



Verschillen in leerresultaten basisscholen

Cito-score zegt niet alles over kwaliteit basisschool
Belangrijke andere verklaringen zijn:

1. verschillen in achtergrond van leerlingen
2. toevallige fluctuaties per jaar
3. verschillen in kenmerken en prestaties van medeleerlingen

Deze drie verklaren samen een aanzienlijk deel van de verschillen tussen basisscholen.

Leerresultaat moet over meerdere jaren worden berekend.

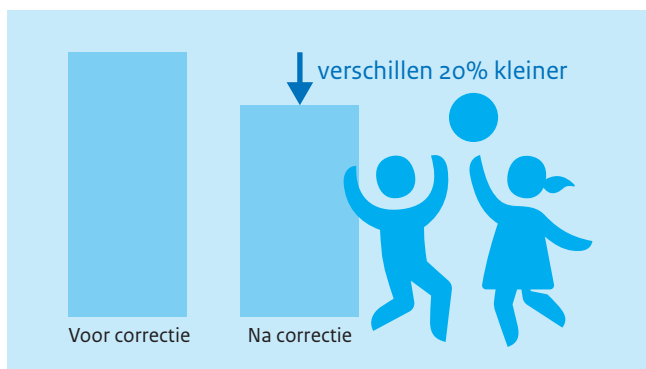
Langere meetperiode geeft een betrouwbaarder beeld van de schoolprestaties en is voor beleidsmakers en ouders relevanter dan een jaarresultaat.

Leerresultaten basisscholen

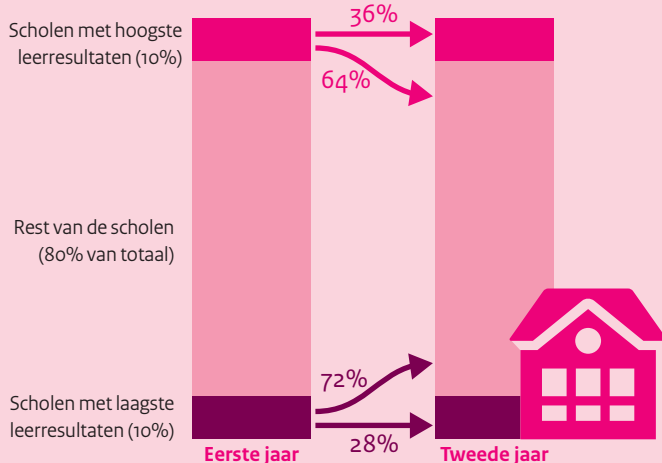
Vier oorzaken van verschillen in leerresultaten tussen scholen

1. Achtergrond leerlingen

Een vijfde van verschillen tussen scholen wordt verklaard door de achtergrond van leerlingen, zoals familie.



Weinig scholen blijven jaar op jaar hoge of lage leerresultaten halen



2. Toeval

Resultaten variëren vaak sterk per school. Hoge en lage scores worden een jaar later meestal niet herhaald. Dit komt doordat het schoolresultaat het gemiddelde is van de eindtoets-score van een beperkt aantal leerlingen. Daardoor spelen toevallige uitschieters een grote rol.

3. Medeleerlingen

Voor leerprestaties van een leerling maakt het uit wie de medeleerlingen zijn en hoe zij presteren.

positief verband negatief verband



Meer leerlingen met ouders met hoog inkomen



Minder leerlingen met laagopgeleide ouders



Grote verschillen in leerresultaten

positief verband



Ervaren leraren

negatief verband



Relatief veel nieuw personeel

4. Kenmerken school

Het laatste deel van de verschillen wordt verklaard door de onderwijskwaliteit.

Samenvatting

Het ministerie van Onderwijs, Cultuur & Wetenschap heeft het CPB gevraagd onderzoek te doen naar de volgende drie onderzoeksvragen:

1. Hoe groot zijn de verschillen in leerresultaten tussen scholen nadat zo goed mogelijk rekening is gehouden met verschillen in leerlingpopulatie?
2. Variëren de schoolverschillen, die overblijven na rekening te houden voor kenmerken van leerlingen, sterk over tijd (voor dezelfde school)?
3. Welke school- en lerarenkenmerken hangen samen met deze verschillen?

Deze drie vragen richten zich op (alternatieve) verklaringen voor de verschillen in leerresultaten tussen scholen. De eerste onderzoeksvraag richt zich op verschillen in leerlingpopulatie als verklaring voor verschillen in leerresultaten (dat wil zeggen de gemiddelde cito-score van een school). De tweede vraag onderzoekt in hoeverre toevallige uitschieters een rol spelen, omdat het leerresultaat wordt berekend als gemiddelde over een klein aantal leerlingen. De laatste onderzoeksvraag gaat in op twee alternatieve verklaringen. Enerzijds op verschillen in leerresultaten als gevolg van groepseffecten (voor de leerresultaten van een individuele leerling maakt het uit wie zijn klasgenoten zijn). Anderzijds kunnen scholen ook in kwaliteit verschillen. We onderzoeken hoe kenmerken van medeleerlingen en kenmerken van de school samenhangen met schoolverschillen.

Verschillen in leerresultaten (gemiddelde cito-score) tussen scholen mogen niet geïnterpreteerd worden als kwaliteitsverschillen. Deze notitie laat zien dat elk van de drie alternatieve verklaringen een aanzienlijke rol speelt in de totstandkoming van verschillen in leerresultaten tussen scholen. Een-vijfde van de verschillen tussen scholen kan verklaard worden door verschillen in achtergrondkenmerken van de leerlingpopulatie. Daarnaast blijkt de voor leerlingpopulatie gecorrigeerde cito-score (het schooleffect) flink te fluctueren van jaar op jaar. Door leerlingen uit meerdere jaren samen te nemen, kan er voor gezorgd worden dat toevallige uitschieters een kleinere rol spelen. Ten slotte blijken ook de kenmerken en prestaties van medeleerlingen invloed te hebben op de prestatie van individuele leerlingen en daarmee op de leerresultaten van de school.

Hoe groot zijn de verschillen in leerresultaten tussen scholen nadat zo goed mogelijk rekening is gehouden met verschillen in leerlingpopulatie?

Door te corrigeren voor individuele leerlingkenmerken nemen de verschillen tussen scholen af met 20%. Na correctie ligt het leerresultaat van een school in 2016 gemiddeld 2,55 cito-punten van het verwachte leerresultaat op basis van de individuele leerlingkenmerken. 10% van de scholen heeft een leerresultaat dat meer dan 5,3 cito-punten afwijkt. Wanneer resultaten op een capaciteitentest beschikbaar zijn, kan nog nauwkeuriger gecorrigeerd worden en nemen verschillen tussen scholen nog wat verder af (ongeveer 5%-punt extra afname). Resultaten van een capaciteitentest zijn echter maar beperkt beschikbaar.

Variëren de schoolverschillen, die overblijven na rekening te houden met populatie, sterk over tijd (voor dezelfde school)?

Het schooleffect (leerresultaat van een school gecorrigeerd voor leerlingkenmerken) kan aanzienlijk fluctueren van jaar op jaar. Omdat de leerresultaten van een school berekend worden als de gemiddelde cito-score van de leerlingen in groep 8, en het aantal leerlingen in groep 8 niet erg groot is (veelal 1 of 2 parallelklassen), is het statistisch gezien logisch dat er over tijd fluctuatie zit in de gemiddelde cito-score van een school. Als een school toevallig in een jaar uitzonderlijk goede leerprestaties van de leerlingen kent gegeven hun achtergrondkenmerken, zal het leerresultaat van de school eenmalig hoog zijn en in de volgende jaren naar verwachting weer rond de voor deze school 'normale' waarde liggen.

Van de 10% scholen met het laagste schooleffect in een jaar, behoort slechts 28% een jaar later weer tot de 10% scholen met het laagste schooleffect. De rest scoort hoger. Zo scoort de 1% scholen met het laagste schooleffect in het daaropvolgende jaar gemiddeld 6,5 punt hoger. Deze fluctuaties zijn er niet alleen bij de scholen met de laagste schooleffecten, maar ook bij de scholen met de hoogste schooleffecten. Van de 10% scholen met het hoogste schooleffect in een jaar, behoort slechts 36% in het jaar daarna weer tot de 10% scholen met het hoogste schooleffect. De rest scoort lager. De 1% scholen met het hoogste schooleffect scoort in het jaar daarop gemiddeld 3,5 punt lager.

Omdat het schooleffect flink fluctueert over tijd, is het schooleffect in een bepaald jaar geen betrouwbare indicator voor de kwaliteit van een school. Voor beleidsmakers en ouders is het veel relevanter hoe een school het gemiddeld over meerdere jaren doet, waarbij toevallige fluctuaties minder een rol spelen. Het verdient daarom aanbeveling verschillende cohorten leerlingen samen te nemen voor het berekenen van het schooleffect. Door leerjaren samen te voegen, nemen de verschillen tussen scholen aanzienlijk af. Nemen we drie leerjaren samen, dan daalt de gemiddelde afwijking van het verwachte leerresultaat naar 1,9 cito-punten. Bij acht leerjaren samen daalt dit verder naar 1,5 cito-punten.

Welke school- en lerarenkenmerken hangen samen met deze verschillen?

Grote verschillen in leerresultaten tussen leerlingen op een school, meer medeleerlingen met laagopgeleide ouders en meer medeleerlingen uit gezinnen met een lager inkomen hangen samen met een lager schooleffect. We vinden, na correctie voor individuele kenmerken en het samenvoegen van gegevens uit meerdere jaren, dat grotere verschillen in leerresultaten van individuele leerlingen binnen een school samenhangen met een lager schooleffect. De 20% scholen met de meeste spreiding hebben gemiddeld een 1 cito-punt lager schooleffect. Daarnaast blijkt er een samenhang te zijn van minder schoolgenoten met laagopgeleide ouders met een hoger schooleffect en ook van meer schoolgenoten met ouders met een hoog inkomen met een hoger schooleffect. In beide gevallen gaat het om een effect van ongeveer 1 cito-punt. Er is geen eenduidige samenhang tussen het schooleffect en het aandeel schoolgenoten met een migratieachtergrond, wanneer ook gecorrigeerd wordt voor opleidingsniveau en inkomen van de ouders van alle leerlingen op school. De verbanden beschreven in deze notitie zijn correlaties en kunnen geen oorzakelijk verband aantonen.

Meer ervaren leerkrachten hangt samen met een hoger schooleffect. Als we goed in staat zijn te corrigeren voor verschillen in samenstelling van de leerlingpopulatie, voor toevallige uitschieters en voor groepseffecten, blijft kwaliteit over als verklaring voor verschillen in leerresultaten tussen scholen. We onderzochten welke schoolkenmerken met het resterende schooleffect samenhangen. Het gaat hier wederom om correlaties en niet om causale verbanden. Meer ervaring van leraren blijkt samen te hangen met een hoger schooleffect, wat in lijn is met de bevinding uit de literatuur dat ervaring in belangrijke mate bijdraagt aan de kwaliteit van het gegeven onderwijs. Het verschil in schooleffecten tussen de scholen met gemiddeld oudste (≥ 50 jaar) en jongste ($< 37,5$ jaar) lesgevend personeel bedraagt ongeveer 0,8 cito-punten. Ook vinden we dat scholen met relatief veel nieuw personeel gemiddeld lagere schooleffecten hebben.

Hoe verhouden deze bevindingen zich met die in de Staat van het Onderwijs 2015/2016?

Deze analyse verrijkt de bevindingen uit de Staat van het Onderwijs. In 2017 observeerde de Inspectie van het Onderwijs in haar jaarlijkse Staat van het Onderwijs 2015/2016 dat de gemiddelde cito-score sterk kan verschillen tussen basisscholen met een vergelijkbaar aandeel gewichtenleerlingen. Hierbij namen de onderzoekers drie leerjaren samen. In ons onderzoek vinden we dat een correctie met *individuele* leerlingkenmerken de schoolverschillen 4%-punt meer verkleint dan een correctie op basis van het aandeel gewichtenleerlingen.¹ Daarnaast hebben we de analyse uitgebreid door de schooleffecten over tijd te volgen en brengen we in kaart hoeveel de schoolverschillen afnemen als we meerdere leerjaren samenvoegen. Ten slotte brengen we in kaart welke groeps- en schoolkenmerken samenhangen met de resterende schooleffecten na correctie voor individuele kenmerken en fluctuaties over tijd.

