

‘Peer group’ effecten op onderwijsprestaties

Een internationaal review van effecten, verklaringen en theoretische en methodologische aspecten

Dr. Geert Driessen

November 2007

ITS – Radboud Universiteit Nijmegen

ITS-project 2007.447

© 2007 ITS, Radboud Universiteit Nijmegen

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvuldigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook, en evenmin in een retrieval systeem worden opgeslagen, zonder de voorafgaande schriftelijke toestemming van het ITS van de Radboud Universiteit Nijmegen.

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

Voorwoord

Recentelijk is door de Organisation for Economic Co-operation and Development (OECD) opgemerkt dat het potentieel onder allochtone leerlingen in Nederland onvoldoende wordt benut. Hierdoor zou hun deelname aan het hoger onderwijs relatief gering zijn. Een van de mogelijke oorzaken van deze onderbenutting is volgens de OECD gelegen in het feit dat leerlingen na het basisonderwijs vroegtijdig worden voorgesorteerd in homogene schooltypes. Tegenover homogene schooltypes staan heterogeen samengestelde types, waarin sprake is van een mix van capaciteiten en prestatieniveaus en – deels daarmee samenhangend – verschillen naar sociaal milieu, etnische herkomst en sekse. De veronderstelling is dat specifieke samenstellingen van de leerlingenpopulaties van klassen en de interacties tussen individuele leerlingen en groepen van medeleerlingen ('peers') binnen die klassen een effect hebben op het prestatieniveau. De vraag is of dergelijke effecten inderdaad optreden, onder welke omstandigheden dat dan gebeurt en hoe groot ze zijn.

Op verzoek van de directie Algemeen Strategische en Economische Advisering (ASEA) van het ministerie van Onderwijs, Cultuur en Wetenschap heeft dr. Geert Driessen van het ITS van de Radboud Universiteit Nijmegen een overzichtsstudie verricht naar dergelijke 'peer group' effecten. De voorliggende rapportage bevat het verslag van deze internationale literatuurstudie waarin peer group effecten vanuit verschillende perspectieven zijn bestudeerd. Niet alleen is het effect onderzocht van de groepssamenstelling in termen van etnische herkomst, sociaal milieu en sekse, maar ook van prestatieniveau en leeftijd. In totaal zijn de resultaten van vele honderden afzonderlijke onderzoeken naar peer group effecten bij elkaar gebracht. Behalve de effecten worden ook de uiteenlopende theoretische verklaringen die daaraan ten grondslag liggen beschreven. In verband met de complexiteit van de materie wordt tevens uitvoerig stil gestaan bij allerlei haken en ogen die met onderzoek naar peer group effecten verbonden zijn.

ITS - Radboud Universiteit Nijmegen

Dr. Jeroen Winkels
Directeur

Inhoudsopgave

Voorwoord	iii
1 Achtergronden, probleemstelling en onderzoeksopzet	1
1.1 Achtergronden	1
1.1.1 Niveaus	1
1.1.2 ‘Selective versus comprehensive’	1
1.1.3 Voor- en nadelen tracking	2
1.1.4 Verklaringen	2
1.2 Probleemstelling	3
1.3 Methode	3
2 Theoretische en methodologische aspecten	5
2.1 Inleiding	5
2.2 Een nadere definitie	5
2.3 Inconsistenties in bevindingen	6
2.4 Oorzaken voor inconsistenties	6
2.4.1 Causaliteit	6
2.4.2 Niet-gemeten kenmerken	7
2.4.3 Selectie	7
2.4.4 Aggregatieniveaus	8
2.4.5 Ontwikkelingsverschillen	8
2.4.6 Kwaliteit interacties	9
2.5 Conclusies	9
3 Verklaringen voor peer group effecten	11
3.1 Inleiding	11
3.2 Specifieke verklaringen	11
3.2.1 Taalcontact	11
3.2.2 Netwerken	12
3.2.3 Nivellering	12
3.3 Algemene theorieën	12
3.3.1 Vergelijkingstheorieën	12
3.3.2 Netwerktheorieën	14
3.4 Dwarsverbanden	15

4 Resultaten	17
4.1 Inleiding	17
4.2 Ability grouping	18
4.3 Combinatieklassen	19
4.4 Single-sex klassen/scholen	21
4.5 Compositie qua milieu en etniciteit	23
4.5.1 Desegregatie in de VS	23
4.5.2 Internationale studies	24
4.5.3 Nederlandse studies	26
4.6 Recente onderwijs-economische studies	30
Conclusies	33
Literatuur	35

1 Achtergronden, probleemstelling en onderzoeksopzet

1.1 Achtergronden

1.1.1 Niveaus

Onderwijskansen worden door een veelheid van factoren bepaald. Die zijn gelegen op het niveau van de leerling (genetisch), de gezinssituatie (milieu, etniciteit, opvoeding), de klas en school (curriculum, pedagogisch-didactische aanpak, organisatie) en het (macro)systeem (het onderwijsstelsel) (Driessen & Dekkers, 2007). Recentelijk is er in internationaal comparatieve studies – opnieuw – op gewezen dat verschillen in de wijze waarop onderwijssystemen zijn ingericht van invloed zijn op de prestaties van leerlingen. Relevant in dit verband is de constatering dat met name de prestaties van allochtone achterstandsleerlingen negatief worden beïnvloed in systemen waar sprake is van vroege selectie ('early tracking') (Hanushek & Wößmann, 2006).

1.1.2 'Selective versus comprehensive'

In feite draait het hier om een oude discussie over de voor- en nadelen van 'selective versus comprehensive' schoolsystemen¹ en – meer in het algemeen – homogeen versus heterogeen groeperen. Wat dit laatste betreft kan dan een onderscheid worden gemaakt naar verschillende 'tracks' (schooltypes) binnen een onderwijsniveau (bv. het voortgezet onderwijs) en verschillende tracks binnen een school. In sommige landen (bv. Duitsland) worden kinderen al op 10-jarige leeftijd voorgesorteerd in voornamelijk op gelijkheid van prestaties en capaciteiten gebaseerde tracks (Imdorf, 2003; Schnepf, 2002). In andere landen (bv. het Verenigd Koninkrijk) blijven de kinderen tot 16-jarige leeftijd bij elkaar en kiezen ze pas daarna voor een vervolgopleiding. Nederland neemt een tussenpositie in: op 12-jarige leeftijd maken de kinderen een keuze voor één van de schooltypes, maar door de combinaties van schooltypes en de basisvorming is er – in principe – nog enige jaren sprake van uitstel van een definitieve schoolkeuze. In de praktijk echter blijken de bij de overgang van basis- naar voortgezet onderwijs gemaakte keuzes toch vrij definitief en is er slechts in beperkte mate sprake van op- en afstroom (Claassen & Mulder, 2006). Een dergelijke op- en afstroom blijkt dan voornamelijk het resultaat te zijn van verkeerde adviezen en keuzes bij de overgang van basis- naar voortgezet onderwijs (de zogenoemde 'over-' en 'onderadvisering'), die – zo werd tot voor kort verondersteld – vooral voor allochtone leerlingen negatieve gevolgen zouden hebben (Driessen & Smeets, 2007).

¹ Deze discussie beleefde in de jaren '70 onder onderwijsminister Van Kemenade zijn hoogtepunt met de vanuit het perspectief van gelijkheid van kansen geïntroduceerde Middenschool.

1.1.3 Voor- en nadelen tracking

Beide genoemde schoolsystemen kennen voor- en nadelen. Een voordeel van 'ability' tracking binnen een 'selective' schoolstelsel is dat leerlingen van vergelijkbaar niveau bij elkaar zitten, waardoor het tempo niet wordt belemmerd door langzame en snelle leerlingen die allemaal extra (individuele) aandacht van de leraar nodig hebben. Voor de leraar maakt dit het lesgeven gemakkelijk (maar interessanter?) en voor de leerlingen houdt dit in dat ze veel sneller door de stof kunnen gaan. Tracking komt daarmee tegemoet aan de behoeften van individuele leerlingen en verbetert de prestaties van vooral de leerlingen in de hogere niveaugroepen. Ook voor hoogbegaafde leerlingen kan tracking een manier zijn waarbij wordt tegemoetgekomen aan hun specifieke capaciteiten en behoeften. Tracking heeft tevens als voordeel dat leerlingen efficiënter worden voorbereid op vervolgstudies. Het heeft zo een sorteer- en allocatiefunctie voor de verschillende soorten vervolgonderwijs, met name ten aanzien van de keuze voor hoger beroepsonderwijs dan wel wetenschappelijk onderwijs. Een nadeel van tracking is dat de laagste niveaugroepen nauwelijks stijgingskansen hebben, doordat ze in een weinig uitdagende leeromgeving zitten en ten gevolge van (stereotype) verwachtingen en 'self-fulfilling prophecies' ook zullen blijven zitten (Van der Hoeven-Van Doornum, Voeten & Jungbluth, 1989). Bij de betreffende leerlingen ontstaan daardoor gevoelens van minderwaardigheid en defaitisme (Slavin, 1987). De meest uitgebreide faciliteiten komen via tracking juist terecht bij die leerlingen, die toch al onder de meest gunstige omstandigheden onderwijs volgen. Er ligt wat dit betreft een duidelijke link met het herkomstmilieu van de leerlingen. Daar komt nog bij dat bij de toelating tot bepaalde tracks niet altijd louter de capaciteiten en prestaties een rol kunnen spelen, maar ook achtergrondkenmerken zoals sociaal milieu en etnische herkomst (Driessen, 2006).

Hanushek & Wößmann (2006) hebben onlangs de resultaten gepresenteerd van een landenvergelijkend onderzoek naar effecten van (early) tracking. Uit hun analyses bleek dat in nagenoeg alle landen (waaronder Nederland) waarin sprake is van een tracking systeem de relatieve ongelijkheid tussen basis- en voortgezet onderwijs toeneemt; de grootste toename vindt plaats in landen met een early tracking systeem. Volgens hen is er mogelijk een samenhang met sociaal-etnische milieukekenmerken (zie ook Entorf & Lauk, 2006; Michaelowa & Bourdon, 2006). Volgens de OECD (2006) is dit early tracking systeem er voor verantwoordelijk dat vooral het potentieel onder allochtone leerlingen onvoldoende benut wordt, waardoor hun deelname aan het hoger onderwijs relatief gering is Nederland.

1.1.4 Verklaringen

Als het gaat om verklaringen voor de negatieve effecten van (early) tracking systemen wordt vaak gewezen op specifieke samenstellingen van de leerlingenpopulaties van klassen en de interacties tussen individuele leerlingen en groepen van leerlingen binnen die klassen. In dat verband wordt wel gesproken over compositie-effecten, concentratie-effecten, 'frog pond' effecten² en 'peer group' effecten. Het betreft steeds aanduidingen van processen die zich binnen een klas afspelen en die worden bepaald door de specifieke samenstelling van de leerlingenpopulatie. Met name de verhoudingen tussen de meerderheid en minderheid lijken daarbij van belang. Daarbij kan bijvoorbeeld gedacht worden in termen van etnische concentratie (het 'zwarte-

2 Ook wel 'big-fish-little-pond' of BFLP effect genoemd (Marsh & Tau, 2003; Zeidner & Schleyer, 1998).

scholeneffect'), sociaal-milieuconcentratie ('arbeidersscholen'-effect), sekse (meisjesscholen), het prestatieniveau (ability) van de klas, en het intelligentieniveau (hoogbegaafden).

1.2 Probleemstelling

In het voorgaande is aangegeven dat een early tracking systeem, wat uitgaat van naast elkaar bestaande homogene schooltypes en klassen, negatieve consequenties kan hebben voor de prestaties van leerlingen in het voortgezet onderwijs, en dan met name voor allochtone kinderen uit lagere sociaal-economische milieus. Ook is aangegeven dat de invloed van de 'peers' in een klas op elkaar mogelijk een verklaring kan bieden voor dit negatieve effect. Over het effect van de peer group bestaat veel onduidelijkheid en discussie. Het doel van het onderhavige, kwantitatieve review is via een internationale literatuurstudie de grootte van dit effect op de onderwijsprestaties, de werking ervan en de voorwaarden waaronder het zich voordoet systematisch in kaart te brengen. Speciale aandacht daarbij krijgt de positie van allochtone leerlingen. Hoewel de aanleiding voor dit onderzoek is gelegen bij veronderstellingen rond early tracking in het voortgezet onderwijs, richt dit review zich in principe op peer group effecten in alle onderwijsniveaus. Daarvoor wordt zowel gebruik gemaakt van de onderwijskundige/sociologische, onderwijs-economische en sociaal-psychologische literatuur. Op basis van het review wordt een wetenschappelijk oordeel gegeven over de robuustheid van de bevindingen van internationaal onderzoek.

1.3 Methode

De kern van het onderzoek bestaat uit een literatuurstudie, waarbij het onderwerp vanuit verschillende invalshoeken wordt bestudeerd. Daarbij worden vier strategieën gevolgd:

- *Een boekenkast search* van de uitvoerder van het onderzoek, die een reeks studies rond het onderhavige thema heeft uitgevoerd.
- *'Colleague consultancy'*. Een aantal andere ITS-collega's heeft eveneens onderzoek verricht op het genoemde terrein; zij zullen worden geconsulteerd.
- *Een literatuur search*. Met behulp van de elektronische zoeksystemen zal naar literatuur worden gezocht rond het thema.
- *De 'snowball-method'*. Op basis van de drie voorgaande strategieën, en dan met name de daarbij verzamelde literatuurreferenties, wordt verder gezocht naar andere bronnen.

1.4 Leeswijzer

In de hoofdstukken hierna wordt eerst ingegaan op een aantal theoretische en methodologische aspecten van onderzoek naar peer group effecten. Daarna volgt een hoofdstuk waarin mogelijke verklaringen voor het optreden van peer group effecten bij elkaar zijn gezet. Het daarna volgende hoofdstuk bevat de kern van het review; daarin worden de bevindingen van onderzoek naar peer group effecten beschreven. Het rapport wordt afgesloten met een samenvatting en conclusies.

2 Theoretische en methodologische aspecten

2.1 Inleiding

In het vorige hoofdstuk is aangegeven dat in dit review peer group effecten centraal staan. Ook is er op gewezen dat er veel discussie is over dergelijke effecten. Die discussie heeft deels te maken met het feit dat niet steeds duidelijk wordt gemaakt wat er nu precies mee wordt bedoeld. Daarom wordt in paragraaf 2.2 eerst geprobeerd tot een begripsverheldering te komen. Een andere reden voor de discussie rond peer group effecten is dat er veel inconsistentie is in onderzoeksbevindingen. In paragraaf 2.3 en 2.4 wordt daarom uitvoerig aandacht besteed aan de mogelijke oorzaken voor deze inconsistentie. Hoewel het gebruikelijk is mogelijke voorbehouden bij resultaten van onderzoek pas op het eind van een rapportage te vermelden, is er hier voor gekozen dat al aan het begin te doen. De verwachting is dat daardoor de hierna gepresenteerde onderzoeksbevindingen beter kunnen worden geplaatst en op hun waarde kunnen worden beoordeeld.

