen nader te verkennen en op significantie te toetsen gaan we kijken naar correlaties tussen subschalen.

Deze correlaties geven informatie over de samenhang tussen de subschalen van de PERCIA-vragenlijst. We presenteren tabel 3.3, 3.4 en 3.5, waarin de correlaties tussen de subschalen op leerlingniveau zijn opgenomen.

*Tabel 3.3: Correlaties tussen subschalen op leerlingniveau, hele bestand (N = 835)*

---

Totale groep

---

	LIK	BIG	SAM	TAKO	IN
LIK	-				
BIG	.37*	-			
SAM	.23*	.16*	-		
TAKO	.50*	.09*	.18*	-	
IN	.33*	.24*	.16*	.24*	-

---

*Tabel 3.4: Correlaties tussen subschalen op leerlingniveau, projectleerlingen (N = 521)*

---

Project leerlingen

---

	LIK	BIG	SAM	TAKO	IN
LIK	-				
BIG	.29*	-			
SAM	.28*	.13*	-		
TAKO	.50*	.07	.19*	-	
IN	.42*	.22*	.10*	.29*	-

---



Tabel 3.5: Correlaties tussen subschalen op leerlingniveau, vergelijkingsleerlingen (N = 314)

Vergelijkingsleerlingen					
	LIK	BIG	SAM	TAKO	IN
LIK	-				
BIG	.47*	-			
SAM	.15*	.04	-		
TAKO	.50*	.09*	.12*	-	
IN	.18*	.18*	.08	.15*	-

De correlaties in tabel 3.3, 3.4 en 3.5 die van een ster zijn voorzien zijn significant op ten minste 5%-niveau. Meestal betreft het echter hogere significantieniveaus  $p < .001$ . Op leerlingniveau bestaan er dus significante correlaties tussen vrijwel alle subschalen onderling. Dat is ook te verwachten gezien de inhoud van de subschalen. Het betreft subschalen die dimensies van één en dezelfde onderwijsleersituatie betreffen. De correlaties zijn doorgaans niet zeer hoog. Dat betekent dat het zinvol is de subschalen als afzonderlijke gehelen in de analyses te betrekken. Overigens maken we in latere analyses ook gebruik van de totale PERCIA-schaal. Deze totaalscore hanteren we als maat voor de mate waarin het programma volgens de bedoelingen is gerealiseerd.

De PERCIA-schaal meet een vijftal aspecten van de onderwijsleersituatie waarmee te werken valt gelet op interne consistentie van en intercorrelaties tussen subschalen.

Inhoudelijk zijn twee opvallende resultaten te melden.

(1) Tussen LIK en TAKO bestaan relatief hoge correlaties die opvallend stabiel zijn over de condities. Er is dus een relatie tussen de mate waarin de leraar duidelijk leiding en instructie geeft aan de klas als geheel (LIK) en de mate waarin leerlingen taakgericht werken (TAKO). Dit is een uitkomst die op grond van gegevens uit de vorige fase te verwachten was.

(2) Relatief laag is de correlatie tussen BIG en TAKO. Ook dit verband is stabiel over de condities. Begeleiding van individuele leerlingen en groepen gaat niet gepaard met een hoge taakgerichtheid van de leerlingen. Ook deze, op het eerste gezicht paradoxale, uitkomst hebben we eerder gevonden. Mogelijk speelt hier een rol dat begeleiding van individuele leerlingen en groepen ook wordt gegeven als reactie op een gebrek aan

taakgerichtheid. Ook is het denkbaar dat een leraar die relatief vaak begeleiding aan individuele leerlingen en groepjes geeft het overzicht over de klas als geheel verliest, waardoor de taakgerichtheid van de leerlingen die op dat moment géén begeleiding krijgen afneemt.

Nu volgen de correlaties op klassenniveau. Het betreft correlaties tussen klassegemiddelden op de 5 subschalen, voor alle klassen gezamenlijk en voor de klassen in de twee condities afzonderlijk. Bij een analyse van deze correlaties is te zien dat er tussen de condities enkele belangrijke overeenkomsten en verschillen bestaan.

*Tabel 3.6: Correlaties tussen klassegemiddelden, alle klassen (N = 33)*

	LIK	BIG	SAM	TAKO
LIK	-			
BIG	.40*	-		
SAM	.26	.23	-	
TAKO	.74*	.18	.50*	-
IN	.36*	.19	.62*	.35*

*Tabel 3.7: Correlaties tussen klassegemiddelden projectklassen (N = 22)*

	LIK	BIG	SAM	TAKO
LIK	-			
BIG	.27	-		
SAM	.47*	.02	-	
TAKO	.78*	.17	.63*	-
IN	.62*	.26	.59*	.47*

Tabel 3.8: Correlaties tussen klassegemiddelden, vergelijkingsklassen (N = 11)

	LIK	BIG	SAM	TAKO
LIK	-			
BIG	.70*	-		
SAM	-.09	-.49	-	
TAKO	.66*	.11	.54*	-
IN	-.08	-.39	.29	.04

De correlaties met een ster zijn alle significant  $p < .05$ .

Ook op dit niveau is er een hoge correlatie tussen LIK en TAKO. Deze correlatie is constant over de condities. Een hoge constante correlatie is er ook tussen SAM en TAKO. We kunnen deze correlaties als volgt verwoorden. (a) In klassen waarin de leraar goed leiding geeft en duidelijk uitlegt, werken de leerlingen vaker taakgericht dan in klassen met een minder goede leiding en instructie. (b) In klassen waarin de leerlingen adequaat samenwerken, is minder vaak sprake van wanorde, rumoer en verveling, dan in klassen waarin de leerlingen minder goed samenwerken.

Er zijn ook enkele opvallende verschillen tussen de condities.

- (a) Bij de vergelijkingsklassen is er een hoge correlatie tussen LIK en BIG (.70); bij de projectklassen is dit verband veel lager (.27). In de vergelijkingsklassen gaat een goede leiding en klassikale instructie door de leraar vaker gepaard met begeleiding van individuen en groepjes dan in de projectklassen.
- (b) Opvallende verschillen tussen de condities doen zich ook voor bij INHOUD (IN). De hoge, meestal significante, correlaties met alle andere gemiddelde waarden bij de projectconditie (zie laatste regel tabel 3.7) maken in de vergelijkingsconditie plaats voor veel lagere en soms zelfs negatieve correlaties.
- (c) In de project-klassen gaat een hoge score op proceskenmerken (LIK, BIG, SAM, TAKO) meestal gepaard met een hoge score op inhoud (en omgekeerd). Met andere woorden er is een duidelijk verband tussen de kwaliteit van het proces en de kwaliteit van de inhoud. Bij de

vergelijkingscholen is dat verband niet aanwezig of niet zo duidelijk: proceskenmerken staan veel "losser" van de inhoudskenmerken. Overigens kunnen deze bevindingen ook een artefact zijn van de onvergelijkbaarheid van de twee dimensies op het punt van de wiskundig-didactische inhoud (contexten).

- (d) Een verschil zien we ook op het punt van samenwerking tussen leerlingen. De correlaties tussen SAM enerzijds en LIK en BIG anderzijds verschillen per conditie.

In de projectklassen is er een duidelijk verband tussen leiding, klassikale instructie en klimaat enerzijds en de samenwerking anderzijds. Bij de vergelijkingsklassen is dat verband er niet.

Bij de vergelijkingsklassen is er een (niet significant) negatief verband (-.49) tussen begeleiding (BIG) en samenwerking (SAM). Bij de projectklassen is dat verband er niet.

Samenvattend: we zien relatief sterke correlaties tussen leerkrachtvariabelen en taakgerichtheid van leerlingen. Dit geldt ook voor samenwerking tussen leerlingen en taakgerichtheid (time on task). Deze correlaties zijn stabiel over de condities. De verschillen tussen de condities spitsen zich toe op twee kenmerken, namelijk inhoud (contexten bij wiskunde) en samenwerking. Dit is niet verwonderlijk omdat dit wellicht de meest onderscheidende kenmerken zijn tussen de projectonderwijsleersituatie en de traditionele vergelijkingssituatie.

### 3.5.2 Verschillen tussen condities, scholen, klassen

Nu volgen de resultaten van enkele variantie-analyses. Het doel van deze analyses is: na te gaan of de verschillen tussen condities, tussen scholen en tussen klassen statistisch significant zijn. Deze verschillen geven informatie over de mate waarin onderwijsleerprocessen volgens de bedoelingen zijn verlopen (dat wil zeggen zoals door de leerlingen gepercipieerd).

Tabel 3.9: Variantieanalyse op verschillen tussen condities

	LIK		BIG		SAM		TAKO		IN		Tot Schaal.		
	N	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Hele be-stand	835	3.64	.62	2.38	.50	3.54	.47	3.39	.55	3.04	.64	3.26	.36
project-leerlingen	521	3.66	.59	2.47	.48	3.66	.45	3.43	.50	3.14	.64	3.33	.34
vergelij-kings-leerlingen	314	3.59	.67	2.25	.52	3.32	.42	3.34	.62	2.86	.60	3.14	.34
F		2.7		40.0		127.1		5.8		38.8		56.26	
(df)		(1;833)		(1;833)		(1;833)		(1;833)		(1;833)		(1;833)	
sign		.10		.000*		.000*		.017*		.000*		.000*	

Uit tabel 3.9 blijkt dat de projectscholen op alle subschalen hoger scoren dan de vergelijkingsscholen. Deze verschillen zijn alle significant (meestal  $p = .000$ ). Het verschil op subschaal LIK is minder groot.

We kunnen zeggen dat volgens de percepties van leerlingen de gewenste vernieuwingskenmerken in de projectscholen vaker en/ of beter zijn gerealiseerd dan in vergelijkingsscholen. De indruk die uit een globale inspectie van de klassegemiddelden naar voren kwam wordt nu dus bevestigd.

Ook tussen project- en vergelijkingsscholen vonden we significante verschillen in leerlingpercepties.

Tabel 3.10: Variantieanalyse op verschillen tussen alle scholen

	LIK		BIG		SAM		TAKO		IN		Tot Schaal		
	N	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
hele be-stand	835	3.64	.62	2.39	.50	3.54	.47	3.40	.55	3.04	.64	3.26	.36
School 1	143	3.36	.65	2.53	.47	3.52	.48	3.21	.50	2.94	.61	3.16	.34
School 2	158	<u>3.91</u>	.41	<u>2.58</u>	.49	<u>3.63</u>	.41	<u>3.62</u>	.40	<u>3.24</u>	.60	<u>3.46</u>	.26
School 3	220	<u>3.68</u>	.56	<u>2.36</u>	.44	<u>3.79</u>	.43	<u>3.44</u>	.51	<u>3.20</u>	.65	<u>3.34</u>	.35
School 4	229	3.46	.69	2.13	.51	3.33	.43	3.29	.65	3.00	.57	3.08	.38
School 5	85	3.93	.49	2.55	.41	3.28	.39	3.46	.53	2.55	.57	3.28	.28
F		27.6		28.7		42.3		14.4		23.6		35.9	
(df)		(4;830)		(4;830)		(4;830)		(4;830)		(4;830)		(4;830)	
sign		.000*		.000*		.000*		.000*		.000*		.000*	



Uit tabel 3.10 blijkt dat tussen de scholen onderling significante ( $p = .000$ ) verschillen bestaan. We gaan nu nog preciezer kijken naar verschillen tussen scholen binnen de condities.

Tabel 3.11: *Verschillen tussen de projectscholen*

	N	LIK		BIG		SAM		TAKO		IN		Tot. Schaal	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Alle proj.	521	3.66	.59	2.47	.48	3.67	.45	3.43	.50	3.14	.64	3.33	.34
School 1	143	3.36	.65	2.53	.47	3.52	.48	3.21	.50	2.94	.61	3.16	.34
School 2	158	3.91	.41	2.58	.49	3.63	.41	3.62	.40	3.24	.60	3.46	.26
School 3	220	3.68	.56	2.36	.44	3.79	.43	3.44	.51	3.20	.65	3.34	.35
F		39.3		11.6		17.5		28.5		10.7		32.8	
(df)		(2;518)		(2;518)		(2;518)		(2;518)		(2;518)		(2;518)	
sign		.000*		.000*		.000*		.000*		.000*		.000*	

Tabel 3.12: *Verschillen tussen vergelijkingscholen*

	N	LIK		BIG		SAM		TAKO		IN		Tot Schaal	
		M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Alle vergl.	314	3.59	.67	2.25	.52	3.32	.42	3.34	.62	2.86	.60	3.14	.37
School 4	229	3.46	.69	2.13	.51	3.33	.43	3.29	.65	3.00	.57	3.08	.38
School 5	85	3.93	.49	2.55	.54	3.28	.39	3.46	.53	2.55	.57	3.28	.28
F		31.9		46.8		.68		4.6		34.1		19.10	
(df)		(1;312)		(1;312)		(1;312)		(1;312)		(1;312)		(1;312)	
sign		.000*		.000*		.41		.03*		.000*		.000*	

De tabellen 3.11 en 3.12 geven significante verschillen te zien tussen scholen binnen elk van de twee condities. Opgemerkt dient te worden dat SAM (samenwerking tussen leerlingen in groepjes) niet differentieert tussen scholen in de vergelijkings conditie, dit in tegenstelling tot de project conditie.

Binnen de groep projectscholen scoort school 2 relatief hoog op vrijwel alle subschalen van de PERCIA. Voor school 1 geldt het omgekeerde. De vernieuwingskenmerken zoals gemeten via de PERCIA-schaal worden dus vaker en beter gerealiseerd in school 2 dan school 1. School 3 neemt vrijwel steeds een middenpositie in.

Van de twee vergelijkingsscholen scoort school 5 relatief hoog op LIK, BIG en TAKO. School 4 scoort vergelijkenderwijs hoog op INHOUD.

In het algemeen blijken er tussen alle vijf scholen significante verschillen te bestaan in de mate van implementatie van de vernieuwingskenmerken. De vergelijking valt vrijwel steeds ten gunste van de projectscholen uit. Ook binnen elke conditie bestaan er verschillen tussen scholen. Tussen alle klassen ( $N = 33$ , zie tabel 3.2) bestaan significante verschillen in gemiddelde scores op de PERCIA-totaalschaal en op de subschalen.

Opvallend zijn ook de verschillen tussen klassen binnen de projectconditie ( $N = 22$ ). De factor klas verklaart 34 procent van de variantie in de Percia-totaalscore. Er zijn op het klasseniveau grote verschillen in de mate van getrouwheid in implementatie van het programma in de projectconditie!

### 3.6 Conclusies

Omtrent het onderwijsleerproces in de klas zoals gezien door de ogen van de leerlingen zijn de volgende conclusies te trekken.

Er zijn verschillen tussen klassen, scholen en condities in de mate van realisering van de beoogde onderwijsleerprocessen.

In de projectscholen worden de processen in de klas meer volgens de bedoelingen gerealiseerd dan in vergelijkingsscholen. Als men de kenmerken van het SLO programma als dimensies voor een vergelijkende beoordeling neemt, dan verlopen processen in projectklassen positiever dan in vergelijkingsklassen. Hoewel deze conclusie voor alle onderscheiden dimensies geldt, zijn de verschillen het opvallendst op twee punten: samenwerking en leerstofinhoud (wiskunde in contexten). Leerlingen werken vaker en beter samen. De wiskunde wordt vaker in een dagelijkse context geplaatst. Projectleerlingen vinden de wiskunde-opdrachten aantrekkelijker en duidelijker dan vergelijkingsleerlingen.

Tussen de subschalen van de PERCIA doen zich enkele opvallend hoge en stabiele correlaties voor.

(a) In klassen waarin de leraar goed leiding geeft en duidelijk uitlegt

blijven de leerlingen in het algemeen beter bij de taak dan in klassen met een minder goede leiding en instructie door de leraar.

(b) In klassen waarin de leerlingen goed samenwerken is de taakgerichtheid in het algemeen beter dan in klassen waarin de leerlingen minder adequaat samenwerken.

Het nieuwe curriculum, Wiskunde voor Iedereen, is dus geen papieren vernieuwing gebleven. De onderwijsleerprocessen zijn anders en beter (in termen van door leerplanontwikkelaars beoogde processen) in projectklassen dan in vergelijkingsklassen waarin met een meer "traditionele" aanpak wordt gewerkt.

Ter relativering van deze conclusies merken we op dat de gevonden verschillen, ondanks de significantie, soms klein zijn. Het vrij grote aantal leerlingen maakt dat kleine verschillen al gauw significant zijn. Statistische significantie impliceert nog niet noodzakelijkerwijs praktische betekenis! De vraag naar de praktische betekenis van deze (kleine) verschillen zal dan ook in het licht van verschillen in toetsprestaties moeten worden gezien. Een interessante vraag is ook of deze leerling-percepties inderdaad informatie geven over het onderwijsleerproces in de klas. Uit een vergelijking van perceptiescores en observatiegegevens blijkt dat die laatste vraag positief kan worden beantwoord (zie onder meer hoofdstuk 6).

## 4. DE HOUDING VAN DE LEERLINGEN

### 4.1 Vraagstelling

De hoofdvraag in dit hoofdstuk is:

Wat is de houding van de leerlingen ten opzichte van verschillende aspecten van het onderwijs? Deze vraag is te specificeren aan de hand van de volgende deelvragen. Wat is de houding van de leerlingen aan het begin van de periode waarin het curriculum is uitgevoerd en wat is de houding aan het eind? Interessant is ook om na te gaan of er verschillen in houding zijn tussen leerlingen in verschillende klassen, scholen en condities.

De vragenlijst is twee keer afgenomen: aan het begin en aan het eind van de periode waarin het wiskunde-programma is uitgevoerd. De beginlijst en de eindlijst zijn identiek, echter aan de eindlijst is een reeks vragen toegevoegd die zijn geïnspireerd op de vragenlijst van Huber (1986)\*. Het betreft een semantische differentiaal omtrent de wiskunde lessen, de leraar en het groepje. (Zie ook bijlage 2).

De vragenlijst als geheel bestaat uit zes subschalen. Steeds gaat het om de waardering van de leerling voor bepaalde aspecten. Deze aspecten (subschalen) zijn: de school, wiskunde, het eigen niveau bij wiskunde zoals ingeschat door de leerling, het belang van wiskunde voor jongens en meisjes (sexe), werkvormen (klassikaal, groepswerk, individueel werken) en contexten. Deze clusters zijn zoals gezegd identiek voor vragenlijst 1 en 2. Voor vragenlijst 2 komen daar nog drie clusters bij (semantische differentiaal). Deze drie clusters betreffen de houding tegenover de les, de leraar en de groep.

De indeling in subschalen is primair vooraf op inhoudelijke gronden tot stand gekomen. Op grond van betrouwbaarheidsanalyses en factoranalyses hebben we enkele kleine wijzigingen aangebracht in de aanvankelijke indeling, terwijl één item is komen te vervallen.

De vraagstelling is als volgt te formuleren: zijn er verschillen in houdingen tussen leerlingen in verschillende klassen, scholen, en condities?

---

\* Niet gepubliceerde vragenlijst

## 4.2 Operationalisering

Om van ongeveer 800 leerlingen gegevens omtrent hun houding te verzamelen is gebruik gemaakt van een schriftelijke vragenlijst. Bij de samenstelling van de vragenlijst is onder meer gebruik gemaakt van items uit andere vragenlijsten, met name de vragenlijsten van Krammer (1984) en van Huber (1986). Daarnaast is een aantal items opgenomen dat betrekking heeft op didactische en inhoudelijke aspecten van het (wiskunde) onderwijs. Deze items zijn in het kader van ons onderzoek ontwikkeld. De vragen hebben de vorm van stellingen waarbij de leerlingen kunnen aangegeven of zij het al of niet eens zijn met de stelling, bijvoorbeeld:

Ik vind wiskunde prettig

- zeer mee oneens
- mee oneens
- niet mee eens
- en niet mee oneens
- mee eens
- zeer mee eens

De items van de vragenlijst zijn opgenomen in bijlage 2.

Bij de analyses zijn de schaalenden van de negatief geformuleerde items omgedraaid, met het oog op correspondentie met de positieve items in dezelfde subschaal.

## 4.3 Betrouwbaarheid

De betrouwbaarheid van de subschalen en van de totale schaal is bepaald met behulp van Cronbach's alpha.

Hieronder volgen de alpha-coëfficiënten voor het hele bestand (ongeveer 800 leerlingen) voor schaal 1 en schaal 2.

Tabel 4.1: *Betrouwbaarheid vragenlijsten houding*

	Aantal items	schaal 1	schaal 2
H-School	7	.73	.73
H-Wiskunde	6	.64	.64
H-Niveau	4	.79	.79
H-Sexe	2	-	-
H-werkvormen	5	.49	.52
H-context	3	.68	.79
Totale Schaal	27	.78	.78



De betrouwbaarheid van de subschalen is matig. De betrouwbaarheid van de subschaal H-werkvormen is niet voldoende.

Tabel 4.1 bevat geen gegevens over de betrouwbaarheid van de subschaal H-sexe. Deze is niet te berekenen omdat deze subschaal slechts 2 items bevat. Ter aanvulling op de gegevens in tabel 4.1 dient opgemerkt te worden dat de betrouwbaarheid van de subschalen in de project-conditie meestal iets hoger ligt dan in de vergelijkingssituatie. Voorts berekenden we de betrouwbaarheden op klasseniveau; dat wil zeggen betrouwbaarheid van klassegemiddelden. Deze liggen gemiddeld .10 hoger dan de waarden in tabel 4.1.

Tenslotte nog een opmerking over de betrouwbaarheid van de semantische differentiaal. De betrouwbaarheid voor deze hele schaal is goed (.81). De betrouwbaarheden van de subschalen zijn: les (.72), leraar (.80) en groep (.68).

De eindconclusie over de betrouwbaarheid van de houdinglijsten is dat de totale schalen betrouwbaar genoeg zijn. De betrouwbaarheid van sommige subschalen laat te wensen over. Met name de subschaal over werkvormen.

#### **4.4 Procedure en analyse**

De vragenlijst naar de houding van de leerlingen inzake verschillende aspecten van het onderwijs werd twee keer afgenomen. Voor alle subschalen zijn gemiddelden en standaarddeviaties bepaald voor verschillende niveaus zoals klas, school en conditie. Er zijn one way-variantie-analyses uitgevoerd op de gevonden verschillen.

#### **4.5 Resultaten**

Hieronder volgen de resultaten van de twee vragenlijsten naar de houding van de leerlingen. In tabel 4.2 staan de gemiddelden en standaardafwijkingen van houdinglijst 1 (beginmeting) per conditie. Tevens zijn gegevens van de variantie-analyse naar verschillen tussen de condities opgenomen.

Tabel 4.2: Gemiddelde scores per conditie subschalen houdingvragenlijst 1 (voormeting)

	School		Wiskunde		Niveau		Sexe		Werkvorm		Context		Hele schaal		N
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	
Hele be- stand	2.86	.60	3.18	.61	3.45	.80	4.37	.82	3.04	.60	3.23	.84	3.20	.39	822
Project- leerlingen	2.87	.60	3.19	.64	3.43	.78	4.39	.81	2.98	.60	3.27	.81	3.20	.110	515
Verge- lijks- leerlingen	2.83	.60	3.16	.55	3.50	.82	4.34	.83	3.13	.60	3.16	.89	3.21	.37	307
F	.987		.350		1.50		.633		12.78		3.08		.049		
(df)	(1:820)		(1:820)		(1:820)		(1:817)		(1:816)		(1:813)		(1:820)		
sign	.321		.554		.221		.427		.000*		.080		.824		

\* is significant op 5 procent niveau

Uit tabel 4.2 is af te leiden dat er tussen de gemiddelde scores van de leerlingen in de twee condities in de beginsituatie nauwelijks verschillen bestaan. Alleen bij de houding ten aanzien van werkvormen is er een significant verschil.

Het verschil bij werkvormen betekent dat de projectleerlingen een positieve houding hebben tegenover groepswork en minder positief staan tegenover andere werkvormen zoals klassikaal en individueel werken. Bij de vergelijkingsleerlingen is het andersom: onbekend maakt onbemind. Overigens moet worden opgemerkt dat deze subschaal onvoldoende betrouwbaar is.

Bij de aanvang van de onderzoeksperiode zijn er dus op de meeste subschalen géén verschillen in houding tussen de leerlingen in de beide condities. Ook over de totale schaal is er geen verschil. We gaan nu kijken naar de houding van de leerlingen na het doorlopen van het project- of het vergelijkingsprogramma. Tabel 4.3 bevat gegevens van vragenlijst 2 (nameting) naar de houding van de leerlingen ten opzichte van verschillende aspecten van de onderwijsleersituatie.

Tabel 4.3: Gemiddelde scores per conditie subschalen houdingvragenlijst 2 (nameting)

	School		Wiskunde		Niveau		Sexe		Werkvorm		Context		Hele schaal		N
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	
Hele be- stand	2.78	.57	3.13	.60	3.43	.79	4.33	.79	2.94	.59	3.14	.90	3.14	.38	790
Project	2.87	.55	3.14	.63	3.42	.79	4.37	.75	2.88	.59	3.22	.88	3.14	.39	489
Verge- lijking	2.77	.59	3.12	.54	3.43	.80	4.27	.85	3.05	.57	3.01	.92	3.14	.37	301
F	.015		.251		.019		2.60		16.01		10.01		.000		
(df)	(1:788)		(1:786)		(1:782)		(1:799)		(1:776)		(1:772)		(1:788)		
sign	.901		.617		.892		.107		.000*		.002*		.998		

Na het doorlopen van het curriculum verschilt de houding van de projectleerlingen op twee punten van die van de vergelijkingsleerlingen. Ook nu zien we bij de projectleerlingen een positievere houding tegenover groepswork (of een negatieve houding tegenover klassikaal en individueel werken) dan bij de vergelijkingsleerlingen. Overigens is deze subschaal (werkvormen) onvoldoende betrouwbaar om gefundeerde conclusies te trekken. Voorts staan projectleerlingen positiever tegenover contexten. Dit laatste is begrijpelijk omdat dit een centraal aspect van het nieuwe wiskundeprogramma betreft. De vraagstelling met betrekking tot contexten is specifiek gericht op het nieuwe programma. Het is mogelijk dat het hier gevonden verschil tussen projectleerlingen en vergelijkingsleerlingen (voor een deel) een artefact is van de meting. Vergelijkingsleerlingen herkennen deze vragen wellicht minder goed.

Overigens blijkt er tussen de condities géén verschil te bestaan in gemiddelde scores op de hele schaal. Als we kijken naar de hele schaal is er dus niets veranderd tussen de voormeting (vragenlijst 1) en de nameting (vragenlijst 2).

We gaan nu kijken naar de resultaten op de semantische differentiaal die onderdeel uitmaakt van vragenlijst 2 naar de houding.

Tabel 4.4: Gemiddelde scores semantische differentiaal (onderdeel van vragenlijst 2 Houding)

	Les		Leraar		Groep		Hele schaal	
	M	SD	M	SD	M	SD	M	SD
Hele bestand	3.12	.82	3.17	.88	3.75	.85	3.31	.62
Projectleerlingen	3.16	.85	3.30	.84	3.86	.84	3.40	.60
Vergelijkingsleerlingen	3.05	.77	2.98	.91	3.58	.85	3.16	.64
F	3.45		23.94		20.55		26.17	
(df)	(1:775)		(1:770)		(1:757)		(1:771)	
sign	.064		.000*		.000*		.000*	

Tabel 4.4 geeft twee significante verschillen te zien waar het gaat om de houding van de leerlingen tegenover de leraar en de groepjes. De conclusie uit tabel 4.2 t/m 4.4 is dat er in de aanvangssituatie nauwelijks verschillen bestaan in houding tussen de leerlingen. Na het doorlopen van de verschillende curricula is er op enkele punten een significant verschil tussen projectleerlingen en vergelijkingsleerlingen. Het gaat daarbij vooral om aspecten die in het project Wiskunde voor Iedereen als vernieuwingsmerken zijn uitgewerkt (werkvormen en contexten).

Een interessant punt is verder dat de leerlingen in de projectscholen hun niveau bij wiskunde niet lager inschatten dan de vergelijkingsleerlingen. In feite is het gemiddeld niveau van projectleerlingen wél lager dan dat van de vergelijkingsleerlingen (zie hoofdstuk 5 en 6). Speelt hier het referentiekader van de groep een rol òf hebben projectleerlingen door het andere onderwijs een positiever beeld over hun niveau? Waarschijnlijk is de houding van de leerlingen ten aanzien van hun niveau bij wiskunde tot standgekomen op grond van verschillende ervaringen en gegevens zoals: kennis, zelfvertrouwen, klassegemiddelde, onderwijsklimaat enz. Overigens is de houding tegenover het eigen niveau (op leerlingniveau) een relatief betrouwbare en relatief invloedrijke predictor van prestaties op de toetsen (zie ook hoofdstuk 6).

Tabel 4.5: Gemiddelde scores per school per subschaal van houdinglijst 1

	School		Wiskunde		Niveau		Sexe		Werkvormen		Context		Hele Schaal		N
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	
Hele bestand	2.86	.60	3.18	.61	3.45	.80	4.37	.82	3.04	.60	3.23	.84	3.20	.39	822
School 1	2.92	.66	3.29	.63	3.00	.76	4.41	.82	3.08	.65	3.29	.90	3.27	.41	144
2	3.06	.52	3.31	.60	3.48	.83	4.56	.70	3.06	.54	3.26	.78	3.31	.36	156
3	2.72	.59	3.02	.65	3.34	.76	4.24	.86	2.86	.58	3.25	.90	3.07	.40	215
4	2.93	.57	3.17	.57	3.51	.81	4.37	.84	3.10	.60	3.23	.81	3.24	.36	224
5	2.60	.61	3.13	.51	3.46	.83	4.25	.78	3.25	.59	2.97	1.04	3.13	.36	83
F	12.38		6.94		1.57		4.13		8.48		2.30		11.80		
(df)	(4;817)		(4;817)		(4;817)		(4;814)		(4;813)		(4;810)		(4;817)		
sign	.000*		.000*		.1810		.0025		.000*		.057		.000*		

Op alle aspecten bestaan significante verschillen tussen de vijf scholen. Op het eerste gezicht valt op dat school 2 relatief hoog scoort. School 3 daarentegen geeft lagere gemiddelde waarden op de houdingsaspecten te zien. Ook als men een one way-variantie-analyse tussen de scholen *binnen* de projectconditie (3 scholen) uitvoert blijft een aantal van deze verschillen significant. Vrijwel steeds scoort school 2 het hoogst en school 3 het laagst. Die conclusie geldt ook voor de hele schaal. Dat wil zeggen dat er tussen de drie projectscholen significante verschillen bestaan in houding van de leerlingen. De leerlingen op school 2 staan relatief positief tegenover hun school en het wiskunde onderwijs. Voor school 3 geldt dat de leerlingen minder positief staan tegenover hun school en het wiskunde-onderwijs.

We gaan nu kijken of deze verschillen tussen scholen in beginsituatie ook terug te vinden zijn in de eindsituatie (houdinglijst 2).



Tabel 4.6: Gemiddelde scores per school per subschaal van houdinglijst 2

	School		Wiskunde		Niveau		Sexe		Werkvormen		Context		Hele Schaal		N
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	
Hele bestand	2.78	.57	3.13	.60	3.43	.79	4.33	.79	2.94	.59	3.14	.90	3.14	.38	790
School 1	2.75	.61	3.17	.64	3.50	.76	4.28	.72	2.89	.60	2.97	.93	3.10	.39	117
2	2.92	.51	3.25	.67	3.37	.85	4.60	.63	2.93	.62	3.24	.81	3.22	.39	158
3	2.69	.54	3.04	.58	3.42	.75	4.23	.82	2.83	.55	3.33	.88	3.09	.39	214
4	2.79	.60	3.13	.56	3.45	.76	4.29	.88	2.98	.58	3.09	.89	3.14	.38	223
5	2.73	.55	3.10	.49	3.36	.92	4.22	.79	3.24	.51	2.78	.99	3.12	.35	78
F	3.82		2.96		.651		6.10		7.64		7.05		3.09		
(df)	(4:785)		(4:783)		(4:779)		(4:785)		(4:773)		(4:769)		(4:785)		
sign	.004*		.0193*		.626		.000*		.000*		.000*		.015*		

Volgens tabel 4.6 zijn de verschillen tussen de vijf scholen significant behalve voor de subschaal niveau. School 2 scoort ook nu op de meeste subschalen en op de totale schaal relatief hoog.

Ook bij een one way-variantie-analyse naar verschillen tussen de drie projectscholen blijven de verschillen significant. Alleen op de houding tegenover contexten scoort school 3 relatief hoog. (Tussen de drie scholen bestaan geen verschillen in gemiddelde scores op niveau en werkvormen).

We concluderen dat de relatief positieve houding van de leerlingen op school 2 óók in de eindsituatie aanwezig is, terwijl school 1 en 3 op gemiddeld hetzelfde (lagere) niveau eindigen.

De nu volgende tabel 4.7 bevestigt deze conclusie. Hier zijn de gegevens per school op de semantische differentiaal samengevat.

Tabel 4.7: Gemiddelde scores semantische differentiaal (onderdeel van houdinglijst 2)

	Les		Leraar		Groep		Hele schaal	
	M	SD	M	SD	M	SD	M	SD
Hele bestand	3.12	.82	3.17	.88	3.75	.85	3.31	.62
School 1	3.19	.93	3.23	.98	3.98	.93	3.39	.73
2	3.22	.89	3.55	.70	3.78	.86	3.50	.50
3	3.11	.78	3.14	.83	3.86	.76	3.33	.60
4	3.03	.76	2.85	.89	3.56	.83	3.11	.63
5	3.10	.80	3.35	.88	3.63	.92	3.33	.64
F	1.34		16.58		6.13		10.68	
(df)	(4;763)		(4;767)		(4;754)		(4;768)	
sign	.252		.000*		.000*		.0000*	

Tussen scholen bestaan geen significante verschillen in houding tegenover de les. Wel bestaan er verschillen in houding tegenover de leraar. De houding van de leerlingen tegenover de leraar is het meest positief op school 2.

School 1 scoort binnen de vijf scholen relatief hoog op de houding tegenover de groep. Als men echter een one way-variantie analyse uitvoert op verschillen in houding tegenover de groep tussen de drie projectscholen dan is het verschil niet (langer) significant.

Conclusie: De leerlingen op school 2 staan relatief positief tegenover hun leraren. Ook over de hele semantische differentiaal komt school 2 relatief hoog uit de bus.

#### 4.6 Conclusies

Bij het begin van de uitvoering van het curriculum werd een vragenlijst afgenomen betreffende de houding van de leerlingen. Dezelfde vragenlijst werd afgenomen na de uitvoering. Aan deze laatste vragenlijst werd nog een semantische differentiaal toegevoegd. De vragenlijsten werden afgenomen in drie projectscholen en twee vergelijkingsscholen.

De conclusies zijn: bij de beginsituatie bestaan er geen verschillen van betekenis in gemiddelde houding tussen leerlingen van projectscholen en vergelijkingsscholen. Wel bestaan er verschillen tussen de scholen. De leerlingen op school 2 hebben, relatief gezien, een positievere houding ten

aanzien van hun school en verschillende aspecten van het (wiskunde) onderwijs.

