



Centraal Bureau
voor de Statistiek

Rapport

Herziening Gewichtenregeling voor het onderwijsachterstandenbeleid Primair Onderwijs

Martine de Mooij

Wouter van Andel

Bart Bakker

Ruben van Gaalen

Rosa Meijer

Magda Slootbeek

Mersiha Tepic

Caroline van Weert

CBS Den Haag
Henri Faasdreef 312
2492 JP Den Haag
Postbus 24500
2490 HA Den Haag
+31 70 337 38 00
www.cbs.nl

projectnummer

1 december 2015

kennisgeving De in dit rapport weergegeven opvattingen zijn die van de auteur(s) en komen niet noodzakelijk overeen met het beleid van het Centraal Bureau voor de Statistiek.

Inhoud

Samenvatting	4
1. Inleiding	6
2. Invloed omgeving op leerprestatie	8
2.1 Individuele kenmerken	8
2.2 Omgevingskenmerken	8
2.3 Meetbare indicatoren	11
3. Onderzoeksmethode	13
3.1 Analysemodel: samenhang onderwijssachterstand en gezins- en omgevingsfactoren	13
3.2 Van analysemodel naar voorspelmodel	19
3.3 Van voorspelmodel naar verdeelmodel	22
4. Risico's en voorwaarden	24
4.1 Risico's	24
4.2 Voorwaarden	25
Bijlage 1. Literatuur	27
Bijlage 2. Overzicht indicatoren	29

Samenvatting

Het onderwijsachterstandenbeleid (OAB) is erop gericht onderwijsachterstanden van kinderen uit zwakkere sociaaleconomische milieus te verminderen. De kern van het OAB voor het primair onderwijs is de zogenaamde gewichtenregeling. Op dit moment houdt deze regeling in dat basisscholen extra financiering krijgen voor leerlingen van wie de ouders een laag opleidingsniveau hebben. Dit opleidingsniveau wordt geregistreerd door de scholen zelf. Scholen die deze extra financiering ontvangen, krijgen daarnaast extra geld als ze in een postcodegebied liggen waar het gemiddelde inkomensniveau laag is en de werkloosheid hoog (de zogenaamde impulsgebieden).

Het ministerie van Onderwijs-, Cultuur en Wetenschap (OCW) heeft het Centraal Bureau voor de Statistiek (CBS) gevraagd te onderzoeken welke mogelijkheden er zijn om de gewichtenregeling te baseren op registraties die aanwezig zijn bij CBS. Het door de scholen geregistreerde opleidingsniveau van ouders blijkt namelijk lang niet altijd correct te zijn. Daarnaast kan zo worden onderzocht of er naast het opleidingsniveau van de ouders meer voorspellers zijn van onderwijsachterstanden die een belangrijke rol spelen. Tenslotte heeft OCW gevraagd in het onderzoek ook buurtkenmerken mee te nemen, zodat de huidige financiering op basis van impulsgebieden mogelijk ook meegenomen kan worden in een nieuwe gewichtenregeling.

Dit rapport bevat de resultaten van het vooronderzoek. In een verkennende literatuurstudie is bekeken wat de belangrijkste voorspellers zijn van leerprestaties. Vervolgens wordt beschreven hoe in een aantal stappen een nieuw model ontwikkeld kan worden dat kan worden gebruikt bij het verdelen van het OAB-budget. In de bijlage wordt een overzicht gegeven van de data die CBS kan gebruiken bij het ontwikkelen van een model.

In dit onderzoek wordt het als onderwijsachterstand beschouwd wanneer een leerling door ongunstige omgevingsfactoren slechter presteert op school dan op grond van het potentieel van het kind (de intelligentie) kan worden verwacht. Bij de omgevingsfactoren wordt een onderscheid gemaakt tussen kenmerken van het gezin (zoals opleidingsniveau van de ouders, inkomen en herkomst), van de buurt (zoals stedelijkheid en armoede) en van de school (zoals concentratie van leerlingen uit zwakkere milieus).

De voorgestelde methode om een nieuw verdeelmodel te ontwikkelen bestaat uit drie fases. In de eerste fase wordt onderzocht wat de samenhang is tussen verschillende individuele en omgevingskenmerken enerzijds en leerprestaties anderzijds door het schatten van een *analysemodel*. Daarbij wordt de intelligentie van het kind als kenmerk meegenomen zodat daadwerkelijk kan worden bepaald wat de omgevingskenmerken toevoegen/afdoen aan de verwachte leerprestatie op basis van de intelligentie van het kind. CBS kan dit analysemodel maken met gegevens uit het cohortonderzoek COOL⁵⁻¹⁸ (steekproefgegevens). Hierin is een meting opgenomen van leerprestaties aan het begin van het basisonderwijs (Cito-toets in groep 2), een meting van de intelligentie (Niet-Schoolse Cognitieve Capaciteiten Test in groep 5) en het opleidingsniveau van de ouders. Dit bestand wordt aangevuld met bij CBS beschikbare andere individuele en omgevingskenmerken. Gekozen is voor de leerprestaties in groep 2 omdat de prestaties op dat moment nog relatief weinig beïnvloed zijn door het OAB. Wel stelt CBS voor om ook analyses uit te voeren met de leerprestaties in groep 8 (Cito-toets) om meer inzicht te krijgen in de invloed van omgevingskenmerken aan het begin versus aan het einde van

de basisschool. Op basis van de uitkomsten kan dan een definitieve keuze worden gemaakt voor het meetmoment van de leerprestaties.

Het analysemodel wordt in drie stappen geschat. Als eerste wordt een model gemaakt waarin alle mogelijke invloedrijke kenmerken worden meegenomen. Daarna wordt dit model stap voor stap teruggebracht tot een model met daarin alleen de werkelijk relevante variabelen. Vervolgens wordt dit model vergeleken met een aantal door OCW voorgestelde, op voorhand interessante, modellen. Bepaald zal worden hoe goed de verschillende modellen bij de data passen en hoe goed onderwijsachterstanden ermee worden voorspeld. Deze informatie kan worden gebruikt om een keuze te maken voor een van de modellen.

Het door OCW gekozen model wordt in fase twee gebruikt als *voorspelmodel*. Door de meest recente integrale gegevens van CBS in te vullen in dit model, kan een voorspelling worden gemaakt van de onderwijsachterstanden per school. Van ruim 40 procent van de ouders van basisschoolleerlingen is bij CBS het opleidingsniveau niet bekend, terwijl deze variabele zeer waarschijnlijk als belangrijke voorspeller uit de bus zal komen. Er zijn twee mogelijke oplossingen om toch een voorspelling te kunnen doen voor deze kinderen: wegen of imputeren. Onderzocht kan worden welke methode in de praktijk het beste uitpakt; CBS verwacht dat dit imputatie zal zijn. Overigens zal in de toekomst van een steeds groter deel van de ouders het opleidingsniveau bekend zijn.

De in fase twee voorspelde achterstanden per school kunnen in de derde fase worden gebruikt om een *verdeemodel* op te stellen voor het toekennen van de OAB-gelden aan scholen. Enkele mogelijke varianten worden kort beschreven; de uiteindelijke keuze zal door OCW moeten worden gemaakt, in overleg met het onderwijsveld.

De beschreven methode om tot een verdeemodel te komen kent een aantal risico's. Zo is op dit moment niet bekend wat de kwaliteit is van de in het COOL⁵⁻¹⁸ onderzoek gemeten schoolprestaties en intelligentie van basisschoolleerlingen. Ook is het mogelijk dat niet voor alle scholen het aantal achterstandsleerlingen betrouwbaar genoeg kan worden voorspeld. Een mogelijke oplossing voor deze scholen is het maken van een voorspelling op schoolbestuurniveau.

Bij het ontwikkelen van een nieuw verdeemodel is het belangrijk vooraf vast te stellen welke verantwoordelijkheid CBS en het ministerie van OCW hebben. CBS kan op grond van empirisch onderzoek bepalen welke kenmerken van belang zijn voor onderwijsachterstanden en wat de omvang van die invloed is. Op basis hiervan is het aan het ministerie van OCW om te bepalen welke kenmerken daadwerkelijk gebruikt worden in het model. CBS kan op basis van dit besluit de geschatte leerachterstand per school berekenen. Het is vervolgens aan het ministerie van OCW om op basis van de geschatte leerachterstand per school budget aan de scholen toe te kennen.

Ten slotte stelt CBS voor om een klankbordgroep in het leven te roepen om ervoor te zorgen dat de methode voor een nieuw verdeemodel breed gedragen wordt binnen de onderwijs- en onderzoekswereld.

1. Inleiding

Nederland voert sinds de jaren zeventig beleid om de onderwijskansen voor kinderen uit achterstandsmilieus te vergroten. Het beleid is erop gericht leerachterstanden onder (basisschool)leerlingen ten gevolge van sociale, economische of culturele oorzaken zoveel mogelijk te voorkomen en om eenmaal opgelopen achterstanden te verminderen. Een belangrijk onderdeel van het onderwijsachterstandenbeleid vormt de zogenaamde “gewichtenregeling”. Deze gewichtenregeling bepaalt hoeveel geld een basisschool krijgt om onderwijsachterstanden weg te werken.

De gewichtenregeling is de afgelopen decennia een aantal keer van vorm veranderd. De indicator voor achterstanden was in eerste instantie het beroep van de vader. Het beleid was erop dat moment vooral op gericht autochtone “arbeiderskinderen” extra te stimuleren. Later werden ook de indicatoren “opleidingsniveau van de ouders” en “etnische herkomst” een onderdeel van de gewichtenregeling. De huidige gewichtenregeling is volledig gebaseerd op het opleidingsniveau van de ouders (Kloprugge en de Wit, 2015). Scholen die extra financiering ontvangen op basis van deze regeling, krijgen daarnaast extra geld als ze in een postcodegebied liggen waar het gemiddelde inkomensniveau laag is en de werkloosheid hoog (de zogenaamde impulsgebieden).

Op dit moment wordt het opleidingsniveau van de ouders door scholen zelf geregistreerd. Uit eerder onderzoek blijkt dat het geregistreerde opleidingsniveau lang niet altijd overeenkomt met het werkelijke opleidingsniveau van de ouders, wat ertoe leidt dat het beschikbare geld niet altijd bij de juiste scholen terecht komt.

Om de gewichtenregeling niet langer afhankelijk te maken van de registraties van de scholen, heeft het ministerie van Onderwijs, Cultuur en Wetenschap aan CBS gevraagd om te onderzoeken welke mogelijkheden er zijn om (een nieuwe variant van) de gewichtenregeling te baseren op registraties die aanwezig zijn bij CBS. Door de registratie van het opleidingsniveau buiten de scholen te leggen, wordt de vergelijkbaarheid van de gegevens verbeterd en levert dit daarnaast een administratieve lastenverlichting voor scholen op.

Ook als het opleidingsniveau goed geregistreerd zou worden, is er nog een tweede aandachtspunt bij de huidige regeling. De vraag die steeds vaker opspeelt is namelijk de vraag of de huidige regeling nog steeds de juiste regeling is om onderwijsachterstanden te bestrijden. Is opleidingsniveau van de ouders nog steeds de beste voorspeller van onderwijsachterstanden, of zijn er meer of andere voorspellers die ook een belangrijke rol spelen bij het tot stand komen van achterstanden? Ook wanneer opleidingsniveau nog steeds een belangrijke voorspeller blijkt te zijn, is het van belang om kritisch te kijken naar de definitie van de groep ouders met een ‘laag opleidingsniveau’, die bijvoorbeeld ook opgerekt zou kunnen worden naar “ouders zonder startkwalificatie”.

Tenslotte heeft OCW gevraagd in het onderzoek ook buurtkenmerken mee te nemen, zodat de huidige financiering op basis van impulsgebieden mogelijk ook meegenomen kan worden in een nieuwe gewichtenregeling

Het is goed vooraf vast te stellen welke verantwoordelijkheid CBS en het ministerie van OCW hebben. CBS kan op grond van empirisch onderzoek bepalen welke kenmerken van belang zijn

voor onderwijsachterstanden en wat de omvang van die invloed is. Op basis hiervan is het aan het ministerie van OCW om te bepalen welke kenmerken daadwerkelijk gebruikt worden. CBS kan op basis van dit besluit de geschatte leerachterstand per school berekenen. Het is vervolgens aan het ministerie van OCW om op basis van de geschatte leerachterstand per school budget aan de scholen toe te kennen.

In dit document wordt ingegaan op de mogelijkheden die CBS ziet bij het verbeteren van de huidige gewichtenregeling. In hoofdstuk 2 wordt een overzicht gegeven van mogelijk kansrijke voorspellers van leerprestatie die naar voren komen uit eerdere onderzoeken. In hoofdstuk 3 wordt vervolgens beschreven op welke manier deze variabelen in een statistisch model kunnen worden gebruikt om voor basisschoolleerlingen tot een voorspelling van onderwijsachterstand te komen en op welke manier deze voorspelling kan worden gebruikt om het onderwijsachterstandenbudget over scholen te verdelen. Voor alle genoemde indicatoren is in de bijlage aangegeven in hoeverre ze integraal beschikbaar zijn voor de doelgroep, wat de bruikbaarheid is en welke bijzonderheden er zijn die gevolgen kunnen hebben voor de toepasbaarheid.

2. Invloed omgeving op leerprestatie

Naar de leerprestaties van kinderen is al veel onderzoek gedaan. In dit hoofdstuk geven we een overzicht van eerdere onderzoeken naar de factoren die van invloed zijn op schoolprestaties. Het doel is om vast te stellen welke factoren de meest invloedrijke voorspellers zijn van leerprestaties bij kinderen. Het onderstaande vormt overigens geen uitputtende lijst, enkel de in de literatuur meest voorkomende factoren worden besproken.