2.2 Een nadere definitie

In de literatuur worden verschillende betekenissen toegekend aan de aanduiding 'peer group effecten'. Behalve 'peer group effecten' circuleren er nog andere benamingen, zoals 'compositie-effecten', 'concentratie-effecten', 'contextuele effecten', 'school mix effecten', 'zwart-schooleffecten' en 'frog pond effecten'. Niet altijd is even duidelijk welke definitie wordt gevolgd (Ammermüller & Pischke, 2006). De aanduiding 'peer group effecten' fungeert als een soort containerbegrip. Wilkinson e.a. (2000) maken echter een onderscheid tussen compositie-effecten en 'echte' peer group effecten. Compositie-effecten zijn effecten van een geaggregeerd kenmerk van een groep leerlingen (bv. het gemiddelde prestatieniveau van een klas of het percentage allochtone leerlingen op een school) op de leerresultaten van een leerling, bovenop de effecten die te herleiden vallen tot individuele kenmerken van leerlingen. Compositie-effecten kunnen bijvoorbeeld het gevolg zijn van verschillen in schoolklimaat, in didactiek en in financiering, maar ook van peer group effecten. 'Echte' peer group effecten betreffen dan de invloeden van leerling tot leerling-interacties en groepsdynamische processen op leerresultaten. Zo gezien vormen kenmerken van de compositie van een school, klas of subgroep de structuur waarbinnen de peer group processen plaats kunnen vinden. Volgens Winston & Zimmerman (2003) is er sprake van peer group effecten, wanneer iemands gedrag wordt beïnvloed door diens interactie met een of meer andere personen die diens 'gelijken' zijn. Gibson & Telhaj (2006) formuleren het iets anders: elk type individueel gedrag waarbij een wederzijdse afhankelijkheid bestaat van gedrag of kenmerken van anderen. Zoals opgemerkt, wordt dit zowel analytische als conceptuele onderscheid in de literatuur vaak niet gemaakt. Hoewel er veel onderzoek is verricht naar peer group effecten in termen van compositieverschillen, is er daarbij weinig aandacht voor het mechanisme dat ten grondslag ligt aan dergelijke effecten, dat wil zeggen de 'echte' peer group effecten (Hanushek, Kain, Markman & Rivkin, 2003; Lavy & Schlosser,

2007). In het onderhavige review wordt doorgaans de aanduiding uit de geciteerde studies overgenomen.

2.3 Inconsistenties in bevindingen

Als er één ding is waarover onderzoekers op het terrein van peer group effecten het eens lijken te zijn is dat het feit dat ze het oneens zijn over het bestaan en de sterkte ervan (o.a. Gibson & Telhaj, 2006; McEwan, 2003; Kang, 2007; Manger & Eikeland, 1997; Manski, 1993; Thrupp, 1995, 1999; Wilkinson e.a., 2000). Het ontbreekt kennelijk aan overtuigend en consistent empirisch bewijs voor het bestaan van dergelijke effecten op uiteenlopende cognitieve, sociale en arbeidsmarktmaten. Daarbij spelen zowel conceptuele als ook methodologische kwesties een rol (Moffitt, 2001; Thrupp, Lauder & Robinson, 2002). Deze problemen roepen vragen op ten aanzien van de interpretatie en waardering van de bevindingen van reeds gepubliceerde studies – zelfs van die waarbij gebruik is gemaakt van geavanceerde econometrische technieken (Gibson & Telhaj, 2006; Hanushek e.a., 2003; Hoxby, 2000). De kritiek gaat soms zover dat getwijfeld wordt aan de correcte schatting van effecten van al het eerder verrichte onderzoek (Webbink, 2005). Hoewel dit voortschrijdend inzicht uiteraard als een positieve ontwikkeling kan worden gezien, maakt het het tegelijkertijd uitermate lastig om de resultaten van de omvangrijke hoeveelheid studies die in de afgelopen decennia³ zijn uitgevoerd op hun waarde te schatten.

2.4 Oorzaken voor inconsistenties

2.4.1 Causaliteit

Een belangrijk probleem is dat de meeste studies uitgaan van een statische (i.c. cross-sectionele) onderzoeksopzet, wat onder meer vragen oproept met betrekking tot de causaliteit van de effecten (Schofield, 1989). De individuele en groepskenmerken (bv. individuele prestaties en daaruit afgeleide geaggregeerde gemiddelde groepsprestaties) zijn dan op een-en-hetzelfde moment verzameld en worden ook gelijktijdig in de analyse opgenomen. Bij dit type onderzoek speelt echter de kip of het ei-kwestie: persoon (of peer group) A beïnvloedt persoon B en vice versa. Steeds meer onderzoek wijst op reciproque effecten of zelfs ‘spiraal’ effecten via een ‘feedback loop’ (vgl. Marsh & Hau, 2003; Sacerdote, 2001; Wilkinson e.a., 2000). Een deel van deze ‘simultaneous equations bias’ (Moffitt, 2001) of dit ‘reflection problem’ (Manski, 1993) zou mogelijk opgelost kunnen worden door gebruik te maken van een dynamische, longitudinale opzet met op meerdere tijdstippen gemeten prestatie- en gedragsindicatoren (vgl. Guldemond, 1994). Binnen een dergelijke opzet kunnen dan bijvoorbeeld groeiscoringen (het verschil tussen huidige en vroegere toetsprestaties) worden berekend. Dat biedt dan overigens misschien wel een oplossing van het probleem op het individuele niveau, maar niet van dat op het (geaggregeerde) groepsniveau (Hanushek e.a., 2003; Peetsma e.a., 2006).

3 Met name na de publicatie van het beroemde Coleman rapport ‘Equality of educational opportunity’ (Coleman e.a., 1966) is de stroom onderzoek naar peer group effecten in het onderwijs goed op gang gekomen.

2.4.2 Niet-gemeten kenmerken

Een ander probleem is dat er veel ongemeten maar wel relevante kenmerken in het spel zijn ('omitted variables bias'; vgl. Hanushek e.a., 2003; Hoxby, 2000; Webbink, 2005). Gedacht kan worden aan mobiliteit van leerlingen (verhuizingen), veranderingen in de omstandigheden van de leerlingen (thuis), van de peers en van school en de reactie daarop van ouders met betrekking tot hun schoolkeuze. Opmerkelijk is ook dat er praktisch geen enkel onderzoek beschikbaar is waarin de dynamiek en kwaliteit van het klasgebeuren wordt verdisconteerd (Thrupp, Lauder & Robinson, 2002). Met name de inhoud, het tempo en de kwaliteit van de instructie, afzonderlijk maar ook in relatie tot specifieke kenmerken van de leerlingen, krijgen nauwelijks aandacht (Hallinan, 1990; Hattie, 2000). De rol van peers onderling (en de interactie met leerkrachten) is uitermate complex. In nagenoeg alle studies wordt echter volstaan met de meting van een beperkt aantal prestatiepunten (meestal taal- en reken/wiskunde prestaties) en achtergrondkenmerken van de leerlingen (sociaal milieu, etnische herkomst), die dan vervolgens op klas- of schoolniveau worden geaggregeerd (de 'black box' benadering). Behalve dat relevante kenmerken niet worden meegenomen, is er overigens ook nog twijfel over de kwaliteit van de wel gemeten kenmerken op zowel individueel als ook geaggregeerd niveau. Om dit probleem aan te pakken wordt wel gewerkt met zogenoemde instrumentele variabelen (Ammermüller & Pischke, 2006; Hanushek e.a., 2003), maar ook aan die benadering zitten veel haken en ogen. Er bestaat daarnaast onduidelijkheid of geaggregeerde kenmerken (analytische groepskenmerken, zoals het gemiddelde sociale milieu van een school) ook dezelfde betekenis hebben als een 'echte' collectieve eigenschap of theoretisch construct. Met de verschuiving in niveau treedt er ook een verschuiving in de betekenis ('shift of meaning') op van de variabele (Guldemond, 1994) en is er mogelijk sprake van 'aggregation bias' (Wilkinson, 2000).

2.4.3 Selectie

Een probleem dat vaak een rol speelt, en moeilijk geheel valt uit te sluiten, is dat van 'selection bias': de compositie van een bepaalde klas (of school) is het product van zowel (al-dan-niet vrijwillige) keuzes van ouders voor een bepaalde wijk en school en van de school zelf die voor een bepaalde verdeling van de leerlingen over de klassen kiest. Wanneer bijvoorbeeld hoogmilieu ouders actief op zoek gaan naar een goede school, kan er vanuit worden gegaan dat dergelijke ouders hun kinderen ook op andere manieren zullen ondersteunen en valt het effect van die – ongemeten – ouderlijke ambitie en ondersteuning moeilijk te onderscheiden van een echt peer group effect (vgl. Hoxby, 2000; Kling, Liebman & Katz, 2007; McEwan, 2003; Robertson & Symons, 2001; Stinebrickner & Stinebrickner, 2006; Winston & Zimmerman, 2003). Dit probleem speelt des te meer in landen met een omvangrijke privé-onderwijs sector (d.w.z. bekostigd door ouders zelf), zoals Engeland en de VS. Niet alleen is er in die landen dan sprake van een selectie-effect voor de betreffende scholen (en des te meer voor – vaak zeer selectieve – universiteiten), maar daardoor ook voor de overige scholen waarvoor (doorgaans minder bemiddelde) ouders eigenlijk niet kunnen kiezen. Mogelijke, maar doorgaans niet haalbare, oplossingen voor het oplossen van dergelijke selectie-effecten zijn het gebruik maken van (natuurlijke) experimenten met random toewijzing van leerlingen aan klassen/scholen (vgl. Hoxby, 2000; Kang, 2006; Sacerdote, 2001; Webbink, 2005). Andere opties om de gevolgen van selectie te reduceren zijn het via 'matching' leerlingen zoveel mogelijk gelijk te schakelen op relevante kenmerken of in de analyses te controleren voor (met selectie samenhangende) achtergrond-

kenmerken van de leerlingen en schoolspecifieke kenmerken, zoals de leerkracht-leerling ratio en de proportie allochtone leerlingen (Entorf & Lauk, 2006; Wilkinson e.a., 2000).

2.4.4 Aggregatieniveaus

Inconsistentie qua bevindingen kan ook het gevolg zijn van verschillen in gehanteerd aggregatieniveau. In sommige studies wordt uitgegaan van klasgemiddelden, in andere van jaar- of groepsgemiddelden en in weer andere van schoolgemiddelden. Soms wordt ook uitgegaan van buurtgemiddelden of – in het geval van onderzoek onder studenten – van gemiddelden van kamer- of ganggenoten (vgl. Ammermüller & Pischke, 2006; Gibson & Telhaj, 2006). Doorgaans wordt vanwege de kwantiteit en kwaliteit van de directe interacties het niveau van de klas als het meest relevante beschouwd (Schneeweis & Winter-Ebmer, 2007; Wilkinson e.a., 2000). Wanneer een school een zeer homogene samenstelling kent wat betreft leerlingachtergronden en -prestaties of wanneer de leerlingen op toevalsbasis over de klassen zijn verdeeld, is het waarschijnlijk een minder groot probleem om in plaats van het klasniveau uit te gaan van het jaar/groeps- of schoolniveau. Wanneer er echter binnen een school (niveau)klassen zijn gevormd op basis van prestaties en/of gedrag (en waarschijnlijk daarmee samenhangend: achtergronden) van de leerlingen, tast dat de validiteit en mogelijkheid tot generalisatie van de conclusies aan (vgl. Hanushek e.a., 2003; Marsh, 1984). In veel scholen zet men bijvoorbeeld allochtone leerlingen met taalproblemen bij elkaar in (kleinere) klassen met leerkrachten die zijn gespecialiseerd in de didactiek van de tweede-taalverwerving. Iets vergelijkbaars gebeurt met betrekking tot leerlingen met gedragsproblemen die bij elkaar worden gebracht in klassen met in de aanpak van die problematiek gespecialiseerde leerkrachten.

Hierboven is al aangegeven dat doorgaans het klasniveau voor de interacties het meest nabije en daarmee relevante niveau is. Wilkinson e.a. (2002) hebben in dit verband een conceptueel model ontwikkeld waarin er vanuit wordt gegaan dat peer group effecten werkzaam zijn via een geneste serie van hiërarchische lagen (school – klas – (sub)groep, binnen verschillende contexten). Veel effecten zijn daarom indirect en zijn/lijken kleiner naarmate ze verder van het concrete instructiegebeuren en de interacties in de klas afstaan; er treden daarbij in zekere zin dus verdunningseffecten op.

2.4.5 Ontwikkelingsverschillen

Wanneer het gaat om sociale vergelijkingsprocessen en het vormen van een referentiegroep/punt is er ook een samenhang met de ontwikkeling en leeftijd van de leerlingen. Heel jonge kinderen hebben nog een beperkt referentiekader (de directe omgeving); bij oudere kinderen is dat kader veel ruimer. Dit heeft tot consequentie dat de invloed van peer group ook per onderwijsfase zal verschillen. Dat er verschillen kunnen optreden tussen basis- en voortgezet onderwijs, heeft wellicht ook te maken met het veranderende karakter van de peer group. In het basisonderwijs zitten de leerlingen doorgaans in vaste groepen bij elkaar, terwijl de samenstelling van de peer group in het voortgezet onderwijs ten gevolge van uiteenlopende vakkenpakketten veel dynamischer is. Sommige onderzoekers wijzen er op dat het effect van klasgenoten naar verwachting sterker zal zijn in het basisonderwijs dan in het voortgezet- en hoger onderwijs (vgl. Guldmond, 1994; Manger & Eikeland, 1997; Marsh, 1987). Anderen gaan er echter van-

uit dat dergelijke effecten juist pas in de hogere jaren optreden (vgl. Hanushek e.a., 2003; Lubbers, 2004).

2.4.6 Kwaliteit interacties

Vanuit een theoretisch (en praktisch) oogpunt kunnen ook vragen worden gesteld over de omvang en diepgang van de interacties die ten grondslag liggen aan de peer group invloeden. Wanneer leerlingen, respectievelijk studenten bij elkaar zitten in een klas, groep, school of samen op een gang of kamer wonen, is dat meestal niet hun eigen keuze. Ze worden vooral in die situatie geplaatst door keuzes van hun ouders, beleid van scholen/universiteiten en (min of meer toevallige) omstandigheden – allemaal factoren waarop ze zelf niet zoveel invloed op kunnen uitoefenen. Hun samen-zijn is dan ook eerder afgedwongen dan op basis van vrijwilligheid tot stand gekomen. Leerlingen/studenten brengen dan weliswaar veel tijd samen door in dezelfde ruimte, maar veel van de interacties zullen ‘out of necessity’ plaats vinden. Daarmee kan dan de vraag worden gesteld hoe sterk die doorwerken op de attitudes en het gedrag van individuele leerlingen/studenten, kortom: kunnen er überhaupt wel sterke peer group effecten worden verwacht (Schneeweis & Winter-Ebmer, 2007)? Een andere vraag die – gelet op de specifieke aard van de relaties/interacties – opkomt, is of er zonder meer vanuit kan worden gegaan dat effecten van een bepaalde relatie in een bepaalde setting ook optreden in een andere setting (vgl. Stinebrickner & Stinebrickner, 2006).

2.5 Conclusies

De conclusie die op basis van bovenstaande kan worden getrokken is dat het inzicht in mogelijke haken en ogen van onderzoek naar peer group effecten in de loop van de jaren duidelijk is toegenomen. Echter, tegelijkertijd wordt duidelijk dat het onderzoek waarin aan alle genoemde bezwaren wordt tegemoet gekomen nog niet is uitgevoerd – voorzover dat uiteraard al mogelijk zou zijn. In dit verband kan afsluitend verwezen worden naar de door Thrupp, Lauder & Robinson (2002) opgestelde criteria, waaraan volgens deze onderzoekers goede studies naar compositie- en peer group effecten (ideaaltypisch) zouden moeten voldoen.⁴ Deze criteria hebben ze geformuleerd met het oog op sociaal-economische compositie-effecten, maar deze gelden mutatis mutandis uiteraard ook voor andere compositieperspectieven. Het betreft de volgende:

1. De steekproef van scholen moet de hele range qua sociaal-economische compositie vertegenwoordigen.
2. De analyse dient uiteenlopende beginkenmerken te bevatten, zoals eerdere prestaties.
3. De drie dimensies van schoolcompositie moeten vertegenwoordigd zijn: peer group processen, didactische processen, en organisatorische en management processen.
4. Er moet een combinatie van compositiekenmerken worden onderzocht, bijvoorbeeld de gemiddelde prestaties en het gemiddelde milieu van een klas.
5. Als mogelijk, zou de steekproef verschillende schooltypes moeten omvatten.
6. Als mogelijk, zou de studie longitudinaal moeten zijn.

4 Overigens merken zij ook op dat dergelijk kwantitatief onderzoek vergezeld zouden moeten gaan van meer fijnmazig beschrijvend of etnografisch onderzoek waarmee de subtielere processen die ten grondslag liggen aan compositie-effecten kunnen worden opgespoord.

7. Er moeten verschillende technieken voor het meten van compositie worden gehanteerd.
8. Er dienen robuuste maten worden genomen voor de compositiekenmerken, zoals sociaal milieu.
9. Er moeten extra gegevens worden verzameld om de interpretatie van effecten beter te kunnen toespitsen.
10. Er dient gebruik te worden gemaakt van multi-level analysetechnieken.

