

Het einde van de klassenstrijd?

Achtergronden bij de discussie over klassengrootte in het basisonderwijs

R.J. Bosker*

Samenvatting

Recente peilingen laten zien dat de gemiddelde klassengrootte in basisscholen groot is vergeleken met andere landen en dat er bovendien een enorme variatie in klassengrootte is binnen, zowel als tussen scholen. Er wordt een samenvattend overzicht geboden van buitenlands onderzoek naar de effecten van klassengrootte op het functioneren van leerlingen en leerkrachten. Verder komt een grootschalig Nederlands onderzoek aan bod waarin klassengrootte centraal stond. Dit onderzoek laat zien dat er een samenhang bestaat tussen klassengrootte en leerlingprestaties op het terrein van taal en rekenen. Leerlingen uit leerjaar 2 die in klassen zitten met 25 of meer leerlingen hebben een grotere kans onder hun niveau te presteren dan vergelijkbare leerlingen uit kleinere klassen. Voor de leerjaren 4, 6 en 8 zijn de resultaten minder eenduidig. Zo is er bijvoorbeeld de bevinding dat in klassen met 30-34 leerlingen relatief hogere rekenprestaties behaald worden dan in kleinere klassen. Desalniettemin is er een duidelijke indicatie dat klassen met 35 of meer leerlingen vergelijken-derwijze niet gunstig zijn. Voor welbevinden van leerlingen, noch voor de onderwijsattitude, noch voor de arbeidssatisfactie van de leerkracht kan een verband met klassengrootte worden aangetoond. In de nabeschuiving, waarbij de effecten van klassengrootte vergele-

ken worden met de effecten van 'Reading Recovery' en het 'Succes for All' interventieprogramma, wordt de hypothese geopperd dat het niet zozeer de klassengrootte is die de genoemde effecten teweeg brengt, maar meer algemeen de leerling/leerkracht ratio.

Inleiding

Van oudsher verheugt het probleem van de klassengrootte zich in een bijzondere belangstelling (zie bijv. Van Gelder, 1959). Het ideaal dat een enkele onderwijskundige in het begin van deze eeuw schetst, namelijk 20 leerlingen per klas (idem, p. 500), komt steeds dichterbij: rond 1900 bevatte meer dan de helft der klassen 40 leerlingen of meer, in 1960 bedroeg de gemiddelde klassengrootte¹ nog 36 leerlingen en sinds 1972 ligt het gemiddelde onder de 30 (Warries, 1985, p. 7).

Met het verschijnen van het inspectierapport 'Groeps-grootte in het basisonderwijs' (Inspectie van het Onderwijs, 1995) werd evenwel duidelijk dat er zich zowel tussen als binnen basisscholen in Nederland enorme verschillen voordoen in klassengrootte. Er zijn klassen met minder dan 10 en klassen met meer dan 40 leerlingen. Op grote scholen (met meer dan 200 leerlingen) bevat de gemiddelde klas 10 leerlingen meer dan op de kleine scholen (met minder dan 100 leerlingen). De gemiddelde klassengrootte bedroeg in mei 1995 25,7 leerlingen. Een kwart van de klassen herbergt meer dan 30 leerlingen gedurende het laatste deel van het schooljaar. Dit impliceert dat, als we even afzien van leerkrachten met een part-time aanstelling, 1 op de 4 leerkrachten met een klas van meer dan 30 leerlingen te maken heeft, en dat – bij benadering – 1 op de 3 leerlingen in een klas van een dergelijke omvang zit.

In dit artikel zullen we ingaan op de volgende vragen:

1. Wat moeten we precies verstaan onder het – op het eerste gezicht zo eenduidig lijkende – begrip klassengrootte?
2. Wat is internationaal gezien de situatie met betrekking tot de grootte van klassen in het basisonderwijs?
3. Hoe komt het dat sommige klassen groter zijn dan andere?
4. Wat is bekend over de gevolgen van klassengrootte?

* Met dank aan J. Hox, K.-J. Polder en S. Veenman voor hun constructieve bijdragen aan de inhoud van dit artikel.

Klassengrootte en de leerling/ leerkracht ratio

Bij de (vaak slechts impliciete) definities van klassengrootte moeten we een onderscheid maken in de aggregatieniveaus en in de mate van specificiteit.

Op het hoogste aggregatieniveau, namelijk het macro-niveau, is de leerling/leerkracht ratio voor een bepaalde sector (in dit geval het basis-onderwijs) gebaseerd op een deling van het aantal leerlingen in de betreffende sector op het aantal personeelsleden in die sector. De kerngegevens primair onderwijs (Ministerie van OC&W, 1996) vermelden voor het basisonderwijs bijvoorbeeld een aantal van 19.3 leerlingen per full-time equivalent personeel in het onderwijs. Op het niveau van de school, het meso-niveau, wordt de leerling/leerkracht ratio analoog berekend. In het wat oudere, economisch getinte onderzoek wordt vaak deze indicator gebruikt, die vervolgens gerelateerd werd aan de prestaties van leerlingen uit een bepaald leerjaar (bijv. Coleman c.s., 1966). Op het micro-niveau is het een telling van het aantal leerlingen per klas, zoals dit ook in het inspectie-onderzoek is gebeurd. De inspectie definieerde een klas als het aantal leerlingen dat voor tenminste 50% van het aantal lessen in de week gelijktijdig les krijgt van één leerkracht (dat hoeft overigens niet één en dezelfde te zijn, gelet op duobanen, deeltijdarbeid, adv-ervanging e.d.). De macro- en mesoniveau benadering leveren een onderschatting van de situatie op: niet alle formatie wordt ook daadwerkelijk ingezet voor instructie en begeleiding (te denken valt aan: managementtaken, taakuren, e.d.). In de praktijk levert de micro-niveau benadering, althans zoals gehanteerd door de inspectie, een overschatting op: vaak worden bijvoorbeeld combinatieklassen gedurende een aantal lessen gesplitst in kleinere instructiegroepen.

Wat betreft specificiteit treden er in de praktijk ook grote verschillen in de gehanteerde definities op. Grofmazig zijn dan de micro-niveau definities (om nog maar te zwijgen over de meso- en macro-niveau definities) die niet tijdseenheden en vakinhouden specificeren.

Veel specifiekier zijn Glass et al. (1982). Deze auteurs hebben de bevindingen rond klassengrootte geïnventariseerd. Als we dit standaardwerk bekijken, dan blijkt dat het vraagstuk van de klassengrootte wordt begrepen als de vraag naar de instructiesituatie: gegeven een criterium (bijvoorbeeld rekenprestaties), wat is dan het effect van de omvang van de groep waarin een leerling in een bepaalde tijdsspanne instructie en begeleiding bij de verwerking van de rekenleerstof krijgt?

Zij die beweren dat op basis van het beschikbare onderzoek de conclusie gerechtvaardigd is dat reductie van de klassengrootte gunstig uitpakt voor de cognitieve ontwikkeling van leerlingen, verwijzen meestal naar de meta-analyse van Glass en Smith (1978). We komen hierop nog terug. Zij die beweren dat klassengrootte er niet zoveel toe doet, verwijzen meestal naar onderzoek waar veel grofmaziger te werk is gegaan; met name Hanushek (1986) wordt in dit verband veel geciteerd. Tenslotte zijn er nog die beweren dat grote klassen beter zijn en zij voelen zich daarbij gesteund door een onderzoek van Van Batenburg, Koster en Ros (1989).

Niet altijd echter wordt dezelfde definitie van klassengrootte gehanteerd. Soms is het de leerling/leerkracht ratio, soms is het de klassengrootte van een jaargroep en dan weer wordt het gepreciseerd naar vak en tijd en instructie-arrangementen. In het verloop zal ook duidelijk worden dat de conclusies van Batenburg c.s. alleen onder zeer bepaalde condities valide zijn, dat Hanushek de situatie onderschat, en dat Glass c.s. de positieve effecten van verkleining van de klassengrootte overschatten.

Internationale gegevens over klassengrootte

Het gemiddelde van bijna 26 leerlingen per klas in Nederland is internationaal gezien hoog. De IEA publiceerde op basis van de 'Reading Literacy' studie internationale gegevens over klassengrootte en leerling/leerkracht ratio (Ross & Postlethwaite, 1994), die in Tabel 1 gepresenteerd worden.

De eerste kolom met gegevens bevat de

Tabel 1
Enkele internationale gegevens over het basisonderwijs (bron: Ross & Postlethwaite, 1994, p. 129)

land	klassengrootte	leerling/ leerkracht ratio
België (Wallonië)	20.3	16.9
Denemarken	17.2	14.9
Duitsland (BRD)	22.5	22.0
Finland	24.6	19.6
Frankrijk	23.5	21.9
Hongarije	23.4	16.0
Ierland	31.0	30.6
Italië	16.0	16.8
Nederland	25.7	22.5
Nieuw-Zeeland	29.7	24.3
Noorwegen	17.7	12.1
Spanje	28.8	24.5
Verenigde Staten	24.6	20.6
Zweden	20.1	13.1
Zwitserland	19.6	14.5

gemiddelde klassengrootte. In een aantal landen (Denemarken, Italië, Noorwegen en Zwitserland) bevat de gemiddelde klas minder dan 20 leerlingen. In Ierland, Nieuw-Zeeland en Spanje ligt dit aantal bijna 10 hoger. Ook Nederland scoort met 25.7 leerlingen aan de hoge kant.