¹ De gewichtenregeling is een belangrijk onderdeel van het onderwijsachterstandenbeleid. Deze regeling bepaalt welke kinderen extra ondersteuning nodig hebben. Deze kinderen krijgen een extra 'gewicht' mee, waarmee de school aanspraak maakt op extra geld om onderwijsachterstanden weg te werken. De bepaling van het gewicht is recent herzien (Centraal Bureau voor de Statistiek, 2016). Deze vernieuwde methode is vergelijkbaar met de correctie op basis van individuele leerlingkenmerken die we in deze notitie uitvoeren.

1 Inleiding

Het ministerie van Onderwijs, Cultuur en Wetenschap heeft het Centraal Planbureau (CPB) verzocht om onderzoek te doen naar de verschillen in leerresultaten tussen basisscholen. In 2017 observeerde de Inspectie van het Onderwijs in haar jaarlijkse Staat van het Onderwijs dat de gemiddelde cito-score sterk kan verschillen tussen basisscholen met een vergelijkbaar aandeel gewichtenleerlingen.²

De verschillen in cito-scores tussen scholen zijn zodanig groot dat ze kunnen leiden tot andere adviezen voor de brugklas. Gemiddeld wijkt een school ongeveer 2,54 cito-punten af van wat je zou verwachten op basis van het aandeel gewichtenleerlingen. 10% van de scholen ligt zelfs meer dan 5,37 cito-punten af van deze verwachtingswaarde. Dat kan bijvoorbeeld het verschil betekenen tussen een vwo-advies en een havo/vwo-advies.

Gegeven de omvang van deze verschillen heeft het ministerie van OCW het CPB gevraagd verder onderzoek te doen naar de volgende drie onderzoeksvragen:

1. Hoe groot zijn de verschillen tussen scholen nadat zo goed mogelijk rekening is gehouden met verschillen in leerlingpopulatie?
2. Variëren de schoolverschillen, die overblijven na rekening te houden voor populatie, sterk over tijd (voor dezelfde school)?
3. Welke school- en lerarenkenmerken hangen samen met deze verschillen?

Al deze drie vragen richten zich op (alternatieve) verklaringen voor de verschillen in leerresultaten (dat wil zeggen de gemiddelde cito-score van een school). Daarom zet paragraaf 2 eerst de mogelijke oorzaken van verschillen in leerresultaten uiteen. Vervolgens toetsen paragrafen 3, 4 en 5 empirisch hoe belangrijk elk van deze alternatieve verklaringen is voor het verschil in leerresultaten tussen scholen.

² De gewichtenregeling is een belangrijk onderdeel van het onderwijsachterstandenbeleid. Deze regeling bepaalt welke kinderen extra ondersteuning nodig hebben. Deze kinderen krijgen een extra 'gewicht' mee, waarmee de school inspraak maakt op extra geld om onderwijsachterstanden weg te werken.

2 Verschillen in leerresultaten verklaren

We onderscheiden vier verklaringen voor verschillen in leerresultaten tussen scholen in een bepaald jaar:

1. Verschillen in de leerlingpopulatie;
2. Toevallige uitschieters omdat de gemiddelde cito-score berekend wordt op een (meestal) klein aantal leerlingen (de leerlingen in groep 8);
3. Groepseffecten (*peer effects*);
4. Kwaliteit van het onderwijs.

Welke van deze verklaringen de grootste bijdrage levert aan verschillen in leerresultaten tussen scholen is van groot belang voor beleidsmakers. Verschillen in samenstelling van de leerlingpopulatie vragen namelijk om heel andere beleidsinterventies dan verschillen in de kwaliteit van het onderwijs. De notitie is zo opgebouwd dat al deze punten in volgorde aan bod komen. Paragraaf 3 laat zien hoe groot de verschillen in leerresultaten zijn als zo goed mogelijk voor verschillen in leerlingpopulatie wordt gecorrigeerd. Paragraaf 4 onderzoekt de fluctuatie over tijd van de gemiddelde cito-score na correctie voor achtergrondkenmerken. Tot slot bespreken we in paragraaf 5 de samenhang tussen leerresultaten en groepskenmerken en maatstaven voor onderwijskwaliteit.

Een belangrijke en voor de hand liggende verklaring voor verschillen in leerresultaten is dat scholen verschillen in de samenstelling van hun leerlingpopulatie. Een school met veel leerlingen met een hoog IQ zal over het algemeen hogere leerresultaten hebben dan een school die overwegend lesgeeft aan leerlingen met een laag IQ. Daarnaast kunnen andere achtergrondkenmerken een rol spelen, zo kan bijvoorbeeld de mate waarin ouders hun kind stimuleren en kunnen ondersteunen een factor van belang zijn.

Het is daarom belangrijk zo goed mogelijk te corrigeren voor achtergrondkenmerken en eigenschappen van leerlingen. In sommige gevallen, bijvoorbeeld aanleg en intelligentie van een kind, is vaak geen informatie over het kenmerk zelf voorhanden, maar is er wel informatie beschikbaar over iets waarvan bekend is dat het sterk samenhangt met dit kenmerk. Voor aanleg en intelligentie is dat bijvoorbeeld het opleidingsniveau van de ouders. Paragraaf 3 laat zien hoe groot de verschillen in leerresultaten zijn, als zo goed mogelijk voor achtergrondkenmerken wordt gecorrigeerd.

Een tweede oorzaak van verschillen in leerresultaten in een bepaald jaar zijn toevallige uitschieters. De leerresultaten van een school in een bepaald jaar zijn een momentopname. Omdat de leerresultaten van een school berekend worden als de gemiddelde cito-score van de leerlingen in groep 8, en het aantal leerlingen in groep 8 niet erg groot is (veelal 1 of 2 parallelklassen), is het statistisch gezien logisch dat er over tijd fluctuatie zit in de gemiddelde cito-score van een school. Als

een school toevallig in een jaar uitzonderlijk goede leerprestaties van de leerlingen kent gegeven hun achtergrondkenmerken, zal het leerresultaat van de school eenmalig hoog zijn en in de volgende jaren naar verwachting weer rond de voor deze school 'normale' waarde liggen. Paragraaf 4 van deze notitie onderzoekt hoeveel jaar op jaar fluctuatie er zit in de gemiddelde cito-score na correctie voor achtergrondkenmerken. Als uitschieters naar boven en beneden geregeld voorkomen, zou dat ervoor pleiten om niet te varen op gegevens over één jaar. In plaats daarvan zou een gemiddelde over meerdere jaren gehanteerd moeten worden.

Een derde verklaring voor verschillen in leerresultaten tussen scholen is dat naast individuele leerlingkenmerken ook kenmerken van de medeleerlingen in de groep ('peers') een rol spelen in de totstandkoming van het leerresultaat van een leerling. Zo kan het zijn dat een slimme leerling beter tot zijn recht komt in een klas met veel slimme kinderen. Of profiteert een minder slimme leerling van een hoog percentage slimme leerlingen in zijn klas, omdat hij zich aan hen kan optrekken. Paragraaf 5 van deze notitie bekijkt de samenhang tussen groepskenmerken en leerresultaten. Het idee is dat de interacties tussen individuele leerlingen en groepen van medeleerlingen ('peers') binnen de groep een effect hebben op de leerresultaten. Hierbij gaat het niet om het effect van individuele leerlingkenmerken, maar het effect van de samenstelling van de groep als geheel. In internationaal onderzoek worden dit soort effecten ook wel *peer effects* genoemd. Paragraaf 5.1 onderzoekt de samenhang tussen groepskenmerken en het schooleffect (leerresultaat van een school gecorrigeerd voor leerlingkenmerken).

Ten slotte kunnen verschillen in leerresultaten ontstaan doordat scholen in kwaliteit verschillen. Een voorbeeld hiervan is de ervaring van leerkrachten. Uit de literatuur is bekend dat leerlingen meer leren van ervaren leerkrachten dan van beginnende leerkrachten. Een school met meer ervaren leerkrachten kan daardoor meer toegevoegde waarde hebben voor een leerling.

Het vaststellen van een oorzakelijk verband tussen schoolkenmerken en de kwaliteit van een school is niet mogelijk gegeven de data. Wel kan de samenhang tussen schoolkenmerken en de leerresultaten van een school (na correctie voor individuele leerlingkenmerken en jaarlijkse fluctuaties in de score) onderzocht worden, en zo handvatten bieden voor verder (causaal) onderzoek. Deze analyse staat centraal in paragraaf 5.2.

3 Hoeveel kunnen achtergrondkenmerken verklaren?

Door te corrigeren voor individuele leerlingkenmerken, nemen de verschillen tussen scholen af met 20%, maar er blijven aanzienlijke verschillen bestaan. Als scholen verschillen in de samenstelling van de leerlingpopulatie, zullen scholen daardoor verschillen in leerresultaten. Wanneer rekening gehouden wordt met verschillen in leerlingpopulatie nemen de verschillen tussen scholen af. Hoewel de verschillen kleiner worden, blijven ze aanzienlijk.

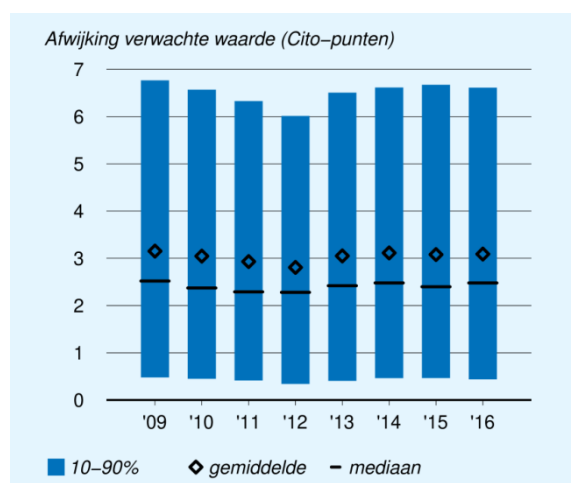
Zonder correctie voor leerlingpopulatie zijn de verschillen in leerresultaat groot. Wanneer geen rekening gehouden wordt met verschillen in leerlingpopulatie, is voor iedere school het verwachte leerresultaat (gemiddelde cito-score van de school³) gelijk aan de gemiddelde cito-score van alle groep8-leerlingen in Nederland. In 2016 was de gemiddelde cito-score 535. Gemiddeld wijkt een school in 2016 3,09 cito-punt⁴ van het verwachte leerresultaat van 535 af (figuur 3.1). De bovenkant van de box geeft weer dat de 10% scholen waarvoor het leerresultaat het meest afwijkt van het verwachte leerresultaat, in 2016 een gemiddelde cito-score heeft die meer dan 6,6 punten van de verwachte waarde afligt. Dat kan bijvoorbeeld het verschil betekenen tussen een vwo-advies en een havo-advies (zie tabel 3.1). De onderkant van de box geeft aan dat 90% van de scholen meer dan 0,44 cito-punten afwijkt. Hoe lager deze punten liggen en hoe korter de verticale balk is, des te kleiner zijn de verschillen tussen scholen. De leerresultaten van scholen liggen dan dichterbij de cito-score die je zou verwachten.

Tabel 3.1 Overzicht toetsadviezen

Interval standaardscore	Brugklastype
501 – 518	vmbo basisberoepsgerichte leerweg
519 – 525	vmbo basis- en kaderberoepsgerichte leerweg
526 – 528	vmbo kaderberoepsgerichte leerweg
529 – 532	vmbo gemengde / theoretische leerweg
533 – 536	vmbo gemengde / theoretische leerweg en havo
537 – 539	havo
540 – 544	havo / vwo
545 – 550	vwo

Bron: terugblik centrale eindtoets 2016.

Figuur 3.1 Afwijking van verwachte waarde van het leerresultaat zonder correctie voor verschillen in leerlingpopulatie



Bron: eigen berekeningen op basis van CBS-data.