3 Verklaringen voor peer group effecten

3.1 Inleiding

Als het gaat om verklaringen voor peer group effecten kunnen zowel algemene als specifieke verklaringen worden gegeven. De specifieke verklaringen gelden met name voor compositie-effecten, terwijl de algemene effecten vooral gericht zijn op de ‘echte’ peer group effecten. De eerste groep van verklaringen is sterk vanuit een onderwijs-sociologisch perspectief geformuleerd, de tweede groep meer vanuit een sociaal- en ontwikkelingspsychologisch perspectief. In de volgende twee paragrafen krijgen beide groepen van verklaringen uitvoerig aandacht. Duidelijk wordt dat er veel overlap bestaat tussen de algemene en specifieke verklaringen. Vandaar dat in een afsluitende paragraaf wordt geprobeerd een verbinding te leggen tussen beide.

3.2 Specifieke verklaringen

In Nederland is het meeste, recente onderwijssociologische onderzoek naar compositie-effecten verricht vanuit het perspectief van etnische concentratie. Het is daarom niet verwonderlijk dat ook de verklaringen voor de gevonden effecten zich vooral binnen dit perspectief bewegen. Deze verklaringen vertonen overigens veel overlap met die welke eerder werden gegeven voor effecten van de concentratie van kinderen uit de lagere sociaal-economische milieus, i.c. kinderen op arbeidersscholen. In de praktijk zijn het geen op zichzelf staande fenomenen: etnische segregatie impliceert immers bijna steeds ook sociale segregatie, aangezien het grootste deel van de allochtonen uit de lagere sociale milieus komt. Daarmee samenhangend is er overigens ook weer een link met groeperingsvormen op grond van capaciteiten (‘abilities’). In verband met deze verklaringen zijn verschillende thesen geformuleerd (zie bv. Driessen, 2002; Driessen e.a., 2003, Tesser e.a., 1995; Westerbeek, 1999). Hierna volgt daarvan een korte beschrijving.

3.2.1 Taalcontact

Een veel genoemde verklaring richt zich op het taalaanbod. Verondersteld wordt dat op zwarte scholen kinderen minder in de gelegenheid zijn om in contact te komen met de Nederlandse taal en vanuit die contacten de Nederlandse taal te leren (de taalcontact-thesen). Op etnisch homogene zwarte scholen zal de meerderheidstaal de voertaal worden – zo is het idee –, terwijl op etnisch heterogene scholen de leerlingen zich terugtrekken in hun eigen subgroep met ieder hun eigen taal. Uit het onderzoek van Driessen & Bezemer (1999) onder islamitische scholen is gebleken dat op een aantal van deze scholen het regel is dat de kinderen onderling Nederlands dienen te spreken. Niettemin blijft het punt staan dat bij gebrek aan van huis uit Nederlands sprekende autochtone medeleerlingen het de vraag is wat de kwaliteit en het niveau van dat Nederlands is. In zekere zin komt deze verklaring overeen met die welke te geven valt voor eventuele negatieve effecten van sociale segregatie. Ook op concentratiescholen van arbeiderskinderen is het taalaanbod een cruciale factor; vergelijk in dit verband bijvoorbeeld het milieu-gerelateerde

onderscheid dat Basil Bernstein in de jaren vijftig al maakte tussen de ‘restricted’ en ‘elaborated code’. Kinderen op arbeidersscholen (c.q. scholen met veel autochtone achterstandskinderen) spreken vaak een streektaal of dialect en kennen een veel minder rijk aanbod qua ABN.

3.2.2 Netwerken

Een tweede verklaring ligt in de lijn van de netwerk-these. Op concentratiescholen vertonen de gezinsstructurele, -culturele en -pedagogische kenmerken sterke overeenkomsten: het betreft immers voornamelijk gezinnen van hetzelfde soort. Als het om concentratie van autochtone kinderen gaat, of van kinderen uit hogere statusgroepen, dan bevordert dit door de werking van sociale netwerken de onderwijskansen van de betreffende kinderen. Voor scholen met een concentratie aan kinderen uit achterstandsgroepen geldt echter het omgekeerde. Op die scholen zijn de mogelijkheden tot sociale stijging via netwerken van ouders relatief beperkt. Met name geldt dit voor veel allochtone gezinnen als gevolg van taalbarrières en gebrek aan inzicht in het Nederlandse onderwijssysteem. Overigens voorspelt de theorie over de functionele gemeenschappen juist een positieve uitwerking van concentratie, voor alle leerlingen ongeacht sociale status.

3.2.3 Nivellering

In een derde verklaring ten slotte, wordt gewezen op nivelleringseffecten (Tesser e.a., 1995) die ten dele weer samenhangen met leerkrachtverwachtingen (vgl. Van der Hoeven-Van Doornum, 1990; Van der Hoeven-Van Doornum, Voeten & Jungbluth, 1989). Volgens Tesser e.a. lijkt het er op dat op concentratiescholen vooral de betere leerlingen worden geremd, hetgeen mogelijk wordt veroorzaakt doordat leerkrachten niveau en tempo van de behandeling van de leerstof aanpassen aan het gemiddeld lage niveau van de leerlingen. Daarnaast is de taakverzwaring op concentratiescholen zodanig, dat het de vraag is of leerkrachten nog gelegenheid hebben de betere leerlingen voldoende te stimuleren.

3.3 Algemene theorieën

Voor een meer algemene verklaring van invloeden van peer group effecten kan aansluiting gezocht worden bij vergelijkingstheorieën en bij netwerktheorieën. Deze zijn met name geformuleerd vanuit een sociaal- en ontwikkelingspsychologische invalshoek. Voor uitgebreidere verhandelingen wordt verwezen naar bijvoorbeeld Davis (1966), Guldmond (1994), Marsh (1984), Marsh & Parker (1984) en Marsh & Tau (2003).

3.3.1 Vergelijkingstheorieën

Een uitvoerig overzicht van de effecten van de cognitieve klassamenstelling en de hiermee samenhangende verklaringen wordt gegeven door De Vries (1994). Centraal staan de verschillen tussen cognitief-homogene en cognitief-heterogene groepen (‘ability grouping’; ‘tracking’; ‘streaming’). Het bestaan hiervan hangt samen met keuzes van de school zelf (voorkeur voor een bepaalde manier van groeperen, bijvoorbeeld via vorming van niveaugroepen), maar ook

met de kenmerken van het leerlingenpubliek. In een heterogene klas wordt een leerling geconfronteerd met leerlingen van een ander cognitief niveau. Volgens voorstanders van heteroogeen groeperen zou de zwakke leerling zich dan kunnen optrekken aan de betere leerlingen en kunnen profiteren van de hogere eisen die de leerkracht stelt. Tegenstanders van heteroogeen groeperen vinden echter dat zowel begaafde als minder begaafde leerlingen nadeel ondervinden van deze groeperingsvorm. Begaafde leerlingen zouden de stimulans van andere begaafde leerlingen missen en daardoor onder hun niveau gaan presteren. De minder begaafde leerling zou door het feit dat hij altijd slechter presteert dan de meeste andere leerlingen gedemotiveerd raken.

Deze strijdige veronderstellingen vinden hun oorsprong in verschillende theorieën die gangbaar zijn voor de verklaring van deze verschijnselen. De eerste theorie wordt gevormd door de referentiegroepen-theorie van Richer (1976). Hierin worden twee vormen van referentieprocessen onderscheiden. Volgens de *normatieve* referentietheorie bepaalt de groep de normen en standaarden waaraan de persoon moet voldoen. Bij de *comparatieve* referentietheorie vormt de standaard van de groep een vergelijkingspunt dat de persoon gebruikt bij het evalueren van zichzelf en van anderen. In een heterogene klas kan de zwak presterende leerling de groep hoogpresterende leerlingen als normatieve referentiegroep kiezen en door de in die groep heersende normen en waarden ten aanzien van leerprestaties positief beïnvloed worden. De zwak presterende leerling kan echter ook de groep hoog presterende leerlingen als een comparatieve referentiegroep kiezen, hetgeen kan leiden tot een negatief zelfbeeld en relatieve deprivatie. Of een zwakke leerling de normatieve of de comparatieve vergelijkingsgroep kiest is volgens Richer afhankelijk van de mate waarin mogelijkheden tot opwaartse mobiliteit gezien worden. Als opwaartse mobiliteit als haalbaar gezien wordt, zullen normatieve referentiegroepsprocessen optreden, maar als dit niet het geval is, zullen comparatieve processen optreden.

Naast de groepsvergelijkingsprocessen spelen ook vergelijkingen met individuen een rol. Vanuit de behoefte aan zelfverbetering zal iemand zich waar het bekwaamheden betreft vergelijken met iemand die enigszins superieur is. Met name in situaties die als bedreigend ervaren worden, bestaat echter ook de behoefte het eigen zelfbeeld te beschermen door neerwaartse vergelijking.

Welk vergelijkingsmotief op een bepaald moment het sterkst is, is afhankelijk van de situatie. In een heterogene groep hebben leerlingen meer mogelijkheden zich op- en neerwaarts te vergelijken dan in een homogene groep. Voor de hoogst presterende leerlingen ontbreekt de mogelijkheid tot opwaartse vergelijking, terwijl voor de laagst presterende leerlingen in een heterogene groep de mogelijkheid tot neerwaartse vergelijking ontbreekt. Bovendien worden de laagst presterende leerlingen voortdurend geconfronteerd met slechte prestaties, wat tot een negatief zelfbeeld en uiteindelijk nog slechtere prestaties kan leiden.

Onderzoek naar de effecten van vergelijkingsprocessen is vooral gericht op het zelfbeeld dat ontstaat ten gevolge van de vergelijking. Het zelfbeeld blijkt echter vaak geen belangrijke determinant van schoolprestaties. De oorzaak hiervan zou kunnen zijn dat het zelfbeeld bepaald wordt door de relatieve positie van de leerling in de klas. Een zwakke leerling zou in een goede klas een negatiever zelfbeeld kunnen hebben dan in een zwakke klas.

Uit de referentiegroepentheorie en uit de sociale vergelijkingstheorie wordt duidelijk dat zowel zwak als goed presterende leerlingen in een heterogene groep zowel positief als negatief beïnvloed kunnen worden door vergelijking met andere leerlingen. De richting van de effecten blijkt

op grond van de theorieën niet voorspeld te kunnen worden. Volgens De Vries is ook de rol van de leerkracht in dit verband van belang. Volgens De Vos (1986) spelen de prestaties van medeleerlingen niet zozeer een rol via een directe vergelijking met de medeleerlingen, maar vooral via de invloed die zij uitoefenen op de standaard die de leerkracht hanteert. Het klasgemiddelde zou van invloed zijn op de standaard die de leerkracht aanlegt. Een groep met een hoog gemiddeld prestatieniveau zou gunstig zijn voor de laag presterende leerlingen, mits de afstand van de groepsstandaard tot het eigen niveau niet te groot is. Een groep met een laag gemiddeld prestatieniveau zou ongunstig zijn voor de hoog presterende leerlingen.

Ook de verwachtingen van leerkrachten van de leerlingen spelen een rol. Hoewel Jungbluth (1985, 2003) en Van der Hoeven-Van Doornum (1990) constateren dat de verwachtingen van leerkrachten ten aanzien van de prestaties van de leerlingen sterk milieu-gebonden zijn, lijkt de invloed van objectief verkregen informatie over de prestaties in het heden en het verleden volgens Dar (1981) groter dan de invloed van 'labeling' op grond van kenmerken als sociaal milieu, sekse en etnische afkomst.

Uit onderzoek van Guldmond (1994) bleek dat de sociaal-hiërarchische structuur van de klas een veel groter effect had op de individuele leerprestaties dan het prestatieniveau van de klasgenoten. Populaire leerlingen die goede prestaties leverden gaven met hun voorbeeldfunctie een positieve impuls aan de schoolvorderingen van hun klasgenoten. De klas diende dus vooral gezien te worden als normatieve referentiegroep. Ook in termen van comparatieve referentie vervulde de groep een functie, maar dit was in veel mindere mate het geval dan de normatieve referentie. Verder bleek dat homogeen groeperen het groepseffect versterkte, terwijl heterogeen groeperen het effect afzwakte. Vooral de laagpresteerders binnen klassen waar homogeen werd gegroepeerd waren hiervan de dupe. Vergelijkingsprocessen binnen de klas worden in dat geval beperkt tot medeleerlingen van een vergelijkbaar prestatieniveau. Door het instellen van niveaugroepen geeft de leerkracht aan dat de eisen en verwachtingen niet voor elke groep hetzelfde zijn. De leerlingen stellen zich vervolgens tevreden met het leveren van prestaties die in overeenstemming zijn met het gangbare prestatieniveau binnen de subgroep. Dit werkt uiteraard vooral negatief voor de zwakke presteerders. Aangezien de groep vooral een normatieve functie vervult, lijken de referentieprocessen vooral conformistisch van aard, in die zin dat een leerling binnen de subgroep niet uit de toon wil vallen.

Samenvattend blijkt uit de verschillende theorieën dat er van de klascompositie tegengestelde effecten kunnen uitgaan, en dat de afstand tot de vergelijkingsgroep, de vergelijkingssander of de standaard die de leerkracht aanlegt daarbij een belangrijke rol speelt.

3.3.2 Netwerkteorieën

In het ecologische ontwikkelingsmodel van Bronfenbrenner (1979) wordt de omgeving als een belangrijke factor beschouwd voor onderwijs en ontwikkeling. Rond het kind bestaat een systeem van concentrische sferen zoals het gezin, de groep leeftijdsgenoten, en de vriendengroep van de ouders. De vorming van het kind komt niet los van de bredere omgeving tot stand, maar wordt direct en indirect beïnvloed door relaties tussen de sferen en de contexten waarin deze zijn opgenomen (zie ook De Groot & Scholte, 1990). Het ecologische ontwikkelingsmodel voorspelt onder meer dat de condities voor ontwikkeling gunstiger zijn naarmate het aantal rela-

ties tussen de sferen rond het kind groter is, het wederzijds vertrouwen tussen de sferen sterker is, en overeenkomende doelen worden nagestreefd. Naarmate gezin, subcultuur en school meer in elkaars verlengde liggen ontstaat er een effectieve sociale context voor opvoeding en onderwijs. Vorming en ontwikkeling profiteren volgens Bronfenbrenner van congruente settings. Een effectieve pedagogische omgeving omvat ook de culturele oriëntaties die gezin, gemeenschap en school verbinden. Het systeem van waarden en normen en de opvattingen over onderwijs reflecteren de sociale en culturele omgeving van het gezin en het geheel aan waarden en gedragingen die daarmee verbonden zijn (Dijkstra, 1999).

Hoewel de ecologische ontwikkelingstheorie veel overeenkomsten vertoont met de functionele gemeenschappentheorie, strekt het ecologische model volgens Dijkstra (1997) verder dan de eerstgenoemde. Het ecologische model vestigt de aandacht op het belang van de sociale context voor de integratiefunctie van het onderwijs. Ontwikkeling van sociale competenties, de vorming van opvattingen en gedragsstandaarden, de overdracht van waarden en normen of de ontwikkeling van het vermogen tot morele reflectie staan niet los van de setting waarin opvoeding en onderwijs plaatsvindt. Congruente settings hebben daarmee tevens betrekking op culturele consistentie, de waardengemeenschap met een set gemeenschappelijke waarden en normen. Deze waardenconsistentie heeft niet alleen betrekking op levensbeschouwelijke aspecten. Voor ouders relevante oriëntaties zijn deel van de leefwereld waarin het gezin is opgenomen en kunnen evenzeer betrekking hebben op het linguïstische, etnische, sociaal-culturele of leefstijldomein. Dijkstra bespreekt een aantal onderzoeken die aantonen dat de leefstijlen van de oudergroepen rond scholen systematisch lijken te verschillen en in onderwijsvoorkeuren worden weerspiegeld. Ook lijken onderwijsoriëntaties opgevat te kunnen worden als onderdeel van meer algemene waardensystemen. Zij liggen ingebed in de culturele leefwereld van ouders.

