

Samenvatting

Sinds 1997 is de Nederlandse overheid begonnen met het verkleinen van de groepen door aan scholen extra formatie beschikbaar te stellen. Met deze investering zijn grote bedragen gemoeid. De vraag is nu, in hoeverre deze extra formatie tot verhoogde prestaties bij leerlingen leidt, en aansluitend, wanneer dit het geval is, hoe een dergelijk effect dan kan worden verklaard. Dit onderzoek richt zich op de effecten van die extra formatie binnen 46 Nederlandse basisscholen. Binnen deze scholen worden leerlingen van groep 1 getoetst, en worden hun interacties met de leerkracht, en hun taakgerichtheid geobserveerd. De resultaten laten zien dat de leerling-leerkrachtratio een positieve invloed heeft op de taalprestaties van de leerling, in tegenstelling tot de groeps grootte en de leerling-volwasseneratio. Verschillende leerling-leerkrachtinteracties en de taakgerichtheid van leerlingen lijken een - bescheiden - rol te spelen in het verklaren van dit verband.

1 Inleiding

Leerlingen, studenten, ouders en beleidsmakers: bijna iedereen heeft wel een mening over de effecten van kleinere groepen. Leerlingen en studenten verwachten dat zij meer individuele aandacht zullen krijgen, leerkrachten verwachten dat zij in staat zullen zijn kwalitatief betere instructie te geven, ouders verwachten al deze voordelen en meer, en ook beleidsmakers verwachten voordelen van kleine klassen. Deze verwachtingen lijken erop te duiden dat er een duidelijke relatie bestaat tussen groeps grootte en leerprestaties: kleine groepen resulteren in hogere prestaties. De vraag is of wetenschappelijk onderzoek deze verwachtingen ondersteunt. Dit onderzoek richt zich op de effecten van kleinere groepen op de taalprestaties en verschillende onderwijsprocessen in het Nederlandse basisonderwijs.

De relatie tussen groeps grootte en prestaties is al lang - en nog steeds - een geliefd onderzoeksthema. Bekende voorbeelden zijn de meta-analyses van Glass en Smith (1979) en de meer recente van Hedges (2000). Het implementeren en evalueren van programma's om de groepen te verkleinen, gebeurt regelmatig. Voorbeelden hiervan zijn het STAR-onderzoek (Student Teacher Achievement Ratio) in Tennessee (Boyd-Zaharias & Bain, 2000; Finn & Achilles, 1999; Krueger & Whitmore, 1999), het Class Size Reduction-project (CSR) in Californië (Bohrnstedt & Stecher, 1999; Stecher & Bohrnstedt, 2000, 2002) en Wisconsin's Student Achievement Guarantee in Education (SAGE) (Molnar et al., 1999, 2000). De meeste van deze projecten hebben laten zien dat kleine groepen positieve, en soms blijvende effecten hebben op de leerprestaties, met name voor etnische minderheden en kinderen uit de lagere sociaal-economische groepen. Jonge kinderen in de onderbouw van het basisonderwijs lijken de grootste voordelen te hebben van een kleinere groep.

Hoewel de uitkomsten van deze programma's de conclusie dat groepsverkleining leidt tot hogere prestaties lijken te ondersteunen, vond Hanushek (1999) dat niet-experimentele studies deze conclusie niet eenduidig onderschrijven. Slechts 15% van de door hem bekeken studies vond een dergelijk positief effect. Hedges (2000) vergeleek drie verschillende onderzoeksopzetten: kleinschalig aselect onderzoek, grootschalig economisch onderzoek en het grootschalige STAR-experiment. Hij stelde vast dat wanneer een groep van 24 leerlingen wordt vergeleken met een groep van 15 leerlingen, de effectgroottes van al deze drie typen onderzoek liggen tussen de 0.13 en 0.18 standaarddeviatie in het voordeel van de kleine groep van 15. Hedges concludeert dan ook dat, hoewel er studies zijn die geen positief effect van kleine groepen vinden op de leerprestaties, de algemene conclusie is dat er wel een positief effect is.

Wanneer overeenstemming bestaat over

de voordelige effecten van kleinere groepen op de leerprestaties, dan zal onderzoek die factoren moeten onderzoeken die dit effect zouden kunnen verklaren. Een dergelijk onderzoek zal zich moeten richten op het groepsniveau, waar de instructie daadwerkelijk plaatsvindt.

De algemene veronderstelling is dat kleinere groepen, leerkrachten in staat stellen hun instructie beter af te stemmen op de behoeften van de individuele leerlingen. Omdat de groepen kleiner zijn, hebben leerkrachten simpelweg meer tijd en mogelijkheden om individuele leerlingen de aandacht te geven die zij nodig hebben (Betts & Shkolnik, 1999; Blatchford, Baines, Kutnick, & Martin, 2001; Blatchford & Mortimore, 1994; Bourke, 1986; Cahen, Filby, McCutcheon, & Kyle, 1983; Glass, Cahen, Smith, & Filby, 1982; Molnar et al., 1999, 2000; Smith & Glass, 1980). Deze simpele verklaring is gebaseerd op het concept *adaptief onderwijs*, dat refereert aan instructie die is aangepast aan de individuele behoeften en mogelijkheden van leerlingen. In het kader van adaptief onderwijs ontvangen sommige leerlingen meer tijd, instructie en/of hulp van de leerkracht (of van anderen, bijvoorbeeld van een remedial teacher) dan andere.

In kleinere groepen hebben leerkrachten meer tijd om met hun leerlingen te interacteren, met als resultaat dat ook de inhoud van deze interacties zal veranderen (Doolaard, Annevelink, Bosker, & Creemers-Van Wees, 2000). Leerkrachten kunnen meer tijd besteden aan de inhoud van het onderwijs (de taken die leerlingen moeten vervullen) en zijn minder tijd kwijt aan klassenmanagement. Wanneer leerkrachten in staat zijn individuele leerlingen meer aandacht te geven, zullen meer interacties plaatsvinden tussen de leerkracht en de leerlingen, en zal de mate waarin leerlingen taakgericht zijn ook worden verhoogd. Dit alles met als resultaat dat de leerprestaties zullen stijgen. In dit kader dient echter een opmerking te worden gemaakt. Het is mogelijk dat leerkrachten hun manier van onderwijs niet aanpassen aan de groeps grootte en dat zij daarmee de voordelen van een kleinere groep niet benutten. Onderzoek heeft uitgewezen dat dergelijke situaties inderdaad kunnen voorkomen.

Blatchford en Mortimore (1994) komen op basis van eerder onderzoek tot de conclusie dat onderzoek weliswaar de suggestie wekt dat zowel leerlingen als leerkrachten voordelen ondervinden van kleinere groepen, maar dat het onderzoek incompleet en inconsistent is. Het is volgens hen niet noodzakelijk dat eventuele voordelige effecten van kleinere groepen ook daadwerkelijk plaatsvinden. Eerder onderzoek wijst namelijk ook uit dat het voorkomt dat groepsverkleiningen géén effect hebben op het gedrag van de leerkracht. Shapson, Wright, Eason en Fitzgerald (1980) hebben onderzoek gedaan naar de invloed van vier verschillende groeps groottes op zowel leerkrachtverwachtingen, prestaties als processen in de groep. De resultaten lieten zien dat leerkrachten bepaalde verwachtingen hebben van kleinere groepen, maar dat de meeste van deze verwachtingen niet werden bevestigd door de resultaten van het onderzoek. Zij concluderen dan ook dat leerkrachten in het algemeen de voordelen van kleine groepen niet benutten door het onderwijs meer te individualiseren.

2 De Nederlandse context

De discussie over groeps grootte en de kwaliteit van het Nederlandse basisonderwijs kreeg in 1994 een impuls met de publicatie van de resultaten van een evaluatie van het basisonderwijs (CEB, 1994). De conclusie van dit onderzoek luidde dat leerkrachten, met name in de onderbouw, onvoldoende in staat waren de instructie aan te passen aan de behoeften en mogelijkheden van hun leerlingen. Schoolleiders en leerkrachten gaven aan dat de groeps grootte hierin een cruciale factor was. Een onderzoek uitgevoerd door de Onderwijsinspectie (1995) liet zien dat de groeps grootte in het Nederlandse basisonderwijs inderdaad groter was dan in vergelijkbare landen en dat deze varieerde tussen en binnen scholen. De tussentijdse instroom van vierjarigen in groep 1 maakte het verschil met name aan het einde van het schooljaar duidelijk. Naar aanleiding van deze uitkomsten stelde de overheid in 1996 een commissie samen die de minister van Onderwijs moest adviseren rondom de groeps grootte

(CKAG, 1996). Deze commissie concludeerde dat groeps-grootte inderdaad een probleem is binnen de organisatie van het basisonderwijs: grote groepen hebben negatieve consequenties voor de onderwijskundige kwaliteit en de leerprestaties. Het advies van de commissie was dan ook de groeps-grootte in de onderbouw van het basisonderwijs te verkleinen en daar een maximale groeps-grootte van 25 leerlingen in te stellen.

Omdat het instellen van een maximale groeps-grootte niet samengaat met de autonomie die Nederlandse scholen genieten, besloot het ministerie geen maximale groeps-grootte in te stellen, maar wel de groepen in de onderbouw te verkleinen. De eerste stap op weg naar groepsverkleining werd gezet in 1997, toen een extra budget van € 120 miljoen per jaar beschikbaar werd gesteld voor extra formatie. Scholen hebben grofweg drie opties om deze extra formatie te gebruiken (PMPO, 1999):

- 1 daadwerkelijk verkleinen van de groeps-grootte door nieuwe groepen te creëren met hun eigen groepsleerkracht;
- 2 extra personeel toewijzen aan bestaande groepen (bijvoorbeeld extra leerkrachten of onderwijsassistenten);
- 3 het creëren van nieuwe takenpakketten en eventuele nieuwe functies die groeps-overstijgend van aard zijn, bijvoorbeeld een onderbouwcoördinator.

