

# Discussiebijdrage

## Schoolverschillen en schooleffecten in het voortgezet onderwijs

K. Lek, R.C.W Feskens en J. Scheerens

### Samenvatting

Op basis van PISA-onderzoek (2015 en 2018) concludeerden o.a. de Inspectie van het Onderwijs dat verschillen tussen Nederlandse scholen in het voortgezet onderwijs “zorgwekkend groot” zijn. Deze conclusie is gebaseerd op de zogenaamde ‘intra-klasse correlatie coëfficiënt’. Echter, het in kaart brengen van verschillen in kwaliteit tussen scholen is veel complexer en genuanceerder dan de intra-klasse correlatie vergelijking op basis van PISA data doet vermoeden. In deze discussiebijdrage gaan we in op de beperkingen van de intra-klasse correlatie en bespreken we enkele alternatieve indicatoren. Deze alternatieve indicatoren zijn bedoeld om in kaart te brengen in welke mate verschillen zijn toe te schrijven aan door de school beïnvloedbare factoren, dan wel door gegeven achtergrond- en contextkenmerken. Voor elk van deze indicatoren wordt de intra-klasse correlatie – die “bruto”, all-in schoolverschillen uitdrukt – verder ‘ontrafeld’ om zo “netto” schooleffecten in beeld te kunnen brengen. De discussiebijdrage bevat illustraties van elk van de alternatieven in de context van internationaal vergelijkend assessment onderzoek. Daarnaast wordt het nut van de intra-klasse correlatie en elk van de alternatieven besproken in het licht van beleids- en praktijkgerichte beslissingen.

**Kernwoorden:** schooleffectiviteit, verschillen in kwaliteit tussen scholen, PISA

### 1 Aanleiding en doelstelling van deze discussiebijdrage

Verschillen tussen scholen in het voortgezet onderwijs (VO) zijn zorgwekkend groot. Dat concludeerde de Inspectie van het Onderwijs in 2017 (Inspectie van het Onderwijs, 2017 p.

6, p. 33), op basis van het PISA-onderzoek, uitgevoerd in 2015 (OECD, 2017a). Het PISA onderzoek is een grootschalig internationaal vergelijkend onderzoek naar de wiskunde-, lees- en natuurkundevaardigheden van 15-jarige leerlingen. Van alle landen die meededen aan dit onderzoek had Nederland namelijk de hoogste intra-klasse correlatie op schoolniveau in 2015. In de recentere PISA vergelijking van 2018 is deze intra-klasse correlatie onveranderd hoog (zie Figuur 1)

Anno 2020 haalt ook McKinsey, in een onderzoek naar de doelmatigheid en de toereikendheid van het funderend onderwijs, dezelfde categorie PISA gegevens uit 2015 en 2018 aan, om te concluderen dat Nederland het grootste verschil tussen scholen VO in leerresultaten heeft van alle OECD-landen (McKinsey, 2020, p. 4, p. 8). In de presentatie van deze gegevens wordt, naar onze mening, onvoldoende gewezen op de discrepantie tussen de ICC- waarden en “echte” kwaliteitsverschillen tussen scholen. Deze bijdrage is bedoeld als een poging om deze discrepantie nader te verduidelijken. Niet in de eerste plaats als kritiek op de presentatie in beide onderzoeken, maar in ruimere zin om bij te dragen aan een beter begrip van de betekenis van schoolverschillen in relatie tot onderwijskwaliteit.

In schooleffectiviteitsonderzoek worden de verschillen tussen scholen als aangeduid door de intra-klasse correlatie ook wel “bruto” schoolverschillen genoemd. Het gevaar bestaat dat deze “bruto” all-in schoolverschillen onterecht geïnterpreteerd worden als kwaliteitsverschillen tussen scholen. In deze discussiebijdrage gaan we in op de beperkingen van de intra-klasse correlatie en bespreken we enkele alternatieve indicatoren. Deze alternatieve indicatoren zijn bedoeld om nauwkeurig in kaart te brengen in welke mate verschillen zijn toe te schrijven aan kwaliteitsverschillen tussen



In de discussiesectie gaan we verder in op de praktische betekenis van deze indicatoren voor beslissingen, beoordelingen en keuzeprocessen van actoren in het onderwijsveld (overheid, Onderwijsinspectie, bevoegd gezag van scholen, ouders van schoolgaande kinderen). Verder geven wij in deze discussiebijdrage enkele getallen-voorbeelden op basis van data verzameld in het PISA survey 2015 en 2018. In de voorbereiding en uitvoering van dit survey wordt veel aandacht besteed aan de operationalisatie en analyse van zowel de cognitieve opbrengstmetingen, als ook aan de achtergrond indicatoren, die betrekking hebben op leerlingen en scholen. Voor informatie over de operationalisatie van de verschillende metingen verwijzen wij kortheidshalve naar OECD (2013), OECD (2017b) en OECD (2019). Uitgebreide informatie rondom de ondersteunende psychometrische analyses en kwaliteit van de operationele variabelen kan teruggevonden worden in de technische rapporten van PISA (OECD, 2017a) en in de appendix van dit artikel. Merk op dat de getallenvoorbeelden in deze discussiebijdrage louter bedoeld zijn ter illustratie en niet als een substantiële aanvullende secundaire analyse van PISA bestanden. We beperken ons in deze illustraties dus tot het VO, binnen de context van internationaal vergelijkend assessment onderzoek.