3.4 Dwarsverbanden

De hierboven genoemde theorieën vormen een bruikbaar kader voor de verklaring van de verschillende peer/compositie-effecten. De referentie- en vergelijkingstheorieën zijn weliswaar alleen uitgewerkt voor groepen die meer of minder cognitief homogeen zijn, maar ze zijn ook bruikbaar als verklaring voor etnische en sociale concentratie: deze impliceert immers ook een bepaalde verdeling naar cognitieve vaardigheden. De netwerktheorieën bieden houvast voor de analyse van concentratie-effecten die samenhangen met sociaal milieu en etnische herkomst.

Aan deze laatste theorieën kunnen voor deze invalshoeken gemeenschappelijke verwachtingen ontleen. Die vinden we in het begrip ‘congruente settings’. Naarmate gezin, subcultuur en school meer in elkaars verlengde liggen ontstaat er volgens deze theorieën een effectievere context voor opvoeding en onderwijs. Congruente settings zijn zowel van belang voor de sociale ontwikkeling als voor het verloop van het onderwijsleerproces op school (Kassenberg, De Vos & Dijkstra, 2000). Dijkstra & Veenstra (2000) hebben deze gedachte onderzocht voor de invalshoek religie. Uit hun studie bleek dat er inderdaad sprake was van een effect van religieuze congruentie op prestaties. Naarmate het levensbeschouwelijk profiel van de school (i.c. de dominante godsdienst op school) en de religieuze achtergrond van de leerling meer overeenstemden, werden betere prestaties behaald. Analoog hieraan is er ook een link met etnische en sociaal-economische congruentie. Dat impliceert aandacht voor het al dan niet behoren tot de dominante groep in een school. Kassenberg e.a. (2000) hebben de congruentievraag onderzocht aan

de hand van het concept ‘belonging’, en daarbij speciale aandacht besteed aan de variabele motivatie. ‘Relatedness’ wordt in deze visie beschouwd als de kern van de motivatie. Daarmee wordt bedoeld op de aanwezigheid van veilige en bevredigende relaties met anderen in iemands sociaal milieu. Hierbij staat ‘belonging’ – het gevoel er bij horen, het toebehoren tot de groep van gelijken – centraal. Het gevoel erbij te horen zou in de schoolse context een belangrijke factor zijn bij het tot stand komen van hoge verwachtingen en van betrokkenheid bij schoolse taken. Iemands motivatie om te presteren is een functie van diens verwachtingen met betrekking tot succes en de waarde die succes voor hem of haar heeft en heeft op zijn beurt effect op de prestaties (Goodenow, 1993; Atkinson, 1964). Op basis van dit theoretisch kader veronderstelden Kassenberg e.a. (2000) dat leerlingen op scholen met meer leerlingen met dezelfde etniciteit, meer medeleerlingen hebben die ze gemakkelijk begrijpen en die hen gemakkelijk begrijpen en sterker het gevoel hebben erbij te horen. Anderzijds verwachtten zij dat op een school met veel verschillende etnische groepen de mogelijkheden tot het realiseren van wederzijds vertrouwen, begrijpen en waarderen bemoeilijkt. Soortgelijke verwachtingen hebben zijn ten aanzien van religie. Inderdaad blijkt uit hun analyses dat er verschillen zijn in het gevoel erbij te horen (‘belonging’) met betrekking tot beide kenmerken.

Dit veralgemeniserend, zou kunnen verwacht dat wanneer een leerling in een groep zit waarvan de meerderheid in bepaalde opzichten (etnisch, sociaal, linguïstisch, religieus, of naar sekse of capaciteiten) aansluit bij zijn of haar eigen achtergronden, hij of zij zich beter op zijn gemak voelt, veiligheid ervaart en betere sociale relaties zal hebben. Dit is in principe bevorderlijk voor de verwachtingen van en betrokkenheid bij de schoolse taken, hetgeen positieve effecten heeft op de prestaties. Dit laatste is echter niet per definitie het geval. Het toebehoren tot een groep kan, zo leren de referentiegroepentheorie en de sociale vergelijkingstheorie, ook negatief uitwerken op prestaties, afhankelijk van de motivatie en inzet van die groep. De richting van die effecten kan niet eenduidig worden voorspeld. De rol van de leerkracht is volgens deze theorie ook van belang, omdat de prestatie-standaard die de leerkracht aanlegt mede wordt bepaald door het klasgemiddelde. Leerkrachtverwachtingen en -gedrag (nivellerings-effecten) zijn bovendien al eerder naar voren gekomen als mogelijke verklaringen voor vooral sociale en etnische compositie-effecten.

4 Resultaten

4.1 Inleiding

Binnen het onderzoek naar peer group effecten kunnen verschillende dimensies/perspectieven worden onderscheiden. Veel onderzoek is gericht op effecten die het gevolg zijn van de indeling van scholen en klassen op basis van de capaciteiten van de leerlingen (ability). Ook is veel onderzoek gedaan naar combinatieklassen op grond van leerjaar dan wel leeftijd, ('multigrade', resp. 'multi-age'). Twee volgende dimensies die onderling sterk samenhangen zijn de schoolcompositie wat betreft etniciteit (het aandeel allochtonen) en sociaal milieu (meestal het gemiddelde opleidingsniveau van de ouders). Die samenhang wordt veroorzaakt door het feit dat – zeker in Nederland – een belangrijk deel van de allochtonen uit de lagere sociaal-economische milieus komt. Een laatste perspectief is dat van de sekse van de leerlingen, waaraan vorm wordt gegeven via aparte meisjesklassen versus gemengde klassen ('single sex' vs. co-educatie) (zie o.m. Hattie, 2002; Thrupp, Lauder & Robinson, 2002; Wilkinson e.a., 2000; 2002).

In onderzoek naar peer group effecten kan een onderscheid worden gemaakt tussen compositionele effecten (van de groepssamenstelling, zoals het gemiddelde prestatieniveau of het aandeel allochtonen, op individuele prestaties) en frog pond effecten (differentiële effecten naargelang de relatieve positie van een leerling in een groep). De veronderstelling is dat een compositioneel effect een eigenstandige (d.w.z. los van individuele effecten) bijdrage levert aan de verklaring van verschillen in bijvoorbeeld individuele prestaties. Hoe een dergelijk effect geïnterpreteerd moet worden is vervolgens een open vraag. Valt het bijvoorbeeld toe te schrijven aan de kwaliteit van het onderwijs, aan de interacties tussen leerlingen (het 'echte' peer group effect) of aan slecht of niet gemeten kenmerken (Guldmond, 1994; Hanushek e.a., 2003). Als het gaat om dergelijke compositie-effecten wordt ook wel gesproken over lineaire, symmetrische baseline modellen (Hoxby, 2000). Het frog pond effect vormt eigenlijk een uitbreiding op het compositie-effect, waarvoor het noodzakelijk is niet alleen het compositiekenmerk zelf (bv. de gemiddelde prestaties van een klas) in de analyse op te nemen, maar ook de relatie daarvan met de relatieve positie die de leerling inneemt ten opzichte van dat compositiekenmerk (bv. boven of onder het klasgemiddelde), dus de (statistische) interactie tussen beide variabelen (Ammermüller & Pischke, 2006). Bij die (niet-lineaire) effecten gaat het vaak om 'equity' effecten, namelijk om de vraag of de winst van bijvoorbeeld homogeen groeperen gelijk over verschillende subgroepen is verdeeld, bijvoorbeeld allochtoon versus autochtoon of high ability versus low ability (Hattie, 2002). In het geval van frog pond effecten spreekt men ook wel over niet-lineaire, asymmetrische of differentiële modellen (Winston & Zimmerman, 2003).

In hoofdstuk 2 zijn uiteenlopende theoretische en methodologische kwesties aan de orde gesteld. Evident is dat veel onderzoek niet (in alle opzichten) voldoet aan met name methodologische eisen (o.m. Thrupp, Lauder & Robinson, 2002; Webbink, 2005). Dat heeft uiteraard consequenties voor de validiteit van de conclusies. Het is evident dat er de laatste jaren veel vooruitgang is geboekt in het denken over eisen die aan dergelijk onderzoek zouden moeten worden gesteld. Dat wil echter niet meteen zeggen dat het optimale onderzoek dan ook is, respectievelijk

lijk kan worden uitgevoerd. Deze overweging dient bij de interpretatie van de bevindingen wel mee te worden genomen. Gelet op de methodologische progressie die is gemaakt, kan in zijn algemeenheid worden gesteld dat latere studies in dat opzicht beter zijn dan eerdere studies.

Hierna volgt een overzicht van de resultaten van kwantitatieve studies naar peer group effecten⁵ op onderwijsprestaties. Daarbij is een indeling aangehouden gebaseerd op de eerder genoemde dimensies/perspectieven. Geprobeerd is de resultaten zoveel mogelijk te kwantificeren. Vanwege het feit dat veel onderzoekers zich beperken tot het presenteren van uiteenlopende effecten zonder daarbij steekproefkarakteristieken te vermelden, is het zeer lastig, zo niet onmogelijk, om de resultaten over de onderzoeken heen te vergelijken. Doorgaans wordt in de publicaties gesteld dat het om al-dan-niet ‘significante’ effecten gaat, zonder daarbij een harde kwalificatie te geven van de grootte/omvang/relevantie van het effect. Omdat de mate van significantie sterk afhangt van de omvang van de steekproef die onderzocht is en in dit type onderzoek juist vaak omvangrijke steekproeven worden geanalyseerd, zegt echter significantie niet veel: bij zeer grote steekproeven zijn al gauw alle effecten significant, maar is het maar zeer de vraag of ze ook enige onderwijskundige of maatschappelijke relevantie hebben (zie Driessen & Doesborgh, 2003). Ook in eerdere overzichtsstudies ontbreekt het vaak aan een kwantificering van de aangetroffen effecten. Ondanks dit handicap wordt geprobeerd trends in de bevindingen te beschrijven (vgl. Schofield, 1989).

4.2 Ability grouping

In het verleden zijn regelmatig meta-analyses uitgevoerd naar peer group effecten. Hattie (2002) en Wilkinson e.a. (2000) hebben de resultaten van een aantal van dergelijke studies samengevat.⁶ Tabel 4.1 geeft een bewerkt overzicht van de meta-meta-analyses waarbij is gekeken naar effecten van ‘ability grouping’ op prestaties. Het gaat daarbij om het plaatsen van leerlingen in homogene (sub)groepen op basis van hun capaciteiten of eerdere prestaties. Bij deze studies worden in principe drie ability niveaus onderscheiden, namelijk laag, midden en hoog.

Tabel 4.1 – Samenvatting resultaten meta-analyses ability grouping

meta-analyse	fase	aantal studies	aantal effecten	effectgrootte
Kulik & Kulik (1984)	PO	31	31	0,19
Kulik & Kulik (1985)	PO/VO	85	85	0,15
Kulik & Kulik (1982/1987)	VO	52	52	0,10
Slavin (1987)	PO	14	17	0,00
Slavin (1990)	VO	29	29	-0,02
Noland & Taylor (1986)	PO/VO	50	570	-0,08
totaal	PO/VO	261	784	-0,03

5 In vrijwel alle studies wordt over ‘effecten’ gesproken, alhoewel het strikt genomen in de meeste gevallen eerder om samenhangen gaat dan echte causale effecten (zie ook hoofdstuk 2).

6 Wilkinson e.a. (2000) komen op basis van hun analyse van alle dimensies tot de conclusie dat er (zeer) kleine compositie-effecten zijn; gemiddeld over alle dimensies heen bedraagt het effect (maximaal) 0,10.

De tabel laat zien dat er in totaal 261 (afzonderlijke) studies zijn beoordeeld, waarbinnen in totaal 784 effecten zijn berekend. In een aantal studies zijn dus meerdere effecten berekend, bijvoorbeeld apart voor taal, rekenen en lezen en dat voor verschillende leerjaren. In de laatste kolom staan de effectgroottes (ES: 'effect size'); deze geven een indruk van de relevantie van een effect. Wat de interpretatie van een effectgrootte betreft, wordt doorgaans de indeling van Cohen (1988) overgenomen, die een coëfficiënt van 0,20 als 'klein' bestempeld, die van 0,50 als 'middelmatic' en die van 0,80 als 'groot'. Deze classificatie dient echter niet al te rigide worden opgevat (Strand, Deary & Smith, in press). Hattie (1999) wijst er op dat een effect moet worden afgezet tegen het gemiddelde effect van onderwijskundige interventies, en dat effect is 0,40. Een positief effect is in het voordeel van homogene ability groepen en een negatief effect is in het voordeel van heterogene groepen. Uit het overzicht blijkt dat de gemiddelde effectgrootte varieert van +0,19 tot -0,08. Het totale (gewogen) effect bedraagt -0,03, met andere woorden het groeperen van leerlingen op basis van hun 'abilities' heeft geen enkel effect op hun prestaties.

Voor zover mogelijk is er binnen die meta-analyses ook nagegaan of de effecten verschillen voor onderscheiden ability groepen. Het kan immers zo zijn dat sommige groepen wel profiteren van de groepssamenstelling, terwijl anderen dat niet doen of er zelfs nadeel van ondervinden. De resultaten van die uitsplitsing staan in Tabel 4.2.

Tabel 4.2 – Samenvatting resultaten meta-analyses ability grouping, naar ability niveau (effectgroottes)

meta-analyse	ability niveau		
	laag	midden	hoog
Kulik & Kulik (1984)			0,49
Kulik & Kulik (1985)	0,14		0,33
Kulik & Kulik (1982/1987)			0,33
Slavin (1987)	0,02	0,02	0,03
Slavin (1990)	-0,02	-0,08	0,01
Noland & Taylor (1986)	0,18		-0,16
Mosteller e.a. (1996)	0,08	-0,04	-0,06

De effectgroottes in deze tabel maken duidelijk dat er hooguit effecten van enige relevantie optreden voor de hoge ability groep, ofschoon ook deze effecten beperkt zijn. Voor de andere twee ability groepen zijn de effecten triviaal.

De algemene conclusie die op basis van deze meta-analyses kan worden getrokken is dat homogene groeperen van leerlingen op basis van hun ability hooguit zeer minimale effecten heeft, die bovendien ook nog soms positief en soms negatief uitpakken.

4.3 Combinatieklassen

Wat betreft combinatieklassen wordt meestal een onderscheid gemaakt naar 'multigrade' en 'multi-age' klassen. Multigrade klassen zijn klassen waarin leerlingen van meer dan één jaar-

groep bij elkaar in dezelfde klas zitten en les krijgen van dezelfde leerkracht. Deze klassen worden bijvoorbeeld op kleine scholen gevormd vanwege geringe aantallen leerlingen per jaargroep. In multi-age klassen worden leerlingen van verschillende leeftijden bij elkaar gezet omdat men er pedagogische voordelen in ziet (bv. op Jenaplan-scholen).

In Tabel 4.3 worden de resultaten van de meest omvattende reviewstudie op dit gebied, die van Veenman (1995), gepresenteerd. Hij analyseerde de resultaten van in totaal 56 studies in het primair onderwijs. In de tabel wordt een onderscheid gemaakt naar organisatievorm, namelijk multigrade en multi-age. De tabel geeft het aantal studies met verschillen tussen combinatieklassen en reguliere klassen, en de mediane effectgrootte.

Tabel 4.3 – Samenvatting resultaten meta-analyse combinatieklassen, naar organisatievorm

organisatievorm	vakgebied	aantal studies met verschil			effectgrootte
		significant positief	geen verschil	significant negatief	
multigrade	lezen	2	19	4	-0,03
	rekenen/wiskunde	2	17	4	-0,01
	taal	2	16	3	0,00
	samengesteld	2	6	1	-0,02
	totaal	4	28	6	0,00
multi-age	lezen	0	8	0	-0,03
	rekenen/wiskunde	0	7	2	-0,10
	taal	0	2	1	-0,02
	samengesteld	1	0	1	-0,24
	totaal	1	9	1	-0,03

De tabel maakt duidelijk dat verreweg de meeste studies geen significante verschillen laten zien. De (mediane) effectgroottes lijken voornamelijk in het nadeel van combinatieklassen, maar gelet op de sterkte van het effect kan geconcludeerd worden dat combinatieklassen niet beter en ook niet slechter presteren dan reguliere klassen.