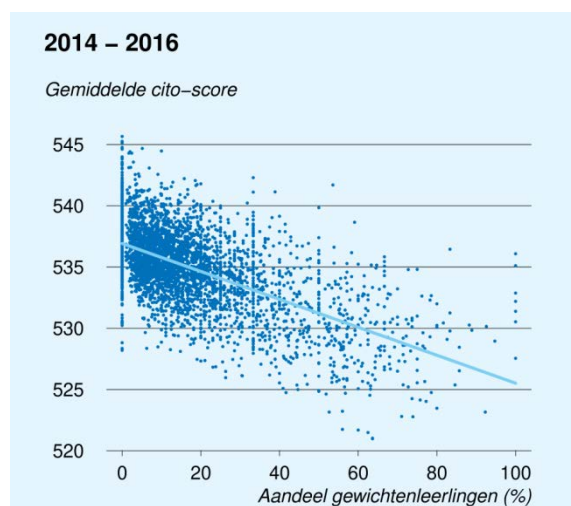
³ Alleen scholen met minstens tien observaties (cito-scores) per jaar worden meegenomen. Ook kijken we alleen naar scholen die in de gehele periode 2009-2016 gebruikmaakten van de Cito-toets als eindtoets.

⁴ Het betreft hier de gemiddelde absolute afwijking.

In de Staat van het Onderwijs 2015/2016 werd het percentage gewichtenleerlingen gebruikt om te corrigeren voor verschillen in de leerlingpopulatie. Onder de gewichtenregeling ontvangen scholen extra geld voor leerlingen waarvan de ouders laag opgeleid zijn. Figuur 3.2 laat zien dat ook als het aandeel gewichtenleerlingen vergelijkbaar is, de gemiddelde score op de eindtoets sterk verschilt tussen scholen. In de figuur is te zien dat de gemiddelde eindscore lager is bij scholen met een hoger aandeel gewichtenleerlingen. Scholen met hetzelfde aandeel gewichtenleerlingen hebben echter nog steeds verschillende leerresultaten: de bolletjes liggen niet op één lijn. De gemiddelde citoscore van scholen zonder gewichtenleerlingen (aandeel 0%) loopt bijvoorbeeld uiteen van 525 tot 546. De correctie voor leerlingpopulatie met behulp van gewichtenleerlingen zal dus geen grote afname in verschillen tussen scholen bewerkstelligen.

Verschillen in leerresultaten nemen 16% af door te corrigeren voor het aandeel gewichtenleerlingen. Rekening houdend met het aandeel gewichtenleerlingen wijkt het daadwerkelijke leerresultaat van een school in 2016 gemiddeld 2,62 cito-punten af van het verwachte leerresultaat op basis van het aandeel gewichtenleerlingen (figuur 3.3). Het verwachte leerresultaat van een school gegeven het percentage gewichtenleerlingen is hierbij het gemiddelde leerresultaat voor scholen met hetzelfde percentage gewichtenleerlingen. De bovenkant van de box geeft weer dat 10% van de scholen meer dan 5,58 cito-punten van de verwachte waarde afligt. De onderkant van de box geeft aan dat 90% van de scholen meer dan 0,37 cito-punten afwijkt. Vergelijking van de verschillen in leerresultaten zonder (figuur 3.1) en met (figuur 3.3) correctie voor het aandeel gewichtenleerlingen leert dat de verschillen in leerresultaten met ongeveer 16% per jaar afnemen (zie figuur 3.4) door het corrigeren voor het leerlinggewicht.⁵

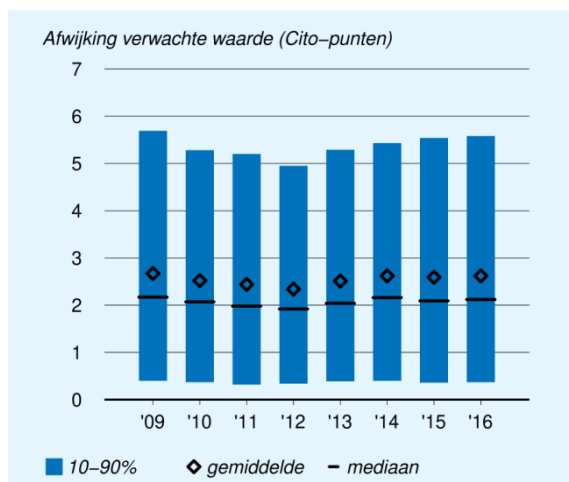
Figuur 3.2 De Staat van het Onderwijs 2015/2016: de gemiddelde score bij de eindtoets is lager op scholen met een hoog aandeel gewichtenleerlingen



Bron: eigen berekeningen op basis van CBS-data.
Opmerking: gemiddelde schoolscore gemeten over 3-jaarsperiode.

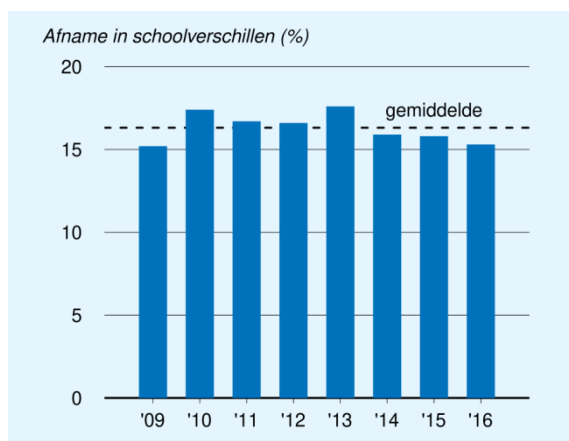
⁵ Dit is berekend als de procentuele afname in de totale som van de absolute afwijkingen.

Figuur 3.3 Afwijking van verwachte waarde van het leerresultaat na correctie voor het aandeel gewichtenleerlingen



Bron: eigen berekeningen op basis van CBS-data.

Figuur 3.4 Door te corrigeren voor het aandeel gewichtenleerlingen nemen de verschillen in leerresultaten met ongeveer 16% per jaar af



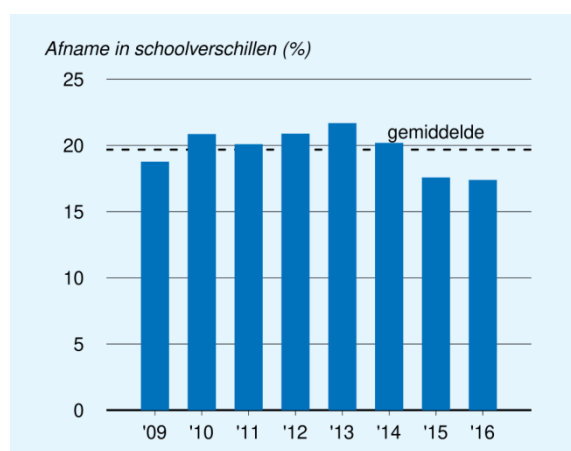
Bron: eigen berekeningen op basis van CBS-data.

Door leerlingkenmerken op individueel niveau te gebruiken, kan nauwkeuriger rekening gehouden worden met verschillen in leerlingpopulatie. Door middel van CBS-microdatabestanden kunnen we individuele leerlingkenmerken gebruiken om de leerlingpopulatie van een school te bepalen. Dit is vergelijkbaar met de nieuwe methode voor het vaststellen van het leerlinggewicht (Centraal Bureau voor de Statistiek, 2016). Door gegevens op individueel (dat wil zeggen leerling) niveau te gebruiken en meer achtergrondkenmerken mee te nemen kan preciezer rekening worden gehouden met verschillen in leerlingpopulatie tussen scholen. De achtergrondkenmerken die we hierin meenemen zijn opleidingsniveau van beide ouders, migratie-achtergrond, verblijfsduur in Nederland van de moeder, schuldenproblematiek, samenstelling van het huishouden, belangrijkste inkomensbron van het huishouden en het besteedbaar inkomen van het huishouden.

De verschillen in leerresultaten tussen scholen nemen met 20% af na correctie voor individuele leerlingkenmerken (figuur 3.5). De verschillen in leerresultaten tussen scholen die overblijven na correctie voor individuele leerlingkenmerken noemen we het schooleffect. Een school wijkt

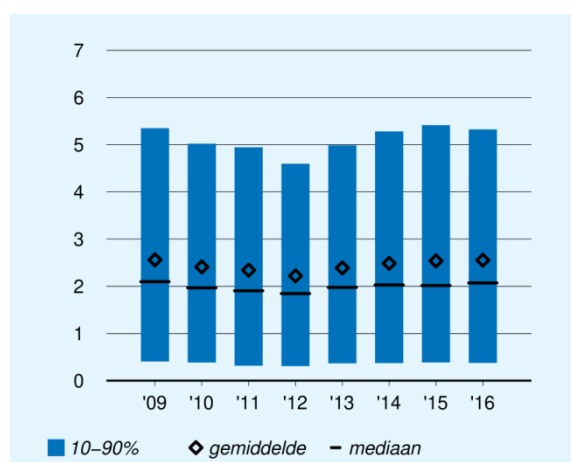
gemiddeld 2,55 cito-punten afwijkt af van wat je zou verwachten op basis van de individuele leerlingkenmerken (figuur 3.6). Dat is dus wat lager dan bij een correctie op basis van leerlinggewicht. 10% van de scholen wijkt meer dan 5,32 cito-punten van de verwachte waarde af, en 90% van de scholen wijkt meer dan 0,38 cito-punten af. In totaal nemen de schoolverschillen ongeveer 20% af ten opzichte van de verschillen zonder correctie (figuur 3.5). Dit is ook terug te zien in de verdeling van de schooleffecten (figuur 3.7). Meer scholen liggen dicht bij de verwachte waarde na controle voor individuele leerlingkenmerken, te zien aan het feit dat de scholendichtheid rond nul toeneemt en afneemt verder weg van de nul (verder dan 4 cito-punten). Er blijven echter nog steeds substantiële verschillen tussen scholen bestaan.

Figuur 3.5 Door te corrigeren voor individuele leerlingkenmerken nemen de verschillen in leerresultaten met ongeveer 20% af (jaar-op-jaar)



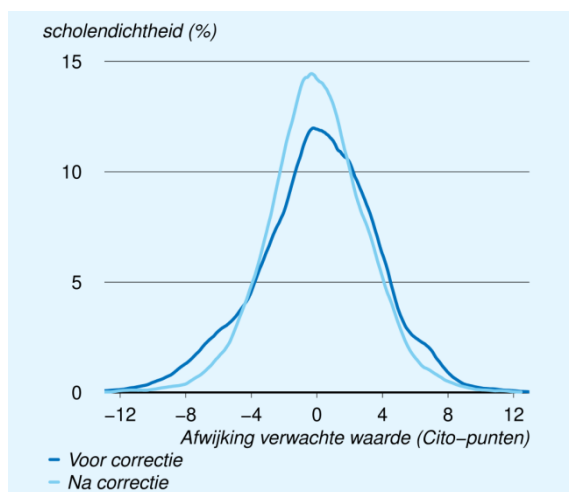
Bron: eigen berekeningen op basis van CBS-data.

Figuur 3.6 Afwijking van verwachte waarde leerresultaat kleiner na correctie voor individuele leerlingkenmerken



Bron: eigen berekeningen op basis van CBS-data.

Figuur 3.7 Aandeel scholen dat dicht om de verwachte waarde van het leerresultaat ligt is hoger na correctie voor individuele leerlingkenmerken



Bron: eigen berekeningen op basis van CBS-data.
Opmerking: Resultaten van leerjaar 2016.

Nauwkeuriger corrigeren voor capaciteiten van leerlingen verkleint verschillen nog iets verder.

Een van de belangrijkste redenen om rekening te houden met de leerlingpopulatie van een school is dat leerlingen verschillen in capaciteiten (zie paragraaf 1). Omdat een maatstaf voor capaciteiten niet beschikbaar is, worden kenmerken die sterk samenhangen met capaciteiten van een leerling gebruikt, zoals opleidingsniveau van de ouders. Voor een beperkt deel van de scholen zijn in 2011 en 2014 resultaten van hun leerlingen op de Niet-Schoolse-Cognitieve-Capaciteiten-Test (NSCCT) beschikbaar. Voor deze scholen is een meer nauwkeurige correctie voor verschillen tussen leerlingen dus mogelijk. Waar op basis van opleidingsniveau van de ouders de verwachte capaciteiten van broertjes en zusjes gelijk zijn, kan de NSCCT wél differentiëren.⁶

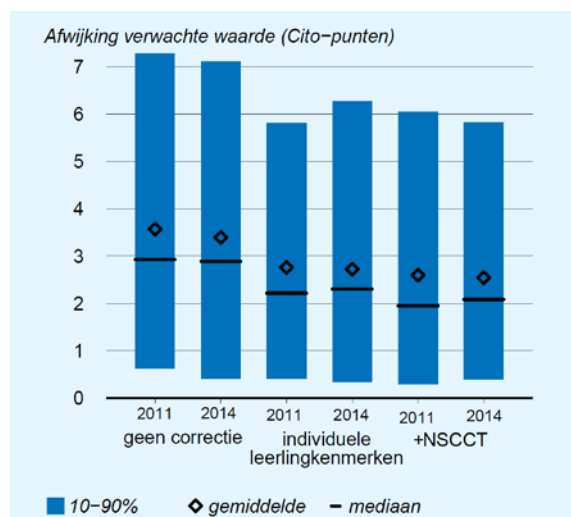
Omdat deze data alleen beschikbaar zijn voor een kleine steekproef in 2011 en 2014, kunnen de analyses met de NSCCT-scores onze algemene analyse niet vervangen. Wel geeft het een idee hoe goed de andere achtergrondkenmerken kunnen corrigeren voor verschillen in capaciteiten door te kijken in welke mate de verschillen in leerresultaten nog extra verminderen als nauwkeuriger gecorrigeerd kan worden voor verschillen in capaciteiten tussen leerlingen.