## 2 Beperkingen van de intra-klasse coëfficiënt als indicator van schoolverschillen<sup>1</sup>

De intra-klasse correlatie drukt uit in hoeverre de resultaten van leerlingen binnen een school dichterbij elkaar liggen dan resultaten van leerlingen die niet bij elkaar op school zitten<sup>2</sup>. Grofweg betekent een intra-klasse correlatie die neigt naar 0 dat resultaten van leerlingen niet structureel samenhangen met schoolkeuze en/ of schooleffectiviteit. Bij een intra-klasse correlatie die neigt naar 1 hangen leerlingresultaten volledig samen met schoolkeuze en/ of schooleffectiviteit. Leerlingen op de ene school scoren dan structureel hoger dan

leerlingen op een andere school. In de intra-klasse coëfficiënt zijn schoolkeuze effecten en kwaliteitsverschillen vervlochten (confounded); verderop bespreken we indicatoren en technieken, waarbij deze uit elkaar kunnen worden gehaald.

De intra-klasse correlatiecoëfficiënt (ICC) kan in formulevorm uitgedrukt worden als:

$$\frac{\sigma^2_{\text{school}}}{\sigma^2_{\text{school}} + \sigma^2_{\text{leerling}}} \quad (1)$$

waar  $\sigma^2_{\text{leerling}}$  de variantie is in leerlingresultaten en  $\sigma^2_{\text{school}}$  de variantie in gemiddelde leerlingresultaten over scholen (zie Koch, 1982). Heeft de ene school relatief veel leerlingen met hoge resultaten en de andere school relatief veel leerlingen met lage resultaten, dan neemt  $\sigma^2_{\text{leerling}}$  relatief toe en gaat de intra-klasse correlatie omhoog. De intra-klasse correlatie kan daardoor gebruikt worden om de vraag te beantwoorden of er verschillen zijn tussen scholen op het gebied van leerlingresultaten. Er zijn echter drie belangrijke beperkingen aan de toepassing van de intra-klasse coëfficiënt als methode om valide schooleffectiviteit in kaart te brengen, omdat deze: 1) ons niet vertelt waar verschillen vandaan komen, 2) niet zonder meer rekening houdt met initiële niveauverschillen en 3) in internationale vergelijkingen niet corrigeert voor schoolstelselverschillen. Deze beperkingen worden hieronder kort toegelicht.

### 2.1 De intra-klasse coëfficiënt vertelt niet waar verschillen vandaan komen

Wat de intra-klasse correlatie ons nadrukkelijk niet vertelt, is waar verschillen tussen scholen op het gebied van leerlingresultaten vandaan komen. Een hoge intra-klasse correlatie betekent niet dat schoolverschillen alleen veroorzaakt worden door verschillen in kwaliteit van het onderwijsbeleid of (on)bekwaamheid van docenten. Verschillen tussen scholen kunnen immers al ontstaan zijn voordat leerlingen op een bepaalde school terecht zijn gekomen. Deze initiële verschillen in schoolpopulatie worden niet meegewogen in berekening van de intra-klasse correlatie van PISA 2015/2018. Daarnaast zijn er veel meer aspecten op

systeem-, school-, docent- en leerling niveau die allemaal de intra-klasse correlatiecoëfficiënt kunnen beïnvloeden, waarvan slechts een gedeelte van deze factoren onder de invloed staat van scholen.

## **2.2 De intra-klasse coëfficiënt houdt niet zonder meer rekening met initiële niveauverschillen**

Over het algemeen is het zo dat leerlingen bij de overgang van primair (PO) naar VO niet dezelfde vaardigheden en capaciteiten hebben. Als leerlingen toevallig naar verschillende scholen voor VO zouden worden verwezen, dan zou dit initiële niveauverschil voor de interpretatie van de intra-klasse correlatie niet uitmaken. Immers, elke school start in dat geval met een even gevarieerde leerling populatie waarvan we verwachten dat zij gemiddeld gezien even veel ontwikkeling laten zien als op elke andere willekeurige school. De toewijzing van leerlingen naar scholen in het VO vindt echter allesbehalve toevallig plaats. Het gevolg is dat de leerling populatie bij aanvang op elke VO-school verschilt. Wanneer er niet wordt gecorrigeerd voor deze initiële niveauverschillen, wordt de intra-klasse correlatie een oninterpreteerbare mengelmoes van echte en initiële verschillen in schoolresultaten.

## **2.3 De intra-klasse coëfficiënt corrigeert niet voor schoolsysteemverschillen**

De mate waarin de leerling populatie verschilt van school tot school wordt gedeeltelijk bepaald door landelijk onderwijsbeleid. Zo zijn er, in het kader van internationale vergelijkingen, landen waar de toewijzing van leerlingen naar VO-scholen niet of nauwelijks afhangt van eerdere leeruitkomsten. Maar zijn er ook landen zoals Nederland en - in mindere mate - Duitsland en België waar leerlingen afhankelijk van hun prestaties op de basisschool verschillende leerwegen volgen in het VO (Pit-Ten Cate, Krolak-Schwerdt & Glock, 2016; Van der Werfhorst & Mijs, 2010). Hoewel deze schoolsysteemverschillen afzonderlijk niet alle verschillen tussen de intra-klasse correlaties kunnen verklaren is er wel een

trend zichtbaar: de (extreem) lage intra-klasse correlaties komen enkel voor in landen met slechts één leerweg en selectie op relatief hoge leeftijd. Andersom is het zo dat (extreem) hoge intra-klasse correlaties zijn voorbehouden aan landen met relatief veel leerwegen en een relatief lage selectieleeftijd. Volgens een rapportage van de OECD (2016) kan 62% van de verschillen in intra-klasse correlaties tussen landen verklaard worden door kenmerken van het schoolsysteem zoals de leeftijd waarop leerlingen binnen verschillende leerwegen worden geselecteerd en het aantal leerwegen dat wordt aangeboden. Wanneer de intra-klasse correlatie niet wordt gecorrigeerd voor dit evidente schoolsysteemverschil hebben we opnieuw te maken met een oninterpreteerbare mengelmoes van échte kwaliteitsverschillen tussen scholen en verschillen die simpelweg veroorzaakt zijn door het schoolsysteem.