In de eindsituatie zijn er evenmin verschillen tussen de condities als we kijken naar de totale schaal. Wel zijn er enkele verschillen op aspecten (subschalen en op de semantische differentiaal). Deze verschillen betreffen de houding ten opzichte van contexten, de leraar en de groep. Deze verschillen vallen steeds in het voordeel uit van de projectleerlingen. Ook op de totale schaal van de semantische differentiaal scoren de projectleerlingen significant hoger dan de vergelijkingsleerlingen. Ook vonden we verschillen tussen scholen, waarbij ook nu weer school 2 relatief hoog scoort.

Hoewel we slechts enkele significante verschillen vonden op enkele subschalen van de houdingvragenlijst moet worden opgemerkt dat de praktische betekenis hiervan nog niet is vastgesteld. In hoofdstuk 6 zullen we deze verschillen in relatie tot andere gegevens bezien.

## 5. LEERRESULTATEN: VOORTOETS EN NATOETS

### 5.1 Vraagstelling

Een belangrijke vraag bij elke onderwijsvernieuwing betreft de leereffecten. Wat steken de leerlingen er van op? Wat preciezer geformuleerd: wat zijn de leereffecten bij de leerlingen, mede gelet op verschillen in uitgangsniveau? Een belangrijke vraag is ook: zijn de leereffecten in de projectscholen beter dan in de vergelijkingscholen? (vergelijk Hypothese B).

### 5.2 Operationalisering

Als methode van onderzoek is gekozen voor schriftelijke toetsen. De toets bestaat uit een combinatie van open en gesloten vragen. Aan de leerplanontwikkelaars van de SLO hebben we gevraagd (ideeën voor) toetsopgaven te leveren die overeenkomen met de doelstellingen van het programma. Daarmee is de inhoudsvaliditeit van de toetsen gewaarborgd. Overigens stuiten we hiermee wèl op de kern van het vraagstuk van de vergelijkbaarheid van twee verschillende programma's: het SLO-programma en een meer traditionele aanpak. In hoofdstuk 1 paragraaf 5 maakten we al een inhoudelijke vergelijking tussen het nieuwe SLO-programma en de traditionele, klassikale aanpak. In de discussie in hoofdstuk 8 komen we hier op terug.

Bij de uiteindelijke keuze van toetsopgaven en bij de constructie van de toets als geheel hebben gegevens uit de vorige onderzoeksrunde eveneens een rol gespeeld. Enkele toetsopgaven zijn door de onderzoekers ontwikkeld, in nauw overleg met de leerplanontwikkelaars. De toetsopgaven zijn eerst op kleine schaal bij leerlingen beproefd. Na uitgebreide discussies en verschillende bijstellingen ontstond toets 1 (de voortoets).

Vervolgens zijn twee "parallelversies" van deze toets gemaakt namelijk toets 2 (de 2e voortoets) en toets 3 (de natoets). Deze toetsen zijn qua doelstellingen identiek, de vormgeving verschilt echter. Door kleine verschillen aan te brengen in de context, de getallen of de alternatieven waaruit gekozen kan worden, ontstonden drie versies van één reeks toetsopgaven. Aan toets 3 (de natoets) zijn drie extra opgaven toegevoegd om eventuele plafondeffecten te voorkomen.

Zo ontstonden toetsen die, naar het oordeel van de ontwikkelaars, de inhouden en doelstellingen van het programma op adequate wijze representeren (zie voor een voorbeeld van de toets bijlage 3). De twee

voortoetsen hebben als titel: Wat weet je al? De natoets is getiteld: Wat weet je nu? De voortoetsen (toets 1 en 2) bevatten elk 16 samengestelde opgaven. De natoets (toets 3) bevat 16 opgaven plus 3 extra opgaven. Meestal bestaat één opgave uit verschillende deelvragen. De opgaven zijn in een zogenaamde 'context' geplaatst. De opgaven gaan over functies, verbanden, grafieken en verhoudingen. Om een indruk te geven van de opbouw, de onderwerpen en de maximaal te behalen score op de samengestelde toetsopgaven volgt hier een kort overzicht van de titels van de opgaven en de maximale score van toets drie (natoets). Zie voorts bijlage 3.

*Tabel 5.1: Toetsitems en maximale scores*

Titel opdracht	maximaal te behalen score
1. Vazen	6 punten
2. Michel en Maureen (fietsen)	4 punten
3. Hardlopen	4 punten
4. Tarief	2 punten
5. Geboorte	7 punten
6. Volleybal	2 punten
7. Kamperen	4 punten
8. Krant	9 punten
9. Video	9 punten
10. Koper*	10 punten
11. Gas	6 punten
12. Balans	7 punten
13. Boek	6 punten
14. Glaasje op	8 punten
15. Taxi	6 punten
16. Mais*	4 punten
17. Foto	5 punten
18. Zagen	4 punten
19. Vlak*	6 punten

---

Totaal 109 punten

\* De drie opgaven met een sterretje zijn extra opgaven die alléén in de natoets voorkomen.



We hebben aan de leerplanontwikkelaars gevraagd een onderscheid aan te brengen tussen communale en differentiële doelstellingen (toets-opgaven). Het antwoord daarop was dat in principe alle opgaven als belangrijk moesten worden beschouwd. Ook hebben we gevraagd naar criteria. Wanneer zijn voor de leerplanontwikkelaars van de SLO de doelstellingen van het curriculum bereikt? Over dit vraagstuk is uitgebreid en herhaaldelijk gediscussieerd. De onderzoekers waren van mening dat er in deze laatste fase van het onderzoek duidelijke evaluatiecriteria zouden moeten worden geformuleerd. Daarover was met de acceptatie van ons onderzoeksplan in principe overeenstemming bereikt. Welnu, deze criteria zijn niet op tafel gekomen. De leerplanontwikkelaars zeiden: wij kunnen geen criteria formuleren. Letterlijk staat in de notulen van de vergadering waarin de criteriumdiscussie werd afgesloten: "... dat het niet zo is dat de ontwikkelgroep niet wil maar dat ze er niet toe in staat is". Dit is een veel voorkomend verschijnsel; zie onder andere de discussie tussen Creemers (1987) en Terwel (1987a).

Door het ontbreken van criteria verviel de mogelijkheid de vooraf geformuleerde als... dan... hypothese te toetsen. Deze hypothese is: Als het curriculum volgens de bedoelingen van de leerplanontwikkelaars is uitgevoerd dan zal tenminste 80% van de leerlingen tenminste 80% van de maximale score behalen op de vooraf aangewezen communale opgaven uit de eindtoets. De leerplanontwikkelaars hebben echter niet in toetsbare termen aangegeven wat het betekent dat alle opgaven als "belangrijk" moeten worden beschouwd. Mede om die redenen hebben we besloten het accent te leggen op het vergelijken van (a) verschillende implementaties van het SLO-programma en (b) het SLO-programma met een meer traditionele aanpak. Daarmee verschoof de onderzoekopzet van een criteriumgerichte opzet naar een vergelijkend design. Zie hoofdstuk 1 voor de opzet van het onderzoek.

Deze vergelijkende opzet maakte een transformatie van de *vorm* van de algemene werkhypothese noodzakelijk. In plaats van een Als... Dan... hypothese formuleren we de hypothese in termen van Hoe... des te.... We merken uitdrukkelijk op dat inhoudelijk-theoretisch géén verandering in de hypothese is aangebracht. Deze inhoudelijk-theoretische achtergrond wordt gevormd door een aantal vernieuwingskenmerken zoals deze ten grondslag liggen aan het SLO-programma. Bij de uitwerking van deze kenmerken is gebruik gemaakt van theoretische inzichten en researchgegevens. Zie hoofdstuk 1 van dit rapport en de "basispublicaties" van het project ID 12-16 in de literatuurlijst zoals opgenomen in dit rapport. Vervolgens zijn uit de algemene werkhypothese twee specifieke hypothesen afgeleid: hypothesen A en B. Hypothese A heeft betrekking op een vergelijking van varianten in de implementatie *binnen* de

projectconditie dat wil zeggen tussen projectklassen. Hypothese B betreft de vergelijking van het SLO-programma met een meer traditioneel programma (vgl. hoofdstuk 1 van dit rapport).

De keuze van een vergelijkende opzet maakt allerlei vergelijkingen mogelijk. Deze opzet biedt veel extra informatie, in vergelijking met een curriculuminterne evaluatie op basis van absolute criteria, terwijl het criterium vraagstuk in beginsel wordt opgelost. Een groot voordeel is ook dat een dergelijke vergelijking door de onderzoekers kan worden uitgevoerd zonder dat medewerking van de leerplanontwikkelaars noodzakelijk is (dit in tegenstelling tot een absolute, criteriumgerichte benadering die onderzoekers afhankelijk maakt van de ontwikkelaars die de criteria moeten leveren).

Een geheel ander probleem betreft het vraagstuk van de leerwinst. Omdat we vanuit de vraagstelling geïnteresseerd zijn in leereffecten mede gelet op verschillen in uitgangsniveau ligt een begrip als leerwinst of veranderingsscore voor de hand (verschilcores tussen voortoets en natoets). Het hanteren van verschilcores kan tenminste drie problemen opleveren.

1. Het statistisch regressie-effect: bijvoorbeeld als men groepen leerlingen wil vergelijken op gemiddelde verschilcores en als deze groepen leerlingen verschillen in voortoetscore, dan kan het verschijnsel van regressie naar het gemiddelde optreden. Dan kan het lijken dat de leerlingen met een lage beginscore relatief veel leerwinst hebben behaald en dat leerlingen met een hoge beginscore weinig zijn vooruitgegaan. Deze verschillen tussen de zwakke en sterke leerlingen kunnen echter het gevolg zijn van regressie naar het gemiddelde. Het is hier niet de plaats uitgebreid in te gaan op dit vraagstuk (zie bijvoorbeeld Oud, 1981 en 1982, Hamers en Ruijsenaars 1984, blz 19 en 20 en Petermann, 1978). Als oplossing van dit vraagstuk adviseert Oud (1982, blz. 85) om meer dan één voormeting toe te passen bijvoorbeeld met behulp van paralleltoetsen. Het regressieprobleem doet zich vanaf de tweede afname niet meer voor. Om die redenen hebben we twee voormetingen afgenomen (voortoets 1 en 2).
2. Schaalproblemen: het is mogelijk dat gelijke afstanden op de toetsschaal in feite verschillende afstanden in beheersing voorstellen. Het zou bijvoorbeeld kunnen zijn dat het vooruitgaan van voortoetscore 20 naar natoetscore 30, niet vergelijkbaar is met een vooruitgang van 30 naar 40 punten. Aan het eind van de schaal, met name op de natoets kunnen plafondeffecten optreden. De goede leerlingen zouden nog wel meer en moeilijker opgaven kunnen maken

maar de schaal laat dat niet toe. Deze plafondeffecten zijn een deel van het schaalprobleem. Wij hebben getracht dit probleem op te lossen door extra opgaven op te nemen in de natoets. Voor het overige moet men alert zijn op het schaalprobleem bij de interpretatie van verschillcores.

3. **Betrouwbaarheidsproblemen:** de betrouwbaarheid van verschillcores is afhankelijk van de betrouwbaarheden en spreidingen van elk van de afzonderlijke scores (voortoets en natoets). In de verschillcores komen de (on)betrouwbaarheden van de afzonderlijke scores samen. Een bijkomende factor is de correlatie tussen voortoets en natoets. Hoe hoger de correlatie hoe lager de betrouwbaarheid van de verschillcores (vgl. Magnusson, 1966 blz. 94). In ons onderzoek is de correlatie tussen voortoets en natoets hoog. Dat is op zichzelf een positieve indicatie voor de betrouwbaarheid van de twee toetsen. De (paradoxe) situatie doet zich echter voor dat die hoge correlatie, een lage betrouwbaarheid van de verschillcores impliceert zelfs bij redelijke tot hoge betrouwbaarheden van de afzonderlijke toetsen. En dat alles is het geval in ons onderzoek. We kunnen dus bij het trekken van conclusies niet blind varen op verschillcores. Het zijn onbetrouwbare indicatoren die steeds in relatie met andere gegevens moeten worden gezien. Daarom maken we ook gebruik van andere gegevens (voortoetsscore, natoetsscore) en andere analyses die niet zijn gebaseerd op het begrip verschillcore (zie ook hoofdstuk 6). Door grote zorgvuldigheid bij de constructie, afname en scoring van de toetsen hebben we geprobeerd de betrouwbaarheid zo hoog mogelijk te doen zijn. Voor het overige moet bij de interpretatie van verschillcores voorzichtig te werk worden gegaan dat wil zeggen we moeten onze conclusies niet uitsluitend op verschillcores baseren. Daarom zullen we ook langs andere wegen zoeken naar gegevens ter bevestiging of weerlegging van de hypothesen.

### 5.3 Betrouwbaarheid

Allereerst is een coderingssysteem ontwikkeld waarmee alle antwoorden van de leerlingen in kaart kunnen worden gebracht. Binnen elke opgave worden deelopgaven onderscheiden, voor elke deelopgave worden punten gegeven. Nadat bij een beperkt deel van de gemaakte toetsen, het coderingssysteem was beproefd zijn enkele verbeteringen in de wijze van coderen aangebracht. Vervolgens is (opnieuw) bij een beperkt deel van de toetsen nagegaan of de intercodeurs-betrouwbaarheid hoog genoeg was. Er was een overeenstemming tussen twee onderzoekers van 95 procent.

Vervolgens zijn alle toetsen zorgvuldig gecodeerd. Van elke toets is de betrouwbaarheid bepaald via een interne consistentie-maat: Cronbach's alpha. Daarbij is als volgt te werk gegaan. Van elke opgave, bijvoorbeeld "Vazen", maakten we een somscore: het aantal punten op alle deelvragen binnen de opgave "Vazen". De betrouwbaarheidsanalyse is dus niet uitgevoerd op onderling samenhangende deelvragen maar we gingen uit van de samengestelde opgaven als eenheid van analyse (16 respectievelijk 19 samengestelde opgaven). De alpha coëfficiënten voor toets 1, 2 en 3 zijn respectievelijk .74, .80 en .83. Mede gelet op de aard van de vragen, dat wil zeggen vragen die vaak een open karakter hebben en die in contexten zijn geplaatst, zijn dat redelijke tot hoge betrouwbaarheden.

#### **5.4 Procedure en analyse**

De twee voortoetsen werden afgenomen aan het begin van de periode waarin het programma werd uitgevoerd. De natoets werd na afloop van die periode afgenomen. Hoewel met alle projectscholen overeenstemming was bereikt dat alle drie toetsen zouden worden afgenomen trok één school (school 3, met 8 klassen) zich terug voor wat betreft toets 2. Men vond het bezwaarlijk om deze toets af te nemen. Om soortgelijke redenen werd in klas 4 van school 2, toets 2 niet afgenomen. Van de twee vergelijkingsscholen, die wij pas in een relatief laat stadium benaderden, kon school 5 (met in totaal 3 klassen) de tweede toets niet inpassen in de reeds gemaakte plannen. Ook deze school viel dus uit bij toets 2. Van toets 2 missen we dus een aanzienlijk deel van de gegevens.

Bij de analyse van de gegevens is gebruik gemaakt van somscores. Er zijn gemiddelden berekend voor verschillende categorieën leerlingen. Ook zijn correlaties bepaald. Door middel van one way-variantie-analyse zijn verschillen getoetst. In alle gevallen waarin we spreken van een significant verschil is getoetst op een significantieniveau van .05, ook al zijn niet altijd F-waarden vermeld.

#### **5.5 Resultaten**

##### **5.5.1 De beginkennis van de leerlingen: toets 1**

De scores op toets 1 voor alle 800 leerlingen liggen in het interval van 4,5 - 71 punten. De maximaal te behalen score is 89 punten. De gemiddelde score is 31,7 en de standaarddeviatie is 10,5.

De volgende frequentieverdeling, volgens intervallen van telkens 10 scorepunten, geeft een beeld van de spreiding van de scores.



Tabel 5.2: Frequentieverdeling toets 1

punten behaald op toets 1	absolute freq	relatieve freq (pct)
0 - 10	11	1.3
10 - 20	77	9.3
20 - 30	308	37.2
30 - 40	276	33.3
40 - 50	97	11.7
50 - 60	31	3.7
60 - 70	10	1.2
70 - 80	1	.1
ontbrekende leerlingen	18	2.2
	----- 829	----- 100.0

Uit tabel 5.2 is af te leiden dat de leerlingen een grote spreiding vertonen qua voorkennis zoals gemeten met de voortoets. De toets differentieert sterk tussen de leerlingen.

Een verdeling van de leerlingen, naar hun score op de voortoets, in Laag, Midden en Hoog maakt het mogelijk om een beeld te vormen van de verschillen tussen condities, scholen en klassen. Het betreft dan verschillen in proporties L,M,H leerlingen. De grenzen voor de categorie "Midden" zijn 27 en 34,5 voortoetsscore. Daaronder is "Laag". Daarboven is "Hoog". Deze grenzen gelden voor het gehele bestand dus voor projectleerlingen en vergelijkingsleerlingen. De totale groep is daarmee in drie ongeveer gelijke delen verdeeld. Soms spreken we in dit rapport van L,M,H leerlingen, soms worden de termen zwakke, middelmatige en sterke leerlingen gebruikt. In beide gevallen is hetzelfde bedoeld: leerlingen met een score op de voortoets vallend in de categorieën L,M of H. In geen geval hanteren we deze categoriën in denigrerende of lovende zin. Het gaat om informatie over leerlingen op basis van hun score op de voortoets, tegen de achtergrond van de scores van alle 800 leerlingen in dit onderzoek. We vermijden hiermee kwalificaties die de scholen zelf hanteren en die van school tot school verschillende betekenissen hebben zoals: LBO-leerlingen, leerlingen met Mavo-advies, L/M-leerlingen, A,B en C niveau enz.

In tabel 5.3 volgt een overzicht per conditie.



Tabel 5.3: *Indeling leerlingen naar voortoetsscore (L,M,H) per conditie*

	Laag	Midden	Hoog	Rij Totaal
Project- leerlingen	216 42.3	162 31.7	133 26.0	511 N 63.0 %
Vergelijking leerlingen	40 13.3	110 36.7	150 50.0	300 N 37.0 %
Kolom Totaal	256 31.6	272 33.5	283 34.9	811 N 100.0 %

Aantal ontbrekende leerlingen = 18

Het percentage leerlingen met een hoge score op de voortoets is in de vergelijkingsscholen 50%. In de projectscholen is dat percentage 26%. Als we kijken naar de leerlingen met een lage score dan zijn die percentages voor de leerlingen in vergelijkingsscholen en projectscholen respectievelijk 13 en 42 procent. Deze getallen illustreren de verschillen in voorkennis bij de leerlingen in beide condities. De gemiddelde voortoetsscores zoals in tabel 5.4 weergegeven corresponderen hiermee.

Tabel 5.4: *Gemiddelde voortoetsscores per conditie*

Conditie	Gemiddelde	Standdaarddeviatie	N
Totaal	31.7	10.5	811
Projectleerlingen	28.9	9.6	511
Vergelijkingsleerlingen	36.4	10.1	300

Uit tabel 5.4 blijkt dat de projectleerlingen gemiddeld 7.5 punten lager scoren op de voortoets dan de vergelijkingsscholen.  
Een kruistabel per school geeft een overzicht van de verdeling van de leerlingen naar L, M, en H over de scholen.

Tabel 5.5: Indeling leerlingen naar voortoetsscore (L,M,H) per school

	Laag 1	Midden 2	Hoog 3	Rij Totaal
School 1	100 70.9	30 21.3	11 7.8	141 N 17.4 %
School 2	39 24.5	67 42.1	53 33.3	159 N 19.6 %
School 3	77 36.5	65 30.8	69 32.7	211 N 26.0 %
School 4	36 16.4	85 38.6	99 45.0	220 N 27.1 %
School 5	4 5.0	25 31.3	51 63.8	80 N 9.9 %
Kolom Totaal	256 31.6	272 33.5	283 34.9	811 N 100.0 %

Aantal ontbrekende leerlingen = 18

Tegen de achtergrond van de scores van alle 800 leerlingen zijn per school opmerkingen te maken. School 1 heeft relatief veel leerlingen met lage scores en relatief weinig leerlingen met hoge scores. Voor de vergelijkingsscholen (4 en 5) geldt tot op zekere hoogte het omgekeerde. De scholen 2 en 3 hebben een tamelijk "evenwichtige" verdeling van leerlingen, dat wil zeggen ze benaderen de verdeling (L,M,H) zoals die is in het totale bestand van 800 leerlingen.

Uit een literatuuroverzicht (Terwel, 1986a) blijkt dat de proporties leerlingen (L,M,H) in een school of klas van betekenis zijn voor de kwaliteit van de onderwijsleerprocessen. En dat kan weer effect hebben op de leerresultaten. Door samenvoeging van de proporties M en H leerlingen hebben we de informatie uit tabel 5.5 gebundeld in één percentage dat de proportie M en H leerlingen aangeeft. Een hoge proportie (M, H) is een belangrijke, overigens geen voldoende, voorwaarde

voor kwalitatief goed onderwijs. Voor de scholen 1 tot en met 5 zijn deze getallen respectievelijk 29, 75, 63, 84 en 95. De voorwaarden voor een goede kwaliteit van de onderwijsleerprocessen zijn op school 1 ongunstig en op school 5 gunstig.

Het is dus van belang om bij de analyse van de leereffecten niet alle scholen over één kam te scheren, maar rekening te houden met verschillen in beginsituatie.

De verschillen in schoolcompositie zoals weergegeven in tabel 5.5 weerspiegelen zich uiteraard ook in gemiddelde voortoetsscores per school

*Tabel 5.6: Gemiddelde voortoetsscores per school*

School	Gemiddelde	Standdaarddeviatie	N
Totaal	31.7	10.5	811
School 1	22.7	8.5	141
School 2	32.0	8.5	159
School 3	30.7	9.2	211
School 4	35.8	10.3	220
School 5	38.2	9.4	80

Er is dus een duidelijke rangorde in de gemiddelde voortoetsscores tussen de scholen. School 1 scoort laag. School 4 en 5 scoren hoog en school 2 en 3 liggen daar tussen in. Overigens is niet alleen de compositie van de leerlingenpopulatie per school van belang. De wijze waarop de school haar klassen samenstelt is ook van betekenis. Daarom volgen nu gegevens per klas.

Tabel 5.7: *Proportie M en H leerlingen en gemiddelde voortoetscore per klas*

Klas	Proportie (M,H)	Gemiddelde voortoets	SD	N
11*	60	28.6	8.7	20
12	19	18.8	9.1	16
13	5	20.2	5.5	19
14	26	22.7	6.2	19
15	30	25.6	7.8	20
16	21	22.4	8.1	19
17	46	23.3	10.7	13
18	27	18.2	7.8	15
21	75	32.2	7.3	28
22	85	33.0	8.6	27
23	85	32.7	7.8	26
24	57	29.0	9.2	28
25	75	31.6	8.6	24
26	77	33.5	9.5	26
31	38	25.2	7.3	21
32	71	30.7	6.4	31
33	97	39.8	8.3	32
34	20	22.0	6.1	20
35	50	27.1	7.0	26
36	97	38.9	8.6	29
38	72	30.2	6.2	28
39	33	25.4	5.4	24
41	50	27.9	6.4	26
42	78	31.9	7.2	23
43	77	31.3	6.0	26
44	85	33.3	7.5	26
45	93	37.7	9.9	29
46	82	31.9	7.0	28
47	100	45.8	11.3	31
48	97	42.7	10.2	31
51	96	39.5	12.1	27
52	100	39.5	6.9	27
53	88	35.6	8.1	26

\* Opmerking: een proportie van 60 bij klas 11 betekent dat in deze klas 60% leerlingen zijn uit de categorieën Midden en Hoog. Dat wil zeggen 40% van de leerlingen heeft een lage voorttoetscore (laag).

Uit tabel 5.7 zijn belangrijke gegevens af te leiden omtrent de samenstelling van de klassen.

De klassen in school 1 hebben een lage proportie leerlingen uit de M en H categorieën. Er is een uitschieter naar onderen (klas 13) en een uitschieter naar boven (klas 11).

School 2 vertoont een evenwichtige verdeling van haar leerlingen over de klassen. Blijkbaar voert deze (midden)school een beleid dat gericht is op de vorming van heterogene klassen. Met heterogene klassen is hier bedoeld dat de totale populatie tweede klassers van deze school zo is verdeeld, dat geen uitgesproken zwakke of sterke klassen ontstaan. Bovendien representeren deze klassen nagenoeg de proporties L-, M- en H-leerlingen zoals we die in de totale onderzoekspopulatie aantreffen. Deze klassen verschillen onderling relatief weinig ook voor wat betreft de gemiddelde score op de voortoets.

School 3 voert een heel ander beleid bij de samenstelling van klassen dan school 2. We zien hier twee klassen met veel M, H leerlingen: klas 33 en 36. Dan zijn er drie klassen met relatief weinig M, H leerlingen: klas 31, 34 en 39. De overige liggen daartussen in.

School 4 vertoont een soortgelijk beeld als school 3, zij het op een iets hoger niveau. Er zijn ook hier drie groepen klassen. Klas 41, 42 en 43 vormen de klassen met de laagste proportie M, H leerlingen. Klas 47 en 48 zijn uitschieters naar boven. De overige klassen liggen hier tussen in.

School 5 heeft overwegend M en H leerlingen. Klas 53 ligt wat lager dan klas 51 en 52.

Er zijn dus aanzienlijke verschillen in samenstelling van de klassen als men kijkt naar de score op de voortoets. Eén school (2) stelt heterogene klassen samen.

Drie scholen (3, 4 en 5) passen een vorm van streaming of setting toe. De andere school (1) ligt hier in bepaald opzicht tussen in: er zou een beleid achter kunnen zitten om bewust één sterke klas en één zwakke klas te maken. Het kan ook toeval zijn. Bij navraag blijkt dat het niet de bedoeling was om een sterke en een zwakke klas samen te stellen.

Wat heeft de verkenning naar de klassecompositie nu opgeleverd? De belangrijkste opbrengst is dat we per klas een maat hebben voor de classesamenstelling. Deze maat is herleidbaar tot één relevant criterium namelijk de scores van de leerlingen op toets 1. Daardoor zijn we niet afhankelijk van de criteria die scholen zelf hanteren bij de samenstelling



van de klassen: deze criteria wisselen van school tot school, terwijl elke school weer andere termen hanteert om zwakke en sterke klassen aan te geven. Met andere woorden er is nu een ondubbelzinnig criterium voor de mate van heterogeniteit of homogeniteit van klassen, los van kwalificaties die scholen daaromtrent zelf verstrekken en die niet onder één noemer te brengen zijn.

Nu de beginkennis door middel van de voortoetsgegevens van de leerlingen in kaart is gebracht gaan we de resultaten van de alle toetsen presenteren en enkele vergelijkingen trekken.

### 5.5.2 De toetsresultaten op de drie toetsen

Nu volgen vergelijkende gegevens over de resultaten op toets 1, 2 en 3.

*Tabel 5.8: Frequentieverdelingen toets 1, 2 en 3, in percentages*

punten behaald op de toetsen	Toets 1			Toets 2			Toets 3		
	Tot.	Proj.	Verg.	Tot.	Proj.	Verg.	Tot.	Proj.	Verg.
0 - 10	1.4	2.2	-	1.6	3.0	-	1.3	1.6	.7
10 - 20	9.5	13.9	2.0	8.2	11.5	4.1	3.9	4.9	2.1
20 - 30	38.0	44.2	27.3	31.2	38.9	21.8	18.1	19.4	15.9
30 - 40	34.0	28.6	43.3	34.1	34.8	33.2	27.5	28.4	26.0
40 - 50	12.0	8.4	18.0	15.5	8.9	23.6	23.9	22.9	25.6
50 - 60	3.8	2.3	6.3	6.7	2.2	12.3	14.6	13.9	15.9
60 - 70	1.2	.4	2.7	2.0	.7	3.6	7.1	6.1	8.7
70 - 80	.1	-	.3	.6	-	1.4	3.1	2.2	4.5
80 - 90	-	-	-	-	-	-	.6	.6	.7

Allereerst vallen de grote verschillen op tussen de leerlingen. Deze verschillen nemen toe van toets 1 tot en met toets 3 als we letten op het totale interval waarbinnen alle scores vallen. Toets 1 laat duidelijke verschillen zien tussen projectleerlingen en vergelijkingsleerlingen: de laatsten zijn beter vertegenwoordigd in de hogere intervallen.

In tabel 5.8 is te zien dat van toets 1 naar toets 3 een verschuiving in de frequentie optreedt van de lagere intervallen naar de hogere intervallen. De leerlingen boeken een duidelijke leerwinst. Dat geldt voor de projectleerlingen maar ook voor de vergelijkingsleerlingen, zij het dat er een inhaaleffect ten gunste van de projectleerlingen optreedt. Deze vooruitgang in het algemeen en voor de projectleerlingen in het bijzonder, is ook duidelijk zichtbaar in de gemiddelde toetsscores in tabel 5.9.

Tabel 5.9: Gemiddelde scores op toets 1, 2 en 3 uitgesplitst naar projectleerlingen en vergelijkingsleerlingen voor de categorieën Laag, Midden en Hoog

		Toets 1			Toets 2			Toets 3		
		Gem	SD	N	Gem	SD	N	Gem	SD	N
Totaal	Totaal	31.7	10.5	811	34.0	11.8	483	41.6	14.7	763
	project	28.9	9.6	511	30.0	9.8	266	40.2	14.5	482
	vergelijking	36.4	10.1	300	38.8	12.3	217	43.9	14.9	281
Laag	totaal	20.9	5.0	256	24.1	7.5	159	29.5	9.0	235
	project	20.4	5.1	261	23.7	7.3	124	29.5	9.0	197
	vergelijking	23.4	3.3	40	25.9	7.7	35	29.2	9.0	38
Midden	totaal	30.5	2.3	272	33.3	7.0	168	40.5	10.9	265
	project	30.2	2.2	162	32.3	6.3	84	41.6	10.9	158
	vergelijking	30.9	2.2	110	34.4	7.5	84	38.8	10.7	107
Hoog	totaal	42.6	7.7	283	44.7	10.5	156	53.5	12.9	263
	project	41.1	6.2	133	40.4	8.3	58	54.9	11.4	127
	vergelijking	44.0	8.6	150	47.3	10.8	98	52.1	14.0	136

Tabel 5.9 levert belangrijke gegevens op. In deze tabel zijn de resultaten op de toetsen 1, 2 en 3 opgenomen. Daarbij zijn verschillende indelingen gemaakt. Allereerst is er een indeling Totaal, Laag, Midden en Hoog. Totaal slaat daarbij op alle ruim 800 leerlingen. Laag, Midden en Hoog betreft zoals gezegd een indeling in drie niveau groepen naar de score op toets 1. Het betreft steeds ongeveer één derde deel van de leerlingen. Vervolgens zijn deze rubrieken verdeeld in projectleerlingen en vergelijkingsleerlingen.

Toets 2 is opgenomen met het oog op het oplossen van het vraagstuk van regressie naar het gemiddelde. Toets 1 zou daarbij fungeren als het criterium voor de indeling van de leerlingen in Laag, Midden en Hoog. Toets 2 zou dan als de eigenlijke voormeting worden gebruikt. Omdat op toets 2 veel leerlingen zijn uitgevallen kan dat plan niet doorgaan. Toets 1 zal dus als voormeting worden gezien. Toch is de uitslag van toets 2 niet onbelangrijk. We gebruiken deze uitslag om een indicatie te krijgen van het regressie-verschijnsel. Als er een regressie-effect is dan valt te verwachten dat de categorie Laag relatief veel "winst" boekt van toets 1 naar 2 en dat bij de categorie Hoog een terugval optreedt. Dit patroon is echter niet duidelijk aanwezig. Vrijwel steeds is er een verschil in

74 19 juni 2012!

ik heb een covariante analyse  
 gedaan: controle geeft sign. effect 4% (.036)  
 Est. Marginal means: controlegr. 38,9  
 Exp. av. 43,1

gemiddelde score tussen toets 1 en 2 van 2 à 3 punten. Er is één uitzondering. De projectleerlingen in de categorie Hoog vallen terug van 41.1 punten naar 40.4 punten voor respectievelijk toets 1 en toets 2. Dit zou op een regressie-effect kunnen duiden. Toch is ook niet uit te sluiten dat deze terugval (mede) veroorzaakt is, door het uitvallen van school 3 op toets 2. Vooralsnog lijkt de conclusie gewettigd dat er niet of nauwelijks sprake is van een regressie naar het gemiddelde.

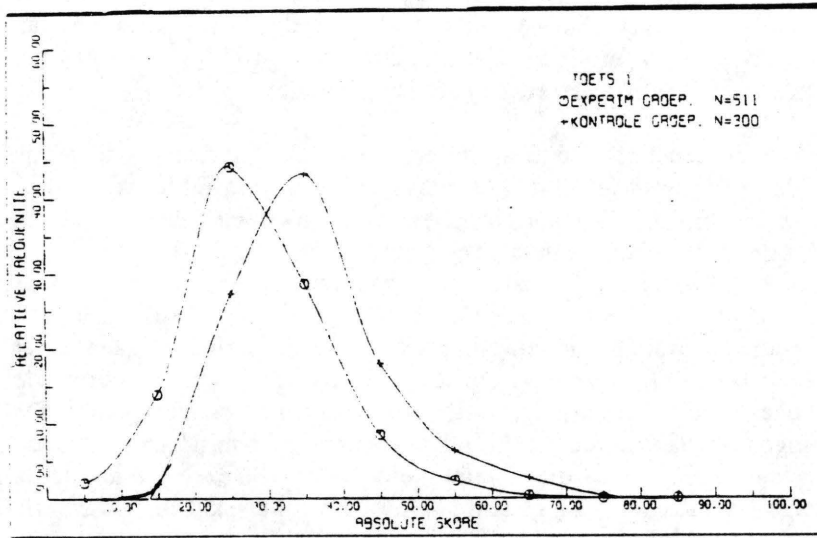
Er zijn drie hoofdconclusies te trekken uit tabel 5.9. Allereerst vallen ook hier weer de grote verschillen tussen de leerlingen op. Deze verschillen nemen toe gelet op de standaarddeviaties. Deze toename van verschillen tussen leerlingen geldt in beide condities en voor alle niveaus.

Een tweede hoofdconclusie betreft de vooruitgang van toets 1 naar toets 3 (Leerwinst). De totale groep leerlingen boekt een vooruitgang van toets 1 naar toets 3 van gemiddeld 10 punten. De H-leerlingen gaan meer vooruit van toets 1 naar toets 3 dan de L-leerlingen. Hoe presteren de projectleerlingen in verhouding tot de vergelijkingsleerlingen? De projectleerlingen behalen gemiddeld 11.3 punten leerwinst en voor de vergelijkingsleerlingen is dit gemiddeld 7.5. Deze gemiddelde leerwinstscores verschillen significant. Ook als we kijken naar de categorieën Laag, Midden, en Hoog valt de leerwinst steeds ten gunste van de projectleerlingen uit. Ook die verschillen zijn statistisch significant. Het verschil in leerwinst tussen de projectleerlingen en vergelijkingsleerlingen is het hoogst in de categorie Hoog: deze projectleerlingen boeken 6 punten méér leerwinst dan de vergelijkingsleerlingen.