Maar ook als nauwkeuriger gecorrigeerd kan worden voor verschillen in capaciteiten tussen leerlingen, resteren substantiële verschillen in schooleffecten. Figuur 3.8 laat zien dat wanneer de NSCCT-scores meegenomen worden in de set van leerlingkenmerken, de verschillen in schooleffecten nog iets verder afnemen. Voor de groep scholen waarvoor de NSCCT-score beschikbaar is, leidt de correctie voor individuele leerlingkenmerken, zoals hierboven, tot een afname van ongeveer 21% in de verschillen tussen scholen, die verder afneemt met ongeveer 5%-punt wanneer de NSCCT-scores meegenomen worden. De kleine groep scholen waarvoor de NSCCT-scores beschikbaar zijn, is niet helemaal representatief voor alle scholen. Onder scholen waarvoor we NSCCT-scores hebben is de gemiddelde afwijking zonder controle voor leerlingpopulatie gemiddeld 3,5 cito-punten (de ruit in de linker balken in figuur 3.8). Door te corrigeren voor

⁶ Omdat de NSCCT in groep 5 wordt afgenomen, meten we in deze analyse de toegevoegde waarde van de school tussen groep 5 en groep 8.

individuele achtergrondkenmerken die voor alle scholen beschikbaar zijn neemt dit af tot gemiddeld 2,75 cito-punten (de ruit in de middelste balken in figuur 3.8). Als daar een correctie voor de NSCCT-score aan wordt toegevoegd, neemt het gemiddelde schooleffect nog iets verder af tot 2,5 cito-punten (de streep in de rechter balken in figuur 3.8).⁷ Voor de analyses met alle scholen betekent dit dat een (klein) deel van de schoolverschillen die we laten zien nog verklaard kan worden door verschillen in capaciteiten tussen leerlingen.

Figuur 3.8 Afwijking van verwachte waarde van het leerresultaat neemt iets verder af na opnemen van maatstaf voor intelligentie (NSCCT-score)



4 De stabiliteit van verschillen in leerresultaten over meerdere jaren

Jaar op jaar stabiliteit van schooleffecten is beperkt. We vinden dat de leerresultaten, na correctie voor de samenstelling van de leerlingpopulatie, fluctueren van jaar op jaar⁸. Dit soort toevallige uitschieters zijn een tweede verklaring voor verschillen in leerresultaten tussen scholen in een bepaald jaar. Omdat de leerresultaten van een school berekend worden als de gemiddelde cito-score van de leerlingen in groep 8, en het aantal leerlingen in groep 8 niet erg groot is (veelal 1 of 2 parallelklassen), is het statistisch gezien logisch dat er over tijd fluctuatie zit in de gemiddelde cito-score van een school. Omdat de leerresultaten van een school in een bepaald jaar een momentopname zijn, is het mogelijk dat een school toevallig in een jaar betere of mindere leerresultaten kent, bijvoorbeeld als de groep leerlingen toevallig uitzonderlijk goed presteert gegeven hun achtergrondkenmerken. In dat geval zal het leerresultaat van de school eenmalig hoog

⁷ In 2011 neemt alleen de 10% scholen met de hoogste afwijking toe. De mediaan, gemiddelde en de waarde van het onderste deciel nemen af. Ook over het geheel gezien (procentuele daling in de som van de absolute afwijkingen) nemen de verschillen tussen scholen af.

⁸ Voor de correlaties tussen schooleffecten over tijd, zie appendix 7.4 (zonder correctie voor leerlingkenmerken) en 7.5 (met correctie voor leerlingkenmerken).

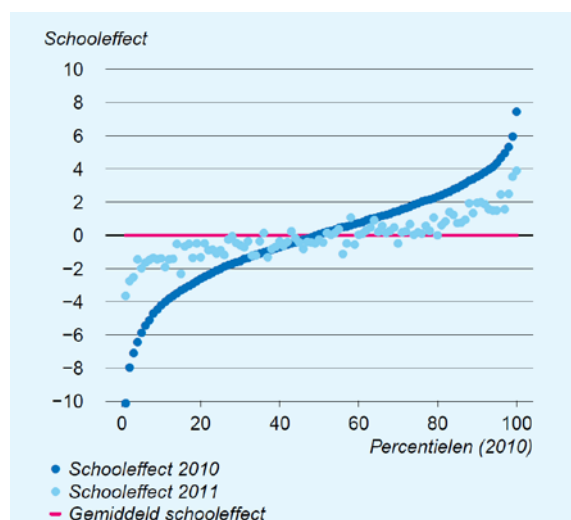
zijn en in de volgende jaren naar verwachting weer gelijk zijn aan die van scholen met een vergelijkbare leerlingpopulatie.

Scholen met de laagste schooleffecten in 2010 presteren gemiddeld een stuk beter in 2011.

Figuur 4.1 hieronder rangschikt scholen op basis van hun schooleffect in 2010. De donkerblauwe bolletjes – elk een groep van 1% (percentiel) van de scholen, ongeveer 35 scholen – tonen het gemiddelde schooleffect van het laagste percentiel tot het hoogste percentiel. De 1% scholen met de laagste schooleffecten in 2010 staan uiterst links en de 1% scholen met de hoogste schooleffecten uiterst rechts. De lichtblauwe bolletjes tonen het gemiddelde schooleffect van dezelfde groep scholen, maar dan in 2011. De scholen die in 2010 een schooleffect hadden in de laagste percentielen, hebben een hoger gemiddeld schooleffect in 2011. Er kunnen hier zowel scholen bij zitten die slechter als beter zijn gaan presteren, maar de scholen die beter zijn gaan presteren overheersen. Hierdoor is het gemiddelde schooleffect binnen deze groep scholen toegenomen.

Het tegenovergestelde geldt voor de scholen met de hoogste schooleffecten in 2010. Die zijn gemiddeld minder gaan presteren: het gemiddelde schooleffect binnen deze groep scholen is afgenomen. Uiteindelijk houden de hoogste percentielen van 2010 wel gemiddeld hogere schooleffecten dan de laagste percentielen van 2010, maar ze liggen dicht bij elkaar. Deze resultaten betekenen overigens niet dat de verschillen tussen scholen steeds kleiner worden: figuur 3.6 liet zien dat de spreiding van de schooleffecten in 2010 en 2016 zeer vergelijkbaar is. Wat deze resultaten wel aangeven is dat scholen met een hoog schooleffect volgend jaar een laag schooleffect kunnen hebben en vice versa. Daardoor blijven verschillen in leerresultaten tussen scholen bestaan, maar zijn het (deels) andere scholen die heel hoge of lage schooleffecten hebben.

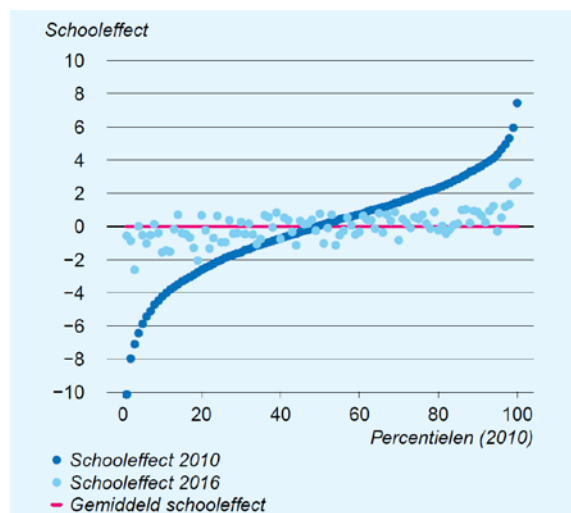
Figuur 4.1 Het schooleffect in 2010 werkt slechts beperkt door in het schooleffect in 2011



Over een langere periode komen scholen met de laagste en hoogste schooleffecten gemiddeld dicht bij elkaar. Een zelfde figuur kan gemaakt worden om te kijken hoe stabiel de schooleffecten over een langere periode zijn. Over een langere periode spelen niet alleen toevallige fluctuaties een rol. Scholen met lage leerresultaten, gegeven de achtergrondkenmerken van de leerlingpopulatie kunnen die resultaten aangrijpen om een verbeterslag te maken. Door over een langere periode te kijken worden ook deze effecten meegenomen. Figuur 4.2 is op dezelfde manier opgebouwd als

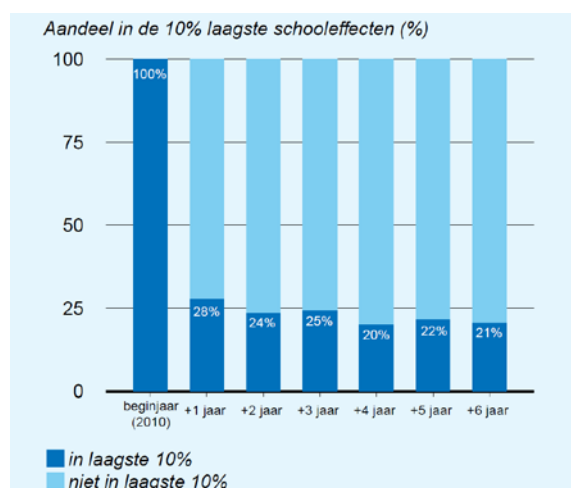
figuur 4.1, maar laat zien hoe stabiel de schooleffecten over 6 jaar zijn (2010 – 2016). Scholen zijn weer op volgorde van het schooleffect in 2010 gerangschikt en elke groep bestaat uit 1% van de scholen, ongeveer 35 scholen per groep. De 1% scholen met het laagste schooleffect in 2010 (uiterst links) en de 1% scholen met het hoogste schooleffect in 2010 (uiterst rechts) hebben in 2016 een gemiddeld schooleffect dat niet heel ver uit elkaar ligt.

Figuur 4.2 Het schooleffect in 2010 werkt slechts beperkt door in het schooleffect in 2016



Maar 28% van de scholen die in 2010 tot de groep met 10% laagste schooleffecten behoorde, behoort in 2011 weer tot die groep. Van de 10% scholen met de laagste schooleffecten in 2010 zit na één jaar slechts 28% opnieuw in het onderste deciel (zie figuur 4.3). Na 6 jaar is dat nog maar 21%. Aan de bovenkant van de verdeling gebeurt hetzelfde. Van de scholen met de 10% hoogste schooleffecten in 2010 blijft na één jaar slechts 36% in het bovenste deciel (zie figuur 4.4). Over een periode van 6 jaar neemt dit af tot net onder de 20%.

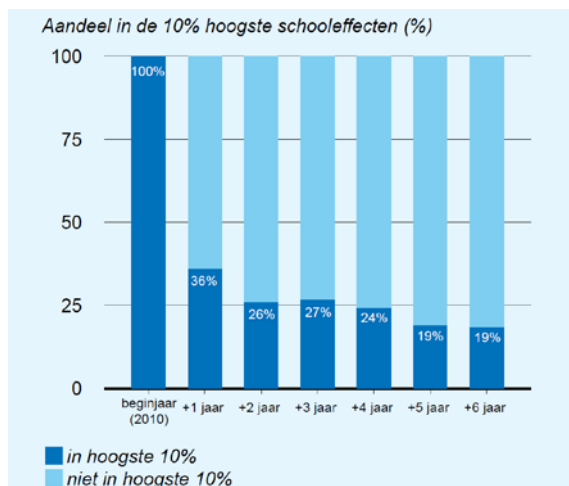
Figuur 4.3 Scholen behoren niet langdurig tot het onderste deciel (laagste 10% schooleffecten)



Opmerking: In deze figuur zijn alle scholen die in 2010 met de laagste 10% schooleffecten over tijd gevolgd. Na 1 jaar zit 28% van deze scholen weer in het laagste deciel. Als een school in een jaar bij de bovenste 90% hoort, maar het jaar daarop weer bij de laagste 10%, dan telt deze school wel mee in bovenstaande statistiek. De scholen hoeven dus niet alle jaren op rij tot de laagste 10% te behoren.