In de tweede kolom staat de leerling/leerkracht ratio. Gemiddeld is er in Nederland op elke 22.5 leerlingen één leerkracht werkzaam. De ongunstigste ratio wordt gerealiseerd in Ierland met 30.6, terwijl België, Denemarken, Hongarije, Italië, Zweden en Zwitserland ratio's van ongeveer 15 leerlingen op 1 leerkracht kennen.

Nu moeten deze cijfers met de nodige voorzichtigheid gehanteerd worden. Ierland, bijvoorbeeld, geeft relatief meer geld uit aan onderwijs dan Nederland, maar omdat de relatieve omvang van de populatie kinderen van 5 tot 14 jaar in Ierland zeer groot is, is er per leerling minder te besteden. Anderzijds hebben landen als Denemarken en Italië het speciaal onderwijs voor een belangrijk deel geïntegreerd in het reguliere onderwijs, zodat hier de klassen om deze reden kleiner kunnen uitvallen. Desalniettemin lijkt zich de conclusie op te dringen: vrij grote klassen in het Nederlandse basisonderwijs. Grotere klassen impliceren minder uitgaven voor leerkrachtsalarissen. Men zou hier dus van een grote doelmatigheid kunnen spreken, als Nederlandse leerlingen tenminste goed zouden presteren. Maar wat betreft het leesvaardigheidsniveau van onze 9-

jarigen verkeert Nederland in de achterhoede, blijktens recente IEA-onderzoeken (zie bijv. De Gloppe & Otter, 1993, p. 15) en de Commissie Evaluatie Basisonderwijs (1994) sprak haar zorgen uit over het rekenpeil.

De oorzaken van verschillen in klassengrootte

Hoe kan het dat sommige klassen groter zijn (of lijken) dan andere? De verschillen in leerling/leerkracht ratio's tussen landen, en dus ook in klassengroottes, zijn het directe gevolg van beleidskeuzen van de overheid. Sommige landen investeren meer in het onderwijs dan andere landen, stellen met andere woorden meer geld en dus meer formatie beschikbaar (Organization for Economic Co-operation and Development, 1995). Vergeleken met andere OESO-landen (gemiddeld 6.1%) besteedt Nederland relatief een gering percentage (5%) van het bruto nationaal product aan het onderwijs. De verdeling van deze middelen over de diverse sectoren volgt echter nauwgezet het gemiddelde OESO-patroon: Voor alle sectoren (primair, secundair en tertiair) geldt dat de uitgaven onder het OESO-gemiddelde liggen. De spreiding in klassengrootte binnen landen is vervolgens afhankelijk van een tweede beleidskeuze: Worden er wettelijke regels gesteld ten aanzien van de klassengrootte? In Tabel 2 zijn voor een aantal landen de wettelijke maximale klassengroottes aangegeven.

Klassen groter dan 35 kunnen in Grieken-

Tabel 2
Wettelijk bepaalde maximale klassengroottes in enkele landen (bron: RISE, 1994; OESO, 1995, p. 287)

land	maximale klassengrootte
Denemarken	28
Duitsland	25 voor jaargroep 1 en 2 29 daarna
Finland	25 voor jaargroep 3 en 4 32 daarna *
Frankrijk	25 jaargroep 1
Griekenland	35
Italië	20 voor klassen met een gehandicapt kind 25 anders
Noorwegen	28
Schotland*	25 voor combinatieklassen 33 anders

*) Niet bij wet geregeld, maar in de CAO vastgelegd.

Tabel 3

Aantal klassen en klassengrootte per type school (bron: Van der Werf & Guldemond, 1994, p. 35)

type school*	aantal klassen	gemiddelde klassengrootte	beschikbare extra formatie per klas
A. Zonder OVB-faciliteiten (schoolscore < 105)	7	25	0
B. Gebiedsscholen met $105 < \text{schoolscore} \leq 115$	6	22	0
C. Niet-gebiedsscholen met $105 < \text{schoolscore} \leq 115$	6	22	0
D. Gebiedsscholen met schoolscore > 115	8	21	0.2
E. Niet-gebiedsscholen met schoolscore > 115	7	22	0.1

*) schoolscore = $100 \times (\text{gewogen leerlingaantal} / \text{ongewogen leerlingaantal} \text{ minus een correctiefactor van } 9\%)$

land dus niet voorkomen, omdat dit wettelijk verboden is. Voor Denemarken en Noorwegen ligt het wettelijk bepaalde maximum zelfs op 28. Andere landen leggen de maxima vast afhankelijk van de jaargroep, met lagere maxima voor de jaargroepen die de jongere leerlingen herbergen. In Schotland is een en ander niet bij wet geregeld maar door middel van afspraken tussen de sociale partners vastgelegd. Bijzonder is bovendien dat in Schotland een apart lager maximum is overeengekomen voor combinatieklassen (25 tegenover 33 voor enkelvoudige klassen).

Dat er verschillen tussen scholen in Nederland bestaan in de gemiddelde klassengrootte, heeft te maken met de formatietoekenning, waarbij zowel de schoolgrootte als de samenstelling van de leerlingpopulatie een rol speelt. Kleine scholen, met minder dan 100 leerlingen, krijgen relatief veel, en grotere scholen relatief weinig formatie. Zouden scholen alle formatie behoeve van de minimalisering van de klassengrootte inzetten, dan kunnen scholen met minder dan 100 leerlingen hun gemiddelde klassengrootte op 18.1 stellen. Voor scholen met 100-200 leerlingen, respectievelijk met meer dan 200 leerlingen liggen deze getallen op 23.2, respectievelijk 24.2. Scholen met achterstandsleerlingen, dat wil zeggen kinderen uit de lagere sociale milieus van allochtone (1.9) of autochtone ouders (1.25), krijgen daarnaast extra formatie. Dit kan ertoe leiden dat op 'zwarte' scholen bijna twee keer zo veel formatie beschikbaar is als op 'witte' scholen. Deze extra formatie wordt in 46% van de gevallen geheel, en in 44% van de gevallen gedeeltelijk

ingezet voor verkleining van de klassengrootte (Inspectie van het Onderwijs, 1995, p. 28).

Scholen hebben de vrijheid de toegekende formatie naar eigen goeddunken in te zetten voor onderwijs, begeleiding, taakuren, managementtaken e.d. Verschillen in klassengroottes binnen een school zijn enkel en alleen het gevolg van door de school zelf gemaakte keuzen:

- werken met een leerstofjaarklassysteem;
- het formeren van combinatieklassen;
- het al dan niet doorbreken van het klassenverband gedurende enkele uren voor wezenlijke onderwijsactiviteiten;
- de specifieke inzet van formatie voor achterstandsbestrijding.

Van der Werf en Guldemond (1994) onderscheiden 5 typen scholen op basis van de leerlinggewichten en participatie in een OnderwijsVoorangsGebied. Per type school hebben ze vervolgens het gemiddeld aantal klassen, de gemiddelde klassengrootte en de hoeveelheid extra beschikbare formatie per klas berekend (zie Tabel 3).

De gegevens van Van der Werf en Guldemond suggereren dat scholen hun extra formatie met name gebruiken om de klassen klein te houden. Dat gebiedsscholen met een schoolscore groter dan 115 gemiddeld 8 klassen formeren, en vervolgens formatie overhouden (die dan wellicht ingezet wordt voor remedial teaching, NT2, OET, e.d.) lijkt erop te wijzen dat scholen een dergelijk formatiegebruik kennelijk effectiever achten dan het formeren van

kleinere combinatieklassen. Belangrijke overwegingen bij de indeling van de klassen zijn: zo veel mogelijk combinatieklassen vermijden, het kunnen opvangen van tussentijdse instroom, het klein houden van jaargroep 3, en het klein houden van klassen in verband met zorgverbreding (Inspectie van het Onderwijs, 1995, p. 31, 32).

Kennelijk achten scholen het in verband met achterstandsbestrijding effectief de klassen klein te houden, en achten ze grotere, enkelvoudige klassen effectiever en/of beter hanteerbaar dan kleinere combinatieklassen.

De gevolgen van klassengrootte voor leerlingprestaties

Enkele inleidende opmerkingen

In deze paragraaf brengen we de effecten van klassengrootte op leerprestaties en het welzijn van leerlingen en het functioneren van leerkrachten in kaart op basis van eerder en zelf (Bosker & Hox, 1996) verricht onderzoek.

De vraag of leerlingen in grotere klassen slechter presteren, meer faalangst hebben en minder gemotiveerd zijn, kan alleen door experimenteel onderzoek afdoende worden beantwoord. Hetzelfde geldt als we de gevolgen van klassengrootte voor het leerkrachtfunctioneren willen vaststellen. In zo'n experiment, met als condities grote en kleine klassen, zouden leerlingen en leerkrachten aselekt aan klassen moeten worden toegewezen. Op die manier worden andere storende invloeden (de kwaliteit van de leraar, de bewuste verdeling van goede en slechte leerlingen in grote en kleine klassen, etc.) geneutraliseerd. En dan nog zijn er problemen:

- er dreigt een Hawthorne-effect (als we gaan experimenteren met kleine klassen, dan gaan deze zich als gevolg van de aandacht die ze krijgen van de onderzoeker gunstiger ontwikkelen) of tegengesteld daaraan;
- een John Henry-effect (de leerkrachten in de controleconditie, d.w.z. met de grote klassen, gaan onder invloed van het onderzoek extra hun best doen).