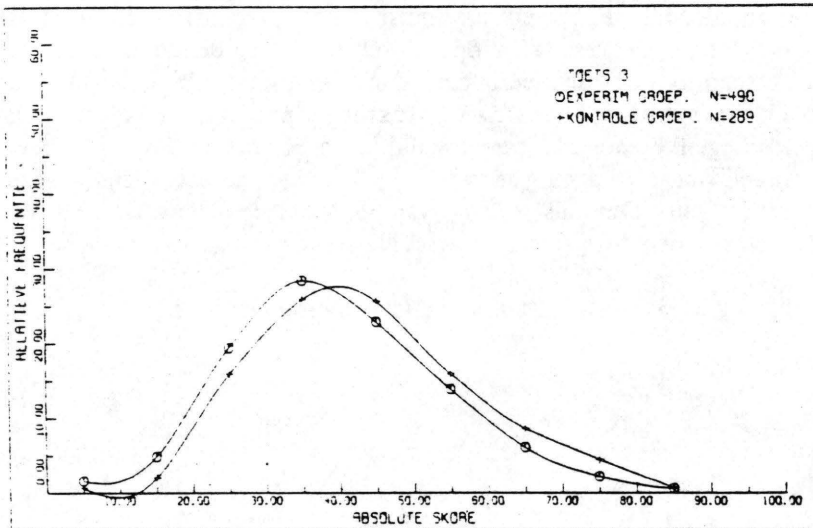
De derde hoofdconclusie betreft de toetsresultaten op toets 1 en toets 3. Als men kijkt naar de resultaten op de voortoets (1) en de natoets (3) per niveau categorie van de leerlingen dan blijkt dat op de voortoets significante verschillen bestaan tussen projectleerlingen en vergelijkingsleerlingen binnen de categorieën L en H, die in het voordeel van de vergelijkingsleerlingen uitvallen. Op de natoets zijn deze verschillen verdwenen. Dus als gevolg van de grotere leerwinst van de projectleerlingen, worden deze aanvangsverschillen ongedaan gemaakt.

Nu volgen enkele grafieken ter illustratie van deze hoofdconclusies.

**Figuur 5.1:** *Relatieve frequenties op toets 1 van projectleerlingen en vergelijkingsleerlingen*



**Figuur 5.2:** *Relatieve frequenties op toets 3 van projectleerlingen en vergelijkingsleerlingen*

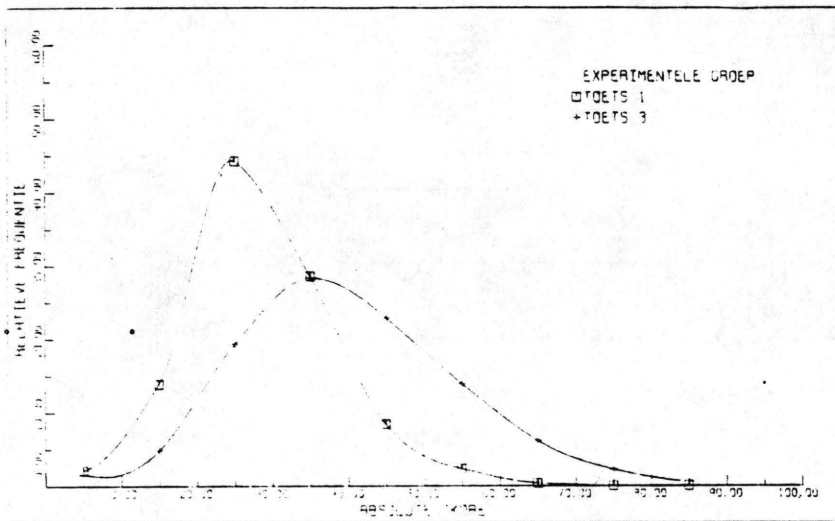


De grafieken in figuur 5.1 en 5.2 illustreren enkele belangrijke conclusies. Allereerst is te zien dat de verschillen tussen de groep projectleerlingen en de groep vergelijkingsleerlingen afnemen van toets 1 naar toets 3 (Inhaaleffect).

De spreiding tussen de leerlingen op toets 1 is groot. Deze spreiding neemt nog aanzienlijk toe tijdens de periode waarin het onderzoek plaats vond. De grote initiële verschillen en de toename hiervan in de tijd zijn bij beide groepen te constateren. De toenemende spreiding is te constateren in beide condities en is dus niet een specifiek effect van het doorlopen van het nieuwe wiskunde-programma.

We kijken nu nog naar een grafiek waarin de verschuiving van voortoets naar natoets tot uitdrukking komt.

**Figuur 5.3:** *Relatieve frequenties van scores op toets 1 en 3 van de groep projectleerlingen*

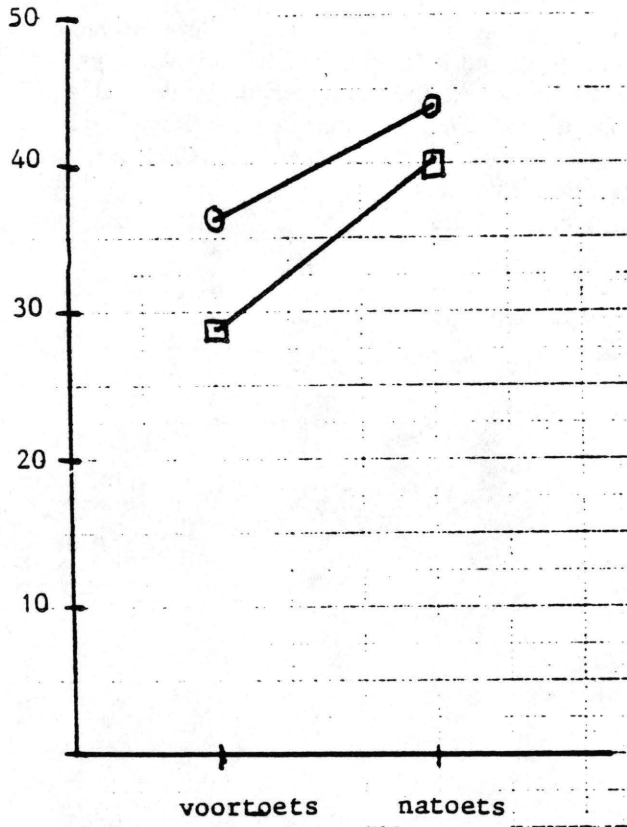


Figuur 5.3 brengt de leerwinst en de toename van de verschillen tussen leerlingen in beeld. Dit verschijnsel is ook te constateren in tabel 5.9.

De hoofdconclusie die uit tabel 5.9 getrokken kan worden komt duidelijk tot uitdrukking in figuur 5.4. In deze figuur zijn de gemiddelde scores op de voortoets (toets 1) en de natoets (toets 3) in een grafiek afgebeeld.



Figuur 5.4: Gemiddelde score voortoets en natoets voor projectleerlingen en vergelijkingsleerlingen



○ = vergelijkingsleerlingen  
□ = projectleerlingen

In figuur 5.4 is het verloop van de gemiddelde scores van voortoets naar natoets aangegeven. De hoogste lijn (met de cirkeltjes) is die van de vergelijkingsleerlingen. De laagste lijn is die van de projectleerlingen (met de vierkantjes). De vergelijkingsleerlingen beginnen hoger en eindigen ook hoger. De projectleerlingen vertonen echter een grotere vooruitgang. Er is een inhaaleffect.

Figuur 5.4 representeert de hoofdconclusie uit de vergelijking tussen de

twee condities: er is een betekenisvolle afname van verschillen tussen projectleerlingen en vergelijkingsleerlingen. Om deze hoofdconclusie te specificeren en met name aan te geven op welke inhoudsaspecten de sterkste effecten te zien zijn, volgt nu een analyse op itemniveau.

Als we de voortoetsgegevens van de projectleerlingen en de vergelijkingsleerlingen op itemniveau met elkaar vergelijken, dan zien we dat de vergelijkingsleerlingen op bijna alle items hoger scoren dan de projectleerlingen (zie bijlage 3a).

Als we de voor- en natoetsgegevens van de projectleerlingen met elkaar vergelijken zien we dat de projectleerlingen op alle items vooruitgaan. Deze vooruitgang is het sterkst op de items die betrekking hebben op lineaire verbanden in de vorm van grafieken (Tarief, Video, Gas). Dit onderdeel komt in 4 van de 6 pakketjes in het vernieuwingsprogramma aan de orde. Deze pakketjes zijn bewerkingen en uitbreidingen van pakketjes die in de vorige rondes van ons onderzoek zijn onderzocht.

Maar ook de vergelijkingsleerlingen gaan vooruit, zij het niet op alle items. Op item Video blijven ze bijvoorbeeld stilstaan en op item Gas gaan ze zelfs achteruit, hoewel lineaire verbanden in de vorm van grafieken ook in de vergelijkingsprogramma's ruim aan bod zijn gekomen.

Als we de natoetsgegevens van de projectleerlingen en de vergelijkingsleerlingen op itemniveau met elkaar vergelijken, dan zien we dat de projectleerlingen de vergelijkingsleerlingen op een aantal items voorbijgaan. Dit is weer het sterkst op de items die betrekking hebben op lineaire verbanden in de vorm van grafieken (Tarief, Gas). (Zie bijlage 3b).

We gaan nu wat gedetailleerder kijken naar de leerwinst. Naar aanleiding van tabel 5.9 trokken we al enkele conclusies over de vooruitgang van de leerlingen, dat wil zeggen de verschilscore tussen toets 3 en toets 1. In de rest van dit hoofdstuk spreken we steeds over de leerwinst van 763 leerlingen. Dat is het aantal leerlingen dat zowel toets 1 als toets 3 heeft gemaakt. Dit geeft enkele kleine verschillen in vergelijking met de eerder gepresenteerde tabel 5.9, omdat deze laatste tabel een verschillende N geeft voor de voortoets en de natoets. Overigens is er géén verschil in conclusies.

Gemiddeld, over alle leerlingen, behalen de leerlingen bijna 10 punten leerwinst ( $SD = 10$ ). Van deze leerlingen gaat 14 procent in score achteruit. De rest, 86 procent dus, blijft gelijk of boekt vooruitgang van toets 1 naar toets 3.

Tabel 5.10 geeft een overzicht van de verdeling van de leerwinstscores over de twee condities.

Tabel 5.10: Percentages leerlingen per categorie leerwinst

Leerwinst in punten	Totaal leerlingen %	Project leerlingen %	Vergelijkings leerlingen %
-30 tot -20	.3	-	.7
-20 tot -10	1.3	.8	2.1
-10 tot 0	14.9	10.6	22.4
0 tot 10	37.1	38.8	34.2
10 tot 20	31.6	33.6	28.1
20 tot 30	11.9	13.5	9.3
30 tot 40	2.4	2.1	2.8
40 tot 50	.5	.6	.4

De geringe leerwinst van de vergelijkingsleerlingen is zichtbaar in de proporties leerlingen in de intervallen. De vergelijkingsleerlingen vallen vaker terug van toets 1 naar 3 (negatieve "leerwinst"), terwijl zij minder goed vertegenwoordigd zijn in vrijwel alle intervallen met positieve leerwinst. Tabel 5.10 is nog verder te specificeren per school. We geven hier die specificatie niet in tabellen maar volstaan met enkele conclusies. Bij de twee vergelijkingsscholen en bij school 1 is er bij 25 procent van de leerlingen géén leerwinst of terugval. Dit verschijnsel doet zich op school 2 en 3 in veel mindere mate voor: respectievelijk 5 en 8 procent van de leerlingen. We gaan nu kijken naar de leerwinst per klas. Om de leerwinstscores per klas in relatie tot de gemiddelde voortoets- en natoetscores te kunnen bezien zijn in tabel 5.11 óók deze gemiddelde toetscores opgenomen. Deze tabel geeft zo een compleet overzicht per klas, waarbij alleen die leerlingen zijn opgenomen die zowel de voortoets als de natoets hebben gemaakt.

Tabel 5.11: Gemiddelde scores voortoets, natoets en leerwinst per klas (N = 763)

SD<sub>2</sub> - SD<sub>1</sub>

Klas	N	Gemiddelde voortoets	SD voortoets	Gemiddelde natoets	SD natoets	Gemiddelde leerwinst	SD leerwinst	
11	19	28.6	9.0	35.5	10.7	6.9	8.8	1,7
12	10	22.0	9.7	32.6	11.5	10.6	5.7	1,8
13	15	18.9	5.0	31.2	10.8	12.3	10.3	5,8
14	19	22.7	6.2	26.4	7.9	3.7	6.3	1,7
15	18	25.0	8.1	33.8	11.0	7.8	6.3	2,9
16	16	22.6	8.9	28.8	10.0	6.2	8.2	1,1
17	11	23.9	10.2	33.9	14.1	10.0	6.5	3,9
18	13	18.2	7.1	13.9	7.8	-4.4	8.0	0,7
<hr/>								
21	28	32.2	7.3	46.3	10.8	14.1	7.4	3,5
22	27	33.0	8.6	47.5	13.9	14.5	9.6	5,3
23	26	32.7	7.8	49.4	15.3	16.6	13.3	7,5
24	28	29.0	9.3	40.2	11.6	11.2	6.9	2,3
25	23	31.4	8.7	44.1	12.5	12.7	8.2	3,8
26	23	32.2	9.1	45.0	14.3	12.8	10.8	5,2
<hr/>								
31	20	25.4 1	7.5	33.8	7.9	8.4	6.2	0,4
32	31	30.7 2	6.4	45.3	11.3	14.6	8.4	4,9
33	31	39.0 3	7.1	55.0	13.0	16.0	9.0	5,9
34	20	21.0 1	6.1	30.2	8.2	8.2	6.8	2,1
35	26	27.1 2	7.0	39.1	10.3	12.0	6.8	3,3
36	29	38.9 3	8.7	53.4	12.1	14.5	9.0	3,4
38	27	29.9 2	6.0	41.5	9.3	11.7	7.3	3,3
39	22	25.4 1	5.6	30.4	8.2	5.0	6.0	2,6
<hr/>								
41	24	27.9	7.0	28.6	10.0	0.7	8.7	
42	23	31.9	7.2	38.7	9.5	6.8	5.4	
43	26	31.3	6.1	30.5	8.5	-.7	8.9	
44	24	34.0	6.9	40.0	10.6	6.1	8.4	
45	28	37.8	10.1	54.7	13.7	16.9	10.5	
46	28	32.0	7.0	38.9	9.1	7.0	9.5	
47	20	45.8	10.8	61.4	14.9	15.6	13.2	
48	30	42.5	10.3	53.2	12.2	10.7	11.1	
<hr/>								
51	26	39.6	12.4	50.4	12.0	10.8	10.6	
52	26	39.3	7.0	53.0	9.4	13.7	8.4	
53	26	35.6	8.2	34.0	8.3	-1.6	7.1	

De gemiddelde leerwinst per klas geeft enkele opvallende resultaten te zien. De klassen in school 1 boeken relatief weinig leerwinst. Bovendien vertonen deze leerwinstscores over de klassen een onrustig (fluctuerend) beeld. Klas 18 springt er helemaal uit met een terugval van 4 punten. Precies het omgekeerde geldt voor school 2: deze klassen vertonen relatief veel leerwinst terwijl het beeld over de klassen gezien stabiel is. School 3 geeft iets minder hoge en wat minder stabiele leerwinstscores over de klassen. In vergelijking met school 1 vertoont school 3 een positiever beeld.

De vergelijkingsklassen liggen gemiddeld wat lager in leerwinstscores dan de projectklassen. De klasgemiddelden van de vergelijkingsgroep geven een fluctuerend beeld: er zijn uitschieters naar boven en naar onderen. De gegevens over gemiddelde leerwinstscores per *klas* geven ook al een eerste indruk van de leerwinst per *school*. De volgende tabel bevat nadere gegevens daaromtrent.

Tabel 5.12: Gemiddelde scores voortoets, natoets en leerwinst per school

School	N	Gemiddelde voortoets	SD voortoets	Gemiddelde natoets	SD natoets	Gemiddelde leerwinst	SD leerwinst
Totaal	763	31.7	10.1	41.6	14.7	9.9	10.1
School 1	121	23.2	8.5	29.7	11.9	6.5	8.8
School 2	155	31.7	8.4	45.4	13.2	13.7	9.6
School 3	206	30.6	9.0	42.5	13.7	11.8	8.3
School 4	203	35.4	9.9	43.2	15.5	7.9	11.2
School 5	78	38.2	9.5	45.8	13.1	7.7	11.0

Tabel 5.12 bevestigt de indruk die al uit tabel 5.11 naar voren kwam. Van de drie projectscholen gaat school 1 relatief weinig vooruit. School 2 geeft de grootste leerwinst te zien. De verschillen tussen de drie projectscholen zijn significant. Tussen de twee vergelijkingscholen bestaan geen verschillen van betekenis.



Tabel 5.13: Gemiddelde scores voortoets, natoets en leerwinst per conditie

School	N	Gemiddelde voortoets	SD voortoets	Gemiddelde natoets	SD natoets	Gemiddelde leerwinst	SD leerwinst
School 1-2-3	482	29.1	9.3	40.2	14.5	11.1	9.3
School 2-3	361	31.1	8.8	43.7	13.6	12.6	8.9
School 4-5	281	36.1	9.9	43.9	14.9	7.8	11.1

Tabel 5.13 laat duidelijk zien dat er een verschil bestaat in gemiddelde leerwinst tussen projectleerlingen en vergelijkingsleerlingen. Dat verschil is significant. Het verschil in gemiddelde leerwinst tussen beide condities is 3.3. De standaarddeviatie van de natoets van de controlegroep is 14.9. Als we het verschil in leerwinst tussen beide condities uitdrukken in standaarddeviatie-eenheden (met de S.D. van de natoets van de controlegroep) dan is dat  $3.3 : 14.9 = .22$ . Deze maat wordt effect size genoemd. In dit geval duidt de effect size op een positief effect, al is het verschil niet groot te noemen. In tabel 5.13 hebben we naast de gebruikelijke conditie-indeling ook een andere indeling gehanteerd. Daarbij zijn school 2 en 3 samen genomen terwijl school 1, die speciale kenmerken heeft, buiten beschouwing blijft. We zien dan een nog sterker effect. School 2 en 3 halen de vergelijkingsscholen volledig in. Hun achterstand op de voortoets van 5 punten is bij de natoets geheel ingelopen! Er is een verschil in leerwinst van 4.8 punten in het voordeel van deze projectleerlingen. De effect size is nu .32. Als men de leerwinst per niveaugroep Laag, Midden, Hoog berekent en vervolgens per niveaugroep vergelijkingen trekt tussen projectleerlingen en vergelijkingsleerlingen valt het verschil steeds ten gunste uit van de eersten. Al deze verschillen zijn significant.

Nu volgt tabel 5.14 waarin de gemiddelde leerwinstscores per school nog weer zijn uitgesplitst naar beginniveau van de leerlingen (Hoog, Midden en Laag).

$0.2 - 0.14 = 0.06 = \text{zwak}$   
 $0.4 - 0.14 = 0.26 = \text{mooi}$   
 $0.6 - 0.14 = 0.46 = \text{sterk}$

\* Slavin:  
 \* Cohen  
 \* Frans Tenbrunsel  
 \* Bloom: The Sigma problem

Tabel 5.14: Leerwinst per school uitgesplitst naar Hoog, Midden en Laag

School	Gemiddelde leerwinst	SD	N
1 Totaal	6.5	8.8	121
Laag	7.4	8.3	84
Midden	2.8	10.0	26
Hoog	8.2	7.4	11
2 Totaal	13.7	9.6	155
Laag	11.2	7.3	39
Midden	15.0	10.7	67
Hoog	13.8	9.2	49
3 Totaal	11.8	8.3	206
Laag	9.5	6.3	74
Midden	11.1	8.1	65
Hoog	15.2	9.6	67
4 Totaal	7.9	11.2	203
Laag	5.6	8.8	34
Midden	8.3	10.3	82
Hoog	8.3	12.7	87
5 Totaal	7.6	11.0	78
Laag	6.1	13.5	4
Midden	6.5	10.6	25
Hoog	8.4	11.2	49

*middelste  
keten*

*afstroom  
groep*

Uit tabel 5.14 blijkt dat de leerwinst per school wordt gereflecteerd in de leerwinst voor de drie categorieën Hoog, Midden en Laag. Met andere woorden de gemiddelde leerwinst per school komt weer terug in de categorieën Hoog, Midden en Laag van elke school.

Met uitzondering van school 1 geldt voor alle scholen dat de leerlingen in de categorie Laag relatief, dat wil zeggen ten opzichte van de M en H leerlingen binnen elke school, weinig leerwinst boeken. Op school 1 gaat de Middengroep opvallend weinig vooruit (2.8 punten). De vraag is echter of dit enige praktische of statistische betekenis heeft. Overigens komen school 2 en 3 in deze tabel relatief gunstig naar voren.

Een vergelijking van school 2 en 3 levert enkele gezichtspunten op die overeenkomen met gegevens uit de literatuur omtrent heterogene versus homogene groepen. School 2, de middenschool met heterogene groepen, boekt meer leerwinst bij de zwakke leerlingen dan school 3 (respectievelijk 11,2 en 9,5).

Gelet op de sterke leerlingen blijkt school 2 wat minder leerwinst te boeken dan school 3 (respectievelijk 13,8 en 15,2). Het gaat echter om zwakke tendenties die niet significant zijn. Omdat ze precies passen in de gegevens uit de research-literatuur vermelden we het toch. Het beeld dat uit recente research naar voren komt is grofweg gezegd: zwakke leerlingen presteren doorgaans meer in heterogene groepen, sterke leerlingen komen tot betere resultaten in homogene groepen. Beide effecten kunnen worden verklaard door de proportie sterke leerlingen in een klas. Hoe hoger die proportie hoe beter de resultaten voor alle leerlingen in die klas (vgl. Terwel 1986a). Daarmee komen scholen voor een moeilijk dilemma te staan. Moeten ze kiezen voor de zwakke leerlingen of voor de sterke leerlingen (Vgl. Terwel 1987b).

Nu volgen gegevens omtrent samenhangen tussen resultaten op de drie toetsen: eerst op individueel niveau dan op klasniveau.

Tabel 5.15: *Correlaties tussen toetsresultaten, individueel niveau  
(N tussen haakjes)*

	Totaal	Project- leerlingen	Vergelijkings- leerlingen
corr. tussen toets 1 en 2	.82 (483)	.80 (266)	.78 (217)
corr. tussen toets 1 en 3	.73 (763)	.78 (482)	.66 (281)
corr. tussen toets 2 en 3	.76 (448)	.76 (244)	.77 (204)

De correlaties op leerlingniveau (tabel 5.15) zijn tamelijk hoog. De hoge correlaties tussen toets 1 en 2 zijn te verwachten. Het betreft twee parallel-toetsen die kort na elkaar zijn afgenomen. De hoge correlaties geven een indicatie van de test-hertest betrouwbaarheid van de toetsen. Ook de correlaties tussen toets 1 en 3 zijn hoog, dat geldt vooral voor de projectleerlingen (.78). Ook tussen toets 2 en 3 zien we tamelijk hoge correlaties zowel voor de hele groep als binnen de twee condities.

Tabel 5.16: Correlaties tussen toetsresultaten, klasniveau (N tussen haakjes)

	Totaal	Project- klassen	Vergelijkings- klassen
corr. tussen toets 1 en 2	.95 (21)	.92 (13)	.98 (8)
corr. tussen toets 1 en 3	.88 (33)	.93 (22)	.92 (11)
corr. tussen toets 2 en 3	.91 (21)	.90 (13)	.97 (8)

Op klassenniveau (tabel 5.16) zijn zéér hoge correlaties aanwezig: vrijwel steeds .90 of hoger. Dat betekent dat in het algemeen klassen met een lage score op de voortoets ook een lage score hebben op de natoets. "Goede" klassen blijven goede klassen, "slechte" klassen blijven slechte klassen. Dit gegeven gaat op in beide condities. Hetzelfde geldt, zij het in mindere mate op leerlingniveau.

## 5.6 Conclusies en reflectie

Er werden drie toetsen afgenomen: toets 1 en 2 aan het begin, toets 3 aan het eind. In de periode tussen begin en einde werden op vijf scholen twee verschillende wiskunde curricula uitgevoerd: een nieuw wiskundeprogramma en een vergelijkingsprogramma. Op school 1, 2 en 3 werd het nieuwe programma uitgevoerd, terwijl school 4 en 5 een meer traditionele benadering volgden. Uit de gegevens zijn de volgende conclusies te trekken.

Tussen de leerlingen bestaan grote verschillen in (voor)kennis. Deze verschillen nemen toe in de loop van de onderzoeksperiode. Deze toename in verschillen geldt voor de projectleerlingen en voor de vergelijkingsleerlingen.



Tussen klassen bestaan grote verschillen in gemiddelde scores op de toetsen. Deze verschillen weerspiegelen het differentiatiebeleid, resp. de differentiatiepraktijk in de scholen. Drie scholen werken met meer of minder gehomogeniseerde klassen, dat wil zeggen klassen waarbij bewust een reductie in heterogeniteit is toegepast. Slechts één school in de categorie projectscholen werkt bewust met volledig heterogene klassen. De verschillen in gemiddelde scores op de toetsen op het niveau van de klassen blijven stabiel bij alle drie toetsen, d.w.z. er zijn hoge correlaties tussen de toetsen op klassenniveau.

Ook tussen scholen bestaan aanzienlijke verschillen in gemiddelde scores op de toetsen. Deze verschillen zijn ook duidelijk zichtbaar in de proportionele verdeling van leerlingen gelet op hun voortoetscore (Laag, Midden, Hoog). Zo heeft school 1 zeventig procent leerlingen met een lage score op de voortoets, op school 5 is dat percentage slechts vijf!

Ook tussen de condities dat wil zeggen tussen projectscholen en vergelijkingscholen bestaan aanzienlijke verschillen in gemiddelde voortoetscores (7.5 punt verschil). Projectscholen hebben relatief veel leerlingen met een lage score op de voortoets. Voor de vergelijkingscholen geldt (dus) het omgekeerde.

De totale groep van 763 leerlingen boekt gemiddeld 10 punten leerwinst van toets 1 naar toets 3. Projectleerlingen behalen ruim 11 punten leerwinst en vergelijkingsleerlingen bijna 8 punten. Projectleerlingen behalen gemiddeld 3,3 punt méér leerwinst dan de vergelijkingsleerlingen. De leerwinst valt ook in de categorieën Laag, Midden en Hoog steeds ten gunste uit van de projectleerlingen. Dit verschil is het duidelijkst in de categorie leerlingen met hoge scores op de voortoets.

Per school zijn er verschillen in gemiddelde leerwinst. School 1 blijkt van de projectscholen de school met de laagste leerwinstscores. School 2 boekt de meeste leerwinst, direct gevolgd door school 3. De twee vergelijkingscholen vertonen een ongeveer gelijke leerwinst. Klassen vertonen onderling grote verschillen in gemiddelde leerwinst. Eén klas behaalt 17 punten leerwinst, terwijl drie klassen een gering leerverlies te zien geven. Klassen met een gemiddeld lage score op de voortoets boeken duidelijk minder leerwinst dan klassen met een hoge score.

Op leerlingenniveau zijn de correlaties tussen de drie toetsen hoog te noemen. Tussen toets 1 en 2 zijn de correlaties voor zowel de projectleerlingen als de vergelijkingsleerlingen ongeveer .80. Tussen toets 2 en 3 is dat getal .76. Alleen bij de correlatie tussen toets 1 en 3 treedt

een betekenisvol verschil op tussen projectleerlingen en vergelijkingsleerlingen respectievelijk .78 en .66.

Op klassenniveau zijn de correlaties tussen de toetsen (klassegemiddelden) zelfs zeer hoog. Meestal is dit .90 of hoger.

De conclusie is dat in scholen en klassen waarin het vernieuwingsprogramma 'Wiskunde voor iedereen' is uitgevoerd, de leerresultaten beter zijn dan in vergelijkingsscholen met een traditioneel programma. Daarmee is hypothese B bevestigd voor wat betreft de toetsresultaten. In hoofdstuk zes volgen nadere analyses ter ondersteuning en nuancering van deze conclusie. Deze analyses zijn ondermeer bedoeld om ook langs andere weg dan via verschillcores tot conclusies te komen. Overigens spreken de tabellen 5.9, 5.12, 5.13 en figuur 5.4 duidelijke taal, ook zonder gebruik te maken van het problematische begrip verschillscore.

Opvallend is voorts dat school 2 (de middenschool) met een evenwichtige, proportionele verdeling van zwakke en sterke leerlingen over de klassen dat wil zeggen met uitdrukkelijk heterogene klassen, de beste resultaten boekt. Overigens dient ter relativering van die resultaten te worden opgemerkt dat deze school in vergelijking met de twee andere projectscholen een kleinere proportie zwakke leerlingen heeft (op schoolniveau).

Het nieuwe wiskunde programma van de SLO dat is bedoeld om in heterogene klassen te worden gebruikt, blijkt in school 2 relatief goed te functioneren. School 3 die een bepaalde vorm van streaming hanteert boekt overigens ook goede resultaten.

Ook los van een vergelijking zijn er enkele vragen te stellen, over de resultaten van de *projectleerlingen*.

Zoals gezegd zijn er door de leerplanontwikkelaars geen absolute criteria geformuleerd, waaraan het eindresultaat voor de projectleerlingen moet voldoen. Dat betekent dat we strikt genomen niet kunnen vaststellen of het programma aan de criteria van de ontwikkelaars voldoet. Overigens is er wel het gegeven dat de makers van het programma hebben verklaard dat de toets hun bedoelingen dekt en dat zij *alle* items van de toets belangrijk vinden. Tegen die achtergrond en gelet op de toetsresultaten, kan men zich afvragen of het ideaal van 'Wiskunde voor Iedereen' is bereikt. Daarbij zijn nog een aantal specifieke vragen te stellen:

1. Bereiken alle leerlingen een acceptabel minimumniveau? (zo'n 20 procent van de leerlingen behaalt een score op de natoets die lager ligt dan 30 punten).
2. Is een gemiddelde leerwinst van 11 punten een voldoende resultaat?

3. Is een gemiddelde score van 40 punten op de natoets een voldoende resultaat?
4. Zijn de topprestaties van voldoende niveau? (bijna één kwart van de leerlingen behaalt een score boven de 50 punten).
5. Hoe dient men toename van verschillen in prestaties tussen de leerlingen te waarderen?

Achteraf gezien is het verklaarbaar waarom de leerplanontwikkelaars geen absoluut criterium konden formuleren. Strikt genomen is een absoluut criterium een illusie. Niemand stelt een criterium zonder daarbij een referentiegroep in gedachten te hebben. Bij ervaren, deskundige personen is die referentiegroep gebaseerd op ervaring, respectievelijk onderzoek van wat in vergelijkbare groepen, bijvoorbeeld in voorafgaande jaren, is gepresteerd.

De leerplanontwikkelaars moesten een criterium formuleren bij (a) grotendeels nieuwe doelen, (b) een grotendeels nieuwe toets waarvan de moeilijkheidsgraad nog niet bekend was en (c) een groep leerlingen waarvan de beginscore nog niet was bepaald. De vraag naar een 'absoluut' criterium was voor de leerplanontwikkelaars een moeilijke opgave. Nog afgezien van de bedreiging die van het opstellen van zo'n criterium kan uitgaan. In situaties als deze lijkt het verstandiger om een vergelijkend onderzoek te doen, dat wil zeggen relatief in plaats van absoluut te evalueren. Dat hebben wij in tweede instantie dan ook gedaan. Op de keper beschouwd is er geen principieel verschil tussen een onderzoek op basis van absolute criteria en een vergelijkend onderzoek op basis van relatieve criteria. In essentie gebeurt hetzelfde: men vergelijkt een uitkomst met een criterium. Het criterium is bij absoluut evalueren ontleend aan een 'abstracte' groep in het hoofd van ervaren, deskundige mensen. Bij relatief evalueren is het criterium ontleend aan een 'reële, bestaande' groep: in ons geval een controlegroep met een traditionele, klassikale aanpak.

Een kernvraag die in een kritische reflectie op de toetsresultaten niet mag ontbreken betreft de moeilijkheidsgraad van de toetsen. Zoals gezegd, dekken de toetsen de doelstellingen van het programma. Op ons uitdrukkelijk verzoek om verwachtingen uit te spreken en criteria te formuleren, is de SLO nooit ingegaan. De SLO is niet verder gegaan dan het verstrekken van toetsopgaven en het legitimeren van de toets. Dit standpunt is als volgt vastgelegd in de notulen: "V. zegt dat ID 12-16 toch de toets heeft, de groep staat achter de toets, deze dekt wat de ontwikkelgroep met de pakketjes voor heeft. Zelfs alle vragen worden belangrijk gevonden. Wat wil ID nog meer?".

Gelet op het belang dat de SLO aan alle opgaven hecht en gelet op de

uitslag van de toets zou men willen weten of de toetsen moeilijker zijn geweest dan de SLO heeft verwacht. Omdat die verwachtingen vóóraf niet zijn geëxpliciteerd zullen we dat nooit weten. De moeilijkheidsgraad van de toetsopgaven is hoog, gelet op de verhouding gemiddelde score en de maximaal te behalen score op de voortoets en de natoets (zie ook bijlagen 3a en 3b). Uiteraard zou het mogelijk zijn geweest items die erg moeilijk blijken te zijn (achteraf) te verwijderen. Bij examens of bij selectiebeslissingen speelt ondermeer de p-waarde een rol bij het handhaven of verwijderen van een item. In ons geval gaat het echter niet om een toets ten behoeve van zak/slaag-beslissingen met betrekking tot individuele leerlingen. In dit onderzoek wordt niet de leerling, maar het programma geëvalueerd. Onze primaire zorg betreft de validiteit en representativiteit van de toets ten opzichte van het programma. Ook de betrouwbaarheid is een belangrijk aandachtspunt. Daarom hebben we achteraf geen items verwijderd. Dat zou immers betekenen dat we bepaalde doelstellingen niet in het onderzoek zouden betrekken. Notabene doelstellingen die door de ontwikkelgroep belangrijk worden geacht. We hebben bewust geen items uit de analyse weggelaten. Het zou verlies van informatie zijn geweest die essentieel is bij programma-evaluatie. Bovendien zou dit een verlaging van de betrouwbaarheid van de toetsen betekenen, daar elk item een positieve bijdrage blijkt te leveren aan die betrouwbaarheid.

Een vraag blijft wel wat het effect is geweest van deze (te?) moeilijke opgaven op de leerlingen. Heeft dit invloed gehad op de motivatie bij het maken van de toets? Waar liggen de grenzen met betrekking tot de moeilijkheidsgraad als men een valide beeld wil krijgen van wat leerlingen van het programma hebben opgestoken? Hadden we misschien met verschillende toetsen moeten werken: bijvoorbeeld voor elke groep (L, M en H leerlingen) een eigen toets? Maar staat dat niet haaks op de intentie van de leerplanontwikkelaars? Men vond alle items belangrijk (Wiskunde voor Iedereen). Achteraf gezien, nu blijkt dat het zelfs voor zeer ervaren leerplandeskundigen toch moeilijk is de moeilijkheidsgraad van een toets op voorhand in te schatten, zou het beter zijn geweest een proefafname op grote schaal uit te voeren. Na item-analyse zou men de toets definitief kunnen vaststellen.