## **3 Alternatieve indicatoren voor schoolverschillen**

Zoals aangegeven in de inleidende paragraaf, kunnen we de verschillen tussen scholen als uitgedrukt door de intra-klasse correlatie zien als een soort “all-in” bruto schoolverschillen. Er zijn verschillende manieren waarop we deze “bruto” schoolverschillen verder kunnen ontrafelen, om zo tegemoet te komen aan de eerder besproken beperkingen van de intra-klasse correlatie. Zo kunnen we de intra-klasse coëfficiënt bepalen per schoolcategorie, om te corrigeren voor schoolsysteemverschillen. Of kunnen we de bruto schoolverschillen corrigeren voor initiële verschillen tussen leerlingen, zoals begin-niveau. Dit wordt in schooleffectiviteitsonderzoek ook wel aangeduid als “netto-1 model”. Dit “netto-1” model zou verder ontrafeld kunnen worden door deze te corrigeren voor door de school beïnvloedbare factoren. Dit wordt ook wel het “netto-2 model” genoemd. Deze verschillende opties worden hieronder verder toegelicht en geïllustreerd aan de hand van PISA data.

Tabel 1  
De ICC's in 2015 en 2018 voor de verschillende Nederlandse onderwijsniveaus op het gebied van wiskunde

Niveau	ICC in 2015	ICC in 2018
pro	0.47	0.03
vmbo 1-2	0.40	0.09
vmbo bb	0.04	0.14
vmbo kb	0.20	0.13
vmbo gt	0.15	0.11
havo 1-3	0.05	0.12
havo 4-5	0.05	0.11
vwo 1-3	0.12	0.08
vwo 4-6	0.10	0.11

### 3.1 De intra-klasse coëfficiënt bepaald per schoolcategorie

Zoals besproken in de vorige sectie wordt de intra-klasse correlatie 'opgeblazen' in landen met een sterke stratificatie, zoals in Nederland (Scheerens, 2019). Dit wordt zichtbaar op het moment dat we de intra-klasse correlatie berekenen voor elk van de zeven leerwegen afzonderlijk; zie Tabel 1 voor deelgebied wiskunde. Te zien is hoe in 2018 de meerderheid van de intra-klasse correlaties voor de verschillende leerwegen redelijk in de buurt lag van die van Noorwegen (.10), Finland (.08) en IJsland (.08). Geen enkele intra-klasse correlatie kwam boven de .15 uit.

Stratificatie naar leerweg heeft dus een direct effect op de hoogte van de intra-klasse correlatie. Immers, hoe meer gestratificeerd een onderwijssysteem is, des te homogener zijn de schoolpopulaties binnen een school. Het is daarom minder informatief om landen te vergelijken op de intra-klasse correlatie wanneer de onderwijssystemen in deze landen verschillend zijn ingericht.

### 3.2 Het netto-1 model: intra-klasse coëfficiënt gecorrigeerd voor begin-niveau en andere achtergrondkenmerken

In de praktijk van eindtoetsing in het onderwijs worden opbrengstindicatoren in het kader van school evaluaties (accountability) gecorrigeerd voor achtergrondkenmerken van leerlingen, bij voorkeur eerdere leerprestaties, om een "eerlijker" beeld te krijgen van de verdienste van een school. Deze praktijk wordt aangeduid als het bepalen van de toegevoegde waarde van het gegeven onderwijs voor de ontwikkeling van de leerlingen<sup>3</sup>. Een dergelijke correctie wordt ook gebruikt in het kader van onderwijs-effectiviteitsonderzoek.

In schooleffectiviteitsonderzoek worden de verschillen tussen scholen zoals uitgedrukt door de intra-klasse correlatie, gezien als een soort 'all-in', bruto schooleffecten. Dit bruto schooleffect moet allereerst gecorrigeerd worden voor allerlei achtergrondkenmerken van leerlingen. Te denken valt aan initiële

Tabel 2  
Percentages verklaarde variantie voor de verschillende modellen Nederland in PISA 2015 en 2018

Afname	Onderdeel	Bruto-effect*	SES en/ of school-structuur**	Netto-1**	Factoren door school bepaald**	Netto-2**
2015	Wiskunde	54.17%	73.30%	14.46%***	6.42%	13.53%****
2018	Wiskunde	55.02%	74.98%	13.77%	2.37%	13.44%
2015	Natuurwetenschappen	53.96%	75.29%	13.33%	23.91%	10.14%
2018	Natuurwetenschappen	54.35%	72.90%	14.73%	-0.50%	14.80%
2015	Leesvaardigheid	54.61%	70.84%	15.92%	8.63%	14.55%
2018	Leesvaardigheid	52.22%	69.86%	15.74%	8.67%	14.38%

\* percentages totale leerling variantie

\*\* gepercenteerd wordt over de tussenschoolse variantie

\*\*\*  $54.17\% - (73.30\% * 54.17\%)$

\*\*\*\*  $14.46\% - (6.42\% * 14.46\%)$

verschillen in prestaties, maar ook bijvoorbeeld aan sociaaleconomische status, algemene leervaardigheid en intelligentie en het culturele kapitaal van de thuissituatie. Op deze manier wordt rekening gehouden met verschillen in de initiële leerling populatie. We kunnen dit ook aanduiden met de term netto schooleffecten (Scheerens & Bosker, 1997, hfst 3). Zoals hieronder verder wordt toegelicht duiden we deze schooleffecten aan als “netto-1 schooleffecten, ter onderscheiden van schooleffecten die worden uitgelicht wanneer vervolgens ook factoren worden meegewogen waar de school invloed op heeft. Deze laatste duiden we aan met de term netto-2 schooleffecten (Scheerens, 2019, p. 49).

