

University of Groningen

Vernieuwing van indicatoren voor leeropbrengsten in het voortgezet onderwijs

Rekers-Mombarg, L.T.M.; Timmermans, A.C.; Bosker, R.J.

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:

2014

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Rekers-Mombarg, L.T. M., Timmermans, A. C., & Bosker, R. J. (2014). *Vernieuwing van indicatoren voor leeropbrengsten in het voortgezet onderwijs: Een dieptestudie*. GION onderzoek/onderwijs.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Vernieuwing van indicatoren voor leeropbrengsten in het voortgezet onderwijs; een dieptestudie

L.T.M. Rekers-Mombarg

A.C. Timmermans

R.J. Bosker

Vernieuwing van indicatoren voor leeropbrengsten in het voortgezet onderwijs; een dieptestudie
L.T.M. Rekers-Mombarg A.C. Timmermans R.J. Bosker

Vernieuwing van indicatoren voor leeropbrengsten in het voortgezet onderwijs; een dieptestudie

L.T.M. Rekers-Mombarg, A.C. Timmermans, R.J. Bosker
GION onderwijs/onderzoek
Rijksuniversiteit Groningen
Juni 2014

ISBN: 978-90-6690-544-3

2014. GION onderwijs/onderzoek
Rijksuniversiteit, Grote Rozenstraat 3, 9712 TG Groningen

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook zonder voorafgaande schriftelijke toestemming van de directeur van het instituut.

No part of this book may be reproduced in any form, by print, foto print, microfilm or any other means without written permission of the director of the institute.

Inhoud

Managementsamenvatting.....	1
1 Inleiding.....	9
1.1 Onderzoeksvragen	11
1.2 Leeswijzer.....	14
2 Achtergrond.....	15
2.1 Definiëring van leerwinst en toegevoegde waarde.....	15
2.2 Conceptuele aspecten van toegevoegde waardebe­paling.....	16
2.3 Toegevoegde waardebe­paling voor toezichtdoeleinden	17
3 Deel I: Validering van beoogde opbrengstenmodel	21
3.1 Inleiding.....	21
3.2 Methode.....	24
3.2.1 Beschrijving van de dataset	24
3.2.2 Berekening bovenbouwrendement.....	25
3.2.3 Beschrijving van modellen en series analyses.....	26
3.2.4 Vergelijking tussen de modellen.....	28
3.2.5 Consistentie van centraal examencijfers voor de vak­kenclusters.....	29
3.3 Resultaten	29
3.3.1 Invloed van ontbrekende data op modellen.....	29
3.3.2 Vergelijking van modellen voor bovenbouwrendement	32
3.3.3 Vergelijking van modellen voor centraal examencijfer.....	34
3.3.4 Consistentie van de centraal examencijfers	36
4 Deel II: Onderbouwrendement	39
4.1 Inleiding.....	39
4.2 Methode.....	41
4.2.1 Huidige berekeningswijze onderbouwrendement	41
4.2.2 Retrospectief onderbouwrendement met individuele data	44
4.2.3 Prospectief onderbouwrendement met individuele data.....	46
4.2.4 Invloed van PO-school op het prospectieve onderbouwrendement.....	49
4.2.5 Compensatie voor lwoo-leerlingen bij bepaling van onderbouwrendement	52
4.3 Resultaten	53
4.3.1 Retrospectief onderbouwrendement met individuele data	53
4.3.2 Prospectief onderbouwrendement met individuele data.....	53
4.3.3 Retrospectief versus prospectief onderbouwrendement met individuele data.....	56
4.3.4 Invloed PO-school op prospectief onderbouwrendement.....	60
4.3.5 Bodem- en plafondeffecten in relatie tot vertekening van het onderbouwrendement	65

4.3.6	Doubleren van lwoo-leerlingen	67
5	Deel III: toegevoegde waarde van vmbo-tl naar havo.....	69
5.1	Inleiding	69
5.2	Methode	70
5.2.1	Beschrijving van de dataset.....	70
5.2.2	Toegevoegde waarde	71
5.2.3	Doorstroomrendement	73
5.3	Resultaten.....	74
5.3.1	Samenhang tussen beginniveau, achtergrondkenmerken en examencijfers havo.....	74
5.3.2	Toegevoegde waarde van havo-afdelingen voor gemiddeld cijfer van doorstromers.....	76
5.3.3	Toegevoegde waarde van havo-afdelingen voor Nederlands van doorstromers.....	79
5.3.4	Toegevoegde waarde van havo-afdelingen voor Engels van doorstromers	82
5.3.5	Toegevoegde waarde van havo-afdelingen voor wiskunde van doorstromers	84
5.3.6	Samenhang tussen toegevoegde waarde gemiddeld cijfer, Nederlands, Engels en wiskunde van havo-afdelingen	86
5.3.7	Doorstroomrendement van havo-afdelingen.....	87
5.3.8	Samenhang tussen toegevoegde waarde en doorstroomrendement van havo-afdelingen.....	88
6	Deel IV: Extern rendement vmbo-kl en bl	91
6.1	Inleiding	91
6.1.1	Onderzoeken van lange termijn effecten	92
6.2	Methode	94
6.2.1	Beschrijving van de dataset.....	94
6.2.2	Extern rendement	94
6.2.3	Multilevel en kruisklassificatie modellen voor extern rendement	96
6.2.4	Correctie voor controle variabelen.....	97
6.3	Resultaten.....	98
6.3.1	Extern rendement volgens huidige rekenregels een jaar na examenklas vmbo-bl of kl.....	98
6.3.2	Extern rendement volgens huidige rekenregels twee jaar na examenklas vmbo-bl of kl.....	101
6.3.3	Isoleren van VO-vestigingseffect in extern rendement een jaar na vmbo-bl of kl	102
6.3.4	Isoleren van VO-vestigingseffect in extern rendement twee jaar na vmbo-bl of kl	104
6.3.5	Vergelijking van extern rendement van VO-vestigingen voor de drie varianten van extern rendement	106
6.3.6	Vergelijking van huidige indicator met meerniveau modellen op basis van percentiel scores 109	
6.3.7	Effecten van gebruik van controle variabelen op extern rendement.....	111
7	Conclusie en discussie.....	115
	Deelproject I: Validering van beoogd opbrengstenmodel.....	115
	Deelproject II: Rendement onderbouw VO.....	118

Deelproject III: Toegevoegde waarde van bovenbouw VO.....	123
Deelproject IV: Extern rendement van VMBO basis- en kaderberoepsgerichte leerweg.....	126
Derde doelstelling: verkenning van meer differentiatie in toezicht.....	128
Literatuur.....	131
Appendix bij hoofdstuk 3.....	135
Appendix bij hoofdstuk 5.....	141
Appendix bij hoofdstuk 6.....	147