Experimenteel onderzoek is een zeldzaamheid in het onderwijs. Meestal wordt gebruik gemaakt van observationele onderzoeksontwer-

pen. Dat wil zeggen: we kijken of er klassen zijn die verschillen in omvang en of de kinderen in de grotere klassen anders functioneren dan de kinderen in de kleinere klassen. In dat geval hebben we geen last van de net beschreven problemen, maar nu zijn er weer andere:

- de context van de klas, in casu de school speelt een rol. Stel dat bijvoorbeeld de ervaren leerkrachten de kleinere klassen hebben; als de kinderen in de kleinere klassen het beter doen, kan dat komen omdat de leerkrachten zo ervaren zijn, en niet omdat de klas zo klein is;
- de leerlingen zelf dragen een onderwijsverleden met zich mee. Als de school bijvoorbeeld besluit de minder gevorderde leerlingen in de kleinere klassen te zetten, dan zouden we kunnen vinden dat in de grotere klassen de kinderen beter presteren.

Aan het slot zullen we op dit probleem van causale attributie (wat is de 'echte' oorzaak?) terugkomen, wanneer de bevindingen worden samengevat en gewogen worden op hun implicaties.

De meta-analyse van Glass en Smith

Honderden studies naar klassengrootte werden door Glass en Smith (1978) geïnventariseerd en aan een nadere meta-analyse onderworpen om te bezien wat nu de beste samenvatting van al die resultaten zou zijn. Het resultaat ten aanzien van klassengrootte-effecten op leerprestaties lijkt in eerste instantie weinig veelzeggend: 40% van de 725 studies toont voordelen voor de grotere klassen en 60% voor de kleinere. De gemiddelde effectgrootte² die Glass en Smith rapporteren, is 1/10e van een standaarddeviatie (0.09 om precies te zijn) in het voordeel van de kleinere klassen. Maar dat voordeel kan tot stand zijn gekomen door de zeer kleine groepen.

De studie van Glass en Smith is niet zonder kritiek gebleven. Met name de in de meta-analyse dominerende rol van de onderzoeksresultaten van studies waar zeer kleine klassen met minder dan vijf leerlingen werden vergeleken met grote klassen werd bekritiseerd. Bovendien zouden meerdere studies door Glass en Smith meermalen zijn gebruikt (omdat bijvoorbeeld zowel reken- als taalprestaties werden onderzocht) en betroffen sommige studies irre-

Tabel 4

Klassengrootte effecten op leerprestaties in primair en secundair onderwijs bij experimentele onderzoeken (bron: Slavin, 1989, p. 249); tussen haakjes staan de uitgevoerde vergelijkingen

	Omvang van de kleine klas (vergelijking)			
	1	3	14-17	20-23
	.65 (1-32)	1.22 (3-25)	.17 (14-30)	.15 (20-28)
	.78 (1-30)		.17 (15-30)	.04 (23-27)
	1.52 (1-25)		.08 (16-37)	.04 (23-30)
	.72 (1-14)		.04 (16-30)	.00 (30-37)
	.30 (1-8)		.05 (16-23)	
	.22 (1-3)		-.29 (17-35)	
mediane effectgrootte	.69		.06	.04

levante domeinen (in één geval ging het om tennislessen). Reden voor Slavin (1989) nog eens naar de oorspronkelijke onderzoeken te kijken. Van de 77 oorspronkelijke studies bleken er slechts in veertien gevallen leerlingen aselekt aan grote dan wel kleine klassen toegevoerd te zijn. Slechts veertien experimentele studies dus. Na verwijdering van de tennislessen-studie, de studies die betrekking hadden op post-secundair onderwijs en studies waar het onderwijs slechts 30 minuten duurde, komt Slavin tot de volgende samenvattende resultaten (zie Tabel 4).

Tabel 4 laat zich als volgt lezen. In de eerste kolom staan de resultaten van studies vermeld waar individueel onderwijs vergeleken is met onderwijs in groepen. Voor de eerste studie waar individueel onderwijs wordt afgezet tegen onderwijs in een klas van 32 leerlingen (links bovenin de tabel terug te vinden als 1-32), is de effectgrootte 0.65.

In de onderste rij wordt de mediane effectgrootte gegeven als de kleine klas vergeleken wordt met de grotere klassen. Op basis van deze gegevens komt Slavin tot de conclusie dat de kleinere klassen (14 tot 17 leerlingen of 20 tot 23 leerlingen) het amper beter doen dan de grotere klassen: 0.04 respectievelijk 0.06 is de effectgrootte. En waar de kleinere klassen het beter doen, blijken, aldus Slavin, op de lange duur de effecten weg te ebbelen. De enige consequente effecten zijn te vinden als een leerling individueel in plaats van in een groep onderwijs ontvangt: 0.69 is hier de effectgrootte.

Sinds de meta-analyse van Glass en Smith en de kritiek die daarop is gekomen is er echter een aantal studies naar klassengrootte uitge-

voerd die nieuw licht werpen op de zaak en/of betrouwbaardere uitkomsten hebben opgeleverd. We gaan kort op deze studies in.

Een Nederlandse studie

In Nederland is in 1989 een onderzoek naar klassengrootte gerapporteerd door Van Batenburg e.a. (1989). De onderzoekers maken gebruik van de Cito-Eindtoetsgegevens van ruim 800 scholen, die ze vervolgens relateren aan de klassengrootte van jaargroep 8. Zij vinden in een analyse op het niveau van de school een positief verband tussen klassengrootte en de Cito-Eindtoetsprestaties, ook als er wordt gecorrigeerd voor de samenstelling van de schoolbevolking in termen van de sociaal-economische positie van de ouders der leerlingen. Zouden we de gerapporteerde resultaten omzetten in effectgroottes, dan zijn deze negatief: -0.10. Om twee redenen hebben deze resultaten een beperkte reikwijdte:

1. De Cito-Eindtoets meet het resultaat van acht jaar basisonderwijs, terwijl de klassengrootte van jaargroep 8 slechts een achtste deel omvat van de 'behandeling' die de kinderen acht jaar lang gekregen hebben. Wellicht worden grotere klassen in jaargroep 8 juist gevormd om kleine klassen in jongere jaargroepen te kunnen maken en vice versa. In dat geval zou het effect dus het gevolg kunnen zijn van de kleinere klasomvang in de jongere jaargroepen. Een dergelijke alternatieve verklaring is slechts uit te sluiten als alleen de leerwinst in het achtste leerjaar geanalyseerd zou worden.
2. Analyses op schoolniveau zijn uitermate gevaarlijk als we er valide conclusies uit

willen trekken over het individuele functioneren van leerlingen. Corrigeren voor de sociaal-economische samenstelling van de schoolbevolking heeft slechts een minimaal effect en is – conceptueel en statistisch – heel iets anders dan corrigeren voor de sociaal-economische achtergrond van individuele leerlingen (zie Bosker & Sniijders, 1990). Dergelijke analyses op schoolniveau worden wel getypeerd als 'gevaarlijk op zijn best, rampzalig op zijn ergst'³.

Een Australische studie

Larkin en Keeves publiceerden in 1984 een onderzoek naar klassengrootte in Australië. Ze maakten gebruik van gegevens uit 1969 over de prestaties van jaargroep 8 leerlingen in 'science' (een vak met onderdelen van natuurkunde, biologie, scheikunde en aardrijkskunde) en rekenen/wiskunde. Het onderzoek is tamelijk standaard in die zin dat het een observationeel onderzoeksontwerp betreft (gekeken wordt of klassen verschillen in grootte – men is dus gebonden aan de natuurlijke variatie die zich in dit opzicht in het onderwijsveld voordoet – en vervolgens wordt nagegaan of in de kleinere klassen andere dingen gebeuren, en of dat samenhangt met prestaties van leerlingen). De conclusies:

- de vaardiger leerlingen werden meestal in de grotere klassen geplaatst (zodat in de kleinere klassen de minder vaardige leerlingen meer begeleiding konden krijgen);
- in grotere klassen lag het prestatieniveau uiteindelijk hoger;
- in kleinere klassen was het verschil tussen oorspronkelijk hoge en lage presteerders uiteindelijk minder groot dan in de grotere klassen.

Plaatsing van de leerlingen in klassen van verschillende omvang blijkt in dit onderzoek dus rechtstreeks gebaseerd te zijn op het vaardigheidsniveau. Iets dergelijks zou zich ook in de Nederlandse situatie kunnen voordoen, waar immers dankzij de gewichtenregeling achterstandsleerlingen een grotere kans hebben in de kleinere klassen terecht te komen. Het onderzoek maakt duidelijk hoe problematisch het is om onderzoek te doen naar de effecten van klassengrootte als de verschillende klassen systematisch verschillen in hun samenstelling.

Onderwijseconomisch onderzoek naar klassengrootte

In onderwijseconomische kringen wordt nogal eens verwezen naar een review van Hanushek (1986), waarin de auteur 112 studies naar leerkracht/leerling ratio op een rij zet (de economisten draaien de actoren om bij de ratioberekening): 14 studies leveren een significant negatief verband op en 9 een significant positief verband. Kortom: onbeslist. Hedges, Laine en Greenwald (1994) onderwierpen de studies aan een statistische meta-analyse. Deze auteurs komen op grond van hun analyses tot de slotsom, dat het niet is uit te sluiten dat een gunstige leerkracht/leerling ratio positieve effecten heeft. Zij trekken uiteindelijk de conclusie dat uitgaven per leerling een duidelijk verband met leerlingprestaties laten zien, maar dat de samenhang tussen de leerkracht/leerling ratio en de leerlingprestaties minder eenduidig is: het hangt kennelijk van de specifieke omstandigheden van een school af hoe het geld het best gebruikt kan worden.