Wellicht was het ook beter geweest als we 66k (beproefde) items betreffende meer traditionele doelen in de toets hadden opgenomen. Dat zou de vergelijkbaarheid ten goede zijn gekomen. Ook zouden dan de effecten van het SLO-programma kunnen worden bepaald met het oog op kosten en baten van traditionele en nieuwe doelen. Overigens merken we op dat dit achteraf-redeneringen zijn die toen volstrekt niet opportuun of haalbaar waren. Aanvankelijk lag het niet in de bedoeling om ook andere dan SLO-programma's in de evaluatie te betrekken. Pas later toen

een absolute evaluatie niet mogelijk bleek en toen ook het design met de varianten niet door kon gaan, zijn controlescholen in het onderzoek betrokken. Toen was er geen tijd meer voor bijstelling van de toets in de richting van meer traditionele doelen en inhoud (zie hoofdstuk 1).



## 6 INTEGRATIE VAN ONDERZOEKSGEGEVENS: RELATIES TUSSEN VARIABELEN

### 6.1 Vraagstelling en hypothese

We herhalen hier in het kort de vraagstellingen en hypothesen zoals omschreven in het eerste hoofdstuk.

In dit onderzoek staan vier vragen centraal:

1. Wat is de beginsituatie van de leerlingen (houding en voorkennis)?
2. Worden de onderwijsleerprocessen volgens de bedoelingen gerealiseerd?
3. Zijn de beoogde leereffecten opgetreden?
4. Wat zijn de relaties tussen beginsituatie, onderwijsleerprocessen en leereffecten.

In de voorgaande hoofdstukken is op de eerste drie hoofdvragen al ingegaan. De beginsituatie van de leerlingen is in kaart gebracht. Het onderwijsleerproces is beschreven en vergeleken. De leereffecten zijn bepaald. Nu gaat het om de vierde vraag: de relaties tussen beginsituatie, onderwijsleerproces en leereffecten. De vraag voor dit hoofdstuk is: wat zijn de relaties tussen (1) beginsituatie (2) onderwijsleerprocessen en (3) leereffecten?

Bij de beantwoording van die vraag gaan we primair uit van de 22 project-klassen waarin het nieuwe SLO-curriculum is gerealiseerd. Voor deze situatie is de vraagstelling, de algemene werkhypothese en de hierna te formuleren hypothese (A) in eerste instantie opgesteld. In tweede instantie betrekken we ook de vergelijkingsleerlingen in de analyse (Hypothese B).

De algemene werkhypothese is: hoe meer de onderwijsleerprocessen volgens de bedoelingen zijn gerealiseerd des te beter zijn de leerresultaten. Om de 'bedoelingen' nader aan te geven is deze hypothese inhoudelijk toegespitst.

- A. In klassen waarin de leraar een banende, taakstellende, (bege)leidende rol vervult en waarin de leerlingen taakgericht samenwerken aan opgaven in contexten, zijn de leerresultaten beter dan in klassen waarin dat minder vaak het geval is. (Zie ook continueringsaanvraag Dekker, Herfs, Terwel 1985 (d), blz. 10). (Zie ook hoofdstuk 1 en 3 in dit eindrapport).

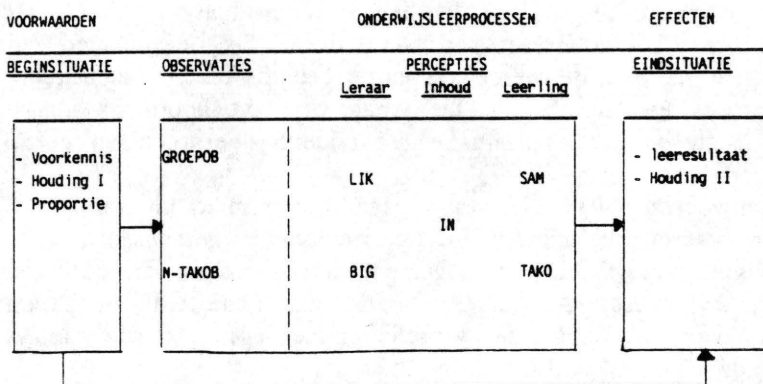
We hebben ook een hypothese geformuleerd omtrent het effect van het SLO-programma in vergelijking met een traditionele aanpak:

B. In onderwijs waarin met het nieuwe wiskundeprogramma van de SLO wordt gewerkt verlopen de onderwijsleerprocessen beter en zijn de leerresultaten beter dan in onderwijs met een meer traditioneel programma. Met 'beter' wordt gerefereerd aan dimensies, zoals aangelegd door de leerplanontwikkelaars.

## 6.2 Operationalisering en procedures

In het nu volgende model zijn de verschillende variabelen aangegeven. Dit model is een specificatie van het onderzoeksmodel zoals gepresenteerd in hoofdstuk 1.

*Figuur 6.1 Overzicht variabelen in model*



Hoewel alle variabelen, behalve de kwantitatieve observaties, in de voorgaande hoofdstukken zijn omschreven, lijkt het nuttig de betekenis kort aan te geven. Figuur 6.2 bevat per hoofdvariabele een korte omschrijving, een typering van het meetinstrument en een vermelding van het hoofdstuk waarin meer gedetailleerde informatie over deze variabele is te vinden.

*Figuur 6.2: Overzicht gegevens en instrumenten*

Gegevens (variabelen)	Omschrijving	Instrument	Hoofdstuk
Voorkennis	Wat leerlingen al weten van de wiskunde (individueel en per klas)	voortoets	5
Houding I	Houding van de leerlingen tegenover de school, wiskunde, enz.	vragenlijst houding I	4
Observaties	Systematische observaties in 22 project klassen	observatie-instrument (kwantitatief)	*
Percepties	Percepties van de leerlingen van het klassegebeuren: o.m. leraar en klas, begeleiding, samenwerking, taakgerichtheid	vragenlijst Percia	3
Leerresultaten	Wat leerlingen weten van de wiskunde ná het doorlopen van het curriculum	natoets	5
Houding II	Houding van de leerlingen tegenover de school, wiskunde, enz.	vragenlijst houding II	4

\* Zie Ruepert (1986) zie voorts de beschrijving onder observaties en bijlage 4.

De gegevens in figuur 6.2 zijn nader uiteen te leggen in variabelen, bijvoorbeeld subschalen op een vragenlijst. We geven deze variabelen kort aan. Voor gedetailleerde gegevens verwijzen we naar de betreffende hoofdstukken.

**Voorkennis:** de score op de voortoets (toets 1). Daarnaast hebben we een maat berekend voor de proportie zwakke en sterke leerlingen in een klas gelet op hun score op de voortoets (PROP). Deze proportiemaat geeft per klas het percentage leerlingen in een klas met een score M (midden) of H (hoog). Als men dit percentage van 100 procent aftrekt heeft men dus een maat voor de proportie zwakke leerlingen in een klas. (Score Laag). Let wel: PROP is steeds gebruikt als percentage M + H leerlingen. PROP slaat dus op het aandeel middelmatige en sterke leerlingen in een klas.

*Houding I:* Deze vragenlijst betreft de houding van de leerlingen. De vragenlijst heeft 6 subschalen: H-school, H-wiskunde, H-niveau, H-sexe, H-werkvormen, H-context. (Zie bijlage 2 voor het instrument en hoofdstuk 4 voor de gegevens).

\* *Observaties:* We gebruiken twee observatie-variabelen voor het proces in de klassen: Groepob en N-takob.

In de 22 projectklassen zijn telkens 2 lessen systematisch geobserveerd. Van de 44 geplande observaties zijn er drie niet doorgegaan. Voor de drie ontbrekende observaties zijn gemiddelde waarden van de betreffende school ingevuld. Aan de hand van een vooraf opgesteld observatie-instrument werd iedere minuut gescoord welke werkvorm op dat moment gehanteerd werd. Tevens werd gescoord of er al of niet taakgericht werd gewerkt. De observaties zijn door T. Ruepert in het kader van zijn M.O.-B studie onderwijskunde verricht en in zijn scriptie beschreven (Ruepert, 1986). Uit dit onderzoek, dat vanaf het begin in nauwe relatie tot het ID 12-16 onderzoek is opgezet, gebruiken we twee variabelen: Groepob en N-takob.

Groepob is het percentage van de tijd dat in een bepaalde klas aan groepswork is besteed. Als men dit percentage aftrekt van 100 procent dan heeft men een maat voor het percentage klassikale activiteiten in een klas (doceren, onderwijsleergesprek, klasgesprek). Het bleek namelijk dat individueel werken nauwelijks voorkwam (volgens Ruepert 1986, blz 7 tabel 1).

N-takob is het percentage (van de tijd aan groepswork besteed) waarin niet-taakgericht werd gewerkt. Niet taakgericht werd gescoord als meer dan 25% van de leerlingen in de klas zichtbaar of hoorbaar met andere zaken dan de taak bezig waren. De gegevens per klas zijn in bijlage 4 opgenomen. De gegevens in bijlage 4 zijn ook op zichzelf al interessant omdat ze duidelijk maken hoe groot de verschillen tussen klassen zijn in tijdsbesteding en taakgerichtheid. Notabene in klassen waarin één en hetzelfde (SLO) curriculum werd uitgevoerd. Overigens zijn van deze gegevens geen betrouwbaarheidsmaten berekend. Er was slechts één observator.

Deze twee variabelen uit observaties zijn in het licht van de hypothese relevant, omdat daarin de verhouding klassikaal werken versus groepswork en de al of niet taakgerichte samenwerking tijdens het groepswork een belangrijke rol spelen. En door Groepob en N-Takob te vergelijken met de klasgemiddelden van de PERCIA-schaal krijgen we zicht op de samenhang tussen observatiegegevens en perceptiegegevens. Beide groepen gegevens zijn bedoeld om informatie te leveren over het feitelijk verloop van de onderwijsleerprocessen in de klassen. Het betreft informatie over in principe hetzelfde object. De operationalisaties, methoden en instru-

menten zijn echter verschillend.

*Percepties:* De Percia - vragenlijst bevat 5 subschalen. LIK: Leiding door de leraar, klassikale instructie en klasseklimaat. BIG: Begeleiding van individuele leerlingen en groepjes door de leraar. SAM: adequate samenwerking in de groepjes. TAKO: taakgerichtheid en orde. IN: wiskundige inhoud, in het bijzonder wiskunde in contexten. Hoe hoger de score op de Percia-vragenlijst, hoe meer het programma volgens de bedoelingen is gerealiseerd. (zie bijlage 1 voor het instrument en hoofdstuk 3 voor de gegevens).

*Houding II:* Deze vragenlijst is identiek aan vragenlijst I. Bij de tweede afname is aan deze vragenlijst een semantische differentiaal toegevoegd met drie subschalen: SD les, SD leraar, SD groep. (zie bijlage 2 voor het instrument en hoofdstuk 4 voor de gegevens).

*Leerresultaten:* De score op de natoets. Tevens is de leerwinst als indicator voor leerresultaten gehanteerd. Het betreft de verschilscore tussen toets 3 en toets 1. (zie bijlage 3 voor het instrument en hoofdstuk 5 voor de gegevens).

### 6.3 Correlaties op klassenniveau

Deze paragraaf bevat gegevens omtrent samenhangen tussen voorwaarden, processen in de klas en leerresultaten. Door middel van een 'advance organizer' is aangegeven welke correlaties achtereenvolgens aan de orde komen.



Figuur 6.3: Schematisch overzicht van te rapporteren correlaties (advance organizer)

	Observaties	Toetsen	Percepties
Observaties			
Toetsen	1e		
Percepties	2e	4e	
Houdingen	3e	5e	6e

- (1e) Eerst volgen gegevens over correlaties tussen observaties en toetsresultaten.
- (2e) Dan gaan we in op correlaties tussen observaties en percepties, enzovoort. Steeds gaat het om correlaties op klassenniveau binnen de categorie projectklassen (N=22). Wellicht ten overvloede merken we op dat correlaties nog geen causale verbanden aangeven.

Tabel 6.1 bevat gegevens over de relatie tussen het proces in de klas zoals geobserveerd en de toetsresultaten.

Tabel 6.1: Correlaties tussen observaties en toetsgegevens (Klassegemiddelden, N = 22)

	voortoets	natoets	leerwinst	prop.
Groepob	-.18	-.11	-.01	-.16
N-takob	-.67*	-.62*	-.46*	-.68*

Uit tabel 6.1 blijkt dat er vrij sterke, significante correlaties op klasseniveau bestaan tussen Niet-taakgerichte activiteiten (N-takob) van de leerlingen en de toetsresultaten.

De conclusie is duidelijk: hoe meer niet-taakgericht gedrag van de leerlingen hoe lager de scores op de toetsen en hoe lager de leerwinst (en omgekeerd). Voorts blijkt de proportionele verdeling van de leerlingen in de klassen samen te hangen met niet-taakgerichte activiteiten. Hoe meer zwakke (L) leerlingen in een klas des te meer niet-taakgerichte activiteiten (of omgekeerd hoe meer PROP (H + M-leerlingen) hoe minder N-takob). In gewoon Nederlands betekent dit dat in klassen met veel zwakke leerlingen meer gepraat, rumoer, ruzie en, niet taakgericht gedrag voorkomt dan in klassen met weinig zwakke leerlingen.

Het percentage van de lestijd besteed aan groepswork (Groepob) correleert negatief met toetsgegevens. Deze correlaties zijn echter laag en niet significant. Er is een zwakke tendens: hoe lager het niveau van de klas des te groter het aandeel groepswork (en des te kleiner het aandeel klassikale activiteiten). Of andersom. Nu volgen gegevens over correlaties tussen observaties en percepties.

Tabel 6.2: *Correlaties tussen observaties en percepties (klassegemiddelden, N=22)*

	Percia totaal	LIK	BIG	SAM	TAKO	INHOUD
Groepob	-.29	-.34	.13	-.27	-.30	-.11
N-Takob	-.52*	-.53*	-.004	-.32	-.54*	-.43*

De eerste rij van tabel 6.2 geeft negatieve, niet-significante correlaties tussen het percentage groepswork (Groepob) in de klassen en alle PERCIA-subschalen behalve begeleiding van individuele leerlingen en groepjes door de leraar (BIG). Er is dus een niet significante tendens dat naarmate het percentage groepswork toeneemt de klassikale, leerkracht-geleide activiteiten (LIK), de samenwerking tussen de leerlingen (SAM) en de taakgerichtheid (TAKO) afnemen. Dat is begrijpelijk (zie hoofdstuk 1 par. 1.3). Ook het feit dat de begeleiding van individuele leerlingen en groepjes toeneemt naarmate het aandeel groepswork in de les toeneemt is

in te zien. Als een leraar veel klassikaal werkt zal hij minder tijd aan begeleiding (BIG) kunnen besteden. Anders gezegd: leraren die veel tijd voor groepswerk uittrekken hebben veel tijd voor begeleiding van individuele leerlingen en groepjes.

De tweede rij van tabel 6.2 geeft significante correlaties in de verwachte richting. Er zijn vier conclusies te trekken over relaties tussen onderwijsleerprocessen zoals gepercipieerd door de leerlingen en niet-taakgerichte activiteiten zoals geobserveerd door een onafhankelijke observator.

In klassen waarin de onderwijsleerprocessen, zoals gepercipieerd door de leerlingen, volgens de bedoelingen verlopen, wordt vaker taakgericht gewerkt dan in klassen waarin de onderwijsleerprocessen minder (goed) volgens de bedoelingen verlopen. En andersom. (Vergelijk de correlatie tussen de perciaschaal als geheel en de observatie van niet-taakgerichte activiteiten: correlatie PERCIA totaal en N-takob,  $r = -.52$ ).

In klassen waarin de leraar duidelijk leiding geeft en relatief vaak klassikale activiteiten organiseert wordt vaker taakgericht gewerkt dan in klassen met een minder duidelijke leiding en minder klassikale activiteiten. (Vergelijk de correlatie tussen LIK en N-Takob,  $r = -.53$ ).

In klassen waarin volgens de leerlingpercepties minder taakgericht wordt gewerkt blijkt (ook) uit de observaties dat daar minder taakgericht wordt gewerkt dan in klassen met een hoge taakgerichtheid (volgens de percepties van de leerling). (Vergelijk de correlatie tussen Tako en N-takob,  $r = -.54$ ).

In klassen waarin de leerlingen de contexten als aantrekkelijk en verhelderend percipiëren wordt vaker taakgericht gewerkt dan in klassen waarin de contexten minder positief worden gewaardeerd. En andersom. (Vergelijk de correlatie tussen Inhoud en N-Takob,  $r = -.43$ ).

Tussen samenwerking (SAM) en Niet-taakgerichte activiteiten (N-Takob) bestaat een correlatie in de verwachte richting. Hoewel deze correlatie niet significant is gaat het in het licht van alle gegevens om een betekenisvolle relatie.

En de correlatie tussen begeleiding van individuele leerlingen en groepjes (BIG) en Niet-taakgerichte activiteiten (N-takob) is vrijwel nul. De afwezigheid van correlatie tussen BIG en N-takob, is te verklaren vanuit de mogelijkheid dat leraren begeleiding geven ook als reactie op niet taakgericht gedrag. Met andere woorden begeleiding helpt wel om één groepje of één leerling binnen een groepje tijdelijk weer aan het werk te

krijgen, maar elders kan dan de taakgerichtheid juist weer afnemen.

Tenslotte nog een opmerking over de validiteit van leerlingpercepties. Geven (gemiddelde) leerlingpercepties inderdaad informatie over het feitelijk verloop van het onderwijsleerproces in de klas? De gegevens in tabel 6.2 duiden op een positief antwoord op die vraag. De significante, negatieve correlatie tussen N-takob en Tako is een indicatie voor de validiteit van de gemiddelde perceptie van de leerlingen in een klas. Interessant is dat Ruepert deze correlatie tussen zijn observatiescore en de perceptiescores al in zijn scriptie voorspelde (Ruepert, 1986 blz 26). Als twee geheel verschillende metingen (observatie door een onafhankelijke observator en een klassegemiddelde van leerlingpercepties) redelijk hoog correleren, dan zegt dat iets over de validiteit van beide metingen. Met andere woorden klassegemiddelden van de Percia-schaal zeggen inderdaad iets over wat feitelijk is gebeurd in de klas. Dat die correlatie niet 1.00 is valt in te zien als we letten op het feit dat de afname van beide instrumenten op verschillende tijdstippen en over verschillende tijdseenheden heeft plaats gevonden. Voorts zijn er aanzienlijke verschillen in operationalisatie en methode. De observaties zijn gericht op kwantiteit van vóórkomen. De percepties slaan behalve op vóórkomen óók op kwaliteit van bepaalde processen. Bovendien bevat elke meting een error-term, dat wil zeggen een foutencomponent waardoor het verband nooit perfect kan zijn. Al met al is te concluderen dat gemiddelde leerlingpercepties informatie geven over het feitelijk verloop van onderwijsleerprocessen in de klas.

Tussen observatie-gegevens en de houding van de leerlingen bestaat een aantal significante verbanden. Zo blijkt het percentage van de tijd aan groepswerk besteed positief, significant te correleren met de houding van de leerlingen. Hoe meer groepswerk hoe positiever de totale houding van de leerlingen (Groepob en Houdinglijst 1: correlatie .46). Voorts is er een positieve niet-significante correlatie met houdinglijst 2.

Het sterkste verband tussen Groepob en aspecten van de houding betreft H-niveau. Hoe meer Groepswerk (Groepob) hoe hoger de leerlingen hun niveau bij wiskunde inschatten (en omgekeerd). Dat verband is bij de eerste meting .58 bij H-niveau en .52 bij de tweede meting van H-niveau.

Ook is er een stabiel significant verband tussen groepswerk en de houding tegenover meisjes en wiskunde. Hoe meer groepswerk des te positiever is de houding tegenover de waarde van wiskunde voor meisjes. (Respectievelijk .39 en .38 op Houdinglijst 1 en 2).

Tussen N-takob en Houding bestaan geen significante verbanden, behalve bij de houding van de leerlingen tegenover contexten. Na het doorlopen van het curriculum is er een significante negatieve correlatie

tussen niet-taakgerichte activiteiten en de houding tegenover contexten (-.36).

Nu de relaties tussen observatiegegevens en andere variabelen in kaart zijn gebracht gaan we naar relaties tussen toetsen en andere variabelen. Eerst volgt een verkenning van de relaties tussen toetsen en percepties van de leerlingen.

Tabel 6.3: *Correlaties tussen toetsen en percepties (klasegemiddelden, N = 22)*

	Percia	LIK	BIG	SAM	TAKO	INHOUD
Voortoets	.58*	.53*	-.17	.52*	.57*	.72*
Natoets	.59*	.55*	-.20	.56*	.56*	.73*
Leerwinst	.51*	.49*	-.14	.48*	.44*	.67*
Proporities	.50*	.49*	-.13	.32	.54*	.60*

Als we de Percia-scores per klas opvatten als een indicatie voor de mate waarin de onderwijsleerprocessen volgens de bedoelingen zijn gerealiseerd dan is een algemene conclusie te trekken.

Hoe meer de onderwijsleerprocessen volgens de bedoelingen zijn gerealiseerd des te beter zijn de resultaten op de toetsen (en omgekeerd). (Percia-totaalscore en natoetsscores: de correlatie is .59) Globaal genomen komt deze algemene conclusie overeen met de algemene hypothese. Deze algemene conclusie is als volgt te specificeren. Daarbij kijken we eerst naar relaties tussen percepties en natoets.

In klassen waarin de leraar relatief vaak klassikale activiteiten organiseert en goed leiding geeft zijn de resultaten op de natoets beter dan in klassen waarin dat minder (vaak) het geval is. (LIK en natoets: de correlatie is .55).

In klassen waarin de leerlingen adequaat samenwerken zijn de resultaten op de natoets beter dan in klassen waarin de leerlingen minder goed samenwerken. (SAM en natoets: de correlatie is .56).

In klassen waarin de leerlingen taakgericht werken zijn de resultaten op de natoets beter dan in klassen waarin veel tijd wordt verspild en waarin



veel gepraat en rumoer is. (Tako en natoets: de correlatie is .56).

In klassen waarin de leerlingen de inhoud (contexten) leuk vinden, de inhoud uit ervaring kennen en in klassen waarin de inhoud de wiskunde verheldert, zijn de resultaten op de natoets beter dan wanneer de onderwerpen minder leuk, onbekend en niet verhelderend zijn. (Inhoud en natoets: de correlatie is .73).

Er is één uitzondering op deze conclusies. Deze heeft betrekking op Begeleiding van individuele leerlingen en groepen (BIG). Hier slaat het teken om van positief naar negatief. De correlatie tussen BIG en natoets is -.20. Ook de correlaties van BIG en voortoets, en van BIG en leerwinst zijn negatief. Hoewel deze negatieve correlaties niet significant zijn gaat het om een consistent patroon dat als volgt te omschrijven is: In klassen met lage natoets resultaten geeft de leraar vaker begeleiding aan individuele leerlingen en groepjes dan in klassen met hoge natoets resultaten. Of andersom: in klassen waarin de leraar relatief vaak begeleiding geeft aan individuele leerlingen en groepen zijn de leerresultaten op de natoets slechter dan in klassen waarin de leraar minder vaak begeleiding geeft.

Het is mogelijk dat dit merkwaardige, overigens niet significante, verband een gevolg is van situaties waarin de leraar gaat begeleiden als reactie op een gebrek aan voorkennis en taakgerichtheid. Men kan zich dat als volgt voorstellen. In klassen met veel zwakke leerlingen is er veel behoefte aan uitleg en begeleiding. Pogingen van de leraar om deze hulp, primair door middel van klassikale uitleg te geven mislukken omdat deze leerlingen geen uitgebreide klassikale behandeling toelaten. Er wordt niet geluisterd en er ontstaan ordeproblemen. Als reactie daarop zegt de leraar: 'begin maar met de opdrachten'. Hij verlaat (te?) snel de centrale, frontale situatie en gaat decentraal verder met het begeleiden van individuele leerlingen en groepjes. Hij verliest de klas als geheel uit het oog. Vervolgens ziet hij geen kans om op alle vragen om begeleiding in te gaan. Er ontstaan wachttijden. Leerlingen gaan zich vervelen en er komen ordeproblemen. De leraar gaat brandjes blussen en komt aan het lesgeven en het begeleiden van het leerproces nog nauwelijks toe. Dan is de klas in een vicieuze cirkel terecht gekomen. De begeleiding, voor zover de leraar er aan toekomt, is op zichzelf wel effectief, maar zet te weinig zoden aan de dijk voor alle leerlingen. Deze begeleiding kan het gebrek aan klassikale uitleg niet compenseren. De leraar vlucht als het ware in een decentrale lesstructuur als reactie op een gebrek aan taakgerichtheid. Die lesstructuur roept vervolgens nog meer taak- en ordeproblemen op. En zo is de cirkel rond.

Er is echter nog een andere interpretatie mogelijk, namelijk dat de begeleiding van individuele leerlingen en groepjes in de projectklassen als

zodanig niet effectief is! Dit is een interessante interpretatie in het licht van de discussie tussen onderzoekers en leerplanontwikkelaars over dit punt (zie hoofdstuk 1 over de inhoudelijk-theoretische achtergrond). Zoals gezegd hebben de leerplanontwikkelaars géén procedures uitgewerkt voor de diagnostische toetsing van leerlingen en van de begeleiding op basis van deze toetsgegevens. Uit een enquête onder de leraren (die we overigens niet in de rapportage hebben opgenomen) blijkt dan ook dat deze procedures niet of nauwelijks in de klaspraktijk voorkomen. De begeleiding die de leraren geven is dus gebaseerd op andere informatie over de individuele voortgang van de leerlingen: bijvoorbeeld op een cijfer van een proefwerk en op observaties in de klas. Het is denkbaar dat deze twee informatiebronnen niet geschikt (proefwerk), respectievelijk te incidenteel (observatie) zijn voor het geven van adequate begeleiding. Dat zou ook een verklaring kunnen zijn voor het merkwaardige verband tussen BIG en toetsgegevens. Het is interessant om de items van BIG bij deze interpretatie te betrekken. In deze items wordt ondermeer gesproken over toetsen die individueel gemaakt moeten worden en over extra werk op basis van die toetsuitslag. Deze items zijn uitdrukkelijk ontwikkeld met het oog op diagnostische toetsing. Omdat dit laatste ontbreekt zullen de leerlingen bij toets dus gedacht hebben, aan proefwerken voor een cijfer (die worden nl. wel gegeven). Dit gegeven kleurt deze vragen bij BIG op een wijze in, die niet door de onderzoekers is beoogd. Zij gingen er bij de vragenlijstconstructie vanuit dat er in de klassen diagnostische toetsing zou plaatsvinden. Zelfs al zou de SLO geen procedures ontwikkelen dan zullen docenten misschien zelf bepaalde maatregelen nemen, dachten we. Dat alles blijkt niet het geval. Daarmee slaan de twee vragen over de toetsen in de perceptie van de leerlingen op (selectieve) proefwerken in plaats van op diagnostische toetsingen. En dat kan invloed op de onverwachte uitkomst hebben gehad. Op dit punt zijn nadere analyses nodig.

De vijf hierboven geformuleerde conclusies met betrekking tot de relaties tussen percepties en natoetsscores gaan ook op voor de relaties tussen percepties en leerwinst. Dat deze op zichzelf interessante (vanuit de hypothese gezien) bevindingen niet in causale zin kunnen worden geduid wordt duidelijk als we kijken naar de correlaties tussen percepties en voortoets. Deze correlaties zijn van dezelfde sterkte en gaan exact in dezelfde richting. Datzelfde geldt voor de correlaties tussen percepties en proporties M en H leerlingen (Percepties en Prop.). Het is dus ook mogelijk de beginsituatie (voortoets en prop.) te zien als determinant van de onderwijsleerprocessen. Met andere woorden: hoe gunstiger de beginsituatie hoe meer de onderwijsleerprocessen volgens de bedoelingen verlopen. Ook is een voor de hand liggende conclusie: hoe gunstiger de

beginsituatie (voortoets) des te gunstiger de eindsituatie (natoets). Er is, op klasseniveau, een zeer hoge correlatie tussen voor- en natoets (.93). We kunnen uit deze correlatie geen causaliteiten afleiden. Een vaak gebezigde uitspraak in dit verband is: 'correlation is no proof of causation'. In hoofdstuk 8 komen we hier op terug. Het zoeken naar verbanden tussen manipuleerbare variabelen (onderwijsleerprocessen) enerzijds en leerresultaten (natoets en leerwinst) anderzijds heeft tot nu toe wel hoge, significante correlaties opgeleverd maar een 'oorzaakelijkheid' is hiermee niet aangegeven.

Hoe staat het nu met de relaties tussen toetsresultaten enerzijds en houding van de leerlingen anderzijds? Tabel 6.4 geeft de correlaties.

*Tabel 6.4: Correlaties toetsresultaten en houding totaal (klassegemiddelden N = 22)*

	Houding I	Houding II	Houding Sem. Diff.
Voortoets	.15	.45*	.24
Natoets	.10	.53*	.30
Leerwinst	-.05	.46*	.20

De conclusie is dat de houding van leerlingen ná het doorlopen van het curriculum (houding 2) positief en significant correleert met voortoets, natoets en leerwinst.

Tussen Houdingen en Percepties van leerlingen bestaan ook correlaties. Tabel 6.5 geeft een overzicht.

*Tabel 6.5: Correlaties Percepties en Houdingen van leerlingen (klassegemiddelden N = 22)*

	Percia (totaal)	LIK	BIG	SAM	TAKO	IN
Houding I	.13	.11	.47*	-.27	.08	.20
Houding II	.39*	.32	.33	.17	.26	.59*
Houding Sem. Diff.	.64*	.59*	.46*	.32	.54*	.39*

Enkele bevindingen naar aanleiding van tabel 6.5 zijn:

In klassen waarin de leraar relatief vaak begeleiding geeft aan individuele leerlingen en groepen is de houding van de leerlingen positiever dan in klassen waarin dat niet het geval is (BIG en Houding 1, BIG en Houding zoals gemeten met de semantische differentiaal).

In klassen waarin de leerlingen de inhoud (contexten) aantrekkelijk en verhelderend vinden is de houding positiever dan in klassen waarin dat niet het geval is (IN en Houding 2, IN en Houding zoals gemeten met de semantische differentiaal).

Opvallend zijn de relatief hoge correlaties tussen Percepties en de houding op de Semantische differentiaal. Hier is vooral de houding ten opzichte van de leraar debet aan. Uit nadere analyses naar de relaties tussen Percia-subschalen en subschalen op de Semantische differentiaal blijkt het volgende. De houding tegenover de leraar correleert hoog met de Percia-totaal score (.68), met LIK (.67) en met TAKO (.67). De conclusie is als volgt te formuleren:

In klassen waarin het onderwijsleerproces volgens de bedoelingen verloopt (zoals gemeten met de Percia) is de houding tegenover de leraar positiever dan in klassen waarin dit niet het geval is (minder goede leiding en klassikale instructie, minder taakgerichtheid). Of andersom.

#### 6.4 Verschillen tussen leerlingen op individueel niveau

In het onderzoek is bijzondere aandacht voor verschillen tussen leerlingen. Twee verschillen tussen leerlingen komen in deze paragraaf aan de orde (1) beginkennis op wiskunde (Laag, Midden, Hoog) en de (2) sexe. De vraagstelling is: gaan verschillen qua voorkennis of sexe samen met verschillen op andere aspecten?

Ter beantwoording van die vraag zijn op alle instrumenten (en subschalen) gemiddelden bepaald. Vervolgens is door middel van univariate variantie-analyses nagegaan of de gevonden verschillen significant zijn. Tabel 6.6 bevat een overzicht van het resultaat van die analyse. In de eerste kolom (beginkennis) is een \* gezet wanneer er significante verschillen zijn tussen de gemiddelde scores van leerlingen uit de categorieën Laag, Midden en Hoog, en wanneer die verschillen positief corresponderen met voorkennis. Bijvoorbeeld als L-leerlingen laag scores op TAKO en als tegelijkertijd H-leerlingen daarop hoog scores en M-leerlingen daar tussen in liggen dan is dat met een asterisk(\*) aangegeven. Als er geen significante verschillen zijn, staat er n.s. Bij sexe is eveneens met een



asterisk(\*) aangegeven als er significante verschillen bestaan tussen jongens en meisjes. Een asterisk zonder toevoeging betekent een significant verschil waarbij jongens het hoogst scoren. Als meisjes het hoogst scoren staat er \* (meisjes).

Tabel 6.6: *Versillen betreffende beginkennis en sexe in relatie tot scores op toetsen, percepties en houdingen projectleerlingen (N = 511)*

	Beginkennis (L,M,H)	Sexe
Toets 1	*	*
Toets 2	*	*
Toets 3	*	*
Percia totaal	*	n.s.
LIK	*	n.s.
BIG	n.s.	*
SAM	n.s.	* (meisjes)
TAKO	*	n.s.
IN	*	*
Houding totaal 1	*	*
H- school 1	n.s.	n.s.
H- wiskunde	n.s.	*
H- niveau	*	*
H- sexe	n.s.	* (meisjes)
H- werkvormen 1	*	*
H- context	*	*
Houding totaal 2	*	*
H- school	n.s.	n.s.
H- wiskunde	n.s.	*
H- niveau	*	*
H- sexe	*	* (meisjes)
H- werkvormen	*	*
H- context	*	*
Houding Sem. Diff. totaal	*	*
Houding SD les	*	*
Houding SD leraar	n.s.	*
Houding SD Groep	n.s.	* (meisjes)



Tabel 6.6 geeft een overzicht van de verschillen tussen leerlingen. Hieronder volgen eerst enkele conclusies in zake verschillen in relatie tot voorkennis (eerste kolom tabel 6.6). Daarna komen verschillen in relatie tot sexe aan de orde.

#### *Toetsen*

Zoals te verwachten valt zijn er tussen leerlingen met lage, middelmatige en hoge scores op de voortoets significante verschillen op alle (andere) toetsen. In het bijzonder bij toets 1 is dit een vanzelfsprekendheid omdat de voortoets het indelingscriterium is. De bevindingen met betrekking tot toets 2 en 3 stroken met de eerder gevonden hoge correlaties tussen de resultaten op de verschillende toetsen.

#### *Percepties*

Over de percepties van de leerlingen is de volgende algemene conclusie te trekken. Leerlingen met een lage score op de voortoets percipiëren het onderwijsleerproces in de klas minder positief dan leerlingen met een hoge score op de voortoets. En omgekeerd. Men kan dit nog anders zeggen. Volgens de percepties van zwakke leerlingen verloopt het onderwijsleerproces minder volgens de bedoelingen dan volgens de percepties van goede leerlingen. Deze algemene conclusie is als volgt te specificeren. De zwakke leerlingen krijgen (volgens hun percepties) minder duidelijke uitleg en leiding van de leraar en het klimaat in de klas is minder positief in vergelijking met de perceptie van sterke leerlingen (LIK). Zwakke leerlingen percipiëren ook vaker een gebrek aan taakgerichtheid (gepraat, rumoer, ruzie) dan sterke leerlingen (TAKO). Opvallend is dat de wiskundig-didactische inhoud door de zwakke leerlingen als minder verhelderend en minder motiverend wordt ervaren dan door de sterke leerlingen. Zwakke leerlingen worden vaker door de rijke context in verwarring gebracht dan sterke leerlingen (IN).