Veenman (1995) heeft de gegevens ook nog uitgesplitst naar sociaal milieu om na te gaan of er wellicht een samenhang bestaat met de leerlingenpopulatie van de scholen. Deze resultaten staan in Tabel 4.4 weergegeven.

Tabel 4.4 – Samenvatting resultaten meta-analyse combinatieklassen, naar organisatievorm en sociaal milieu

organisatievorm	sociaal milieu	aantal studies met verschil			effectgrootte
		significant positief	geen verschil	significant negatief	
multigrade	hoog	0	4	1	-0,06
	midden	0	8	0	0,03
	laag	1	4	1	-0,10
	gemengd	3	5	4	-0,02
	onbekend	0	6	0	-0,00
multi-age	hoog	1	2	0	0,01
	midden	0	5	1	-0,01
	laag	0	1	0	-0,10
	gemengd	-	-	-	-
	onbekend	0	1	0	-

De naar sociaal milieu uitgesplitste resultaten wijken niet af van die van de totale steekproef en laten ook geen noemenswaardige verschillen zien tussen de onderscheiden milieus.

Naar aanleiding van Veenmans meta-analyse is er een ware polemieek ontstaan. Mason & Burns (1996) stellen dat Veenman in zijn analyse geen rekening heeft gehouden met selectiebias; volgens Mason & Burns hebben multigrade klassen weliswaar betere leerlingen, maar is in die klassen tegelijkertijd ook de kwaliteit van de instructie lager. Hierdoor zou er eigenlijk een klein negatief effect van -0,10 verwacht kunnen worden. Naar aanleiding daarvan heeft Veenman een kritische heranalyse van zijn data uitgevoerd (met een uitbreiding met enkele studies). De resultaten bleven echter onveranderd: er zijn wat betreft de cognitieve uitkomsten geen verschillen tussen combinatieklassen en reguliere klassen. Mason & Burns (1997) houden vervolgens op basis van kwalitatief onderzoek (observaties, interviews) echter vast aan hun hypothese dat combinatieklassen kleine, negatieve effecten (zouden moeten) hebben. Veenman (1997) vindt echter op grond van een nieuw uitgevoerde best-evidence synthese en meta-analyse opnieuw geen empirisch bewijs voor de veronderstelde selectie-effecten.

De eindconclusie luidt daarmee dat er qua cognitieve uitkomsten geen verschillen zijn tussen combinatieklassen en reguliere klassen, en dat er daarbij ook geen verschillen optreden naar het sociale herkomstmilieu van de leerlingen op die twee typen klassen.

4.4 Single-sex klassen/scholen

Een volgende invalshoek wat betreft de samenstelling van de leerlingenpopulatie is die van single-sex klassen en scholen. In de Nederlandse onderwijspraktijk komt dit nauwelijks voor en is daarmee op zich – vanuit praktisch oogpunt – minder relevant. Een uitzondering op de algemene regel is dat op sommige islamitische basisscholen in Nederland meisjes in aparte klassen les krijgen of binnen een klas in aparte groepjes/rijen worden gezet; met name gebeurt dit in de hogere jaargroepen (Driessen & Bezemer, 1999). Daarnaast wordt op enkele scholen voor voortgezet onderwijs geëxperimenteerd met aparte meisjesklassen voor de bètavakken.

In Tabel 4.5 worden de resultaten van een meta-analyse van Mael e.a. (2005) samengevat. In totaal selecteerden zij 40 studies met 112 effecten die aan hun methodologische criteria voldeden. De studies hadden betrekking op zowel het primaire als voortgezette onderwijs. Ze onderscheidden tussen huidige en langetermijn cognitieve opbrengsten, met daarbij een uitsplitsing naar vier mogelijke resultaten: pro single sex, pro co-educatie, geen verschil tussen single sex en co-educatie, en gemengde/tegengestelde effecten. In de tabel staat per opbrengst het percentage studies dat tot een van de vier genoemde resultaten leidt. Overigens dient bij die percentages enige terughoudendheid betracht te worden, aangezien de aantallen waarop ze betrekking hebben niet erg groot zijn.

Tabel 4.5 – Samenvatting resultaten meta-analyse single sex versus co-educatie, naar onderwijsfase

fase en effectmaten	aantal studies	effecten			
		% pro single sex	% pro co-educatie	% geen verschil	% gemengd resultaat
<i>PO en VO</i>					
prestaties alle vakken	9	67	11	22	0
prestaties rekenen/wiskunde	14	22	0	56	22
prestaties 'science'	8	25	0	62	13
prestaties taal	10	30	0	70	0
rapportcijfers	1	0	0	100	0
prestaties 'social studies'	1	100	0	0	0
subtotaal	43	35	2	53	10
<i>na het VO</i>					
prestaties	2	50	0	50	0
voltooien 'college'	1	0	0	100	0
studeren aan universiteit	1	0	0	100	0
subtotaal	4	25	0	75	0

De tabel laat zien dat er nauwelijks effecten zijn ten gunste van co-educatie. Meer dan de helft van de studies in het PO en VO en driekwart van de studies die betrekking hebben op de fase daarna laten geen verschillen zien. Een derde van de studies in het PO en VO wijzen op effecten in het voordeel van single sex klassen. Er is dus enige ondersteuning voor single sex klassen. Van belang is overigens dat dit positieve effect volgens Mael e.a. (2005) voor zowel jongens als voor meisjes geldt.

Een probleem bij studies naar single sex klassen is dat er vaak sprake is van selectie-effecten, zowel wat betreft de leerlingen als leerkrachten. Kelly (1996) vond dat gescheiden scholen voor zowel meisjes als jongens betere prestaties opleveren dan gemengde scholen, maar hier was geen rekening gehouden met instroomverschillen. In een studie van Marsh & Rowe (1996) werd daar wel voor gecontroleerd. Deze onderzoekers vonden op geen van de effectmaten verschillen tussen meisjes in meisjes- en gemengde klassen; en dat gold eveneens voor jongens. Ook andere onderzoekers die corrigeerden voor verschillen in instroom vonden weinig tot geen effecten voor de prestaties van jongens of meisjes (o.a. Daly, 1996; Harker, 2000; zie ook Van Langen & Driessen, 2006; Driessen & Van Langen, 2007). Onlangs heeft Van de gaer (2006) een nieuw

element toegevoegd aan de discussie over het effect van gescheiden onderwijs dat de zaak er niet eenvoudiger op maakt. Volgens haar analyses, waarin terdege rekening is gehouden met instroomkenmerken en andere relevante variabelen, presteren jongens beter bij Nederlands in gemengde klassen dan in jongensklassen, maar presteren meisjes beter bij wiskunde in meisjes-scholen dan in gemengde scholen.

In zijn algemeenheid kan dus worden geconcludeerd dat er nauwelijks of geen bewijs is voor compositie-effecten die het gevolg zijn van plaatsing van een leerling in een single sex of gemengde klas.

4.5 Compositie qua milieu en etniciteit

4.5.1 Desegregatie in de VS

Een specifieke invalshoek is die van onderzoek naar effecten van desegregatie dat in de VS is uitgevoerd. Vóór 1954 was scheiding van leerlingen in aparte scholen op basis van hun ras wettelijk vastgelegd. Naar aanleiding van een gerechtelijke uitspraak ('Brown v. Board of Education') is er een proces van desegregatie in gang gezet waarbij zwarte leerlingen ook naar (eerder strikt) witte scholen konden. (Later gold dit ook voor andere minderheidsgroepen, zoals de Hispanics.) De verwachting was dat de prestaties van zwarte leerlingen, die flink achterliepen op die van blanke leerlingen, zouden verbeteren, zonder dat dit ten koste zou gaan van de prestaties van hun blanke medeleerlingen. Er is zeer veel onderzoek verricht naar de gevolgen van de vorming van meer etnisch-gemengde scholen. Schofield (1989) heeft een uitvoerig review geschreven waarin de resultaten van een aantal eerdere reviews en meta-analyses zijn samengevat; in totaal omvat haar review enkele honderden afzonderlijke studies. Het betreft hier dus onderzoek naar effecten van de etnische compositie in termen van het aandeel zwarte of Hispanic leerlingen op de prestaties van deze en blanke medeleerlingen. Van belang is dat Schofield meldt dat veel van deze (voornamelijk in de jaren '70 en '80 uitgevoerde) studies methodologische tekortkomingen vertonen. In Tabel 4.6 zijn de resultaten samengevat.

Tabel 4.6 – Samenvatting resultaten reviews en meta-analyses desegregatie

review	aantal studies	effecten
<i>effecten op lees- en wiskunde-prestaties van zwarte leerlingen</i>		
St. John (1975)	>60	geen gevolgen
Weinberg (1977)	61	meerderheid studies verbetering; veel studies geen verschil; geen negatieve effecten
Stephan (1978)	?	meerderheid studies geen verschil; veel studies positief effect
Bradley & Bradley (1977)	?	geen effect of positief effect
Krol (1978)	?	positief effect (effectgrootte 0,10)
Mahard & Crain (1983)	93	positief effect (effectgrootte 0,08)
Mahard & Crain (1983)	23	positief effect (effectgrootte 0,25)
Cook (1984)	6	geen negatieve effecten; geen effect op wiskunde-prestaties; positief effect op leesprestaties (effectgroottes 0,06 – 0,26)
<i>effecten op prestaties van Hispanic leerlingen</i>		
Gerard & Miller (1975)	1	geen gevolgen voor geen van de groepen
Kimball	1	positief effect
Adelman (1969)	1	positief effect
<i>effecten op prestaties van blanke leerlingen</i>		
St. John (1975)	24	meerderheid studies geen gevolgen; meer studies met positief dan met negatief effect
Weinberg (1977)	61	geen negatieve effecten
Orfield (1978)	?	geen negatieve effecten
Singer, Gerard & Redfeam (1975)	?	geen effect
Patchen (1982)	?	hooguit klein negatief effect
<i>effecten op deelname en voltooiing universitair onderwijs van zwarte studenten</i>		
Crain & Weisman (1972)	1	klein positief effect op deelname van mannen; positieve effecten voltooiing mannen en vrouwen
Crain & Mahrad	1	positief effect op deelname en voltooiing
Crain & McPartland (1982)	1	zwak positief effect op voltooiing

In zijn algemeenheid geldt dat er geen of hooguit kleine effecten optreden en voorzover er al effecten optreden zijn die soms positief en soms negatief.

4.5.2 Internationale studies

Thrupp, Lauder & Robinson (2002) hebben naar aanleiding van hun kritische studie naar compositie- en peer group effecten tien criteria opgesteld waaraan dergelijk (grootschalig kwantitatief) onderzoek ideaaltypisch zou moeten voldoen; zie hoofdstuk 2. Op basis van deze criteria hebben zij een aantal studies met betrekking tot de sociaal-economische compositie van scholen uit de Verenigde Staten, Engeland, Nieuw-Zeeland en België onder de loep genomen. De resultaten van dat review staan in Tabel 4.7.

Tabel 4.7 – Resultaten en waardering internationale studies naar sociaal-economische compositie-/peer group effecten aan de hand van een aantal methodologische criteria

studie	land	effect	methodologische criteria						
			longi- tudinaal	eerdere prestaties	3 dimen- sies	range com- positie	schoo- l- types	robuuste meting milieu	meer technieken
Smith & Tomlinson (1989)	UK	-	-	+	-	+	-	-	-
Gray e.a. (1990)	UK	-	-	+	-	+	nv*	-	-
Thomas & Mortimore (1994)	UK	-	-	+	-	+	+	-	-
Sammons e.a. (1997)	UK	+	+	+	-	+	+	-	-
Robertson & Symons (1996)	UK	+	-	+	-	+	nv	-	-
Strand (1998)	UK	-	-	+ / -	-	+	nv	-	-
Chubb & Moe (1990)	VS	+	-	+	-	+	+	-	-
Bryk e.a. (1993)	VS	+	-	+	+	+	+	-	-
Ho & Willms (1996)	VS	+	-	-	-	+	-	-	-
Pong (1998)	VS	+	-	+	-	+	+	-	-
Zimmer & Toma (2000)	VS	+	-	+	-	+	+	-	-
Harker & Nash (1996)	NZ	- / +	-	+	-	- / +	- / +	-	-
Lauder e.a. (1999)	NZ	+	-	+	-	-	-	+ / -	-
Opdenakker & Van Damme (2001)	B	+	-	+	+	+	+	-	-

* nv=niet vermeld.

De minnen en plussen in de kolom ‘effect’ geven aan of er een effect is aangetroffen. Grofweg blijkt dat in de Engelse studies meestal geen effecten zijn gevonden, terwijl dat voor de Amerikaanse studies wel steeds het geval is. (Onduidelijk is overigens hoe sterk deze effecten zijn.) Verder blijkt dat slechts één van de studies longitudinaal van opzet is. Wel is in praktisch alle studies in de analyse rekening gehouden met eerdere prestaties. In slechts twee studies zijn ook instructie-, organisatorische en managementkenmerken meegenomen. Ook is er in (bijna) geen van de studies sprake van een robuuste meting van sociaal milieu en van het op verschillende manier meten van het compositiekenmerk. Doorgaans is wel de hele range van scholen qua sociaal milieu in de steekproef opgenomen. Over de vertegenwoordiging van verschillende schooltypes bestaat onduidelijkheid. Al met al scoren deze 14 studies op 58 van de in totaal 98 te behalen punten (=60% van de laatste 7 kolommen) negatief. Uiteraard gaat het bij die waardering volgens Thrupp, Lauder & Robinson om ideaaltypische criteria, maar het geeft wel aan dat het ideale onderzoek nog niet verricht is.

4.5.3 Nederlandse studies

In Tabel 4.8 wordt een overzicht gepresenteerd van recente Nederlandse studies naar de effecten van klascompositie. Het accent ligt op studies waarover de laatste tien jaar is gepubliceerd en waarvoor gebruik is gemaakt van gegevens die ook in die periode zijn verzameld. Nagenoeg al die studies zijn gericht op de situatie in het primair onderwijs. In één (longitudinale) studie is nagegaan wat het effect is van klascompositie in het primair onderwijs op de onderwijspositie van de leerlingen vier jaar later in het voortgezet onderwijs (Roeleveld, 2005). In een andere (eveneens longitudinale) studie in het voortgezet onderwijs gaat het niet zozeer om compositie-effecten, maar om de onderlinge relaties tussen leerlingen (Lubbers, 2004). Omdat deze studie in veel opzichten uniek is, zeker voor Nederland, is zij toch opgenomen in dit overzicht. Alle studies maken zonder uitzondering gebruik van zeer omvangrijke databestanden met gegevens van tussen de 200 en 600 klassen (resp. groepen of scholen), en per analyse tussen de (bijna) 3000 en 12000 leerlingen.⁷ Vermeldenswaardig is ook dat, op één uitzondering na (Driessen, Mooij & Doesborgh, 2007; zie ook Driessen & Mooij, 2007), in alle studies gebruik wordt gemaakt van de voor dit type geclusterde data (leerlingen binnen klassen) meest adequate analyse-techniek, namelijk multilevel analyse.⁸ De te verklaren variabelen zijn steeds de (voorbereidende) taal-, lees- en rekenprestaties en in enkele gevallen het advies voortgezet onderwijs dat de leerlingen in groep 8 krijgen en de onderwijspositie die zijn in het voortgezet onderwijs hebben bereikt. Qua compositie worden verschillende invalshoeken onderscheiden: etniciteit (aandeel allochtonen; aandeel niet-Nederlands sprekende ouders), sociaal milieu (gemiddeld opleidingsniveau ouders; percentage laagopgeleide ouders), achterstandssituatie (gemiddelde 'weefactor' uit het Onderwijsachterstandenbeleid: middelbaar en hoger opgeleide ouders: 0,0; laagopgeleide autochtone ouders: 0,25; laagopgeleide allochtone ouders: 0,90), taalvaardigheid (gemiddeld taalniveau), rekenvaardigheid (gemiddeld rekenniveau), 'ability' (gemiddeld intelligentie- of prestatieniveau). Op de eerder genoemde uitzondering na wordt in alle studies bij het bepalen van compositie-effecten eerst rekening gehouden met (i.c. gecontroleerd voor) een serie kenmerken op individueel niveau. Met andere woorden: het gaat uiteindelijk steeds om netto-effecten van klascompositie. De kern van de kenmerken op individueel niveau waarvoor wordt gecontroleerd, zijn ook vrijwel steeds dezelfde: etnische herkomst, sociaal milieu, geslacht, leeftijd, en intelligentie. In een aantal studies zijn ook instructie- en schoolkenmerken meegenomen (o.m. Driessen & Claassen, 1996; Driessen & Doesborgh, 1998; Koopman, Van Schooten & Ledoux, 2002; Roeleveld, 2005). Enkele studies zijn longitudinaal van opzet (Koopman, Van Schooten & Ledoux, 2002; Peetsma, Van der Veen, Koopman & Van Schooten, 2003, 2006; Roeleveld, 2005; Van der Slik, Driessen & De Bot, 2006; Tesser & Iedema, 2001). In die studies wordt dan ook nog gecontroleerd voor eerdere prestaties en is bij de bepaling van de klascompositie niet uitgegaan van het laatste maar van het eerste onderzochte jaar. Dergelijke

7 Nagenoeg alle data zijn in de periode 1995-2005 verzameld bij de zes metingen van het zogenoemde PRIMA-cohortonderzoek. In hoofdstuk 2 is gewezen op het probleem van selectie-bias in peer effecten onderzoek (vgl. Webbink, 2005). Een van de oplossingen voor dit soort problemen is het gebruik van zogenoemde instrumentele variabelen. Dobbelsteen, Levin & Oosterbeek (2002) analyseren in hun studie dezelfde PRIMA-data en vergelijken de resultaten van twee methodes, namelijk gewone regressie-analyse en analyse met een instrumentele variabele. Voor twee van de drie onderzochte jaargroepen blijken de resultaten identiek. De onderzoekers concluderen op basis hiervan dat door de rijkdom van deze PRIMA-data een groot deel van het mogelijke probleem wordt opgelost.