Het schooleffect in een bepaald jaar is geen betrouwbare indicator voor de kwaliteit van een school. Het schooleffect kent namelijk flinke fluctuaties in de tijd. Voor beleidsmakers en ouders is het veel relevanter hoe een school het gemiddeld over meerdere jaren doet, waarbij toevallige fluctuaties minder een rol spelen.

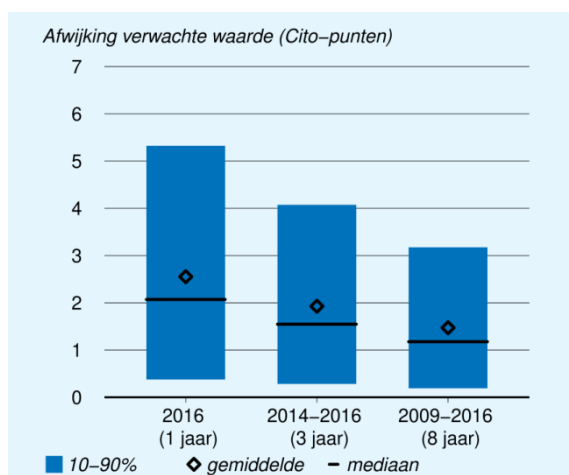
Figuur 4.4 Scholen behoren niet langdurig tot het bovenste deciel (hoogste 10% schooleffecten)



Opmerking: In deze figuur zijn alle scholen die in 2010 met de hoogste 10% schooleffecten over tijd gevolgd. Na 1 jaar zit 36% van deze scholen weer in het hoogste deciel. Als een school in een jaar bij de onderste 90% hoort, maar het jaar daarop weer bij de hoogste 10%, dan telt deze school wel mee in bovenstaande statistiek. De scholen hoeven dus niet alle jaren op rij tot de hoogste 10% te behoren.

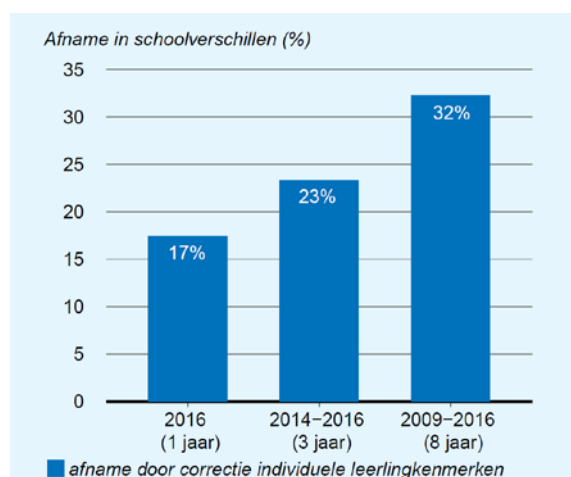
Door cohorten leerlingen samen te voegen en het schooleffect te berekenen over meerdere jaren, spelen uitschieters in een specifiek jaar een veel minder grote rol. Figuur 4.5 laat zien dat wanneer de schooleffecten (gecorrigeerd voor achtergrondkenmerken) over de periode 2009-2016 berekend worden, de verschillen tussen scholen veel kleiner worden. De linker staaf geeft de spreiding in schoolresultaten voor 2016 weer, net zoals in figuur 3.5. De rechterstaaf geeft de spreiding in schoolresultaten weer als het schooleffect berekend wordt over de periode 2009-2016. Het gemiddelde schooleffect neemt af van 2,6 naar 1,5. Vooral de grootte van uitschieters is afgenomen: in 2016 had 90% van de scholen een schooleffect kleiner dan 5,3, het schooleffect over de periode 2009-2016 is voor 90% van de scholen kleiner dan 3,2.

Figuur 4.5 Afwijking van verwachte waarde leerresultaat, gecorrigeerd voor achtergrondkenmerken, neemt af als meerdere jaren bijeen worden genomen

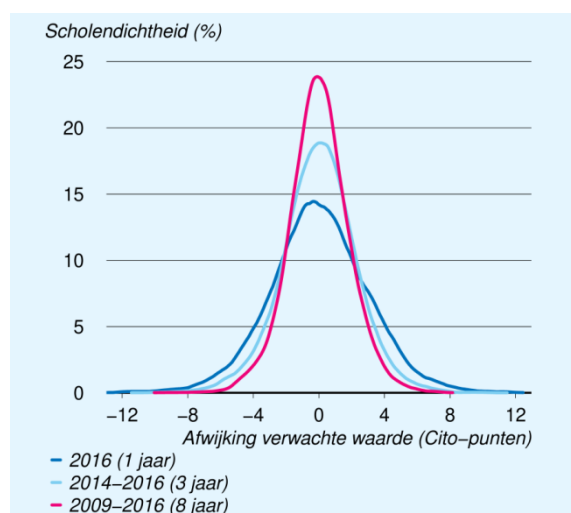


Wanneer 8 jaren samengevoegd worden, kan 32% van de verschillen in leerresultaten verklaard worden uit individuele achtergrondkenmerken. Wanneer 3 jaren worden samengevoegd is dat 23% (figuur 4.6). Omdat uitschieters een minder grote rol spelen wanneer meerdere jaren samengenomen worden, kan met achtergrondkenmerken een groter deel van de verschillen tussen scholen verklaard worden. Dit is ook terug te zien in de verdeling van de schooleffecten (figuur 4.7). Meer scholen liggen dicht bij de verwachte waarde als er meer jaren worden samengevoegd. Dit is te zien aan de toenemende scholendichtheid rond nul en afnemende dichtheid verder weg van de nul (verder dan ongeveer 3 cito-punten).

Figuur 4.6 Als meerdere jaren bijeen worden genomen zorgt corrigeren voor verschillen in achtergrondkenmerken voor een grotere afname totale van verschillen in leerresultaten



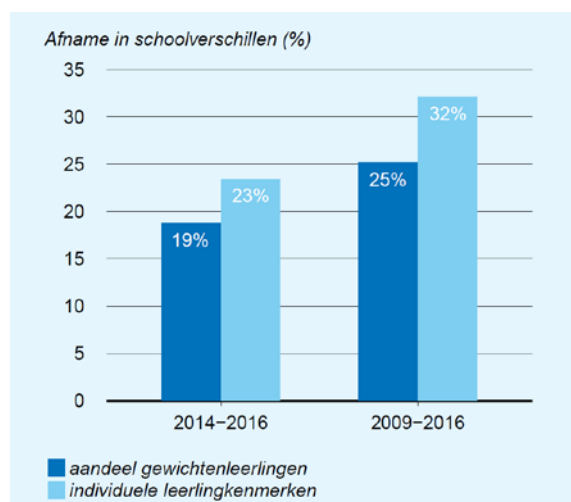
Figuur 4.7 Aandeel scholen dat dicht om de verwachte waarde van het leerresultaat ligt is groter als er meer jaren samen worden genomen



Verschillen in leerresultaten tussen scholen nemen met slechts 19% af wanneer 3 jaren worden samen genomen en alleen gecorrigeerd wordt voor aandeel gewichtenleerlingen Deze analyse is vergelijkbaar met de Staat van het Onderwijs 2015/2016. Dat is dus 4%-punt minder dan bij een correctie voor individuele leerlingkenmerken over dezelfde periode (figuur 4.8). Bij het samenvoegen van 8 jaren zorgt de correctie enkel op basis van leerlinggewicht voor een 25% afname. Ook dat is kleiner dan een vergelijkbare correctie op basis van individuele leerlingkenmerken. In dit

geval is het verschil is 7%-punt.

Figuur 4.8 Corrigeren voor individuele leerlingkenmerken leidt tot een grotere afname in schoolverschillen



5 Samenhang van schooleffecten met groeps- en schoolkenmerken

Deze laatste paragraaf onderzoekt de overige twee verklaringen voor de verschillen in leerresultaten tussen scholen: **groepseffecten en kwaliteit**. Deze paragraaf neemt als uitgangspunt de schoolverschillen na correctie voor achtergrondkenmerken over een periode van 8 leerjaren (2009 tot 2016). De verschillen in deze schooleffecten tussen scholen zijn weergegeven in Figuur 4.5. Paragraaf 5.1 onderzoekt hoe deze schooleffecten samenhangen met groepskenmerken en paragraaf 5.2 onderzoekt hoe deze schooleffecten samenhangen met schoolkenmerken. Deze samenhangen onderzoeken we zowel voor de totale groep scholen, als voor de subgroepen van scholen met de hoogste en laagste schooleffecten. In al deze analyses kunnen we geen oorzakelijke (ofwel causale) verbanden vaststellen, alleen correlaties.

Deze verbanden komen grotendeels overeen met de bevindingen uit wetenschappelijke literatuur. Zie voor een uitgebreid overzicht Kansrijk onderwijsbeleid (CPB, 2016). Details over de gebruikte data, schattingen en ruwe regressieresultaten staan in de appendix.⁹

⁹ Voor een overzicht van de categorische variabelen gebruikt in de schattingen, zie appendix 7.6. Voor de tabel met regressie-uitkomsten, zie appendix 7.7.

5.1 Groepskenmerken

Naast de kenmerken van de leerling zelf kunnen ook de kenmerken van de andere leerlingen op school invloed hebben op de leerresultaten van een leerling. We kijken naar twee groepskenmerken: de mate waarin de leerresultaten binnen een klas verschillen en de sociaaleconomische achtergrond van de leerlingpopulatie.¹⁰ Uitgangspunt zijn de schoolverschillen na correctie voor individuele achtergrondkenmerken van leerlingen, over een periode van 8 leerjaren.

Grotere verschillen in leerresultaten binnen een school hangen samen met een lager schooleffect. Scholen waar binnen groep 8 het leerresultaat van individuele leerlingen sterk uiteen loopt, hebben gemiddeld een lager schooleffect. Dit is consistent met de bestaande onderwijsliteratuur, waaruit blijkt dat betere leerresultaten worden behaald in groepen die meer homogeen zijn in (verwachte) leerprestatie (zie Kansrijk Onderwijsbeleid (CPB, 2016) voor een overzicht van de Nederlandse en internationale literatuur op dit gebied). De 20% scholen met de grootste spreiding hebben gemiddeld een 1 cito-punt lager schooleffect. Dit negatieve verband tussen de spreiding in leerresultaten en het schooleffect vinden we niet voor scholen met relatief lage leerresultaten, alleen voor scholen met relatief goede leerresultaten.

Minder medeleerlingen met laagopgeleide ouders hangt samen met een hoger schooleffect. Scholen met een relatief hoog gemiddeld leerlinggewicht hebben over het algemeen een lager schooleffect. Dit verband is vooral sterk bij scholen met relatief lage leerresultaten. Het gemiddelde leerlinggewicht is een indicator voor het opleidingsniveau van de ouders, omdat alleen leerlingen van ouders met maximaal vmbo-basis/kader (of vergelijkbaar) een gewicht groter dan nul krijgen toegekend. Omdat de schooleffecten al gecorrigeerd zijn voor individuele kenmerken, suggereert dit dat in een klas zitten met relatief veel medeleerlingen met laagopgeleide ouders veelal samengaat met een lager schooleffect. Omdat het een correlatie betreft is niet duidelijk hoe de causaliteit loopt. Het verschil tussen de 20% scholen met het hoogste leerlinggewicht en scholen zonder gewichtenleerlingen is 1 citopunt op het schooleffect.

Meer medeleerlingen uit gezinnen met een hoger inkomen hangt samen met een hoger schooleffect. Scholen met relatief veel leerlingen uit gezinnen met een hoger inkomen hebben gemiddeld een hoger schooleffect. Hierbij geldt dat scholen met leerlingen waarvan het gemiddeld huishoudinkomen boven de € 60.000 ligt een schooleffect hebben dat 1 tot 1,5 cito-punten hoger ligt dan scholen met een gemiddeld huishoudinkomen onder de € 40.000.