#### *Illustratie*

Specifiek kunnen de ‘netto-1’ (en later de ‘netto-2’) schooleffecten worden bepaald op basis van een verdere decompositie van de totale variantie op leerling niveau met behulp van multi-level analyses<sup>4</sup>. Na schatting van de ICC in het ‘lege’ model, wordt een modelvariant getoetst die de controle variabelen bevat. Een voorbeeld aan de hand van de PISA 2015 en 2018 data is te zien in Tabel 2<sup>5</sup>.

De derde kolom toont de intra-klasse correlatie \* 100%; het bruto-effect. Oftewel: het percentage van alle variantie die zich manifesteert tussen scholen. Opvallend is dat voor zowel 2015 als 2018 en alle drie de vakgebieden het percentage zo rond de 50-55% schommelt. De volgende kolom toont hoeveel procent van het bruto school-effect verklaard kan worden door contextuele factoren. In dit voorbeeld kon gecorrigeerd worden voor sociaaleconomische status op leerling- en schoolniveau en het gevolgde onderwijsniveau. Net als voor het bruto-effect geldt dat de waarden in 2015 en 2018 dichtbij elkaar liggen; zo’n 70-75% van het bruto-effect kan verklaard worden vanuit de eerdergenoemde contextuele invloeden (SES en schooltype). Het netto-1-effect is gedefinieerd als het bruto-effect min dat deel van het bruto-effect dat door contextuele factoren verklaard kan worden. Voor dit netto-1-effect geldt dat dit voor alle vakken grofweg tussen de 14- en 16% schommelt.

### **3.3 Het netto-2 model: ook correctie voor factoren die beïnvloedbaar zijn door de school**

In het kader van onderwijs-effectiviteitsonderzoek kan het bepalen van de toegevoegde waarde van scholen door te corrigeren voor achtergrondkenmerken van leerlingen en andere context-variabelen worden gezien als voorbereiding op het schatten van de effecten van factoren die wel door de school beïnvloed kunnen worden. Eventueel kan men dit als opeenvolgende stappen zien: eerst wordt het netto-schooleffect berekend en vervolgens wordt nagegaan hoe dit netto-effect verklaard wordt door beheersbare factoren. Het gaat om een voortgezette decompositie van de variantie. Eventueel kunnen de door de school te beïnvloeden variabelen ook worden aangegeven als “treatment” variabelen. Op elk niveau van de decompositie van variantie is er sprake van onverklaarde variantie. Op het niveau van de treatment variabelen is dat het deel van de netto-1 schooleffecten dat niet verklaard wordt door de variabelen die in het treatment model zijn opgenomen. Eventueel is deze variantie-component aan te duiden als het netto-2 schooleffect, waarin het bruto schooleffect niet alleen gecorrigeerd wordt voor context- en achtergrondeffecten, maar ook voor de invloed van in het model opgenomen door de school manipuleerbare variabelen. Het is overigens de vraag of dit een behulpzame term is, omdat de primaire interesse uitgaat naar het percentage variantie dat gebonden wordt door beïnvloedbare factoren, en niet zozeer naar een soort “zuiver” schooleffect, dat overblijft als men voor alle bekende factoren heeft gecorrigeerd. In feite is dat restant het “onverklaarde” schooleffect, de onzekerheid die overblijft als men zich, bijvoorbeeld bij schoolkeuze, heeft laten leiden door alles wat bekend is over de schoolpopulatie en een bepaalde set door de school te manipuleren, effectiviteitsbevorderende kenmerken.

#### *Illustratie*

Net als voor het netto-1 effect kan het netto-2 schooleffect worden bepaald met behulp van multi-level analyses. Na schatting van het

netto-1 model, wordt een modelvariant getoetst met door de school te beïnvloeden factoren. Een voorbeeld van een dergelijke netto-1 naar netto-2 analyse aan de hand van de PISA 2015 en 2018 data is ook te vinden in Tabel 2. De zesde kolom toont het percentage van de nog overgebleven variantie dat verklaard kan worden door factoren waar de school invloed op heeft. In deze illustratie wordt gebruik gemaakt van enkele organisatorische schoolkenmerken die PISA beschikbaar stelt. Er kan hier echter ook bijvoorbeeld gebruik worden gemaakt van variabelen die zich richten op pedagogische en/of didactische schoolkenmerken. Opvallend is dat de verklaarde variantie door de organisatorische schoolkenmerken in deze illustratie vele malen kleiner is dan het effect van contextfactoren en dat deze ook sterker fluctueert van 2015 naar 2018 en van vakgebied naar vakgebied. Het netto-2-effect wordt ten slotte gegeven in Tabel 2; het netto-1-effect min dat deel van het netto-1-effect dat door organisatorische schoolkenmerken verklaard kan worden. Voor dit netto-2-effect geldt dat dit voor alle vakken grofweg tussen de 10- en 14% schommelt. Wat opvalt is dat de door de school te beïnvloeden variabelen de netto-1 variantie maar weinig reduceren. Dit betekent dat kwaliteitsverschillen tussen scholen, op deze manier uitgedrukt, zeer klein zijn. Een nadere toelichting op het netto-2 effect wordt gegeven in de appendix.

## 4 Discussie

In deze discussie sectie gaan we in op de twee onderzoeksvragen uit de inleidende paragraaf.

### *Beperkingen van de intra-klasse correlatie coëfficiënt*

De afgelopen 30 jaar is er veel onderzoek gedaan dat relevant is voor de vraag hoe kwaliteitsverschillen tussen scholen het beste in kaart gebracht kunnen worden. Ondanks deze “state of the art” wordt in de praktijk nog vaak gebruik gemaakt van de intra-klasse correlatie coëfficiënt, welke louter een soort “all-in” bruto schoolverschillen reflecteert.