De tweede, derde en vierde stap van het groepsverkleiningsbeleid werden geïntroduceerd in augustus 2000, 2001 en 2002, met als doel de groeps-grootte in de onderbouw te verkleinen naar één leerkracht voor elke 20 leerlingen in schooljaar 2002/2003. Bovendien zullen scholen met veel achterstandsleerlingen, vanwege de gewichtenregeling in staat zijn de groepen nog kleiner te maken. Met ingang van het schooljaar 2003/2004 zijn scholen vrij om te bepalen waar zij de extra formatie willen inzetten. Omdat de data voor het onderhavige onderzoek zijn verzameld in schooljaar 2000/2001 tot en met 2002/2003, heeft dit geen effect op dit onderzoek.

In het kader van de extra formatie is het inzetten van bevoegde leerkrachten wellicht effectiever dan het inzetten van anderen (bijvoorbeeld onderwijsassistenten of ouders).

Differentiële effecten van verschillende manieren waarop de formatie wordt ingezet, kunnen worden besproken in termen van de daadwerkelijke groeps-grootte, de leerling-leerkrachtratio (LLR), en leerling-volwas-seneratio (LVR). Bijvoorbeeld: wanneer een extra leerkracht aan een bestaande groep wordt toegevoegd, dan heeft dit geen invloed op de groeps-grootte, maar wel op de LLR en LVR, die daardoor dalen. Wanneer echter een extra onderwijsassistent aan een groep wordt toegevoegd, daalt de LVR, maar heeft dit geen invloed op de groeps-grootte en de LLR, omdat onderwijsassistenten niet bevoegd zijn zelfstandig voor een groep te staan. Kortom: het toevoegen van extra personeel aan bestaande groepen is niet hetzelfde als het verkleinen van de groeps-grootte. Dit onderzoek richt zich dan ook op zowel de effecten van de groeps-grootte als op de effecten van de LLR en de LVR. Daarnaast richt het zich ook op de vraag of het zinvol is een onderscheid tussen deze drie indicatoren te maken.

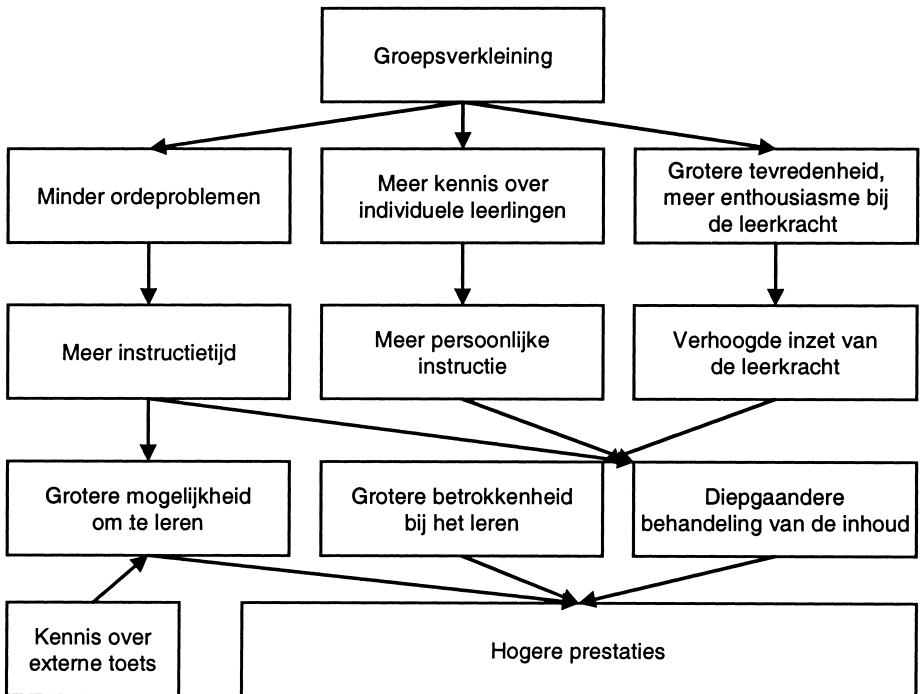
3 Theoretische achtergrond

3.1 Het model van Anderson

Anderson (2000) richtte zich op de vraag *waarom* groepsverkleiningen zouden leiden tot verhoogde leerprestaties. Om het verband tussen groeps-grootte en prestaties te kunnen verklaren, ontwikkelde Anderson een causaal model dat een mogelijke verklaring biedt voor het positieve effect van een kleinere groep op de prestaties van de leerlingen. Het model start met groepsverkleining en eindigt met verhoogde prestaties (zie Figuur 1).

Het model voorspelt dat kleinere groepen een direct effect zullen hebben op de volgende drie variabelen: ordeproblemen, kennis over de leerling, en tevredenheid en enthousiasme van de leerkracht. Elk van deze variabelen begint vervolgens een eigen traject:

- Minder ordeproblemen resulteren in meer instructietijd, die in combinatie met meer kennis over de externe toets (meer kennis over dié curriculumstandaarden die in gestandaardiseerde toetsen zullen worden getoetst, zodat leerkrachten beter in staat zijn om leerlingen datgene bij te brengen



Figuur 1. Een mogelijke verklaring van de invloed van groeps grootte op prestaties (Anderson, 2000).

dat zij nodig hebben om op deze toetsen goed te presteren) resulteert in een grotere mogelijkheid tot leren. In combinatie met meer persoonlijke instructie en een verhoogde inzet van leerkrachten, leidt meer instructietijd ook tot een verhoogde betrokkenheid bij leerlingen, en een meer diepgaande behandeling van de inhoud van het onderwijs.

- Meer kennis over de leerling leidt tot meer persoonlijke instructie, die in combinatie met meer instructietijd en een verhoogde inzet van leerkrachten leidt tot een verhoogde betrokkenheid bij leerlingen, en een meer diepgaande behandeling van de inhoud van het onderwijs.
- Grotere tevredenheid en meer enthousiasme bij de leerkracht resulteren in een verhoogde inzet van de leerkracht, die in combinatie met meer instructietijd en meer persoonlijke instructie leidt tot meer betrokkenheid bij leerlingen, en een meer diepgaande behandeling van de inhoud van het onderwijs.
- Uiteindelijk zijn verhoogde leerprestaties het resultaat van de combinatie van drie variabelen: grotere mogelijkheid om te

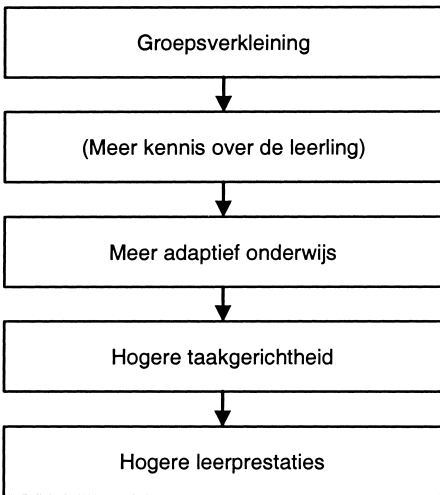
leren, grotere betrokkenheid bij leerlingen, en een meer diepgaande behandeling van de inhoud van het onderwijs.

3.2 Het model van Anderson als startpunt van dit onderzoek

Omdat het betrekken van alle variabelen uit het model van Anderson de omvang en de praktische mogelijkheden van elke studie te boven zou gaan, richt dit onderzoek zich op het middelste gedeelte van het model: het gedeelte van meer kennis over de leerling via meer persoonlijke instructie en leerlingbetrokkenheid naar verhoogde leerprestaties. Deze keuze is gebaseerd op de uitkomsten van eerder onderzoek en de wens om de focus bij de leerling te leggen. Wanneer het gaat om het verklaren van het verband tussen groeps grootte en prestaties, ligt de focus dus bij de mate waarin de leerkracht in staat is leerlingen instructie te bieden die is afgestemd op hun individuele kenmerken. Persoonlijke instructie is een aspect van het bredere concept *adaptief onderwijs*. Leerkrachten die meer persoonlijke instructie willen geven, zullen hun activiteiten moeten afstemmen op de behoeften van hun individuele leerlingen.

Instructie bestaat uit verschillende interacties tussen de leerkracht en zijn leerlingen. Om de instructie af te stemmen op de behoeften van bijvoorbeeld de zwakke leerling, zal de leerkracht meer taakgerichte interacties met deze leerling moeten aangaan om deze leerling zo een uitgebreide uitleg te geven over hetgeen de leerkracht van de leerling verwacht. In dit onderzoek is de betrokkenheid van de leerling gemeten als de *taakgerichtheid van de leerling*. Leerlingen die veel tijd aan hun taak besteden, zullen naar verwachting beter presteren.

Samenvattend zijn de sleutelvariabelen van deze studie: adaptief onderwijs, leerling-leerkrachtinteracties en taakgerichtheid. Dit onderzoek richt zich op de taalprestaties. De variabele *meer kennis van de leerling* wordt in dit onderzoek niet specifiek gemeten. Er wordt eenvoudigweg aangenomen dat leerkrachten hun leerlingen in kleinere groepen beter kennen dan in grotere groepen. In Figuur 2 is het model van Anderson weergegeven zoals het in deze studie zal worden onderzocht.



Figuur 2. Model van Anderson (Anderson, 2000), aangepast aan dit onderzoek.