Geen samenhang tussen schooleffect en aandeel medeleerlingen met migratieachtergrond. We vinden geen eenduidige associatie tussen het aandeel leerlingen met een migratieachtergrond en het schooleffect, wanneer ook gecorrigeerd wordt voor opleidingsniveau en inkomen van de ouders van alle leerlingen op school.

¹⁰ Om de associaties tussen het schooleffect en de verschillen groeps- en schoolkenmerken vast te stellen hebben we rekening gehouden met het feit dat de schooleffecten niet exact gemeten worden. Omdat het schattingsresultaten zijn, zit er een foutmarge omheen. Om daarvoor te corrigeren voeren we een *Feasible Generalised Least Squares*-schatting (FGLS-schatting) uit. Voor meer informatie over deze methode, zie appendix 7.2.

5.2 Schoolkenmerken

Deze paragraaf onderzoekt met welke schoolkenmerken het resterende schooleffect samenhangt. Als we goed in staat zijn te corrigeren voor verschillen in samenstelling van de leerlingpopulatie, voor toevallige uitschieters en voor groepseffecten, blijft kwaliteit over als verklaring voor verschillen in leerresultaten tussen scholen. Een uitgebreide beschrijving van de analyses staat wederom in de appendix (7.3).

Meer ervaring van leraren hangt samen met hoger schooleffect. De kwaliteit van het onderwijsgevend personeel is van groot belang voor de leerprestaties (zie Kansrijk onderwijsbeleid (CPB, 2016) en Effectieve interventies leerachterstanden in het primair onderwijs (CPB, 2018)). Omdat daarnaast bekend is dat ervaring in belangrijke mate bijdraagt aan de kwaliteit van docenten (leerlingprestaties zijn hoger als leerlingen les hebben van ervaren docenten (Gerritsen, Plug, & Webbink, 2014)), onderzoeken we de samenhang van het schooleffect met de gemiddelde leeftijd van het onderwijsgevend personeel¹¹, waarbij leeftijd als indicator dient voor de ervaring van een docent. Scholen met meer ervaren docenten blijken vaker een hoger schooleffect te hebben. Het verschil in schooleffecten tussen de scholen met gemiddeld het oudste (≥ 50 jaar) en jongste lesgevend personeel ($< 37,5$ jaar) bedraagt ongeveer 0,9 cito-punten.

Scholen met relatief veel nieuw personeel hebben iets lagere schooleffecten. Het aandeel nieuw onderwijsgevend personeel (gemiddeld per jaar) op een school hangt samen met een iets lager schooleffect: scholen met relatief veel nieuw personeel hebben een 0,4 cito-punten lager schooleffect. Een hoog aandeel nieuw onderwijsgevend personeel kan duiden op veel verloop onder het onderwijsgevend personeel, maar ook op uitbreiding van de school door een groeiend leerlingenaantal. Een andere indicator voor verloop is het aandeel onderwijsgevend personeel met een tijdelijke aanstelling. Hiervan vinden we geen verband met het schooleffect.

Scholen met heel weinig leerlingen per docent hebben gemiddeld een lager schooleffect. Een heel laag aantal leerlingen per fte onderwijsgevend personeel hangt samen met een lager schooleffect, vooral bij scholen met relatief lage leerresultaten. Scholen met minder dan 12 leerlingen per fte onderwijsgevend personeel hebben gemiddeld een 0,4-0,6 cito-punten lager leerresultaat. Ook een heel hoge groepsgrootte (28+ leerlingen) hangt samen met een -0,4 cito-punten hoger leerresultaat. Dit lijkt contra-intuïtief, omdat uit de literatuur bekend is dat (substantieel) kleinere klassen een positief effect hebben op leerprestaties (CPB 2016, CPB 2018). Extra onderwijsgevend personeel zal echter vaak betaald worden uit extra middelen die een school ter beschikking staan vanuit de gewichtenregeling, en voor de invoering van passend onderwijs ook de leerlinggebonden financiering. Hierdoor hebben scholen met een laag aantal leerlingen per fte onderwijsgevend personeel en kleine groepsgrootte relatief vaker een moeilijke leerlingpopulatie.

Geen eenduidige samenhang schooleffect met schoolgrootte en stedelijkheid. Het schooleffect is groter voor kleine scholen (< 100 leerlingen), maar alleen voor de subset van scholen met relatief

¹¹ Alle gemiddelden die betrekking hebben op het onderwijsgevend personeel zijn gewogen naar het aantal fte.

hoge schooleffecten. Voor de andere subset vinden we juist een positieve associatie als de scholen zeer groot zijn (> 300 leerlingen). Omdat de samenhang tegengesteld is zien we in de volledige steekproef geen significante associatie. Met de overige schoolkenmerken vinden we ook geen eenduidig significante associaties, zoals verstedelijking en functiemix.

6 Referenties

Centraal Bureau voor de Statistiek, 2016, Herziening gewichtenregeling primair onderwijs-Fase I. Den Haag.

CPB, 2016, *Kansrijk onderwijsbeleid*, Den Haag: Centraal Planbureau.

CPB, 2018, Effectieve interventies leerachterstanden in het primair onderwijs. Den Haag: Centraal Planbureau.

Gerritsen, S., E. Plug en D. Webbink, 2014, Teacher quality and student achievement. Evidence from a Dutch sample of twins, Den Haag: Centraal Planbureau.

Inspectie van het Onderwijs, 2017, De staat van het onderwijs 2015/2016, Den Haag: ministerie van Onderwijs, Cultuur en Wetenschap.

Ost, B., 2014, How Do Teachers Improve? The Relative Importance of Specific and General Human Capital, *American Economic Journal: Applied Economics*, vol. 6(2): 127-151.

7 Appendix

7.1 Data

7.1.1 Individuele leerlingkenmerken

Om onze analyse uit te voeren moeten we een prestatiemaat en set van leerlingkenmerken kiezen. Als prestatiemaat nemen we de CITO-score van groep 8, zoals dat ook is gedaan in de Staat van het Onderwijs 2015/2016 (Inspectie van het Onderwijs, 2017) en in de herziening van de gewichtenregeling (Centraal Bureau voor de Statistiek, 2016). We nemen de volgende individuele leerlingkenmerken mee in onze analyse:

1. Opleiding moeder & vader.
2. Migratieachtergrond.
3. Verblijfsduur moeder.
4. Ouder(s) in schuldsanering.
5. Huishouden: alleenstaand of paar.
6. Belangrijkste inkomensbron huishouden.
7. Besteedbaar inkomen huishouden.
8. Testscore in groep 5 (NSCCT).

De laatste variabele is echter niet voor elk jaar beschikbaar. We moeten onze analyse dus opsplitsen in twee onderdelen: met en zonder deze maatstaf. Zonder de intelligentiemaatstaf kunnen we de schoolprestaties schatten van 2009 t/m 2016. Als we de analyse uitvoeren mét de intelligentiemaatstaf, dan hebben we alleen data beschikbaar voor 2011 en 2014.

Het opleidingsniveau van de moeder en vader is voor een aanzienlijk deel van de leerlingpopulatie onbekend. Aan het begin van de dataset (2009) is van 55% van de vaders en 48% van de moeders de opleiding onbekend. Aan het eind (2016) is dat 42% en 35%. Dit komt doordat centrale opleidingsregisters in Nederland pas recent worden bijgehouden. Het opleidingsniveau is vaker onbekend voor lager opgeleiden, omdat het hoger onderwijs als eerste is begonnen met registreren (in de jaren 80). Registers over voortgezet onderwijs kwamen beschikbaar vanaf eind jaren 90, over het mbo vanaf schooljaar 2004/'05 en een centraal register van het basisonderwijs is er pas sinds 2010. Personen met een hoog opleidingsniveau zijn daardoor oververtegenwoordigd. Hierdoor is de categorie 'opleiding onbekend' een selectieve categorie: de groep bevat relatief veel lager opgeleiden. Daarom nemen we deze categorie op als een apart opleidingsniveau in onze schattingen. Daarmee nemen we de ouders, van wie de opleiding onbekend is op in de analyse.

Overzicht van de dataset: aantal leerlingen in de steekproef

Jaar	Volledig sample	Cito-score bekend (elk jaar)	Alle leerlingkenmerken bekend (behalve opleiding ouders)	Ook NSCCT-score bekend
2009	165869	91000	79356	0
2010	180121	95936	82866	0
2011	186846	100685	88549	3813
2012	189386	104823	91591	0
2013	190351	109999	95344	0
2014	188070	109356	94028	4177
2015	186255	107809	92029	0
2016	181945	102209	86433	0

Overzicht van de dataset: aantal scholen in de steekproef

Jaar	Volledig sample	Cito-score bekend (elk jaar)	Alle leerlingkenmerken bekend (behalve opleiding ouders)	Ook NSCCT-score bekend
2009	6461	3255	3142	0
2010	6749	3348	3213	0
2011	6748	3396	3298	184
2012	6691	3433	3332	0
2013	6603	3482	3369	0
2014	6503	3472	3335	190
2015	6358	3420	3286	0
2016	6299	3327	3183	0

7.1.2 Kenmerken op schoolniveau

Daarnaast zijn we geïnteresseerd in de samenhang tussen de prestatiemaat en school- en personeelskenmerken. De informatie over lerarenkenmerken aggregeren we naar schoolniveau, omdat we niet weten elke leraren en leerlingen een klas vormen. We weten alleen op welke school het personeel werkzaam is en welke leerlingen daar naar school gaan.

De variabelen die niet in de CBS-RA-omgeving beschikbaar zijn, zijn toegevoegd uit het functiemixbestand van DUO. De variabelen groeps- en schoolkenmerken die we in deze analyse gebruiken zijn:

Groepskenmerken

- Spreiding in leerresultaten binnen groep 8
 - o Gemeten als spreiding (60ste – 40ste percentiel) in het gestandaardiseerd aantal juiste antwoorden op de Cito-toets als indicator voor de spreiding in niveau van de leerlingen in groep 8

- o Het aantal juist antwoorden op de Cito-toets is gestandaardiseerd met een Box-Cox-transformatie om de *skewness* te verminderen
- o Gestandaardiseerd aantal juiste antwoorden $= \frac{(\text{Aantal juiste antwoorden})^{L-1}}{L}$, met $L \approx 2,014$
- Aandeel leerlingen met een migratieachtergrond in groep 8
- Gemiddeld leerlinggewicht van de school
 - o als indicator voor het opleidingsniveau van de ouders over de hele school gemeten.
- Gemiddeld huishoudinkomen van de school
 - o Op basis van het huishouden waar de leerling in woont

Schoolkenmerken:

- Gemiddelde leeftijd van het onderwijsgevend personeel op de school
 - o Als maatstaf voor ervaring van het onderwijsgevend personeel. Gewogen naar fte.
- Het aandeel tijdelijke contracten van het onderwijsgevend personeel
 - o Gewogen naar fte
- Het aandeel nieuw onderwijsgevend personeel
 - o Gewogen naar fte
- Aantal leerlingen per fte onderwijsgevend personeel
- Aantal leerlingen in groep 8
 - o Groepsgrootte
- Totaal aantal leerlingen op de school
 - o Schoolgrootte
- Aantal inwoners in de gemeente waar de hoofdlocatie van de school staat
 - o Maatstaf voor verstedelijking
- Aandeel directie, onderwijsgevend personeel, ondersteunend personeel en leraren in opleiding
 - o Functiemix

7.2 Methode voor het schatten van de schooleffecten

Het doel van onze analyse is het maken van een maatstaf voor de schoolprestatie ten opzichte van wat op basis van omgevingskenmerken en intelligentie van de leerlingen kan worden verwacht. Deze maatstaf wordt vervolgens voor meerdere jaren berekend (van 2009 t/m 2016), zodat we de schoolprestatie over tijd kunnen volgen.

We verklaren de prestatiemaat (Y) van leerling i op school s met een set verklarende leerlingkenmerken ($X_{s,i}$). Dit doen we door een fixed-effects-model te schatten:

$$(1) \quad Y_{s,i} = \beta_0 + \beta_k X_{k,s,i} + \mu_s + \varepsilon_{s,i}$$

Hierbij is β_0 het algemene gemiddelde van de prestatiemaat. $X_{k,s,i}$ is een vector met leerlingkenmerken, die de prestatiemaat beïnvloeden. Het effect van deze leerlingkenmerken is in dit model onafhankelijk van de school waar een leerling naartoe gaat. De deviatie van de scholen

van het algemene gemiddelde wordt gemeten in μ_s (*School fixed effect*). Dit is dus onze maatstaf: de schoolprestatie gecorrigeerd voor alle leerlingkenmerken in ons model, ten opzichte van het gemiddelde over de hele steekproef. Door dit model voor elk jaar te schatten krijgen we een beeld van de schoolprestatie over tijd.