Het gevaar bestaat dat deze bruto schoolverschillen te absoluut worden geïnterpreteerd als kwaliteitsverschillen, waardoor verwarrende of zelfs misleidende signalen ontstaan. In deze discussiebijdrage is daarom stilgestaan bij enkele beperkingen van de intra-klasse coëfficiënt en alternatieven uit schooleffectiviteitsonderzoek.

### *Decompositie van bruto schoolverschillen*

Twee besproken alternatieven zijn het zogenaamde netto-1 en netto-2 model, waarin de bruto schoolverschillen achtereenvolgens gecorrigeerd worden voor initiële verschillen tussen leerlingen en door de school beïnvloedbare factoren. Ondanks de veelbelovendheid van deze alternatieven is het goed te benadrukken dat zowel bij het bepalen van de netto-1 als de netto-2 schooleffecten is er sprake van belangrijke uitdagingen bij de modelspecificatie. Op dit moment bestaat er in de literatuur geen consensus over het aantal variabelen en de specifieke variabelen die geselecteerd zouden moeten worden om in optie 1 de netto-2- schooleffecten succesvol te bepalen (Manzi, San Martin, & Van Belleghem, 2010). Wel is er een groeiende consensus over de voorkeur voor het in aanmerking nemen van eerdere prestaties (Marks, 2020, Timmermans en Van der Werf, 2018). Ook het belang van het opnemen van intelligentie in de set controle variabelen is een punt van discussie (vergelijk Detterman, 2016). Een vergelijking tussen landen – zoals gedaan door PISA - maakt het nog ingewikkelder. Dit kan namelijk alleen eerlijk gebeuren wanneer er een set variabelen is die even succesvol initiële verschillen tussen scholen corrigeert in het ene als in het andere land. In een land als het Verenigd Koninkrijk kan het bijvoorbeeld zinvol en ‘eerlijk’ zijn mee te wegen of een leerling populatie relatief vaak of weinig gebruikmaakt van een gratis schoolmaaltijd (Goldstein, 2001), terwijl deze informatie niets toevoegt voor landen waarin een dergelijke optie niet aanwezig is. Daarnaast moet deze set van achtergrondkenmerken een gelijke betekenis hebben in de verschillende landen (dit wordt aangeduid als meetinvariantie en wordt hier niet verder besproken. Zie echter, Davidov et al., 2014; Vandenberg &



Lance, 2000). Ook de selectie en operationalisatie van door de school en de leerkracht te beïnvloeden variabelen is een bijzondere opgave. Hier gaat het om niet minder dan de soliditeit van de kennisbasis van onderwijs-effectiviteit. Ondanks een zekere consensus over factoren die “werken”, laten meta-analyses grote diversiteit zien in effectgroottes en zijn er weinig gestandaardiseerde instrumenten om “proces kwaliteit” te meten. In het kader van deze discussiebijdrage kunnen wij niet dieper op deze “state of the art” ingaan (zie Scheerens, 2016, 2017). Wel memoreren we enkele onderzoeken die zich uitspreken over de grootte van “treatment effecten” (Aloisi & Tymms, 2017; Borman et al., 2003, Detterman, 2016; Lortie-Forgues & Inglis, 2019).

*De praktische betekenis van de intra-klasse coëfficiënt en alternatieve indicatoren voor beslissingen en keuzeprocessen in het onderwijsveld*

De verschillende typen schooleffecten die wij hebben besproken hebben elk andere praktische implicaties. Hieronder worden deze implicaties besproken samen met toepassingssituaties voor verschillende actoren in het onderwijs.

*Het bruto effect (intra-klasse correlatie \* 100%)*

Weergave van de ongecorrigeerde bruto verschillen tussen scholen in het VO drukt ons met de neus op de consequenties van het sterk gestratificeerde scholenstelsel, dat wil zeggen de kenmerken van ons gedifferentieerde vo-stelsel (vmbo, havo, vwo). Op de nadelen hiervan is herhaaldelijk gewezen, maar de politieke wil heeft tot nu toe ontbroken om hier echt iets aan te doen, hoewel de recente Nota Brede Maatschappelijke Heroverweging (2020) opnieuw op dit probleem wijst. Aan het gestratificeerde scholenstelsel zijn natuurlijk ook belangrijke keuzeprocessen bij ouders en scholen verbonden.

*De intra-klasse correlatie bepaald per schoolcategorie*

De schoolkeuze is, zoals hierboven

aangegeven, in Nederland in de eerste plaats een keuze tussen schoolcategorieën. Maar, zeker in stedelijke gebieden, kan er binnen schoolcategorieën gekozen worden tussen verschillende scholen. De gemiddelde ongecorrigeerde bruto schoolverschillen per schoolcategorie zijn relevante indicatoren voor de schoolkeuze. De bruto schoolverschillen worden zowel bepaald door goed onderwijs als door de samenstelling van de leerling populatie. Maar voor ouders hoeft het niet relevant te zijn om die te kunnen onderscheiden bij hun beslissing.

*Het netto-1 model*

Het corrigeren van bruto schoolverschillen voor achtergrondkenmerken van leerlingen, zoals SES, is te zien als een eerste stap om een indruk te krijgen in welke mate verschillen al dan niet binnen de invloedssfeer van de school liggen. In het kader van meritocratische beoordelingen van de kwaliteit van scholen, bijvoorbeeld door de Onderwijsinspectie, zijn “value-added” indicatoren van belang omdat ze een meer “faire” beoordeling van scholen mogelijk maken.