8 Het onderwijssociologische onderzoek lijkt in dit opzicht geavanceerder te werk te gaan dan het onderwijspsychologische en (doorgaans toch zeer sophisticated en state-of-the art) onderwijseconomische onderzoek, waar deze techniek nog relatief weinig wordt toegepast (zie ook Lubbers, 2004; Thrupp, Lauder & Robinson, 2002).

studies benaderen het ideaalbeeld van een studie op basis waarvan causale effecten kunnen worden bepaald. In een paar studies worden ook differentiële effecten getoetst. Het gaat dan om de vraag of compositie-effecten steeds op identieke wijze uitpakken voor onderscheiden categorieën van leerlingen of dat het effect bijvoorbeeld voor hoog-milieu leerlingen anders uitpakt dan voor laag-milieu leerlingen, afhankelijk van de eigen positie van de leerling (vgl. het frog pond effect). Binnen die analyses zijn dan de (statistische) interacties berekend tussen bijvoorbeeld het individuele sociale milieu en het gemiddelde sociale milieu van de klas, de individuele etnische herkomst en het aandeel allochtonen in de klas, of de individuele taalprestaties en het gemiddelde prestatieniveau van de klas.⁹ Om de tabel overzichtelijk te houden zijn de gegevens over de aantallen klassen en leerlingen en de kenmerken waarop in de analyses is gecontroleerd achterwege gelaten; deze zijn immers toch voor alle studies in grote lijnen identiek. Wat resteert in de tabel is de bron van de gegevens, het jaar of de jaren dat de gegevens verzameld zijn, de te verklaren kenmerken, en ten slotte de manier waarop compositie is geoperationaliseerd en de effecten van deze compositiekenmerken.

Tabel 4.8 – Samenvatting resultaten Nederlandse studies naar effecten klascompositie

studie	jaar	groep/ jaar	te verklaren	compositie=C / effecten=E
Driessen & Claassen (1996); Driessen (1997a, b)	1995	2	lezen, rekenen	C=Gemiddelde weegfactor. Ook instructiekenmerken meege- nomen. E=Zwak negatief effect.
Driessen & Doesborgh (1998); Driessen & Slegers (2000)	1995	4, 6, 8	taal, rekenen	C=Gemiddelde weegfactor. Ook cross-level interacties onder- zocht individuele x gemiddelde weegfactor. Ook instructie- kenmerken meegenomen. E=Voor alle groepen: zwak negatief effect. Ook zwak negatief cross-level interactie-effect voor taal (niet voor rekenen): de individuele weegfactor speelt in klassen met veel achterstands- leerlingen een grotere rol dan in klassen met weinig achter- standsleerlingen.
Driessen (2002a, b); Driessen & Doesborgh (2003)	1999	4, 8	taal, rekenen	C=7 categorieën o.b.v. milieu en etniciteit ouders; diversiteit (aantal etnische groepen). Ook cross-level interacties onder- zocht individuele x milieu/ethniciteitscategorie. E=Groep 4: sterk negatief effect taal en rekenen; groep 8 sterk negatief effect taal: zwarte ($\geq 50\%$ laag-milieu allochtoon) t.o.v. witte scholen (hoog-milieu). Zwak negatief interactie- effect milieu x zwarte scholen: op witte scholen telt milieu sterker mee dan op zwarte scholen (0,1% extra verklaarde variantie).
Driessen & Doesborgh (2003); (zie ook Dries- sen, 2002c)	1999	4, 8	taal, rekenen	C=Relatieve positie in de klas qua milieu en etniciteit: minder- heid, meerderheid niet-dominant, meerderheid dominant. E=Geen effect, m.u.v. groep 8: zwak effect etnische positie op taal (meerderheid niet-dominant > meerderheid dominant).

9 Omdat het om de statistische interactie gaat tussen het lagere niveau (dat van de leerling) en het hogere niveau (dat van de de klas) worden deze aangeduid als cross-level interacties.

studie	jaar	groep/ jaar	te verklaren	compositie=C / effecten=E
Driessen & Doesborgh (2005); Driessen (2006)	2003	8	advies	C=% laag-milieu autochtoon, % laag-milieu allochtoon. E=Zeer klein negatief effect % laag-milieu autochtoon (0,3% extra verklaarde variantie).
Driessen & Smeets (2007)	2005	8	advies	C=% laag-milieu autochtoon, % laag-milieu allochtoon. E=Geen effect.
Driessen, Mooij & Doesborgh (2007); Driessen & Mooij (2007)	1997, 1999, 2001, 2003, 2005	2, 4, 6, 8	IQ taal, rekenen	C=% allochtonen, % laag-milieu, gemiddeld IQ, sd* IQ, gemiddeld taal, sd taal, gemiddeld rekenen, sd rekenen. Alleen bivariate samenhangen onderzocht, dus geen controle voor individuele kenmerken. E=Intelligentie: zwakke samenhang (eta <0,30) met gemiddeld IQ, taal en rekenen. Taal: zwakke samenhang met % allochtonen en % laag-milieu, gemiddeld IQ, sd taal, gemiddeld rekenen; middelmatige samenhang (eta <0,50) met gemiddeld taal. Rekenen: zwakke samenhang met gemiddeld IQ, taal en rekenen. Absoluut onderpresteren taal: zwakke samenhang met % allochtonen en % laag-milieu en gemiddeld taal. Absoluut onderpresteren rekenen: zwakke samenhang met gemiddeld rekenen. Relatief onderpresteren taal en rekenen: geen samenhang.
Gijsberts (2003)	2001	8	advies	C=% allochtonen. E=Klein positief effect % allochtonen, maar alleen voor de categorie 10-25%.
Gijsberts (2006); (zie ook Gijsberts, 2003)	1989, 1995, 1997, 1999, 2001, 2003	4, 6, 8	taal, rekenen	C=% allochtonen. E=Kleur van de school heeft relatief weinig invloed op de prestaties: 10% van de verschillen tussen scholen in taalprestaties en 4% van de verschillen in rekenprestaties hangen samen met het % allochtonen. Deze effecten zijn het sterkst op scholen met $\geq 50\%$ allochtonen. Effect neemt tussen 1989-2003 af.
Koopman, Van Schooten & Ledoux (2002)	1997-1999	6-8	taal, rekenen	C=Gemiddeld milieu en % allochtonen, gemiddelde taal, rekenen, IQ. Ook cross-level interacties onderzocht individuele etniciteit x % allochtonen, individueel milieu x gemiddeld milieu; taal x gemiddeld taal, rekenen x gemiddeld rekenen. Longitudinaal, met controle voor eerdere prestaties. Ook instructie- en schoolkenmerken meegenomen. E=Taal: effect van gemiddelde rekenen (!). Rekenen: geen effect. Cross-level: geen enkel differentieel effect.
Lubbers (2004)	2000-2002	1-3vo	positie VO	C=Peer relaties (peer acceptatie, vriendschap). Longitudinaal met controle voor eerdere prestaties. E=Peer acceptatie heeft effect op afstroom en doubleren. Geen effect van vriendschapsrelaties op wiskundepunten en klein effect op taalpunten. In het eerste jaar geen verschillen naar geslacht; in het tweede jaar is het aantal vrienden voor meisjes belangrijker dan voor jongens. De effecten van peer relaties worden niet gemedieerd door het gevoel er bij te horen.

studie	jaar	groep/ jaar	te verklaren	compositie=C / effecten=E
Peetsma, Van der Veen, 1997- Koopman & Van Schooten (2003; 2006)	1999	6-8	taal, rekenen	C=Gemiddeld milieu en % allochtonen. Ook cross-level interacties onderzocht individuele etniciteit x % allochtonen, individueel milieu x gemiddeld milieu. Longitudinaal, met controle voor eerdere prestaties. E=Rekenen: geen effect % laag milieu; klein negatief effect % allochtonen, maar alleen in combinatie met interactie met etniciteit: allochtone leerlingen presteren het best in klassen met een meerderheid van allochtonen, terwijl autochtone leerlingen het best presteren in klassen met meerderheid autochtone leerlingen (0,5% extra verklaarde variantie). Taal: zeer klein negatief effect % laag-milieu; zeer klein interactie-effect (van één van de categorieën). Geen effect van % allochtonen, zeer klein interactie-effect % allochtonen (0,4% extra verklaarde variantie).
Roeleveld (2005)	1997- 2001	8bao- 4vo	positie VO	C=7 categorieën o.b.v. milieu en etniciteit ouders. Longitudinaal met controle voor eerder prestaties en advies. Ook instructiekenmerken meegenomen. E=Klein negatief effect voor scholen met meer dan de helft laag-milieu allochtone leerlingen (ES**=0,17) en klein positief effect voor scholen met meer dan de helft hoog-milieu leerlingen (0,3% extra verklaarde variantie).
Van der Slik, Driessen & De Bot (2006)	1995- 1997	4-6	taal	C=% niet-Nederlands thuistaal, % werk, gemiddelde opleiding en inkomen; sd opleiding en inkomen. Longitudinaal groep 6 na controle groep 4. E= Klein negatief effect % niet-Nederlands thuistaal en spreiding in inkomen (0,8% extra verklaarde variantie).
Tesser & Iedema (2001)1999		2, 8; 4-6, 6-8	lezen, taal, rekenen	C=% allochtone leerlingen. Deels longitudinaal, met controle voor eerdere prestaties via verschillcores. E=Groep 2: zwakke negatieve effecten (max. 0,25 sd op lezen [bij >75% allochtonen 0,25 sd] en 0,17 sd op rekenen. Groep 8: zwakke negatieve effecten (max. 0,25 sd op lezen [bij >50% allochtonen 0,25 sd] en 0,19 op rekenen. Longitudinaal van groep 4 naar 6 en 6 naar 8: geen significante effecten.

*= standaarddeviatie; **=effectgrootte

Wat kan nu, samenvattend, op basis van de in Tabel 4.8 beschreven studies worden geconcludeerd? Allereerst moet worden geconstateerd dat deze studies doorgaans van een hoog methodologische gehalte zijn: er wordt met zeer omvangrijke steekproeven gewerkt, gecontroleerd voor een groot aantal achtergrondkenmerken (vaak ook instructie- en schoolkenmerken) en gebruik gemaakt van state-of-the-art analysetechnieken (multi-level). Daarnaast zijn een aantal studies longitudinaal van aard, met controle voor eerdere prestaties. En bovendien worden in enkele studies via de berekening van cross-level interacties ook differentiële effecten bepaald, of, anders gezegd, of compositie-effecten eventueel verschillen voor subgroepen.

Uit de analyses blijkt dat er doorgaans geen tot hooguit zwakke effecten optreden van de milieu- en etnische samenstelling van de leerlingenpopulatie van een klas. Meer lager milieu en/of meer allochtone leerlingen gaan dus hooguit in beperkte mate samen met lagere prestaties en adviezen. De toegevoegde verklaringskracht van de compositiekenmerken, dat wil zeggen bovenop de invloed van de achtergronden en eerdere prestaties van de leerlingen, is minimaal: meestal minder dan 1%. De grootste verschillen worden – vanzelfsprekend – aangetroffen wanneer de compositievariabele is uiteengelegd in een aantal categorieën en extreme categorieën worden vergeleken, bijvoorbeeld witte met zwarte klassen (bv. Driessen & Doesborgh, 2003; Roeleveld, 2005). Er zijn aanwijzingen dat de compositie-effecten in de loop van de onderzochte jaren af aan het nemen zijn. Er treden ook geen of hooguit zwakke differentiële effecten op. Al met al lijkt er dus nauwelijks of geen sprake van sociale of etnische compositie-effecten. Daar hoort overigens wel een relativering bij. Wanneer zich compositie-effecten voordoen, dan hebben ze meteen betrekking op een aantal leerlingen binnen een school, groep of klas. Daarnaast cumuleren deze effecten over de hele schoolloopbaan van een leerling. Deze op zich kleine effecten tellen dan wel op tot een groter effect. Analyses van Gibson & Telhaj (2006) maken echter duidelijk dat ook dergelijke cumulatieve effecten gering zijn.

4.6 Recente onderwijs-economische studies

In Tabel 4.9 staan de resultaten van een aantal recente, internationale studies naar peer group effecten (i.c. compositie-effecten), waarbij gebruik is gemaakt van rijke datasets en geavanceerde schattingsmethoden. Het betreft voornamelijk onderwijs-economische studies waarbij qua compositie uiteenlopende perspectieven (soms meerdere tegelijk) zijn gehanteerd. De basis van deze tabel wordt gevormd door overzichtstabellen van Ammermüller & Pischke (2006) en Gibson & Telhaj (2006). Deze tabellen zijn aangevuld met informatie uit de originele publicaties en uitgebreid met een aantal nog niet door hen opgenomen studies.