De opleiding van moeder en vader, herkomst, verblijfsduur moeder, ouder(s) in schuldsanering, huishoudtype en belangrijkste inkomensbron zijn allemaal opgenomen als categorische variabelen. Het natuurlijk logaritme van huishoudinkomen en NSSCT-score zijn beide opgenomen als continue variabelen.

7.3 Methode voor het schatten van de samenhang tussen het schooleffect en groeps- en schoolkenmerken

De volgende stap is het in kaart brengen van de samenhang tussen schoolprestaties en school- en personeelskenmerken. We kunnen leraren aan scholen koppelen, maar niet aan leerlingen (klassen). We hebben dus twee niveaus van waarnemingen: leerlingen en scholen. De samenhang tussen de schoolkenmerken en schoolprestatie kunnen we meten door de schooleffecten uit vergelijking (1) te verklaren met de schoolkenmerken:

$$(2) \quad \mu_s = \beta_0 + \beta_j X_{j,s} + \epsilon_s$$

Met $Var(\epsilon_s) = \sigma$.

$X_{j,s}$ is een vector met schoolkenmerken en β_j meet de samenhang daarvan met de prestatie maat. Door deze schoolkenmerken op te nemen in het model wordt een deel van de *School fixed effect* verklaard. Alle variabelen, met uitzondering van het aandeel per functiemix, zijn omgezet naar categorische variabelen. Daardoor nemen we vooraf niet aan dat de verbanden lineair zijn. Zie de tabel met resultaten in paragraaf 7.5 voor een overzicht van de categorieën die we hebben opgenomen.

In deze analyse moeten we rekening houden met het feit dat de schooleffecten schattingsresultaten zijn en dus niet exact gemeten worden (er zit een foutmarge omheen). We observeren slechts de het geschatte resultaat van het schooleffect μ_s^* :

$$(3) \quad \mu_s^* = \mu_s + \rho_s$$

Met $E(\rho_s) = 0$ en $Var(\mu_s) = \omega_s$. In feite schatten we dus:

$$(4) \quad \mu_s = \beta_0 + \beta_j X_{j,s} + \rho_s + \epsilon_s$$

Als we $\sigma + \omega_s$ kennen, dan kunnen we vergelijking (4) schatten met *Weighted Least Squares* (WLS) uit te voeren. De gewichten zijn dan:

$$w_s = \frac{1}{\sqrt{\sigma^2 + \omega_s^2}}$$

Doorgaans kennen we σ^2 echter niet vooraf. Daarom kunnen we deze schatting niet simpelweg uitvoeren. We kunnen dan toch WLS uitvoeren met de gewichten $w_s = \frac{1}{\omega_s}$, maar dan nemen we dus impliciet aan dat $\sigma^2 = 0$. Dat impliceert weer dat vergelijking (1) perfect moet passen (*overall* $R^2 = 1$) en dat is bij ons bij lange na niet het geval.

Een andere oplossing is schatten met *Ordinary Least Squares*, maar dan moeten we aannemen dat ω_s constant is over alle observaties (scholen). Maar ook dat is vaak niet het geval, waardoor de OLS-schatting inconsistent is. Dit kunnen we corrigeren door te schatten met robuuste standaardfouten. De OLS-schatting blijft dan echter inefficiënt.

Om efficiënt en correct te corrigeren voeren we een *Feasible Generalised Least Squares*-schatting (FGLS-schatting) uit. In feite een pragmatische manier om de WLS-regressie uit te voeren. Het FGLS-model, geïntroduceerd in Hanusek (1974), maakt gebruik van het feit dat ω_i bekend is (de variantie van het schooleffect). Door dat aan te nemen kunnen we σ^2 schatten ($\hat{\sigma}^2$), door eerst vergelijking (4) te schatten zonder gewichten.

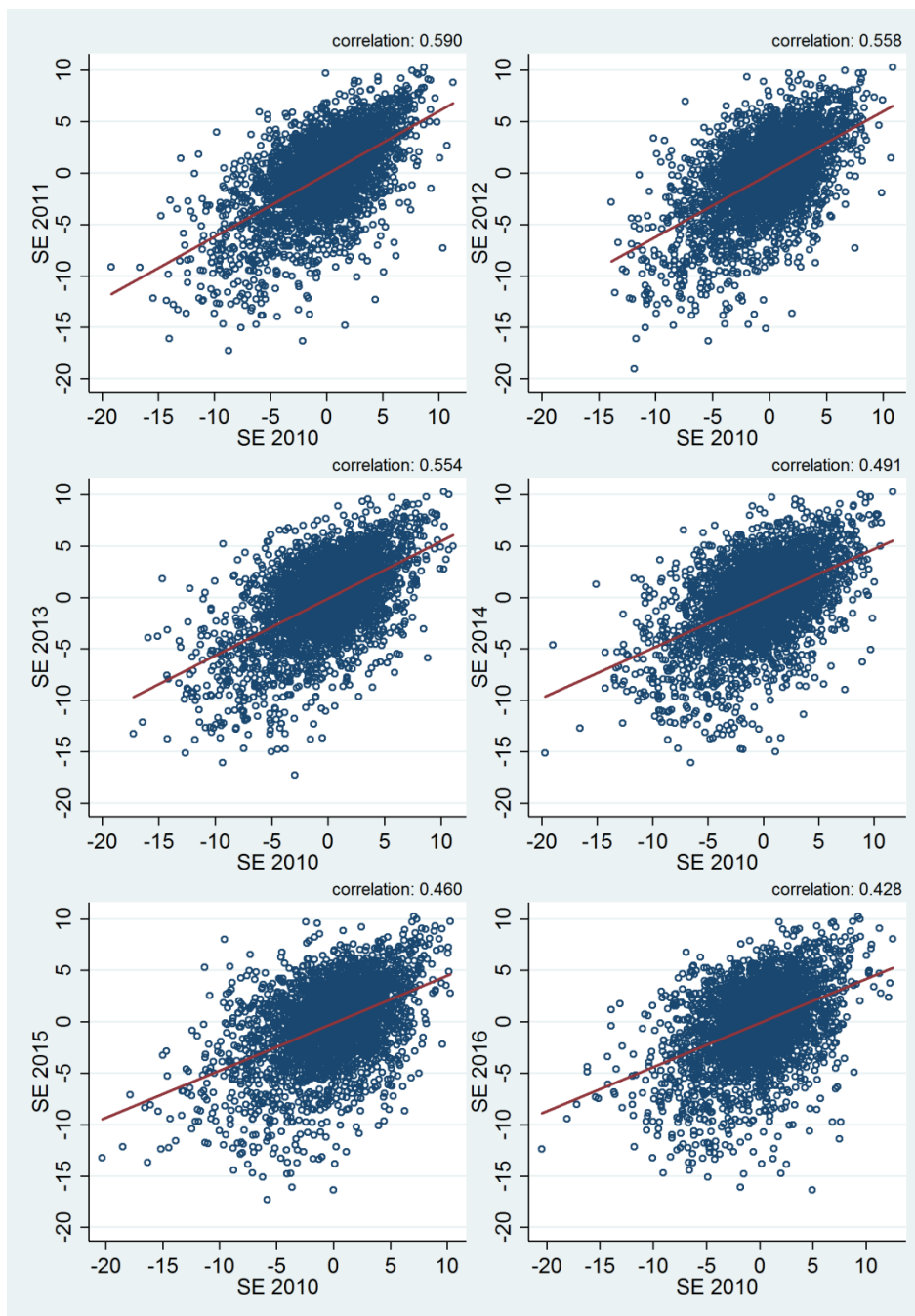
Daarna kunnen we de vergelijking nogmaals – correct – schatten met de gewichten gelijk aan:

$$w_s = \frac{1}{\sqrt{\hat{\sigma}^2 + \omega_s^2}}$$

De resultaten van deze FGLS-schatting staan in de tabel met resultaten in paragraaf 7.5.

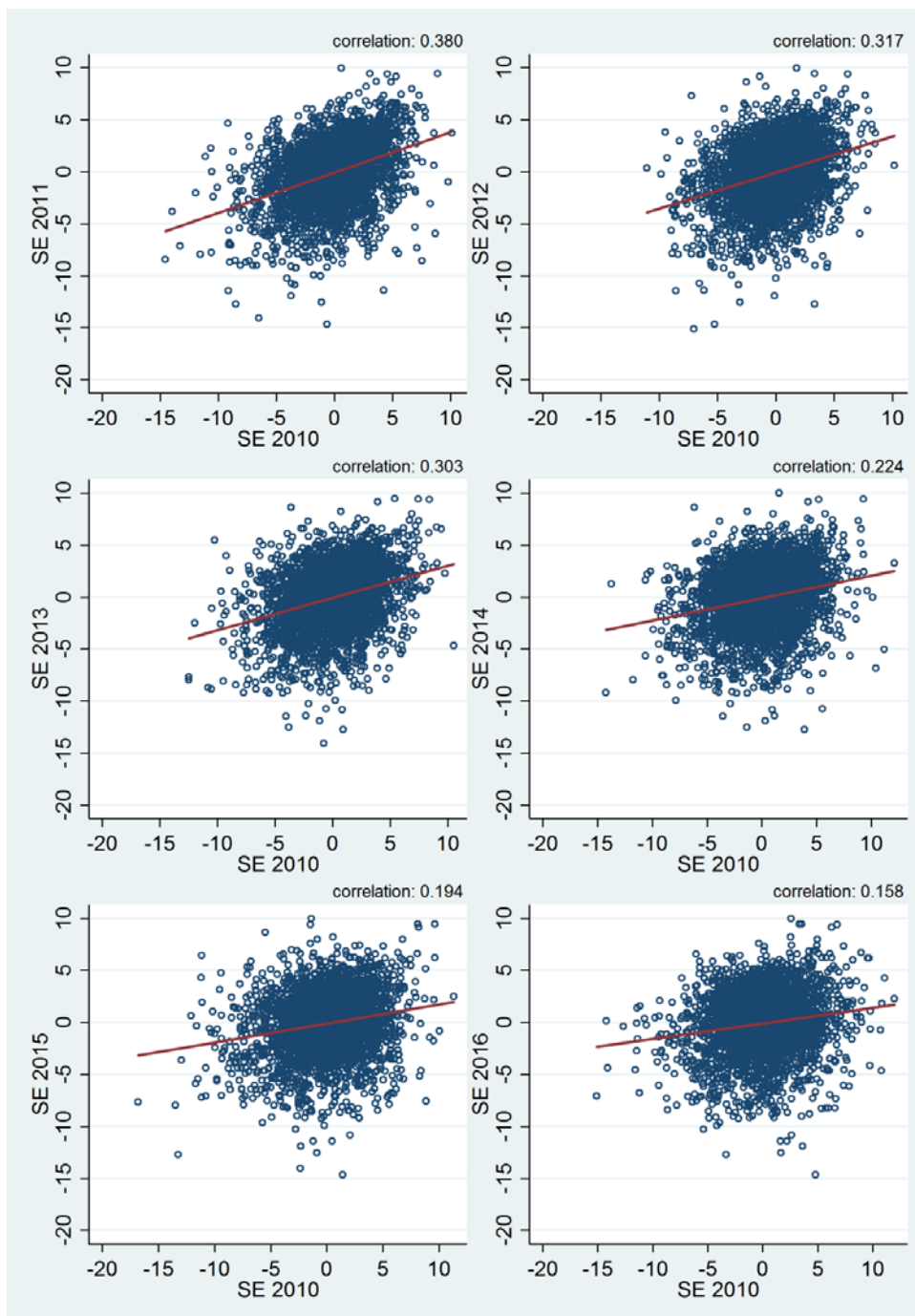
7.4 Scatterplots zonder controle voor individuele leerlingkenmerken

Zonder correctie voor individuele leerlingkenmerken is de correlatie tussen schooleffecten over tijd aanzienlijk. De correlatie tussen het schooleffect van 2010 en 2011 bedraagt 0,590 (*scatterplot* linksboven). Na verloop van tijd wordt deze correlatie steeds zwakker: de correlatie tussen het schooleffect van 2010 en 2016 bedraagt nog 0,428 (*scatterplot* rechtsonder).