De vraag is echter of deze verschillen in percepties tussen zwakke en sterke leerlingen moeten worden opgevat als individuele subjectieve reacties van leerlingen of dat het percepties zijn die herleidbaar zijn tot een verschil in objectieve onderwijsleersituaties. Wellicht spelen individuele, subjectieve impressies en objectieve, voor alle leerlingen geldende omstandigheden een rol. Wellicht is voor zwakke leerlingen, dat wil zeggen leerlingen met weinig voorkennis, de onderwijsleersituatie minder positief in objectieve zin. Alleen al de samenstelling van de klassen en subgroepjes is hierbij een factor van belang. Op twee van de drie projectscholen (1 en 3) is sprake van meer of minder homogene klassen. Dat impliceert dat in bepaalde klassen relatief veel zwakke leerlingen aanwezig zijn, terwijl andere klassen relatief veel sterke leerlingen

bevatten. Dat verklaart, tenminste voor een deel, waarom zwakke leerlingen hun situatie minder positief percipiëren. Hun onderwijsleersituatie is minder stimulerend, motiverend, ordelijk enzovoort dan voor de sterke leerlingen in de betere klassen. Dit effect kan nog worden versterkt door de veelal vrije keuze van subgroepjes. Zonder ingrijpen door de leraar kunnen zwakke en sterke groepjes ontstaan. Overigens is er geen significant verschil ten aanzien van BIG en SAM. Wellicht valt hier het verschil weg omdat zwakke leerlingen juist op deze twee aspecten hulp ervaren. Misschien compenseert dit de algemene negatieve perceptie van de zwakke leerlingen. Bij de leerlingen met een hoge voortoets-score kan het omgekeerd liggen.

Overigens blijkt dat de correlaties op leerlingniveau tussen voortoets en BIG en tussen voortoets en SAM, laag en niet significant zijn.

### *Houdingen*

Leerlingen met een lage voortoetscore vertonen een minder positieve houding tegenover verschillende aspecten van het onderwijs dan leerlingen met een hoge voortoetscore. En omgekeerd. Er is nog een interessante bevinding. Leerlingen met een hoge score op de voortoets staan positiever tegenover H-werkvormen. Het betreft hier een serie vragen die een voorkeur voor klassikale of individuele werkvormen betreft en een afwijzing van groepswork. Sterke leerlingen staan dus positief tegenover klassikale en individuele werkvormen terwijl ze groepswork minder waarderen. Voor zwakke leerlingen geldt het omgekeerde, zij waarderen juist het werken in groepjes. Overigens is deze subschaal (H-werkvormen) onvoldoende betrouwbaar.

De semantische differentiaal laat ook een positieve houding zien van sterke leerlingen in vergelijking met zwakke leerlingen. Overigens zijn er geen significante verschillen ten aanzien van de houding tegenover de leraar en de groep.

Meisjes en jongens vertonen betekenisvolle en statistisch significante verschillen op een groot deel van de variabelen.

(Zie de tweede kolom (sexe) in tabel 6.6). Uit deze tabel is een algemene conclusie te trekken over verschillen in toetsprestaties, percepties en houding tussen jongens en meisjes. Meisjes presteren minder op de toetsen, ze percipiëren hun leeromgeving minder positief en ook hun houding tegenover het wiskunde onderwijs is minder positief in vergelijking met de jongens. We specificeren deze algemene conclusie en geven vervolgens enkele opvallende uitzonderingen aan

### *Toetsen*

Allereerst valt op dat de jongens hoger scoren op alle toetsen dan de meisjes. De verschillen liggen in de orde van grootte van 3 tot 5 punten. Op de natoets zijn de gemiddelde scores voor meisjes en jongens respectievelijk 27.5 en 30.7 punten. Op de natoets is de gemiddelde score voor meisjes 38.0 punten terwijl de jongens hierop gemiddeld 43.0 punten halen. Deze verschillen zijn niet uniek voor dit programma. In de vergelijkingsscholen is de situatie verhoudingsgewijs niet anders. Het vraagstuk van meisjes en wiskunde is dus ook met dit nieuwe programma niet opgelost. Meisjes beginnen met een achterstand in beginkennis en ze lopen deze niet in. Het verschil lijkt eerder nog toe te nemen. In de vergelijkingssituatie lijken de verschillen tussen jongens en meisjes wat af te nemen. Het SLO-programma is dus in elk geval niet beter op dit punt dan het traditionele programma.

Ook als we kijken naar andere variabelen dan de toetsen zijn er belangrijke verschillen tussen jongens en meisjes, die doorgaans in het voordeel van de jongens uitvallen.

### *Percepties*

Gelet op de percepties van de leerlingen krijgen jongens vaker begeleiding in de groepjes en individuele hulp (BIG). De wiskunde inhoud (contexten) is voor meisjes minder vaak verhelderend en motiverend dan voor jongens.

### *Houdingen*

Ook is de houding van de meisjes doorgaans minder positief dan die van de jongens. Meisjes staan minder positief tegenover de wiskunde, de les, de leraar en de contexten uit het dagelijks leven. Ze schatten bovendien hun niveau bij wiskunde lager in dan de jongens. Die schatting correspondeert dus met hun scores op de toetsen.

### *Afwijkingen en vergelijkingen*

Interessant zijn enkele afwijkingen van dit patroon. Meisjes percipiëren adequatere samenwerking in de groepjes en staan ook positiever tegenover hun groepje dan jongens (vergelijk SAM,D Groep en H- werkvormen). We hebben uit observaties en gesprekken de indruk gekregen, dat meisjes en jongens vaker in gescheiden groepjes werken dan in gemengde groepjes. Wellicht is de samenwerking en de sfeer in meisjesgroepjes inderdaad beter dan in jongensgroepjes. Dit kan tot gevolg hebben dat de leraar meisjesgroepjes minder begeleiding geeft zoals uit de percepties van meisjes blijkt, terwijl meisjes vanwege hun geringere voorkennis juist wiskundig gezien wel extra begeleiding zouden kunnen gebruiken. Al deze

gegevens samen geven een consistent beeld: meisjes werken graag in groepjes, werken beter samen en waarderen hun groepje hoger dan jongens. Een opvallende afwijking van het patroon dat jongens hoger scoren dan meisjes, betreft H-sexe. Meisjes scoren hier hoger op dan jongens. Zij vinden wiskunde voor meisjes vaker belangrijk dan jongens. Dit kan van invloed zijn op de houding van jongens ten opzichte van de inbreng van meisjes bij wiskunde.

Ter nuancering van de, overigens zeer consistente, verschillen tussen jongens en meisjes, is een vergelijking tussen de twee kolommen in tabel 6.6 verhelderend. Op de 27 regels zijn er 16 overeenkomsten in dezelfde richting in de twee kolommen. (Beginkennis en sexe in tabel 6.6). De vraag blijft dus of en in hoeverre de lagere beginkennis van meisjes bepalend is voor de verschillen. Omdat onder de meisjes relatief veel leerlingen met lage scores zitten kunnen verschillen tussen jongens en meisjes op bijvoorbeeld houdingen en percepties evengoed terug gaan op voorkennis als op sexe(-specifieke socialisatie). De vraag blijft uiteraard of en in hoeverre voorkennis het resultaat is van sexe specifieke socialisatie processen. Die socialisatie vindt overigens voor een belangrijk deel op school plaats. Verder zijn er aanwijzingen dat het SLO programma meisjes relatief weinig aanspreekt en dat ook de leraren minder rekening houden met meisjes dan met jongens. Let bijvoorbeeld op IN, H-context 1 en 2 en Houding tegenover leraar en les. Er zijn dus nog aanknopingspunten genoeg om via manipuleerbare, curriculaire variabelen de prestaties van meisjes te verbeteren. (Zie voor een reeks van verbeteringssuggesties Meeder e.a. 1984).

Tenslotte kan men de kwantitatieve gegevens over individuele verschillen (voorkennis en sexe) nog vergelijken met de kwalitatieve gegevens uit hoofdstuk 2 en gegevens uit vragenlijst en gesprekken met leraren. Uit deze observatiegegevens blijkt dat weinig rekening wordt gehouden met deze verschillen. Geen van de beide leraren treft speciale didactische maatregelen voor leerlingen die achterblijven. Deze leraren vormen geen uitzondering. Uit een schriftelijke vragenlijst en uit gesprekken met leraren blijkt dat diagnostische toetsing en het geven van extra hulp aan leerlingen die de doelen niet bereikt hebben niet of nauwelijks vòdrkomen. Dat is niet verbazingwekkend. De SLO heeft dit aspect in het materiaal voor leerlingen en leraren niet uitgewerkt. De leraren staan met lege handen waar het gaat om differentiatie-maatregelen voor specifieke groepen. Het enige differentiatie-principe is het uitlokken en benutten van verschillende oplossingswegen en verschillende oplossingen. Dit belangrijke principe van Freudenthal, namelijk de niveaus in het leerproces in de heterogene groep, is blijkaar niet voldoende. Uit de



kwalitatieve observaties blijkt bovendien dat één van de twee leraren er nauwelijks in slaagt de niveau-gedachte te realiseren. Het idee van de niveaus in het leerproces is een waardevol maar kwetsbaar principe. Nu kan men terecht tegenwerpen: "Alles van waarde is kwetsbaar". Dan nog lijkt het niet verstandig om alléén op dit principe de differentiatie in de klas te baseren. Er zijn onzes inziens daarnaast meer robuuste maatregelen nodig. Dat wil zeggen vooraf geplande maatregelen die niet zo'n zware wissel trekken op het vakmanschap van de leraar en op het momentane, spontane interactie-proces tussen leerlingen en leraren. Men moet óók meer systematisch, door middel van regelmatige diagnostische toetsen bij elke individuele leerling en door gerichte remediërende of aanvullende maatregelen op verschillen tussen leerlingen inspelen.

Het SLO-curriculum differentieert te weinig. Het kan in de praktijk gemakkelijk resulteren in een diffuse onderwijsleersituatie waarin niet wordt ingegaan op verschillen tussen leerlingen. Ondertussen gaat het selectieproces gewoon door. Want alle scholen geven wél selectieve proefwerken met cijfers die later voor determinatie en selectie van leerlingen worden gebruikt. Dat leerlingen met geringe (voor)kennis en vooral meisjes dan gedupeerd kunnen worden blijkt zonneklaar uit de gegevens die we in dit rapport presenteren.

### **6.5 Verschillen in effecten tussen projectklassen (Hypothese A)**

Hypothese A betreft de relatie tussen verschillen in implementatie van het 'experimentele' programma van de SLO en de leereffecten van de leerlingen. Deze verschillen in implementatie hebben we gemeten door middel van de Percia-vragenlijst. Zoals blijkt uit de gegevens in hoofdstuk 3, bestaan er significante verschillen tussen klassen. Een belangrijke vraag is nu of deze verschillen een relatie vertonen met leerresultaten. In dit hoofdstuk (6) is gebleken dat er vrij hoge, significante correlaties bestaan tussen de gemiddelde percepties van leerlingen van het onderwijsleerproces en de leerresultaten op de natoets (zie bijvoorbeeld tabel 6.3).

Er is echter ook een correlatie tussen voortoets en gemiddelde leerlingpercepties. Dat roept de vraag op naar een evt. causale richting van de verbanden. Zijn de verschillen in onderwijsleerprocessen (zoals gemeten via gemiddelde percepties van leerlingen) de oorzaak van verschillen op de natoets? Of zijn die verschillen op de natoets volledig veroorzaakt door verschillen op de voortoets? Determineert de voortoets de onderwijsleerprocessen in de klas en de leereffecten? Draagt de kwaliteit van het onderwijsleerproces als een zelfstandige factor bij aan de prestaties op de natoets?

Door middel van regressie-analyse hebben we geprobeerd een antwoord



op deze vragen te geven. Daartoe hanteren we de gemiddelde score op de totale PERCIA-schaal als indicator voor de mate waarin het onderwijsleerproces volgens de bedoelingen is verlopen. Alle leerlingen in een bepaalde klas kregen dezelfde gemiddelde PERCIA-score van die klas. Met andere woorden: de gemiddelde PERCIA-score hanteren we in de analyse als "treatment" voor alle leerlingen in een bepaalde klas. Om te kunnen zien of verschillen in kwaliteit van onderwijsleerprocessen (PERCIA-gemiddelde) iets toevoegen aan de variantie die reeds door verschillen in voortoets wordt verklaard, wordt de voortoetsscore van de leerlingen als eerste stap ingevoerd en de PERCIA-score als tweede stap. De resultaten van die regressieanalyse zijn samengevat in tabel 6.7. Het betreft hier een regressieanalyse met natoets als afhankelijke variabele.

*Tabel 6.7: Regressie van de natoets op voortoets en PERCIA (N = 482)\**

stap	variabele	R <sup>2</sup> tot	F1	(df)	R <sup>2</sup> ch	F2	(df)	Beta
1	toets 1	.60967	749.7	** (1;480)				.75
2	mpercia	.51849	388.3	** (2;479)	.00883	11.08	** (1;479)	.10

\* Betekenis:

R<sup>2</sup>tot: tot en met de betreffende stap verklaarde varianties

F1 : F-ratio behorende bij R<sup>2</sup>tot

R<sup>2</sup>ch :in de betreffende stap toegevoegde verklaarde variantie (R<sup>2</sup> change)

F2 : F-ratio behorende bij R<sup>2</sup>ch

\*\* : sign. op 1%

Beta : Beta-coëfficiënt

Uit tabel 6.7 blijkt dat de gemiddelde perceptie van een klas (toegekend aan elke individuele leerling van die klas) significant bijdraagt aan de verklaarde variantie op de natoets (de interactie tussen voortoets en gemiddelde percia-score is te verwaarlozen). Dit gegeven is een ondersteuning van hypothese A. Al is de toename van verklaarde variantie als gevolg van de 'treatment', zoals geoperationaliseerd door de gemiddelde percia score per klas, gering. Wat leerlingen aan voorkennis bezitten (toets 1) draagt veel meer bij aan de variantie in de natoets dan de 'treatment'. We gaan nu verder met een verkenning van de relaties tussen conditie en leereffecten (hypothese B).

## 6.6 Verschillen in leereffecten tussen condities (Hypothese B)

In hypothese B is de verwachting uitgesproken dat de leerresultaten van leerlingen die het SLO-programma hebben doorlopen beter zijn dan de leerresultaten van leerlingen die een meer traditioneel programma hebben gevolgd. In hoofdstuk 5 werd al op basis van uitsluitend de toetsresultaten geconcludeerd dat hypothese B is bevestigd.

Deze (voorlopige) conclusie uit hoofdstuk 5 verdient nadere nuancering. De conclusie uit hoofdstuk 5 was voor een deel gebaseerd op verschillen in gemiddelde leerwinst (verschilscore voortoets en natoets) tussen beide groepen. Omdat het werken met verschilcores problematisch kan zijn, lijkt het nuttig ook langs andere weg te zoeken naar gegevens ter toetsing van de hypothese.

Een voor de hand liggende analysetechniek is covariantie-analyse, waarbij de voortoets als covariaat in de analyse wordt meegenomen. Uit tabel 6.8 blijkt echter, dat er tussen voortoets en conditie een weliswaar kleine, maar significante interactie bestaat, wat betekent dat de regressielijnen in beide condities niet parallel lopen.

Tabel 6.8: Regressie van de natoets op voortoets, conditie en interactie daartussen (N = 763)\*

stap	variabele	R <sup>2</sup> tot	F1	(df)	R <sup>2</sup> ch	F2	(df)	Beta
1	toets 1	.531	862 **	(1;761)				.69
2	conditie	.548	460 **	(2;760)	.01670	28.07 **	(1;760)	-.10
3	interactie	.552	312 **	(3;759)	.00450	7.63 **	(1;759)	.22

\* voor de betekenis van R<sup>2</sup>tot, F1, R<sup>2</sup>ch en F2 zie het onderschrift bij tabel 6.7.

In principe vervalt daarmee de covariantie-analyse als geschikte techniek.

In plaats daarvan hebben we voor beide condities aparte regressievergelijkingen opgesteld:

projectconditie:  $\text{Natoets} = 4.90 + 1.21 \text{ Toets 1}$

controleconditie:  $\text{Natoets} = 7.80 + 1.00 \text{ Toets 2}$

In figuur 6.4 zijn deze lijnen in beeld gebracht.

$$Y = 4.90 + 1.21 X$$

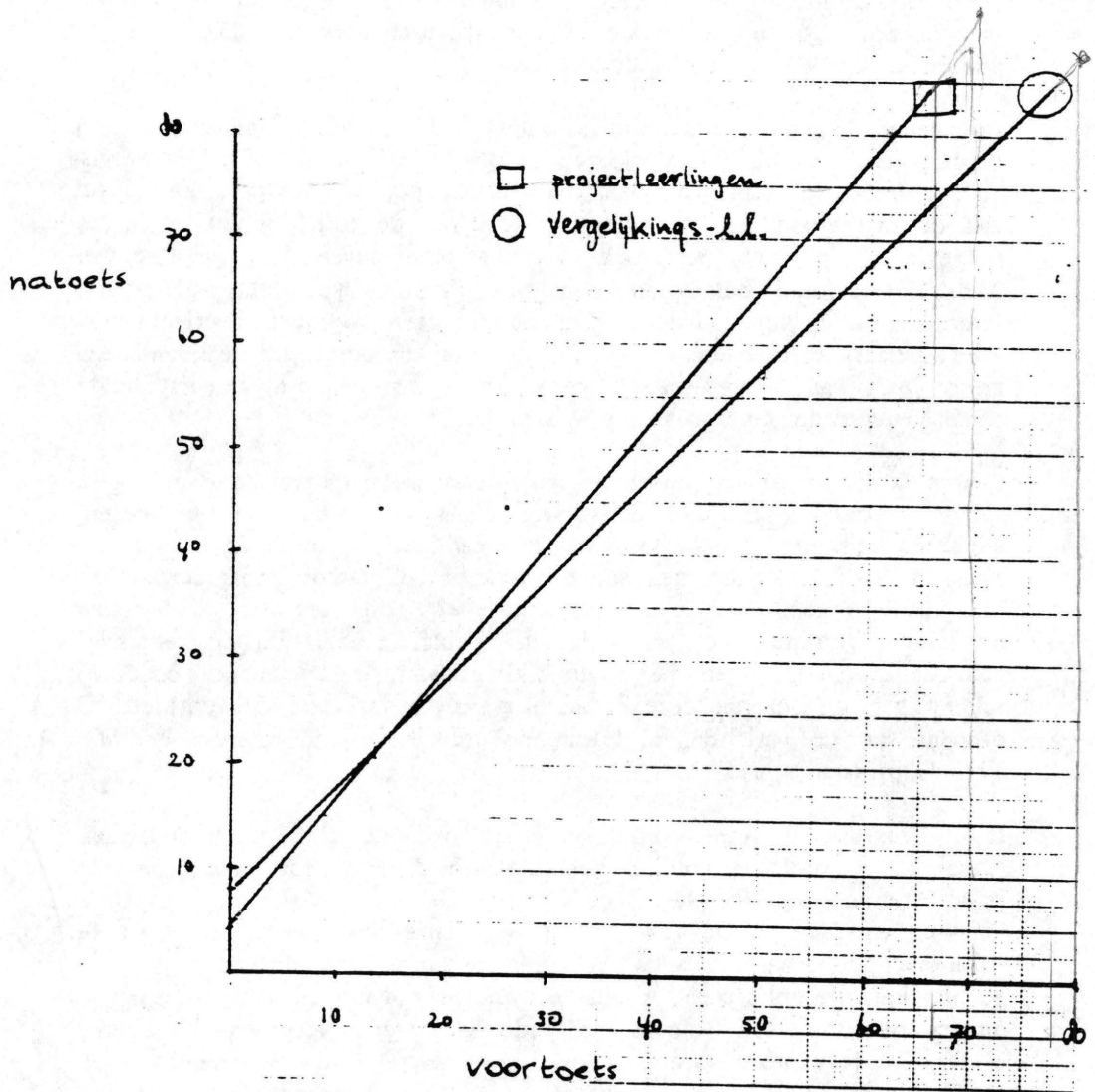
$$Y = 7.80 + 1.00 X$$

is dan niet geschikt?

? Regressieanalyse

zoek op Pedhazur

Figuur 6.4: Grafische voorstelling regressielijnen projectleerlingen en vergelijkingsleerlingen



Met behulp van de Johnson-Neyman techniek hebben we zogenaamde "areas of significance" bepaald, gebieden waar de natoetsscores in beide condities significant van elkaar verschillen. Een significant verschil in natoetsscore blijkt op te treden bij een voortoetsscore van 23.6 of hoger (zie Pedhazur 1982, p. 469 voor de formules).

Dat betekent dus dat het voor leerlingen met een voortoetsscore tot 23.6 punten geen verschil maakt of men in de projectconditie of in de vergelijkingsconditie is geplaatst. Daarboven maakt het wél verschil, ten gunste van de projectconditie. Het grootste deel van de totale groep leerlingen (ongeveer 80 procent) heeft een voortoetsscore boven de grenscore van 23.6. Dat is een belangrijke conclusie die de hypothese ondersteunt. Interessant is echter dat de project-conditie geen voordeel (overigens ook geen nadeel) biedt voor de groep van zwakke leerlingen beneden deze grens (29% van de projectleerlingen). Dit is een enigszins teleurstellende conclusie tegen de achtergrond van 'Wiskunde voor Iedereen'.

Omdat 80% van de leerlingen een voortoetsscore heeft van 23.6 of hoger, valt te overwegen om tóch een gewone covariantie-analyse uit te voeren. De interactie term uit tabel 6.8 was wel significant, maar bijzonder laag.

Toets 1 is daarbij als covariaat en conditie als factor gehanteerd. Ook deze analyse geeft een significant hoofdeffect voor conditie na controle op initiële verschillen op de voortoets. Echter evenals bij de regressie-analyses in tabel 6.7 en 6.8 is de bijdrage van de 'treatment' (conditie) relatief gering. Ook bij deze covariantie-analyse verklaart de voortoets 53 procent van de variantie in de natoets terwijl de bijdrage van conditie slechts 2 procent is.

In de regressie-analyse waarin het effect van de conditie is nagegaan (tabel 6.8), fungeert conditie als 'black box'. Hetzelfde geldt voor de covariantie-analyse. Hoe het proces in de klas is verlopen blijft in deze analyses een vraag. Daarom voeren we eenzelfde analyse uit zoals in paragraaf 6.5, maar nu voor het hele bestand. Ook nu vatten we de gemiddelde perceptiescore op als een indicatie voor de mate waarin de onderwijsleerprocessen volgens de bedoelingen zijn verlopen. Tabel 6.9 geeft de resultaten van een regressie-analyse van de natoets, op respectievelijk voortoets en 'toegeschreven' PERCIA-gemiddelde (voor alle leerlingen in de twee condities).

Tabel 6.9: Regressie van natoets op voortoets en gemiddelde PERCIA (N = 763)\*

stap	variabele	R <sup>2</sup> tot	F1	(df)	R <sup>2</sup> ch	F2	(df)	Beta	
1	toets 1	.53133	862	**	(1;761)			.73	
2	mpercia	.54471	455	**	(2;760)	.01359	23 **	(1;760)	.12

\* voor de betekenis van R<sup>2</sup>tot, F1, R<sup>2</sup>ch en F2 zie het onderschrift bij tabel 6.7.

Ook deze tabel laat zien dat de onderwijsleerprocessen in de klas (zoals gezien door de ogen van de leerlingen) een significante bijdrage leveren aan de variantie op de natoets. Uit hoofdstuk 3 bleek dat er tussen de twee condities significante verschillen zijn in de kwaliteit van onderwijsleerprocessen zoals gemeten via de PERCIA-vragenlijst. In dit hoofdstuk blijkt dat deze verschillen tussen condities inderdaad significant bijdragen aan de variantie op de natoets, ook als men rekening houdt met verschillen in beginsituatie (voortoets). Daarmee is ook hypothese B bevestigd.

## 6.7 Conclusies

De centrale vraag in dit hoofdstuk is: wat zijn de relaties tussen (1) de beginsituatie van de leerlingen (2) het onderwijsleerproces in de klas en (3) de leereffecten? Er is ook een algemene hypothese geformuleerd: hoe meer de onderwijsleerprocessen volgens de bedoelingen zijn gerealiseerd des te beter zijn de leereffecten. Deze algemene hypothese is inhoudelijk toegespitst ten aanzien van leraar- en leerling-activiteiten en de wiskundig - didactische inhoud. We vonden verschillende betekenisvolle, significante verbanden die vrijwel steeds overeenkomen met de algemene hypothese.

Voorts blijkt de beginsituatie van de leerlingen, met name voorkennis, van grote betekenis te zijn voor de onderwijsleerprocessen in de klas en de leerresultaten aan het eind. Echter ook als we, zoals in de regressie-analyses, rekening houden met verschillen in voorkennis is er een significante bijdrage van het onderwijsleerproces aan de variantie in de natoets. Dat geldt voor verschillen in onderwijsleerprocessen binnen de experimentele conditie (tussen projectklassen) maar ook voor verschillen



tussen de twee condities. De onderzoeksresultaten bevestigen onze hypothesen A en B.

Leerlingen met een lage score op de voortoets percipiëren hun onderwijsleersituatie minder positief dan leerlingen met een hoge score op de voortoets. Datzelfde geldt voor hun houding. Een lage of hoge score op de voortoets blijkt in hoge mate bepalend te zijn voor de leereffecten aan het eind. De leerresultaten van zwakke leerlingen zijn relatief gering. Voor zwakke leerlingen met een voortoetsscore lager dan 23.6 punten maakt het geen verschil of ze in de projectconditie dan wel in de vergelijkingsconditie zijn geplaatst. (Let wel: deze zwakke leerlingen zijn een deel van de L-leerlingen, de grensscore voor de L-leerlingen ligt bij 27 punten op de voortoets, zie hoofdstuk 5).

Meisjes presteren duidelijk minder dan jongens in termen van voortoets en natoets. Meisjes scoren ook lager op de perceptieschaal betreffende de begeleiding door de leraar. De wiskundig didactische inhoud spreekt hen minder aan. Ook is de houding van de meisjes minder positief. Een belangrijk verschil tussen jongens en meisjes betreft hun houding tegenover hun eigen niveau bij wiskunde. Meisjes vinden vaker dan jongens dat ze de wiskunde niet begrijpen, dat ze niet zo goed zijn in wiskunde enz. en dat correspondeert met hun toetsresultaten. Er moet dus nog veel verbeteren voor er sprake is van wiskunde voor iedereen.

## 7 SAMENVATTING EN EINDCONCLUSIES

### 7.1 Opzet en vraagstelling

Dit onderzoek is gericht op beantwoording van de vraag naar de implementatie en het effect van onderwijs in heterogene groepen in de eerste fase van het voortgezet onderwijs bij wiskunde.

In het onderzoek zijn drie projectscholen, twee brede scholengemeenschappen en één middenschool, betrokken die werken met een nieuw programma voor intern gedifferentieerd onderwijs. Daarnaast zijn gegevens verzameld over het onderwijs in twee vergelijkingsscholen (mavo, havo, vwo scholengemeenschappen) die op een meer traditionele wijze werken. In projectscholen werd het nieuwe programma uitgevoerd in 22 klassen. Er waren ruim 500 leerlingen bij betrokken. In de vergelijkingsscholen deden 11 klassen en ongeveer 300 leerlingen mee aan het onderzoek. In totaal waren dus 5 scholen, 33 klassen en ruim 800 leerlingen in het onderzoek betrokken. Het onderzoek vond plaats in de tweede klas over de periode van oktober 1985 tot en met april 1986.

Het nieuwe programma is ontwikkeld door de Stichting voor de Leerplanontwikkeling (SLO) te Enschede. De leerplanontwikkelaars zijn daarbij uitgegaan van enkele basisprincipes zoals: klassikale instructies, samenwerken in heterogene groepen, gebruik maken van rijke contexten en niveaus in het leerproces. Zij hebben hun ideeën en materialen aangevoerd onder de vlag van 'Wiskunde voor Iedereen'. De taak van de onderzoekers is ondermeer na te gaan of die vlag de lading dekt. Naast die beschrijvende en evaluatieve functie beoogt het onderzoek een bijdrage te leveren aan theorieontwikkeling omtrent vraagstukken van heterogeniteit, interne differentiatie, didaxologie en curriculum. Het onderzoek is uitgegaan van vier hoofdvragen en een algemene hypothese.

1. Wat is de beginsituatie van de leerlingen? (houding en voorkennis)
2. Worden de onderwijsleerprocessen volgens de bedoelingen gerealiseerd?
3. Zijn de beoogde leereffecten opgetreden?
4. Wat zijn de relaties tussen (1) beginsituatie, (2) onderwijsleerprocessen en (3) leereffecten?

De algemene hypothese die betrekking heeft op de vierde vraag is: Hoe meer de onderwijsleerprocessen volgens de bedoelingen zijn gerealiseerd, des te beter zijn de leereffecten. Deze algemene hypothese is toegespitst in twee specifieke hypotheses:

- A. In klassen waarin de leraar een banende, taakstellende, (bege)leidende rol vervult en waarin de leerlingen taakgericht samenwerken, aan de oplossing van opgaven in contexten, zijn de leerresultaten beter dan in klassen waarin dat minder (vaak) het geval is.
- B. In onderwijs waarin met het nieuwe wiskundeprogramma van de SLO wordt gewerkt verlopen de onderwijsleerprocessen beter en zijn de leerresultaten beter dan in onderwijs met een meer traditionele aanpak.

Hypothese A heeft betrekking op de 22 project klassen in de drie projectscholen. Hypothese B betreft een vergelijking tussen de project-situatie en de vergelijkingssituatie.

In het onderzoek is een breed scala van methoden en instrumenten gehanteerd: toetsen, observatiemethoden en vragenlijsten voor het meten van de houding en de percepties van de leerlingen. Deze methoden en instrumenten zijn voor een deel in eerdere fasen van het onderzoek ontwikkeld. Van alle instrumenten behalve van de observatiemethoden, is de betrouwbaarheid vastgesteld. De betrouwbaarheid van alle instrumenten bleek redelijk te zijn. Van de observaties is door vergelijking met andere instrumenten een indicatie verkregen over de betrouwbaarheid. De validiteit van de instrumenten is zo veel mogelijk gewaarborgd door zorgvuldigheid bij de constructie en door regelmatig overleg met de leerplanontwikkelaars, onder meer bij doelstellingen-discussies, toets- en vragenlijstconstructies. De eindconclusies van het onderzoek hebben we gerangschikt aan de hand van de vier hoofdvragen en de hypothesen.

## 7.2 Wat is de beginsituatie?

1. Tussen leerlingen, klassen en scholen bestaan grote verschillen in voorkennis zoals gemeten met de voortoets.
2. Ook tussen projectscholen en vergelijkingsscholen onderling bestaan aanzienlijke verschillen in voorkennis bij de leerlingen. De gemiddelde voortoetsscore bij alle leerlingen is 32 punten. Projectleerlingen behalen gemiddeld 29 punten en vergelijkingleerlingen 36 punten. De projectscholen hebben dus vóór de uitvoering van het curriculum relatief veel leerlingen met een lage score op de voortoets. Dit geldt in het bijzonder voor projectschool 1.
3. Als men een indeling maakt van de leerlingen in drie niveaugroepen Laag, Midden en Hoog (L, M, H) op basis van hun score op de voortoets, dan blijkt dat de leerlingen niet evenredig verdeeld zijn over scholen en klassen. Tussen scholen en tussen klassen (binnen

scholen) bestaan aanzienlijke verschillen in de proportionele verdeling van L-, M-, en H-leerlingen. Alleen projectschool 2, een middenschool, heeft een evenwichtige verdeling van L-, M- en H-leerlingen over alle klassen. Alle andere scholen hebben klassen met veel of weinig zwakke (L) leerlingen. Meestal is dit een gevolg van het beleid van de scholen of van omstandigheden waarop men geen invloed kan of wil uitoefenen, bijvoorbeeld instroom van veel zwakke leerlingen bij projectschool 1.

4. Tussen projectleerlingen en vergelijkingsleerlingen bestaan aan het begin géén verschillen van betekenis in houding ten aanzien van de school en verschillende aspecten van het wiskunde onderwijs. Wel zijn er verschillen tussen scholen. Van de projectleerlingen hebben de leerlingen op school 2 (middenschool) een positievere houding in vergelijking met de leerlingen van de andere projectscholen.

### 7.3 Worden de onderwijsleerprocessen gerealiseerd?

Omtrent het onderwijsleerproces in de klas zoals gezien door de ogen van de leerlingen zijn de volgende conclusies te trekken.

1. Er zijn verschillen tussen klassen, scholen en condities in de mate van realisering van de beoogde onderwijsleerprocessen.
2. In de projectscholen worden de processen in de klas meer volgens de bedoelingen gerealiseerd dan in vergelijkingscholen. Als men de kenmerken van het SLO-programma als dimensies voor een vergelijkende beoordeling neemt, dan verlopen processen in projectklassen positiever dan in vergelijkingsklassen. De verschillen zijn het opvallendst op twee punten: samenwerking en leerstofinhoud (wiskunde in contexten). Projectleerlingen werken vaker en beter samen dan vergelijkingsleerlingen. De wiskunde wordt vaker in een dagelijkse context geplaatst en dat wordt door de leerlingen als aantrekkelijk en verhelderend ervaren.
3. Er doen zich enkele hoge en stabiele correlaties voor, tussen leraars- en leerlingenactiviteiten.  
In klassen waarin de leraar goed leiding geeft en duidelijk uitlegt blijven de leerlingen beter bij de taak dan in klassen met een minder goede leiding en instructie door de leraar. En omgekeerd.  
In klassen waarin de leerlingen goed samenwerken is de taakgerichtheid beter dan in klassen waarin de leerlingen minder adequaat samenwerken. En omgekeerd.

4. Tussen projectscholen onderling en tussen klassen binnen projectscholen komen verschillen voor in de realisatie van het nieuwe programma. School 2 scoort relatief hoog op vrijwel alle aspecten. Voor school 1 geldt het omgekeerde.
5. Uit observaties blijkt dat de ene leraar het differentiatieprincipe "niveaus in het leerproces" niet realiseert en de andere wel.
6. Voorts blijkt dat andere differentiatiemaatregelen zoals diagnostische toetsingen en aanvullende didactische maatregelen voor zwakke en sterke leerlingen niet of nauwelijks voorkomen. Dit geldt niet alleen voor vergelijkingscholen maar ook voor projectscholen.
7. Tussen projectscholen en tussen projectklassen bestaan zeer grote verschillen in de tijdsbesteding. School 1 besteedt 72% van de tijd aan groepswerk. Voor school 2 en 3 zijn die percentages respectievelijk 57 en 38. Per klas variëren deze percentages aanzienlijk. In de ene klas komt groepswerk nauwelijks voor terwijl in de andere klas 96% van de tijd aan groepswerk wordt besteed.
8. Ook zijn er grote verschillen tussen projectscholen en tussen projectklassen in taakgerichtheid (time-on task). Dit blijkt uit percepties van leerlingen en uit kwantitatieve observaties in alle 22 projectklassen. Tussen deze perceptiescores en observaties bestaat een vrij sterk verband. Op school 1, die relatief veel zwakke leerlingen heeft, wordt weinig taakgericht gewerkt. Gedurende 50 procent van de lestijd wordt niet taakgericht gewerkt. Voor school 2 en 3 zijn die getallen respectievelijk 5 en 15 procent van de lestijd. Een grote proportie groepswerk in verhouding tot klassikaal werken blijkt onder bepaalde condities niet samen te gaan met een hoge taakgerichtheid.
9. Het nieuwe programma 'Wiskunde voor Iedereen', is geen papieren vernieuwing gebleven. De onderwijsleerprocessen zijn anders en beter (in termen van door leerplanontwikkelaars beoogde processen) in projectklassen dan in vergelijkingsklassen waarin met een meer "traditionele" aanpak wordt gewerkt. Toch is er nog wel veel variatie in de uitvoering van het nieuwe curriculum. De vraag is of al deze varianten wel passen binnen de bedoelingen van de leerplanontwikkelaars. Omdat eenduidige, absolute criteria van de zijde van de SLO ontbreken, kunnen we als onderzoekers ook geen absoluut antwoord op die vraag geven. Wel hebben we geconstateerd dat in bepaalde scholen en bepaalde klassen de kenmerken van de SLO méér volgens de bedoelingen zijn gerealiseerd dan in andere scholen en klassen.