Het is belangrijk om een duidelijk onderscheid te maken tussen onderwijsachterstand en leerprestatie. Onderwijsachterstand betekent in het OAB dat de schoolprestatie slechter is dan mogelijk was gegeven de individuele potentie. In de meeste onderzoeken wordt achterstand niet als zodanig gemeten, maar wordt van achterstand gesproken als een groep slechter gepresteert dan gemiddeld. Het zijn dus de leerprestaties die worden gemeten, bijvoorbeeld met de Cito toetsscore.

Leerprestaties van een kind worden op verschillende niveaus beïnvloed. Op het individuele niveau spelen aangeboren kenmerken een rol. Aan de andere kant zijn er allerlei invloeden waar een kind mee te maken heeft en die het kind in de loop der tijd beïnvloeden. Eén daarvan is het gezin waarin het kind opgroeit. Daarnaast kunnen buurtkenmerken een rol spelen. Zodra het kind naar school gaat, krijgen de schoolinvloeden ook een rol.

2.1 Individuele kenmerken

De schoolprestatie van een kind hangt voor een deel samen met het aangeboren intelligentieniveau van het kind. Onderzoek onder autochtone leerlingen laat zien dat 'niet-schoolse cognitieve capaciteiten' (een grove maat voor intelligentie) de grootste bijdrage aan de verklaring van leerprestatie leveren (Mulder et al., 2014).

Er zijn ook andere individuele kenmerken die een rol kunnen spelen. De leerprestatie is mogelijk mede-afhankelijk van het geslacht: meisjes doen het vaak beter op school dan jongens. Daarnaast kan de gezondheid en het sociaal-emotioneel functioneren van een leerling van invloed zijn op de schoolprestaties.

2.2 Omgevingskenmerken

2.2.1 Gezin

Naast de biologische kenmerken, zoals de intelligentie van het kind, zijn er tal van factoren die de leerprestatie van het kind beïnvloeden (Mulder et al., 2014). Het gezin is een van de belangrijkste invloedsferen van het kind en daarmee medebepalend voor de leerprestatie. In dit kader wordt er vaak gesproken over sociaaleconomische achtergrond of milieu van herkomst: dit geeft een indicatie van de economische en sociale positie van een individu of gezin in de maatschappij. De drie meest voorkomende componenten die in dit verband in de literatuur worden genoemd zijn: de opleiding van de ouders, het beroep of de beroepsstatus van de ouders en het inkomen van het gezin (Reynders et al., 2005). Deze indicatoren worden meestal gezamenlijk gehanteerd om de sociaaleconomische achtergrond weer te geven. Van deze drie blijken opleiding en beroep de meeste voorspellende waarde te hebben (Reynders et al., 2005).

De leerprestatie van een kind wordt bepaald door een combinatie van verschillende factoren (Mulder et al., 2014). Veel van deze factoren vormen onderdeel van gezinsomstandigheden. Problematische omstandigheden komen vaker voor in gezinnen met een lage sociaaleconomische status of in een achterstandsmilieu. Uit onderzoek naar autochtone leerlingen met laag opgeleide ouders blijkt dat deze leerlingen relatief vaak opgroeien in een taalarme omgeving, minder ouderlijke hulpbronnen hebben, vaker uit multi-probleem gezinnen komen, over minder niet-schoolse capaciteiten beschikken en minder vaak een voor- en vroegschoolse educatie(vve) instelling bezoeken. Ook gezinnen met één ouder of een groot aantal kinderen, psychische problemen van de opvoeder, huwelijksproblemen, een autoritaire opvoedingsstijl en ingrijpende gebeurtenissen als ontslag en ziekte van gezinsleden verhogen het risico tot laag presteren, zeker wanneer ze in combinatie met elkaar voorkomen (Leseman, 2007).

Deze kenmerken hebben te maken met de stabiliteit en capaciteit van de ouders om een stimulerende omgeving aan hun kind te bieden. Hierbij zijn de opvoedingsprincipes die een positieve invloed hebben op de cognitieve ontwikkeling en een betere voorbereiding op het basisonderwijs van belang. Een belangrijk principe hierbij is 'informeel leren', zoals taalspelletjes of klokkijken: dit verhoogt het kennisniveau van kinderen en stimuleert hun leren en denkvermogen (Mesman, 2011). Bij de stimulering is de kind-ouder interactie van belang. Bij het voorlezen gaat het dan bijvoorbeeld om napraten over het voorgelezen verhaal of samen met het kind reflecteren op bijvoorbeeld de functie van bepaalde lettertekens (Vedder & Bekkers, 1993). De verschillen in informeel leren in gezinnen met een andere sociaaleconomische achtergrond leiden tot grote verschillen in het cognitieve functioneren op vierjarige leeftijd (Meijnen, 2003). Opvoeders in middenklassenmilieus blijken hun kinderen meer te stimuleren tot abstract en symbolisch denken dan opvoeders in de lagere sociale milieus, die vaker concrete opdrachten geven of handelingen zelf uitvoeren. Daarnaast geven opvoeders uit de middenklassenmilieus vaker verbale en expliciete instructies en suggesties met specialistische woorden, terwijl dit bij de opvoeders uit de lagere sociale milieus vaker op non-verbale manier gebeurt, bijvoorbeeld door dingen aan te wijzen of minder specialistische woorden te gebruiken (Leseman, 2007).

Verder stellen Mulder et al. (2014) dat de ouderlijke betrokkenheid bij het onderwijs een deel van de taal- en rekenachterstand verklaart. Interviews onder directeuren en leerkrachten wijzen op de cognitieve aspecten binnen het gezin als verklaringen voor prestatie. Als onderdeel hiervan worden (taal)vaardigheden genoemd, maar ook betrokkenheid en ambities van de ouders en problemen in het gezin.

Hoger opgeleide ouders zijn vaker in staat een stimulerende omgeving voor de kinderen te creëren (Dronkers, 2007; Meijnen, 2003). Volgens Fettelaar en Smeets (2013) laat het opleidingsniveau als afzonderlijke indicator de hoogste voorspellende waarde in groep 8 en groep 5 zien: in deze groepen wordt 13 tot 14 procent van de variantie in leerlingresultaten verklaard door het opleidingsniveau van de ouders. Herkomst als afzonderlijke indicator voorspelt het best de leerresultaten in groep 2 (ruim 9 procent). Inkomen en herkomst hebben meer voorspellende waarde in de lagere groepen (groep 2 en groep 5) dan in groep 8: deze kenmerken verklaarden samen 11 procent van de verschillen in leerprestaties in groep 2, 10 procent in groep 5 en ruim 5 procent in groep 8. Het opleidingsniveau van de ouders hangt sterker samen met verschillen in leerresultaten op het gebied van taalvaardigheid dan met rekenvaardigheid. Dit is nog meer het geval voor herkomst: allochtone ouders hebben immers zelf vaak een taalachterstand (Fettelaar en Smeets, 2013).

Factoren als armoede en slechte woonomstandigheden zijn ook risicofactoren die tot lagere leerprestaties kunnen leiden (Leseman, 2007). Het gegeven of een leerling in een impuls- of armoedeprobleemaccumulatiegebied woont is echter, evenals het gestandaardiseerd inkomen van het huishouden, als een afzonderlijke indicator een veel minder goede voorspeller van leerresultaten, dan bijvoorbeeld opleidingsniveau. Wel kan de combinatie van inkomen en herkomst gebruikt worden om leerprestaties te voorspellen, waarbij geldt dat deze combinatie een beter voorspeller is voor de verschillen op taalgebied dan op rekengebied (Fettelaar en Smeets, 2013).

Het achterblijven van veel allochtone kinderen hangt in het algemeen samen met hun relatief gedepriveerde sociaaleconomische achtergrond. Veel niet-westerse immigranten behoren tot de lagere sociale milieus. Het niveau van de opleiding van de ouders speelt hierbij een rol en de daarmee gepaard gaande beperkte begeleiding en stimulatie van het kind. Een andere risicofactor is het feit dat Nederlands niet de moedertaal van deze kinderen is. Onder andere Driessen (2003) concludeert dat een betere taalbeheersing van de ouders een positief effect heeft op de taalprestatie van kinderen en dat een andere thuistaal dan het Standaardnederlands meestal een negatief effect heeft. Taalbeheersing neemt gemiddeld genomen af naarmate het opleidingsniveau van de ouders lager is en in de lagere sociale milieus wordt minder vaak standaard Nederlands gesproken.

Uit grootschalige studies blijkt echter dat er toch grote prestatieverschillen zijn tussen autochtone en allochtone kinderen van ouders met een vergelijkbaar opleidingsniveau (Ledoux et al., 2015). Dit wijst erop dat herkomst op zichzelf een invloedrijke factor voor leerprestatie is. Kloprogge en de Wit (2015) geven daarnaast aan dat behalve de verschillen in prestaties, ook de vraag gesteld kan worden of de oorzaken van minder presteren bij beide groepen wel dezelfde zijn en of er verschillende beleidsingrepen nodig zijn om onderprestatie aan te pakken.

2.2.2 Buurt/regio

In het onderzoek van Mulder et al. (2014) wordt in interviews met directeuren en leerkrachten ook gewezen op de rol van de omgeving. Kenmerken van de omgeving die een rol kunnen spelen betreffen de aanwezigheid van laagopgeleiden, werkloosheid, taalarmoede en een belemmerende dorps/straatcultuur (Mulder et al., 2014). Daarnaast kan onveiligheid van de omgeving een rol spelen (Leseman, 2007). Het is met name een opeenstapeling van deze aspecten die negatief uitpakt voor leerlingen. Het ministerie van Onderwijs, Cultuur en Wetenschap kijkt voor het onderwijsachterstandenbeleid in het voortgezet onderwijs dan ook naar armoedeprobleemcumulatiegebieden (apc-gebieden). Dit zijn gebieden waarin een opeenstapeling van 'negatieve' effecten/factoren voorkomt. Apc-gebieden zijn namelijk gebieden waar mensen met lage inkomens, niet-westerse allochtonen en uitkeringsontvangers een aanzienlijk deel van de bevolking vormen.

Onderzoek onder autochtone laag presterende leerlingen wijst uit dat de prestaties van leerlingen het hoogst zijn in het noorden en het laagst in het westen van het land. Deze resultaten hangen samen met de stedelijkheid van de omgeving. In zeer stedelijke gebieden staan de meeste scholen waar autochtone leerlingen slecht presteren, terwijl zich in de plattelandsregio's de meeste scholen bevinden waar de leerlingen juist goed presteren. Scholen in stedelijke gebieden hebben gemiddeld genomen meer voorzieningen (schakelklassen, VVE) maar een complexere samenstelling (o.a. grotere diversiteit aan herkomstgroepen) van leerlingen. Scholen in plattelandsregio's hebben minder vaak te maken met een complexe

leerlingsamenstelling, maar meer met taalproblemen en lage verwachtingen (Mulder et al., 2014).

2.2.3 School

De omgeving van het kind is vaak beperkt tot het gezin en de buurt waarin het gezin woont totdat het kind naar school gaat. Op dat moment beginnen de invloeden van de school een rol te spelen. Scholen kunnen de bestaande verschillen in cognitieve vaardigheden tussen kinderen vergroten of verkleinen (Van Rooijen en Zoon, 2012). Onderzoek onder autochtone leerlingen met lage leerprestaties laat zien dat deze leerlingen vaker op een school met ongunstige kenmerken zitten (Mulder et al., 2014). Dit zijn vaak kleine scholen en scholen met een concentratie van leerlingen met lage leerprestaties. Vaak hebben de leerkrachten relatief lage verwachtingen van deze leerlingen en tonen hun ouders minder betrokkenheid bij het onderwijs (Mulder et al., 2014). Volgens het onderzoek van Mulder et al. (2014) dragen de verwachtingen van de leerkrachten ook sterk bij aan de verklaring van leerprestaties van kinderen.

De omgeving van een school is vaak bepalend voor de samenstelling van de school en de klas waarin het kind zich bevindt. Segregatie is dan ook een belangrijke risicofactor op schoolniveau (Van Rooijen en Zoon, 2012). Vaak gaat het om concentratie van allochtone leerlingen waardoor ook wel de zogenoemde 'zwarte scholen' ontstaan. Zoals eerder beschreven, is het opleidingsniveau van (niet-westerse) allochtone ouders vaak laag. Mede daardoor kampen allochtone kinderen voorschools al met achterstanden op het gebied van bijvoorbeeld taalvaardigheden, wat zich kan uiten in lagere leerprestaties. De concentratie van deze leerlingen op 'zwarte scholen' versterkt het negatieve effect van herkomst op leerprestatie. De mechanismen die op scholen (in groepen) een rol spelen, pakken voor deze groep negatief uit. Er zijn bijvoorbeeld onvoldoende mogelijkheden om te 'mobiliseren' ofwel om over te gaan naar een betere groep. De stimulans van goede leerlingen ofwel leerlingen die het goede voorbeeld geven, ontbreekt. Daarnaast stellen de leraren de doelstellingen naar beneden bij in een dergelijke omgeving, waardoor ook de druk om te presteren wordt afgezwakt. Ook is het zelden het geval dat de beste leerkrachten worden ingezet voor de zwakste niveaugroep (Van Rooijen en Zoon, 2012; Houtveen et al., 2011).

2.3 Meetbare indicatoren

Vaak wordt in onderzoek gesproken over factoren die een effect hebben, maar niet direct gemeten kunnen worden. Ook in dit hoofdstuk hebben we factoren in dit licht besproken, en niet zozeer de werkelijk meetbare indicatoren. In bijlage 2 geven we een overzicht van meetbare indicatoren die mogelijk geschikt zijn om in ons uiteindelijke statistisch model op te nemen.

Per meetbare indicator wordt in de bijlage het volgende overzicht gegeven:

Indicator	Naam indicator
Definitie	Wat is de definitie van de betreffende indicator?
Bron	De naam van de bron die nodig is voor het samenstellen van deze indicator. Dit kunnen ook meerdere bronnen zijn.
Beschikbaarheid	Hier wordt aangegeven voor welke jaren de gegevens, die nodig zijn

	voor de indicator, beschikbaar zijn op moment van schrijven (november 2015). Ook wordt beschreven hoe compleet de data is en wat de dekking is (representativiteit).
Bijzonderheden	Als er verder nog bijzonderheden zijn worden die in deze rubriek vermeld. Bijvoorbeeld wat betreft de kwaliteit of aansluiting op de afbakening.