De review van Hanushek was voor Akerhielm (1995) aanleiding te onderzoeken of in observationeel onderzoek kleinere klassen niet goed uit de verf komen, omdat zij vaker de minder vaardige leerlingen herbergen. Een eerste heranalyse van een dataset waarin meer dan 1000 scholen vertegenwoordigd zijn met elk 24 leerlingen uit jaargroep 8, laat zien dat voor klassen tot 20 leerlingen het klassengrootte-effect positief is. In het gebied tussen de 21 en 30 is er geen effect. Wanneer de onderzoekster echter controleert voor de verschillen in klassensamenstelling qua oorspronkelijk vaardigheidsniveau blijken er omgekeerde effecten op te treden voor twee van de vier onderzochte vakken (namelijk voor geschiedenis en science, maar niet voor rekenen en taal): kleinere klassen doen het dan iets beter in geschiedenis en science. De richting van het verband tussen klassengrootte en prestaties is dus soms inderdaad afhankelijk van het feit of rekening wordt gehouden met de samenstelling van de klas.

Het Prime Time onderzoek

In Indiana (VS) werd in 1984 Prime Time ingevoerd, een beleidsprogramma bedoeld om het kennisniveau van de leerlingen te verhogen door de klassen te verkleinen. In het schooljaar

84/85 werd begonnen met de reductie van de klassengrootte in jaargroep 3 ('1st grade'), waarna in 85/86 jaargroep 4 en 86/87 jaargroep 5 en jaargroep 2 ('Kindergarten') voor het eerst onder het beleidsprogramma vielen. Door de kleinere klassen (met 19.1 leerlingen gemiddeld) te vergelijken met de grotere klassen (met 26.4 leerlingen gemiddeld) van vóór de implementatie van Prime Time kan een schatting van het effect van reductie van de klassengrootte worden gemaakt. In de controlescholen die altijd al kleine klassen hadden gehad, bleek in twee jaar tijd niets veranderd te zijn (McGivern, Gilman & Tilitzki, 1989). In de scholen waar de klassengrootte was verkleind, deden de kleinere klassen het duidelijk beter na twee jaar. Als we de vastgestelde effectgrootte voor rekenen en taal samen bezien, dan is deze 0.34.

Het STAR-experiment

Een experimenteel onderzoek naar klassengrootte is uitgevoerd in Tennessee in de Verenigde Staten van Amerika (Word, 1990). De aanleiding vormde een politieke discussie over de enorme kosten van verkleining van de klassengrootte in verhouding tot de onzekere opbrengsten daarvan. Om uit de impasse te geraken werd besloten tot een nadere studie, die anders dan gebruikelijk in onderwijsonderzoek, als een echt experiment werd opgezet. 79 scholen werden – allesbehalve aselekt, namelijk op basis van vrijwilligheid en voldoende grootte van de jaargroepen – geselecteerd voor het onderzoek. Binnen elke school – dit is qua onderzoeksontwerp een sterk punt, omdat de grote en kleine klassen functioneren binnen dezelfde context – werden drie typen kleuterklassen ('Kindergarten') geformeerd: een klas van 13-17 leerlingen, een klas van 22-25 leerlingen en een klas van 22-25 leerlingen waar een full-time onderwijs-assistent zou gaan werken. Leerlingen en leerkrachten werden aselekt aan de klassen toegewezen. De klassen bleven vier jaar lang (tot en met '3rd grade', ofwel jaargroep 5) grotendeels intact (tussentijdse instroom of uitval had geen invloed op de resultaten van het experiment). De statistische analyses zijn geheel comme-il-faut, en doen dus recht aan het feit dat toeval een rol speelt op het niveau van leerlingen én klassen én scholen. De resultaten in termen van effectgroottes zijn weergegeven in Tabel 5.

Tabel 5

Het voordeel van kleine klassen volgens het STAR-experiment (naar Word, 1990, p. 24)

	taal	rekenen
jaargroep 2	.21	.17
jaargroep 3	.34	.33
jaargroep 4	.26	.23
jaargroep 5	.24	.21

Tabel 5 bevat de gegevens over één generatie leerlingen die vier jaar lang gevolgd zijn op hun weg door het onderwijs. De grote positieve effecten van de kleine klas aan het einde van jaargroep 3 (0.34 voor rekenen en 0.33 voor taal) houden niet ten volle stand tot aan het eind van jaargroep 5, maar zijn nog steeds aanzienlijk (0.24 voor rekenen en 0.21 voor taal).

Zowel voor rekenen als voor taal zijn de resultaten in het voordeel van kleine klassen (gemiddeld 15 leerlingen). Grote klassen (met gemiddeld 23 leerlingen) met en zonder onderwijsassistent (die onderwijsassistent heeft dus geen effect in deze studie) doen het beide slechter dan de kleine klassen. Vervolgens zijn analyses uitgevoerd om te onderzoeken of het effect van klassengrootte verschillend is:

- voor scholen in verstedelijkte en plattelandsgebieden,
- voor jongens en meisjes,
- voor oorspronkelijk (d.w.z. in het begin van de kleuterklas) hoog en laag presterende leerlingen,
- voor hoog- en laag-milieu leerlingen,
- voor alloctonen en autoctonen.

Kleine klassen blijken het onder alle condities beter te doen dan grote klassen. Finn en Achilles (1990) laten bovendien zien dat bij het STAR-experiment de effecten van kleine klassen voor kansarme leerlingen zelfs nog gunstiger uitpakken dan voor hun beter bedeelde medeleerlingen.

Vervolgonderzoek (Achilles et al., 1993) op het STAR-project bracht bovendien nog twee interessante zaken aan het licht:

1. Ook nadat de leerlingen uit de kleine klassen na het 5e leerjaar teruggekeerd waren naar klassen met een voor Tennessee normale omvang bleef hun voorsprong op de kinderen uit de van oorsprong grotere klassen behouden;
2. Omdat STAR succesvol bleek, ging men

over tot de invoering van zeer kleine klassen (13-17 leerlingen) in 17 arme districten met veel risicoleerlingen. Scoorden deze districten qua taal- en rekenprestaties altijd ver onder het gemiddelde van Tennessee, na de klassenverkleining scoorden ze daar (net) boven (Mosteller, 1995).

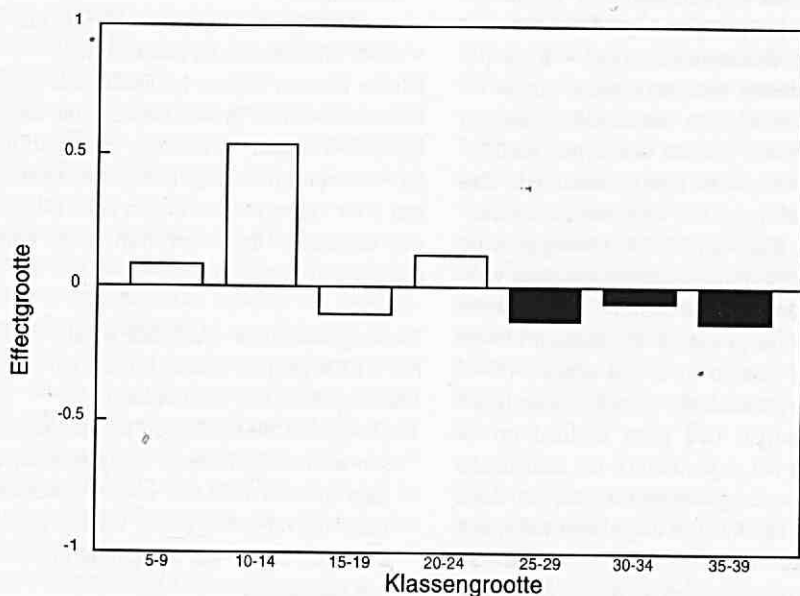
Klassengrootte en leerprestaties: de Nederlandse situatie

Om zicht te krijgen op de samenhang tussen klassengrootte en leerprestaties in Nederland is door Bosker en Hox (1996) een heranalyse uitgevoerd op de gegevens, zoals die verzameld worden in het kader van het zogenaamde PRIMA-cohort onderzoek (zie bijv. Jungbluth, Peetsma & Roeleveld, 1996). Het PRIMA-onderzoek heeft onder andere betrekking op 416 aselekt gekozen basisscholen die samen de zogenaamde referentiesteekproef vormen. In het schooljaar 94/95 werden de leerkrachten die het schooljaar daarvoor les hadden gegeven aan de jaargroepen 1, 3, 5 en 7 schriftelijk bevraagd, o.a. over de klassengrootte. Van alle leerlingen werd in het begin van het schooljaar 94/95 (maar voor de jaargroep-2 leerlingen pas halverwege het schooljaar) in kaart gebracht: leerlinggewicht, sekse, reken- en taalprestaties. Bovendien werd het IQ van de leerlingen in de