Tabel 4.9 – Samenvatting recente, geavanceerde studies naar peer group effecten

studie	land	fase	te verklaren	compositie/ methode	effectgrootte (1 sd → x sd)*	aanvullende opmerkingen
Hoxby (1998)	VS	PO	(ontwikkeling in) prestaties (lezen, schrijven, rekenen)	compositie qua sekse en ras / natuurlijke random-variatie	% zwarte leerlingen: 0; % meisjes: (per 10%) -0,30 – 0,20	Rekenen: -0,30; schrijven: 0,20; lezen: 0,10.
Hoxby (2000)	VS	PO	prestaties (lezen, rekenen)	compositie qua sekse en ras / cohortvariatie	0,15 – 0,40	Geen niet-lineaire effecten.
Sacerdote (2001)	VS	HO	prestaties	prestaties kamergenoten / random toewijzing	- 0,07	Controle voor leerlingkenmerken. Ook interacties leerling x groep onderzocht: zwakke, wisselende effecten.

studie	land	fase	te verklaren	compositie/ methode	effectgrootte (1 sd → x sd)*	aanvullende opmerkingen
McEwan (2003)	Chili	PO	prestaties (taal, rekenen)	milieu / school fixed effecten	0,27	Controle voor leerlingken- merken.
Hanushek e.a. (2003)	VS	PO	prestaties (rekengroei)	compositie qua milieu, gemiddelde prestaties / en sd prestaties / school x jaar fixed effecten	0,08 (van prestaties) ¹⁰	Controle voor leerling- en schoolkenmerken. Ook interacties etniciteit leerling x groep onderzocht: geen niet-lineaire effecten.
Zimmerman (2003)	VS	HO	prestaties (taal- en wiskunde)	prestaties kamerge- noten / random toewijzing	0,05	Controle eerdere prestaties, etniciteit en sekse. Sterker verband met taal dan wis- kunde.
Cullen, Jacob & Levitt (2003)	VS	VO	prestaties	toewijzing via loterij	0	Geen niet-lineaire effecten.
Sanbonmatsu e.a. (2004)	VS	6-20 jaar	o.m. prestaties	random toewijzing	0	
Goux & Maurin (2005)	Frank- rijk	?	doubleren	instrumentele vari- abelen (leeftijd buren)	0,10	
Ammermüller & Pischke (2006)	Europa	PO	prestaties (lezen)	boeken thuis, oplei- ding ouders / schoolverschillend fixed effecten; instrumentele vari- abelen	0,11 (maar per land)	Controle voor leerling-, leerkracht- en schoolken- merken. Ook interacties leerling x groep onderzocht: niet significant.
Schindler- Rangvid (2003)	Dene- marken	VO	prestaties (lezen)	milieu	0,07	
Gibbons & Tel- haj (2005)	UK	VO	prestaties (taal en wiskunde)	gemiddelde presta- ties / random varia- tie; instrumentele variabelen	0,05-0,08	Controle voor leerlingken- merken.
Graham (2004)	VS	PO	prestaties (lezen en rekenen)	random toewijzing	50 ^e percentiel toename: 0,90- 1,10 sd	
Schneeweiss & Winter-Ebmer (2005)	Oosten- rijk	VO	prestaties (lezen en wiskunde)	milieu / school fixed effects	0,06	Controle voor leerling- , leerkracht- en schoolken- merken. Ook interacties leerling x groep onderzocht: niet relevant.
Angrist & Lang (2002)	VS	PO	prestaties (taal lezen, rekenen, algemene vaar- digheden)	% zwarten / fixed effects; instrumen- tele variabelen	geen effect	Controle voor klaskenmer- ken.

10 De effectgrootte verschilt nogal tussen de onderscheiden bronnen, van 0,02 (Gibbons & Telhaj, 2006) en 0,05 (Ammermüller & Pischke, 2006) tot 0,20 (Hanushek e.a., 2003), maar is waarschijnlijk 0,08.

studie	land	fase	te verklaren	compositie/ methode	effectgrootte (1 sd → x sd)*	aanvullende opmerkingen
Kang (2007)	Zuid-Korea	VO	prestaties ` (wiskunde)	gemiddelde en sd prestaties / random toewijzing; school fixed effects; instrumentele variabelen	0,26	Controle voor leerling- en leerkrachtkenmerken. Geen niet-lineaire effecten.
Lavy & Schlosser (2007)	Israel	PO, VO	prestaties (taal, wiskunde en science)	sekse (% meisjes) / cohortvariatie	verschillend per onderwijs-fase, vak en sekse	Controle voor leerling- en leerkracht/klaskenmerken. PO: positief effect voor j. en m.** op bètavakken, niet op taal. Middle schools: effect op wiskunde, taal en Engels voor m. High schools: positief effect voor j. en m.; niet-lineair.
Lefgren (2007)	VS	PO	prestaties (lezen en wiskunde)	gemiddelde prestaties / school fixed effects; instrumentele variabelen	0,02	Controle voor leerling- en buurtkenmerken en eerdere prestaties.
Stinebrickner & Stinebrickner (2006)	VS	HO	prestaties (taal en wiskunde)	prestaties en inkomsten kamergenoten	net-significant effect voor m. (niet voor j.)	

* = standaarddeviatie; ** j=jongens, m=meisjes

Hoewel voor de in deze tabel beschreven studies doorgaans gebruik is gemaakt van geavanceerde statistische modellen, kan ook daarmee niet alle methodologische kritiek worden weggenomen (zie hoofdstuk 2) – mede omdat de gebruikte databestanden hun beperkingen kennen. Opmerkelijk is overigens dat slechts sporadisch gebruik is gemaakt van multi-level analysetechnieken, iets wat in onderwijssociologisch onderzoek de laatste tien jaar juist min of meer standaard gebeurt. Van belang is tevens dat soms krachtige uitspraken lijken te worden gedaan op basis van de significantie van effecten. Omdat – zoals al eerder opgemerkt – de mate van significantie afhankelijk is van de omvang van een steekproef en hier voornamelijk grote steekproeven zijn gebruikt, is al gauw alles significant. Dit zegt echter feitelijk nog niets over de (praktische, onderwijskundige, maatschappelijke) relevantie van een dergelijk effect (vgl. Carver, 1978; Cahan, 2000; Levin & Robinson, 2000; Schagen & Elliot, 2004). Daarvoor kan beter gekeken worden naar de effectgroottes (vgl. Thompson, 1998); die zijn niet afhankelijk van de steekproefomvang en, omdat het gestandaardiseerde coëfficiënten betreft, voor verschillende indicatoren rechtstreeks met elkaar vergelijkbaar. Wat deze tabel duidelijk maakt is dat, vergeleken over verschillende landen, onderwijsniveaus en peer groups, op een enkele uitzondering na peer group effecten zeer gering of zelfs non-existent zijn. Ook is er daarbij doorgaans geen sprake van niet-lineaire (of differentiële of asymmetrische) effecten.

Conclusies

In het onderhavige onderzoek is nagegaan wat de invloed is van de specifieke samenstelling van de groep medeleerlingen (de ‘peer group’) op de prestaties van individuele leerlingen. Een vraag is bijvoorbeeld of het voor leerlingen ongunstig is voor hun prestaties of ze in een klas zitten met veel allochtone medeleerlingen. Concreet zijn peer group effecten onderzocht vanuit de volgende perspectieven:

- ‘ability’ (niveaugroepen; de gemiddelde prestaties in de groep);
- combinatieklassen (van meerdere leerjaren of leeftijden);
- sekse (meisjesklassen/scholen vs. co-educatie; het percentage meisjes in de groep);
- sociaal milieu (het gemiddelde opleidingsniveau van de ouders van de leerlingen in de groep; het percentage leerlingen van laagopgeleide ouders in de groep);
- etnische herkomst (het percentage allochtone leerlingen in de groep; zwarte scholen vs. witte scholen).

Voor het onderzoek zijn de bevindingen van vele honderden Nederlandse en buitenlandse studies bij elkaar gebracht. Dit type onderzoek blijkt buitengewoon complex en er zitten dan ook veel haken en ogen aan. Om die reden wordt uitvoerig stil gestaan bij allerlei theoretische en methodologische kwesties. Op grond daarvan lijkt enige terughoudendheid op zijn plaats bij het trekken van harde conclusies op basis van de resultaten van deze studies. Dergelijke conclusies kunnen pas worden getrokken wanneer er adequaat ingericht onderzoek beschikbaar komt.

Uitgebreide aandacht wordt besteed aan mogelijke verklaringen voor peer group effecten. Er kunnen daarbij specifieke en algemene theorieën worden onderscheiden. Onder de eerste categorie vallen de taalcontact-, netwerk- en nivellerings- theorieën, onder de tweede vergelijkings- en netwerktheorieën.

Gegeven de huidige theoretische en methodologische state-of-the-art rond peer group onderzoek, lijkt in zijn algemeenheid te kunnen worden geconcludeerd dat er hooguit kleine effecten optreden van de klassamenstelling. Dit betekent dus dat het niet uitmaakt of leerlingen in klassen zitten met veel medeleerlingen met eenzelfde prestatieniveau of in klassen met een sterk gemengd prestatieniveau, of ze in klassen zitten met leerlingen van dezelfde leeftijd of in combinatieklassen en of ze in klassen zitten met veel of weinig meisjes. Evenmin maakt het uit of ze in klassen zitten met veel of weinig leerlingen uit eenzelfde sociaal milieu of etnische groep.

Ook blijkt dat er doorgaans geen differentiële effecten worden gevonden, ofwel verschillen in effecten tussen subgroepen. Dit betekent dat het bijvoorbeeld niet alleen niet uitmaakt voor allochtone leerlingen of ze in klassen met veel andere allochtone leerlingen zitten, maar evenmin voor hun autochtone medeleerlingen in die klassen.

Opgemerkt zou kunnen worden dat al die kleine effecten cumulatief mogelijk wel enige relevantie hebben. Immers, niet alleen tellen die effecten voor alle leerlingen binnen een groep op,

maar daarnaast ook voor de hele schoolloopbaan van elke leerling afzonderlijk (Hanushek e.a., 2003). Onderzoek naar mogelijke cumulatieve effecten laat echter zien dat ook die effecten klein zijn (vgl. Gibson & Telhaj, 2006).

Het feit dat er geen substantiële peer group effecten worden gevonden betekent dat het voor het verbeteren van de onderwijsprestaties van leerlingen waarschijnlijk geen zin heeft de klassamenstelling aan te passen (vgl. Lefgren, 2004; Wilkinson e.a. 2000). Omdat er ook geen differentiële peer group effecten optreden, heeft wijzigen van de klassamenstelling ook weinig zin om eventuele prestatieverschillen tussen bepaalde groepen van leerlingen op te heffen. Deze beide bevindingen houden dus in dat er zowel vanuit een 'equity' als een 'efficiency' perspectief weinig verwacht kan worden van realloceren van groepen van leerlingen (vgl. Hanushek & Wößmann, 2006).

Afsluitend kan de vraag worden gesteld waarom er geen of hooguit zeer kleine peer group effecten worden aangetroffen. Op de eerste plaats natuurlijk omdat ze er gewoonweg niet zijn. Kenmerken op het individuele leerlingniveau, zoals sociaal milieu en etnische herkomst, blijken veel belangrijker dan die op het klas- en schoolniveau. Bovendien is er ten aanzien van peer group kenmerken sprake van verdunningseffecten: de effecten op de hogere niveaus zijn indirect en nemen daardoor met elk lager niveau in sterkte af (vgl. Hattie, 2002; Wilkinson e.a., 2002). Een veronderstelling is ook dat leerkrachten bij een gewijzigde klassamenstelling hun didactiek aanpassen. Daar zijn volgens verschillende onderzoekers echter geen aanwijzingen voor (Hattie, 2002; Hoxby, 1998). Dit neemt overigens niet weg dat in specifieke leerarrangementen (voor bv. hoogbegaafde leerlingen) wel duidelijk positieve effecten optreden van klassamenstelling (Mooij, Hoogeveen, Driessen, Van Hell & Verhoeven, 2007). Een mogelijkheid is tevens dat er verschillende peer group effecten tegelijkertijd werkzaam zijn, maar dan in tegengestelde richting. Dit betekent dat ze tegen elkaar wegvallen en dat het er dus op lijkt alsof er geen effecten zijn (Guldemond, 1994; Koopman, Van Schooten & Ledoux (2002). Een laatste punt waarop (nogmaals) moet worden gewezen is dat onderzoek naar peer group effecten bijzonder gecompliceerd is. Dat er geen effecten worden gevonden, heeft misschien ook (deels) te maken met allerlei methodologische en theoretische en conceptuele tekortkomingen. In ieder geval is duidelijk dat er nauwelijks onderzoek is verricht waarin op adequate wijze de belangrijkste theoretische variabelen vanuit het frog pond effect zelf zijn opgenomen. Naar die onderliggende (interactie) processen zou daarom veel meer onderzoek gedaan moeten worden, bijvoorbeeld in de vorm van fijnmazig beschrijvend of etnografisch onderzoek (Thrupp, Lauder & Robinson, 2002).

Literatuur

- Ammermüller, A., & Pischke, J.-S. (2006). *Peer effects in European primary schools: Evidence from PIRLS*. Mannheim: Zentrum für Europäische Wirtschaftsforschung.
- Angrist, J., & Lang, K. (2002). *How important are classroom peer effects? Evidence from Boston's Metco program*. Cambridge, MA: National Bureau of Economic Research.
- Atkinson, J. (1964). *An introduction to motivation*. New York: American Book.
- Cahan, S. (2000). Statistical significance is not a 'kosher certificate' for observed effects: A critical analysis of the two-step approach to the evaluation of empirical results. *Educational Researcher*, 29, (1), 31-34.
- Carver, R. (1978). The case against statistical significance testing. *Harvard Educational Review*, 48, (3), 378-399.
- Claassen, A., & Mulder, L. (2006). *Na vier jaar voortgezet onderwijs*. Nijmegen: ITS.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Coleman, J., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966) *Equality of educational opportunity*. Washington, D.C.: US Government Printing Office.
- Daly, P. (1996). The effects of single-sex and coeducational secondary schooling on girls' achievement. *Research Papers in Education*, 11, (3), 289-306.
- Dar, Y. (1981). *Homogeneity and heterogeneity in education: Interaction between personal resources and the learning environment in their effect on scholastic achievement*. Jerusalem: Hebrew University Institute for Innovation in Education.
- Davis, J. (1966). The campus as a frog pond: An application of the theory of relative deprivation to career decisions of college men. *American Journal of Sociology*, 72, 17-31.
- DeWayne, A., & Burns, R. (1997). Reassessing the effects of combination classes. *Educational Research and Evaluation*, 3, (1), 1-53.
- Dijkstra, A., & Veenstra, R. (2000). Functionele gemeenschappen, godsdienstigheid en prestaties in het voortgezet onderwijs. *Mens & Maatschappij*, 75, (2), 129-150.
- Dobbelsteen, S., Levin, J., & Oosterbeek, H. (2002). The causal effect of class size on scholastic achievement: Distinguishing the pure class size effect from the effect of changes in class composition. *Oxford Bulletin of Economics and Statistics*, 64, (17), 17-38.
- Driessen, G. (1991). Discrepancies tussen toetsresultaten en doorstroomniveau. Positieve discriminatie bij de overgang basisonderwijs – voortgezet onderwijs? *Pedagogische Studiën*, 68, 27-35.
- Driessen, G. (1997a). Pre-reading and pre-arithmetic instruction in infant education in the Netherlands: A multilevel analysis approach. *Early Child Development and Care*, 134, 1-21.
- Driessen, G. (1997b). Voorbereidend lezen en rekenen in kleutergroepen: werkwijzen en resultaten. *Tijdschrift voor Onderwijsresearch*, 22, (3), 145-160.
- Driessen, G. (2002a). Sociaal-etnische schoolcompositie en onderwijsresultaten: Effecten van positie, concentratie en diversiteit. *Pedagogische Studiën*, 79, (3), 212-230.
- Driessen, G. (2002b). School composition and achievement in primary education: A large-scale multilevel approach. *Studies in Educational Evaluation*, 28, (4), 347-368.