Managementsamenvatting

Aanleiding

Sinds 2012 werkt de Inspectie van het Onderwijs aan het vernieuwen van het opbrengstenmodel voor het VO. In overleg met het onderwijsveld is een begin gemaakt met de ontwikkeling van opbrengstindicatoren waarmee een betere indicatie van de toegevoegde waarde van VO-scholen verkregen kan worden dan nu het geval is. Als basis voor deze opbrengstindicatoren kunnen landelijke databestanden van individuele leerlingen dienen. Verder wil men de nieuwe opbrengstindicatoren voorzien van absolute normen, waarbij in het geval van een ongunstige leerlingpopulatie deze norm naar beneden kan worden bijgesteld. De aandacht voor het ontwikkelen van toegevoegde waarde indicatoren komt voort uit het regeerakkoord van Rutte I en Rutte II. De regering wil toewerken naar een kwaliteitsnorm voor scholen, waarbij een toegevoegde waarde indicator zwaarder mee zal wegen en bij de beoordeling van scholen zal op termijn meer gelet worden op toegevoegde waarde.

Het begrip toegevoegde waarde heeft in het onderwijsveld geen eenduidige betekenis. In het huidige onderzoek wordt toegevoegde waarde gedefinieerd als de bijdrage van een school aan de eindprestaties van leerlingen, gegeven het aanvangsniveau van leerlingen. Daarbij is het van belang om zo veel mogelijk rekening te houden met kenmerken die de eindprestaties van leerlingen wel beïnvloeden maar waar de school geen invloed op heeft, zoals sociaal milieu van de leerling, samenstelling van de schoolpopulatie of de wijk waarin de school staat. Een goede meting van het beginniveau van leerlingen is essentieel. Het probleem van de Inspectie van het Onderwijs is dat ze meestal niet kan beschikken over een goede meting van het beginniveau van leerlingen in het voortgezet onderwijs. In het huidige opbrengstenmodel wordt daarom gewerkt met schoolgemiddelde proxies van het beginniveau (sociaal economische status, leerwegondersteuning en armoedeprobleemcumulatiegebied). Dit is een suboptimale oplossing die ongewenste consequenties kan hebben als VO-scholen onderling vergeleken gaan worden. In het beoogde opbrengstenmodel wordt gewerkt met individuele data van leerlingen. Hiermee kan potentieel het beginniveau van leerlingen beter geschat worden. Onduidelijk of met het beoogde model een goede indicatie van de toegevoegde waarde van een school kan worden verkregen.

Doelen

Er zijn drie doelen geformuleerd voor dit onderzoeksproject. Ten eerste is dat de validering van het beoogde opbrengstenmodel aan de hand van databestanden van het grootschalige VOCL'99 cohortonderzoek. Op basis van gedetailleerde informatie omtrent beginniveau wordt een toegevoegde waarde model gemaakt, gegeven de mogelijkheden die het VOCL'99 onderzoek daarvoor biedt. Hiermee kunnen de huidige en beoogde opbrengstindicatoren

vergeleken worden. Het valideringsonderzoek richt zich op twee opbrengstindicatoren, te weten het bovenbouwrendement en het gemiddelde centraal examencijfer.

Het tweede doel van dit project is het op landelijke schaal uitproberen van een beoogde nieuwe maat voor het onderbouwrendement en de ontwikkeling van toegevoegde waarde bepalingen van VO-scholen voor twee specifieke groepen leerlingen in het VO, namelijk de leerlingen die vanuit vmbo theoretische leerweg doorstromen naar havo en de leerlingen die vanuit vmbo basis- en kaderberoepsgerichte leerweg doorstromen naar het mbo. Het gaat hier om respectievelijk het doorstroomrendement en het externe rendement. De vernieuwing voor het onderbouwrendement betreft niet alleen in de beschikbaarheid van individuele data, maar ook de mogelijkheid om de berekeningen prospectief uit te voeren (uitgaan van leerlingen die in het eerste leerjaar van het VO zitten en deze twee jaar lang volgen in hun schoolloopbaan) in plaats van de huidige retrospectieve berekeningswijze (uitgaan van leerlingen in het derde leerjaar van het VO en informatie verzamelen over hun VO-schoolloopbaan tot nu toe).

Het eerste en het tweede doel zijn in vier verschillende deelprojecten onderzocht. Het derde doel is het verkennen van de mogelijkheden of de verkregen informatie uit de grootschalige berekeningen ook ingezet kan worden voor een meer gedifferentieerd toezicht. De Inspectie van het Onderwijs streeft naar het maken van een onderscheid in drie groepen scholen: scholen met zwakke opbrengsten, scholen met voldoende opbrengsten maar met mogelijkheden tot verbetering en scholen met goede opbrengsten.

Conclusies deelproject 1

In deelproject 1 is de validering van het beoogde opbrengstenmodel met VOCL'99 data aan de orde. Er worden vier modellen onderling vergeleken door ze in een spectrum te plaatsen. Het ene uiteinde van het spectrum wordt gevormd door een model zonder correctie van het beginniveau en het andere uiteinde door een model waarin zo optimaal mogelijk voor het beginniveau en relevante achtergrondkenmerken van leerlingen is gecorrigeerd. Dit laatste wordt gedaan door de Cito-eindtoetsscore, etniciteit, sociaal economische status en sekse van de leerling in het model op te nemen. Omdat in dit optimale model zowel een goede meting van het beginniveau als relevante achtergrondkenmerken van leerlingen is opgenomen, is het optimale model te beschouwen als een model waarmee de toegevoegde waarde van een school kan worden bepaald. Onderzocht is waar het huidige en beoogde opbrengstenmodel geplaatst moeten worden in dit spectrum. In zowel het huidige als het beoogde model worden proxies voor het beginniveau van leerlingen opgenomen, zoals instroom in het derde leerjaar VO, armoedeprobleemcumulatiegebied (apcg) en leerwegondersteuning (lwoo). Het verschil tussen beide modellen is het niveau waarop deze proxies worden meegenomen; leerlingniveau of schoolniveau. Voor de bepaling van het bovenbouwrendement lijken de resultaten van de vier modellen sterk op elkaar. Desondanks is het beoogde model een betere keuze dan het huidige model, omdat het huidige model methodologisch minder sterk en de samenhang met de andere modellen het zwakst is. Voor het gemiddelde centraal examencijfer voor alle

vakken zijn wel duidelijke verschillen tussen de modellen gevonden. Het huidige model lijkt sterk op een model zonder correctie en helemaal niet op het optimale model. Het beoogde model neemt een positie in tussen het huidige en optimale model, maar staat toch vrij ver af van het optimale model. Ook voor het gemiddelde centraal examencijfer kan het beste voor het beoogde model gekozen worden, maar men moet zich bedenken dat het hooguit een indicatie is van de toegevoegde waarde bepaling van een VO-vestiging is.