#### 7.4 Zijn de beoogde leereffecten opgetreden?

In het onderzoek zijn twee soorten effecten bepaald:

(a) de houding van de leerlingen ten aanzien van de school en verschillende aspecten van de onderwijsleersituatie en

(b) de resultaten op de natoets. Deze effecten kunnen uiteraard niet los worden gezien van de beginsituatie. We vergelijken daarom de houding (voormeting) en de toetsresultaten (voortoets) aan het begin met de houding (nameting) en de toetsresultaten (natoets) aan het eind. Tevens vergelijken we de resultaten van projectleerlingen met die van vergelijkingsleerlingen.

##### *Houding*

In de beginsituatie bestaan er geen verschillen van betekenis in de houding van projectleerlingen en vergelijkingsleerlingen. Wel zijn er verschillen tussen scholen. Daarbij scoort school 2 relatief hoog.

In de eindsituatie is de houding van de projectleerlingen positiever ten aanzien van drie aspecten: contexten, de leraar en de groep. Zowel aan het begin als aan het einde blijken de leerlingen van school 2 - relatief gezien - een positieve houding te bezitten.

##### *Toetsresultaten*

De totale groep leerlingen gaat van de voortoets naar de natoets zo'n 10 punten vooruit, dat wil zeggen van een gemiddelde van 32 punten naar een gemiddelde van 42 punten. De projectleerlingen gaan ruim 11 punten vooruit: van 29 punten op de voortoets naar 40 punten op de natoets. De vergelijkingsleerlingen gaan bijna 8 punten vooruit, van ruim 36 punten op de voortoets naar bijna 44 punten op de natoets. Het verschil in leerwinst tussen projectleerlingen is statistisch significant en van praktische betekenis. Projectleerlingen boeken dus meer leerwinst dan vergelijkingsleerlingen.

Als we kijken naar de leerlingen met lage, middelmatige en hoge scores (L, M, H) op de voortoets dan ligt de leerwinst van de M-leerlingen het dichtst bij het gemiddelde voor de hele groep, namelijk 10 punten. De leerlingen met een lage voortoetsscore (L) behalen 8,5 punt leerwinst. Terwijl de leerlingen met een hoge voortoetsscore 11 punten vooruit gaan. Er is dus een zwakke trend dat sterke leerlingen meer vooruit gaan dan zwakke leerlingen.

De tendens dat projectleerlingen in het algemeen meer leerwinst boeken, is ook aanwezig binnen de groepen L, M en H leerlingen. Die tendens is het sterkst bij de leerlingen met een hoge score op de

voortoets. De H-leerlingen van de projectscholen gaan 14 punten vooruit, terwijl de H-leerlingen van de vergelijkingscholen slechts 8 punten leerwinst boeken.

Bij gebrek aan absolute criteria kunnen we geen definitief antwoord geven op de vraag of de beoogde effecten zijn bereikt. Wel kan men zeggen dat de projectleerlingen na het doorlopen van het nieuwe programma een hoger niveau hebben bereikt dan daarvoor.

De leerresultaten van de projectleerlingen zijn beter dan die van de vergelijkingsleerlingen:

- De projectleerlingen boeken meer leerwinst dan de vergelijkingsleerlingen. Er is sprake van een inhaaleffect van projectleerlingen ten opzichte van de vergelijkingsleerlingen.
- Dit inhaaleffect reduceert de gemiddelde voorsprong van de vergelijkingsleerlingen, zoals die bestond bij de voortoets, met ruim 3 punten. Voor de totale groep, dat wil zeggen niet uitgesplitst naar L,M,H, leerlingen blijft er echter een significant verschil op de natoets ten gunste van de vergelijkingsleerlingen. Met andere woorden de voorsprong van de vergelijkingsleerlingen in de beginsituatie wordt in de eindsituatie niet volledig ingelopen door de projectleerlingen.
- Als men kijkt naar de resultaten op de voortoets en de natoets per niveau-categorie van de leerlingen (L, M, H) dan blijkt echter dat op de *voortoets* significante verschillen bestaan die steeds in het voordeel van de vergelijkingsleerlingen uitvallen. Op de natoets zijn deze verschillen verdwenen. Er bestaat géén significant verschil meer tussen projectleerlingen en vergelijkingsleerlingen voor de categorieën L en H leerlingen en voor de M-leerlingen is de achterstand op de voortoets omgezet in een voorsprong bij de natoets in het voordeel van de projectleerlingen. Dus als gevolg van de grotere leerwinst van de projectleerlingen, worden de aanvangsverschillen ongedaan gemaakt of zelfs in een voorsprong omgezet, als men kijkt binnen de categorieën Laag, Midden en Hoog.
- Van de drie scholen in de project-conditie boekt school 2 (de middenschool) relatief veel leerwinst. School 1 boekt weinig leerwinst; deze school heeft veel 'zwakke' leerlingen en relatief veel allochtone leerlingen.
- De houding van de projectleerlingen is in de eindsituatie, dus na het doorlopen van het nieuwe programma, positiever dan die van de vergelijkingsleerlingen.

De vergelijking tussen de projectscholen en de vergelijkingsscholen valt ten gunste van de projectscholen uit.

*Hypothese B* is bevestigd: In onderwijs waarin met het nieuwe wiskunde-programma van de SLO wordt gewerkt, verlopen de onderwijsleerprocessen beter en zijn de leerresultaten beter dan in onderwijs met een meer traditionele aanpak.

### 7.5 Wat zijn de relaties tussen beginsituatie, processen en effecten?

In hoofdstuk 6 staat deze vraag centraal. In deze samenvatting volstaan we met enkele hoofdconclusies.

*Hypothese A is:* In klassen waarin de leraar een banende, taakstellende, (be)geleidende rol vervult en waarin de leerlingen taakgericht samenwerken aan opgaven in contexten zijn de leerresultaten beter dan in klassen waarin dat minder (vaak) het geval is.

We vonden verschillende betekenisvolle, significante verbanden die vrijwel steeds overeenkomen met hypothese A. Dat is een bevestiging van hypothese A. Uit regressie-analyses blijkt dat de verschillen op de voortoets een belangrijk deel van de variantie op de natoets verklaren. Na 'correctie' voor verschillen in voortoets blijkt echter dat de onderwijsleerprocessen waarop wordt gedomd in hypothese A, een significante bijdrage leveren aan de variantie op de natoets. Ook in die zin is hypothese A bevestigd.

Nadere regressie-analyses met betrekking tot hypothese B maken een nuancering van de conclusie mogelijk. Er is een significant effect van de verschillende condities op natoetsresultaten, ook na 'correctie' voor verschillen in voortoets. Er is echter ook een significant interactie-effect van voortoets en conditie. Projectleerlingen met een middelmatige of hoge score op de voortoets profiteren meer van het experimentele programma dan leerlingen met een lage voortoetscore. Voor zwakke leerlingen met een voortoetscore beneden 23.6 punten is er géén verschil tussen beide condities.

Leerlingen met een lage score op de voortoets percipiëren hun onderwijsleersituatie minder positief dan leerlingen met een hoge score op de voortoets. Datzelfde geldt voor hun houding. Een lage of hoge score op de voortoets blijkt in hoge mate bepalend te zijn voor de leereffecten aan het eind.

Meisjes presteren duidelijk minder dan jongens in termen van voortoets en natoets. Meisjes percipiëren minder begeleiding van de leraar. De wiskundig didactische inhoud spreekt hen minder aan. Ook is de houding van de meisjes minder positief dan die van de jongens.

De conclusie met betrekking tot zwakke leerlingen en meisjes gelden voor beide condities. Het verdient aanbeveling om bij de vernieuwing van het wiskundeonderwijs speciale aandacht aan deze categorieën leerlingen te geven.

Voor een discussie over de resultaten van het onderzoek gaan we nu over naar hoofdstuk 8. Daarin komen de beperkingen van het onderzoek aan de orde.

## 8 DISCUSSIE EN AANBEVELINGEN

In dit hoofdstuk wordt ingegaan op de betekenis van de onderzoeksgegevens tegen de achtergrond van het gehanteerde design en de gekozen analyse technieken. Het design zoals dat uiteindelijk is gerealiseerd is een compromis tussen wat wij als onderzoekers wenselijk achtten en wat praktisch haalbaar bleek. Dit compromis leverde verschillende beperkingen op:

1) We moesten werken met bestaande groepen leerlingen, klassen, scholen, en onder condities die in aanvang (grote) verschillen vertoonden op relevante variabelen bijvoorbeeld in beginkennis bij wiskunde. Daarmee ontstond het probleem van (on)vergelijkbaarheid.

Is het wel mogelijk conclusies te trekken over het effect van het SLO programma (in vergelijking met dat van een meer traditionele aanpak) als er op voorhand verschillen bestaan tussen leerlingen, klassen en scholen? We hebben er in dit rapport voor gekozen de gegevens van voortoetsen, natoetsen, leerwinst uitgebreid te beschrijven. Ook hebben we gekeken naar verschillende subgroepen leerlingen zoals leerlingen met een lage of een hoge beginscore. Dan blijkt dat de projectleerlingen een lagere beginscore hebben dan de vergelijkingsleerlingen, maar dat deze achterstand voor een belangrijk deel is ingehaald bij de eindtoets. Er is dus een inhaaleffect, dat wil zeggen een grotere leerwinst bij de projectleerlingen in vergelijking met de leerlingen die een meer traditioneel programma hebben gevolgd.

Het hanteren van leerwinstgegevens (verschilscore tussen voortoets en natoets) brengt echter allerlei problemen mee. In hoofdstuk 5 hebben we drie hoofdpijnen genoemd: regressie naar het gemiddelde, schaalproblemen en het vraagstuk van de betrouwbaarheid. Dit betekent dat men uiterst voorzichtig met verschilcores moet omgaan. Sommige auteurs adviseren zelfs om helemaal niet met verschilcores te werken. Wij hebben wél met verschilcores gewerkt, maar onze conclusies zijn niet uitsluitend daarop gebaseerd. Behalve door het gebruik van verschilcores hebben we door multiple regressie-analyses gecorrigeerd voor verschillen in voorkennis (voortoets) tussen leerlingen. Hoewel ook deze analyses niet alle problemen zoals we die bij verschilcores tegenkomen ondervangen, geven ze door de andere invalshoek toch extra informatie. Tevens hebben we gekeken naar mogelijke interactie-effecten tussen voortoets en conditie. De effecten op de natoets van (a) voortoets (b) conditie en (c) de interactie tussen a en b zijn in kaart gebracht. Ook uit die analyse blijkt een significant effect van de projectconditie. Met andere woorden het nieuwe programma blijkt effectief ook wanneer we rekening houden met verschillen in voorkennis tussen projectleerlingen en vergelijkingsleer-



lingen.

2) Een tweede punt van onvergelijkbaarheid betreft het verschil in inhouden en doelstellingen van het wiskunde-onderwijs in de twee condities. De toetsen zijn een operationalisering van de doelen van het nieuwe SLO-programma. Daarmee zijn de projectleerlingen in het voordeel. Zij zijn getoetst op doelen die in de projectklassen centraal hebben gestaan. Ze waren vertrouwd met opgaven die in de toetsen zijn opgenomen. De uitslag zou mogelijk anders zijn geweest als een toets met traditionele opgaven zou zijn voorgelegd. Mogelijk zouden dan zelfs de vergelijkingsleerlingen meer leerwinst hebben behaald. Overigens is er een grote mate van overeenkomst in wiskundige inhouden, zoals lineaire verbanden, tussen beide programma's (zie hoofdstuk 1). Het zou interessant zijn geweest als traditionele opgaven naast SLO gegevens in de toets waren opgenomen. Dan zouden de kosten en baten naar twee kanten kunnen worden bepaald.

Ook is er discussie mogelijk over de waarde van de doelen en inhouden van het nieuwe wiskundeprogramma in vergelijking met die van traditionele benaderingen. Zijn de SLO-doelen en inhouden nastrevenswaardiger dan die van traditionele programma's? Dat is geen empirische vraag. We doen daarover in dit rapport geen uitspraken. Wel kan men op grond van gegevens uit ons onderzoek concluderen, dat als men de SLO-doelen belangrijk acht, dan is het SLO-programma een beter middel ter realisering van deze doelen dan een meer traditionele aanpak.

Wat hierboven is gezegd over de verschillen in leereffecten, geldt mutatis mutandis óók voor verschillen in onderwijsleerprocessen, zoals gemeten met de vragenlijst leerlingpercepties. Ook deze vragenlijst is eenzijdig toegespitst op het experimentele SLO-programma.

3) Een discussiepunt is ook de verhouding voortoets en natoets. Er is gewerkt met voortoetsen en natoetsen die qua inhoud en doelstellingen nagenoeg identiek waren. Tussen voortoets en natoets blijken op leerlingen en klassenniveau hoge correlaties te bestaan. De voortoets 'verklaart' dus een groot deel van de variantie op de natoets. Dit impliceert dat er voor andere variabelen weinig variantie te verklaren over blijft, als men de voortoetsgegevens in de analyse betreft.

Eenzijds is dit een belangrijke conclusie uit het onderzoek: wat leerlingen vóóraf aan beginkennis bezitten is véél invloedrijker ten aanzien van hun resultaten op de natoets dan andere variabelen zoals instructie in de klas. Beginkennis blijkt een sterke predictor te zijn van leereffecten aan het eind. Dit is een conclusie die theoretisch te verwachten valt.

Anderzijds dient men de vraag te stellen in hoeverre deze bevinding een

artefact is van de toetsingsprocedure. Zijn de resultaten op de natoets mede beïnvloed door het maken van de voortoets? Hebben de leerlingen iets geleerd door het maken van de voortoets? Zijn ze testwijzer geworden? Kunnen herinneringseffecten een rol hebben gespeeld? Hebben ze op wiskundig gebied hun kennis vermeerderd tijdens het doorlopen van het programma of simpelweg door het maken van de voortoets. Dit soort effecten zouden dan in rekening moeten worden gebracht. Het gehanteerde design maakt een onderscheiding in dergelijke bij-effecten en hoofdeffecten niet mogelijk. Omdat de periode tussen de afname van de voortoets en de afname van de natoets vrij lang is geweest lijkt het effect van 'testing' niet waarschijnlijk.

4) Er zijn nog andere factoren die een rol kunnen spelen. De betere resultaten van de projectleerlingen kunnen, in meer of mindere mate, veroorzaakt zijn door een effect van de voortoets in relatie tot het daarna gevolgde programma. Dit effect staat bekend als een interactie effect tussen voortoets en treatment. Het betreft een effect van 'sensitizing' dat alleen optreedt voor de projectleerlingen.

Het maken van de voortoets in combinatie met het SLO-programma kan de projectleerlingen voordeel hebben opgeleverd. Mogelijk hebben zij door het maken van de voortoets meer geprofiteerd van het daaropvolgende onderwijs dan de vergelijkingsleerlingen. Misschien zijn de projectleerlingen meer bewust geworden van centrale doelstellingen, begrippen en problemen van het onderwijs dat ze daarna hebben ontvangen. Het gaat hier dus om grotere bewustwording en frequentere herkenning als gevolg van de voortoets. De voortoets is echter alleen een onderzoeksinstrument; het maakt geen onderdeel uit van de treatment.

Het verschil in leerwinst tussen projectleerlingen en vergelijkingsleerlingen kan dus (voor een deel) een artefact zijn van de voormeting. Het gekozen design biedt geen mogelijkheid dit eventuele effect te constateren.

5) Er zijn ook factoren die niet onder controle zijn gehouden en die in het voordeel van de vergelijkingsleerlingen kunnen hebben gewerkt. Zo kunnen vergelijkingsleerlingen een groot voordeel hebben gehad van het feit dat zij relatief vaak in klassen hebben gezeten met weinig zwakke leerlingen. Dat levert doorgaans een hogere taakgerichtheid op en maakt dat de leraar méér leerstof kan behandelen. Er kunnen dus allerlei factoren werkzaam zijn geweest die de leereffecten in beide condities hebben beïnvloed zonder dat deze door de onderzoekers zijn gecontroleerd. Er bestaat een positieve correlatie tussen het klasgemiddelde op de voortoets en de gemiddelde leerwinst per klas. Dit is een gegeven dat niet alleen voor de totale groep geldt, maar ook binnen de condities

opgaat. Daarmee is de vergelijkingsgroep in het voordeel.

Om causale uitspraken te kunnen doen over de effecten van een nieuw wiskundeprogramma al of niet in vergelijking met een meer traditionele benadering, zouden deze factoren in een theoretisch model moeten worden opgenomen. Via Padanalyse of LISREL-analyse kunnen causale relaties worden geanalyseerd en getoetst (vgl. Pedhazur, 1982). Op dit moment werken we aan de ontwikkeling van zo'n causaal model en aan de analyse van de gegevens in het licht van dat model. Het gaat om uiterst ingewikkelde vraagstukken. Dwars daar doorheen spelen multi-level problemen. Deze laatste problemen hebben we in dit onderzoek nog niet opgelost. In sommige regressie-analyses hebben we gewerkt met een quasi-multi-level analyse. Er zijn echter meer verfijnde analyses beschikbaar om met multi-level vraagstukken om te gaan (vergelijk Koehler en Van den Eeden, 1987). In dit rapport hebben we ons beperkt tot het exploreren van verbanden langs correlationele weg. Bovendien zijn enkele regressie-analyses uitgevoerd. Hoewel de gevonden correlaties en regressies een ondersteuning vormden voor de hypothese, mogen we daaruit geen causale relaties afleiden, want: "Correlation is no proof of causation". Nader onderzoek is dus gewenst.

6) Er is nog een andere factor die differentiële effecten in de beide condities kan hebben veroorzaakt. Dat is de factor "rijping" in combinatie met selectie. De leerlingen in de vergelijkingsconditie bezitten een hoger niveau van beginkennis bij wiskunde dan de leerlingen in de projectconditie. Daarmee corresponderen de schooladviezen en wellicht ook de intelligentieniveaus. De projectleerlingen hebben adviezen van LBO t/m VWO. De adviezen van de vergelijkingsleerlingen variëren van MAVO t/m VWO. Als men twee groepen selecteert van verschillend niveau dan zou men de hypothese kunnen opwerpen dat alleen al op grond van rijping, dus zonder invloed van welk programma dan ook, te verwachten is dat de groep met het hoge niveau meer vooruit gaat dan de lagere groep. Aanvangsverschillen tussen leerlingen nemen bij die hypothese in de loop der tijd toe (vgl. Cook en Campbell, 1979, blz 104 en 110).

In ons onderzoek nemen de verschillen tussen beide condities af. De projectleerlingen halen de vergelijkingsleerlingen gedeeltelijk in. Dat is het omgekeerde van wat men op grond van rijping zou verwachten. Ook dat duidt op de positieve invloed van het experimentele programma.

Nu zou dit verschijnsel ook verklaard kunnen worden door plafondeffecten bij de toets. Daarvan is echter geen sprake, gelet op de moeilijkheidsgraad van de toets en gezien het feit dat de natoets drie extra (moeilijke) items bevat die eventuele plafondeffecten voorkomen.

Ook een verklaring door plafondeffecten in de cognitieve ontwikkeling bij de sterkere leerlingen lijkt ons niet aannemelijk. Toch is het niet geheel

uitgesloten dat de leerwinst van de projectleerlingen geflatteerd is door het feit dat zij gemiddeld lager op de voortoets scoren. Het is misschien wel gemakkelijker om leerlingen die nog weinig afweten van wiskunde in korte tijd nieuwe (basale) wiskunde-opgaven te leren maken. Met andere woorden vergelijkingsleerlingen moeten hun leerwinst behalen op een moeilijker deel van de wiskunde. Deze laatste leerlingen bevinden zich relatief vaak op een 'moeilijker' deel van de toetsschaal (zie tabel 5.9 in hoofdstuk 5).

7) Men kan zich nog afvragen of het inhaaleffect veroorzaakt kan zijn door regressie naar het gemiddelde. Hoewel we met dit regressie-probleem uitdrukkelijk rekening hebben gehouden in het design (door de afname van een tweede voortoets) zijn we door uitval van een groot aantal klassen op deze tweede voortoets niet in onze opzet geslaagd. Er blijft een mogelijkheid dat het inhaal-effect is versterkt door regressie naar het gemiddelde. De verhoudingen tussen de gemiddelde prestaties van de projectleerlingen en die van de vergelijkingsleerlingen liggen zodanig dat er statistische regressie naar het gemiddelde *kán* optreden (zie figuur 5.4 in hoofdstuk 5). Bij zo'n inhaaleffect, waarbij verschillen tussen de projectgroep en de vergelijkingsgroep afnemen van voortoets naar natoets, moet men alert zijn op regressie-effecten. Cook en Campbell (1979, blz. 110) zeggen dat het in die situaties belangrijk is na te gaan *waarom* een experimenteel programma aan een groep is gegeven.

Als de projectscholen voor het SLO-programma hebben gekozen *omdat* zij relatief veel zwakke leerlingen hebben, dan is het regressie-effect te verwachten. We sluiten niet uit dat dit motief inderdaad heeft gespeeld bij de keuze voor het SLO-programma. Omgekeerd kan het zijn dat de vergelijkingsscholen voor een meer traditioneel wiskunde-programma hebben gekozen omdat zij relatief veel sterke leerlingen hebben.

Het regressieprobleem blijkt nog gecompliceerder als men zich realiseert dat het niet vanzelfsprekend is om er vanuit te gaan dat de hele onderzoekspopulatie (projectleerlingen en vergelijkingsleerlingen) één groep vormt en dus naar één gemiddelde trekt. Men kan ook stellen dat de leerlingen in de twee condities twee verschillende groepen vormen. Dan heeft elke groep een eigen gemiddelde. Het regressie verschijnsel kan dan een effect in de andere richting hebben. De keuze voor het ene of andere uitgangspunt is niet vanzelfsprekend. Er blijven dus onzekerheden op dit punt.

Het lijkt ons echter niet aannemelijk dat het inhaaleffect door regressie naar het gemiddelde is veroorzaakt. Van belang is ook om op te merken dat er geen plafondeffecten aanwijsbaar zijn. Deze plafondeffecten kunnen in combinatie met regressie-verschijnselen tot grote interpretatieproblemen leiden. Die combinatie doet zich gelukkig niet voor in ons onderzoek.



8) De moeilijkheidsgraad van de toetsen verdient ook aandacht in een discussie over de resultaten van het onderzoek. De toetsen waren moeilijk en differentieerden zeer sterk tussen de leerlingen. We hebben enkele aanwijzingen dat daardoor de toetssituatie voor de zwakke leerlingen demotiverend kan zijn geweest. Vooral op school 1, met een relatief groot percentage zwakke leerlingen kan dit verschijnsel een rol hebben gespeeld. Hoewel de toetsen zorgvuldig zijn opgebouwd van gemakkelijke opgaven naar moeilijke opgaven, kan de moeilijkheidsgraad er toe hebben geleid dat zwakke leerlingen voortijdig het bijltje er bij neergegoid hebben. Meer in het algemeen kan de moeilijkheidsgraad van de toetsen het beeld van het niveau van de projectleerlingen hebben vertekend: onder de projectleerlingen treffen we aanzienlijk meer zwakke leerlingen aan dan onder de vergelijkingsleerlingen. Het is echter denkbaar dat demotivering van zwakke leerlingen vooral bij de voortoets en in minder mate bij de natoets is opgetreden. In dat geval heeft dat een vertekend effect op de leerwinst (verschilscore). Een leerling die bij de voortoets gedemotiveerd raakt door de moeilijkheidsgraad van de voortoets, heeft daar bij de natoets wellicht minder last van. Hij scoort op de natoets méér overeenkomstig zijn werkelijke prestatieniveau dan op de voortoets. Zo is zijn gemeten verschilscore hoger dan de werkelijke leerwinst. Dat zou de leerwinstscores van de projectleerlingen kunnen flatteren. Omgekeerd kan de herinnering aan het falen op de voortoets(en) de prestatie op de natoets in negatieve zin beïnvloeden en dat kan juist negatief zijn geweest voor de projectleerlingen. Achteraf gezien zou het misschien beter zijn geweest om, in overleg met de docenten, de toetsresultaten mee te laten wegen in een cijfer voor de leerlingen. Vóór het maken van de toets zou dit aan de leerlingen meegedeeld kunnen worden. Dat kan de motivatie verhogen. Nu was de toets iets extra's zonder consequenties voor de leerlingen. Die situatie is misschien wel nadeliger geweest voor de vergelijkingsleerlingen dan voor de projectleerlingen. Deze laatsten konden weten dat het maken van toetsen onderdeel vormt van het experiment aan hun school. Ook al telde de toets niet mee voor een cijfer, de projectleerlingen konden er toch betekenis aan toekennen.

9) Ook Hawthorne-effecten kunnen een rol hebben gespeeld. De project-scholen hebben meer aandacht gehad van ontwikkelaars van de SLO en van onderzoekers. School 3 fungeerde jarenlang als ontwikkelschool van de SLO. School 2 is als participant in het vernieuwingsproces middenschool als voorhoede-school aan te merken. Wellicht heeft deze school ook nog het voordeel dat de werkwijze bij andere vakken ondersteunend werkt, bijvoorbeeld door hantering van soortgelijke werkvormen en het gebruik van situaties uit de leefwereld van de leerlingen. Misschien trekt



deze school andere leraren, leerlingen en ouders aan. Mogelijk heeft dat invloed op het totale klimaat in de school en in de klas. Het zou een te simpele verklaring zijn om de goede resultaten van school 2 eenzijdig toe te schrijven aan het werken met heterogene klassen. Het is evenmin aannemelijk dat deze school het zo goed doet alleen omdat het een middenschool is. In de vorige ronde van ons onderzoek participeerde een middenschool die relatief geringe resultaten boekte. Deze school had veel van de kenmerken die we hier bij school 1 zien en de resultaten waren navenant. School 1 heeft van de drie projectscholen de ongunstigste uitgangspositie. Men werkt nog maar kort met het nieuwe SLO-programma. Men heeft relatief veel onervaren docenten, zwakke leerlingen en allochtone leerlingen. Bovendien heeft men tijdens de uitvoering van het programma met ziekte van docenten te kampen gehad. Er zou wat voor te zeggen zijn om school 1 even buiten haakjes te zetten. Dan zien we veel sterkere effecten (bij school 2 en 3). (zie hoofdstuk 5).

10) Tenslotte moet nog iets worden gezegd over verschillen in afnamecondities. De toetsen zijn niet door de onderzoekers maar door de leraren afgenomen. De onderzoekers zijn bij die afname niet altijd aanwezig geweest. De afname heeft weliswaar in nauw overleg met de onderzoekers, volgens duidelijke schriftelijke instructie, plaatsgevonden maar er waren enkele situatieverschillen bij de afname van de toetsen. Meestal werden de toetsen in de klaslokalen tijdens de normale lessen afgenomen. Bij school 3 zijn de toetsen (voor een deel) in een grote zaal afgenomen, zoals bij een examen het geval is. School 4 nam een bijzondere positie in. Deze school heeft bij de toezegging, om als vergelijkingsschool deel te nemen, als voorwaarde gesteld dat de toetsen (voor een deel) in uitgevalen lessen zouden worden gemaakt. De implicaties hiervan waren tweeledig: (a) de toetsen werden door een leraar afgenomen die een (zieke) collega verving (b) een grotere spreiding in afnamedata in vergelijking met de andere scholen. Deze twee factoren kunnen enigszins afbreuk hebben gedaan aan de vergelijkbaarheid van de toetsresultaten omdat de toetsituatie niet volledig onder controle is gehouden. Het is denkbaar dat de motivatie in school 4 lager is geweest door de bijzondere situatie. In een aantal klassen van school 4 is de voortoets enkele weken later afgenomen dan in de overige scholen. Dat laatste kan de voortoetsresultaten in positieve zin hebben beïnvloed en tegelijk de verschilscore kleiner hebben gemaakt. Overigens is ook op school 4 zorgvuldig met de toetsprocedures omgegaan. Het lijkt niet waarschijnlijk dat verschillen in afnamecondities een grote rol hebben gespeeld.

In de voorgaande discussie zijn niet alle beperkingen van het onderzoek aan de orde geweest. We volstaan met deze tien hoofdpunten die tot een

nuancering van de conclusies kunnen leiden. Men dient de conclusies van ons onderzoek steeds tegen de achtergrond van het gehanteerde design te zien. Er zijn factoren die in het voordeel van de projectleerlingen kunnen hebben gewerkt. Er zijn ook factoren waarvoor het omgekeerde geldt. Wat de vergelijking betreft houden we vast aan de hoofdconclusies (volgens hypothese B). Dat betekent dat in onderwijs waarin het SLO-programma wordt gebruikt:

- (a) de onderwijsleerprocessen meer volgens de bedoelingen verlopen dan in scholen waarin een traditionele aanpak wordt gehanteerd.
- (b) de leerlingen een positievere houding hebben tegenover verschillende aspecten van het onderwijs.
- (c) de leerlingen de doelen van het nieuwe programma sneller en beter bereiken dan leerlingen die met een traditionele aanpak, zijn geconfronteerd.

Of die verschillen praktisch van grotere of kleinere betekenis zijn, lijkt discutabel. Overdreven verwachtingen van een nieuw wiskundeprogramma als zodanig zijn misplaatst. Ook aan dit nieuwe programma zijn nog verschillende problemen verbonden als we uitgaan van 'Wiskunde voor Iedereen'. We denken daarbij in het bijzonder aan het vraagstuk van de zwakke leerlingen.

Het SLO-programma slaat in vergelijking met een meer traditioneel programma geen slecht figuur. De positieve effecten van het SLO-programma krijgen nog meer reliëf tegen de achtergrond van recente vergelijkende effectstudies. In het overgrote deel van de Amerikaanse onderzoeken vond men géén significant verschil in prestaties tussen nieuwe wiskundeprogramma's waarin met kleine groepjes wordt gewerkt en meer traditionele, klassikale benaderingen (Terwel 1986a). In ons onderzoek zien we significante verschillen in percepties, houdingen en prestaties die in het voordeel van het vernieuwingsprogramma uitvallen. Toch is de intentie 'Wiskunde voor Iedereen' nog niet (volledig) gerealiseerd. Is het niveau na het doorlopen van het wiskunde programma wel hoog genoeg? Blijven er niet teveel leerlingen op een te laag niveau steken? Daarbij denken we in het bijzonder aan leerlingen met een laag beginniveau. Het nieuwe wiskundeprogramma rekent ook niet af met het vraagstuk van meisjes en wiskunde. Net als bij de traditionele programma's blijkt dat meisjes achter blijven in prestaties bij de jongens en dat het wiskundeonderwijs hen minder aanspreekt.

Het SLO-programma zou expliciete, didactische maatregelen moeten bevatten voor het opsporen en begeleiden van leerlingen die achterblijven.

Het ontbreken van diagnostische toetsen en het ontbreken van programmaonderdelen voor zwakke leerlingen zien we als een tekort van het SLO-programma.

Voor wat betreft het differentiatievraagstuk heeft de SLO alles op één kaart gezet: het principe van de niveaus in het leerproces in de heterogene groep (Freudenthal 1973 a en b en 1980). Hoewel ook wij het niveauprincipe een zéér belangrijk 'instrument' vinden bij het onderwijs in heterogene groepen, zien we niet in waarom dit andere differentiatiemaatregelen zou moeten uitsluiten. Vooral als blijkt dat het niveauprincipe 'kwetsbaar' is, lijkt het niet verstandig alles op een kaart te zetten. Uit onze observaties (hoofdstuk 2) blijkt dat het 'niveauprincipe' niet door alle leraren wordt gerealiseerd. De ene leraar (in klas B) lukt het wel (vooral in de klassikale reflectie) de andere leraar (in klas A) lukt het niet of nauwelijks. Misschien is het te moeilijk of vereist het een visie die niet alle leerkrachten delen. Wellicht speelt hierbij ook de weinig expliciete wijze, waarop dit niveauprincipe in de leerlingenteksten is uitgewerkt, een rol. We merken uitdrukkelijk op dat onze conclusies met betrekking tot het niveauprincipe betrekking hebben op kwalitatieve observaties in slechts twee klassen. De unieke mogelijkheden en problemen met de niveaus in het leerproces hebben we echter ook in eerdere fasen van ons onderzoek ontmoet. Zo gezien staan de conclusies niet op zichzelf, al hebben we geen kans gezien het niveauprincipe in een groot aantal klassen te meten (zie hoofdstuk 1). Bij de huidige stand van het ontwikkelingswerk en het onderzoek met betrekking tot het niveauprincipe komt men al gauw in een situatie terecht waarbij op verschillen tussen leerlingen niet wordt ingespeeld. Dan komen leerlingen tekort. In het gunstigste geval, zoals in school 2, neemt de docent, naast het niveauprincipe, didactische maatregelen voor een deel van de leerlingen. Hij vangt tekorten in het SLO-materiaal zelf op. Hij neemt ontwikkelings-taken op zich waarvoor de meeste leraren geen tijd en expertise bezitten.

Het is denkbaar dat bij verspreiding van het SLO-programma op grotere schaal de ideeën van Freudenthal als 'alibi' gaan fungeren. Men kan zich dat als volgt voorstellen. Scholen gaan heterogene klassen vormen, maar men ziet op tegen allerlei didactische maatregelen die nodig zijn om op verschillen tussen leerlingen in te spelen. Men formuleert ook geen duidelijke communale en differentiële doelen. Daar is immers het niveauprincipe voor in de plaats gekomen? Het registreren van de individuele voortgang door middel van diagnostische toetsen en het treffen van aanvullende maatregelen voor specifieke categorieën leerlingen, wordt afgewezen. Freudenthal heeft toch zelf het denken in termen van fundamentum en additum gekritiseerd? (zie hoofdstuk 1).

Het alibi zou hierin kunnen bestaan dat leraren zeggen: wij wensen geen

kunstmatige onderwijssituatie, wij werken met 'natuurlijke differentiatie' volgens het niveauprincipe van Freudenthal. Vooral daar waar men het inzicht en de ervaring mist om het niveauprincipe didactisch vorm te geven ontstaat een te weinig gedifferentieerde onderwijsleersituatie. In die situatie komen vooral de zwakke en de sterke leerlingen tekort.

In Engeland waar men in het Sherwood Project over een lange reeks van jaren onderzoek heeft gedaan naar 'mixed ability teaching' in 'comprehensive schools' blijkt dat in heterogene klassen veel te weinig wordt ingespeeld op verschillen tussen leerlingen. Uit systematische observaties blijkt dat het niveau van de uitleg en de discussies in de klas vaak bedroevend laag liggen. Er gaat veel tijd verloren. Leerlingen vervelen zich, ze tonen weinig interesse omdat de taken te moeilijk of te makkelijk voor hen zijn. Er zijn ordeverstoringen. Dit is nadelig voor alle leerlingen maar in het bijzonder voor de leerlingen aan de uiteinden van de 'ability range'. (Kerry, 1982).