- Driessen, G. (2002c). The effect of religious groups' dominance in classrooms on cognitive and noncognitive educational outcomes. *International Journal of Education and Religion*, 3, (1), 46-68.
- Driessen, G. (2005). De totstandkoming van de adviezen voortgezet onderwijs: Invloeden van thuis en school. *Pedagogiek*, 25, 279-298.
- Driessen, G. (2006). Het advies voortgezet onderwijs: Is de overadvisering over? *Mens & Maatschappij*, 81, (1), 5-23.
- Driessen, G., & Bezemer, J. (1999). *Islamitisch basisonderwijs. Schipperen tussen identiteit en kwaliteit?* Nijmegen: ITS.
- Driessen, G., & Claassen, A. (1996). *Voorbereidend lezen, rekenen en schrijven. Toetsresultaten en leerling- en leerkrachtkenmerken in de kleutergroepen van het basisonderwijs.* Ubergen: Tandem Felix.
- Driessen, G., & Dekkers, H. (2007). Educational inequality in the Netherlands: Policy, practice, and effects. In R. Teese, S. Lamb & M. Duru-Bellat (Eds.), *International studies in educational inequality, theory and policy. Volume 3. Inequality: Educational theory and public policy* (pp. 257-274). Dordrecht: Springer.
- Driessen, G., & Doesborgh, J. (1998). *Consistentie in didactische werkwijze binnen de school en prestaties van leerlingen in het basisonderwijs.* Nijmegen: ITS.
- Driessen, G., & Doesborgh, J. (2003). Compositie van de schoolpopulatie en meerder- en minderheidspositie. In G. Driessen, J. Doesborgh, G. Ledoux, I. van der Veen, & M. Vergeer, *Sociale integratie in het primair onderwijs. Een studie naar de relatie tussen de sociale, etnische, religieuze en cognitieve schoolcompositie en de cognitieve en niet-cognitieve positie van verschillende groepen leerlingen. Analyses bij het PRIMA-cohortonderzoek, derde meting* (pp. 105-164). Nijmegen/Amsterdam: ITS/SCO-Kohnstamm Instituut.
- Driessen, G., & Doesborgh, J. (2005). Relaties tussen achtergrondkenmerken en competenties van leerlingen en hun advies voor voortgezet onderwijs. In G. Driessen, J. Doesborgh, G. Ledoux, M. Overmaat, J. Roeleveld, & I. van der Veen, *Van basis- naar voortgezet onderwijs. Voorbereiding, advisering en effecten* (pp. 39-70). Nijmegen/Amsterdam: ITS/SCO-Kohnstamm Instituut.
- Driessen, G., & Langen, A. van (2007). Sekseverschillen in het onderwijs: 'The boys' problem' in internationaal perspectief. *Mens & Maatschappij*, 82, (2), 109-132.
- Driessen, G., & Slegers, P. (2000). Consistency of teaching approach and student achievement: An empirical test. *School Effectiveness and School Improvement*, 11, (1), 57-79.
- Driessen, G., & Smeets, E. (2007). De relatie tussen prestaties en advies: onder- of overadvisering bij de overgang van basis- naar voortgezet onderwijs? In Inspectie van het Onderwijs (Ed.), *Onderadvisering in beeld* (pp. 59-81). Utrecht: Inspectie van het Onderwijs.
- Driessen, G., & Mooij, T. (2007). Onderpresteren, begaafdheid en leerling- en groepskenmerken in het basisonderwijs. *Tijdschrift voor Orthopedagogiek*, 46, (9), 366-378.
- Driessen, G., Mooij, T., & Doesborgh, J. (2007). *Hoogbegaafdheid van leerlingen in het primair onderwijs. Ontwikkelingen en samenhangen met kenmerken van thuis, de groep en de school.* Nijmegen: ITS.
- Driessen, G., Doesborgh, J., Ledoux, G., Veen, I. van der, & Vergeer, M. (2003). *Sociale integratie in het primair onderwijs.* Nijmegen/Amsterdam: ITS/SCO-Kohnstamm Instituut.
- Entorf, H., & Lauk, M. (2006). *Peer effects, social multipliers and migration at school: An international comparison.* Hamburg: HWWI.
- Gibbons, S., & Telhaj, S. (2006). *Peer effects and pupil attainment: Evidence from secondary school transition.* London: London School of Economics.

- Gijsberts, M. (2003). Minderheden in het basisonderwijs. In J. Dagevos, M. Gijsberts & C. van Praag (Eds.), *Rapportage minderheden 2003. Onderwijs, arbeid en sociaal-culturele integratie* (pp. 63-109). Den Haag: SCP.
- Gijsberts, M. (2006). De afnemende invloed van etnische concentratie op schoolprestaties in het basisonderwijs, 1988-2002. *Sociologie*, 2, (2), 157-177.
- Goodenow, C. (1993). Classroom belonging among early adolescent students: Relationships to motivation and achievement. *Journal of Early Adolescence*, 12, (1), 21-43.
- Guldemond, H. (1994). *Van de kikker en de vijver. Groepseffecten op individuele prestaties*. Leuven/Apeldoorn: Garant.
- Hallinan, M. (1990). The effects of ability grouping in secondary schools: A response to Slavin's best-evidence synthesis. *Review of Educational Research*, 60, (3), 501-504.
- Hanushek, E., Kain, J., Markman, J., & Rivkin, S. (2003). Does peer ability affect student achievement? *Journal of Applied Economics*, 18, 527-544.
- Hanushek, E., & Wößmann, L. (2006). Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. *The Economic Journal*, 116, 63-76.
- Harker, R. (2000). Achievement, gender and the single-sex/coed debate. *British Journal of Sociology of Education*, 21, 203-218.
- Hattie, J. (2002). Classroom composition and peer effects. *International Journal of Educational Research*, 37, 449-481.
- Hoeven-Van Doornum, A. van der (1990). *Effecten van leerlingbeelden en streefniveaus op schoolloopbanen*. Nijmegen: ITS.
- Hoeven-Van Doornum, A. van der, Voeten, R., & Jungbluth, P. (1989). The effect of aspiration levels set by teachers for their pupils on learning achievement. *School Effectiveness and School Improvement*, 1, 231-239.
- Hoxby, C. (1998). *The effects of class size and composition on student achievement: New evidence from natural population variation*. Cambridge, MA: National Bureau of Economic Research.
- Hoxby, C. (2000). *Peer effects in the classroom: Learning from gender and race variation*. Cambridge, MA: National Bureau of Economic Research.
- Imdorf, C. (2003). *Organisational perspective on the transition from primary to secondary school: Tracking decisions depending on gender and ethnicity in the Swiss education system*. Paper UNESCO Conference on Intercultural Education, 15-18 June 2003, Jyväskylä, Finland.
- Jungbluth, P. (1985). *Verborgene differentiatie. Leerlingbeeld en onderwijsaanbod op de basisschool*. Nijmegen: ITS.
- Jungbluth, P. (2003). *De ongelijke basisschool. Etniciteit, sociaal milieu, sekse, verborgen differentiatie, segregatie, onderwijskansen en schooleffectiviteit*. Nijmegen: ITS.
- Kang, C. (2007). Classroom peer effects and academic achievement: Quasi-randomization evidence from South Korea. *Journal of Urban Economics*, 61, 458-495.
- Kassenberg, A., Vos, H. De, & Dijkstra, A. (2000). *Gemeenschap binnen en rond scholen voor voortgezet onderwijs en het gevoel 'bij de school te horen'*. Paper ORD, 2-26 mei 2000, te Leiden.
- Kelly, A. (1996). Comparing like with like. *Education*, 117, 14-15.
- Kristen, C. (2000). *Ethnic differences in educational placement: The transition from primary to secondary schooling*. Mannheim: MZES.

- Kling, J., Liebman, J., & Katz, L. (2007). Experimental analysis of neighborhood effects. *Econometrica*, 75, (1), 83-119.
- Koopman, P., Schooten, E. van, & Ledoux, G. (2002). *Effecten van klassencompositie*. Amsterdam: SCO-Kohnstamm Instituut.
- Kulik, C.-L., & Kulik, J. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal*, 19, (3), 415-428.
- Langen, A. van, & Driessen, G. (2006). *Sekseverschillen in onderwijsloopbanen. Een internationaal comparatieve trendstudie*. Nijmegen: ITS.
- Lavy, V., & Schlosser, A. (2007). *Mechanisms and impacts of gender peer effects at school*. Cambridge, MA: National Bureau of Educational Research.
- Lefgren, L. (2004). Educational peer effects and the Chicago public schools. *Journal of Urban Economics*, 56, 169-191.
- Levin, J., & Robinson, D. (2000). Rejoinder: Statistical hypothesis testing, effect-size estimation, and the conclusion coherence of primary research results. *Educational Researcher*, 29, (1), 34-36.
- Lubbers, M. (2004). *The social fabric of the classroom. Peer relations in secondary education*. Groningen: GION.
- Mael, F., Alonso, A., Gibson, D., Rogers, K., & Smith, M. (2005). *Single-sex versus coeducational schooling: A systematic review*. Washington, DC: U.S. Department of Education.
- Manger, T., & Eikeland, O.-J. (1997). The effect of social comparison on mathematics self-concept. *Scandinavian Journal of Psychology*, 28, 237-241.
- Manski, C. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60, (3), 531-542.
- Marsh, H. (1984). Self-concept, social comparison, and ability grouping: A reply to Kulik and Kulik. *American Educational Research Journal*, 21, (4), 799-806.
- Marsh, H., & Hau, K.-T. (2003). Big-fish-little-pond effect on academic self-concept. A cross-cultural (26-country) test of negative effects of academically selective schools. *American Psychologist*, 58, (5), 364-376.
- Marsh, H., & Parker, J. (1984). Determinants of student self-concept: Is it better to be a relative large fish in a small pond even if you don't learn to swim as well? *Journal of Personality and Social Psychology*, 47, (1), 213-231.
- Mason, D., & Burns, R. (1996). 'Simply no worse and simply no better' may simply be wrong: A critique of Veenman's conclusion about multigrade classes. *Review of Educational Research*, 66, (3), 307-322.
- Mason, D., & Burns, R. (1997). Reassessing the effects of combination classes. *Educational Research and Evaluation*, 3, (1), 1-53.
- McGee, C., Ward, R., Gibbons, J., & Harlow, A. (2003). *Transition to secondary school: A literature review*. Hamilton, NZ: University of Waikato.
- McEwan, P. (2003). Peer effects on student achievement: Evidence from Chile. *Economics of Education Review*, 22, 131-141.
- Michaelowa, K., & Bourdon, J. (2006). *The impact of student diversity in secondary schools*. Hamburg: HWWI.
- Moffit, R. (2001). Policy interventions, low-level equilibria, and social interactions. In S. Durlauf & H. Young (Eds.), *Social dynamics* (pp. 45-82). Washington, DC: Brookings Institution Press.

- Mooij, T., Hoogeveen, L., Driessen, G., Hell, J. van, & Verhoeven, L. (2007). *Succescondities voor onderwijs aan hoogbegaafde leerlingen. Eindverslag van drie deelonderzoeken*. Nijmegen: Radboud Universiteit, ITS / Centrum voor Begaafdheidsonderzoek / Afdeling Orthopedagogiek.
- OECD (2006). *Where immigrant students succeed. A comparative review of performance and engagement in PISA 2003*. Paris: OECD.
- Peetsma, T., Veen, I. van der, Koopman, P., & Schooten, E. van (2003). Kan een eenzijdige klassamenstelling de ontwikkeling van leerlingen afremmen? Effecten van sociale en etnische klassamenstelling en verschillen hierin tussen groepen. *Pedagogische Studiën*, 80, (5), 339-357.
- Peetsma, T., Veen, I. van der, Koopman, P., & Schooten, E. van (2006). Class composition influences on pupils' cognitive development. *School Effectiveness and School Improvement*, 17, (3), 275-302.
- Richer, S. (1976). Reference-group theory and ability grouping: A convergence of sociological theory and educational research. *Sociology of Education*, 49, (1), 65-71.
- Robertson, D., & Symons, J. (2003). Do peer groups matter? Peer group versus schooling effects on academic attainment. *Economics*, 70, 31-53.
- Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics*, 116, (2), 681-704.
- Schagen, I., & Elliot, K. (Eds.) (2004). *But what does it mean? The use of effect sizes in educational research*. Berkshire: NFER.
- Schneeweis, N., & Winter-Ebmer, R. (2007). Peer effects in Austrian schools. *Empirical Economics*, 32, 387-409.
- Schnepf, S. (2002). *A sorting hat that fails? The transition from primary to secondary school in Germany. Innocenti Working Paper No. 92*. Florence: UNICEF Innocenti Research Centre.
- Schnepf, S. (2004). *How different are immigrants? A cross-country and cross-survey analysis of educational achievement. IZA Discussion Paper No. 1398*. Bonn: IZA.
- Schofield, J. (1989). *Review of research on desegregation's impact on elementary and secondary school students*. University of Pittsburgh.
- Slavin, R. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research*, 60, (3), 471-499.
- Slavin, R. (1990). Ability grouping in secondary schools: A response to Hallinan. *Review of Educational Research*, 60, (3), 505-507.
- Slik, F. van der, Driessen, G., & Bot, K. de (2006). Ethnic and socioeconomic class composition and language proficiency: A longitudinal multilevel examination in Dutch elementary schools. *European Sociological Review*, 22, (3), 293-308.
- Stinebrickner, R., & Stinebrickner, T. (2006). What can be learned about peer effects using college roommates? Evidence from new survey data and students from disadvantaged backgrounds. *Journal of Public Economics*, 90, 1435-1454.
- Strand, S., Deary, I., & Smith, P. (in press). Sex differences in cognitive ability test scores: A UK national picture. *British Journal of Educational Psychology*.
- Tesser, P., & Iedema, J. (2001). *Rapportage minderheden 2001. Deel I Vorderingen op school*. Den Haag: SCP.
- Tesser, P., Praag, C. van, Dugteren, F., Herweijer, L., & Wouden, H. van der (1995). *Rapportage minderheden 1995. Concentratie en segregatie*. Den Haag: VUGA.

- Thompson, B. (1998). *Five methodology errors in educational research: The pantheon of statistical significance and other faux pas*. Invited address AERA annual meeting, San Diego, US, April 1998.
- Thrupp, M. (1995). The school mix effect: the history of an enduring problem in educational research, policy and practice. *British Journal of Sociology of Education*, 16, 183-203.
- Thrupp, M. (1999). *Schools making a difference: let's be realistic! Schoolmix, school effectiveness and the social limits of reform*. Buckingham/Philadelphia: Open University Press.
- Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research*, 37, 483-505.
- Van de gaer, E. (2006). *Gender differences in academic achievement. The role of school engagement, group composition, and educational choices*. Leuven: Katholieke Universiteit.
- Veenman, S. (1995). Cognitive and noncognitive effects of multigrade and multi-age classes: A best-evidence synthesis. *Review of Educational Research*, 65, (4), 319-381.
- Veenman, S. (1996). Effects of multigrade and multi-age classes reconsidered. *Review of Educational Research*, 66, (3), 323-340.
- Veenman, S. (1997). Combination classes revisited. *Educational Research and Evaluation*, 3, (3), 262-276.
- Vergeer, M., Driessen, G., & Ledoux, G. (2003). Theoretische verkenning. In G. Driessen, J. Doesborgh, G. Ledoux, I. van der Veen, & M. Vergeer, *Sociale integratie in het primair onderwijs. Een studie naar de relatie tussen de sociale, etnische, religieuze en cognitieve schoolcompositie en de cognitieve en niet-cognitieve positie van verschillende groepen leerlingen. Analyses bij het PRIMA-cohortonderzoek, derde meting* (pp. 11-50). Nijmegen/Amsterdam: ITS/SCO-Kohnstamm Instituut.
- Vos, H. de, (1986). Een model van het interactiesysteem van de schoolklas. *Mens en Maatschappij*, 2, 147-167.
- Vries, A. de (1994). Effecten van de classesamenstelling: Een theoretische verkenning. *Pedagogisch Tijdschrift*, 19, (2), 93-108.
- Webbink, D. (2005). Causal effects in education. *Journal of Economic Surveys*, 19, (4), 535-560.
- Westerbeek, K. (1999). *The colours of my classroom. A study into the effects of the ethnic composition of classrooms on the achievement of pupils from different ethnic backgrounds*. Rotterdam: CED.
- Wilkinson, I., Hattie, J., Parr, J., Townsend, M., Fung, I., Ussher, C., Thrupp, M., Lauder, H., & Robinson, T. (2000). *Influence of peer effects on learning outcomes: A review of the literature*. Auckland, New Zealand: Auckland UniServices Limited.
- Wilkinson, I., Parr, J., Fung, I., Hattie, J., & Townsend, M. (2002). Discussion: Modelling and maximizing peer effects in school. *International Journal of Educational Research*, 37, 521-535.
- Winston, G., & Zimmerman, D. (2003). *Peer effects in higher education*. Cambridge, MA: National Bureau of Economic Research.
- Zeidner, M., & Schleyer, E. (1998). The big-fish-little-pond effect for academic self-concept, test anxiety, and school grades in gifted children. *Contemporary Educational Psychology*, 24, 305-329.
- Younger, M., & Warrington, M. (2005). *Raising boys' achievement*. Nottingham: DfES.
- Zimmer, R., & Toma, E. (2000). Peer effects in private and public schools across countries. *Journal of Policy Analysis and Management*, 19, (1), 75-83.

Zimmerman, D. (2003). Peer effects in academic outcomes: Evidence from a natural experiment. *The Review of Economics and Statistics*, 85, (1), 9-23.