Het streven naar toepassing van het optimale model uit dit project in de toezichtpraktijk lijkt vooralsnog minder belangrijk voor het bovenbouwrendement, dan voor het gemiddelde centraal examencijfer. Voor een goede benadering van de toegevoegde waarde van een VO-vestiging moeten de belangrijkste indicatoren voor individuele leerprestaties bij aanvang in ieder geval meegenomen worden in de modellen. Van de toegepaste correctiefactoren in het optimale model is de Cito-eindtoetsscore de belangrijkste. Dit kenmerk is inmiddels voor een ruime meerderheid van de Nederlandse leerlingen beschikbaar in landelijke onderwijsnummer-bestanden waar de Inspectie gebruik van kan maken. Hoewel veel leerlingen in de landelijke onderwijsnummer-bestanden nu nog geen eindexamen hebben gedaan, biedt dit wel perspectief voor de toekomst. Gebleken is verder dat het potentiële leervermogen (IQ) – een andere mogelijke belangrijke maat voor het aanvangsniveau – niet noodzakelijker wijze in het optimale model hoeft te worden opgenomen. Wat dit betreft kan men met de Cito-eindtoetsscore volstaan. Hiermee komt de haalbaarheid van het optimale model een stukje dichterbij.

Conclusies deelproject 2

In deelproject 2 is op twee manieren het onderbouwrendement van een vestiging berekend met individuele data van de meeste recente Bron-bestanden. De retrospectieve variant volgt zo veel mogelijk de huidige werkwijze van de Inspectie van het Onderwijs. De prospectieve variant zou de Inspectie van het Onderwijs in de toekomst toe kunnen passen, omdat inmiddels in de landelijke databestanden voldoende informatie van individuele leerlingen beschikbaar is. Een prospectieve werkwijze heeft in principe de voorkeur omdat de definitie van het begrip onderbouwrendement – het deel van de leerlingen in de eerste klas VO dat volgens verwachting doorstroomt naar de derde klas VO - een dergelijk werkwijze suggereert.

Met de prospectieve werkwijze kan voor 75% van de VO-vestigingen het onderbouwrendement voldoende betrouwbaar berekend worden. Dat is 5% hoger dan voor de retrospectieve werkwijze. De samenhang tussen de twee varianten is vrij hoog, maar er is sprake van een klein, maar systematisch verschil. Op het vmbo basis- en kaderberoepsgerichte leerwegen zijn deze verschillen het grootst. Waarschijnlijk is de oorzaak gelegen in het algemene gegeven dat er meer leerlingen afstromen naar een lagere schoolsoort dan er leerlingen opstromen naar een hogere schoolsoort. Op het vmbo basis en kader is de afstroom van leerlingen beduidend meer aan de orde dan de opstroom, hetgeen tot gevolg heeft het retrospectieve onderbouwrendement – waarin ook afstromende leerlingen meegenomen

worden – negatief beïnvloed wordt. Dit kan leiden tot het ongewenst effect dat vmbo basis en kader vestigingen worden ‘afgerekend’ op het feit dat ze afstromende vmbo theoretische of gemengde leerweg leerlingen een plek geven op hun school. Op grond van bovenstaande bevindingen is geconcludeerd dat een prospectieve berekening van het onderbouwrendement te verkiezen is boven een retrospectieve berekening.

Verder is nog onderzocht in dit deelproject of mogelijke lange termijneffecten van toeleverende basisscholen van invloed zijn op het prospectieve onderbouwrendement van een VO-vestiging. Dit is inderdaad het geval. Het klein deel van de verschillen in onderbouwrendement tussen leerlingen is toe te schrijven aan de aan de PO-vestiging waar ze vandaan komen. Dit aandeel is echter ongeveer even groot als het aandeel van de VO-vestiging die ze bezoeken. Dit impliceert dat de Inspectie van het Onderwijs in het beoogde opbrengstenmodel voor het onderbouwrendement niet alleen rekening zou moeten houden met de VO-vestigingen van de leerlingen, maar ook met de toeleverende PO-vestigingen.

Een ander mogelijk basisschooleffect is dat VO-scholen ‘last’ of ‘profijt’ hebben van basisscholen die structureel respectievelijk te hoge of te lage adviezen afgeven. Dit leidt tot een foutieve indicatie van het instroomniveau van leerlingen en daarmee tot een vertekening van het onderbouwrendement. Karakterisering van VO-vestigingen die ‘last’ hebben van toeleverende basisscholen leert dat dit scholen zijn die in ieder geval een vwo-afdeling hebben en waar vrij veel doublures voorkomen. Profiterende VO-vestigingen zijn vooral vmbo basis en kader vestigingen en vmbo breed-havo-vestigingen waar niet of nauwelijks doublures voorkomen en relatief veel lwoo-leerlingen zitten.

Er zijn door de Inspectie van het Onderwijs nog vier losse aspecten naar voren gebracht die ook bij het vernieuwen van het opbrengstenmodel voor het onderbouwrendement aan de orde zijn en die nadere bestudering behoeven. Eén daarvan is de normering van het onderbouwrendement. In het huidige opbrengstenmodel worden hiervoor groepen van vrijwel gelijke schoolsoorten gebruikt. Het verdient aanbeveling om dergelijke vergelijkingsgroepen ook bij het beoogde opbrengstenmodel te blijven gebruiken. Daar zijn twee belangrijke redenen voor. Ten eerste is de invloed van bodem- en plafondeffecten bij het berekenen van het onderbouwrendement nooit uit te sluiten. Het plafondeffect ontstaat doordat leerlingen met een vwo-advies alleen maar overeenkomstig hun advies kunnen doorstromen of afstromen naar een lagere schoolsoort. Bij een bodemeffect is het omgekeerde aan de hand: leerlingen met een vmbo praktijkonderwijs-advies kunnen alleen maar overeenkomstig hun advies doorstromen of opstromen naar een hogere schoolsoort. Deze vormen van vertekening zijn alleen maar onder controle te houden door vergelijkingsgroepen toe te passen bij de normering. Ten tweede is aangetoond dat basisscholen wel degelijk invloed hebben op het onderbouwrendement van een VO-vestiging, maar deze invloed is niet voor alle schoolsoorten even sterk. Overwogen kan worden of het niet minder dan huidige zeven vergelijkingsgroepen

zouden moeten zijn. Daarbij is het van belang dat het aantal VO-vestigingen in een vergelijkingsgroep een minimaal grootte heeft.