3. Onderzoeksmethode

Het uiteindelijke doel van het te ontwikkelen verdeelmodel is om geld dusdanig over de basisscholen in Nederland te verdelen dat de aanwezige onderwijssachterstanden zoveel mogelijk worden verkleind. Hoewel het moeilijk is om een eenduidige definitie van “onderwijssachterstanden” te geven, zullen wij ons in dit stuk baseren op het rapport van Kloprugge en de Wit (2015), waarin wordt gesproken van onderwijssachterstanden wanneer “*leerlingen door een ongunstige economische, sociale of culturele omgeving (m.n. de thuissituatie) op school slechter presteren dan ze bij een gunstiger situatie zouden kunnen*”. “Kunnen” wijst hierbij op het potentieel van leerlingen en daarmee vooral naar hun intelligentie. Daarbij wordt dus teruggerepen naar het concept van verborgen talent zoals dat bij de aanvang van het onderwijssachterstandenbeleid reeds werd gebruikt (zie voor dit concept Van Heek, et al., 1968)

Om op basis van deze definitie tot een juiste verdeling van het beschikbare geldbedrag te komen, moeten verschillende stappen gezet worden.

1. Allereerst moet worden onderzocht op welke manier onderwijssachterstand samenhangt met gezins- en omgevingsfactoren. Om deze samenhang in kaart te brengen zullen we een *analysemodel* opstellen waarmee we de leerprestatie (de zogenaamde afhankelijke variabele) schatten op basis van zowel gezins- en omgevingsfactoren (de onafhankelijke variabelen) als van intelligentie. Intelligentie van het kind wordt meegenomen om in kaart te brengen wat de omgevingsvariabelen toevoegen/afdoen aan de leerprestatie die je zou verwachten op basis van de aangeboren potentie van het kind. Op deze manier wordt aan de hand van de omgevingsvariabelen dus daadwerkelijk de eerder gedefinieerde onderwijssachterstand geschat.
2. Wanneer het analysemodel is ontwikkeld, moet er een model komen dat voor alle kinderen in het basisonderwijs een schatting kan maken van onderwijssachterstand (of leerprestatie, als wordt besloten intelligentie buiten beschouwing te laten) op basis van de gezins- en omgevingsfactoren. Hiervoor is een *voorspelmodel* nodig. Dit voorspelmodel zal sterk lijken op het eerder ontwikkelde analysemodel, maar het kan voorkomen dat er in het analysemodel variabelen zijn meegenomen die niet in het uiteindelijke voorspelmodel zullen worden opgenomen. Dit zijn bijvoorbeeld variabelen waarin we op zichzelf niet geïnteresseerd zijn, maar die ervoor zorgen dat we het effect van de variabelen die er echt toe doen nauwkeuriger kunnen schatten.
3. Uiteindelijk zal op basis van de leerling-populatie en de geschatte onderwijssachterstand per leerling, voor een school als geheel moeten worden bepaald hoeveel extra geld er naar deze school toe moet. Op basis van de uitkomsten van het voorspelmodel moet er dus een *verdeelmodel* worden ontwikkeld dat, gegeven de voorspelde leerachterstanden binnen de school, een geldbedrag aan de school toekent. Welke bedragen uiteindelijk worden uitgekeerd aan de scholen is een politieke beslissing. Gegeven de politieke beslissing kan CBS op basis van data uit registers de bedragen per school uitrekenen.

In het vervolg van deze methodebeschrijving wordt op de drie stappen verder ingegaan.

3.1 Analysemodel: samenhang onderwijssachterstand en gezins- en omgevingsfactoren

Met het analysemodel wordt onderzocht hoe we de leerprestatie kunnen schatten met zowel gezins- en omgevingsfactoren als intelligentie. Om het analysemodel te kunnen ontwikkelen is een dataset nodig waarin zowel de waarden van de onafhankelijke als van de afhankelijke variabelen bekend zijn. We kiezen ervoor om dit model te ontwikkelen op basis van de data die zijn verzameld tijdens het cohortonderzoek COOL⁵⁻¹⁸ (CohortOnderzoek OnderwijsLoopbanen van 5 tot 18 jaar) (Driessen en anderen, 2009). In deze dataset zijn leerprestaties op verschillende momenten tijdens de basisschoolperiode gemeten en zijn ook gegevens over het opleidingsniveau van de ouders opgenomen. Omdat CBS niet over integrale gegevens over opleidingsniveau beschikt is het een voordeel dat deze gegevens in de COOL⁵⁻¹⁸ cohortstudie apart zijn verzameld. De validiteit en volledigheid van deze gegevens is echter niet op voorhand vast te stellen, omdat het gaat om gegevens op basis van door de ouders ingevulde vragenlijst, indien nodig aangevuld door de schooladministratie.

Ook is er in de COOL⁵⁻¹⁸ cohortstudie per kind een maat voor “aangeboren intelligentie” beschikbaar op basis van de in groep 5 afgenomen Niet-Schoolse Cognitieve Capaciteiten Test (NSCCT; Van Batenburg & Van der Werf, 2004). Deze test is speciaal ontwikkeld om leerlingen op te kunnen sporen die onder hun niveau presteren. In de komende paragraaf zal verder worden ingegaan op de voor- en nadelen van het gebruiken van deze informatie.

Leerprestatie

Voor de afhankelijke variabele, oftewel leerprestatie, zijn op basis van de literatuur, in principe drie mogelijkheden:

- de Cito-score in groep 2,
- de Cito-score in groep 5,
- de Cito-score in groep 8, en
- het schoolniveau, zoals gemeten in de derde klas van de middelbare school.

De laatste twee variabelen worden het vaakst genoemd, maar wanneer we één van deze twee variabelen als uitkomstvariabele nemen lopen we tegen het probleem aan dat het huidige achterstandenbeleid invloed heeft op deze uitkomsten. In het hypothetische geval dat het achterstandenbeleid de volledige onderwijsachterstanden weg weet te werken - met als resultaat dat ook kinderen met lager opgeleide ouders goed scoren op de Cito toets - zal opleidingsniveau van de ouders niet meer als relevante voorspeller van onderwijsachterstand naar voren komen, terwijl het dit wel is.

De Cito-score in groep 8 is in dit licht nog te verkiezen boven het schoolniveau op de middelbare school, omdat er op de middelbare school weer andere invloeden zijn die de onderwijsachterstanden beïnvloeden.

Om een uitkomstvariabele te hebben die niet of nog slechts in geringe mate is beïnvloed door het huidige onderwijsachterstandenbeleid, zou je een meting van prestatie willen hebben op het moment dat het kind voor het eerst naar school gaat of een moment dat daar in de buurt ligt. In het COOL⁵⁻¹⁸ cohortonderzoek is al in groep 2 een Cito toets afgenomen die leerprestaties meet.

Omdat deze toets zo vroeg in de basisschoolperiode is afgenomen is het waarschijnlijk dat het huidige achterstandenbeleid nog geen grote invloed op deze uitkomsten zal hebben. Precies om deze reden stellen we voor om voor het ontwikkelen van het analysemodel (en het

bijbehorende voorspel- en verdeelmodel) de Cito toets in groep 2 als maat voor leerprestaties te gebruiken.

Deze Cito toets bestaat uit twee onderdelen, te weten het onderdeel Taal voor Kleuters en het onderdeel Ordenen. In de toets Taal voor Kleuters zijn aspecten van de taalontwikkeling (conceptueel bewustzijn) en de ontluikende geletterdheid (metalinguïstisch bewustzijn) opgenomen. In de toets Ordenen worden drie vaardigheden getoetst: classificeren, seriëren en vergelijken en tellen. Een aandachtspunt is de betrouwbaarheid van deze vroege meting. Het zou kunnen dat de leerprestaties pas betrouwbaar geschat kunnen worden op latere leeftijd. Ook is het goed om stil te staan bij de vraag of deze toets echt de leerprestatie toetst, of dat ze eigenlijk grotendeels intelligentie toetst.

Om een inschatting te kunnen maken van de impact die de keuze van de afhankelijke variabele heeft op het uiteindelijke model, is er de mogelijkheid om een model te maken waarin naast de Cito score in groep 2 ook de score in groep 8 wordt meegenomen, en eventueel de score in groep 5. Op deze manier kan bekeken worden of de gezins- en omgevingsfactoren op dezelfde manier doorwerken in onderwijsachterstand in groep 2 als in onderwijsachterstand in groep 8 (en groep 5). Het is niet onwaarschijnlijk dat er factoren zullen zijn die meer invloed hebben in groep 2 dan in groep 8 en omgekeerd. Hoewel de insteek in de eerste plaats zou zijn om het daadwerkelijke verdeelmodel te baseren op Cito scores in groep 2, geeft een breder analysemodel inzicht in onderwijsachterstand en kan zo'n model ook gebruikt worden om de keuze voor een afhankelijke variabele niet slechts op theoretische argumenten te baseren, maar ook te kijken naar de uitwerking daarvan in de analyses. Het CBS raadt aan om deze uitgebreidere analyse uit te voeren.

Intelligentie

Het voorstel is om in het model de factor intelligentie, zoals gemeten door de in groep 5 afgenomen Niet-Schoolse Cognitieve Capaciteiten Test (NSCT), in het analysemodel op te nemen. Op deze manier kan een inschatting gemaakt worden van de, op basis van intelligentie, te verwachten leerprestatie. Deze kan dan worden vergeleken met de werkelijke leerprestatie om tot een maat van onderpresteren/overpresteren te komen. Een dergelijke analyse sluit heel mooi aan bij het beoogde doel van het onderwijsachterstandenbeleid: het beschikbaar stellen van extra geld voor kinderen die, ongeacht hun prestatieniveau, onderpresteren. In de huidige onderzoeken naar het effect van gezins- en omgevingsfactoren op onderwijsachterstand/leerprestatie, is deze aanpak echter ongebruikelijk.

Een mogelijk risico van de voorgestelde aanpak is dat het zou kunnen dat de NSCT niet slechts aangeboren, niet-schoolse, intelligentie meet, maar ook "schoolse vaardigheden" zoals leesvaardigheid. Wanneer een intelligent kind niet in staat is de testvragen te begrijpen, omdat de in de vraag voorkomende woorden bij hem/haar niet bekend zijn, kan ten onrechte geconcludeerd worden dat dit kind niet over de juiste potentie beschikt en dus geen extra aandacht (en geld) nodig heeft. Mede doordat de test is afgenomen in groep 5, kan de testscore mede beïnvloed zijn door het eerder gevolgde onderwijs. Statistisch gezien kan men dit opvatten als een meetfout. De testscore heeft dan een meetfout waardoor de correlatie met de werkelijke intelligentie lager dan 1 zal zijn.

Dit risico moet afgewogen worden tegen het nadeel van het niet meenemen van deze variabele. In dat geval is de enige overgebleven relevante uitkomstmaat immers de

“leerprestatie”. Het achterstandenbeleid zal dan in de eerste plaats gericht zijn op het extra financieel ondersteunen van slecht presterende kinderen, ongeacht of deze kinderen de potentie hebben om beter te presteren. Wanneer de factoren die van invloed zijn op “slecht presteren” vergelijkbaar zijn met de factoren die van invloed zijn op “onderpresteren”, zullen ook kinderen die gemiddeld tot goed presteren, maar niet hun maximale niveau halen, profiteren van het beleid, maar of dit daadwerkelijk het geval zal zijn is van tevoren niet in te schatten.

Om een model te kunnen maken dat daadwerkelijk aansluit op de achterliggende gedachte van het onderwijsachterstandenbeleid, heeft het onze voorkeur om de NSCCT score wel mee te nemen. In het vervolg van dit stuk zullen we daarom uitgaan van een analyse waarin intelligentie, zoals gemeten door de NSCCT, wordt meegenomen. Het is eenvoudig om de analyse alsnog uit te voeren zonder deze variabele of zowel de variant mét als de variant zonder deze variabele te bekijken.

Populaties

Binnen de COOL⁵⁻¹⁸ cohortstudie worden leerlingen uit groep 2, 5 en 8 door de jaren heen gevolgd. Tot nu toe zijn in de schooljaren 2007/2008, 2010/2011 en 2013/2014 metingen verricht in groep 2, 5 en 8. Omdat wij in ons model voor iedere leerling zowel de Cito score in groep 2 als de score op de Niet-Schoolse Cognitieve Capaciteiten Test in groep 5 willen gebruiken, kunnen we slechts 2 cohorten in onze studie meenemen, namelijk de leerlingen die in schooljaar 2007/2008 of 2010/2011 in groep 2 zaten. Van deze leerlingen hebben we immers ook de gegevens uit groep 5, afgenomen in respectievelijk schooljaar 2010/2011 en 2013/2014. Wanneer we in ons model ook het verband tussen Cito score in groep 2 en groep 8 mee willen nemen kunnen we alleen het cohort dat in 2007/2008 in groep 2 zat gebruiken. De groep-2-cohorten uit 2007/2008 en 2010/2011 bestaan uit respectievelijk 13.842 en 11.995 leerlingen. Het aantal wat hiervan daadwerkelijk kan worden gebruikt om het model te schatten zal echter een stuk lager liggen, omdat we alleen kinderen mee kunnen nemen die ook in groep 5 (en eventueel groep 8) in de COOL studie aanwezig waren.