4 Probleemdefinitie

Met betrekking tot de relatie tussen groeps-grootte, LLR en LVR aan de ene kant en taalprestaties aan de andere kant, zullen de volgende onderzoeksvragen worden beantwoord:

- Zijn de taalprestaties van leerlingen in kleinere groepen (of in groepen met een lagere LLR of LVR) hoger dan van leerlingen in grotere groepen (of in groepen met een hogere LLR of LVR)?
- In hoeverre is het analytisch gezien zinvol een onderscheid te maken tussen de maten groeps-grootte, LLR en LVR? (leiden de verschillende maten tot dezelfde of verschillende effecten?)

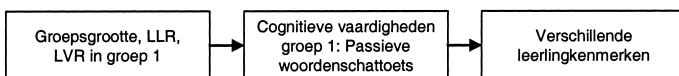
Deze vragen zullen worden beantwoord voor leerlingen die gedurende schooljaar 2000/2001 in groep 1 van het basisonderwijs zaten. Het onderzoekskader dat in deze studie wordt gebruikt, is weergegeven in Figuur 3.

Gebaseerd op eerder onderzoek is de verwachting dat de variabelen groeps-grootte, LLR en LVR de taalprestaties van leerlingen zo zullen beïnvloeden dat hoe kleiner of lager de groeps-grootte, LLR of LVR, des te hoger de taalprestaties van de leerlingen zullen zijn.

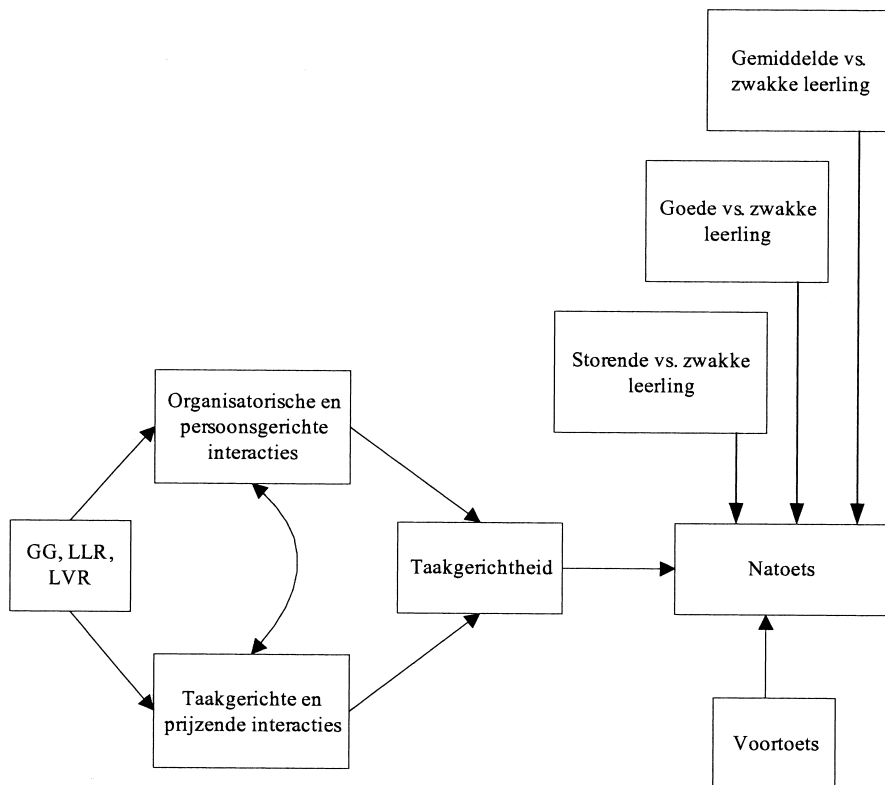
Wanneer een significante, negatieve relatie wordt gevonden tussen groeps-grootte, LLR of LVR en de prestaties in groep 1, zal de volgende onderzoeksvraag worden beantwoord:

- Kunnen de voordelige effecten van kleinere groepen (of groepen met een lagere LLR of LVR) op de prestaties worden verklaard door veranderingen in de hoeveelheid leerling-leerkrachtinteracties en vervolgens door een verhoogde taakgerichtheid?

Het conceptuele model dat zal dienen als startpunt voor de analyses om de hierboven geformuleerde vraag te beantwoorden, is weergegeven in Figuur 4. Dit model is gebaseerd op de literatuur en eerder onderzoek naar de relatie tussen een verkleining van de



Figuur 3. Onderzoekskader¹.



Figuur 4. Het conceptuele, causale model².

groeps-grootte (GG) (of een verlaging van de LLR of LVR) en leerprestaties.

Het algemene idee achter dit conceptuele, causale model is dat wanneer de groeps-grootte, LLR of LVR wordt verlaagd, het aantal interacties tussen de leerkracht en de leerling zal veranderen. Er zullen minder organisatorische en persoonsgerichte interacties zijn, omdat leerkrachten minder tijd kwijt zullen zijn met klassenmanagement.³ Daarnaast zullen er meer taakgerichte en prijzende interacties zijn, omdat leerkrachten meer tijd zullen hebben om zich met het leren, en daarmee met de instructie van hun leerlingen bezig te houden. Deze veranderingen in de hoeveelheid interacties zullen leiden tot een situatie waarin leerlingen taakgerichter zijn. Deze verhoogde taakgerichtheid zal vervolgens leiden tot verhoogde prestaties.

Verder zullen de prestaties van de leerlingen op de natoets afhankelijk zijn van de prestaties op de voortoets. De hoeveelheid organisatorische en persoonsgerichte interac-

ties zal positief correleren met de hoeveelheid taakgerichte en prijzende interacties. Dit vanuit het idee dat leerkrachten van elkaar verschillen aangaande de hoeveelheid interacties die zij in het algemeen met hun leerlingen hebben; sommige leerkrachten interacteren nu eenmaal meer met hun leerlingen dan andere leerkrachten. Wanneer leerkrachten veel organisatorische en persoonsgerichte interacties aangaan met hun leerlingen, dan zullen zij ook veel taakgerichte en prijzende interacties met hun leerlingen hebben. Met betrekking tot leerlingkenmerken zal de nadruk liggen bij het type leerling. De verwachting is dat gemiddelde en goede leerlingen beter zullen presteren dan zwakke leerlingen.

Methode

5.1 Steekproef en representativiteit

In 1999/2000 is een eerste onderzoek uitgevoerd om de implementatie van het over-

Tabel 1

Schoolgrootte en samenstelling van de leerlingpopulatie van de scholen in de steekproef (S) in vergelijking met de populatie Nederlandse basisscholen (P)

Schoolgrootte (aantal leerlingen)	Samenstelling van de leerlingpopulatie (gemiddeld leerlinggewicht)							
	1.0-1.1		1.1-1.2		>1.2		Totaal	
	S% (n)	P(%)	S% (n)	P(%)	S% (n)	P(%)	S% (n)	P(%)
Klein (1-150)	13.0 (6)	19.9	32.6 (15)	8.5	.0(0)	3.5	45.6 (21)	31.9
Gemiddeld (151-250)	10.9 (5)	18.4	15.2 (7)	9.4	4.3 (2)	8.4	30.4(14)	36.1
Groot (> 250)	19.6 (9)	15.4	2.2 (1)	8.4	2.2 (1)	8.2	23.9 (11)	32.0
Totaal	43.5 (20)	53.7	50.0 (23)	26.2	6.5 (3)	20.1	100 (46)	100

heidsbeleid rond groepsgrootte te evalueren. Van de 185 deelnemende scholen hadden er 118 een groep 1 die, in vergelijking met groep 2, 3 en 4 binnen de betreffende school, relatief groot of relatief klein was. Deze scholen zijn benaderd voor het onderhavige onderzoek. Van de 118 benaderde scholen wilden er 46 (39%) deelnemen. In Tabel 1 is een overzicht gegeven van de representativiteit van de steekproef wat betreft schoolgrootte, en de samenstelling van de leerlingpopulatie uitgedrukt in het gemiddelde leerlinggewicht.

Schoolgrootte was verdeeld in drie categorieën: *klein* (minder dan 151 leerlingen), *gemiddeld* (tussen de 151 en 250 leerlingen), en *groot* (meer dan 250 leerlingen). De samenstelling van de leerlingpopulatie was verdeeld in drie categorieën op basis van het gemiddelde leerlinggewicht. De laagste categorie wordt gevormd door de scholen met weinig of geen allochtone leerlingen en autochtone leerlingen met ouders met een lage opleiding (een gemiddeld leerlinggewicht van 1.0 tot 1.1). De hoogste categorie bestaat uit de scholen waar het grootste deel van de leerlingen van een dergelijke afkomst is (een gemiddeld leerlinggewicht boven de 1.2). De steekproef bleek representatief met betrekking tot schoolgrootte ($\chi^2 = 4.08$, $df = 5$, $p = .13$), maar niet met betrekking tot de samenstelling van de leerlingpopulatie ($\chi^2 = 15.06$, $df = 5$, $p = .00$). Scholen met een hoog gemiddeld leerlinggewicht zijn ondervertegenwoordigd in de steekproef. Met dit gegeven zal rekening gehouden moeten worden bij het trekken van conclusies op basis van dit on-

derzoek. Het onderzoek is uitgevoerd in het schooljaar 2000/2001. Er deden 829 leerlingen van groep 1 aan mee, verdeeld over 73 groepen.

5.2 Variabelen en instrumenten

Leerlingenkenmerken

In het begin van het schooljaar 2000/2001 is de leerkrachten gevraagd de volgende achtergrondgegevens over de leerlingen te verstrekken: geslacht, sociaal-economische status (SES) uitgedrukt in het leerlinggewicht, thuistaal (Nederlands, niet Nederlands of dialect), ingeschatte intelligentie (laag, gemiddeld of hoog) en de geboortedatum.