Een tweede aspect is dat bij het prospectieve onderbouwrendement geen rekening wordt gehouden met leerlingen die in de loop van leerjaar 1 of 2 instromen of uitstromen. Analysetechnisch kan dergelijke in- en uitstroom wel meegenomen worden, maar dit is hier niet aanbevelenswaardig. Het aandeel van leerlingen die dit betreft is erg beperkt. De toename in complexiteit van de berekeningen en het feit dat zeer waarschijnlijk de resultaten niet of nauwelijks veranderen, zijn de onderbouwing voor deze aanbeveling.

Het derde aspect is de compensatie voor leerlingen met leerwegondersteuning in de berekening van het onderbouwrendement. In het huidige opbrengstenmodel is het zo dat 1 op de 4 lwoo-leerlingen mag doubleren, zonder dat dit negatieve gevolgen heeft voor het onderbouwrendement van een VO-vestiging. Het is niet verstandig om VO-vestigingen te blijven compenseren voor lwoo-leerlingen op de manier waarop dit tot nu toe gedaan is. Lwoo-leerlingen doubleren zeker niet vaker dan andere leerlingen: ze blijven helemaal niet zitten (vmbo-bl en kl) of juist minder vaak (vmbo-tl en gl). Maar hier moeten twee zaken wel uit elkaar gehouden worden. Als lwoo-leerlingen niet doubleren dankzij de extra ondersteuning die ze krijgen, dan is de financiële compensatie voor lwoo-leerlingen effectief gebleken en zou in stand gehouden moeten worden. Als vestigingen bij de berekening van het onderbouwrendement gecompenseerd worden voor het feit dat ze lwoo-leerlingen hebben, dan is dit op grond van de resultaten van dit onderzoek feitelijk een overcompensatie. Het zou afgeschaft kunnen worden omdat vestigingen geen 'last' hebben van lwoo-leerlingen. De vergelijking tussen vestigingen zou eerlijker worden.

Het vierde aspect is dat het onderbouwrendement in belangrijke mate bepaald wordt door de kwaliteit van het advies dat de basisschool afgeeft. Met name brede adviezen (meer dan twee schoolsoorten breed) zouden problematisch kunnen zijn. Ze zorgen voor een grotere mate van onnauwkeurigheid van de meting van het startniveau van leerlingen. De vraag is of leerlingen die naar brede brugklassen in VO-1 gaan, ook brede adviezen hebben gekregen. Gevonden is dat brede adviezen maar in een klein percentage van het totaal voorkomen en dat de samenhang tussen de breedte van het advies en de breedte van de brugklas zwak is. Dit maakt dat met dit aspect in het beoogde opbrengstenmodel geen rekening hoeft te worden gehouden.

Conclusies deelproject 3

In deelproject 3 is onderzoek gedaan naar de mogelijkheden voor het ontwikkelen van een maat voor de toegevoegde waarde van VO-scholen voor leerlingen die naar havo doorstromen na het afronden van vmbo-tl. Deze maat kan maar voor ruim de helft van de leerlingen en vestigingen bepaald worden. Hiervoor zijn twee redenen van algemene aard aan te wijzen: 1) zeker een op de vijf gediplomeerde vmbo-tl-leerlingen maakt het doorstroomtraject niet af en doet dus ook geen havo-eindexamen en 2) vrij veel vestigingen voldoen niet aan de eis dat van

tenminste tien leerlingen alle nodige informatie beschikbaar is. Gevonden is dat er verschillen in centraal examencijfers bestaan tussen havo-afdelingen van VO-vestigingen voor deze specifieke groep leerlingen. Dit geldt zowel voor het gemiddelde centraal examencijfer voor alle vakken, als voor de kernvakken Nederland, Engels en wiskunde. Om de toegevoegde waarde voor het centraal examencijfer havo te bepalen is alleen correctie voor het beginniveau van leerlingen (centraal examencijfer vmbo-tl) voldoende. Als ook achtergrondkenmerken worden meegenomen in de modellen dan leidt dit niet tot een grote verschuivingen. Dat betekent dat voor deze specifieke groep leerlingen (vmbo-tl-leerlingen die naar havo opstromen) gekozen kan worden voor een relatief simpele toegevoegde waarde maat met een model waarin alleen het beginniveau van de leerlingen is opgenomen. Daarnaast is ook gevonden dat havo-vestigingen met hoog gemiddeld centraal examencijfer voor deze specifieke groep leerlingen in het algemeen ook een hoog doorstroomrendement hebben. De vmbo-tl-gediplomeerde leerlingen doubleren er relatief weinig van havo-4 naar havo-5 en slagen relatief vaak. Echter, omdat deze samenhang vrij zwak is, is het onvoldoende om de opbrengsten van een havo-afdeling als doorstroom-afdeling te bepalen op basis van slechts één van de twee opbrengstindicatoren. Dit zou strategisch gedrag van scholen in de hand kunnen werken.

Conclusies deelproject 4

In dit deelproject zijn de mogelijkheden verkend voor bepalen van het externe rendement van VO-vestigingen voor de leerlingen uit de basis- en kaderberoepsgerichte leerwegen van het vmbo. Het externe rendement betreft hier de mate waarin VO-vestigingen deze groep leerlingen succesvol weet te plaatsen in het mbo of in een hogere leerweg van het vmbo. Deze bepaling is mogelijk voor 90% van de basis- en 95% van de kader-afdelingen van VO-vestigingen. De vestigingen die afvallen voldoen voornamelijk niet aan de eis dat van tenminste tien leerlingen alle nodige informatie beschikbaar is.

Om VO-vestigingen onderling op een eerlijke manier te kunnen vergelijken, is het zaak om de unieke bijdrage van de VO-vestiging vast te stellen in het externe rendement. In dit deelproject is dat gedaan door met geavanceerde modellen het effect van de mbo-instellingen uit het externe rendement te filteren. De modellen laten niet alleen zien dat succesvolle plaatsing van vmbo-bl en kl-leerlingen mede afhankelijk is van de mbo-instelling, maar ook dat verschillen in succesvolle plaatsing van bl en kl-leerlingen één jaar na de examenklas sterker gerelateerd zijn aan de mbo-instelling waar de leerlingen naar toe gaan, dan aan de vmbo-vestiging waar de leerlingen van afkomstig zijn. Als het effect van de mbo-instellingen eruit is gehaald, zijn er nog steeds verschillen in extern rendement tussen VO-vestigingen aanwezig. Een andere belangrijke bevinding is dat het meenemen van de invloed van mbo-instellingen in de berekening van het externe rendement leidt tot zeer grote veranderingen van de waarde van het externe rendement van VO-vestigingen. Dit betekent dat bepaling van het externe rendement van vmbo-bl en kl-afdelingen van VO-vestigingen zonder daarbij de

invloed van mbo-instellingen te betrekken, leidt tot inaccurate bepalingen van het externe rendement. Met het huidige model van de Inspectie kan derhalve (nog) niet de toegevoegde waarde op een valide manier bepaalde worden. Een valide externe rendementsbepaling van basis- en kaderberoepsgerichte leerwegen in het vmbo zal dus relatief complex zijn en is niet eenvoudig te rekenen door scholen.