*Het netto-2 model*

Het vaststellen in welke mate schoolverschillen daadwerkelijk kunnen worden toegeschreven aan wat McKinsey (2020) noemt de “doelmatige beslissingen van besturen, directeuren en leraren”, is van belang omdat deze indicator een beeld geeft van de beïnvloedbaarheid van onderwijsprestaties door de direct betrokken actoren. De grootte van deze indicator kan worden gezien als een conservatieve schatting van de maakbaarheid van het onderwijs, en de kans op succes van hervormings- en verbeterprogramma’s. Het is een conservatieve schatting vanwege de complexiteit van het studieobject en tal van methodologische uitdagingen bij het onderzoek ernaar, waardoor “onder specificatie” van de modellen als een tamelijk chronisch probleem moet worden gezien. Een boven schatting van de “maakbaarheid” is de effect grootte van het type netto-1. Hiervoor leidt de ook daar niet zo gemakkelijk



te vermijden onder specificatie (bijvoorbeeld door het frequent ontbreken van eerdere prestaties als correctie factor) juist tot overschatting van de maakbaarheid. De illustratieve bewerkingen op basis van PISA-gegevens maken overigens duidelijk dat, zelfs bij een niet al te sterke set van correctie factoren op leerling niveau, ook de netto-1 marges betrekkelijk smal zijn.

## Noten

1. De tekst in paragraaf 2 en paragraaf 3 bouwt voort op tal van publicaties, die in deze discussiebijdrage niet uitvoerig geciteerd kunnen worden. Wel noemen we enkele relevante publicaties in de referentielijst. Fundamentele publicaties zijn van Raudenbush, & Bryk (1986), Raudenbush & Willms (1995) en Willms & Raudenbush (1989). Overige publicaties zijn van Luyten (2007), Opdenakker & Van Damme (2000) en Scheerens & Bosker (1997, hfst 3).
2. Meer in algemene zin is een ICC in principe te interpreteren als een correlatie tussen twee willekeurig getrokken leerlingen binnen een eenheid op hoger niveau.
3. Toegevoegde waarde is een indicator voor de relatieve prestaties (of ontwikkeling) van leerlingen in een school in vergelijking tot andere scholen in dezelfde steekproef, waarbij een correctie plaatsvindt voor verschillen tussen leerlingen bij binnenkomst van de school die een bijdrage kunnen leveren aan de ontwikkeling van leerlingen” (Timmermans, 2012, p. 192). Wanneer alleen eerdere prestaties als correctie factor worden opgenomen kan men de term “leerwinst” gebruiken.
4. Zie Raudenbush & Willms (1995). De relatie tussen schooleffecten van het type A en het type B, enerzijds en “netto 1” en “netto 2” effecten wordt uitgelegd in de appendix.
5. Reden om te kiezen voor PISA data is het gegeven dat de nadrukkelijke vermelding van Nederland als koploper van grote schoolverschillen in de rapporten van de Inspectie en McKinsey, (zie par. 1) voor wat betreft het Voortgezet Onderwijs beide op bruto schoolverschillen in PISA zijn gebaseerd.

## Literatuur

- Aloisi, A., & Tymms, P. (2017). PISA trends, social changes, and education reforms. *Educational Research and Evaluation*, 23(5-6), 180-220.
- Birnbaum, A. (1968). “Some latent trait models and their use in inferring a student’s ability”, In F. M. Lord and M. R. Novick (eds.). *Statistical Theories of Mental Test Scores*, Reading, MA, Addison-Wesley, pp. 397-479.
- Borman, G.D., Hewes, G.M., Overman, L.T., & Brown, S. (2003). Comprehensive school reform and achievement: A meta-analysis. *Review of Educational Research*, 73(2), 125-230.
- Bosker, R.J., and Witziers, B. (1995). School effects: Problems, solutions, and a meta-analysis. Paper presented at the Annual meeting of the Internationals Congress for School Effectiveness and School Improvement (ICSEI). Leeuwarden, the Netherlands.
- Davidov, E., Meuleman B., Cieciuch, J., Schmidt, P., & Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, 40(1), 55-75.
- Detterman, D. K. (2016). Education and intelligence: Pity the poor teacher because student characteristics are more significant than teachers or schools. *The Spanish Journal of Psychology*, 19(e93), 1-11.
- Dumay, X., Coe, R., & Anumendem, N.D., (2014). Stability over time of different methods of estimating school performance. *School effectiveness and school improvement*, 25(1), 64-82
- Ganzeboom, H. & D. Treiman (2003). “Three Internationally Standardised Measures for Comparative Research on Occupational Status”, In J.H.P Hoffmeyer-Zlotnik and C. Wolf C. (eds.). *Advances in Cross-National Comparison*, Boston Springer US, pp. 159-193.
- Goldstein, H. (2001). Using pupil performance data for judging schools and teachers: scope and limitations. *British Educational Research Journal*, 27(4), 433-442.
- Hill, P.W., & Rowe, K.J. (1996). Multilevel modelling in school effectiveness research. *School Effectiveness and School Improvement*, 7, 1-34.
- Inspectie van het Onderwijs. (2017). *De Staat van het Onderwijs 2015/2016*. Utrecht: Inspectie van het Onderwijs.
- Koch, G. (1982). “Intraclass correlation coefficient”. In Samuel Kotz and Norman L. Johnson