De leerkrachten is gevraagd de intelligentie van de leerlingen in te schatten op basis van de volgende instructie:

Het gaat om een inschatting van het intelligentieniveau ten opzichte van het intelligentieniveau van alle leerlingen in Nederland, ervan uitgaande dat 50% van alle leerlingen in Nederland een gemiddeld intelligentieniveau heeft, 25% een laag intelligentieniveau en 25% een hoog intelligentieniveau. Het gaat er dus niet om dat u een oordeel geeft over de intelligentie van de leerlingen ten opzichte van de andere leerlingen in de klas. Zo kan het bijvoorbeeld zo zijn dat u alle leerlingen een hoog of gemiddeld intelligentieoordeel geeft.

Er zijn verschillende onderzoeken uitgevoerd naar de betrouwbaarheid van inschattingen

van leerkrachten van de intelligentie van leerlingen, waarbij de correlatie wordt bepaald tussen de inschatting van de leerkracht en de intelligentie zoals gemeten met een gestandaardiseerde IQ-test (Biesheuvel & Flim, 2001; Follman, 1991; Tellegen, Winkel, Wijnberg-Williams, & Laros, 1998; Wild, 1993) Al deze onderzoeken laten zien dat leerkrachten redelijk in staat zijn om een inschatting van de intelligentie van leerlingen te maken, maar dat het niet voldoende is om de ingeschatte intelligentie te beschouwen als een vervanging van intelligentie zoals gemeten met een gestandaardiseerde IQ-test. Daarom wordt voor deze variabele in dit onderzoek de term *ingeschatte intelligentie* gebruikt.

Taalprestaties

In dit onderzoek wordt als effectmaat de taalprestaties van leerlingen gebruikt. De keus voor de gebruikte toets is gemaakt op basis van drie interviews met experts op het terrein van taalonderwijs. Er is gekozen voor het taalonderdeel "woordenschat". (Passieve) woordenschat is een goede voorspeller voor een succesvol verloop van het leren lezen in groep 3. Na een afweging van voor- en nadelen van beschikbare toetsen is gekozen voor een toets voor passieve woordenschat van de Citogroep. Deze is het eerste deel van de toets "Taal voor Kleuters" voor jongste kleuters (Van Kuyk, 1996). Deze toets wordt afgenomen in januari (voortoets) en juni (natoets) van groep 1 en bestaat uit 32 opgaven.

Een voordeel van deze toets is dat veel scholen de toets al gebruiken. Verder kost het niet veel tijd om de toets af te nemen en kan de toets bij een groep leerlingen tegelijk afgenomen worden. De toets kan afgenomen worden door de school of door een getrainde toetsleider van de universiteit. Na de toetsafname is een aantal gegevens geregistreerd ten aanzien van de afname, zoals de datum, het aantal leerlingen dat tegelijk getoetst is, wie de toetsleider was en of er speciale omstandigheden zijn geweest die de uitkomst zouden kunnen beïnvloeden.

De kwaliteit van toetsen en testen wordt beoordeeld door de COTAN (Evers et al., 2002) op basis van de volgende criteria: constructie van de toets, kwaliteit van het toets-

materiaal, kwaliteit van de handleiding, normen, betrouwbaarheid, constructvaliditeit en criteriumvaliditeit. Zes van de zeven criteria worden als voldoende beoordeeld. Over de criteriumvaliditeit is geen oordeel gegeven, omdat de toets niet bedoeld is om voorspellingen te doen en op dit terrein daarom geen onderzoek is gedaan. In situaties waarin toch voorspellingen worden gedaan op basis van deze toets, moet de criteriumvaliditeit beschouwd worden als "onvoldoende".

Groepsgrootte, leerling-volwasseneratio en leerling-leerkrachtratio

Groepsgrootte is gedefinieerd als het aantal leerlingen in de groep waarvoor een leerkracht elke dag verantwoordelijk is. Voor dit onderzoek is deze maat niet afdoende, omdat er situaties voorkomen waarin meerdere leerkrachten tegelijk in een groep werken of groepen voor een bepaalde periode worden gesplitst. De leerling-volwasseneratio (LVR) wordt berekend door het aantal leerlingen te delen door het aantal aanwezige volwassenen. De LVR wordt kleiner als het aantal volwassenen (bijvoorbeeld leerkrachten, assistenten, ouders) toeneemt. Een groep van 24 leerlingen met een leerkracht heeft een LVR van 24, een groep van 30 leerlingen, een leerkracht en een assistent heeft een LVR van 15. Naast de LVR wordt ook gebruikgemaakt van de leerling-leerkrachtratio (LLR). In dit geval worden alleen die volwassenen meegeteld die (bijna) gediplomeerd zijn op het niveau van een leerkracht zoals een groepsleerkracht, een intern begeleider, een remedial teacher, een NT2-, een OALT-leerkracht of een lio (leraar in opleiding).

Om de benodigde informatie voor het berekenen van de ratio's te verzamelen, is de leerkrachten gevraagd aan te geven wie betrokken is bij het onderwijs aan de groep op dinsdag en donderdag. Deze dagen zijn daartoe verdeeld in 15 periodes van een half uur. De leerkracht had daarbij de keus tussen een extra groepsleerkracht, een onderwijsassistent, een lio, NT2-leerkracht, een OALT-leerkracht, een intern begeleider, een ouder of een ander. Eerst zijn per half uur de ratio's berekend, vervolgens zijn door de gemiddelden te berekenen van de ratio's van alle periodes op een dag *per dag* de ratio's berekend

(periodes tussen-de-middag zijn buiten beschouwing gelaten). Dit is ook afzonderlijk voor alleen de ochtenden gedaan. Uiteindelijk zijn in de analyses alleen deze ochtendratio's (LVR- en LLR-gemiddelde over dinsdag- en donderdagochtend) gebruikt, omdat taalonderwijs meestal in de ochtend plaatsvindt en de variatie in de ochtendratio's groter was dan in de ratio's gemeten over de hele dag.

Interacties en taakgerichte leertijd

De observaties zijn uitgevoerd in februari en maart 2001 door getrainde observatoren. De gebruikte instrumenten zijn gebaseerd op instrumenten die gebruikt zijn voor de evaluatie van WSNS (Houtveen, Pijl, Pijl, Reezigt, & Vermeulen, 1998) en longitudinaal onderzoek in de laagste groepen van het basisonderwijs (Reezigt, Guldmond, & Ros, 1995). Doolaard e.a. (2000) hebben deze instrumenten aangepast om ze te gebruiken in eerder onderzoek naar de effecten van de groeps grootte. In het onderhavige onderzoek zijn de instrumenten op dezelfde manier gebruikt.

In elke deelnemende groep is gedurende 30 minuten een taalactiviteit geobserveerd, waarbij gekeken is naar taakgerichtheid, de tijdsbesteding, de activiteit (inhoud) en de interacties tussen de leerlingen en de leerkracht. Taakgerichtheid, tijdsbesteding en activiteit zijn elke drie minuten gescoord (in elke observatie van een half uur zijn deze variabelen daarmee 11 maal gescoord, beginnend op tijdstip 0). Zowel de tijdsbesteding als de activiteit zullen in dit artikel niet aan de orde komen en derhalve niet verder worden besproken (voor meer informatie zie Annevelink, 2004).

Taakgerichtheid kende vier categorieën: *taakgericht*, *niet taakgericht*, *uit de klas* en *wachtend op de leerkracht*. Per leerling is uitgerekend in hoeveel procent van de geobserveerde momenten de leerling taakgericht was. De interacties zijn continu gescoord. Er is onderscheid gemaakt tussen interacties geïnitieerd door de leerkracht en interacties geïnitieerd door de leerling. Vervolgens is een onderscheid gemaakt tussen taakgerichte interacties, interacties gericht op de organisatie en persoonsgerichte interacties. Een door

de leerkracht geïnitieerde interactie kon ook nog prijzend van aard zijn. Omdat niet elke observatie precies 30 minuten duurde, is voor het aantal interacties een correctie toegepast. Om de uitkomsten vergelijkbaar te maken, is het aantal gescoorde interacties gedeeld door het aantal geobserveerde meetmomenten en vermenigvuldigd met 11.

Tijdens de observaties is gekeken naar leerlingen. Vooraf zijn door de leerkracht vier leerlingen geselecteerd op basis van hun cognitief en affectief functioneren. Deze leerkrachten moesten een zwakke, gemiddelde en goede leerling selecteren en daarnaast een leerling met storend gedrag. Hiervoor ontvingen zij de volgende aanwijzing (gebaseerd op Werkman & Faber, 2000):

Een leerling vertoont storend gedrag als hij/zij tenminste twee van de onderstaande gedragingen vertoont: loopt vaak weg van het werk/de tafel, roept of schreeuwt door de klas, roept vaak andere kinderen, zoekt vaak ruzie, bedreigt of daagt uit, zit aan andermans spullen.

Vooraf werd de leerkrachten niet verteld waar de observator tijdens de observatie op ging letten. De leerkracht wist daarmee niet dat de interacties en de taakgerichtheid de focus van de observatie waren.

Omdat de instrumenten al eerder gebruikt zijn, was informatie beschikbaar over de betrouwbaarheid ervan (Doolaard et al., 2000; Houtveen et al., 1998; Reezigt et al., 1995). De interbeoordelaarsovereenstemming van de meting van interacties wordt als "voldoende" beoordeeld. Consensus over de inhoud van de door de leerkracht geïnitieerde interacties is "redelijk" ($\kappa = .57$), en consensus over de inhoud van de door leerlingen geïnitieerde interacties is "aanzienlijk" ($\kappa = .67$). Consensus over wie de initiator is van een interactie is bijna perfect ($\kappa = .91$). Ook de interbeoordelaarsovereenstemming van het instrument voor taakgerichtheid is voldoende.