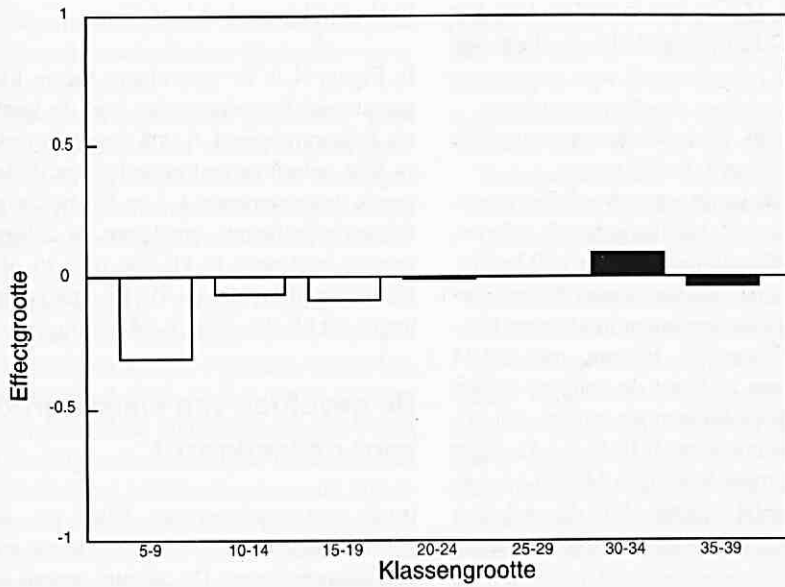
jaargroepen 4, 6 en 8 gemeten (voor de constructie van deze toetsen zie: Vierke, 1995). De samenhangen tussen klassengrootte en leerlingprestaties die hier gerapporteerd worden, zijn alle vastgesteld na correctie voor de achtergrondkenmerken van leerlingen en de samenstelling van de klas. In feite wordt dan de samenhang onderzocht tussen de klassengrootte en de mate waarin leerlingen boven of onder (de statistische) verwachting presteren, gegeven het IQ, het leerlinggewicht en de sekse van de leerling en gegeven de samenstelling van de klas qua IQ, leerlinggewicht en sekse. Nagegaan wordt dan of er een omslagpunt te vinden is in de klassengrootte waarboven de prestaties van leerlingen achterblijven bij vergelijkbare leerlingen uit de kleinere klassen.

In Figuur 1 is de samenhang tussen klassengrootte en rekenprestaties van de jaargroep-2 leerlingen afgebeeld.

In Figuur 1 zijn op de horizontale as 7 verschillende klassengroottes afgebeeld: 5-9, 10-14, 15-19, 20-24, 25-29, 30-34, en 35 of meer leerlingen. Op de verticale as loopt de effectgrootte. Als referentiepunt is gekozen: het gemiddelde niveau van klassen ter grootte van 20-24 en 25-29 leerlingen. Daarom ligt de staaf voor klassen met 25-29 leerlingen ook net zoveel onder de as als de staaf voor klassen met 20-24 leerlingen er boven ligt. Witte balken



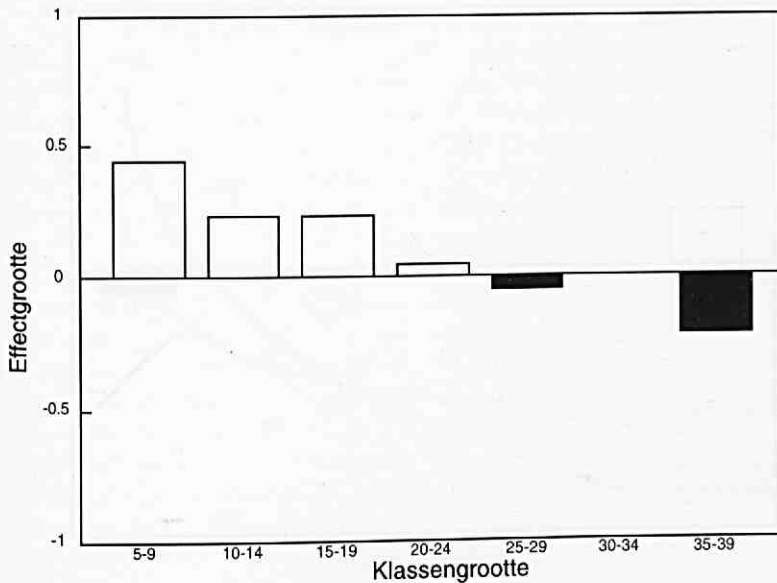
Figuur 1. Klassengrootte en rekenvaardigheid, jaargroep 2 (naar Bosker & Hox, 1996)



Figuur 2. Klassengrootte en rekenvaardigheid, jaargroep 4, 6 en 8 (naar Bosker & Hox, 1996)

geven aan dat de verschillen niet en zwarte balken dat de verschillen wel statistisch significant zijn. Statistische analyses laten zien dat jaargroep-2 leerlingen in klassen ter grootte van 25 of meer leerlingen minder goed rekenen dan leerlingen in kleinere klassen dan we, op basis van hun IQ, leerlinggewicht, sekse en de samenstelling van de klas waarin ze zitten, zouden mogen verwachten. Afgezet tegen de

rekenprestaties van de klassen met 20-24 leerlingen blijven leerlingen in klassen met 25-29 leerlingen -0.25 , leerlingen in klassen met 30-34 -0.18 en leerlingen in klassen met 35 of meer leerlingen -0.26 achter. Het rekenresultaat van de kinderen in klassen met 10-14 leerlingen lijkt boven dat van de overige klassen uit te steken, maar het gaat hier om een verschil dat niet de statistische toets der significantie



Figuur 3. Klassengrootte en taalvaardigheid, jaargroep 2 (naar Bosker & Hox, 1996)

kan doorstaan. Met andere woorden: hier kan heel goed toeval in het spel zijn in plaats van systematiek.

In Figuur 2 zijn de resultaten weergegeven voor de jaargroepen 4, 6 en 8.

Bij de hogere jaargroepen treedt het omgekeerde op van wat bij jaargroep 2 gebeurt. Leerlingen in klassen met meer dan 20 leerlingen doen het qua rekenprestaties boven verwachting beter dan leerlingen in kleinere klassen. De leerlingen in klassen met 30-34 leerlingen komen zelfs tot de hoogste scores: vergeleken met de klassen ter grootte van 20-24 leerlingen is hun score 0.10 hoger. De klassen met 35 of meer leerlingen blijven vervolgens weer -0.04 achter bij de gekozen referentieklassen (20-24 en 25-29 leerlingen).

In Figuur 3 zijn de taalprestaties van de leerlingen uit jaargroep 2 afgebeeld in samenhang met de klassengrootte.

Qua taalprestaties van de jaargroep-2 leerlingen blijken er twee breekpunten te zijn: bij klassen met 25-29 leerlingen (die -0.09 achterblijven bij klassen met 20-24 leerlingen) en bij klassen met 35 of meer leerlingen (de achterstand is hier -0.27). Ook hier geldt weer dat het verleidelijk is aan de taalprestaties in de hele kleine klassen aandacht te besteden, maar zulks is niet gerechtvaardigd, omdat het hier 'toeval-

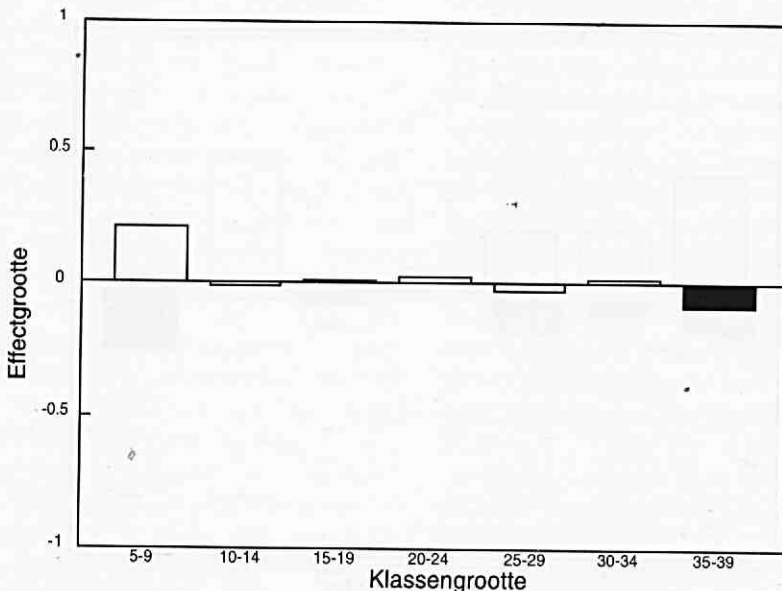
lige' afwijkingen kan betreffen.

In Figuur 4 is de samenhang tussen klassengrootte en de taalprestaties van de leerlingen uit de jaargroepen 4, 6 en 8 weergegeven.

Wat betreft de taalprestaties van de leerlingen in de jaargroepen 4, 6 en 8 is het enige statistisch significante breekpunt de achterstand van de leerlingen in klassen met 35 of meer leerlingen (deze blijven -0.11 achter op de leerlingen uit klassen met 20-24 leerlingen).