Uit de eerste twee edities van de COOL studie blijkt dat ongeveer de helft van de voormalige COOL1 leerlingen niet meer bij COOL2 betrokken is. Zo blijkt dat van 13.842 groep 2 leerlingen die in 2007/2008 in de studie zaten, er in 2010/2011 nog 5877 worden teruggevonden in de tweede COOL studie. In het technisch rapport dat bij de tweede meting hoort wordt hiervoor de volgende verklaring gegeven: *“Dat komt deels door de uitval van scholen en deels doordat leerlingen in de tussentijd zijn blijven zitten, naar een andere basisschool zijn gegaan of naar het speciaal (basis)onderwijs zijn verwezen. (...) Deze ‘verdwenen’ leerlingen worden in ieder geval via het Onderwijsnummer en zo mogelijk via de nieuw school verder gevolgd. Ook de leerlingen die zijn uitgevallen omdat de school niet meer aan COOL mee wilde doen, worden via het Onderwijsnummer verder gevolgd. De uitgevallen leerlingen blijven hier verder buiten beschouwing; er wordt in een aparte rapportage aandacht aan deze groep leerlingen besteed.”* Het is in deze fase van het onderzoek bij CBS nog niet duidelijk hoeveel informatie er daadwerkelijk beschikbaar is over de uitgevallen leerlingen. Ook is het ons nog niet bekend hoeveel leerlingen die in 2010/2011 in groep 2 zaten, terug te vinden zijn in groep 5 bij de 3^e meting. Voordat we de gegevens gaan gebruiken, evalueren we de selectiviteit van de beschikbare informatie. Indien nodig worden de beschikbare data gewogen om deze representatief te maken.

In eerste instantie stellen we voor om bij het ontwikkelen van het analysemodel zoveel mogelijk informatie mee te nemen, ook als dat betekent dat er kinderen uit verschillende cohorten worden meegenomen. Het zou echter kunnen dat het verband tussen de afhankelijke en onafhankelijke variabelen over de tijd verandert. Dit is te onderzoeken door per cohort een model te maken en deze modellen vervolgens met elkaar te vergelijken. Op deze manier kan ook een inschatting gemaakt worden van de toekomstbestendigheid van het uiteindelijke model.

Op basis van het analyse model zal een voorspelmodel worden ontwikkeld dat gebruikt zal worden om aan het begin van het schooljaar voor *alle* kinderen die op een reguliere basisschool zitten een voorspelling van onderwijsachterstand te doen. Deze voorspellingen zullen daarna in het verdeelmodel gebruikt worden om tot een bedrag per school te komen.

Peilmoment

De onafhankelijke variabelen (zoals het opleidingsniveau van de ouders, het gezinsinkomen, et cetera) kunnen op meerdere momenten gemeten worden. In het huidige model, waarin alleen het opleidingsniveau van de ouders wordt gebruikt, wordt dit opleidingsniveau gemeten op het moment dat het kind voor het eerst naar school gaat of op het moment dat het kind van school verandert.

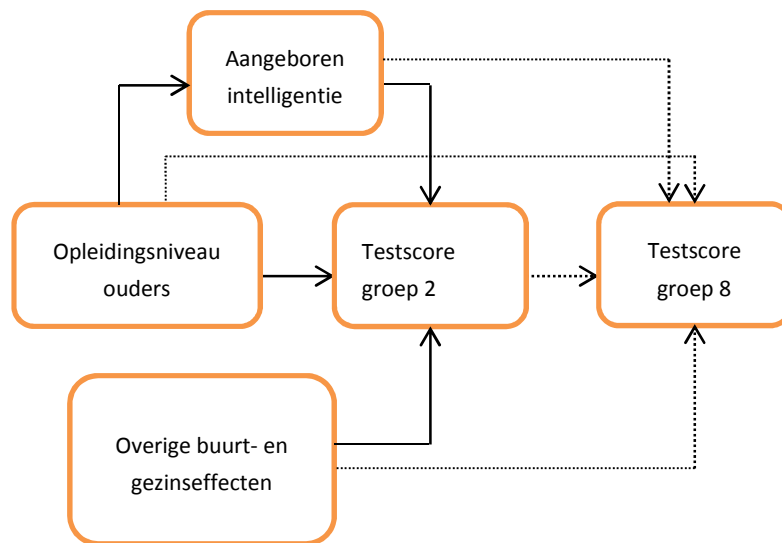
Voor het nieuwe voorspel- en verdeelmodel stellen we wederom voor om de onafhankelijke variabelen op een vast moment te meten, namelijk op het moment dat het kind voor het eerst naar school gaat. Ook voor kinderen in hogere klassen, of kinderen die van school veranderen, gaan we uit van de waardes van de onafhankelijke variabelen toen het kind in groep 1 zat.

In het analysemodel zullen we ook uitgaan van onafhankelijke variabelen die zijn gemeten op het moment dat het kind in groep 1 zat. Deze onafhankelijke variabelen zullen uit CBS registraties komen. Een uitzondering zal het opleidingsniveau van de ouders vormen. In de CBS registers is dit gegeven niet voor iedereen bekend, terwijl het in de COOL⁵⁻¹⁸ cohortstudie apart is gemeten. Om voor het analysemodel over zo compleet mogelijke data te kunnen beschikken, zullen we dus voor het opleidingsniveau van de ouders van de COOL-gegevens uitgaan (gemeten in groep 2). Ter controle kunnen we deze gegevens nog naast de bij CBS aanwezige gegevens over opleidingsniveau leggen.

Om ook veranderingen in de onafhankelijke variabelen mee te kunnen nemen (te denken valt bijvoorbeeld aan echtscheidingen of een veranderd gezinsinkomen) zou ook een dynamisch model kunnen worden ontworpen waarin veranderingen van de gezinssituatie tijdens de schooltijd kunnen worden meegenomen. Zo'n model zou in dat geval wel gemaakt moeten worden met een afhankelijke variabele die aan het eind van de basisschoolperiode gemeten wordt, omdat eventuele veranderingen in de thuissituatie invloed moeten kunnen hebben op deze variabele. Een dergelijk dynamisch model heeft dus het nadeel dat het alleen gemaakt kan worden op basis van een variabele die al is beïnvloed door het huidige achterstandenbeleid (zie ook paragraaf 1.3). Verder is het een complex model om te ontwerpen. We stellen dit voorlopig dan ook alleen voor als mogelijke uitbreiding in een later stadium wanneer er al een eerste model is ontwikkeld op basis van gezins- en omgevingsfactoren zoals gemeten aan het begin van de basisschoolperiode.

3.1.1 Onderliggende structuur analysemodel

Het voorgestelde analysemodel is schematisch als volgt weer te geven:



In eerste instantie zal er worden gekeken naar de effecten die zijn weergegeven als doorgetrokken lijnen. Het gaat dan dus om een (regressie)model waarin Cito scores in groep 2 worden voorspeld aan de hand van intelligentie, opleidingsniveau van de ouders en overige buurt- en gezinseffecten. Wat betreft de buurt- en gezinseffecten komen alle variabelen die uit het literatuuronderzoek naar voren zijn gekomen als potentiële voorspellers van leerprestatie/onderwijsachterstand (zie hoofdstuk 2), in aanmerking om in het analysemodel opgenomen te worden.

In het schema zijn ook stippellijnen opgenomen. De bijbehorende verbanden hebben geen direct effect op de vervolgmogelijken (dat wil zeggen het voorspelmodel en het verdeelmodel) maar geven wel meer inzicht in het onderliggende fenomeen onderwijsachterstand. Wanneer blijkt dat er buurt- en gezinseffecten zijn die anders doorwerken op de Cito scores in groep 2 dan op de Cito scores in groep 8, zou dit bovendien reden kunnen zijn om de keuze om in de eerste plaats Cito scores in groep 2 te voorspellen te heroverwegen.

3.1.2 Opbouw en keuze voor specifiek analysemodel

Om alle effecten te kunnen schatten, moet vervolgens het juiste regressiemodel gekozen worden. In een standaard regressiemodel wordt er vanuit gegaan dat alle waarnemingen onafhankelijk zijn. In ons geval verwachten we dat twee kinderen uit hetzelfde gezin, dezelfde buurt of van dezelfde school meer overeenkomsten vertonen dan twee willekeurig gekozen kinderen, waardoor deze aanname van onafhankelijkheid niet langer opgaat. Daardoor worden de varianties foutief geschat en worden effecten te gemakkelijk significant verklaard. Om voor die afhankelijkheden te corrigeren wordt een clustercorrectie uitgevoerd.

Omdat we de invloed van gezins- en buurtfactoren op onderwijsachterstand willen schatten, zullen we al deze factoren als onafhankelijke variabelen in het regressiemodel opnemen. Wanneer we alleen deze omgevingsfactoren in het model opnemen is er echter een kans dat het werkelijke omgevingseffect overschat wordt, omdat er een onderliggende variabele bestaat die zowel invloed heeft op de omgevingsfactor als op de leerprestatie. Een dergelijke variabele

wordt binnen de statistiek een “confounder” genoemd. Het zou bijvoorbeeld kunnen dat, omdat het kind ernstige gedragsproblemen heeft, de ouders minder gaan werken. Hierdoor zal het gezinsinkomen omlaag gaan. Het kind presteert vervolgens slecht op school, voornamelijk door de gedragsstoornis. Omdat de variabele gedragsstoornis niet in ons model is opgenomen, maar gezinsinkomen wel, lijkt het nu of de schoolprestatie wordt beïnvloed door het gezinsinkomen, terwijl het in feite door de gedragsstoornis wordt beïnvloed.

Om de werkelijke effecten van omgevingsvariabelen op onderwijsachterstand te kunnen schatten, zullen mogelijke confounders dus meegenomen moeten worden in het model. Hierbij dient te worden opgemerkt dat er van tevoren goed nagedacht moet worden over welke mogelijke confounders er zijn, omdat het willekeurig toevoegen van variabelen ook het omgekeerde effect kan hebben en er juist toe kan leiden dat de werkelijke omgevings- en gezinseffecten onderschat worden. Ook is er de optie om bijvoorbeeld zorgleerlingen buiten de analyse te houden, om effecten op die manier zuiverder te kunnen schatten.

Uiteindelijk zal het model in meerdere fases worden geschat:

1. Allereerst zal een zo volledig mogelijk model worden geschat waarin alle mogelijke invloedrijke variabelen worden meegenomen. Omdat veel van deze variabelen in de praktijk toch weinig invloed zullen hebben, of sterk gecorreleerd zullen zijn met andere variabelen, zal dit model zeer waarschijnlijk te uitgebreid zijn.
2. In de volgende stap zal het model stap voor stap teruggebracht worden tot een model waarin alleen de werkelijk relevante variabelen overblijven.
3. Dit model zal vervolgens vergeleken worden met een aantal, op voorhand interessante, deelmodellen. De deelmodellen waar in ieder geval naar gekeken zal worden zijn de door het ministerie van OCW voorgestelde modellen met de volgende variabelen (+ clustercorrectie, eventuele confounders en intelligentie):
 1. Opleidingsniveau en inkomen
 2. Opleidingsniveau en herkomst
 3. Opleidingsniveau, inkomen en herkomst
 4. Opleidingsniveau, inkomen, herkomst en buurtgegevens
 5. Opleidingsniveau, inkomen, buurtgegevens en verblijfsduur en status
 6. Inkomen, herkomst, buurtgegevens
 7. Inkomen, herkomst, buurtgegevens en verblijfsduur en status.

Voor deze deelmodellen zal gekeken worden of ze net zo goed schatten als het voorgestelde model in stap 2. Dit wordt beoordeeld aan de hand van een goodness of fit test waarin wordt bepaald hoe goed een model bij de gegevens past en aan de hand van de grootte van de verklaarde variantie in de afhankelijke variabele.

Om een beter inzicht te krijgen in de betekenis van het uiteindelijke analysemodel, zal nagedacht worden over visualisaties waarin inzichtelijk wordt gemaakt welke factoren nu precies leiden tot onderwijsachterstanden, en hoe dat terug te zien is wanneer de leerprestaties van verschillende groepen kinderen met elkaar worden vergeleken.

3.2 Van analysemodel naar voorspelmodel

3.2.1 Onbekende of ongewenste factoren niet meenemen.

Om van het uiteindelijke analysemodel tot een voorspelmodel te komen, willen we alleen de variabelen uit het analysemodel behouden die we zowel *willen* als *kunnen* gebruiken in het voorspelmodel. Wanneer we in het analysemodel confounders zouden hebben opgenomen, die op zichzelf geen gezins- of omgevingsfactor zijn, willen we deze variabelen niet in het voorspelmodel meenemen. Deze confounders waren immers alleen in het analysemodel opgenomen om de werkelijke effecten van de gezins- en omgevingsfactoren goed te kunnen schatten.

Ook intelligentie is een factor die waarschijnlijk wel in het analysemodel is meegenomen, maar die niet meegenomen hoeft te worden in het voorspelmodel. Ten eerste kan deze variabele niet worden meegenomen voor nieuwe leerlingen omdat hij niet bekend is voor leerlingen die niet binnen de COOL⁵⁻¹⁸ cohortstudie vielen. Ten tweede zal het voorspelmodel zonder deze factor precies onderwijsachterstand schatten (en dus niet langer leerprestatie) omdat de regressiecoëfficiënten in het analysemodel gecorrigeerd zijn voor intelligentie. Het overgebleven voorspelmodel zal dus precies het deel van de leerprestatie voorspellen dat uit gezins- en omgevingsfactoren en niet uit intelligentie verklaard kan worden. Een negatieve waarde zal dus duiden op een omgeving die leidt tot onderpresteren, terwijl een positieve waarde duidt op een kind dat beter presteert dan alleen op basis van intelligentie te verwachten valt.

3.2.2 Onvolledigheid van opleidingsniveau

Het is zeer waarschijnlijk dat het uiteindelijke analysemodel het opleidingsniveau van de ouders als onafhankelijke variabele zal bevatten. Ditzelfde zal dan gelden voor het voorspelmodel. Om met dit voorspelmodel voor nieuwe kinderen een voorspelling te kunnen doen over de mate van onderwijsachterstand zal het opleidingsniveau van de ouders dus nodig zijn.