- (ed.). *Encyclopedia of Statistical Sciences*. 4. New York: John Wiley & Sons, pp. 213–217.
- Lortie-Forgues, H., & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? *Educational Researcher*, 48(3), 158–166.
- Luyten, H. (2007). Assessing the effect of schooling, alternatives for the focus on school differences. In : Scheerens, J., Luyten, H., Steen, R., & Y. Luyten-de Thouars (2007). *Review and meta-analyses of school and teaching effectiveness*. Enschede: University of Twente, Department of Educational Organization and Management
- Manzi, J., San Martín, E., & Van Belleghem, S. (2011). School System Evaluation by Value-Added Analysis Under Endogeneity. *Psychometrika*, 79(1), 130-153.
- Marks, G.N. (2020). Should value-added school effects models include student- and school-level covariates? Evidence from Australian population assessment data. *British Educational Research Journal*. (in press): DOI:10.002/berj.3684.
- McKinsey & Company (2020). *Een verstevigd fundament voor iedereen. Een onderzoek naar de doelmatigheid en toereikendheid van het funderend onderwijs (primair en voortgezet)*. Amsterdam: McKinsey & Company Netherlands.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm, *Applied Psychological Measurement*, 16(2), 159-177.
- Nota Brede Maatschappelijke Heroverweging (2020). Den Haag: Ministerie van OC&W.
- OECD (1999). *Classifying Educational Programmes: Manual for ISCED-97 Implementation in OECD Countries*, Paris: OECD Publishing.
- OECD (2013). *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, Paris: OECD Publishing.
- OECD (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*, Paris: OECD Publishing.
- OECD (2017a). *PISA 2015 Technical Report*. Paris: OECD Publishing.
- OECD (2017b). *PISA 2015 Assessment and Analytical Framework: Science, Reading, Mathematical, Financial Literacy and Collaborative Problem Solving*, revised edition, Paris: OECD Publishing.
- OECD (2019). *PISA 2018 Assessment and Analytical Framework*, PISA, OECD Publishing, Paris.
- OECD (2020). *PISA 2018 2015 Technical Report*, chapter 16: Scaling procedure and construct validation of context questionnaire data, Paris: OECD Publishing.
- Opendakker, M.-C., & Van Damme, J. (2000). Effects of schools, teaching staff and classes on achievement and well-being in secondary education: similarities and differences between school outcomes. *School Effectiveness and School Improvement*, 11(2), 165-196.
- Pit-Ten Cate, I., Krolak-Schwerdt, S., & Glock, S. (2016). Accuracy of teachers' tracking decisions: Short- and long-term effects of accountability. *European Journal of Psychology of Education*, 31(2), 225-243.
- Raudenbush, S. W. & Bryk, A. S. (1986). A Hierarchical model for studying school effects. *Sociology of Education*, 59, 1-17.
- Raudenbush, S., & Willms, J.D. (1995). The estimation of school effects. *Journal of Educational and Behavioral Statistics*, 20(4), 307–335.
- Raudenbush S.W. (2004). What are value-added models estimating and what does this imply for statistical practice? *Journal of Educational and Behavioral Statistics*, 29(1), 121-129.
- Rekers-Mombarg, L.T.M., Timmermans, A.C., & Bosker, R.J. (2014). *Vernieuwing van indicatoren voor leeropbrengsten in het voortgezet onderwijs: Een dieptestudie*. Groningen: Gion onderzoek/onderwijs.
- Scheerens, J., & Bosker, R. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Scheerens, J. (2016). *Educational Effectiveness and Ineffectiveness*. Dordrecht, Heidelberg, NewYork, London: Springer.
- Scheerens, J. (2017). The perspective of “limited malleability” in educational effectiveness. Treatment effects in schooling. *Educational Research and Evaluation*, 23(5-6) 247- 266
- Scheerens, J. (2019) (red.). *Niveau van onderwijsprestaties en verschillen tussen scholen. Enkele indicatoren van het Nederlandse funderend onderwijs*. Utrecht: Oberon.
- Timmermans, A. C. (2012). *Value added in educational accountability: possible, fair and useful?* GION, Gronings Instituut voor Onderzoek van Onderwijs, Rijksuniversiteit Groningen.

- Timmermans, A. C., & Thomas, S. M. (2015). The impact of student composition on schools' value-added performance: A comparison of seven empirical studies. *School Effectiveness and School Improvement*, 26(3), 487-498.
- Timmermans, A.C. & Van der Werf, G. (2017). Accounting for previous performance of students by means of growth curves analyses to estimate the size, stability and consistency of school effects. *Educational Research and Evaluation*, 23(5-6), 221 -246.
- Van De Werfhorst, H., & Mijs, J.J.B. (2010). Achievement inequality and the institutional structure of educational systems: A comparative perspective. *Annual Review of Sociology*, 36, 407-28.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices and recommendations for organizational research. *Organizational Research Methods*, 3(1), 4-70.
- Willms, J. D., & Raudenbush, S. W. (1989). A longitudinal hierarchical linear model for estimating school effects and their stability. *Journal of Educational Measurement*, 26(3), 209-232.

## Abstract

### Differences between schools and school effects in Dutch secondary education

As a result of PISA studies (2015 and 2018), the Inspectorate of Education concluded that differences between Dutch schools in secondary education are “worryingly large”. Anno 2020 bureau McKinsey did the same (McKinsey, 2020). This conclusion is based on the so-called ‘intra-class correlation coefficient’. However, mapping out differences in quality between schools is more complex and nuanced than the intra-class correlation comparison based on PISA data suggests. This article discusses some limitations of using the intra-class correlation coefficient as a measure for differences in quality between schools. Additionally, alternatives are discussed which express to what extent differences between schools are attributable to factors that can be influenced by the schools and factors that are due to background- and context characteristics; the so-called ‘bruto’ and ‘netto’

models. This discussion article contains illustrations for each of the alternatives and lists the situations in which one would opt for each of the alternatives.

Keywords: school effectiveness, school differences in quality, PISA

## Auteurs

**Kimberley Lek**, Wetenschappelijk onderzoeker, Cito. **Remco Feskens**, Hoofd psychometrisch onderzoek en dienstverlening, Cito. **J. Scheerens**, Emeritus hoogleraar Universiteit Twente.