7.5 Scatterplots met controle voor individuele leerlingkenmerken

Na correctie voor leerlingkenmerken wordt de correlatie tussen schooleffecten over tijd kleiner. De correlatie tussen het schooleffect van 2010 en 2011 bedraagt 0,380 (*scatterplot* linksboven). Na verloop van tijd wordt deze correlatie steeds zwakker: de correlatie tussen het schooleffect van 2010 en 2016 bedraagt nog 0,158 (*scatterplot* rechtsonder).



7.6 Categorijsche variabelen in FGLS-schatting

Schoolgrootte								
Aantal leerlingen op de school	< 100	100-200	200-300	300-400	400-500	>= 500		Totaal
Aantal scholen	421	1114	1123	577	284	205		3724
% scholen	11,3%	29,9%	30,2%	15,5%	7,6%	5,5%		100%
Groepsgrootte								
Aantal leerlingen in de groep waar de leerling in zit	< 12	12-16	16-20	20-24	24-28	>= 28		Totaal
Aantal scholen	283	338	523	966	1303	311		3724
% scholen	7,6%	9,1%	14,0%	25,9%	35,0%	8,4%		100%
Leerlingen per FTE onderwijsgevend personeel								
Aantal leerlingen per FTE	< 12	12-14	14-16	16-18	>= 18			Totaal
Aantal scholen	511	470	940	1436	367			3724
% scholen	13,7%	12,6%	25,2%	38,6%	9,9%			100%
Percentage tijdelijk								
Percentage tijdelijk onderwijsgevend personeel	< 1,0%	1,0-2,5%	2,5-5,0%	5,0%-7,5%	7,5-10,0%	10,0-15,0%	>= 15,0%	Totaal
Aantal scholen	414	757	1064	628	358	322	181	3724
% scholen	11,1%	20,3%	28,6%	16,9%	9,6%	8,6%	4,9%	100%
Gemiddelde leeftijd								
Gemiddelde leeftijd onderwijsgevend personeel	< 37,5	37,5-40	40-42,5	42,5-45	45-47,5	47,5-50	>= 50	Totaal
Aantal scholen	190	348	663	989	892	504	138	3724
% scholen	5,1%	9,3%	17,8%	26,6%	24,0%	13,5%	3,7%	100%
Aandeel nieuw onderwijsgevend personeel								
Aandeel nieuw onderwijsgevend personeel aan het begin van het schooljaar	< 2,5%	2,5-5,0%	5,0%-7,5%	7,5%-10,0%	10,0-12,5%	12,5%-15,0%	>= 15,0%	Totaal
Aantal scholen	521	1175	962	516	273	146	131	3724
% scholen	14,0%	31,6%	25,8%	13,9%	7,3%	3,9%	3,5%	100%
Verstedelijking								
Aantal inwoners in gemeente	< 20.000	20.000-50.000	50.000-100.000	100.000 - 150.000	150.000 - 250.000	>= 250.000		Totaal
Aantal scholen	460	1587	605	305	261	435		3653
% scholen	12,6%	43,4%	16,6%	8,3%	7,1%	11,9%		100%
Gemiddeld leerlinggewicht								
Gemiddeld leerlinggewicht	1-1,01	1,01-1,025	1,025-1,05	1,05-1,10	1,10-1,15	>= 1,15		Totaal
Aantal scholen	498	755	896	748	253	574		3724
% scholen	13,4%	20,3%	24,1%	20,1%	6,8%	15,4%		100%
Migratieachtergrond								
Aandeel leerlingen groep 8 met een migratieachtergrond	< 2,5%	2,5-5,0%	5,0-10,0%	10,0-20,0%	20,0-30,0%	30,0-40,0%	>= 40,0%	Totaal
Aantal scholen	288	368	778	1003	486	211	590	3724
% scholen	7,7%	9,9%	20,9%	26,9%	13,1%	5,7%	15,8%	100%
Spreadingsmaat								
Verschil tussen p60 en p40 in aantal goede antwoorden op de Cito-toets (getransformeerd)	< 1500	1500-1750	1750-2000	2000-2250	2250-2500	2500-3000	>= 3000	Totaal
Aantal scholen	154	319	681	953	847	669	101	3724
% scholen	4,1%	8,6%	18,3%	25,6%	22,7%	18,0%	2,7%	100%
Gemiddeld huishoudinkomen								
Gemiddeld huishoudinkomen ouders	< 40.000	40.000-45.000	45.000-50.000	50.000-55.000	55.000-60.000	60.000-65.000	>= 65.000	Totaal
Aantal scholen	604	485	792	861	462	251	269	3724
% scholen	16,2%	13,0%	21,3%	23,1%	12,4%	6,7%	7,2%	100%

7.7 Tabel uitkomsten Samenhang Schooleffecten en school-/leraarkenmerken

		I	II	III
		<i>Totale steekproef</i>	<i>Steekproef: alleen 50% hoogste schooleffecten</i>	<i>Steekproef: alleen 50% laagste schooleffecten</i>
	<1500	0	0	0
Spreidingsmaat aantal goede antwoorden op de Cito-toets (getransformeerd)	1500-1750	-0.239 (-1.00)	-0.507** (-2.46)	0.263 (1.09)
	1750-2000	-0.547** (-2.47)	-0.693*** (-3.62)	0.282 (1.26)
	2000-2250	-0.707*** (-3.20)	-0.760*** (-3.98)	0.274 (1.23)
	2250-2500	-0.783*** (-3.54)	-0.889*** (-4.64)	0.339 (1.52)
	2500-3000	-0.906*** (-4.06)	-0.980*** (-5.11)	0.218 (0.97)
	1,00-1,01	0	0	0
Gemiddeld leerlinggewicht van de school	1,01-1,025	-0.0569 (-0.58)	0.0134 (0.15)	0.0531 (0.57)
	1,025-1,05	-0.235** (-2.36)	0.0418 (0.45)	-0.0804 (-0.88)
	1,05-1,10	-0.396*** (-3.50)	0.00922 (0.09)	-0.201** (-1.99)
	1,10-1,15	-0.915*** (-5.21)	-0.0892 (-0.51)	-0.495*** (-3.50)
	1,15+	-1.244*** (-6.09)	0.0407 (0.22)	-0.843*** (-4.76)
	<12	0	0	0
	<40.000	0	0	0
Gemiddeld huishoudinkomen ouders	40.000-45.000	0.309* (1.89)	0.204 (1.30)	0.0728 (0.51)
	45.000-50.000	0.278* (1.70)	0.0255 (0.17)	0.162 (1.14)
	50.000-55.000	0.363** (2.12)	0.136 (0.87)	0.278* (1.88)
	55.000-60.000	0.400** (2.22)	-0.00720 (-0.05)	0.231 (1.41)
	60.000-65.000	0.588*** (2.96)	0.250 (1.39)	0.221 (1.26)
	65.000+	1.442*** (7.24)	0.749*** (4.12)	0.447** (2.46)
	0-2,5%	0	0	0
Aandeel leerlingen groep 8 met een migratieachtergrond	2,5-5,0%	-0.115 (-0.78)	-0.210* (-1.68)	-0.0611 (-0.46)
	5,0-10,0%	-0.0421 (-0.32)	-0.147 (-1.24)	0.0481 (0.42)
	10,0-20,0%	0.0666 (0.50)	0.0156 (0.13)	0.0514 (0.44)
	20,0-30,0%	0.174 (1.14)	0.0992 (0.67)	0.134 (1.01)
	30,0-40,0%	0.452** (2.36)	0.0756 (0.42)	0.425** (2.34)
	40,0%+	0.298 (1.37)	0.0540 (0.26)	0.215 (1.12)

	<37,5	0	0	0
Gemiddelde leeftijd onderwijsgevend personeel	37,5-40	-0.00525 (-0.03)	0.0790 (0.52)	-0.183 (-1.23)
	40-42,5	0.112 (0.73)	0.122 (0.88)	-0.188 (-1.31)
	42,5-45	0.144 (0.98)	0.123 (0.90)	-0.117 (-0.85)
	45-47,5	0.238 (1.54)	0.232 (1.59)	-0.240 (-1.64)
	47,5-50	0.458*** (2.79)	0.196 (1.32)	-0.00978 (-0.06)
	50+	0.920*** (4.00)	0.570*** (2.81)	-0.0923 (-0.44)
	<2,5%	0	0	0
Aandeel nieuw onderwijsgevend personeel aan het begin van het schooljaar	2,5-5,0%	-0.0717 (-0.78)	0.0959 (1.14)	0.117 (1.29)
	5,0%-7,5%	-0.182* (-1.88)	-0.0127 (-0.14)	0.0439 (0.46)
	7,5%-10,0%	-0.284** (-2.44)	-0.0128 (-0.12)	0.00302 (0.03)
	10,0-12,5%	-0.489*** (-3.34)	-0.110 (-0.80)	-0.0781 (-0.60)
	12,5%-15,0%	-0.377** (-2.06)	-0.0947 (-0.48)	0.164 (1.05)
	15,0%+	-0.426* (-1.93)	-0.145 (-0.87)	-0.149 (-0.66)
	<1,0%	0	0	0
Aandeel tijdelijk onderwijsgevend personeel	1,0-2,5%	-0.0983 (-0.91)	-0.131 (-1.31)	-0.0780 (-0.85)
	2,5-5,0%	-0.00767 (-0.07)	0.00806 (0.08)	-0.0823 (-0.94)
	5,0%-7,5%	-0.120 (-0.99)	-0.0526 (-0.45)	-0.140 (-1.38)
	7,5-10,0%	-0.0892 (-0.62)	0.139 (1.01)	-0.134 (-1.14)
	10,0-15,0%	-0.0657 (-0.46)	-0.0977 (-0.70)	0.00857 (0.07)
	15,0% +	-0.202 (-1.05)	-0.183 (-1.03)	-0.368** (-2.00)
	<12	0	0	0
Leerlingen per FTE onderwijsgevend personeel	12-14	0.512*** (3.47)	0.0848 (0.55)	0.137 (1.16)
	14-16	0.608*** (3.88)	0.116 (0.74)	0.296** (2.29)
	16-18	0.580*** (3.33)	0.0218 (0.13)	0.288** (2.00)
	18+	0.619*** (3.13)	0.119 (0.61)	0.389** (2.38)
	<12	0	0	0
Groepsgrootte	12-16	-0.215 (-1.12)	-0.0935 (-0.53)	-0.178 (-1.14)
	16-20	0.162 (0.84)	0.208 (1.07)	-0.129 (-0.79)
	20-24	0.208 (1.11)	0.0534 (0.29)	0.0534 (0.34)
	24-28	0.219 (1.13)	0.104 (0.55)	-0.00889 (-0.05)
	28+	0.430** (2.05)	0.195 (0.95)	0.102 (0.56)
	<100	0	0	0
Schoolgrootte	100-200	-0.348** (-2.08)	-0.511*** (-3.30)	0.163 (1.09)

	200-300	-0.289 (-1.54)	-0.457*** (-2.61)	0.264 (1.55)
	300-400	-0.303 (-1.54)	-0.592*** (-3.20)	0.326* (1.84)
	400-500	-0.174 (-0.82)	-0.671*** (-3.45)	0.392** (2.00)
	500+	-0.150 (-0.68)	-0.597*** (-2.94)	0.485** (2.50)
	<20.000	0	0	0
	20.000- 50.000	-0.119 (-1.22)	-0.226** (-2.34)	-0.0276 (-0.34)
Verstedelijking (aantal inwoners in gemeente)	50.000- 100.000	-0.130 (-1.11)	-0.209* (-1.83)	-0.0582 (-0.60)
	100.000 - 150.000	-0.350** (-2.47)	-0.324** (-2.35)	-0.209* (-1.75)
	150.000 - 250.000	-0.447*** (-2.85)	-0.192 (-1.29)	-0.326** (-2.40)
	>= 250.000	-0.214 (-1.29)	0.0540 (0.33)	-0.408*** (-2.67)
	Percentage FTE onderwijsgevend personeel	-0.0440 (-1.73)	-1.315* (-1.26)	-0.857 (-0.89)
	Percentage FTE directie	-0.134 (-0.17)	-0.211 (0.21)	0.227 (-0.33)
	Constante	-0.493 (1.72)	1.373* (4.21)	3.019*** (-2.11)
	R-sq	0.191	0.143	0.228
	N	3653	1834	1819

t statistics in parentheses

** p<0.1, ** p<0.05, *** p<0.01*