De gevolgen van klassengrootte voor de leerkracht

In de meta-analyses van Glass c.s. is ook gekeken naar de relatie klassengrootte en leerkrachtfunctioneren. De auteurs maken onderscheid tussen het algemene leerkrachtfunctioneren (waaronder: organisatie en planning, 'morale', houding tegenover leerlingen, tevredenheid, 'workload', afwezigheid, professionele groei) en instructiegedragingen (zoals: kennis omtrent leerlingen, hoeveelheid directe interactie, hoeveelheid aandacht voor individuele leerlingen, doelgerichtheid, gebruik van leermiddelen, adaptief onderwijs, contacten met ouders, gebruik van ruimte, innovatie e.d.). In totaal zijn 30 studies samengevat naar het algemene leerkrachtfunctioneren en 155 naar



Figuur 4. Klassengrootte en taalvaardigheid, jaargroep 4, 6 en 8 (naar Bosker & Hox, 1996)

instructiegedrag. In Figuur 5 zijn de resultaten van de geanalyseerde studies afgebeeld.

In Figuur 5 zijn twee lijnen getrokken. Voor beide lijnen geldt dat het willekeurige referentiepunt wederom een klas met 30 leerlingen betreft. De lijn die rechtsonder eindigt geeft de gevonden samenhang weer tussen klassengrootte en het algemene leerkrachtgedrag. Vergelijken met een klas van 30 leerlingen is de effectgrootte voor een leerkracht in een klas met 20 leerlingen qua tevredenheid, werkbelasting, absentie, e.d. 0.47. Anders dan in de voorgaande figuren blijkt klassengrootte ook boven de 30 leerlingen een duidelijke samenhang te vertonen met het leerkrachtgedrag: de effectgrootte voor een klas met 35 leerlingen is -0.13.

De tweede lijn in de grafiek geeft het verband tussen de klassengrootte en het instructiegedrag weer. De lijn loopt redelijk vlak in de buurt van de klassengroottes van 20 tot 40 leerlingen: er zijn verschillen ten gunste van de kleinere klassen binnen deze range, maar die verschillen zijn beperkt. De samenhang verandert echter snel naarmate de klassenomvang verder afneemt: de kleine klassen doen het duidelijk beter in termen van instructiegedrag.

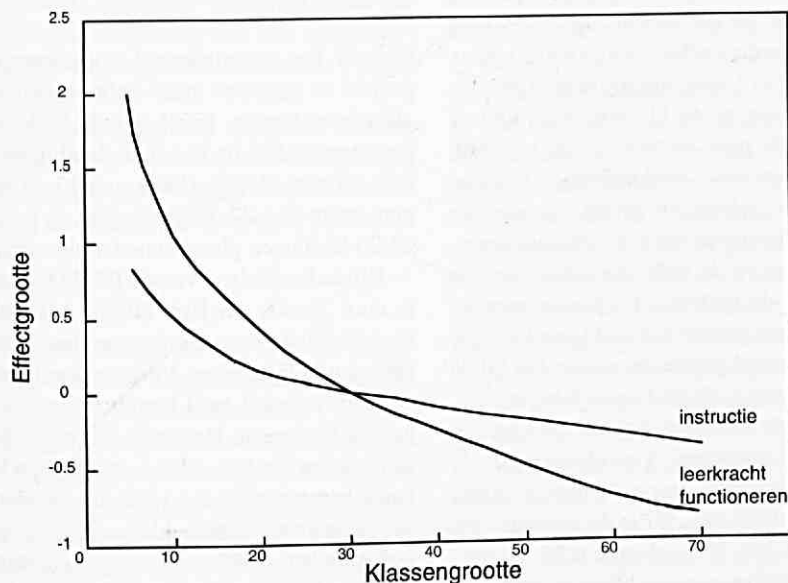
Ook ten aanzien van het leerkrachtfunctioneren zijn de PRIMA-gegevens geanalyseerd door Bosker en Hox (1996). De leerkrachten van de

jaargroepen 1, 3, 5 en 7 zijn namelijk bevraagd naar hun arbeidssatisfactie en hun onderwijsattitude (zie bijv. Overmaat & Ledoux, 1996):

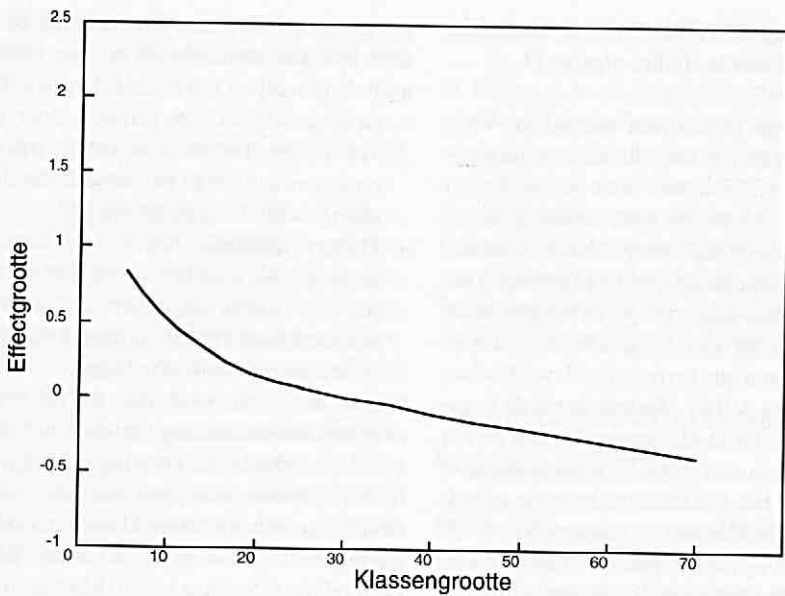
- Arbeidssatisfactie is gemeten door middel van vragen als 'ik voel me in mijn werk ondergewaardeerd' en 'ik vind de sfeer bij ons op school bijzonder prettig'.
- Onderwijsattitude betrof een schaal met items als 'ik vind het moeilijk eisen te stellen aan zwakke leerlingen' en 'ik heb vaak een hard hoofd in het vermogen van sommige leerlingen om wat te leren'.

Omdat niet is uit te sluiten dat de scores op deze schalen samenhang vertonen met de sekse van de leerkracht, de ervaring e.d., zijn dergelijke variabelen weer constant gehouden alvorens de samenhang tussen klassengrootte en het leerkrachtfunctioneren te schatten. Er blijkt geen relatie te bestaan tussen klassengrootte en de arbeidssatisfactie. Wel treedt er een verband op tussen de klassengrootte en de onderwijsattitude.

Het verband tussen klassengrootte en de onderwijsattitude blijkt met name veroorzaakt te worden door de afwijkende opvattingen van leerkrachten die klassen met minder dan 10 leerlingen onder hun hoede hebben. Aangezien dit een relatief zeldzame groep betreft, is er geen reden aan deze bevinding verder aandacht te schenken.



Figuur 5. Relatie tussen klassengrootte en instructie en klassengrootte en leerkracht functioneren (naar Glass c.s., 1982)



Figuur 6. Relatie tussen klassengrootte en affectief functioneren van leerlingen (naar Smith & Glass, 1980)

Klassengrootte en andere effecten

Klassengrootte en het affectief functioneren van leerlingen

In de meta-analyse van Glass c.s. (1982) is ook gekeken naar het affectief functioneren van leerlingen. Daarover is apart gerapporteerd door Smith en Glass (1980). Ze maken gebruik van 59 studies op dit terrein, die betrekking hebben op meerdere affectieve uitkomstmaten. Zo ontstaan er 172 vergelijkingsmogelijkheden tussen kleine en grote klassen, waarbij o.a. gekeken wordt naar motivatie, self-concept, houding tegenover leerkrachten, houding tegenover de school/de groep, geestelijke gezondheid, divergent denken, studiegewoonten, gerichtheid op de taak, discipline, apathie etc. 85% van de onderzochte studies rapporteert effecten ten gunste van de kleine klas. De resultaten zijn veel geprononceerder dan bij de cognitieve maten zoals uit Figuur 6 blijkt.

In Figuur 6 is als referentiepunt een klas van 30 leerlingen genomen. Vervolgens zijn de andere klassengroottes daartegen afgezet in termen van effectgroottes. Zo is de effectgrootte voor een klas met 20 leerlingen 0.13. Anders dan bij de cognitieve maten blijkt er geen volledige afname van het effect te zijn bij toene-

mende klassengrootte: hoe groter de klas des te lager de scores op de attitudeschalen (voor een klas met 40 leerlingen is de effectgrootte -0.08).

Voor de experimenteel opgezette onderzoeken die recent zijn uitgevoerd, zijn de resultaten minder spectaculair, maar dit effect wordt gecompenseerd doordat in het primair onderwijs de effecten sterker zijn dan in het secundair onderwijs.

Ook in het experimenteel opgezette STAR-project is gekeken naar opbrengsten in het affectieve domein. In dit opzicht traden echter geen verschillen op tussen de leerlingen uit de drie soorten klassen (klein met 13-17 leerlingen, groot met 22-25 leerlingen, en groot met 22-25 leerlingen plus een onderwijsassistent).

Bij de heranalyse van de PRIMA-gegevens is door Bosker en Hox (1996) ook gekeken naar de relatie klassengrootte en het welbevinden van de leerlingen. Voor de leerlingen van de jaargroepen 2 en 4 betreft het een oordeel van de leerkracht. De leerlingen uit de hogere jaargroepen hebben zelf een vragenlijst betreffende hun welbevinden ingevuld. Wederom is de samenhang tussen klassengrootte en het welbevinden onderzocht onder constanthouding van IQ, sekse, leerlinggewicht en de klas-samenstelling.