Omdat het analysemodel wordt gemaakt met behulp van de COOL data, waarin (voor zover we nu weten) opleidingsniveau van de ouders per kind bekend is, is er tijdens het maken van het analysemodel nog geen probleem met missende waardes. In de voorspelfase zal echter, voor nieuwe kinderen, CBS registerdata gebruikt worden om per kind het voorspelmodel in te vullen. Opleidingsniveau is bij CBS helaas niet voor iedereen bekend. De reden van een onbekend opleidingsniveau is dat het diploma behaald is voordat de behaalde diploma's integraal geregistreerd werden, of omdat de ouders hun opleiding in het buitenland hebben afgerond. Voor de ouders van kinderen die op dit moment op de basisschool zitten is van ruim 40% geen opleidingsniveau bij CBS bekend. Wel zien we dat voor jongere kinderen het opleidingsniveau van hun ouders steeds beter bekend is.

Om toch een voorspelling te kunnen doen voor kinderen van wie het opleidingsniveau van één of beide ouders onbekend is, zijn er twee mogelijke oplossingen, te weten *wegen* of *imputeren*.

Wegen

Het achterliggende idee van wegen is om per school alleen voor de leerlingen waarvan het opleidingsniveau van de ouders wel bekend is, de onderwijsachterstand te schatten. Van deze schattingen wordt vervolgens een gewogen som genomen om tot een verdeling van onderwijsachterstand op schoolniveau te komen welke zou moeten lijken op de verdeling die je

had gehad wanneer je wel voor alle leerlingen alle informatie over het opleidingsniveau van de ouders zou hebben. Het idee is dus niet om per kind alsnog een schatting voor opleidingsniveau van de ouders en leerprestatie te hebben, maar om op schoolniveau een goed beeld te krijgen van het aantal leerlingen met een onderwijsachterstand en van de omvang van die onderwijsachterstand.

Per kind van wie het opleidingsniveau van de ouders bekend is, moet dus een “gewicht” bepaald worden. Op die manier worden kinderen van wie het opleidingsniveau van de ouders onbekend is, gerepresenteerd door andere kinderen die “op hen lijken” en waarvoor het opleidingsniveau van de ouders en daarmee ook de onderwijsachterstand bekend is. De effectiviteit van een wegingsprocedure staat of valt met de beschikbaarheid van geschikte *hulpvariabelen*. Tijdens de weging wordt gezorgd dat de waargenomen data na weging representatief zijn voor de gekozen hulpvariabelen, in de verwachting dat dat ook leidt tot representatieve uitkomsten voor de doelvariabele (in dit geval onderwijsachterstand). Hulpvariabelen moeten variabelen zijn waarvoor we op schoolniveau weten hoe de verdeling is, en waarvan we verwachten dat ze sterk gecorreleerd zijn met de doelvariabele. Beschikbare hulpvariabelen zijn bijvoorbeeld het gestandaardiseerd huishoudensinkomen, de herkomst van het kind en de duur van het verblijf in Nederland (van de ouders en/of het kind).

Het is in dit stadium van het onderzoek nog niet in te schatten hoe goed een weegprocedure zal werken. Wanneer er scholen zijn waar voor vrijwel geen enkele leerling iets bekend is over het opleidingsniveau van de ouders, zal het voor deze scholen onmogelijk zijn om een weegprocedure toe te passen. Ook voor kleine scholen zal een weegprocedure lastig uitvoerbaar zijn. Aangezien het aantal leerlingen per basisschool op dit moment varieert van 5 tot 1105 is het niet makkelijk om iets te voorspellen voor “een gemiddelde school”. Voor kleine scholen (zeg scholen met minder dan 100 leerlingen, wat geldt voor ongeveer 20% van de Nederlandse scholen) zal het bij een weegprocedure waarschijnlijk nodig zijn om meerdere scholen samen te nemen en dan op dat overkoepelende niveau het aantal en de omvang van leerachterstanden te schatten.

Imputeren

Waar bij wegen de nadruk niet ligt op het schatten van de missende opleidingsniveaus, maar op het direct schatten van de onderwijsachterstanden op schoolniveau, wordt bij imputeren wel eerst een schatting gemaakt van het missende opleidingsniveau om daarna per kind een schatting van onderwijsachterstand te kunnen maken.

Voor ieder kind van wie het opleidingsniveau van één of beide ouders ontbreekt, wordt dit opleidingsniveau dus geschat/geïmputeerd. Deze schatting kan op verschillende manieren tot stand komen. Net als bij de schatting van onderwijsachterstand, kan er ook voor opleidingsniveau een regressiemodel gemaakt worden dat op basis van verschillende onafhankelijke variabelen (vergelijkbaar met de hulpvariabelen in de weegprocedure) een schatting maakt van de missende waarde voor opleidingsniveau. Ook zijn er varianten denkbaar waarbij voor een leerling met een ontbrekend opleidingsniveau van de ouders wordt gezocht naar een leerling die “vergelijkbaar” is en waarvoor dit opleidingsniveau wel bekend is, om vervolgens voor beide leerlingen met dit bekende opleidingsniveau verder te rekenen.

Net als bij het weegmodel is het wederom niet op voorhand in te schatten hoe goed een imputatiemethode zal werken. Voor beide varianten zal getest moeten worden hoe goed het in de praktijk uitpakt.

3.2.3 Testen van het voorspelmodel

Bij het opstellen van het analysemodel is gekeken hoe goed leerprestaties te voorspellen zijn op basis van intelligentie en gezins- en omgevingsfactoren. Op basis van de voorspellende waarde op de COOL dataset is vervolgens een keuze gemaakt voor “het best voorspellende model”. Het is echter nodig om ook inzicht te hebben in de voorspellende waarde van het model wanneer het gebruikt wordt op een dataset waarin andere leerlingen zitten dan de leerlingen die gebruikt zijn om het model te ontwikkelen.

Om te kunnen zien hoe goed een model voorspelt, is het nodig om de voorspelling naast een werkelijke waarde te kunnen leggen. Omdat we voor nieuwe leerlingen niet op de hoogte zijn van hun Cito scores in groep 2, kunnen we het model alleen valideren met behulp van de COOL data. Om te zorgen dat de data niet dubbel gebruikt worden, dus zowel om het model te maken als om het model te testen, is het mogelijk om de data op te delen in *test data* en *training data*. Met de training data wordt het model gemaakt, terwijl het model getest wordt op de test data. Om niet te afhankelijk te zijn van één willekeurige opdeling in training en test data, kan deze stap verschillende keren herhaald worden.

Omdat het voorspelmodel onderwijsachterstand voorspelt (en dus niet leerprestatie) en de werkelijke onderwijsachterstand niet gemeten is, kan voor dit model de voorspelling niet direct naast de werkelijke waarde gelegd worden. Wel kan ook hier uitgegaan worden van de voorspellende waarde van het analysemodel, maar dan in de situatie waarin het opleidingsniveau van de ouders niet voor iedereen bekend zal zijn. Om een realistische dataset te hebben waarin de Cito scores in groep 2 wel voor iedereen bekend zijn, maar de opleidingsniveaus van de ouders niet, kan het opleidingsniveau uit de COOL⁵⁻¹⁸ cohortstudie vervangen worden door de opleidingsniveau-gegevens van CBS. Op deze manier zal er een patroon van missende waarden ontstaan dat ook representatief is voor kinderen die niet in de COOL⁵⁻¹⁸ cohortstudie zijn opgenomen. Op deze manier kan worden gekeken hoe goed de voorspellingen op basis van weging/imputatie in de buurt komen van de echte Cito-scores en ook in hoeverre ze afwijken van de voorspellingen op basis van complete data.

Als blijkt dat de voorspellingen op basis van bekende opleidingsniveaus van de ouders veel beter zijn dan de voorspellingen op basis van geïmputeerde waarden of op basis van een weging, kan er besloten worden om alsnog voor een analyse- en voorspelmodel te kiezen waarin opleidingsniveau van de ouders geen rol speelt. Ook kan er mogelijk nagedacht worden over een variant waarin voor de ene leerling opleidingsniveau van de ouders wel wordt meegenomen en voor de andere niet.

3.3 Van voorspelmodel naar verdeelmodel

Wanneer er een voorspelling van onderwijsachterstand op kindniveau (in het geval van imputatie) of schoolniveau (in het geval van een weging) gemaakt kan worden, moet er vervolgens nog bepaald worden hoeveel geld er precies naar elke school toe moet gaan.

De manier om van een voorspelling van onderwijsachterstand tot een geldbedrag te komen, is zeker niet eenduidig en uiteindelijk een politieke keuze. Het is immers niet wetenschappelijk vastgesteld hoeveel geld het kost om onderwijsachterstanden daadwerkelijk weg te werken. Hoewel de definitieve keuze voor een verdeelmodel dus niet door CBS genomen zal worden, zullen we hieronder kort enkele mogelijke varianten bespreken.

Een vrij simpele manier zou zijn om de (gewogen) voorspelde onderwijsachterstanden voor alle kinderen in Nederland in te delen in decielgroepen. Op die manier kan per kind bepaald worden in welke decielgroep het valt. Per school kan vervolgens bepaald worden hoeveel kinderen er in de verschillende decielgroepen vallen, en per decielgroep kan er een gewicht/bedrag worden vastgesteld. Voor kinderen in de eerste decielgroep, dus kinderen behorende tot de laagst scorende 10% (waar een lagere score wil zeggen dat het gaat om een grotere onderwijsachterstand), kan vervolgens een bepaald bedrag worden uitgetrokken. Ook kan er bijvoorbeeld naar de tweede en derde decielgroep gekeken worden, maar voor deze kinderen zal waarschijnlijk een minder hoog bedrag nodig zijn.

In plaats van naar decielgroepen kan ook naar kleinere of grotere percentielgroepen gekeken worden. Ook moet de keuze gemaakt worden of het geld volledig verdeeld wordt over de slechtst scorende groep, of dat het wordt besteed aan verschillende percentielgroepen.

Een andere mogelijkheid, meer in lijn met de huidige gewichtenregeling, zou zijn om de belangrijkste voorspellende variabelen uit het voorspelmodel te pakken en kinderen op basis daarvan te classificeren in groepen waarvoor een grote, middelmatige of geen onderwijsachterstand wordt verwacht. Op basis van een aantal eenvoudige beslisregels (ieder voor zich vergelijkbaar met de huidige beslisregel waarin uit het opleidingsniveau van de ouders direct het gewicht van het kind kan worden bepaald) zou dan bepaald kunnen worden welk extra geldbedrag er voor een kind moet worden gereserveerd.

In bovenstaande varianten wordt geopperd om aan groepen kinderen hetzelfde gewicht (en bijbehorende geldbedrag) toe te kennen, maar het is ook mogelijk om unieke kindgewichten te berekenen op basis van de eveneens unieke voorspellingen van onderwijsachterstand.

Bovenstaande lijst is zeker niet compleet en moet slechts beschouwd worden als stof tot nadenken, niet als een waardeoordeel van CBS. Uiteindelijk gaat het hier om een keuze die het ministerie van OCW zal moeten maken, in overleg met het onderwijsveld.

Het effect van het uiteindelijke verdeelmodel zal niet direct te toetsen zijn. Mocht het model daadwerkelijk ingevoerd worden, dan kan het effect van het nieuwe beleid op langere termijn worden vergeleken met de effecten van het huidige beleid.

4. Risico's en voorwaarden

4.1 Risico's

Het voorgestelde plan van aanpak is niet zonder risico's. Hieronder wordt kort ingegaan op een aantal mogelijke risicofactoren:

1. In deze methodebeschrijving is er vanuit gegaan dat we de beschikking zullen hebben over de (in eerste instantie niet geanonimiseerde) datasets van de COOL⁵⁻¹⁸ cohortstudie. Alhoewel we verwachten dat we met deze data zullen mogen werken, zijn deze data niet in beheer van CBS en zal er aparte toestemming voor gegeven moeten worden.
2. Omdat we de COOL data nog niet hebben kunnen inzien, is het op dit moment niet mogelijk om een uitspraak te doen over de kwaliteit van deze data. Met name de kwaliteit van de variabele "opleidingsniveau van de ouders" is nog onduidelijk. Uit de technische rapporten behorende bij de eerste twee studies (in 2007/2008 en 2010/2011) valt af te leiden dat het opleidingsniveau van de ouders niet altijd door de ouders is verstrekt en dat de school ook uitgaat van eigen administratie op dit gebied. Uit eerder onderzoek naar de functionaliteit van de huidige gewichtenregeling kwam al naar voren dat de gegevens uit deze schooladministratie niet altijd overeenkomen met de werkelijke situatie. De data zullen naast de eigen CBS data gelegd moeten worden om een inschatting te kunnen maken van de betrouwbaarheid ervan. Wanneer de data onbetrouwbaar blijken zou wellicht toch gebruik gemaakt kunnen worden van de bij CBS aanwezige informatie over opleidingsniveau. In dat geval kan ervoor worden gekozen om het analysemodel alleen te schatten op basis van kinderen waarvan de opleidingsniveaus van beide ouders bekend zijn. Deze data zal dan wel worden gewogen om representativiteit te kunnen garanderen.
3. Op dit moment is er door CBS nog geen uitgebreid (literatuur)onderzoek gedaan naar de bruikbaarheid van de Cito score in groep 2 als maat voor leerprestatie en naar de NSCT score als maat voor intelligentie.
4. Het is op dit moment nog niet duidelijk hoeveel kinderen we mee kunnen nemen in het schatten van het analysemodel. Dit hangt er vanaf hoeveel kinderen in 2 (of zelfs 3) opeenvolgende COOL studies aanwezig waren. Alhoewel de verwachting is dat er genoeg gegevens zullen zijn om het analysemodel goed te schatten, is dit niet met 100% zekerheid te zeggen.
5. Het is op voorhand niet goed te voorspellen hoe betrouwbaar de uiteindelijke voorspellingen zullen zijn. Dit is van verschillende factoren afhankelijk, waaronder de uiteindelijke keuze voor het beste model, de kwaliteit van de weging/het imputeermodel en van de verdeling van het aantal missende opleidingsniveaus per school. Het is zeer goed mogelijk dat de voorspelling van het aantal achterstandsleerlingen per decielgriep op schoolniveau niet altijd betrouwbaar genoeg zal zijn om de daadwerkelijke geldverdeling op te baseren. In dat geval moet gekeken worden naar de mogelijkheid om deze aantallen niet op schoolniveau maar op bijvoorbeeld schoolbestuursniveau aan te leveren.
6. Er is een grote kans dat het percentage verklaarde variantie van het uiteindelijke analysemodel aan de lage kant zal zijn. Zo blijkt uit het recent gepubliceerde artikel van Driessen (2015) dat een model met Cito score in groep 2 als afhankelijke variabele en de (op basis van significantie geselecteerde) belangrijkste drie gezinsfactoren als onafhankelijke variabelen nog geen 10% van de totale variantie verklaarde. Eenzelfde

soort percentage (14%) kwam ook uit een eerder onderzoek van CBS naar de invloed van gezinsfactoren op leerprestatie (van Gaalen et al., 2014). Hoewel deze modellen niet 1 op 1 vergelijkbaar zijn met het door ons voorgestelde model, is er een reële kans dat ook in ons onderzoek slechts een klein percentage van de variantie van leerprestatie/onderwijsachterstand verklaard zal kunnen worden. Toch hoeft dit op zichzelf zeker geen teleurstellende conclusie te zijn, omdat het leidt tot dieper inzicht in de mogelijkheden om onderwijsachterstanden te voorspellen uit gezins- en omgevingsvariabelen.