5.3 Data-analyse

Vooraf is met behulp van een covariantie-analyse vastgesteld welke covariabelen samenhangen met de afhankelijke variabelen,

zodat hiervoor gecontroleerd kon worden. Omdat de gegevens een geneste structuur hebben (leerlingen zitten in groepen en groepen bevinden zich in scholen) zijn vervolgens meerniveau-analyses uitgevoerd met behulp van het MLwiN-softwarepakket. De afhankelijke variabele (*taalprestaties*) is genormaliseerd, evenals de taalprestaties op de voor-toets. Een normaal verdeelde afhankelijke variabele is noodzakelijk in een meerniveau-analyse.

Overeenkomstig het Amerikaanse STAR-onderzoek is de effectgrootte berekend voor de situatie waarin een groep wordt verkleind van 24 naar 15 leerlingen (zie bijv. Boyd-Zaharias & Bain, 2000). Hiervoor is de volgende formule gebruikt: (groepsgrootte 24-15) * (regressiecoëfficiënt van groepsgrootte, LVR of LLR / $\sqrt{\text{variantie tussen leerlingen}}$). Anders geformuleerd: $(9 * B) / \sigma$.

De relaties tussen groepsgrootte (c.q. LVR of LLR), de interacties, taakgerichtheid en taalprestaties zijn geanalyseerd met behulp van een padanalyse. Met deze analyse kunnen de best passende relaties tussen meerdere variabelen tegelijk worden vastgesteld. Het conceptuele model wordt hierbij als uitgangspunt genomen. Er is gebruikgemaakt van het Mplus-softwarepakket.

6 Resultaten

6.1 Enkele beschrijvende gegevens

In Tabel 2 wordt een aantal beschrijvende gegevens gepresenteerd over de groepsgrootte, LLR en LVR zoals deze in dit onderzoek zijn gemeten.

Zoals te zien is, was de variatie in groepsgrootte, LLR en LVR groot. Er waren erg kleine groepen met minder dan 10 leerlingen, maar ook erg grote groepen met meer dan 30

Tabel 2
Beschrijvende gegevens met betrekking tot groepsgrootte (GG), LLR, LVR in groep 1

Groep 1	Min.	Max.	M	SD
GG	9.0	38.0	25.1	5.3
LLR	7.4	35.7	22.7	5.5
LVR	7.4	34.0	20.0	6.1

en soms zelfs meer dan 35 leerlingen. Hetzelfde geldt voor de LLR en de LVR die een range van 5 tot 30 leerlingen per leerkracht of volwassene besloegen.

Groepsgrootte, LLR en LVR verschillen daadwerkelijk van elkaar. Zo is de groepsgrootte altijd gelijk aan of groter dan de LLR en is de LLR altijd gelijk aan of groter dan de LVR. Correlaties tussen de drie maten leiden ook tot de conclusie dat het gaat om verschillende maten (zie Tabel 3). Wanneer groepsgrootte, LLR en LVR één en hetzelfde zouden meten, dan zouden de correlaties tussen de maten onderling gelijk moeten zijn aan 1.0. Dit is echter niet het geval; de correlaties bevinden zich tussen de 0.55 en 0.82.

Tabel 3
Correlaties tussen groepsgrootte (GG), LLR en LVR in groep 1

	LLR	LVR
GG	.82*	.55*
LLR		.76*
LVR		

*Significant op niveau 0.01.

6.2 Groepsgrootte en leerlingprestaties in groep 1

De eerste vraag waar we ons hier op richten, is of kleinere groepen (of groepen met een lagere LLR of LVR) samengaan met een grotere taalvaardigheid. In Tabel 4 worden de uitkomsten van analyses van vijf modellen gepresenteerd. De modellen GG0 en R0 zijn de basismodellen, waarbij enkel de significante effecten van de covariaten geslacht, SES, ingeschat IQ en de voortoets worden gemodelleerd. In model GG1 wordt groepsgrootte toegevoegd als onafhankelijke variabele, zoals in model LLR de leerling-leerkrachtratio wordt toegevoegd aan model R0, en in model LVR de leerling-volwasseneratio.

Alleen het model LLR bleek significant beter bij de data te passen dan het bijbehorende basismodel. De samenhang tussen LLR en taalvaardigheid is daarbij negatief: hoe meer leerlingen per leerkracht, hoe lager de taalvaardigheid van de leerling. De samenhang tussen groepsgrootte en LVR enerzijds en de taalvaardigheid anderzijds is weliswaar in de verwachte richting, maar niet

Tabel 4

Effecten van groepsmaat, LLR en LVR op de prestaties in groep 1

	Model GG0 (N = 709)		Model GG1 (N = 709)		Model R0 (N = 637)		Model LLR (N = 637)		Model LVR (N = 637)	
	Est.	SE	Est.	SE	Est.	SE	Est.	SE	Est.	SE
Vaste effecten										
Intercept	-.37	.11	-.04	.28	-.36	.12	.12	.26	-.12	.22
Meisjes vs. jongens	.15	.05	.16	.05	.13	.06	.13	.06	.13	.06
SES (1.25 vs. 1.0)	-.02	.09	-.02	.09	-.01	.10	-.00	.10	-.00	.10
SES (1.9 vs. 1.0)	-.55	.12	-.56	.12	-.56	.14	-.57	.14	-.57	.14
Gem. vs. laag IQ	.34	.10	.34	.10	.34	.11	.36	.11	.35	.11
Gem./hoog vs. laag IQ	.49	.17	.50	.17	.47	.18	.49	.18	.48	.18
Hoog vs. laag IQ	.55	.13	.56	.13	.56	.14	.58	.14	.57	.14
Voortoets	.50	.03	.50	.03	.48	.04	.48	.04	.48	.04
GG/LLR/LVR (*100)			-1.34	1.04			-2.11	1.02	-1.19	.91
Variatie componenten										
Tussen groepen	.11	.03	.11	.03	.12	.03	.12	.03	.12	.03
Tussen leerlingen	.47	.03	.47	.03	.49	.03	.49	.03	.49	.03
Percentage verklaarde variantie										
Tussen groepen			1.5				2.5		2.6	
Tussen leerlingen			.5				-.1		.8	
Deviantie										
	1558		1556		1424		1419		1422	
Verbetering in model- passing (p)			.21				.04		.20	
Effectgrootte (24 vs. 15)			-.18				-.27		-.15	

significant: een kleinere groep of een lagere LVR gaat niet samen met een grotere woordenschat. Anders gezegd: het realiseren van meer bevoegd personeel in de groep leidt wel tot, en het toevoegen van onbevoegd personeel of het verkleinen van de groep gaat niet gepaard met een gunstigere ontwikkeling van de taalvaardigheid van de leerling in groep 1. De geschatte effectgrootte voor een verlaging van de LLR van 24 naar 15 is -0.27: een verlaging van de leerling-leerkrachtratio van 24 op 1 naar 15 op 1 gaat samen met een toename van 0.27 standaarddeviatie in de taalvaardigheidsscore (dit correspondeert met één antwoord meer correct op de 32 in de toets gestelde vragen).

In Figuur 5 wordt een grafische weergave van dit verband gepresenteerd.

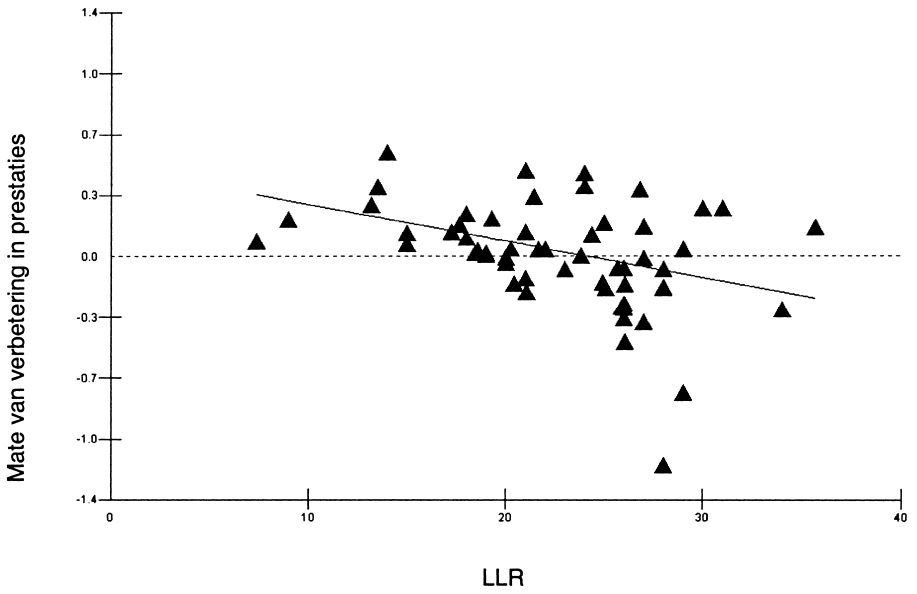
Zoals uit de grafiek blijkt, is er een aantal klassen met een extreem hoge of lage LLR of met een extreem hoge of lage taal-

vaardigheidsontwikkeling. Om na te gaan of deze groepen een belangrijke rol speelden in de analyseuitkomsten, zijn alle analyses nog eens uitgevoerd, maar nu zonder deze extreme gevallen. De resultaten veranderden amper, en het significante effect van LLR bleef overeind, zodat we concluderen dat de uitkomsten tamelijk robuust zijn.