Het derde doel van dit project betreft het verkennen van de mogelijkheden voor een meer gedifferentieerd toezicht op basis van de beoogde opbrengstindicatoren. De Inspectie van het Onderwijs wil in de toekomst drie groepen van scholen onderscheiden: scholen met zwakke opbrengsten, scholen met voldoende opbrengsten maar met mogelijkheden tot verbetering en scholen met goede opbrengsten. De data-analyses van die in de deelprojecten zijn uitgevoerd leiden in alle gevallen tot opbrengstindicatoren die ingezet kunnen worden om scholen onderling te vergelijken. Deze opbrengstindicatoren volgen steeds een verdeling die het mogelijk maakt om op een of andere manier drie groepen scholen te onderscheiden. Daarbij moeten afkapgrenzen gekozen worden. Waar de grens tussen zwakke scholen, scholen met voldoende opbrengsten maar met verbeteringsmogelijkheden en scholen met goede opbrengsten moet komen te liggen, is in zekere zin arbitrair. In de huidige toezichtpraktijk is de bolletjes-systematiek een geaccepteerde manier om dit te doen. Het is ook mogelijk om bij een normale verdeelde opbrengstindicator over te gaan op afkapgrenzen die gebaseerd zijn op de spreiding rondom het gemiddelde. Op welke manier men de grenzen ook trekt, er kleven altijd bezwaren aan. Of een school als zwak beoordeeld wordt zal deels afhankelijk zijn van waar men de grens getrokken heeft. Verder is de vergelijking steeds relatief: er zijn altijd zwakke en goede scholen aan te wijzen.

1 Inleiding

In het regeerakkoord van Rutte I staat dat zal wordt toegewerkt naar een kwaliteitsnorm voor scholen waarbij een indicator voor de toegevoegde waarde zwaarder meeweegt dan voorheen. Tevens wordt vermeld dat bij de beoordeling van scholen op termijn meer zal worden gelet op de leerwinst en toegevoegde waarde. Het kabinet Rutte II heeft deze prioritering gehandhaafd. In aansluiting hierop is het ontwikkelen van een goede maat voor de toegevoegde waarde een belangrijk aandachtspunt geworden voor de Inspectie van het Onderwijs bij de verbetering van hun huidige beoordelingssystematiek. De Inspectie is in het voorjaar van 2012 gestart met het vernieuwen van het opbrengstenmodel voor het VO (zie notitie: Prototype: Opbrengsten Model VO, september 2012). In samenspraak met het onderwijsveld is gezocht naar bruikbare proxies voor de toegevoegde waardebepaling van VO-scholen. De kern van de vernieuwing is dat uitgegaan wordt van data van individuele leerlingen (Bron-bestanden VO en PO en 1-CijferVO bestanden) en dat de opbrengstindicatoren worden voorzien van absolute normen. Indien er sprake is van een ongunstige leerlingpopulatie kan de norm naar beneden worden bijgesteld.

De begrippen leerwinst en toegevoegde waarde hebben in het onderwijsveld geen eenduidige betekenis. De toegevoegde waarde wordt in dit rapport beschouwd als de bijdrage van een school aan de eindprestaties van leerlingen, gegeven het beginniveau. Daarbij is het van belang om rekening te houden met kenmerken die de eindprestaties van leerlingen wel beïnvloeden, maar waar de school geen invloed op heeft. Het gaat dan om achtergrondkenmerken van leerlingen zoals sociaal milieu of etnische afkomst, of om kenmerken van scholen zoals de samenstelling van de leerlingpopulatie of de wijk waarin de school staat. Een goede meting van het beginniveau van leerlingen is in ieder geval essentieel. De leerprestaties van leerlingen aan het eind worden dus in verband gebracht met het prestatieniveau aan het begin. De Inspectie van het Onderwijs kan echter voor diverse opbrengstindicatoren (nog) niet beschikken over een goede meting van het beginniveau van VO-leerlingen. Daarom werkt de Inspectie in het huidige opbrengstenmodel veelal met proxies voor het beginniveau. Het zijn algemene indicatoren voor de leerachterstand van leerlingen: sociaal economische status, leerwegondersteuning en achterstandsgebiedsbepaling. Dit is een suboptimale oplossing die consequenties kan hebben voor de vergelijking van de toegevoegde waarde van VO-scholen onderling. VO-scholen kunnen hierdoor ten onrechte een slechte of goede beoordeling van hun toegevoegde waarde krijgen.

Het huidige onderzoek richt zich op diverse aspecten, effecten en mogelijkheden van de geplande vernieuwing van het opbrengstenmodel VO. Het dient ter onderbouwing van de keuzes die gemaakt moeten worden voor het nieuwe opbrengstenmodel VO. De gevolgen van bepaalde keuzes zullen zo goed mogelijk inzichtelijk gemaakt worden. De mogelijke

consequenties van het gebruik van proxies voor het beginniveau van leerlingen is één daarvan. De huidige vijf rendementsindicatoren (opstroom-afstroom in de onderbouw, doorstroom in de onderbouw, doorstroom in de bovenbouw, opstroom/afstroom in de bovenbouw en extern rendement) en de prestatie-indicatoren gemiddeld centraal examencijfer voor alle vakken en gemiddeld centraal examencijfer per vak of vakkencluster dienen als uitgangspunt. Mogelijk worden de twee rendementsindicatoren voor de bovenbouw gecombineerd tot één indicator; door- en opstroom bovenbouw (zie notitie *Verkenning prototype opbrengstenmodel voor door- en opstroom bovenbouw*, 11 maart 2013).