In ons onderzoek hebben we aspecten van die problematiek ook aangetroffen. Met name bij school 1 zijn er problemen met de taakgerichtheid en zijn ook de resultaten minder goed. Soortgelijke problemen hebben we in een middenschool gezien in een eerdere fase van ons onderzoek. (Vergelijk Dekker e.a. 1985, tweede interimrapport). Een opvallend verschil met de studie van Kerry en ons onderzoek is dat in Engeland zeer veel klassikaal wordt gewerkt. In ons onderzoek wordt juist op school 1 relatief véél in groepjes gewerkt. Misschien wel te veel. Gemeenschappelijk in beide situaties is echter de weinig expliciete wijze waarop de differentiatie wordt gepland en gerealiseerd. Oppervlakkig gezien is er geen sprake van separatie van leerlingen. Maar als men de effecten toetst blijven er leerlingen achter, zonder dat daar op wordt ingespeeld. Dan verzeilt men in een situatie waarin geen recht wordt gedaan aan de pedagogische hoofdmotieven van Freudenthal, namelijk het voorkomen van separatie en het creëren van een menswaardige, stimulerende onderwijssituatie voor allen. Zo kan 'natuurlijke differentiatie' omslaan in 'natuurlijke selectie'.

Overigens merken we op dat het selectievraagstuk veel meer omvat dan het al of niet nemen van didactische maatregelen voor specifieke groepen leerlingen. Met ons voorstel voor bepaalde didactische maatregelen is selectie uiteraard niet uitgebannen.

De praktische realisering zal nog veel ontwikkelingswerk en onderzoek vergen. Het zal gaan om een juist evenwicht. Wie teveel diagnostische en remediërende maatregelen neemt komt inderdaad terecht in een kunstmatige situatie die moeilijk te organiseren valt. Wat men wint aan extra aandacht voor bepaalde leerlingen, verliest men aan leertijd en begeleidingstijd omdat er zoveel tijd en energie gaat zitten in de organisatorische kant van het klassegebeuren (toetsing, hergroepering enz.). Het is



denkbaar dat door een overdosis aan differentiatie en organisatie de taakgerichtheid afneemt. Dan spant men het paard achter de wagen. Het model Freudenthal, zoals uitgewerkt door de SLO, blijkt in vergelijking met een traditionele aanpak goed te werken. Veel anders hoeft het niet! Het kan wellicht nog beter! Voor dat laatste hebben wij concrete suggesties gedaan in de vorm van het AGO-model (Terwel, 1986a). Dit model wordt in het vervolgproject AGO 12-16 (SVO) nader tot ontwikkeling gebracht en op uitvoerbaarheid en effectiviteit onderzocht. Daarbij zal speciaal worden gelet op de begeleiding van zwakke en sterke leerlingen in heterogene klassen.

Uit de gesprekken met de leerplanontwikkelaars is ons gebleken dat zij huiverig stonden tegenover expliciete maatregelen voor leerlingen die achterblijven of voor leerlingen die juist meer aan kunnen. Daarbij waren hoogstaande ethische overwegingen in het geding. Men vreesde dat het registreren van zwakke en sterke leerlingen tot selectie zou leiden. Daarmee zou hun ideaal van onderwijs in heterogene klassen in gevaar komen. Docenten kunnen de diagnostische toetsen voor selectie-doeleinden gaan gebruiken, zo redeneerde men. Ons onderzoek laat echter zien dat toepassing van het SLO-programma in de huidige vorm en in de huidige onderwijssituatie selectie geenszins uitsluit. De scholen gebruiken selectieve toetsen en men past vormen van homogenisering toe. Slechts op één van de drie SLO scholen was sprake van een evenwichtige verdeling van zwakke en sterke leerlingen over de klassen. Blijkbaar zijn er scholen die vinden dat het wiskunde onderwijs in homogene klassen of stromen gegeven moet worden, ook bij hantering van het SLO-programma. In de praktijk blijkt dat het SLO-programma, hoewel primair ontworpen voor onderwijs in heterogene groepen, niet altijd in deze heterogene situaties wordt gebruikt.

Beslissingen over homogene of heterogene groepering worden vaak op andere gronden genomen dan op onderwijskundige of vakdidactische overwegingen. Vaak ontbreekt een wetenschappelijke fundering. Zo zegt de WRR dat het bij ondermeer wiskunde en nederlands niet haalbaar en niet wenselijk is de leerlingen over een lange periode gemeenschappelijk onderwijs te geven. Men pleit voor aparte stromen in of na het eerste leerjaar. Inmiddels is uit de gedegen case-study van Bonset (1987) gebleken dat het bij Nederlands wél haalbaar is om intern gedifferentieerd onderwijs in een heterogene brugklas te geven. Daarmee weerlegt Bonset de vooronderstellingen van de WRR op empirische gronden.

Ons onderzoek vormt een ondersteuning en uitbreiding van de bevindingen van Bonset. Ons onderzoek toont aan dat het haalbaar is om in de tweede klas met heterogene groepen te werken bij wiskunde. School 2 die de klassen bewust heterogeen samenstelt blijkt over het geheel gezien



evengoede resultaten te boeken als school 3, die een vorm van streaming toepast. Op verschillende aspecten bleek school 2 zelfs betere effecten op te leveren. Ook in het geheel van alle vijf scholen komt school 2 relatief positief naar voren. We hopen met ons onderzoek een bijdrage te leveren aan een rationele discussie op empirische gronden over de haalbaarheid van het werken met heterogene klassen in de eerste fase van het voortgezet onderwijs. We zeggen uitdrukkelijk niet dat school 2 goede resultaten behaalt *omdat* ze werkt met heterogene klassen (vgl. discussiepunt 9 in deze paragraaf). Wel levert ons onderzoek het existentiebewijs van een school die met heterogene klassen werkt en goede resultaten behaalt. Tegelijk moet daar aan worden toegevoegd dat school 3 die met een milde vorm van streaming (banding) werkt ook goede resultaten boekt. Deze school doet al voor de tweede keer mee in ons onderzoek. Beide keren behaalde deze school goede resultaten. Deelnemers aan de (vaak gepolariseerde) discussie over homogeen of heterogeen groeperen lijken de complexiteit van dit vraagstuk nogal eens over het hoofd te zien. Dan raakt men het zicht kwijt op de kernvraag. Deze vraag luidt: hoe komt het dat deze twee scholen relatief goede resultaten behalen en welke rol speelt de classesamenstelling in het totale didactische arrangement? Deze vraag kan nog nader worden gespecificeerd naar effecten op specifieke categorieën leerlingen (sterk/zwak, jongens/meisjes).

Ook de conclusies die betrekking hebben op de effectiviteit van verschillende implementaties van het SLO-programma blijven in het licht van de voorgaande discussie overeind (vergelijk hypothese A). Dat betekent dat in klassen waarin de leraar een banende, taakstellende, (bege)leidende rol vervult en waarin de leerlingen taakgericht samenwerken aan de oplossing van opgaven in contexten, de leerresultaten beter zijn dan in klassen waarin dat minder vaak het geval is.

Nadere analyse is echter nodig voor het opsporen van causale verbanden en het bepalen van de relatieve invloed van de afzonderlijke variabelen. Het is niet alleen de leraar die het onderwijs maakt of breekt. De classesamenstelling, dat wil zeggen het aandeel zwakke leerlingen in verhouding tot de gehele leerlingen-populatie in een klas, lijkt mede bepalend te zijn voor de mate waarin de leraar in staat is om een gewenst handelingsverloop te realiseren. De instroom-kenmerken van leerlingen in scholen en klassen verdienen grote aandacht in de discussie over de kwaliteit van het onderwijs.

## LITERATUUR

### a. Basispublicaties m.b.t. het project ID 12-16

- Terwel, J., *Onderzoeksplan Interne Differentiatie wiskunde 12-16*. Plan voor een onderzoek naar het functioneren van een curriculum voor intern gedifferentieerd wiskunde-onderwijs voor heterogene groepen in het voortgezet onderwijs, Utrecht, IPAW, Vakgroep Onderwijskunde, 1981.
- Dekker, R., Herfs, P., Terwel, J., *Wiskunde voor iedereen*. Eerste interim-rapport project ID 12-16, Vakgroep Onderwijskunde Utrecht, 1983 (a).
- Terwel, J., *Onderwijs maken. Naar ander onderwijs voor 12-16 jarigen*. SVO-reeks nr 77, Harlingen, Flevodruk, 1984 (dissertatie).
- Dekker, R., Herfs, P., Terwel, J. en D. van der Ploeg, *Interne Differentiatie in heterogene brugklassen bij wiskunde*. Tweede interimrapport project ID 12-16. Den Haag, SVO Selectareeks, 1985.
- Posthuma de Boer, M.W., *Werken met heterogene groepen*. Den Haag, SVO Selectareeks, 1986.
- Dekker, R., Herfs, P., Terwel, J., *Continueringaanvraag voor het project ID 12-16 voor de periode 1 maart 1985 - 31 december 1986*. Vakgroep Onderwijskunde Utrecht, 1985 (d).

### b. Overige literatuur waaronder overige projectpublicaties

- Bonset, H., *Onderwijs in heterogene groepen*. Purmerend: Muuses, 1987. Dissertatie.
- Cook, T.D., Campbell, D.T., *Quasi-Experimentation; Design & Analysis Issues for Field Settings*. Chigago, Rand McNally, 1979.
- Creemers, B.P.M., *Formatieve curriculumevaluatie, een heroverweging van de CURVO-strategie*. In: *Pedagogische Studiën* 64, 1987, 2, 59-65.
- Cronbach, L.J., *Course Improvement Through Evaluation*. In: R. Tyler, R. Gagné, M. Scriven, *Perspective of Curriculum-evaluation*. AERA, Chigago, Rand McNally, 1967.
- Dekker, R., Herfs, P., Terwel, J., *Wiskunde voor iedereen op papier en in actie*. In: Halkes, R. en R. Wolbert, (red.): *Docent en Methode*. Lisse, Swets en Zeitlinger, 1985 (a); 89-101.
- Dekker, R., Herfs, P., Terwel, J., *Contexten en begripsontwikkeling bij wiskunde 12-16*. In: *Tijdschrift voor Didactiek der B-Wetenschappen*, 1985 (b); 3, 169-183
- Dekker, R., Herfs, P., Terwel, J., *Groepswerk en klasgesprek essentieel voor nieuw wiskunde-onderwijs*. In: *Didaktief* 15, 1985 (c), 4, 27-28.

- Dekker, R., Herfs, P., Terwel, J., *Continueringaanvraag voor het project ID 12-16 voor de periode 1 augustus 1983 - 31 januari 1985*. IPAW, Vakgroep Onderwijskunde Utrecht, 1983 (b)
- Dekker, R., Herfs, P., Terwel, J., Niveaus in het leerproces als differentiatieprincipe bij het werken in heterogene groepen bij wiskunde. In: Reints, A.J.C. en P. Span (red.), *Differentiatie in het Onderwijs*. Lisse, Swets en Zeitlinger, 1986; 131-141.
- Dekker, R., Real Mathematics. In: L. Filipe & N. Vaz (Ed.), *Didactique de la Mathématique et Réalité Scolaire et Sociale*. Lisboa: CIEAEM, 1983.
- Dekker, R., Meester, F., ATM, girls and maths. *Euclides*, 1984, 59, 449-450.
- Dekker, R., José (spreek uit Gôsé). *Nieuwe Wiskrant*, 1984, 4, 36-39.
- Dekker, R., Maths for all? Just have a look. In: L. Streefland (Ed.), *Proceedings of the Ninth International Conference for the Psychology of Mathematics Education*. Utrecht: OW & OC, 1985.
- Dekker, R., Meisjes samen? *Euclides*, 1985, 60, 214-217.
- Dekker, R., Roos, Rik en Jane. *Willem Bartjens*, 1985, 4, 77-80.
- Dekker, R., Wiskunde voor allen? Bekijk het maar. *Nieuwe Wiskrant*, 1985, 4, 11-12.
- Dekker, R., Mathematical reasoning for all? Just have a look. In: J. de Lange Jzn (Ed.), *Mathématiques pour tous ... à l'age de l'ordinateur*. CIEAEM XXXVII. Utrecht: OW & OC, 1986.
- Dekker, R., Eén context, vijf leerlingen. *Euclides*, 1986, 61, 168-177.
- Dekker, R., Make your own maths. In: P. Bowie (Ed.), *Mathematics for those between 14 and 17. Do they really need it?* Southampton: CIEAEM, 1986.
- Dekker, R., Roos and José, two children in a mixed ability group. *Educational Studies in Mathematics*, 18, 317-324.
- Eijkelhof, H.M.C., Holl, E., Pelupessy, Valk, A.E. van der, Verhagen, P.A.J., Wierstra, R.F.A., *Op weg naar vernieuwing van het natuurkundeonderwijs*. Den Haag, SVO Selectareeks, 1986.
- Fischer, D.L., Fraser, B.J., Validity and use of the Classroom Environment Scale. In: *Educational Evaluation and Policy Analysis*, 5, 1983, 3, 261-271.
- Fraser, B.J., Learning Environment in Curriculum Evaluation: a review. In: *Evaluation in Education: An International Review Series*. Oxford, 1981.
- Freudenthal, H., De niveaus in het leerproces en de heterogene leergroep, met het oog op de middenschool. In: *Gesamtschule conferentie 1973*. Amsterdam, Purmerend, APS, Muuses, 1973a.
- Freudenthal, H., *Mathematics as an educational task*, Dordrecht, 1973b.
- Freudenthal, H., Differentiatie binnen het wiskunde-onderwijs, *Wiskrant* 2, febr. 1976, nr 2, p. 1-2.
- Freudenthal, H., *Weeding and Sowing*. Dordrecht/Boston, Reidel, 1980.

- Grift, W. van der, Mulder, H.P., Het meten van leerlingpercepties van het onderwijsklimaat in de wiskundeles. In: Voogt, J.C. en Reints, A (red), *Naar beter onderwijs*. Zwijzen, 1986, Tilburg, 144-159.
- Hamers, J.H.M., Ruijsenaars, A.J.J.M., *Leergeschiktheid en Leertests*. SVO-reeks 81. Flevodruk Harlingen, 1984. Dissertatie.
- Herfs, P., Speelpenning, J., Terwel, J., Verstappen, P., Leerplanontwikkeling en ontwikkelingsonderzoek. In: *Onderwijskundige Notities* 5, 1985, 2, 13-19.
- Herfs, P., De schommel, een wiskundig rijke context. In: *Nieuwe Wiskrant* 4, 1985 (a), 2, 37-40.
- Herfs, P., Een kwadratische functie aan de leerling(e) gebracht. In: *Nieuwe Wiskrant* 5, 1985 (b), 1, 10-14.
- Horst, P., A generalized expression for the reliability of measures. In: *Psychometrika*, 14, 1949, 21-31.
- Kerry, T., The demands made on Pupils 'Thinking in Mixed Ability Classes'. In: M. Sands en T. Kerry (eds) *Mixed Ability Teaching*. London 1982.
- Krammer, H.P.M., *Leerboek en Leraar*. SVO-reeks 82. Flevodruk Harlingen, 1984. Dissertatie.
- Koehler, S. en P. van den Eeden. *De invloed van classesamenstelling op schoolloopbanen*. Een multilevel onderzoek bij het algemeen voortgezet onderwijs in Amsterdam. Amsterdams samenwerkingsverband. Vakgroep Methoden en Technieken, Vrije Universiteit, publikatie 2, 1987.
- Magnusson, D., *Test Theory*. London. Addison-Wesley, 1966.
- Oud, J.H.L., Onderzoek van orthopedagogische en onderwijskundige interventies aan de hand van tijdreeksen: een MANOVA-procedure. In: *Tijdschrift voor Onderwijsresearch*, 6, 1981, 6, 267-291.
- Meeder, M., F. Meester, R. Dekker, C. Geysel, T. de Poel, *VrouwWiskundig*. Amsterdam: Groep Vrouwen en Wiskunde, 1984.
- Oud, J.H.L., Enige oplossingen voor het regressieprobleem. In: *Tijdschrift voor Onderwijsresearch*, 7, 1982, 2, 83-85.
- Petermann, F., *Veränderungsmessung*. Stuttgart. Verlag W. Kohlhammer, 1978.
- Pedhazur, E.J., *Multiple regression in behavioral research; explanation and prediction*. CBS College Publishing. New York, 1982.
- Ruepert, A.P.A., *Activiteiten van leraren en leerlingen*. Vakgroep Onderwijskunde, R.U.U. 1986. (MOB-scriptie).
- Scriven, M. The Methodology of Evaluation. In: R. Tyler, R. Gagné, M. Scriven, *Perspectives of Curriculumevaluation*, AERA, Chicago, Rand McNally, 1967.
- SLO, *Uitvoeringsplan Visiedeel aug. 1982 - aug. 1984 Wiskunde 12-16*. Enschede, SLO, 1982.



- SLO, *In verband met .....*, een introductie op functies via verbanden. Enschede, SLO, 1983 (a).
- SLO, *Naar aanleiding van .....*, het werken in (kleine) heterogene groepen. Enschede, SLO, 1983 (b).
- SLO, *Situatiebeschrijvingen in wiskundeteksten*. Enschede, SLO, 1984.
- SLO, *Revijs in ontwikkeling*, lesgeven aan heterogene klassen. Enschede, SLO, 1985.
- Terwel, J., Samen verder na de basisschool. In: *Pedagogische Studiën*, 1983 (a), 60, 131-143.
- Terwel, J., Wetenschap of ervaring: een dilemma? In: Creemers, B.P.M. (red.), *De kwaliteit van het onderwijs*. Groningen, RION-congresbundel, 1983 (b), 1-12.
- Terwel, J., Andere kijk op het curriculum voor 12-16 jarigen. In: *Pedagogisch Tijdschrift*, 1985, 10, 426-437.
- Terwel, J., Basisvorming en het ontwerpen van onderwijsleersituaties voor 12-16 jarigen. In: *Pedagogisch Tijdschrift*, 1986 (a), 45-57
- Terwel, J., Leren in coöperatieve groepen. In: Voogt, J. en A.J.C. Reints, (red), *Naar beter onderwijs*. Tilburg, Zwijsen, 1986 (b), 71-100.
- Terwel, J., *Mathematics for all: between dream and reality*. Paper presented at the Annual Meeting of the AERA San Francisco, 1986 (c).
- Terwel, J., *Effecten van Adaptief Groepsonderwijs (AGO 12-16)*. Plan voor een vergelijkend onderzoek naar de implementatie en de effectiviteit van een curriculum voor intern gedifferentieerd onderwijs in de eerste fase voortgezet onderwijs. Utrecht, Vakgroep Onderwijskunde. R.U.U. 1986 (d).
- Terwel, J., Curriculum evaluatie als wetenschappelijke activiteit. In: *Pedagogische Studiën* 64, 1987 (a), 2, 67-73.
- Terwel, J., Basisvorming voor iedereen? Individuele verschillen in het licht van drie researchtradities. In: *Comenius* 7 (1987b), 259-281.
- Terwel, J., Intervisie en leerlingbegeleiding in heterogene klassen. In: N. Deen (red.). *Handboek leerlingbegeleiding*. Samson, Alphen a/d Rijn, juli 1987 (c), blz. 5150-1 t/m 5150-19.
- Terwel, J., *Implementation and effects of a program for mixed ability teaching in secondary mathematics education*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, 1988.
- Terwel, J., Real mathematics in Cooperative Groups. In: N. Davidson (ed.) *Mathematics in Cooperative Groups* (in press).
- Thompson, A.G., The relationship of teachers' conceptions of mathematics and mathematics teaching to instructional practice. In: *Educational Studies in Mathematics*, 1984, 15, 105-127.
- Treagust, D.F. en Fraser, B.J. (e.a.). *Validity and use of classroom and school environment assessments*. AERA symposium. San Francisco, 1986.



Wubbels, Th. (e.a.) Perceptie van de leraar- leerlingrelatie; constructie en kenmerken van een instrument. In: *Tijdschrift voor Onderwijsresearch*, 12, 1987, 1, 3-16.



**ITEMS PERCIA - VRAGENLIJST**

**Leraar, Instructie, klasseklimaat (LIK)**

Onze leraar geeft uitleg aan de hele klas, vóórdat we de opdrachten maken.

Mijn leraar heeft niet in de gaten dat ik veel meer kan presteren.

\*Onze leraar treedt aarzelend op.

Als de opdrachten zijn gemaakt, bespreekt onze leraar die met de hele klas.

Onze leraar is bereid iets opnieuw uit te leggen.

\*Mijn leraar ziet niet dat ik het niet begrijp.

Als ik iets niet snap, heeft mijn leraar daar begrip voor.

De sfeer in onze klas is prettig.

Onze leraar informeert naar wat wij begrepen hebben.

Onze leraar vertelt precies hoe de opdracht gemaakt had moeten worden.

Onze leraar is vriendelijk tegenover zijn leerlingen.

Onze leraar heeft door wat er in de klas gebeurt.

Onze leraar kan goed leiding geven.

Als wij een fout antwoord geven, legt onze leraar uit waarom het fout is.

Onze leraar legt duidelijk uit.

\* = Schaaleinden omgekeerd van negatief item

## **Begeleiding van individuele leerlingen en groepjes (BIG)**

Onze leraar geeft regelmatig toetsen die wij individueel moeten maken.

Bij de start van een nieuw onderwerp vertellen wij over onze ervaringen met het nieuwe onderwerp.

Als ik een toets niet goed gemaakt heb, krijg ik extra werk.

Ik krijg extra hulp van mijn leraar.

Zwakke leerlingen krijgen speciale aandacht en hulp.

Onze leraar controleert of alle leerlingen hun werk gemaakt hebben.

Wij maken afspraken over de manier waarop wij samenwerken.

Onze leraar geeft hulp aan verschillende groepjes.

Onze leraar vraagt hoe wij gewerkt hebben.

Wij vertellen aan de hele klas hoe ons groepje de opdracht uitgevoerd heeft.

Onze leraar helpt sommige leerlingen apart.

### **Samenwerking tussen leerlingen (SAM)**

In ons groepje schrijven we allemaal hetzelfde antwoord op.

\*Wij werken geheel zelfstandig (dus los van elkaar) in de groepjes.

Wij vragen elkaar hulp als we iets niet begrijpen.

De leerlingen lossen opdrachten op verschillende manieren op.

\*Ik begin al met een opdracht als de andere leerlingen uit mijn groepje nog bezig zijn met de vorige opdracht.

Wij helpen elkaar bij het zoeken van een oplossing.

\*Ik doe niet mee aan groepswork.

Bij het uitleggen maakt onze leraar gebruik van goede antwoorden van leerlingen.

Wij begrijpen elkaar.

Wij praten met elkaar over de manier waarop we de opdracht uitgevoerd hebben.

In mijn eentje zou ik bepaalde oplossingen niet gevonden hebben.

\*Ik kan niet goed opschieten met de leerlingen uit mijn groepje.



## Taakgerichtheid en orde (TAKO)

\*Er is rumoer in de klas.

\*Wij praten in ons groepje ook over dingen, die niets met de opdrachten te maken hebben.

\*Wij schrijven het antwoord van elkaar over zonder te controleren of dat antwoord juist is.

\*Wij maken ruzie in ons groepje.

\*Knappe leerlingen vervelen zich in de wiskundelessen.

Wij zijn voortdurend met de opdrachten bezig.

\*In mijn groepje beginnen we al te praten voordat we goed nagedacht hebben.

\*Leerlingen lopen door de klas.

\*Wij rotzooien maar wat aan.

\*Wij begrijpen elkaar niet.

\*Onze leraar laat zich voor de gek houden.

## **Inhoud, contexten (IN)**

De onderwerpen zoals "remweg", "buttons", "autohuren", "luchtfoto", enz. zijn leuk.

De opdrachten gaan over onderwerpen, waar ik wel iets vanaf weet.

De onderwerpen die in de opdrachten gebruikt worden, maken het vak wiskunde aantrekkelijk.

De onderwerpen van de opdrachten maken duidelijk waarover de opdrachten gaan.

\*De onderwerpen maken de opdrachten verwarrend.



## ITEMS VRAGENLIJST HOUDING

### Houding tegenover school (H-school)

De fijnste uren van de dag zijn de uren die ik op school doorbreng.

- \* Ik heb een hekel aan het werk dat ik voor school moet doen.

Ik vind alles wat met school te maken heeft prettig.

Ik wil in mijn leven zoveel mogelijk onderwijs volgen.

- \* Op school is het niet zo fijn.
- \* Het enige wat ik leuk vind aan school is dat ik daar mijn vrienden/vriendinnen kan ontmoeten.

De schooljaren zijn de gelukkigste jaren in het leven.

### Houding tegenover wiskunde (H-wiskunde)

Ik wil erg graag echt goed zijn in wiskunde.

Ik zou graag meer wiskundelessen willen hebben.

- \* Ik besteed liever niet veel van mijn vrije tijd aan wiskunde.

Ik vind wiskunde prettig.

Om een goede baan te krijgen, is het belangrijk dat je wiskunde kent.

- \* Ik kan mij best redden in het dagelijks leven zonder gebruik te maken van wiskunde.

- \* = schaaleinden omgekeerd van negatief item

### **Houding tegenover eigen niveau bij wiskunde (H-niveau)**

Ik begrijp meestal wel waarover het gaat in de wiskundelessen.

- \* Ik ben niet zo goed in wiskunde.
- \* Wiskunde is voor mij moeilijker dan voor de meeste leerlingen in mijn klas.
- \* Ook al doe ik nog zo mijn best, ik ben toch niet goed in wiskunde.



### Houding tegenover belang wiskunde voor jongens en meisjes (H-sexe)

- \* Jongens moeten meer wiskunde kennen dan meisjes.
- \* Jongens zijn beter in wiskunde dan meisjes.

### Houding tegenover verschillende werkvormen (H-werkvormen)

Ik heb een hekel aan het werken in groepjes bij wiskunde.

- \* Als de leerlingen iets uitleggen snap ik het beter dan wanneer de leraar iets uitlegt.

Ik leer het meest van de uitleg die de leraar voor de klas geeft.

Als de leraar mij individueel helpt, snap ik het direct.

Ik werk het liefst alleen aan de wiskunde-opgaven.

### Houding tegenover contexten (H-context)

Het werken met onderwerpen zoals fietsen, de huurprijs van auto's, de remweg van auto's enz. maakt de wiskundelessen aantrekkelijk.

Het werken met onderwerpen zoals fietsen, de huurprijs van auto's enz. vergemakkelijkt het leren van wiskunde.

- \* De onderwerpen zoals fietsen enz. brengen mij in de war.

### Semantische Differentiaal

Semantische Differentiaal wiskundelessen

* boeiend	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	vervelend
moeilijk	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	makkelijk
* ontspannen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	gespannen
* goed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	slecht
* plezierig	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	onplezierig

Semantische Differentiaal wiskundeleraar of lerares

* vriendelijk	o	o	o	o	o	onvriendelijk
* opwekkend	o	o	o	o	o	teleurstellend
gejaagd	o	o	o	o	o	kalm
* duidelijk	o	o	o	o	o	onduidelijk
rumoerig	o	o	o	o	o	rustig
rommelig	o	o	o	o	o	ordelijk

Semantische Differentiaal Groepje bij wiskunde

* samenwerken	o	o	o	o	o	alleen werken
* ijverig	o	o	o	o	o	lui
elkaar afwijzen	o	o	o	o	o	elkaar accepteren
* elkaar helpen	o	o	o	o	o	egoïstisch

Toets 3 (natoets)

WAT WEET JE NU ?

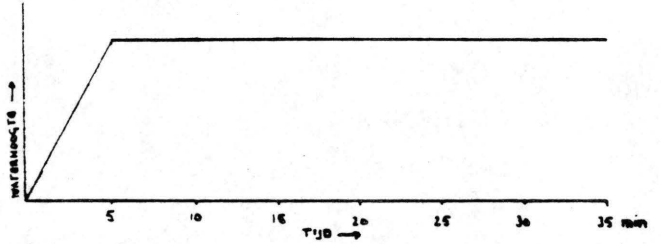
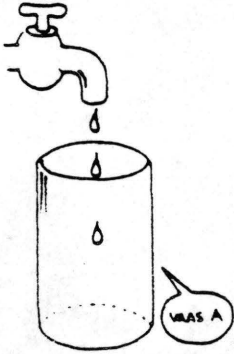
VOORNAAM :  
ACHTERNAAM :  
KLAS :  
SCHOOL :  
DATUM :

LET OP:

Deze toets lijkt op de toetsen die je al gemaakt hebt. Toch zijn de opgaven meestal anders dan de vorige keer. Bovendien zijn er geheel nieuwe opgaven bijgekomen. Het is verstandig om elke opdracht goed te lezen.

VAZEN

a. Hieronder zie je een vaas onder een kraan staan. De kraan druppelt heel regelmatig. Daarnaast staat een grafiek die het verloop van de waterhoogte in de vaas beschrijft.

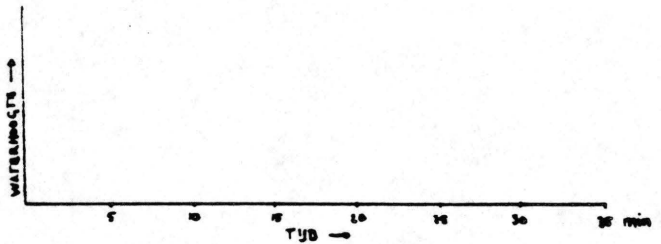
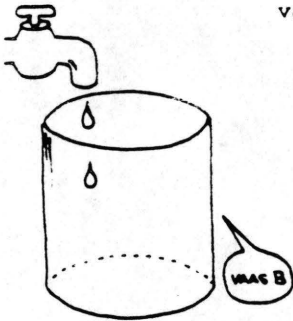


Na hoeveel minuten is de vaas vol?

na ..... minuten

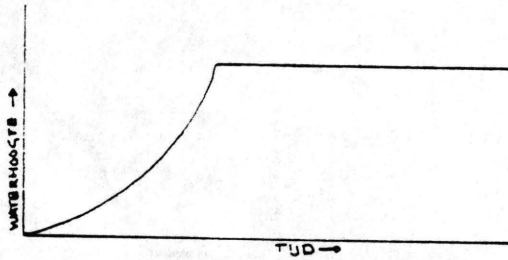
b. In vaas B (hieronder) kan drie keer zoveel water als in vaas A (hierboven). Na hoeveel minuten is vaas B vol?

vaas B is na .....minuten vol



c. Teken hierboven de grafiek die het verloop van de waterhoogte in vaas B beschrijft.

d. Hieronder staat een grafiek die het verloop van de waterhoogte in een vaas beschrijft. De vaas staat onder een kraan die heel regelmatig druppelt.

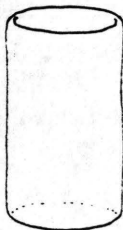


Welke van de 4 hieronderstaande vazen hoort bij de grafiek?

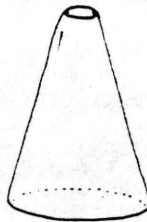
- vaas A hoort bij de grafiek
- vaas B hoort bij de grafiek
- vaas C hoort bij de grafiek
- vaas D hoort bij de grafiek



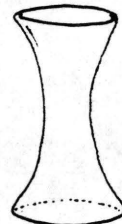
A



B



C

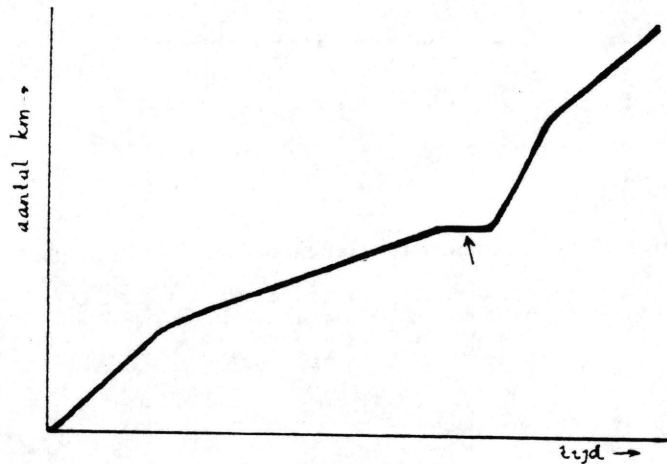


D



M & M

Michel fietst nogal eens van zijn huis naar het huis van zijn vriendin Maureen. Zijn fietstocht ziet er meestal zo uit, in een grafiek:

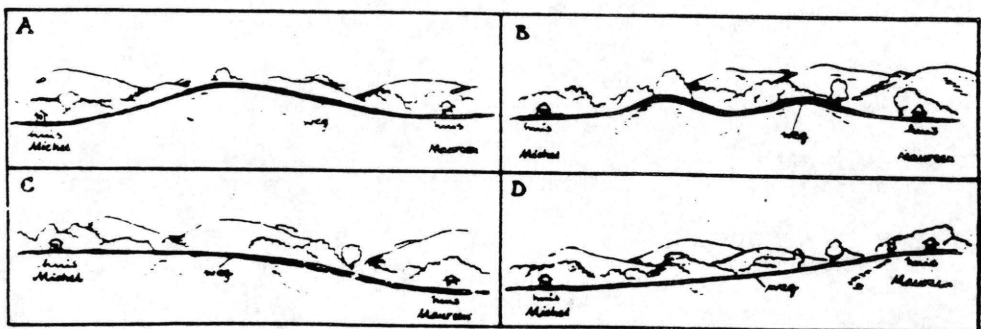


a. Zoals je ziet zit er in de grafiek een horizontaal stukje. (zie pijltje)  
Wat betekent dat? Kruis het goede antwoord aan.

- Hij loopt even
- Hij rijdt heel langzaam
- Hij staat even stil
- Hij fietst een stukje terug

b. Welk kaartje past volgens jou het best bij de weg die Michel fietst?

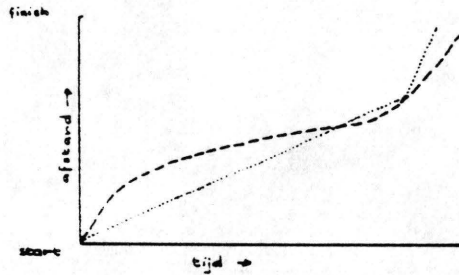
- kaartje A
- kaartje B
- kaartje C
- kaartje D



## HARDLOPEN

Marion en Ruud doen op koninginnedag mee aan een hardloopwedstrijd om het voetbalveld bij hen in de buurt. Ze gaan tegelijk van start. Ruud begint in een hoog tempo, maar krijgt al snel pijn in zijn zij en komt haast niet meer vooruit. Toch weet hij zich te herstellen en krijgt langzaam zijn snelheid weer te pakken.

Marion begint rustig aan en houdt dit tempo lang vol. Als ze de finish ziet gaat ze zelfs nog een stuk harder lopen.



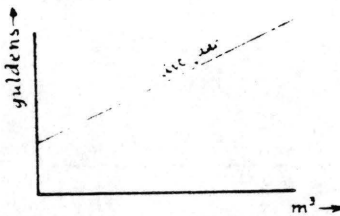
- a. Vul in ,welke lijn hoort bij wie?  
de stippellijn hoort bij .....  
de streepjeslijn hoort bij .....
- b. Hoe vaak gaan ze elkaar onderweg voorbij?
- 0 keer
  - 1 keer
  - 2 keer
  - 3 keer
- c. Wie is het eerst bij de finish?
- Marion
  - Ruud
  - ze komen gelijk aan

TARIEFSWIJZIGING VOORGESTELD



Het gemeentelijk waterbedrijf in Apeldoorn wil voor volgend jaar het tarief voor geleverd water veranderen. Dit tarief bestaat uit de kosten van het water dat je per jaar verbruikt en een vast bedrag dat je ook elk jaar moet betalen. Dit vaste bedrag noemen ze het vastrecht.