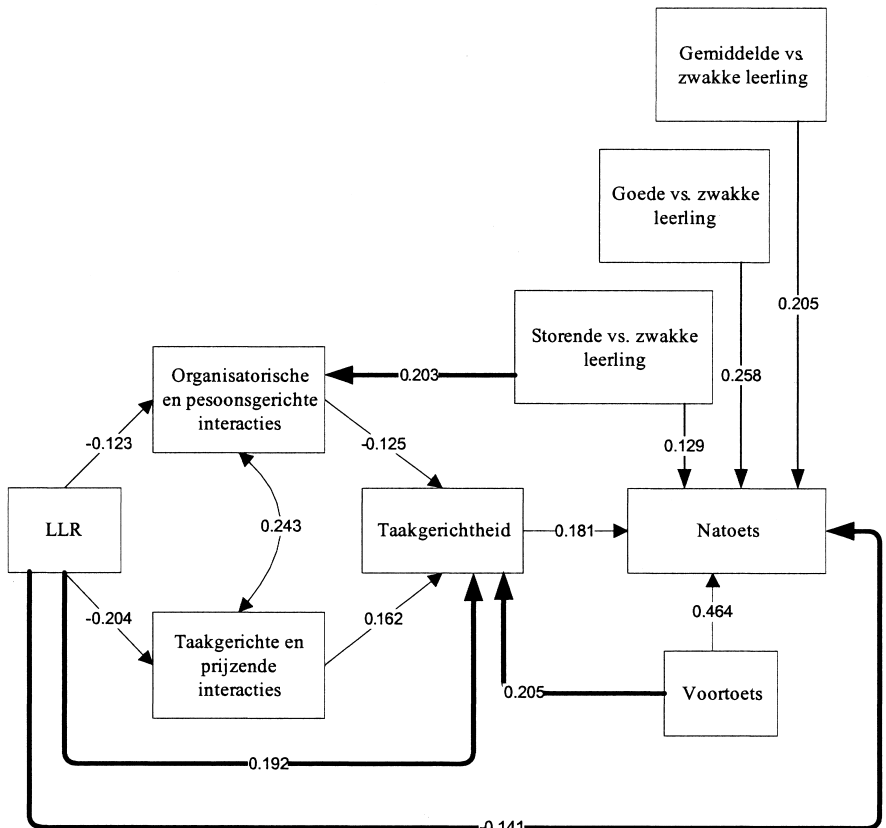
6.3 Groepsmaat: de verbinding tussen onderwijzen en leren

LLR en taalprestaties in groep 1

Om de vraag te kunnen beantwoorden, waarom een lagere LLR samengaat met een toename in taalvaardigheid, is nagegaan of de data passen bij het conceptuele model zoals gepresenteerd in Figuur 4. Dat bleek niet het geval te zijn voor de gehypothetiseerde oorzakelijke samenhang tussen de leerling-leer-



Figuur 5. Relatie tussen LLR en de mate van verbetering in prestaties.



Figuur 6. Uiteindelijk model voor LLR en prestaties in groep 1.

krachtratio en de taalvaardigheid ($\chi^2=37.71$, $p = .00$, $CFI = .86$, $RMSEA = .08$). Daarop is het model enigszins aangepast, totdat een acceptabele passing werd bereikt.⁴ Het uiteindelijke model wordt weergegeven in Figuur 6.⁵ De modelfit is acceptabel ($\chi^2 = 6.29$, $p = .90$, $CFI = 1.00$, $RMSEA = .00$). Alle weergegeven relaties zijn significant ($\alpha \leq .10$).

Uitgaande van het conceptuele model, werd verwacht dat een lagere LLR zou resulteren in een toename in taakgerichte en prijzende interacties, en tevens in een afname van het aantal organisatorische en persoonsgerichte interacties. Tegen deze verwachting in, bleek een lagere LLR tot een toename in alle soorten interacties te leiden. Bovendien bleken deze typen interacties onderling positief samen te hangen. Verder werd verwacht - en dit werd door de gegevens bevestigd - dat een afname van het aantal organisatorische en persoonsgerichte interacties en een toename in taakgerichte en prijzende interacties zouden resulteren in meer taakgerichte leertijd. En, zoals verwacht, bleek een toename in taakgerichtheid samen te gaan met een grotere taalvaardigheid.

In aanvulling op de tot nu toe beschreven relaties, bleek in het uiteindelijke model ook een aantal directe relaties toegestaan te moeten worden, zoals tussen LLR en taakgerichtheid (merkwaardigerwijze: hoe groter de LLR, hoe taakgerichter de leerling) en tussen LLR en taalvaardigheid (conform de verwachte richting: hoe hoger de LLR, hoe lager de taalvaardigheid). Twee andere nieuwe relaties hadden betrekking op effecten van covariaten: de voortoets bleek ook een direct effect op de taakgerichtheid te hebben, en leerlingen met storend gedrag hadden meer organisatorische en persoonsgerichte interacties met de leerkracht dan de leerlingen met een laag ingeschat IQ.

Om te bezien of het model de relatie tussen LLR en taalvaardigheid inderdaad kan verklaren, is deze relatie uiteengelegd in directe effecten en indirecte, door andere variabelen gemedieerde effecten: 81.7% van deze relatie is direct en 18.3% verloopt via de gehypothetiseerde tussenliggende schakels.

Modellen voor groeps grootte en LVR in groep 1

Het conceptuele model is ook getoetst voor de relaties tussen groeps grootte en de leerling-volwasseneratio enerzijds en de leerlingprestaties anderzijds. In beide gevallen bleek het model niet goed bij de gegevens te passen (groeps grootte model: $\chi^2 = 26.61$, $p = .05$, $CFI = .93$, $RMSEA = .06$; LVR model: $\chi^2 = 44.23$, $p = .00$, $CFI = .82$, $RMSEA = .10$). Ten aanzien van de kern van het model (groeps grootte, LVR, interacties, taakgerichtheid en leerlingprestaties) vertoonden de uiteindelijke, wel passende modellen, grote gelijkenis met het hierboven gepresenteerde model voor de relatie LLR en taalvaardigheid. Zo bleken ook groeps grootte en de LVR tegen de verwachting in positief samen te hangen met beide onderscheiden typen interacties. Een afwijking bestond hierin dat organisatorische en persoonsgerichte interactie in het LVR-model weliswaar negatief, maar niet significant bleken samen te hangen met taakgerichtheid. Voor het overige verlopen de belangrijkste relaties conform het gehypothetiseerde model.

Teneinde te geraken tot een model met een acceptabele passing, werden extra effecten van covariaten toegestaan: de voortoets met een direct effect op de taakgerichtheid, en storend gedrag met een direct effect op de organisatorische en persoonsgerichte interacties. Had LLR een onverwacht positief direct effect op taakgerichtheid, bij groeps grootte en LVR trad deze anomalie niet op. Het LVR-model week op een ander punt ook af van het LLR-model: LVR bleek geen direct effect op leerlingprestaties te hebben. Voor groeps grootte trad dit directe effect echter wel weer op: hoe kleiner de groep, hoe groter de taalvaardigheid. De modelpassing voor zowel het groeps grootte- als het LVR-model was acceptabel. Voor groeps grootte: $\chi^2 = 6.70$, $p = .92$, $CFI = 1.00$, $RMSEA = .00$ en voor LVR: $\chi^2 = 6.42$, $p = .93$, $CFI = 1.00$, $RMSEA = .00$. De uiteindelijke causale modellen voor de relatie tussen groeps grootte en LVR enerzijds en leerlingprestaties anderzijds zijn opgenomen in Appendix 1.

7 Conclusies en discussie

7.1 Groepsgrootte, LLR, LVR en leerlingprestaties

In het begin van dit artikel hebben we twee vragen geformuleerd:

- 1 Zijn de taalprestaties van leerlingen in kleinere groepen (of in groepen met een lagere LLR of LVR) hoger dan van leerlingen in grotere groepen (of in groepen met een hoge LLR of LVR)?
- 2 In hoeverre is het analytisch gezien zinvol een onderscheid te maken tussen de maten groepsgrootte, LLR en LVR?

De resultaten van de meerniveau-analyses brachten slechts één significante relatie aan het licht: er is een samenhang tussen LLR en de passieve woordenschat van leerlingen in groep 1. Hoe meer bevoegde leerkrachten bij het onderwijs in een groep betrokken zijn, hoe meer woorden de leerling passief beheerst.

In dit onderzoek is ervoor gekozen diverse vormen van inzet van personeel en ouders op hun effecten te bekijken, en niet enkel groepsgrootte. Dat dit een zinvolle aanpak was, blijkt wel uit de gedifferentieerde resultaten. Zo vinden we dus wel een effect van LLR, maar niet van LVR en groepsgrootte. Het is ons inziens derhalve zinvol in toekomstig onderzoek dit analytisch onderscheid te blijven hanteren.

Er is een drietal methodologische kanttekeningen bij het gepresenteerde onderzoek te plaatsen. De gebruikte testen hadden een dusdanig beperkte omvang dat er mogelijk plafondeffecten kunnen zijn opgetreden.

Omdat we echter gecontroleerd hebben voor de scores die leerlingen op de pretest behaald hadden, speelt dit geen wezenlijke rol. Voorts zouden mogelijke uitschieters (groepen met een extreme omvang of met extreme taalresultaten) de resultaten van het onderzoek beïnvloed kunnen hebben. Derhalve zijn de analyses ook nog eens uitgevoerd zonder deze extreme groepen, en de resultaten bleken tamelijk robuust te zijn. Een laatste opmerking betreft het onderscheidingsvermogen. Gelet op het beperkte aantal scholen (46) en groepen (73), is de kans dat een in de populatie aanwezig verband tussen groepsgrootte, LLR en LVR enerzijds en leerlingprestaties anderzijds in een dergelijke, kleine steekproef niet wordt aangetroffen groter dan wenselijk.