In opdracht van het Ministerie van OCW en de Inspectie van het Onderwijs zijn recentelijk twee pilotstudies geïnitieerd over leerwinst en toegevoegde waarde; één in het primair onderwijs (*Pilot leerwinst en toegevoegde waarde PO*) en één in het voortgezet onderwijs (*Pilot toegevoegde waarde/leerwinst VO*). De pilot in het PO is inmiddels afgerond (Janssens, Rekers-Mombarg, & Lacor, 2014). In deze pilot zijn methoden ontwikkeld om de leerwinst en toegevoegde waarde van PO-scholen inzichtelijk in kaart te brengen. Hierbij is gebruik gemaakt van individuele gegevens van leerlingen die regulier door de scholen zelf verzameld worden. De Pilot in het VO zal eind 2014 worden afgerond. Deze pilot heeft als doel VO-scholen te ondersteunen bij het inzichtelijk maken van hun leerwinst (zie ook <http://www.onderwijsinspectie.nl/onderwijs/Voortgezet+onderwijs/leerwinst-en-toegevoegde-waarde>). Voor de Inspectie van het Onderwijs zijn deze pilots niet direct bruikbaar omdat leerwinst en toegevoegde waarde wordt benaderd vanuit een schoolverbeteringsperspectief, in plaats van een schoolaccountabilityperspectief. De pilotstudies zijn vooral gericht op inzichtelijkheid en bruikbaarheid voor scholen zelf. Bovendien liggen aan de berekeningen vaardigheidsscores van toetsen uit het Cito-leerlingvolgsysteem over meerdere leerjaren ten grondslag. De Inspectie beschikt niet over dergelijke data en kan dus niet op soortgelijke wijze de toegevoegde waarde berekenen.

In de dissertatie van Timmermans (2012) is onderzoek gedaan naar de assumpties achter modellen waarmee de toegevoegde waarde van scholen kan worden geschat. Ze heeft voor een deelonderzoek binnen het VO gebruik gemaakt van het grootschalig cohortonderzoek VOCL'99. Voor de Inspectie zijn de toepassingsmogelijkheden van deze dissertatie beperkt omdat de Inspectie in de huidige toezichtpraktijk niet kan beschikken over gedetailleerde informatie over het beginniveau en achtergrondkenmerken van leerlingen als in VOCL'99. Ook is de inzichtelijkheid van de berekeningen voor scholen beperkt; opbrengstindicatoren zouden bij voorkeur na te rekenen moeten zijn door scholen en inspecteurs. Wel zou dergelijke gedetailleerde data van individuele leerlingen inzicht kunnen bieden in wat de toegevoegde waarde van een school is, als dit berekend wordt met een zo optimaal mogelijke dataset. De bestaande databestanden van VOCL'99 bevatten diverse relevante achtergrondkenmerken van leerlingen en kenmerken waarmee het beginniveau van leerlingen nauwkeurig geschat kan worden. De bepaling van de toegevoegde waarde van een school op

basis van VOCL'99 data kan dan als 'gouden standaard' dienen voor het vergelijken van de toegevoegde waarde van een school op basis van de huidige en beoogde berekeningswijzen van de Inspectie.

Voor alle duidelijkheid, de toegevoegde waarde van een school zoals die in dit project aan de orde is, is geen meting van de 'zuivere' bijdrage van een school aan de leerprestaties van leerlingen (schooleffect). Daarvoor is zijn de toegepaste modellen te beperkt en kan er niet worden gecorrigeerd voor alle relevante niet-schoolse kenmerken, inclusief onbekende of niet te meten kenmerken van leerlingen, ouders, leeftijdsgenoten of schoolcontext. Het ultieme doel, de zuivere bijdrage van een school vaststellen, is met de huidige methoden en beschikbare informatie vooralsnog een brug te ver. De toegevoegde waarde in dit project is daarom te beschouwen als een indicatie van de invloed die scholen hebben op de prestaties van hun leerlingen. Dit is in overeenstemming met recente publicaties over dit onderwerp (Harris, 2011; Bosker, 2012; Timmermans, 2012; Janssens, Rekers-Mombarg, Lacor, 2014).

Ter ondersteuning van de beoogde vernieuwing van het opbrengstenmodel VO is het huidige onderzoek opgezet. Er wordt voortgeborduurd op het onderzoek van Timmermans (2012). Door Inspectie van het Onderwijs zijn drie doelen geformuleerd:

- Het valideren van het beoogde opbrengstenmodel VO door een vergelijking te maken tussen de indicatorberekening volgens het huidige model, het beoogde opbrengstenmodel VO en een model waarin de toegevoegde waarde berekening zo optimaal mogelijk gedaan wordt.
- Het uitproberen van grootschalige vormen van toegevoegde waardeberekening op basis van Bron-bestanden en 1-cijfer-VO-bestanden wat betreft het onderbouwrendement en speciale groepen in het VO waar zowel informatie over het beginniveau als het eindniveau aanwezig is. In overleg met de opdrachtgever is besloten om de uitwerking voor de speciale groepen in het VO te beperken tot alleen vmbo theoretische leerweg-leerlingen (vmbo-tl) die doorstromen naar het havo, en vmbo basisberoepsgerichte leerweg (vmbo-bl) en kaderberoepsgerichte leerweg (vmbo kl) leerlingen die hun opleidingen voortzetten in een hogere vmbo-leerweg of het mbo.

Het verkennen van de mogelijkheden hoe de verkregen informatie kan worden ingezet voor de differentiatie in het toezicht, waarbij onderscheid wordt gemaakt tussen scholen met zwakke opbrengsten, scholen met voldoende opbrengsten maar met mogelijkheden tot verbetering en scholen met goede opbrengsten.

1.1 Onderzoeksvragen

Op basis van de hierboven geformuleerde doelen is het onderzoek op te splitsen in vier deelprojecten met de volgende negen hoofdonderzoeksvragen.

Deelproject 1: Validering van beoogd opbrengstenmodel

1. Leidt het toepassen van modellen die verschillen in de manier en mate voor correctie voor het instroomniveau van leerlingen tot verschillen in het rendement bovenbouw van een VO-vestiging?
2. Leidt het toepassen van deze modellen tot verschillen in het gemiddelde centraal examencijfer voor alle vakken van een VO-vestiging?

In verlengde van de tweede vraag over de validiteit van het gemiddelde centraal examencijfer ligt de vraag naar de betrouwbaarheid van deze opbrengstindicator en of dit verschillend is voor de vier modellen die onderling vergeleken worden. De bijbehorende aanvullende onderzoeksvragen van deelproject 1 zijn:

- In hoeverre vormt de indicator ‘gemiddeld cijfer centraal examen’ een goede afspiegeling van de onderliggende centraal examencijfers per vak of vakkencluster?
- Is de interne consistentie van deze indicator verschillend voor de vier modellen?