Voor de jaargroepen 2 en 4 blijkt er geen samenhang te bestaan tussen het welbevinden en de klassengrootte. Voor jaargroep 6 en 8 blijkt het welbevinden van de leerlingen in de grotere klassen (met 30 of meer leerlingen) boven verwachting positief te zijn. De effectgrootte bedraagt ca. 0.20.

Klassengrootte en de samenhang met zittenblijven en verwijzing naar het speciaal onderwijs

In verband met de kwestie van de kosteneffectiviteit van verkleining van de klassengrootte is recent uitgevoerd onderzoek in Nederland (Van der Werf, Reezigt & Guldmond, 1996) van belang. Dit laat zien dat juist in de kleinere klassen het zittenblijven vaker voorkomt, ook na controle voor achtergrondkenmerken van de leerlingen. De auteurs vermelden echter niet hoe groot de gevonden effecten zijn. Het tegenargument dat het hier niet om een direct effect van klassengrootte kan gaan, daar zittenblijven het gevolg is van een beslissing van de leerkracht waarbij objectieve criteria niet voor 100% de beslissing bepalen, geldt echter ook de volgende STAR-uitkomsten.

De STAR-onderzoekers betogen namelijk dat het aantal zittenblijvers in de kleine klassen geringer was (19.8 procent tegenover 27.4 in de grote klassen), zodat de kosten van de kleine klassen uiteindelijk weer gedrukt worden. De onderzoekers komen uiteindelijk tot de conclusie dat, alhoewel verkleining van de klassengrootte duurder is dan een onderwijs-assistent, het eerste kosteneffectiever is (Word, 1990, p. 29).

Het inspectie-onderzoek (Inspectie van het Onderwijs, 1995) liet zien dat verwijzingen naar het speciaal onderwijs iets frequenter voorkomen bij de scholen met de grotere klassen.

Conclusies

Welke conclusies zijn uit het hiervoor gepresenteerde overzicht aangaande klassengrootte in het basisonderwijs te trekken?

In de eerste plaats hebben we geconstateerd dat de klassen in het Nederlandse basisonderwijs, met 25.7 leerlingen gemiddeld, internationaal gezien groot zijn. Een belangrijke oorzaak lijkt gelegen te zijn in de geringe uitgaven voor het onderwijs in Nederland. Opvallend is voorts dat zeer grote klassen in enkele landen

niet in basisscholen kunnen voorkomen, omdat er wettelijk vastgestelde maxima worden gehanteerd. In Nederland zit één op de drie leerlingen nog in een klas die meer dan 30 leerlingen bevat.

Alhoewel meer recent onderzoek wenselijk zou zijn, kan op grond van Glass c.s. geconcludeerd worden, dat er sprake is van een negatieve samenhang tussen klassengrootte en het didactisch handelen van de leerkracht. Voorts is vermeldenswaard dat leerkrachten volgens het Nederlandse onderzoek niet echt lijken te 'lijden' onder grote klassen. Immers, hun arbeidssatisfactie en onderwijsattitude vertoont geen samenhang met de omvang van de klas.

Er zijn enige indicaties dat klassenverkleining of reductie van de leerling/leerkracht ratio gunstig uitpakt in termen van minder zittenblijven en minder verwijzing naar het speciaal onderwijs. Er zij echter op gewezen, dat beide verschijnselen sterk afhankelijk zijn van beslissingen van de leerkracht en niet voor 100% te herleiden zijn op het werkelijke functioneren van leerlingen.

Het overzicht van het onderzoek naar de samenhangen tussen klassengrootte en het affectief functioneren van leerlingen laat eenduidige conclusies niet toe. Directe relaties worden zelden of nooit gerapporteerd. Alhoewel de meta-analyse van Glass c.s. een verband laat zien, worden noch in het STAR-project noch in de analyse van het PRIMA-cohort samenhangen op dit gebied aangetoond.

Tenslotte vatten we de uitkomsten samen voor de relatie klassengrootte en leerprestaties. Er lijkt een trend waarneembaar dat met name voor de jongere leerlingen kleine klassen gunstig uitpakken. Maar is er hier werkelijk sprake van een oorzaak en gevolg relatie? De evidentie gevonden in het STAR-project is methodologisch gezien het sterkst en wijst duidelijk in het voordeel van kleine klassen. De tegenwerping dat juist het functioneren in een experimentele setting het effect veroorzaakt heeft en niet de klassenverkleining zelf, zou men kunnen pareren door te verwijzen naar de algemene trend die uit de meta-analyse van Glass c.s. blijkt, en naar de bevindingen uit Indiana (Prime Time) en Nederland (het PRIMA-cohort). Deze zijn immers ontleend aan gegevens die in een natuurlijke setting verzameld zijn.

De effecten die gerapporteerd zijn voor de jaargroep-2 leerlingen in Nederland zijn behoorlijk groot. Om een vergelijkingsbasis te geven: leerlingen in de grotere klassen (met 25 of meer leerlingen) hebben een achterstand die vergelijkbaar is met die van de arbeiderskinderen ten opzichte van de autochtone leerlingen uit de midden en hogere sociaal-economische milieus.

Zijn voor oudere leerlingen de resultaten minder geprononceerd, voor de jongere leerlingen lijkt een forse reductie van de klassengrootte samen te gaan met betere prestaties. Correa (1993) maakt aannemelijk dat het pas bij forse reducties van de klassengrootte voor rationeel handelende leerkrachten nuttig wordt meer tijd te besteden aan individuele hulp van leerlingen: tijd en vervolgens voldoende kwaliteit van de mogelijk geworden leerkracht-leerling interactie is dan de crux waarom het zou draaien.

Klassenverkleining in de onderbouw lijkt het meest kansrijk, maar onduidelijk is nog in welke jaargroep de klassenverkleining het meest effectief kan worden ingezet. De minder beklievende effecten van Prime Time ten opzichte van STAR lijken te wijzen op een verkleining van de klassengrootte voor de allerejongsten, maar bedacht dient te worden dat in STAR ook een veel sterkere reductie van de klassenomvang is ingezet (naar gemiddeld 15) dan in Prime Time (naar gemiddeld 19).

Desalniettemin zijn vele auteurs niet zeker over de kosteneffectiviteit van klassenverkleining. Met name Slavin (1989) spreekt de sterkste twijfels uit. Zelf propageert hij de omnibus aanpak van Success for All: een speciale lees-aanpak, 'tutors', een voorschools traject, coöperatief leren, regelmatige voortgangscontrole (eens per 8 weken), gezinsondersteuning en deskundige hulp voor het schoolteam. Het probleem met zijn alternatieve aanpak is echter, dat een onafhankelijke evaluatie van het Success for All programma nog niet heeft plaatsgevonden. De door de auteur zelf gerapporteerde resultaten lijken gunstig (Madden, Slavin, Karweit, Dolan & Wasik, 1993; Slavin, 1996), maar zijn wellicht sterk opgeblazen door mogelijke Hawthorne-effecten. Slavin (1996) spreekt zelf de veronderstelling uit dat voor risicoleerlingen op scholen met veel achterstandsleerlingen de omnibus aanpak van Suc-

cess for All het meest effectief is, terwijl voor risicoleerlingen op andere scholen enkel en alleen 'Reading Recovery' voldoende is. De benadering van Slavin heeft met klassenverkleining gemeen, dat een reductie van de leerling/leerkracht ratio noodzakelijk is. Kennelijk is dat toch de algemene cruciale factor.

Bosker en Hox (1996) laten zien dat scholen hun formatie, bedoeld voor achterstandsbestrijding, voor een belangrijk deel inzetten voor een algehele reductie van de klassenomvang in alle jaargroepen. Deze bevinding, gecombineerd met de magere resultaten van reductie van de klassengrootte voor de prestaties van leerlingen van 8 jaar en ouder, zou een mogelijke verklaring kunnen bieden voor de tot nu toe weinig bevredigende resultaten van het OnderwijsvoorrrangsBeleid (zie: Mulder, 1996). Het roept ook de vraag op of er niet toch scholen zijn die gericht hun formatie gebruiken, hetzij voor een forse klassenverkleining, hetzij voor een Slavin-achtige aanpak, voor de jongere leerlingen, en wat daar dan het effect van is.

Gericht – bij voorkeur experimenteel – onderzoek naar de effectiviteit van deze aanpakken op reguliere scholen en op scholen met veel achterstandsleerlingen zou hier uitsluitsel kunnen geven.

Voorts is nader onderzoek gewenst naar combinatieklassen in relatie tot de klassengrootte problematiek. De 'best-evidence' reviews van Veenman (1995, 1996) laten zien dat combinatieklassen niet ongunstig uitpakken, noch voor het cognitief noch voor het affectief functioneren van leerlingen. En onder verwijzing naar vernieuwende scholingsconcepten wordt regelmatig het einde gevraagd van het leerstofjaarklassensysteem. Maar zowel groot-schalig goed opgezet onderzoek naar combinatieklassen in Nederland (gebruik makend van adequate statistische analysemethoden) als goed opgezet evaluatie-onderzoek naar de effecten van vernieuwingsscholen als Dalton, Jenaplan, Montessori e.d. ontbreekt. Klassenverkleining lokt leeftijdsheterogene combinatieklassen of andere groeperingsvormen uit. Maar wat voor de leerkracht haalbaar en voor de leerling wenselijk is, blijft voor alsnog een open vraag. De discussie over klassengrootte in het basisonderwijs mag dan zijn voorlopig einde bijna lijken te hebben bereikt (Commissie Kwalitatieve Aspecten van Groeps-grootte,

1996), de discussie over en het onderzoek naar een pedagogisch-didactisch verstandig, efficiënt, effectief, en gelijke kansen bevorderend gebruik van formatie in basisscholen kan ons inziens nu beginnen.