7.2 Groepsgrootte, LLR, LVR: de verbinding tussen onderwijzen en leren

De derde en laatste vraag die we ons stelden betrof de aard van het oorzakelijk verband tussen groepsgrootte, LLR en LVR enerzijds en leerlingprestaties anderzijds: hoe komt ons daarbij gericht op leerling-leerkrachtinteracties en taakgerichte leertijd als mogelijke mediërende factoren. De te toetsen veronderstelling was dat in kleinere groepen (of groepen met een lagere LLR of LVR) minder organisatorische of persoonsgerichte interacties zouden zijn, en meer interacties gericht op de taak, of prijzend. Dit zou dan leiden tot meer taakgerichtheid, met als gevolg dat de taalvaardigheid van de leerling zou toenemen.

Dit conceptuele model werd getoetst in

Tabel 5

Samenvatting van de belangrijkste relaties in de verschillende uiteindelijke modellen

Relatie	GG	LLR	LVR
GG/LLR/LVR - organisatorische en persoonsgerichte interacties	-.14	-.12	-.13
GG/LLR/LVR - taakgerichte en prijzende interacties	-.22	-.20	-.14
Organisatorische en persoonsgerichte interacties - taakgerichtheid	-.14	-.13	-.11*
Taakgerichte en prijzende interacties - taakgerichtheid	.13	.16	.16
Time on task - natoets	.16	.18	.16
GG/LLR/LVR - taakgerichtheid19	...
GG/LLR/LVR - natoets	-.09	-.14	...

* Niet significant.

groep 1. Het gehypothetiseerde model moest enerzijds aangepast worden om een acceptabele fit te bereiken. In de samenvattende Tabel 5 worden de resultaten gepresenteerd voor de belangrijkste verbanden.

Voor alle modellen geldt dat de samenhang tussen groeps grootte, LLR en LVR enerzijds en het aantal organisatorische en persoonsgerichte interacties anderzijds, negatief is: hoe kleiner de groeps grootte (resp. de LLR of de LVR), hoe meer dergelijke interacties optreden. Dit was niet conform de verwachting. Op basis van eerder onderzoek en het theoretisch raamwerk werd de hypothese geformuleerd dat deze samenhang positief zou zijn. Daarentegen werd wel ondersteuning gevonden voor de verwachte positieve samenhang tussen groeps grootte, LLR en LVR enerzijds en het aantal taakgerichte en prijzende interacties anderzijds. Voorts blijken conform de verwachting in het algemeen de organisatorische en persoonsgerichte interacties gepaard te gaan met minder, en taakgerichte, respectievelijk prijzende interacties met meer taakgerichte leertijd.

De laatste belangrijke relatie die in dit onderzoek werd bestudeerd, betrof die tussen taakgerichtheid en taalvaardigheid. De gegevens ondersteunen de idee dat deze samenhang positief zou moeten zijn: naarmate leerlingen meer taakgericht zijn, zijn hun taalprestaties hoger.

Samenvattend kunnen we stellen dat leerling-leerkrachtinteracties en de taakgerichtheid van leerlingen inderdaad een rol lijken te spelen als mediërende variabelen in de relatie tussen groeps grootte, LLR en LVR enerzijds en leerlingprestaties anderzijds: 18.3% van het geconstateerde verband tussen LLR en leerlingprestaties in groep 1 blijkt via deze paden te verlopen. Toch blijkt het verklaren van de relatie tussen groeps grootte, LLR en LVR en de taalvaardigheid van leerlingen complexer dan in het model was gehypothetiseerd. Zo gaan kleinere groepen of lagere LLR's en LVR's samen met meer leerling-leerkrachtinteracties, ook als deze niet op het onderwijs zelf betrekking hebben. De kern van het model liet zien dat kleinere groepen (of groepen met een lagere LLR of LVR) samengingen met meer leerling-leerkrachtinteracties - ongeacht of deze interacties nu

onderwijsinhoudelijk of meer organisatorisch gericht zijn - en dat de onderwijsinhoudelijke interacties leidden tot een grotere mate van taakgerichtheid. Die taakgerichtheid ging in groep 1 dan weer samen met een grotere taalvaardigheid.

Alles overziend, lijkt nader onderzoek naar het verband tussen de groeps grootte, LLR, LVR en prestaties noodzakelijk om een verantwoorde uitspraak te kunnen doen over de kwestie of het nu al dan niet is aan te bevelen de groepen te verkleinen of de LLR of LVR te verlagen. Echter, het onderwijsproces zelf, en met name het handelen van de leerkracht zou wellicht meer aandacht moeten krijgen. Zij zouden zich wellicht meer bewust moeten zijn van hun handelen en van de voordelen van kleinere groepen en zij zouden toegerust moeten worden met meer kennis over *hoe* men het onderwijs zo kan aanpassen dat de voordelen van een kleinere groep (of een lagere LLR of LVR) optimaal worden benut. Met name ook, omdat binnen de scholen die participeerden in dit onderzoek geen duidelijke relatie is gevonden tussen groeps grootte, LLR, LVR en de mate waarin leerkrachten adaptief onderwijs geven (zie Annevelink, 2004; dit gedeelte van het onderzoek is hier echter niet beschreven).

Noten

- 1 Voor meer informatie over de leerlingkenmerken: zie het eerste deel van paragraaf 5.2.
- 2 Voor meer informatie over de soorten interacties: zie het laatste deel van paragraaf 5.2.
- 3 Persoonsgerichte interacties blijken gedurende de observaties binnen dit onderzoek nauwelijks voor te komen.
- 4 Exploratieve analyses lieten zien dat, gegeven de voortoets, SES, thuistaal en leeftijd geen significante effecten hadden. Om deze reden maken deze covariaten geen deel uit van het model.
- 5 Gepresenteerd worden de gestandaardiseerde coëfficiënten.

Literatuur

- Anderson, L. W. (2000). Why should reduced class size lead to increased student achievement? In M. C. Wang & J. D. Finn. *How small classes help teachers do their best* (pp. 3-24). Philadelphia: Temple University Center for Research in Human Development and Education.
- Annevelink (2004). *Class size: Linking teaching and learning*. Enschede: Printpartners Ipskamp.
- Betts, J. R., & Shkolnik, J. L. (1999). The behavioral effect of variations in class size: The case of math teachers. *Educational Evaluation and Policy Analysis, 21*(2), 193-213.
- Biesheuvel, M., & Flim, G. (2001). Het vermogen van leerkrachten de intelligentie van kleuters in te schatten. In W. J. Van der Linden & S. Oosterloo (Reds.), *Proceedings van het symposium in het kader van het vak onderzoeksopdracht (196037)*. Enschede: Faculty of educational science and technology. Department of educational measurement and data-analysis. University of Twente.
- Blatchford, P., & Mortimore, P. (1994). The issue of class size for young children in schools: What can we learn from research? *Oxford Review of Education, 20*(4), 411-428.
- Blatchford, P., Baines, E., Kutnick, P., & Martin, C. (2001). Classroom contexts: Connections between class size and within class grouping. *British Journal of Educational Psychology, 71*(2), 283-302.
- Bohrnstedt, G. W., & Stecher, B. M. (Eds.). (1999). *Class size reduction in California: Early evaluation findings, 1996-1998. (CSR Research Consortium, Year 1 Evaluation Report)*. Palo Alto, CA: American Institutes for Research.
- Bourke, S. (1986). How smaller is better: some relationships between class size, teaching practices and student achievement. *American Educational Research Journal, 23*, 558-571.
- Boyd-Zaharias, J., & Bain, H. P. (2000). Early and New Findings from Tennessee's project STAR. In M. C. Wang & J. D. Finn, *How small classes help teachers do their best*. (pp. 65-97). Philadelphia: Temple University Center for Research in Human Development and Education.
- Cahen, L. S., Filby, N. N., McCutcheon, G., & Kyle, D. W. (1983). *Class size and instruction*. New York: Longman.
- CEB (Commissie Evaluatie Basisonderwijs). (1994). *Zicht op kwaliteit*. Den Haag: SDU.
- CKAG (Commissie Kwalitatieve Aspecten van Groepsgrootte in het Basisonderwijs). (1996). *Klassenverkleining. Advies over de betekenis van klassenverkleining voor beter basisonderwijs*. Den Haag: SDU.
- Doolaard, S., Annevelink, E., Bosker, R. J., & Cremers-van Wees, L. M. C. M. (2000). *Organisatie en vormgeving van het onderwijs in de onderbouw. Formatie-inzet en differentiatie op microniveau na de eerste stap op weg naar verkleining van de groepsgrootte*. Enschede: Universiteit Twente, OCTO.
- Evers, A., Vliet-Mulder, J. C. van, Resing, W. C. M., Starren, J. C. M. G., Alphen de Veer, R. J. van, & Boxtel, H. van. (2002). *COTAN Testboek voor het onderwijs*. Amsterdam: NDC-Boom.
- Finn, J. D., & Achilles, C. M. (1999). Tennessee's class size study: Findings, implications, misconceptions. *Educational Evaluation and Policy Analysis, 21*(2), 97-100.
- Follman, J. (1991). Teachers' estimates of pupils' IQs and pupils' tested IQs. *Psychological Reports, 69*, 350.
- Glass, G. V., & Smith, M. L. (1979). Meta-analyses of research on class size and achievement. *Educational Evaluation and Policy Analysis, 1*(1), 2-16.
- Glass, G. V., Cahen, L. S., Smith, M. L., & Filby, N. N. (1982). *School class size*. London: Sage Publications.
- Hanushek, E. A. (1999). Some findings from independent investigation of the Tennessee Star Experiment and from other investigations of class size effects. *Educational Evaluation and Policy Analysis, 21*(2), 215-229.
- Hedges, L. V. (2000). Using Converging Evidence in policy Formation: The Case of Class Size Research. *Evaluation and Research in Education, 14*(3&4), 193-205.
- Houtveen, A. A. M., Pijl, S. J., Pijl, Y. J., Reezigt, G. J., & Vermeulen, C. J. (1998). *Adaptief onderwijs. Stand van zaken in het WSNS-proces*. De Lier: Academisch Boeken Centrum.
- Houtveen, A. A. M., & Reezigt, G. J. (2000a). *Succesvol adaptief onderwijs, ontwikkelingen in het denken over adaptief onderwijs*. Paper gepresenteerd op de Onderwijs Research