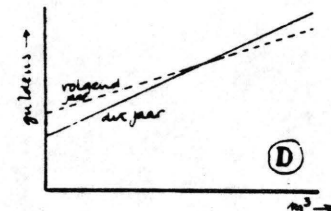
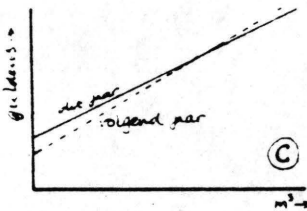
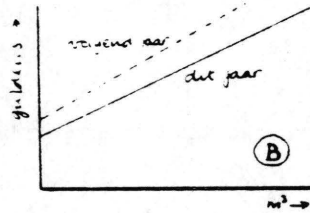
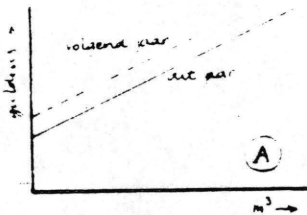
"We willen het vastrecht wat laten zakken, want we hebben weinig onderhoudskosten gehad. Daar staat tegenover dat we de prijs per m<sup>3</sup> water wel wat omhoog laten gaan", aldus de directie van het waterbedrijf. In grafiek nr. 1 zie je het tarief zoals dit jaar wordt gehanteerd.



grafiek nr. 1

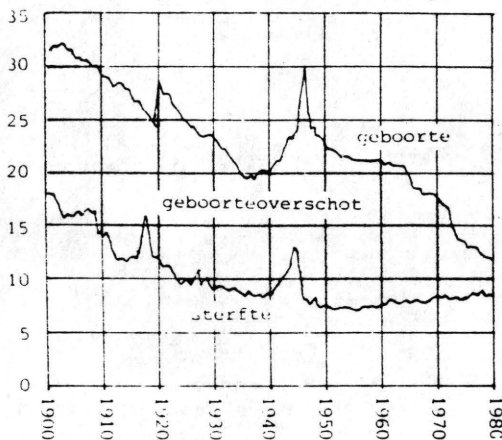
Welke van de grafiekjes A tot en met D geeft het voorstel van de directie weer?

- grafiek A geeft het voorstel voor volgend jaar weer
- grafiek B geeft het voorstel voor volgend jaar weer
- grafiek C geeft het voorstel voor volgend jaar weer
- grafiek D geeft het voorstel voor volgend jaar weer



## GEBORTEOVERSCHOT

Deze grafiek geeft een overzicht van het aantal geboortes en sterfgevallen per 1000 inwoners in Nederland vanaf 1900 tot 1980.



Geboorte en sterfte (en geboorteoverschot) in Nederland per 1000 inwoners, sinds 1900 .(bron: CBS)

a. Hoeveel kinderen werden er in 1950 per 1000 inwoners geboren?

..... kinderen

b. En hoeveel in 1975?

..... kinderen

c. In 1950 waren er 10 miljoen (10.000.000) inwoners in Nederland. Hoeveel kinderen zijn er toen in totaal in ons land geboren?

..... kinderen

d. Met hoeveel personen is de bevolking van Nederland in 1950 toegenomen? Maak een schatting. Houd rekening met sterfte! Laat zien hoe je aan je antwoord gekomen bent.

..... personen

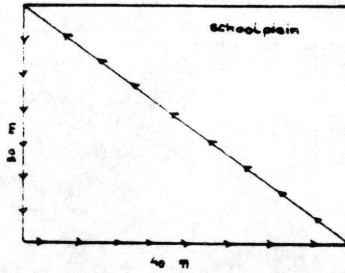
e. Verklaar waarom het woord 'geboorteoverschot' tussen beide grafieken staat.

f. Als er in een jaar meer mensen overlijden dan er kinderen worden geboren, hoe zou je dat dan aan de grafiek kunnen zien?



VOLLEYBAL

Als we buiten gaan volleyballen, beginnen we altijd met hardlopen op schoolplein. Zo lopen we dan:

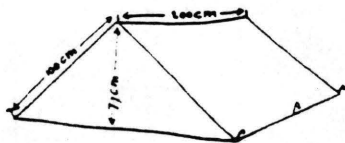


Als we de driehoek 2 keer lopen, hoeveel meter denk je dat we dan lopen?

..... meter

## KAMPEREN

Shirley en Peggy gaan dit jaar samen op vakantie met dit tentje:



Ze kunnen van iemand 2 luchtbedden lenen van 65 cm breed en 180 cm lang.

- a. Wat denk je, passen de bedden naast elkaar in de tent?
- ja, ze passen
  - nee, ze passen niet
- b. Leg uit, waarom wel of niet.

KRANTENKNIPSEL

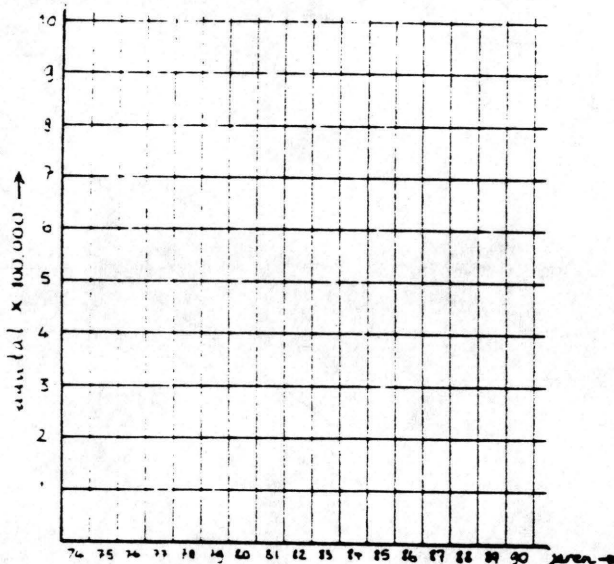
Lees het krantenknipsel uit de Voorburgse Courant goed door.



Ine van der Steen: huogevol gestolen

derd vooral de laatste jaren  
 We zijn er niet oostevreden over  
 aldus de woordvoerder. zeker als  
 je bedenkt dat het er naar uitgaat  
 dat het tot een miljoen zou oplo-  
 pen. Eind '74 waren het er onge-  
 veer 100 000, eind '75 al 160 000.  
 Tot eind '80 bleef het min of meer  
 constant op een niveau van 200 000.  
 Toen begon het enorm op te lopen.  
 Eind '81 waren het er al 300 000  
 en in het volgende jaar verdubbel-  
 de dat aantal zelfs. Daarna ging  
 het wat snider snel, maar het liep  
 toch op tot ruim 625 000 eind '84.  
 Gelukkig was toen het hoogste punt  
 bereikt en is er sindsdien een  
 kleine daling waarneembaar. We  
 verwachten 60 000 snider in '85  
 snieschies zelfs iets meer, aldus  
 een zichtbaar huogevol gestolde  
 mevrouw dr. Ine van der Steen.

uit: Voorburgse Courant, met foto



a. Maak in het rooster een grafiek waarin je het verloop kunt zien vanaf 1974 tot nu.

b. Als de daling ook in de volgende jaren net zo is als de verwachting van mevrouw Van der Steen, hoeveel zouden er dan in 1988 zijn?

.....

c. Waarover zou het artikeltje eigenlijk gaan? Maak een keuze uit de volgende vier mogelijkheden.

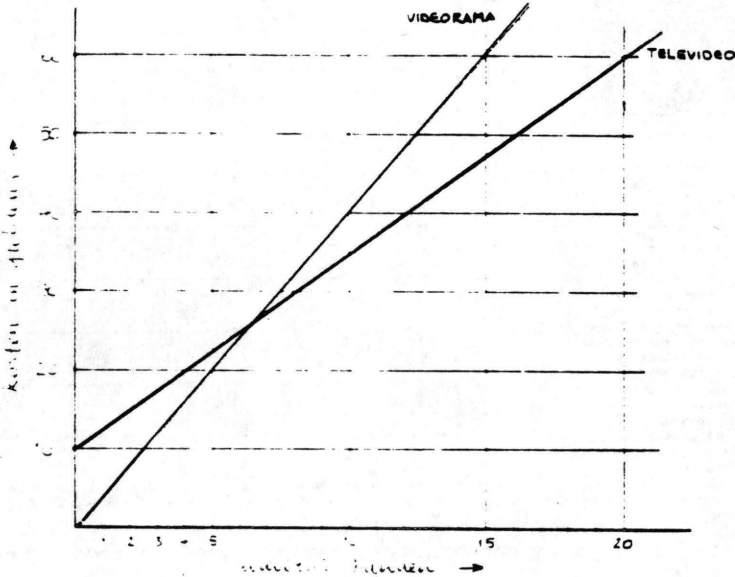
- het aantal gestolen auto's in Nederland
- het aantal nieuw gebouwde huizen in Nederland
- het aantal werklozen in Nederland
- het aantal bezoekers van Ponymark Slagharen

d. Leg uit waarom je die mogelijkheid hierboven gekozen hebt.

VIDEOBANDEN

Bij ons in de wijk zijn twee winkels waar je videobanden kunt huren, TELEVIDEO en VIDEORAMA.

Bij de een moet je de eerste keer dat je een videoband wilt huren eerst een pasje kopen, bij de ander hoeft dat niet en betaal je alleen een bedrag per band. Beide rekenen een tarief per dag. De kostengrafieken zie je hieronder.



- a. Bij welke winkel moet je eerst een pasje kopen?  
.....
- b. Hoe kun je dat aan de grafiek zien?
- c. En hoeveel moet je voor dat pasje betalen?  
..... gulden
- d. Hoeveel is de huurprijs per videoband (per dag) bij VIDEO-  
RAMA ?  
..... gulden
- e. En hoeveel bij TELEVIDEO ?  
..... gulden

f. Zoals je kunt zien, snijden de grafieken elkaar. Wat betekent dat voor mensen die vaak videobanden huren?

g. VIDEORAMA heeft een speciale aanbieding deze maand:

3 banden halen maar 2 betalen
----------------------------------

Hoeveel moet je betalen als je -deze maand- 15 banden huurt?

..... gulden

h. Als je deze maand bij TELEVIDEO 40 banden huurt, hoeveel moet je dan betalen? Je hebt al een pasje.

..... gulden



## KOPER

Paul heeft van z'n oom een paar stukjes koper gekregen. Reststukjes die overbleven bij het gieten van koper. Ruw en raar van vorm maar wel mooi. En flink zwaar. Paul heeft ze zo nauwkeurig mogelijk gewogen:

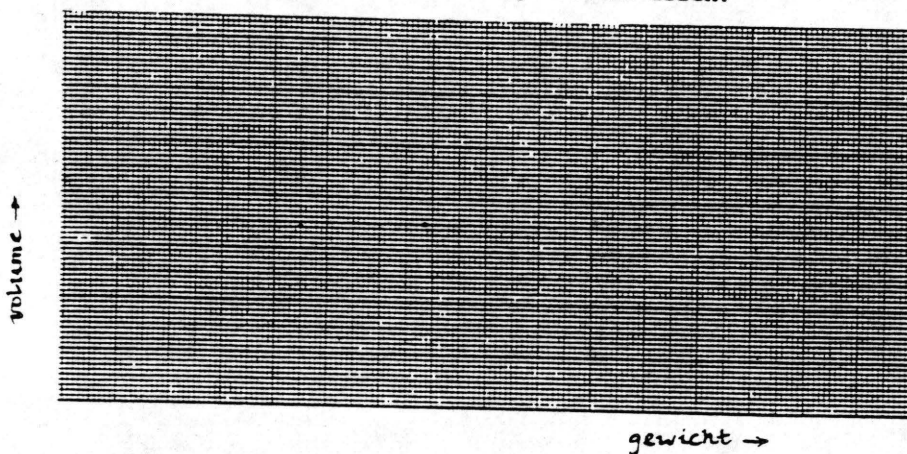
75,4 gram, 52,5 gram 12,4 gram en, de zwaarste, 130,4 gram.

Door ze onder te dompelen in water kon hij ook het volume vaststellen van de verschillende hompjes.

De kleinste was 2,1 cm<sup>3</sup>. Verder een van 6,4 cm<sup>3</sup>, een van 9,2 cm<sup>3</sup> en, tenslotte, een van 15,9 cm<sup>3</sup>.

a. Zet de gegevens die bij elkaar horen in het onderstaande rooster.

Kies zelf een handige schaalverdeling voor de assen.



"Hmm.." zegt Paul al hij naar het rooster kijkt.  
"Ziet er naar uit dat er iets fout is.."

b. "Als het gewicht van dat ene hompje koper goed is, dan zou je een volume verwachten van ..... cm<sup>3</sup>."

c. Maar als het volume goed is, dan zou je verwachten dat het ..... gram weegt."

d. Gebruik de informatie die jullie nu hebben om te bepalen wat het gewicht van 1 cm<sup>3</sup> koper is.

..... gram

GASREKENEN

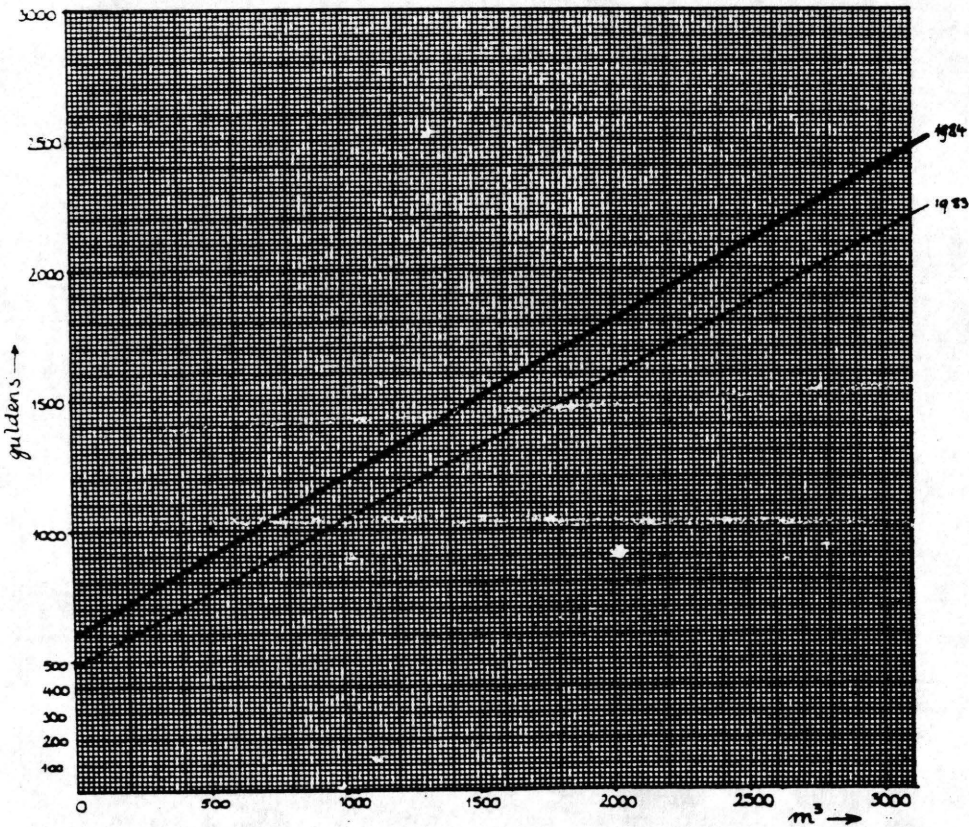
Onderaan deze bladzijde staan twee grafieken getekend. Ze laten zien hoe hoog de gasrekening was, afhankelijk van het aantal m<sup>3</sup> gas dat is verbruikt. De ene gaat over 1983, de ander over 1984. De gasrekening bestaat uit de kosten van het gas dat per jaar verbruikt wordt en een vast bedrag. Dit vaste bedrag noemen ze het vastrecht

a. Met hoeveel gulden is het vastrechtbedrag in 1984 verhoogd?

..... gulden

b. Bij Ellen thuis gebruiken ze elk jaar zo'n 3000 m<sup>3</sup> gas. Hoeveel moesten ze daarvoor in 1984 betalen?

..... gulden



c. En welk gedeelte van dat bedrag was voor vastrecht in 1984?  
..... gulden

d. Hoe hoog was in 1984 de prijs per m<sup>3</sup> gas?  
..... gulden

e. Ellen zegt: "Die grafieken snijden elkaar niet. Volgens mij betekent dat dat iedereen in 1984 meer voor gas moest betalen dan in 1983."  
Ben je het met Ellen eens?

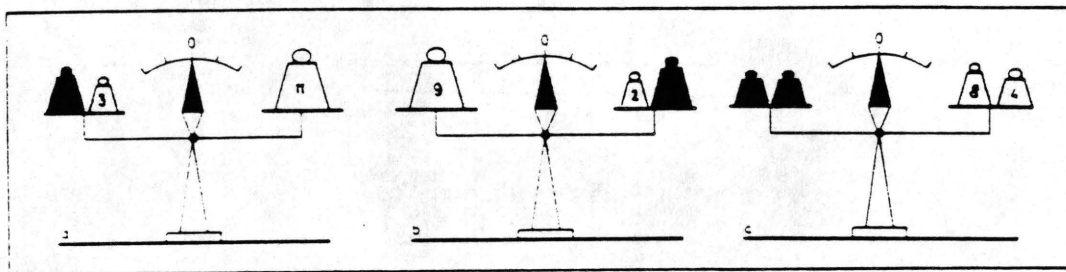
- ja
- nee

Licht je antwoord toe.

BALANS

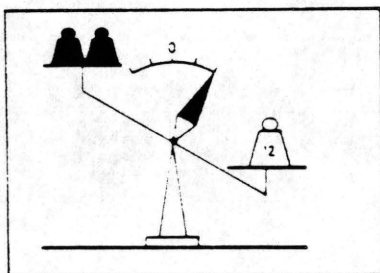
a. In dit plaatje (plaatje 1) zijn 3 balansen getekend. Hoeveel wegen de zwarte gewichten als de balansen in evenwicht zijn? Zwarte gewichten die even groot zijn getekend, zijn ook even zwaar.

Bij a weegt het zwarte gewicht .....  
 Bij b weegt het zwarte gewicht .....  
 Bij c weegt elk zwart gewicht .....



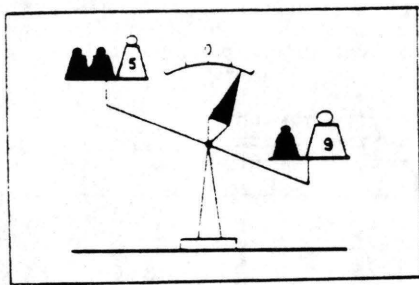
PLAATJE 1

b. De balans in plaatje 2 is niet in evenwicht. Wat weet je van elk zwart gewicht?



PLAATJE 2

c. De balans in plaatje 3 is niet in evenwicht. Wat weet je van elk zwart gewicht?



PLAATJE 3

d. Als je deze opdrachten niet snapt mag je ze overslaan.

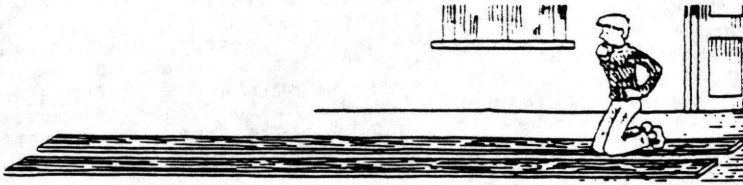
$$?x \mid 2x + 3 > x - 3$$

$$?x \mid 2x + 3 = x - 3$$

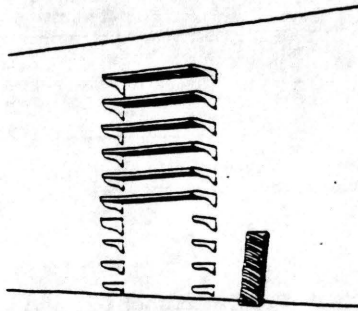


## BOEKENKAST

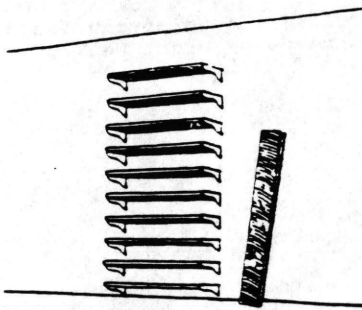
Mark heeft twee houten planken die even lang zijn. Hij wil een boekenkast maken met 10 planken van gelijke lengte.



Hij zaagt 6 stukken van de eerste plank. Hij houdt 24 cm over.



Hij zaagt 4 stukken van de tweede plank en houdt 146 cm over.



- Hoe lang is elke boekerplank?  
..... cm
- Schrijf op hoe je dat uitgerekend hebt.

GLAASJE OP? LAAT JE .....

Alcoholische dranken worden meestal in speciale glazen geschonken. Je weet vast wel hoe een bierglas er uitziet, en je hebt ongetwijfeld wel eens van wijnglazen, sherryglazen, borrelglazen enzo gehoord. Wat je misschien wel is opgevallen is dat 'hoe sterker de drank is, hoe kleiner het glaasje'. De 'Warenwet' \*) schrijft dat ook voor. Die zegt namelijk, dat een standaard-glas met een of andere alcoholische drank een vaste hoeveelheid zuivere alcohol moet bevatten. Daarom worden sterke dranken, die 'weinig verdund zijn', in kleine glaasjes geserveerd en worden de zogenaamde 'licht alcoholische dranken' in grotere glazen geschonken.

Hieronder zie je het allemaal wat nauwkeuriger:

drank	alcoholpercentage	inhoud standaard-glas in cm <sup>3</sup>
bier	5	240
Bordeaux wijn	12	100
Champagne	20	60
Wodka	50	24

a. Probeer 's een voorspelling te doen over het alcoholpercentage van Tequilla, een Mexicaanse drank, waarvan het standaard-glas 30 cm<sup>3</sup> inhoudt.

Tequilla bevat ..... % alcohol

b. Shandy, een drankje dat je misschien wel 's geproefd hebt, heeft 1% alcohol in zich. Het Shandyglas heeft een inhoud van 200 cm<sup>3</sup>. Hoe groot zou het Shandyglas volgens de Warenwet moeten zijn?

..... cm<sup>3</sup>

c. Er is een verband tussen alcoholpercentage en inhoud van het standaard-glas bij de genoemde dranken. Welk?

---

\*) De Warenwet is een wet die bijv. zegt aan welke eisen etenswaren en dranken moeten voldoen die verkocht worden.

TAXI VOOR OMA

Omdat ik niet goed ter been ben, heb ik nogal eens een taxi nodig om bij mijn kleinkinderen te komen.

Ik ben er nog steeds niet achter hoe de taxichauffeurs het bedrag bepalen dat je betalen moet. Wat ik wel heb begrepen is dat het bedrag afhangt van het aantal kilometers dat de rit lang is en van het aantal minuten dat de rit duurt.

Dat laatste is trouwens ook wel redelijk, vind ik. Want als de taxi in een file terecht komt, dan kan een kort ritje heel lang duren. En dan zou de chauffeur nauwelijks iets verdienen. Vandaar niet alleen een afstands-tarief, maar ook een tijds-tarief.

Hier zie je de afstand en duur van mijn laatste 4 ritten:

nr	afstand in km	tijd in min.	kosten in gld
1	12	16	20
2	10	30	25
3	15	20	25
4	10	20	20

a. Bij een van de ritten kwamen we in een opstopping terecht. Welke zou dat geweest zijn?

rit nummer .....

b. Hoe zou het taxitarief in elkaar zitten?

c. Wat moet je betalen voor een rit over 6 kilometer die 10 minuten duurt?

..... gulden

d. 'Verzin' een rit die precies een tientje kost.

.....kilometer en ..... minuten

MAIS

In een laboratorium wordt hard gewerkt aan het verbeteren van de groei van maisplantjes. Om te kijken wat het beste is heeft men verschillende bakken met plantjes. De plantjes in bak 1 krijgen alleen water, die in bak 2 water met kalk, bak 3 geeft men water met ijzer en de plantjes in bak 4 krijgen water vermengd met kalk en ijzer.

De resultaten van bak 4 zijn als volgt:

gemiddelde groei in 2 dagen in mm	Bij toediening van	
	kalk in grammen	ijzer in grammen
26	2	5
17	1	4
21	4	1

Meneer van de Berg, die in het laboratorium werkt, kijkt eens goed naar die getallen en zegt: "Volgens mij is er een duidelijk verband te vinden, want als ik 5 keer het aantal grammen kalk neem en daar 3 keer het aantal grammen ijzer bij optel, dan is de uitkomst daarvan precies gelijk aan de gemiddelde groei van de plantjes".

a. Leg uit waarom dat verband niet goed kan zijn.

b. Ga er even van uit dat het verband wel klopt, dus dat 5 maal het aantal grammen kalk plus 3 keer het aantal grammen ijzer het aantal millimeters geeft dat het plantje gegroeid is. Als een plantje 5 cm gegroeid is, zoek dan eens uit hoeveel gram kalk en hoeveel gram ijzer men dat plantje gegeven heeft.

..... gram kalk  
 ..... gram ijzer

c. Bij vraag c is er niet één antwoord. Je kunt ook andere mogelijkheden vinden.  
Laat al die mogelijkheden eens zien.





FOTO

Maak -zo je wilt- gebruik van de foto om te achterhalen of de volgende uitspraken waar zijn.

a. Dingen, die in de werkelijkheid even groot zijn, zijn op een foto altijd ook even groot

- waar
- onwaar

b. Lijnen, die in de werkelijkheid evenwijdig zijn, zijn op een foto altijd ook evenwijdig

- waar
- onwaar

c. Lijnen, die elkaar in de werkelijkheid snijden, snijden elkaar op een foto ook altijd.

- waar
- onwaar

d. Hoeken, die in de werkelijkheid recht zijn, zijn op een foto ook altijd recht (90 graden).

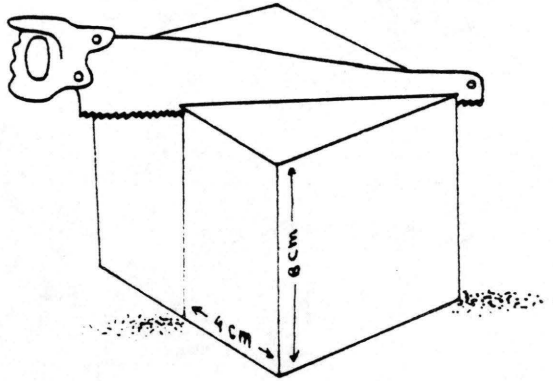
- waar
- onwaar

e. Lijnen, die in de werkelijkheid recht zijn, zijn op een foto ook altijd recht.

- waar
- onwaar

ZAGEN

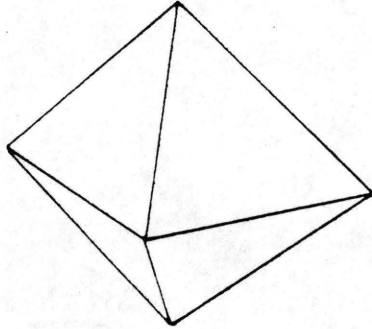
Van een houten kubus wordt een hoek afgezaagd. Als de hoek eraf is komt er op de kubus een nieuw vlak te voorschijn, het zaagvlak.



Teken het zaagvlak op ware grootte.

## VLAK

Dit is een plaatje van een regelmatig acht-vlak. De zijvlakken zijn allemaal regelmatige driehoeken. Elke driehoek heeft dus drie zijden die even lang zijn.



Schets een bouwplaat van zo'n regelmatig acht-vlak





Relatieve scores voortoets (moeilijkheidsgraad)

---

Deze scores geven per item het gemiddelde percentage van de maximaal te behalen score (zie hoofdstuk 5, tabel 5.1).

Relatieve scores per item op de voortoets, hele bestand (N = 811).

---

Items	Gemiddelde (%)	Standaarddeviatie
vazen	72.7	29.0
Michel en Maureen (fietsen)	47.8	39.0
hardlopen	68.5	23.6
tarief	47.6	50.0
geboorte	24.0	18.9
volleybal	19.9	39.9
kamperen	12.0	21.5
krant	44.5	22.8
video	43.0	23.6
gas	28.3	18.8
balans	44.4	17.7
boek	17.3	36.9
glasje op taxi	10.7	18.0
taxi	25.0	22.7
foto	56.4	26.9
zagen	6.6	24.9

---

Relatieve scores per item voortoets, projectleerlingen (N = 511)

Items	Gemiddelde (%)	Standaarddeviatie
vazen	70.2	30.2
Michel en Maureen (fietsen)	48.0	39.4
hardlopen	67.3	24.7
tarief	45.0	49.0
geboorte	20.5	17.9
volleybal	21.5	41.1
kamperen	14.5	22.7
krant	40.8	22.4
video	37.4	22.1
gas	24.9	16.7
balans	38.1	14.7
boek	11.7	31.0
glaasje op	7.4	14.6
taxi	22.8	20.5
foto	54.7	26.5
zagen	3.7	18.9

Relatieve scores per item voortoets, vergelijkingsleerlingen (N = 300)

Items	Gemiddelde (%)	Standaarddeviatie
vazen	77.0	26.5
Michel en Maureen (fietsen)	47.7	38.4
hardlopen	70.5	21.6
tarief	52.0	50.0
geboorte	30.0	19.1
volleybal	17.0	37.6
kamperen	7.7	18.5
krant	50.7	22.2
video	52.5	22.9
gas	34.3	20.6
balans	55.2	17.2
boek	27.0	43.5
glaasje op	16.3	21.5
taxi	28.8	25.6
foto	59.4	27.4
zagen	11.7	32.2

Relatieve scores natoets (moeilijkheidsgraad)

Deze scores geven per item het gemiddelde percentage van de maximaal te behalen score (zie hoofdstuk 5 tabel 5.1).

Relatieve scores per item op de natoets, hele bestand (N = 799)

Items	Gemiddelde (%)	Standaarddeviatie
vazen	81.8	24.5
Michel en Maureen (fietsen)	53.5	40.1
hardlopen	72.1	23.8
tarief	68.2	46.6
geboorte	26.8	20.3
volleybal	34.9	47.7
kamperen	18.5	29.5
krant	51.1	20.7
video	53.6	23.9
koper *	11.5	14.1
gas	38.6	22.8
balans	50.1	19.6
boek	29.5	44.2
glaasje op	14.9	21.9
taxi	33.5	29.7
mais *	15.7	27.5
foto	62.7	26.2
zagen	14.6	34.4
vlak *	20.1	30.0

\*Extra items die niet in de voortoets voorkomen maar wel in de natoets

Relatieve scores per item natoets, projectleerlingen (N = 490)

Items	Gemiddelde (%)	Standaarddeviatie
vazen	81.2	25.1
Michel en Maureen (fietsen)	51.7	40.1
hardlopen	72.9	23.0
tarief	74.1	43.9
geboorte	24.9	19.8
volleybal	30.8	46.2
kamperen	18.6	28.1
krant	49.8	20.4
video	53.9	24.0
koper *	11.3	13.5
gas	41.9	23.3
balans	45.2	16.7
boek	22.1	40.0
glaasje op	11.9	19.9
taxi	29.7	25.9
mais *	14.1	26.9
foto	61.0	27.3
zagen	17.6	38.1
vlak *	20.3	28.5



Relatieve scores per item, natoets vergelijkingsleerlingen (N = 289)

Items	Gemiddelde (%)	Standaarddeviatie
vazen	82.9	23.3
Michel en Maureen (fietsen)	56.6	39.9
hardlopen	70.8	25.2
tarief	58.1	49.4
geboorte	30.0	20.7
volleybal	41.9	49.4
kamperen	18.3	31.9
krant	53.3	21.1
video	53.1	23.8
koper *	11.7	15.1
gas	33.2	20.8
balans	58.7	21.2
boek	42.9	48.2
glaasje op	20.0	24.0
taxi	40.1	34.3
mais *	18.3	28.4
foto	63.9	24.0
zagen	9.7	29.6
vlak *	19.8	32.4

\*Extra items die niet in de voortoets voorkomen maar wel in de natoets.

OBSERVATIEGEGEVENS PROJECTKLASSEN

Klasnummer	Groepswerk per klas in % (Groepob)	Niet-taakgerichte momenten tijdens groepswerk in % per klas (N-Takob)*
11	78	69
12	61	57
13	82	75
14	55	47
15	67	29
16	96	61
17	73	29
18	61	48
21	80	7
22	37	.4
23	51	6
24	57	5
25	57	2
26	61	7
31	23	19
32	42	20
33	55	10
34	43	17
35	51	41
36	49	4
38	40	10
39	0	0

Gemiddelde Groepob  
per school in %

Gemiddelde N-Takob  
per school in %

School 1	72	52
School 2	57	5
School 3	38	15

\* N-Takob is "gecorrigeerd" voor het percentage groepswork dat in een bepaalde klas voorkomt.

Dat wil zeggen  $N\text{-Takob} = \frac{\text{het percentage niet-taakgerichte momenten}}{\text{van de totale tijd aan groepswork besteed maal Groepob}}$ .

Door deze correctie is N-Takob te lezen als percentage van de totale lestijd.





Dit onderzoek betreft de uitvoering en de effecten van interne differentiatie in het voortgezet onderwijs bij wiskunde. Het onderzoek is uitgevoerd in het tweede leerjaar van (brede) scholengemeenschappen voor voortgezet onderwijs. Daarbij is een vergelijking gemaakt tussen scholen met een nieuw programma en scholen met een meer traditionele aanpak.

In het nieuwe programma wordt de wiskunde in situaties uit het dagelijks leven geplaatst. En er wordt gewerkt met een combinatie van klassikale instructie en groepswerk. Het nieuwe programma is ontwikkeld door het Instituut voor de Leerplanontwikkeling in Enschede (SLO). Het onderzoek is uitgevoerd door de vakgroep Onderwijskunde aan de Rijksuniversiteit Utrecht.

De conclusie is dat de leerlingen die het nieuwe programma hebben doorlopen, betere resultaten boeken dan de leerlingen in het traditionele programma. Niet alleen de leereffecten zijn beter, ook de houding van de leerlingen tegenover aspecten als wiskunde in contexten, de leraar en de groep is positiever. Bovendien werken de experimentele leerlingen beter samen en blijven zij meer bij de taak.

Er werden aanzienlijke verschillen gevonden in de manier waarop het nieuwe programma in de klas werd uitgevoerd. In klassen waarin het nieuwe programma volgens de bedoelingen is uitgevoerd waren de resultaten beter dan in klassen waarin de uitvoering minder adequaat verliep.

De uitvoering van het nieuwe programma vond plaats onder verschillende condities van heterogeniteit. Het bleek dat in die situaties waarin met volledig heterogene klassen werd gewerkt (minstens) evengoede resultaten werden behaald dan in situaties waarbij de leerlingen in meer homogene stromen waren geplaatst. Het blijkt mogelijk, in het tweede leerjaar, onderwijs van goede kwaliteit te realiseren in volledig heterogene klassen. Dat is een belangrijke conclusie vanuit het gezichtspunt van „Basisvorming voor iedereen”.

Toch blijven er nog wensen over. Ook bij dit nieuwe programma blijven sommige leerlingen achter. Speciale aandacht is nodig voor leerlingen met een ongunstige startpositie en voor meisjes.



het Instituut voor  
Onderzoek van het Onderwijs

Sweelinckplein 14, 2517 GK 's-Gravenhage  
Telefoon: 070 - 469679

---