7. Een ander punt van aandacht is de toekomstbestendigheid van het te ontwikkelen model. Uit het artikel van Geert Driessen (2015) blijkt dat de correlatie tussen gezinsvariabelen en leerprestaties over de loop van de jaren sterk is teruggelopen. Wanneer dit een trend is die zich voortzet, is het mogelijk dat het door ons ontwikkelde model over de loop der jaren aan voorspelkracht inboet. Wel moet hierbij opgemerkt worden dat daarmee niet gezegd is dat het model op dat moment aangepast moet worden. Het kan immers zo zijn dat de in het model meegenomen variabelen nog steeds de beste voorspellers voor onderwijsachterstand zijn, ook al is onderwijsachterstand steeds slechter te voorspellen op basis van omgevingsfactoren.

4.2 Voorwaarden

Daarnaast zullen op het onderzoek een aantal voorwaarden van toepassing zijn:

1. Bij de ontwikkeling van een nieuw verdeelmodel is het van groot belang dat de nieuwe methode breed gedragen wordt binnen de onderwijs- en onderzoekswereld. CBS stelt daarom voor dat het ministerie van OCW bij aanvang van het onderzoek een klankbordgroep in het leven roept met daarin experts op het gebied van onderwijskunde en sociologie en vertegenwoordigers uit het onderwijsveld. CBS kan desgewenst een voorstel doen voor de bemensing van de klankbordgroep.
2. CBS gaat ervanuit dat scholen geen bezwaar hebben tegen berekening van een voorspelde leerachterstand per school en verstrekking van die gegevens aan het ministerie van OCW en/of publicatie. In dat geval ziet CBS geen belemmering in de levering van gegevens aan het ministerie van OCW.
3. CBS levert geen gegevens aan het ministerie waaruit informatie over individuele (ouders van) leerlingen kan worden afgeleid. CBS kan daarom geen voorspelde leerachterstand leveren voor scholen met zeer weinig (gewicht)leerlingen.
4. CBS kan op grond van empirisch onderzoek bepalen welke kenmerken van belang zijn voor onderwijsachterstanden en wat de omvang van die invloed is. Op basis hiervan is het aan het ministerie van OCW om te bepalen welke kenmerken daadwerkelijk gebruikt worden. CBS kan op basis van dit besluit de geschatte leerachterstand per school berekenen. Het is vervolgens aan het ministerie van OCW om op basis van de geschatte leerachterstand per school budget aan de scholen toe te kennen.
5. Bij het bepalen van de geschatte leerachterstand zal CBS onder andere uitgaan van de inschrijvingen in het basisonderwijs die DUO jaarlijks aanlevert aan CBS. CBS ontvangt de inschrijvingen in het basisonderwijs gedurende het schooljaar. Het is dus niet mogelijk om al voorafgaand aan het schooljaar de geschatte leerachterstand per school te bepalen. De geschatte leerachterstand die CBS per school levert aan het ministerie van OCW zal dus betrekking hebben op het voorgaande schooljaar of het jaar daarvoor.

6. De actualiteit van de in het model gebruikte gegevens kan per variabele verschillen. Er wordt naar gestreefd om voor alle onafhankelijke variabelen de waarde van deze variabele te gebruiken uit het jaar dat het kind in groep 1 zat. Mocht het echter voorkomen dat deze waardes nog niet beschikbaar zijn op het moment dat de onderwijssachterstand geschat moet worden, dan kan ervoor gekozen worden om de waardes van een voorgaand jaar te nemen.

Bijlage 1. Literatuur

- Batenburg, T. van, & Werf, M. van der (2004). NSCCT: Verantwoording, normering en handleiding. Groningen: GION.
- Driessen, G. (2003). Family and Child characteristics, Child-Rearing Factors, and Cognitive Competence of Young Children. *Early Child Development and Care*, vol. 173, 323-339.
- Driessen, G. (2015) De wankel empirische basis van het onderwijsachterstandenbeleid. *Mens & Maatschappij*, vol 90, nr 3.
- Driessen, G., Mulder, L., Ledoux, G., Roeleveld, J. en van der Veen, I. (2009) Cohortonderzoek COOL⁵⁻¹⁸. Technisch rapport basisonderwijs, eerste meting 2007/08. ITS.
- Dronkers, J. (2007). Ruggengraat van ongelijkheid: Beperkingen en mogelijkheden om ongelijke onderwijskansen te veranderen. Amsterdam: Mets & Schilt uitgevers/Wiardi Beckman Stichting.
- Fettelaar, D. & Smeets, E. (2013). Mogelijke indicatoren van schoolgewichten. Onderzoek naar de voorspellende waarde. Nijmegen: ITS, Radboud Universiteit Nijmegen.
- Galen, R. van, Bakker, B., Laan, J. van der, Westerman, S. en Scholtus, S. (2014) Hoeveel ouderlijk gezang klinkt door in het gepiep van de jongen? Den Haag: CBS.
- Heek, F. van, et al. (1968), Het verborgen talent. Milieu, schoolkeuze en schoolgeschiktheid (Meppel: Boom)
- Houtveen. A.A.M. & Van der Velde, V., m.m.v. Van Niekerk, J.C.A. (2011). Opbrengst van taalbeleid: Eindrapportage pilots taalbeleid onderwijsachterstanden. Utrecht: Hogeschool Utrecht/Zuidam en Uithof Utrecht BV.
- Klopprogge, J. & de Wit, W. (2015). Het onderwijsachterstandenbeleid na 2015. Literatuurstudie t.b.v. expertbijeenkomst OAB september 2015. Nationaal Regieorgaan Onderwijsonderzoek.
- Ledoux, G., Roeleveld, J., Veen, A., Karssen, M., van Daalen, M., Blok, H., Kuiper, E. & Dijkers, L. (2015). Het onderwijsachterstandenbeleid onderzocht. Werkt het zoals bedoeld? ITS, Radboud Universiteit Nijmegen | Kohnstamm Instituut Amsterdam.
- Leseman, P.P.M. (2007). Achterstandenbeleid: voorbij de voor- en vroegschoolse periode. In P.A.H. Van Lieshout, M.S.S. Van der Meij & J.C.I. De Pree (Red.), *Bouwstenen voor betrokken jeugdbeleid*. Amsterdam: Amsterdam University Press.
- Meijnen, G.W. (2003). *Onderwijsachterstanden in basisscholen*. Leuven: Garant.
- Mesman, J. (2011). Oud geleerd, jong gedaan. *Het Jonge Kind*, 11, 9-12.
- Mulder, L., Fettelaar, D., Schouwenaars, I., Ledoux, G., Dijkers, L. & Kuiper, E. (2014). De achterstand van autochtone doelgroepopleerlingen. Oorzaken en aanpak. Nijmegen: ITS, Radboud Universiteit Nijmegen.

Reynders, T., Nicaise, I. & Van Damme, J. (2005). Longitudinaal onderzoek in het basisonderwijs. De constructie van een SES-variabele voor het SiBO-onderzoek. LOA-rapport nr. 31. Leuven: SIBO.

Rooijen, K. van & Zoon, M. (2012). Risicofactoren voor Onderwijsachterstanden. Nederlands Jeugd Instituut:

http://www.nji.nl/nji/dossierDownloads/Risicofactoren_Onderwijsachterstanden.pdf

Vedder, P. & Bekkers, B. (1993). Opstap: Onderzoek en praktijk. Amsterdam: Averroès Stichting.

Bijlage 2. Overzicht indicatoren

Overzicht indicatoren ten behoeve van het onderzoek gewichtenregeling basisonderwijs op basis van registers binnen CBS. De beschikbaarheid die is vermeld geldt op 30 november 2015.

Schoolprestaties

1. Cito score groep 2
2. Cito score groep 5
3. Citoscores van groep 8

Kenmerken van kind/ouders/huishouden

4. Opleidingsniveau (vader, moeder)
5. Inkomen (vader, moeder, huishouden)
6. SEC (vader, moeder)
7. NSCCT (Niet-Schoolse Cognitieve Capaciteiten Test) groep 5
8. Leeftijd moeder van leerling
9. Leeftijd leerling
10. Geslacht leerling
11. Gescheiden ouders
12. Overlijden ouders
13. Type huishouden
14. Herkomst
15. Vluchtelingkind
16. Generatie leerling
17. Verblijfsduur
18. Verblijfstatus
19. Thuistaal
20. Taalbeheersing ouders
21. Schuldsanering ouders
22. Ouders ooit verdacht van misdrijf
23. Indicatie leerlinggebonden financiering (LGF)
24. Ontvangen jeugdzorg (leerling)
25. Gebruikte zorg onder basisverzekering Zvw (leerling, ouders)
26. Verstrekte geneesmiddelen (leerling, ouders)
27. Ziekenhuisopnamen (leerling, ouders)
28. Indicatie / gebruik langdurige zorg (leerling, ouders)
29. Melding kindermishandeling

Kenmerken buurt

30. Armoede
31. Woningkenmerken
32. Cultureel aanbod
33. Stedelijkheid
34. Veiligheid

Kenmerken van de school

35. Groepsgrootte
36. Cumulatie

Schoolprestaties

Indicator	1. Cito score groep 2
Definitie	De score die een leerling heeft gehaald op de Citotoets Ordenen en Citotoets Taal voor kleuters in groep 2 van de basisschool.
Bron	Cito-Leerlingvolgsysteem
Beschikbaarheid	Vanaf schooljaar 2007/2008 tot schooljaar 2013/2014. Het bestand bevat alleen gegevens van groep-2-leerlingen die deelnamen aan het Cito-Leerlingvolgsysteem, van scholen die meedoen aan de steekproef. Het gaat hierom een steekproef voor groep 2: referentiesteekproef: 10069 leerlingen en totale steekproef groep 2: 13842 leerlingen.
Detailniveau	
Bijzonderheden	Variabele komt uit het COOL-onderzoek en CBS beschikt nog niet over deze data.

Indicator	2. Cito score groep 5
Definitie	De score die een leerling heeft gehaald op de Citotoets Ordenen en Citotoets Taal voor leerlingen in groep 5 van de basisschool.
Bron	Cito-Leerlingvolgsysteem
Beschikbaarheid	Vanaf schooljaar 2007/2008 tot schooljaar 2013/2014. Het bestand bevat alleen gegevens van groep-5-leerlingen die deelnamen aan het Cito-Leerlingvolgsysteem, van scholen die meedoen aan de steekproef. Het gaat hierom een steekproef voor groep 5: referentiesteekproef: 9288 leerlingen en totale steekproef groep 8: 12609 leerlingen.
Detailniveau	
Bijzonderheden	Variabele komt uit het COOL-onderzoek en CBS beschikt nog niet over deze data.

Indicator	3.Citoscore groep 8
Definitie	De standaardscore die een leerling heeft gehaald op de Citotoets in groep 8 van de basisschool.
Bron	Eindtoets basisonderwijs Cito
Beschikbaarheid	01-09-2005 tot 1-9-2014 (schooljaar 2013/2014).
Detailniveau	De Citostandaardcode kan nog uitgesplitst worden naar de score voor rekenen, studievoordigheden en wereldoriëntatie(de laatste is optioneel omdat dit onderdeel niet verplicht is).
Bijzonderheden	Het bestand bevat alleen gegevens van groep-8-leerlingen die deelnamen aan de Citotoets. De school moet toestemming hebben gegeven aan Cito om de data te leveren aan CBS. Dus leerlingen op scholen die niet mee doen of geen toestemming geven zitten niet in dit bestand. Gebruik van de Cito data door CBS kan in overleg met Cito.