- Dagen, Leiden 2000.
- Houtveen, A. A. M., & Reezigt, G. J. (2000b). *Succesvol adaptief onderwijs. Handreikingen voor de praktijk*. Alphen aan den Rijn: Samson.
- Inspectie van het Onderwijs. (1995). *Groeps-grootte in het basisonderwijs*. Utrecht: Inspectorate of Education.
- Krueger, A. B., & Whitmore, D. M. (1999). *The effect of attending a small class in the early grades on college-test taking and middle school test results: Evidence from project STAR*. [Working paper no. 427]. Princeton, NJ: Princeton University.
- Kuyk, J. J. van. (1996). *Taal voor kleuters. Hand-leiding*. Arnhem: Cito.
- Molnar, A., Smith, Ph., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. (1999). Evaluating the SAGE Program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational Evaluation and Policy Analysis*, 21(2), 165-177.
- Molnar, A., Smith, P., Zahorik, J., Palmer, A., Halbach, A., & Ehrle, K. (2000). Wisconsin's Student Achievement Guarantee in Education (SAGE) class size reduction program: Achievement effects, teaching, and classroom implications. In M. C. Wang & J. D. Finn, *How small classes help teachers do their best* (pp. 227-277). Philadelphia: Temple University Center for Research in Human Development and Education.
- PMPO (Procesmanagement Primair Onderwijs). (1999). *Groepsverkleining in de onderbouw. De tweede stap*. Den Haag: Procesmanagement Primair Onderwijs.
- Reezigt, G. J., Guldmond, H., & Ros, A. A. (1995). *Effectief onderwijs aan jonge kinderen*. Groningen: RION, Instituut voor Onderwijsonderzoek, Rijksuniversiteit Groningen.
- Shapson, S. M., Wright, E. N., Eason, G., & Fitzgerald, J. (1980). An experimental study of the effects of class size. *American Educational Research Journal*, 17(2), 141-152.
- Smith, M. L., & Glass, G. V. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Educational Journal*, 17(4), 419-433.
- Stecher, B. M., & Bohrnstedt, G. W. (Eds.). (2000). *Class size reduction in California: The 1998-1999 evaluation findings*. Sacramento, CA: California Department of Education.
- Stecher, B. M., & Bohrnstedt, G. W. (Eds.). (2002). *Class size reduction in California: Findings from 1999-00 and 2000-01*. Sacramento, CA: California Department of Education.
- Tellegen, P., Winkel, M., Wijnberg-Williams, B., & Laros, J. (1998). *Snijders-Oomen Niet-verbale intelligentietest. SON-R 2?-7. Handleiding en verantwoording. Speciale Uitgave*. Lisse: Swets Test Publishers.
- Werkman, E., & Faber, S. (2000). *De sociaal-emotionele ontwikkeling evalueren. Drie instrumenten voor signalering en gedragsbeoordeling*. Oegstgeest: STEVIN Uitgeverij-Adviesbureau.
- Wild, K. P. (1993). *Studies in Educational Psychology. No. 5. Teacher-judgements and the identification of gifted pupils*. München: Institut für Erziehungswissenschaft und Pädagogische Psychologie. Universität der Bundeswehr München.

Manuscript aanvaard: 1 november 2004

Auteurs

Elvira Folmer-Annevelink is IOWO-onderzoeker voor onderwijs, beleid en organisatie aan de Radboud Universiteit Nijmegen.

Roel J. Bosker is als hoogleraar Onderwijskunde en directeur van het GION, Instituut voor Onderzoek van Onderwijs, Opvoeding en Ontwikkeling, verbonden aan de Rijksuniversiteit Groningen.

Simone Doolaard is als universitair docent werkzaam aan het hetzelfde instituut.

Correspondentieadres: E. Folmer-Annevelink, Radboud Universiteit Nijmegen, Postbus 6540, 6503 GA Nijmegen, e.folmer@iowo.ru.nl

Abstract

Additional staffing, classroom processes and achievement

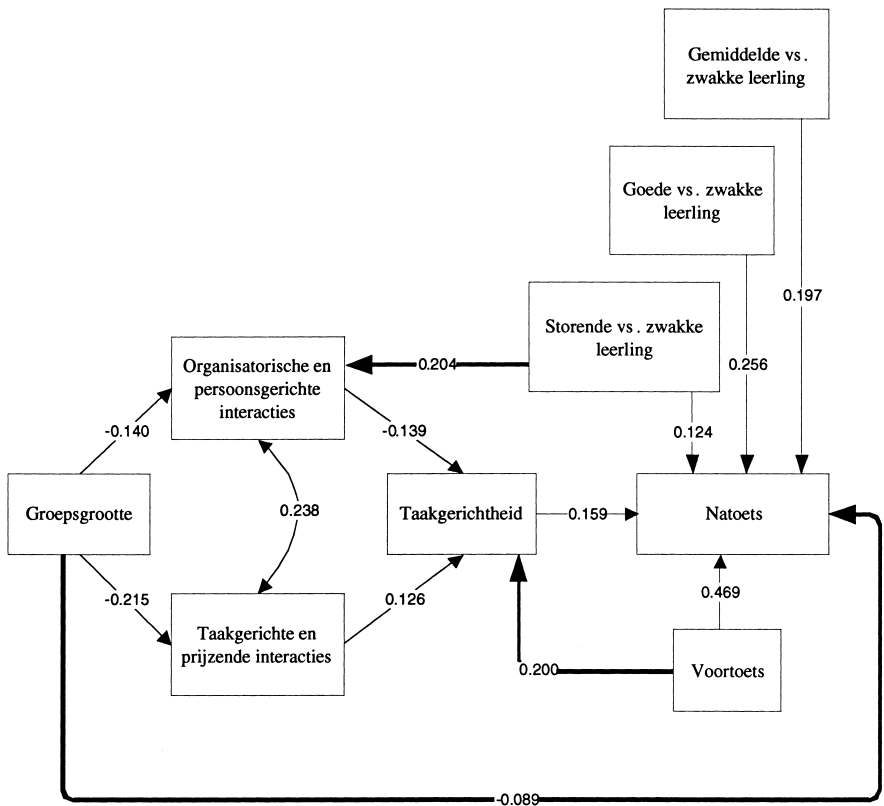
In 1997 the Dutch Ministry started with reducing class sizes by making a supplemental budget for additional staffing available. With this investment

a large amount of money was involved. Now, the question is whether this additional formation results in higher pupil achievement and consequently, when this is the case, how such an effect can be explained. This research studied the effects of additional formation within 46 Dutch primary schools. Within these schools pupils of Grade 1 were tested. Also, their interactions with

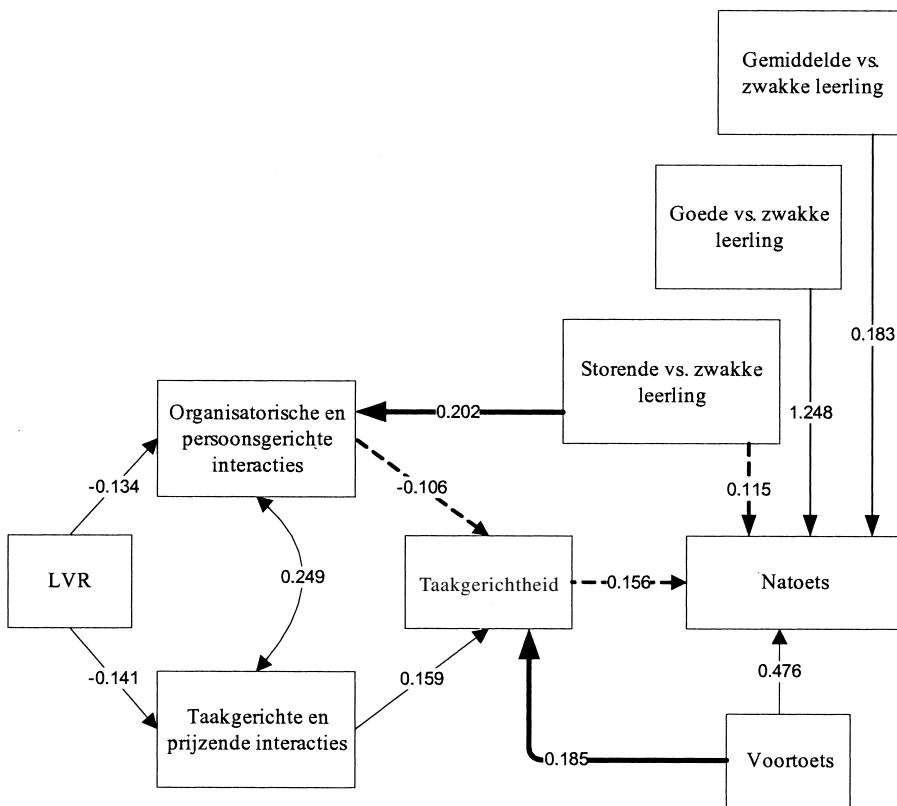
the teacher and the amount of time they spent on task were observed. Results show that the pupil-teacher ratio had a positive influence on language achievement, contrary to class size and pupil-adult ratio. Pupil-teacher interactions and time on task seemed to play a, modest, role in explaining this relationship.

Appendix 1

Uiteindelijk modellen voor groeps-grootte en LVR in groep 1



Figuur A1. Uiteindelijk model voor groeps-grootte en prestaties in groep 1.



Figuur A2. Uiteindelijk model voor LVR en prestaties in groep 1.