*Correspondentieadres:* Kimberley Lek, Postbus 1034, 6801 MG Arnhem; E-mail: Kimberley.lek@cito.nl

## Appendix

### Technische toelichting op PISA

De belangrijkste cognitieve domeinen die in elke PISA cyclus worden gemeten zijn wiskunde, natuurwetenschappen en leesvaardigheid (OECD, 2017a). Elke cyclus staat een van deze domeinen centraal in de afname en wordt het construct opnieuw geoperationaliseerd. In 2012 was wiskunde het centrale hoofddomein, in 2015 was dit natuurwetenschappen en in 2018 leesvaardigheid. Een beschrijving van de operationalisaties zijn te vinden in OECD (2013), OECD (2017b) en OECD (2019). De responses op de items die de cognitieve vaardigheden meten worden geschaald met behulp van modellen uit de itemresponstheorie. Vanaf 2015 worden het 2-parameter logistisch model (Birnbaum, 1968) en het *generalized partial credit model* (Muraki, 1992) gebruikt (OECD, 2017a). De belangrijkste eenvoudige index die in het artikel gebruikt is, is het opleidingsniveau van de leerlingen. Leerlingen hebben zelf aangeven welk opleidingstype zij op dat moment volgen. De opleidingsniveaus zijn geclassificeerd op basis van de *International Standard Classification of Education (ISCED) 1997* (OECD, 1999). In PISA wordt de sociaaleco-

nomische status van leerlingen gemeten door een index die de economische-, sociale- en culturele status (ESCS) van leerlingen schat. Deze index is opgebouwd uit vier onderliggende constructen: een indicator voor het opleidingsniveau van de ouders; een indicator voor de arbeidspositie van de ouders; ruim een dozijn variabelen die huishoudbezittingen meten en een aantal vragen over het aantal onderwijskundige hulpmiddelen in het huishouden van de leerling. Het opleidingsniveau van de ouders is ook geclassificeerd op basis van ISCED 1997. De arbeidspositie van de ouders is geclassificeerd naar de *International Socio-Economic Index of Occupational Status* (ISEI) (Ganzeboom & Treiman, 2003). Het generalized partial credit model is ook gebruikt voor het schalen van constructen die door middel van meerdere vragen in de achtergrondvragenlijst zijn gemeten (OECD, 2017a; OECD, 2020).

### **Toelichting op type A, type B effecten in relatie tot “netto-1” en “netto-2” modellen**

In de literatuur wordt ook vaak gesproken over Type A en B effecten in de context van schooleffectiviteit (Raudenbush & Willms, 1995). Het type A effect is vooral van belang voor ouders die een school voor hun kind willen selecteren; het type B effect is meer van toepassing voor beleidsmakers die willen weten welk deel van leerprestaties toe te schrijven zijn aan door de school manipuleerbare factoren (Raudenbush, 2004). De leerprestaties kunnen beïnvloed worden door leerling eigenschappen, een toevallige component, de omgeving van een school (waar een school zelf niet of nauwelijks invloed op heeft) en uiteindelijk ook het schoolbeleid (wat wel onder de invloedssfeer van een school valt). Terwijl ouders niet perse geïnteresseerd zullen zijn om een onderscheid te kunnen maken of leerprestaties van hun kinderen nu verbeterd worden door de omgeving of door het beleid van de school (Type A effect), is het voor beleidsmakers wel van belang om vast te kunnen stellen welk van de leerprestaties nu bepaald zijn door het schoolbeleid (Type B effect). Type B effecten zijn daardoor geschikt om tot een faire meritocratische beoordeling te komen, door de scholen

zelf, of bijvoorbeeld ook door de Onderwijsinspectie. Juist omdat omgevingseffecten en schoolbeleid gecorreleerd kunnen zijn, is het niet eenvoudig een Type B effect goed vast te stellen (Raudenbush, 2004). In deze bijdrage gaan we in feite van hetzelfde regressiemodel uit, maar interpreteren we bruto en netto schooleffecten in termen van de grootte van variantie-componenten (met andere woorden in termen van het percentage variantie dat ze wel verklaren en niet in termen van de variantie die ze niet verklaren- de residuele schooleffecten uit het hierboven weergegeven model. Verder gaan we uit van een twee niveau model, school – en leerling niveau, waarbij de effecten van achtergrondkenmerken van leerlingen op leerlingniveau worden bepaald. De variantie-componenten die worden bepaald zijn achtereenvolgens:

1) De proportie van de totale variantie die bepaald wordt door de vraag op welke school een leerling zit, oftewel de “bruto tussenschoolse variantie”, de ICC uit een leeg multi-level model.

2) De proportie variantie die wordt bepaald door de achtergrondkenmerken op leerlingniveau, wanneer de schatting van deze component wordt afgetrokken van de schatting van de bruto tussenschoolse variantie. Wij noemen dit “netto-1” schooleffecten.

3) De proportie van de totale variantie die bepaald wordt door de onafhankelijke variabelen op schoolniveau (omgevingseffecten en schoolbeleid) uit het model. Door omgevings- en schoolbeleidseffecten af te trekken van de schatting van het netto-1 schooleffect vindt een verdere reductie plaats van het schooleffect. Dit zou een consequente interpretatie van een “netto-2 schooleffect” zijn (zie ook Tabel 2). Men zou er ook voor kunnen kiezen alleen de effectschatting van de schoolbeleidseffecten component aan te duiden als netto-2 effect. Dit effect is op te vatten als een “treatment effect” in schooleffectiviteitsonderzoek, omdat het de invloed weergeeft van de door de school beheersbare maatregelen om de opbrengst van het onderwijs te optimaliseren. Omgevingseffecten zijn contextueel en niet beïnvloedbaar door de school.