Noten

1. Met de introductie van de Wet op het Basisonderwijs, is het woord klas vervangen door groep, zodat het usance is geworden om waar vroeger werd gesproken over klassengrootte (bijv. Van Gelder, 1959; Warries, 1985) nu te spreken over groeps-grootte. Deze term is echter misleidend, omdat één (combinatie)klas meerdere jaargroepen kan omvatten. Derhalve houden we vast aan de term klassengrootte.
2. De effectgrootte is in navolging van Cohen (1988), uitgedrukt als het effect van een conditie ten opzichte van een referentiegroep in termen van de residuele (dus niet-triviale) binnen-groepen variantie. Effectgroottes van ca. 0.10 worden door Cohen klein genoemd, maar kunnen desalniettemin relevant zijn (zie ook: Bosker & Witziers, 1996).
3. Citaat ontleend aan: Aitkin en Longford (1986, p. 42).

Literatuur

Achilles, C.M., Nye, B.A., Zaharias, J.B., & Fulton, B.D. (1993). Creating successful schools for all children: a proven step. *Journal of School Leadership*, 3, 606-621.

Aitkin, M.A., & Longford, N.L. (1986). Statistical modelling issues in school effectiveness studies. *Journal of the Royal Statistical Society, A*, 149, 1-43.

Akerhielm, K. (1995). Does class size matter? *Economics of Education Review*, 14(3), 229-241.

Batenburg, T. van, Koster, D., & Ros, A. (1989). Kleinere groepen, betere prestaties? In J. Scheerens & J.C. Verhoeven (Red.), *Schoolorganisatie, beleid en onderwijskwaliteit* (pp. 55-62). Bijdragen aan de onderwijsresearch. Lisse: Swets & Zeitlinger.

Bosker, R.J., & Hox, J. (1996). Klassengrootte en functioneren van leerlingen en leerkrachten; een onderzoek op basis van de PRIMA-cohort gegevens. In Commissie Kwalitatieve Aspecten

van Groeps-grootte, *Klassenverkleining* (pp. 107-118). Den Haag: SDU.

Bosker, R.J., & Sniijders, T.A.B. (1990). Statistische aspecten van multi-niveau onderzoek. *Tijdschrift voor Onderwijsresearch*, 15, 317-329.

Bosker, R.J., & Witziers, B. (1996) *The magnitude of school effects*. New York: AERA-paper.

Cohen, J. (1988) *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Lawrence Erlbaum.

Coleman, J.S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). *Equality of educational opportunity*. Washington, DC: US Government Printing Office.

Commissie Evaluatie Basisonderwijs (1994). *Zicht op kwaliteit. Evaluatie van het basisonderwijs*. Den Haag: SDU.

Commissie Kwalitatieve Aspecten van Groeps-grootte (1996). *Klassenverkleining*. Den Haag: SDU.

Correa, H. (1993). An economic analysis of class size and achievement in education. *Education Economics*, 1(2), 129-135.

Finn, J.D., & Achilles, C.M. (1990). Answers and questions about class size: A statewide experiment. *American Educational Research Journal*, 27(3), 557-577.

Gelder, L. van (1959). De klasgrootte bij het lager onderwijs. *Pedagogische Studiën*, 36, 498-519.

Glass, G.V., & Smith, M.L. (1978). *Meta-analysis of research on the relationship of class size and achievement*. San Francisco: Far West Laboratory.

Glass, G.V., Cahen, L.S., Smith, M.L., & Filby, N.N. (1982). *School class size*. London: Sage Publications.

Glopper, K. de, & Otter, M.E. (1993). *Nederlandse leesprestaties in internationaal perspectief*. Amsterdam: SCO.

Hanushek, E.A. (1986). The economics of schooling: production and efficiency in public schools. *Journal of Economic Literature*, 24, 1141-1177.

Hedges, L.V., Laine, R.D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher*, 23(3), 5-14.

Inspectie van het Onderwijs (1995). *Groeps-grootte in het basisonderwijs*. Den Haag: SDU.

Jungbluth, P., Peetsma, T., & Roeleveld, J. (1996). Leerlingprestaties en leerlinggedrag in het primair onderwijs. Ubbergen: ITS/SCO-Kohnstamm/Tandem Felix.

Larkin, A.I., & Keeves, J.P. (1984). *The class size*

- question: A study at different levels of analysis. Hawthorn: Australian Council for Educational Research.
- Madden, N.A., Slavin, R.E., Karweit, N.L., Dolan, L., & Wasik, B.A. (1993). Succes for all: Longitudinal effects of a restructuring program for inner city elementary schools. *American Educational Research Journal*, 30(2), 123-148.
- McGiverin, J., Gilman, D., & Tilitzki, C. (1989). A meta-analysis of the relation between class size and achievement. *The Elementary School Journal*, 90(1), 47-56.
- Ministerie van OC&W (1996). *Kerngegevens primair onderwijs*. Zoetermeer: MOCW.
- Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children*, 5(2), 1-13.
- Mulder, L. (1996). *Meer voorrang, minder achterstand? Het onderwijsvoorrangsbeleid getoetst*. Academisch proefschrift, Katholieke Universiteit Nijmegen.
- Organization for Economic Co-operation and Development (1995). *Education at a Glance*. Parijs: OECD.
- Overmaat, M., & Ledoux, G. (1996). *School- en klaskenmerken basisonderwijs en speciaal onderwijs*. Ubbergen: SCO-Kohnstamm/Tandem Felix.
- RISE (1994). *Class size regulation. A dossier of international comparisons*. London: The Research and Information on State Education Trust.
- Ross, K.N., & Postlethwaite, T.N. (1994). Differences among countries in school resources and achievement. In W.B. Elley (Ed.), *The IEA study of reading literacy: Achievement and instruction in thirty-two school systems*. Oxford: Elsevier Science Ltd.
- Slavin, R.E. (1989). Achievement effects of substantial reductions in class size. In R.E. Slavin (Ed.), *School and classroom organization* (pp. 247-257). Hillsdale, N.J.: Lawrence Erlbaum.
- Slavin, R.E. (1996). *Education for All*. Lisse: Swets & Zeitlinger.
- Smith, M.L., & Glass, G.V. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Educational Research Journal*, 17(4), 419-433.
- Veenman, S. (1995). Cognitive and noncognitive effects of multigrade and multi-age classes: A best-evidence synthesis. *Review of Educational Research*, 65(4), 319-382.
- Veenman, S. (1996). Effects of multigrade and multi-age classes reconsidered. *Review of Educational Research*, 66(3), 323-340.
- Vierke, H. (1995). *De PRIMA-toetsen gecalibreerd*. Nijmegen: ITS.
- Warries, E. (1985). *Klassegrootte en onderwijskwaliteit*. Enschede: Universiteit Twente.
- Werf, M.P.C. van der, & Guldemon, H. (1994). *De tweede meting van de OVB evaluatie. School- en klaskenmerken in schooljaar 1990/1991*. Groningen: RION/ITS/LEO.
- Werf, M.P.C. van der, Reezigt, G.J., & Guldemon, H. (1996). *Zittenblijven in het basisonderwijs: omvang, effecten en relaties met leerling- en schoolkenmerken*. Groningen: GION.
- Word, E. (1990). *Student/Teacher Achievement Ratio (STAR). Tennessee's K-3 Class Size Study*. Washington: ERIC Report ED320 692.

Auteur

R.J. Bosker is als universitair hoofddocent werkzaam bij de vakgroep Onderwijsorganisatie & -management van de faculteit der Toegepaste Onderwijskunde van de Universiteit Twente.

Correspondentie-adres: Vakgroep O&M, Faculteit TO, Universiteit Twente, Postbus 217, 7500 AE Enschede.

Abstract

The end of class struggle? Background to the discussion on class size in primary education.

R.J. Bosker. *Pedagogische Studiën*, 1997, 74, 210-227.

Recent surveys show that class size in Dutch elementary schools is on average large as compared to other countries, and moreover varies enormously between as well as within schools. This article presents a review of foreign studies into the effects of class size on the functioning of teachers and students. Moreover, it also presents results from a recent large scale Dutch research study in this area. This study shows that there is an association between class size and student achievement in the areas of language and arithmetic. Grade-2 students (five year olds) in classes with more than twenty five students are more likely to be

underachievers than comparable students from smaller classes. For grades 4, 6, and 8 the results are less clear cut (showing, for example, a relatively high level of overachievement for arithmetic in classes sized 30-34), but there is an indication that class sizes of thirty five and up may be detrimental to student achievement.

Neither in the affective domain, nor with respect to the functioning of teachers this Dutch study could show a relationship with class size.

Comparing the results with programs like 'Reading Recovery' and 'Succes for All' it is hypothesized that not class size per se is the factor causing the effects mentioned, but a more general factor, namely the teacher-to-pupil ratio.