Deelproject 2: Onderbouwrendement VO

1. In welke mate wijkt de huidige retrospectieve bepaling van het onderbouwrendement af van een prospectieve bepaling van het onderbouwrendement met individuele data?
2. In hoeverre hebben de toeleverende PO-vestigingen invloed op de waarde van het onderbouwrendement van een VO-vestiging?
3. Welke rol spelen over- en onder advisering hierbij?

Na overleg met opdrachtgever zijn bij dit deelproject een viertal aanvullende (discussie)vragen opgesteld.

- Is het wenselijk om VO-vestigingen met lwoo-leerlingen te compenseren voor doubleren?
- Hoe kan het beste omgegaan worden met de diverse combinaties van onderwijstypen in de onderbouw van VO-vestigingen?
- Hoe kan het beste omgegaan worden met instromende leerlingen in leerjaar 1 en 2? Aan welke VO-vestiging moeten deze leerlingen worden toegekend?
- In welke mate is het advies van leerlingen afhankelijk van het onderwijsaanbod op de ontvangende VO-vestigingen?

Deelproject 3: Toegevoegde waarde van vmbo-tl naar havo

4. Wat is de toegevoegde waarde van een VO-vestiging op het gemiddelde centraal examencijfer havo voor vmbo-tl gediplomeerde leerlingen die doorgestroomd zijn naar havo?

5. Verschilt de toegevoegde waarde van havo-afdelingen van VO-vestigingen tussen de verplichte vakken Nederlands, Engels en wiskunde?

Naast deze hoofdvraagstellingen is er nog een aanvullende vraag van de opdrachtgever:

- Hoe verhoudt de toegevoegde waarde van een VO-vestiging op het gemiddelde centraal examencijfer zich tot het rendement voor vmbo-tl gediplomeerde leerlingen die doorgestroomd zijn naar het havo?

Deelproject 4: Extern rendement vmbo-bl en kl

6. Wat is het externe rendement na één en twee jaar mbo van vmbo basis- en kaderberoepsgerichte leerweg-afdelingen van VO-vestigingen, gegeven het examenklasniveau van hun leerlingen?
7. Wat zijn de effecten van de voorgestelde normering voor het percentage succesvolle plaatsingen?
8. In hoeverre is de succesvolle plaatsing van leerlingen in het eerste jaar na het verlaten van een vmbo basis- of kaderberoepsgerichte leerweg gerelateerd aan de VO-vestiging enerzijds en mbo-instelling anderzijds?
9. In hoeverre leidt het corrigeren voor invloeden van mbo-instellingen tot veranderingen van de schattingen van het externe rendement voor VO-vestigingen?

Derde doelstelling: verkenning van meer differentiatie in toezicht

Naast deze zeven hoofdvraagstellingen zijn er nog enkele implementatie- en evaluatievragen die voortvloeien uit de derde doelstelling van dit onderzoek. Het gaat om de volgende vragen:

- Kan met de beoogde opbrengstindicatoren op een verantwoorde wijze meer variatie in de toezichtarrangementen en de benadering van VO-vestigingen worden aangebracht (bijvoorbeeld door het maken van onderscheid in goede, matige en zwakke scholen)?
- Kunnen de beoogde opbrengstindicatoren ook indicaties geven voor de gewenste schoolverbetering voor specifieke groepen van leerlingen, te weten naar havo doorstromende vmbo-tl leerlingen en leerlingen die vmbo-bl of kl eindexamen hebben gedaan?
- In hoeverre kan met de beoogde opbrengstindicatoren op verantwoorde wijze inzicht verkregen worden in de longitudinale of trendontwikkeling van de resultaten van VO-vestigingen?

Voor de beantwoording van deze vragen is de empirische data van dit onderzoek niet toereikend. Daarom zullen deze vragen in de algemene discussie aan het eind van dit rapport aan de orde komen (Hoofdstuk 7).

1.2 Leeswijzer

Het volgende hoofdstuk - hoofdstuk 2 - gaat nader in op het begrip toegevoegde waarde. De theorie wordt behandeld, evenals de verschillende manieren waarop dit kan worden berekend voor een school. Speciale aandacht gaat uit naar toegevoegde waardebeoordelingen die in de VS en de UK al in de toezichtpraktijk worden toegepast.

In hoofdstuk 3 komt het eerste deelproject aan de orde. Aan dit deelproject is medewerking verleend door Dr. Henk Guldemond. Het betreft hier het valideren van de toegevoegde waardebeoordeling van het gemiddelde centraal examencijfer en rendement bovenbouw met behulp van individuele data van een grootschalig landelijk cohortonderzoek (VOCL'99). De methode en de resultaten worden hier besproken. De conclusies en discussie staan in het laatste hoofdstuk van dit rapport.

Hoofdstuk 4 gaat over het tweede deelproject. Op basis van landelijke bestanden wordt de beoogde prospectieve berekeningswijze van het onderbouwrendement vergeleken met de huidige retrospectieve berekeningswijze van het onderbouwrendement. Ook wordt ingegaan op de invloed van toeleverende PO-scholen op het onderbouwrendement en de compensatie voor lwoo-leerlingen. De methode en de resultaten van dit deelproject worden in dit hoofdstuk besproken, de discussie volgt in het laatste hoofdstuk.

Hoofdstuk 5 betreft het derde deelproject. Het gaat om de toegevoegde waardebeoordeling van de doorstroom van vmbo-tl leerlingen naar het havo, waarbij het vmbo-tl eindexamen als meting van het beginniveau dient en het havo eindexamen als meting van het eindniveau. Ook hier worden in dit hoofdstuk alleen de methode en resultaten van dit deelproject besproken.

Het vierde deelproject, gepresenteerd in Hoofdstuk 6, is vooral exploratief van aard. Het betreft verkenningen van het extern rendement van vmbo-bl en kl leerlingen die hun opleiding voortzetten in een hogere leerweg van vmbo of op het mbo. De potentie van deze maat als opbrengstindicator wordt onderzocht. In dit hoofdstuk komen de opzet van de analyses en de resultaten van dit deelproject aan de orde.

Per deelproject worden in Hoofdstuk 7 de conclusies getrokken en resultaten bediscussieerd. In dit hoofdstuk wordt ook ingegaan op de aanvullende discussievragen over onderbouwrendement van deelproject 2, de algemene implementatie- en evaluatievragen van de opdrachtgever en het derde doel van dit onderzoek.