Kenmerken van kind/ouders/huishouden

Indicator	4.Opleidingsniveau (vader, moeder)
Definitie	Gedetailleerde typering van hoogst behaalde opleiding of hoogst gevolgde opleiding op peilmoment (laatste vrijdag van september van jaar JJJJ), dit is het opleidingsnummer.
Bron	<p>Opleidingsniveaubestand. Doelpopulatie is GBA-bevolking op peilmoment. Bron bevat alleen records van personen waarvan de hoogste opleiding bekend is. De informatie is voor een deel van de bevolking integraal en voor de overige personen op steekproefbasis.</p> <p>Naast het GBA, is voor het COOL-onderzoek in een vragenlijst aan de ouders van de gevolgde leerlingen gevraagd naar het opleidingsniveau (gevolgd en behaald).</p>
Beschikbaarheid	<p>Het opleidingsbestand is beschikbaar vanaf 24-9-1999. Meest recente dataset is 2013 (definitief).</p> <p>Opleidingsniveaubestand is deels op steekproefbasis, oorzaak: Aangezien vooral de onderwijsregistraties nog niet zo heel lang bestaan, ontbreekt er helaas registerinformatie over veel oudere Nederlanders. Ook particuliere opleidingen vallen buiten het bereik van de onderwijsregisters. Daarnaast is van veel allochtonen die hun opleiding in het buitenland hebben gevolgd op dit moment (nog) onvoldoende betrouwbare registerinformatie beschikbaar. Dit betekent dat er voor het samenstellen van het Opleidingsniveaubestand een aanvullend beroep moet worden gedaan op de EBB. Anders dan bij de onderwijsregistraties zijn de gegevens van de EBB alleen op steekproefbasis beschikbaar. Er wordt dus gebruik gemaakt van gewichten.</p> <p>Bij leerlingen die 1 jaar staan ingeschreven, is van 66,7% van de moeders en 56,6% van de vaders het opleidingsniveau bekend. Bij leerlingen die 8 jaar staan ingeschreven, is van 51,5% van de moeders en 46,2% van de vaders is het opleidingsniveau bekend. Van leerlingen in groep 1 is het opleidingsniveau dekking van de ouders al 10-15% beter dan van leerlingen in groep 8.</p>
Bijzonderheden	De data over opleidingsniveau uit het COOL-onderzoek is nog niet beschikbaar bij CBS.

Indicator	5. Inkomen (vader, moeder, huishouden)
Definitie	<p><u>Persoonlijk</u>: Bruto-inkomen van een persoon uit de volgende bestandsdelen:</p> <ul style="list-style-type: none"> - inkomen uit arbeid - inkomen uit eigen onderneming - uitkering inkomensverzekeringen - uitkering sociale voorzieningen (m.u.v. kinderbijslag). <p><u>Huishouden</u>: Het bruto inkomen van het huishouden is gelijk aan het primair inkomen plus het overdrachtsinkomen. Voor alle personen in het huishouden worden het primair inkomen en het overdrachtsinkomen samengeteld. Vervolgens worden deze inkomens geaggregeerd op huishoudensniveau om zo het bruto inkomen van het huishouden te bepalen.</p>
Bron	<p>Persoonlijk inkomen bestand (vader en moeder apart).</p> <p>Huishoudinkomen bestand (huishouden op jaarbasis).</p> <p>Bestanden zijn gebaseerd op voornamelijk registers afkomstig van de Belastingdienst (fiscaal basisregister) en de bevolkingsregisters van gemeenten (GBA). Daarnaast worden gegevens verzameld uit de registraties van de studiefinanciering van de Dienst Uitvoering Onderwijs (DUO).</p>
Beschikbaarheid	<p>Bestand is beschikbaar vanaf 01-01-2003. Meest recente dataset is 2014 (voorlopig). Bestanden betreft personen die behoren tot de Nederlandse bevolking, ingeschreven in de gemeentelijke bevolkingsregisters, op 31 december van een onderzoeksjaar. In de bevolkingsregisters zijn niet opgenomen de in Nederland wonende personen waarvoor uitzonderingsregels gelden met betrekking tot opneming in de bevolkingsregisters (bijvoorbeeld asielzoekers, diplomaten en NAVO militairen) en personen die niet legaal in Nederland verblijven.</p>
Bijzonderheden	<p>Persoonlijk inkomen afgekapt op +/- 1 miljoen euro.</p>

Indicator	6.SEC Sociaal economische categorie ouders																								
Definitie	Sociaal economische categorie in een bepaalde maand																								
Bron	Meerdere inkomensbronnen																								
Beschikbaarheid	01-09-2005 tot 1-9-2014 (schooljaar 2013/2014). Alle personen die zijn ingeschreven in het GBA, en personen die niet in GBA zijn ingeschreven maar die wel een bekend onderwijsnummer of sofinummer hebben.																								
Detailniveau	<p>Uit te splitsen naar de volgende kenmerken:</p> <table border="1"> <tr><td>11</td><td>Werknemer</td></tr> <tr><td>12</td><td>Directeur-grotoaandeelhouder</td></tr> <tr><td>13</td><td>Zelfstandige</td></tr> <tr><td>14</td><td>Overig actief</td></tr> <tr><td>21</td><td>Ontvanger werkloosheidsuitkering</td></tr> <tr><td>22</td><td>Ontvanger bijstandsuitkering</td></tr> <tr><td>23</td><td>Ontvanger uitkering sociale voorz.overig</td></tr> <tr><td>24</td><td>Ontvanger uitkering ziekte/AO</td></tr> <tr><td>25</td><td>Ontvanger pensioenuitkering</td></tr> <tr><td>26</td><td>Nog niet schoolg./schol./stud. met ink.</td></tr> <tr><td>31</td><td>Nog niet schoolg./schol./stud. geen ink.</td></tr> <tr><td>32</td><td>Overig zonder inkomen</td></tr> </table>	11	Werknemer	12	Directeur-grotoaandeelhouder	13	Zelfstandige	14	Overig actief	21	Ontvanger werkloosheidsuitkering	22	Ontvanger bijstandsuitkering	23	Ontvanger uitkering sociale voorz.overig	24	Ontvanger uitkering ziekte/AO	25	Ontvanger pensioenuitkering	26	Nog niet schoolg./schol./stud. met ink.	31	Nog niet schoolg./schol./stud. geen ink.	32	Overig zonder inkomen
11	Werknemer																								
12	Directeur-grotoaandeelhouder																								
13	Zelfstandige																								
14	Overig actief																								
21	Ontvanger werkloosheidsuitkering																								
22	Ontvanger bijstandsuitkering																								
23	Ontvanger uitkering sociale voorz.overig																								
24	Ontvanger uitkering ziekte/AO																								
25	Ontvanger pensioenuitkering																								
26	Nog niet schoolg./schol./stud. met ink.																								
31	Nog niet schoolg./schol./stud. geen ink.																								
32	Overig zonder inkomen																								
Bijzonderheden	Geen bijzonderheden																								

Indicator	7. NSCT (Niet-Schoolse Cognitieve Capaciteiten Test) groep 5
Definitie	Toets die is bedoeld om leerkrachten op de basisscholen een objectieve second opinion te geven over de niet-schoolse capaciteiten van hun leerlingen.
Bron	COOL-bestand.
Beschikbaarheid	Vanaf schooljaar 2007/2008 tot schooljaar 2013/2014. Het bestand bevat alleen gegevens van groep-5-leerlingen van scholen die meedoen aan de steekproef. Het COOL-bestand bevat voor groep 5 gegevens van 9288 leerlingen in de referentiesteekproef en 12609 in de totale steekproef. In de referentiesteekproef heeft 95.1% van de leerlingen aan de testafname meegedaan en in de totale steekproef 90.1%.
Detailniveau	
Bijzonderheden	Variabele komt uit het COOL-onderzoek en CBS beschikt nog niet over deze data.

Indicator	8. Leeftijd moeder van leerling
Definitie	Leeftijd moeder op een bepaald tijdstip.
Bron	Gemeentelijke basisadministratie (GBA)
Beschikbaarheid	Vanaf 1-1-1995 tot en met 2014.
Bijzonderheden	Geen bijzonderheden

Indicator	9. Leeftijd leerling
Definitie	Indicatie leeftijd basisschoolleerling op een bepaald tijdstip.
Bron	Gemeentelijke basisadministratie (GBA)
Beschikbaarheid	Bestand is beschikbaar vanaf 1-10-1994. Meest recente versie 2015 (voorlopig), 2014 (definitief).
Bijzonderheden	Geen

Indicator	10. Geslacht leerling
Definitie	Geslacht basisschoolleerling.
Bron	Gemeentelijke basisadministratie (GBA)
Beschikbaarheid	Bestand is beschikbaar vanaf 1-10-1994. Meest recente versie 2015 (voorlopig), 2014 (definitief).
Detailniveau	Man/vrouw
Bijzonderheden	Geen

Indicator	11. Gescheiden ouders
Definitie	Burgerlijke staat: Formele positie van een persoon waarbij wordt verwezen naar het huwelijk en het geregistreerd partnerschap. Per 1 januari 1998 is het geregistreerd partnerschap ingevoerd.
Bron	Gemeentelijke basisadministratie (GBA) Vragenlijst COOL-onderzoek.
Beschikbaarheid	01-01-1995 tot en met 2014
Bijzonderheden	Geen bijzonderheden

Indicator	12. Overlijden ouders
Definitie	
Bron	Gemeentelijke basisadministratie (GBA)
Beschikbaarheid	01-01-1995 tot en met 2014
Bijzonderheden	Geen bijzonderheden

Indicator	13.Type huishouden
Definitie	Typering van een huishouden op basis van de onderlinge relaties van de personen binnen het huishouden.
Bron	Gemeentelijke basisadministratie (GBA)
Beschikbaarheid	Bestand is beschikbaar vanaf 1 oktober 1994. Meest recente versie 2014 (definitief).
Bijzonderheden	Geen bijzonderheden

Indicator	14.Herkomst (vader, moeder, kind)
Definitie	Land waarmee persoon verbonden is op basis van het geboorteland van de ouders of van zichzelf.
Bron	Gemeentelijke basisadministratie (GBA)
Beschikbaarheid	Bestand is beschikbaar vanaf 1-10-1994. Meest recente versie 2015 (voorlopig), 2014 (definitief).
Bijzonderheden	De publicatie van landen moet volgens een standaard landenpublicatielijst plaatsvinden. Deze lijst bevat per publicatiejaar de actuele landen. De geboortelandencodes moeten worden gehercodeerd naar deze landenlijst, met een referentiebestand (bij CBS aanwezig).

Indicator	15.Vluchteling kind
Definitie	Indicatie of een basisschoolleerling een vluchtelingkind is.
Bron	Gemeentelijke basisadministratie (GBA)
Operationalisering	Indien uitgegaan wordt van niet in GBA ingeschreven maar wel in INSCHRWPO (basisschool, doorgaans kinderen tot en met 12 jaar) en het land van herkomst/geboorteland tot een oorlogsgebied behoort, dan mag men deze leerling als vluchteling beschouwen. Het percentage kinderen in het basisonderwijs dat niet in het GBA staat ingeschreven wisselt per jaar. Ter indicatie: in het jaar 2013 komt 99,6% van de basisschoolleerlingen voor in het GBA.
Beschikbaarheid	Bestand is beschikbaar vanaf 1-10-1994. Meest recente versie 2015 (voorlopig), 2014 (definitief).
Samenstellen indicator	Op basis van GBA gegevens land van herkomst/geboorteland
Detailniveau	Indeling naar landen
Bijzonderheden	Geen

Indicator	16.Generatie
Definitie	Tot welke generatie (allochtoon) behoort de basisschoolleerling.
Bron	Gemeentelijke basisadministratie (GBA)
Beschikbaarheid	01-09-2005 tot 1-9-2014 (schooljaar 2013/2014)
Bijzonderheden	<p>Autochtoon: Persoon van wie de beide ouders in Nederland zijn geboren.</p> <p>Eerstegeneratieallochtoon: Persoon die in het buitenland is geboren met ten minste één in het buitenland geboren ouder.</p> <p>Tweedegeneratieallochtoon: Persoon die in Nederland is geboren met ten minste één in het buitenland geboren ouder.</p> <p>Er is ook de derde generatie; persoon in Nederland geboren maar ten minste één in het buitenland geboren grootouder. Derde generatie valt wel onder autochtoon.</p>

Indicator	17.Verblijfsduur (vader, moeder, kind)
Definitie	De periode dat een persoon, die niet in Nederland is geboren, in Nederland staat ingeschreven bij de gemeentelijke basisadministratie (GBA).
Bron	Gemeentelijke basisadministratie (GBA), versie met migratie kenmerken
Beschikbaarheid	01-09-2005 tot 1-9-2014 (schooljaar 2013/2014)
Bijzonderheden	Geen bijzonderheden

Indicator	18.Verblijfsstatus
Definitie	In de bevolkingsregisters van Nederlandse gemeenten ingeschreven immigranten met een buitenlandse nationaliteit.
Bron	In de bevolkingsregisters van Nederlandse gemeenten ingeschreven immigranten met een buitenlandse nationaliteit. Afkomstig van gegevens van de Immigratie- en Naturalisatiedienst (IND) van het ministerie van Veiligheid en Justitie. Tot en met 2003 was deze informatie ontleend aan het Centraal Register Vreemdelingen (CRV) van de IND, een registratie met gegevens over alle in Nederland wonende niet-Nederlanders. Vanaf 2004 zijn de gegevens ontleend aan INDIS, het informatiesysteem van de IND.
Beschikbaarheid	<p>01-09-1999 tot en met 2012.</p> <p>Van een aantal niet-Nederlandse immigranten is het migratiemotief van de IND niet bekend of niet ingevuld. Deze 'onbekende motieven' worden bijgeschat op basis van de verdeling van de bekende motieven. Bij deze schatting is uitgegaan van de variabelen die het sterkst samenhangen met het migratiemotief, te weten geslacht, leeftijd, burgerlijke staat, het jaar van vestiging in Nederland en</p>

	geboorteland.
Bijzonderheden	Geen bijzonderheden

Indicator	19.Thuistaal
Definitie	De taal/talen die een basisschoolleerling thuis spreekt.
Bron	Eindtoets basisonderwijs Cito
Beschikbaarheid	01-09-2005 tot 2014 (definitief) Voor de jaren 2012, 2013 en 2015 niet gemeten, 2014 niet betrouwbaar.
Bijzonderheden	Het bestand bevat alleen gegevens van leerlingen op scholen die toestemming hebben gegeven aan Cito om de data te leveren aan CBS. Gebruik van de Cito data door CBS kan in overleg met Cito.

Indicator	20.Taalbeheersing ouders
Definitie	De taal/talen beheersing van de ouders van een basisschoolleerling.
Bron	COOL-bestand (vragenlijst)
Beschikbaarheid	Vanaf schooljaar 2007/2008 tot schooljaar 2013/2014. Het bestand bevat alleen gegevens van ouders van leerlingen op scholen die meedoen met de steekproef en die de vragenlijst hebben ingevuld.
Bijzonderheden	De data van het COOL-project zijn nog niet beschikbaar bij CBS.

Indicator	21.Schuldsanering ouders
Definitie	Betrokkenheid ouders van de leerlingen in het basisonderwijs bij een schuldsanering in een bepaald jaar.
Bron	Register waarin opgenomen alle natuurlijk personen die een bepaald jaar in de schuldsanering hebben gezeten.
Beschikbaarheid	Vanaf 01-01-1998 tot en met 2014
Bijzonderheden	Geen bijzonderheden

Indicator	22. Ouders ooit verdacht van misdrijf
Definitie	Indicator of ouders verdachte zijn geweest bij een misdrijf waarvoor een proces-verbaal is opgemaakt.
Bron	Register herkenningssysteem (HKS) van de politie
Beschikbaarheid	Vanaf 1-1-1996 tot en met 2014
Bijzonderheden	Toestemming voor deze data moet CBS intern worden verkregen.

Gezondheid

Indicator	23.Indicatie leerlinggebonden financiering (LGF)																																				
Definitie	Leerlingen met een indicatie binnen het basisonderwijs																																				
Bron	Inschrijvingen in het basisonderwijs																																				
Beschikbaarheid	01-09-2008 tot 1-9-2014. Dit bestand bevat alleen leerlingen die in een schooljaar (speciaal) basisonderwijs volgen, en niet leerlingen die in een schooljaar uitsluitend speciaal onderwijs volgen. De bestanden over 2008/'09 en 2009/'10 zijn niet volledig gevuld. Vanaf 2010/'11 worden de scholen bekostigd op basis van de onderwijsnummerbestanden en zijn de bestanden volledig.																																				
Detailniveau	<table border="1"><tr><td>00</td><td>Leerling heeft geen LGF</td></tr><tr><td>11</td><td>onderwijs aan visueel gehandicapte leerlingen</td></tr><tr><td>12</td><td>onderwijs aan visueel gehandicapte leerlingen die tevens zeer moeilijk lerend zijn</td></tr><tr><td>21</td><td>onderwijs aan dove kinderen</td></tr><tr><td>22</td><td>onderwijs aan slechthorende kinderen</td></tr><tr><td>23</td><td>onderwijs aan kinderen met ernstige spraakmoeilijkheden</td></tr><tr><td>24</td><td>onderwijs aan dove kinderen die tevens zeer moeilijk lerend zijn</td></tr><tr><td>25</td><td>onderwijs aan slechthorende kinderen die tevens zeer moeilijk lerend zijn</td></tr><tr><td>26</td><td>Doof/visueel gehandicapt [WEC]</td></tr><tr><td>31</td><td>onderwijs aan langdurig zieke kinderen met een lichamelijke handicap</td></tr><tr><td>32</td><td>onderwijs aan lichamelijk gehandicapte kinderen</td></tr><tr><td>33</td><td>onderwijs aan zeer moeilijk lerende kinderen</td></tr><tr><td>34</td><td>onderwijs aan kinderen met een lichamelijke handicap die tevens zeer moeilijk lerend zijn</td></tr><tr><td>35</td><td>Syndroom van Down</td></tr><tr><td>40</td><td>cluster 4: onderwijs aan langdurig zieke kinderen anders dan met een lichamelijke handicap of aan zeer moeilijk opvoedbare kinderen of kinderen in scholen verbonden aan pedologische instituten</td></tr><tr><td>44</td><td>Justitiële Jeugdinstelling</td></tr><tr><td>13</td><td>Blind</td></tr><tr><td>14</td><td>Slechtziend</td></tr></table>	00	Leerling heeft geen LGF	11	onderwijs aan visueel gehandicapte leerlingen	12	onderwijs aan visueel gehandicapte leerlingen die tevens zeer moeilijk lerend zijn	21	onderwijs aan dove kinderen	22	onderwijs aan slechthorende kinderen	23	onderwijs aan kinderen met ernstige spraakmoeilijkheden	24	onderwijs aan dove kinderen die tevens zeer moeilijk lerend zijn	25	onderwijs aan slechthorende kinderen die tevens zeer moeilijk lerend zijn	26	Doof/visueel gehandicapt [WEC]	31	onderwijs aan langdurig zieke kinderen met een lichamelijke handicap	32	onderwijs aan lichamelijk gehandicapte kinderen	33	onderwijs aan zeer moeilijk lerende kinderen	34	onderwijs aan kinderen met een lichamelijke handicap die tevens zeer moeilijk lerend zijn	35	Syndroom van Down	40	cluster 4: onderwijs aan langdurig zieke kinderen anders dan met een lichamelijke handicap of aan zeer moeilijk opvoedbare kinderen of kinderen in scholen verbonden aan pedologische instituten	44	Justitiële Jeugdinstelling	13	Blind	14	Slechtziend
00	Leerling heeft geen LGF																																				
11	onderwijs aan visueel gehandicapte leerlingen																																				
12	onderwijs aan visueel gehandicapte leerlingen die tevens zeer moeilijk lerend zijn																																				
21	onderwijs aan dove kinderen																																				
22	onderwijs aan slechthorende kinderen																																				
23	onderwijs aan kinderen met ernstige spraakmoeilijkheden																																				
24	onderwijs aan dove kinderen die tevens zeer moeilijk lerend zijn																																				
25	onderwijs aan slechthorende kinderen die tevens zeer moeilijk lerend zijn																																				
26	Doof/visueel gehandicapt [WEC]																																				
31	onderwijs aan langdurig zieke kinderen met een lichamelijke handicap																																				
32	onderwijs aan lichamelijk gehandicapte kinderen																																				
33	onderwijs aan zeer moeilijk lerende kinderen																																				
34	onderwijs aan kinderen met een lichamelijke handicap die tevens zeer moeilijk lerend zijn																																				
35	Syndroom van Down																																				
40	cluster 4: onderwijs aan langdurig zieke kinderen anders dan met een lichamelijke handicap of aan zeer moeilijk opvoedbare kinderen of kinderen in scholen verbonden aan pedologische instituten																																				
44	Justitiële Jeugdinstelling																																				
13	Blind																																				
14	Slechtziend																																				
Bijzonderheden	Geen bijzonderheden																																				

Indicator	24. Ontvangen jeugdzorg (leerling)
Definitie	Ontvangen zorg in het kader van de Wet op de jeugdzorg (t/m 2014) resp. de Jeugdwet (vanaf 2015)
Bron	Registraties door Bureaus Jeugdzorg en jeugdzorgaanbieders (t/m 2014) resp. organisaties Jeugd en Opvoedhulp en gecertificeerde instellingen (vanaf 2015)
Beschikbaarheid	Wet op de jeugdzorg 1-1-2011 tot en met 2014; Jeugdwet eerste half jaar 2015
Detailniveau	Jeugdhulp, jeugdbescherming, jeugdreclassering.
Bijzonderheden	Geen bijzonderheden

Indicator	25. Gebruikte zorg onder basisverzekering Zvw (leerling, ouders)
Definitie	Door de basisverzekering van de Zorgverzekeringswet (Zvw) vergoede zorg
Bron	Registratie door Vektis C.V.
Beschikbaarheid	2009 tot en met 2013
Detailniveau	Per zorgvorm
Bijzonderheden	Per jaar wordt de inhoud van het basispakket, hetgeen vergoed wordt in het kader van de basisverzekering, door de politiek vastgesteld.

Indicator	26. Verstrekte geneesmiddelen (leerling, ouders)
Definitie	Verstrekte geneesmiddelen (vergoed door de Zvw); excl. verstrekkingen in ziekenhuizen en verpleeghuizen
Bron	Registratie door Zorginstituut Nederland
Beschikbaarheid	2006 tot en met 2013
Detailniveau	Naar ATC-code, eerste 4 digits
Bijzonderheden	Geen bijzonderheden

Indicator	27. Ziekenhuisopnamen (leerling, ouders)
Definitie	Ziekenhuisopnamen en diagnoses
Bron	Registraties door DHD
Beschikbaarheid	1995 tot en met 2012
Detailniveau	Diagnoses op gedetailleerd ICD-niveau
Bijzonderheden	Medio 2016 komen ziekenhuisopnamen van 2013 en 2014 beschikbaar. In de loop van 2016 komen daarnaast DBC's voor

medisch specialistische zorg en GGZ beschikbaar.

Indicator	28. Indicatie / gebruik langdurige zorg (leerling, ouders)
Definitie	Indicatie voor / gebruik van AWBZ-zorg (t/m 2014) resp. ontvangen zorg Wlz (vanaf 2015)
Bron	Registraties door CIZ, BIZ's en CAK en Vektis (t/m 2014) resp. SVB
Beschikbaarheid	AWBZ-zorg 2009 tot en met 2014
Detailniveau	AWBZ-zorg: grondslag zorgfuncties (zorg zonderverblijf) en zorgzwaartepakketten (zorg met verblijf)
Bijzonderheden	Gebruik van AWBZ-zorg alleen vanaf 18 jaar. Ook gegevens beschikbaar over Wmo-zorg. Gegevens over Wlz-zorg 2015 komen naar verwachting in 2016 beschikbaar; waarschijnlijk is de registratie dan nog niet compleet.

Indicator	29. Melding kindermishandeling (leerling)
Definitie	Melding van kindermishandeling
Bron	Registratie door AMK's
Beschikbaarheid	2011 tot en met 2014
Detailniveau	Individuele meldingen
Bijzonderheden	Deze gegevens zijn recent beschikbaar gekomen en naar verwachting begin 2016 bruikbaar voor onderzoek.

Kenmerken van de buurt

Indicator	30.Armoede
Definitie	Gemiddeld inkomen huishoudens op buurtniveau
Bron	De Inkomensstatistiek. De Inkomensstatistiek maakt gebruik van gegevens die ontleend zijn aan administraties, waarvan de belangrijkste dataleverancier de Belastingdienst is
Beschikbaarheid	Bestand is beschikbaar vanaf 01-01-2003. Meest recente dataset is 2014 (voorlopig). Bestanden betreft personen die behoren tot de Nederlandse bevolking, ingeschreven in de gemeentelijke bevolkingsregisters, op 31 december van een onderzoeksjaar. In de bevolkingsregisters zijn niet opgenomen de in Nederland wonende personen waarvoor uitzonderingsregels gelden met betrekking tot opname in de bevolkingsregisters (bijvoorbeeld asielzoekers, diplomaten en NAVO militairen) en personen die niet legaal in Nederland verblijven. Gemiddeld persoonlijk inkomen van een buurt kan afgeleid worden van de persoonlijke inkomens.
Bijzonderheden	De Inkomensstatistiek maakt gebruik van gegevens die ontleend zijn aan administraties, waarvan de belangrijkste dataleverancier de Belastingdienst is.

Indicator	31.Woningkenmerken
Definitie	Kenmerken van de woning van de leerling (koop of huur, aantal kamers, type woning, vierkante meters, etc.)
Bron	Registers woningvoorraad, type woning en WOZ-waarde
Beschikbaarheid	Registers zijn gelijktijdig beschikbaar van 1-1-1999 tot en met 2011.
Detailniveau	Bouwjaar, oppervlakte, status verblijfsobject, type woning, WOZ-waarde, aantal kamers, soortwoonruimte etc.
Bijzonderheden	Geen bijzonderheden

Indicator	32.Cultureel aanbod
Definitie	Cultureel aanbod in de nabijheid van de woning van de leerling.
Bron	Register over afstand van woning tot cultuur en vrijetijdsbesteding. De voorzieningen betreffen bibliotheken, musea, podiumkunsten, poppodia, bioscopen, attracties, sauna, zwembaden, kunstijsbanen, zonnebanken.
Beschikbaarheid	Geldig vanaf 1-1-2006. Meest recente versie 2012 (definitief). Voor minder dan 0.1% van de bewoonde objecten is geen afstand opgenomen. Het betreft hier de afstand vanaf alle objecten met herkomst BAG of historisch objecten, voor zover het bewoonde objecten betreft volgens de GBA.

Bijzonderheden	Geen bijzonderheden
----------------	---------------------

Indicator	33.Stedelijkheid
Definitie	Indeling van gemeenten naar stedelijkheid is gebaseerd op de omgeving-adressendichtheid van de gemeente. De vijf stedelijkheidsklassen zijn gebaseerd op klassegrenzen van 2500, 1500, 1000 en 500 adressen per km ² .
Bron	Registers regionale indelingen
Beschikbaarheid	Vanaf 2010
Bijzonderheden	Geen bijzonderheden

Indicator	34.Veiligheid
Definitie	Aantal incidenten in een buurt
Bron	Incidentenbestanden afkomstig van politieregistraties.
Beschikbaarheid	Vanaf 1-1-2005 tot en met 2012. Het bestand betreft alle personen die slachtoffer zijn geworden van een misdrijf dat in het betreffende jaar is gepleegd en dat op enig moment bij de politie is gemeld.
Detailniveau	Op persoonsniveau zijn de incidenten geregistreerd, aggregatie naar buurt niveau is mogelijk.
Bijzonderheden	Plek van het misdrijf is niet per definitie in de buurt van de verblijfplaats van de persoon.

Schoolkenmerken

Indicator	35.Groepsgrootte
Definitie	Groepsgrootte van het basisonderwijs
Bron	Inschrijvingen in het basisonderwijs.
Beschikbaarheid	01-09-2008 tot 1-9-2014. Van alle leerlingen in het basisonderwijs wordt op basis van registratie de groepsgrootte meegenomen.
Bijzonderheden	Geen bijzonderheden

Indicator	36.Cumulatie
Definitie	Samenvattende indicatoren per school, zoals gemiddeld inkomen van de ouders, gemiddeld opleidingsniveau, etc.
Bron	
Beschikbaarheid	
Bijzonderheden	Geen bijzonderheden