2 Achtergrond

2.1 Definiëring van leerwinst en toegevoegde waarde

De begrippen leerwinst en toegevoegde waarde hebben geen eenduidige betekenis in het onderwijsveld en worden veelal door elkaar gebruikt. In de wetenschappelijke literatuur wordt wel onderscheid gemaakt tussen deze begrippen, maar voor elk begrip hanteert men verschillende definities. Onder leerwinst wordt in het algemeen verstaan de toename van vaardigheden, kennis en/of competenties van leerlingen. De leerwinst van een leerling wordt gewoonlijk berekend door het verschil tussen de leerprestaties op twee verschillende momenten van de betreffende leerling. Hierbij wordt verondersteld dat het twee metingen op dezelfde onderliggende schaal zijn.

De leerwinst van leerlingen is niet alleen op het conto van de school te schuiven. Andere niet-schoolse invloeden, zoals sociaal milieu en etniciteit, laten ook hun invloed gelden op de prestatieontwikkeling van leerlingen. Het begrip toegevoegde waarde doelt op de bijdrage van de school aan de leerwinst van leerlingen. Door de OECD (2008) wordt toegevoegde waarde omschreven als *'the contribution of a school to student's progress towards stated of prescribed education objectives (e.g. cognitive achievements)'*. Sammons, Thomas en Mortimore (1997) gaan een stap verder en definiëren toegevoegde waarde als *'an indication of the extent to which any given school has fostered the progress of all students in a range of subjects during a particular time period in comparison to the effects of schools in the same sample'*. De definitie van toegevoegde waarde in dit project sluit aan bij deze laatste omschrijving. Het gaat om het bepalen van de bijdrage van een school aan de eindprestaties van leerlingen, gegeven het beginniveau. Daarbij is het van belang om te corrigeren voor kenmerken die de leerprestaties van leerlingen wel beïnvloeden, maar waar de school geen invloed op heeft. Ook is het belangrijk om met de structuur van de data rekening te houden: leerlingen zijn gegroepeerd binnen scholen. Samenvattend, de berekening van de toegevoegde waarde in dit project vereist meerniveau modellen met een goede begin- en eindmeting van de leerprestaties op leerlingniveau, en met correctie voor relevante kenmerken van de achtergrond van de leerlingen en de context van de school. De toegevoegde waarde is bovendien een relatieve maat; het gaat om een vergelijking tussen scholen in dezelfde steekproef.

Bij de toegevoegde waardebepaling van een school is dus tenminste een begin- en eindmeting nodig, maar die hoeven niet noodzakelijkerwijs op dezelfde meetlat te liggen. Het komt de inzichtelijkheid en bruikbaarheid ten goede als dit wél het geval is, zoals in de *Pilot leerwinst en toegevoegde waarde PO* (Janssens, e.a., 2014). Het beginniveau van leerlingen wordt in dit project zo nauwkeurig mogelijk benaderd door een combinatie van leerlingkenmerken. Welke

combinatie wordt gebruikt is afhankelijk van de opbrengstindicator waar het betrekking op heeft. Het kan gaan om kenmerken voor de leerachterstand op schoolniveau, om het effect van de advies, Cito-eindtoetscore en/of motivatie van individuele leerlingen of om de centraal examencijfers van de schoolsoort waar de gediplomeerde leerlingen van afkomstig zijn.

2.2 Conceptuele aspecten van toegevoegde waardebeoordeling

In de jaren tachtig van de vorige eeuw startten methodologen een discussie over hoe de toegevoegde waarde van een school het beste gemeten zou kunnen worden. Er werd met name gewezen op het belang van het adequaat modelleren van meerniveau gegevens: uitspraken over scholen worden immers gedaan op basis van prestatiegegevens van een selectie van leerlingen die die scholen bezoeken (Aitkin & Longford, 1986; Raudenbush & Bryk, 1986; Snijders & Bosker, 2012).

Sinds het openbaar maken van de opbrengstgegevens van scholen, begin jaren negentig, wordt er ook publiekelijk debat gevoerd over opbrengstgegevens van scholen. Dit debat begon in Engeland met het publiceren van de zogenaamde League Tables, maar zette zich voort in de VS en, na de eerste Trouw-publicatie in 1997 met ranglijsten van scholen, ook in Nederland. (Bosker, Lam, Luyten, Steen, & Vos, 1998). Hier werd de vraag opgeworpen in hoeverre dergelijke maten werkelijk de inspanningen van scholen weerspiegelden, dan wel mede de effecten van andere niet-schoolse invloeden (Dijkstra, Karsten, Veenstra, & Visscher, 2001).

In de loop der jaren zijn deze twee discussies grotendeels geconvergeerd. Men werd het erover eens dat een valide en eerlijke maat voor de toegevoegde waarde van een school gebaseerd zou moeten zijn op tenminste een voor- en nameting van de leerprestaties van voldoende leerlingen. Deze leerprestaties moeten valide en betrouwbaar gemeten zijn en de data moet geanalyseerd worden met statistische modellen die de meerniveau-structuur adequaat modelleren.

Tegenwoordig wordt er gediscussieerd over of er wel of niet gecorrigeerd zou moeten worden voor niet-schoolse invloeden, en zo ja, welke kenmerken dat dan zouden moeten zijn en op welk niveau de correctie moet plaatsvinden (leerling- en/of schoolniveau). (Castellano & Ho, 2013; McCaffrey, Lockwood, Koretz, Louis & Hamilton, 2004; Willms & Raudenbusch, 1989). Men lijkt inmiddels toch een zeker mate van overeenstemming te hebben bereikt over de kenmerken op leerlingniveau. Afhankelijk van het gebruiksdoel van de toegevoegde waardebeoordeling wordt er al dan niet gecorrigeerd voor kenmerken van leerlingen die wel van invloed zijn op de leerprestaties maar waar de school zelf geen invloed op heeft. Er zijn grofweg drie gebruiksdoelen te onderscheiden: schoolkeuze, schoolverbetering en schooltoezicht. Deze worden hieronder nader toegelicht. Voor het corrigeren voor kenmerken

van de schoolcontext is die overeenstemming er zeker niet, daar is nog steeds veel discussie over.

Voor ouders die een school willen kiezen voor hun kinderen is een controle voor de niet-schoolse factoren niet noodzakelijk, omdat het voor hen niet uitmaakt hoe die bijdrage van een school tot stand komt: als dat ook komt door het sociaal milieu van medeleerlingen is dat prima. Het gaat er vooral om dat hun kind veel leert in de tijd dat het naar de basisschool gaat.

Voor schoolverbeteringsdoeleinden wordt correctie voor leerlingkenmerken in het algemeen wel relevant bevonden. Uit een evaluatieonderzoek van Janssens, Rekers-Mombarg & Lacor (2014) is gebleken dat schoolleiders van mening zijn dat, door rekening te houden met de moeilijkheidsgraad van de leerlingpopulatie, de gemiddelde leerwinst op hun school op een eerlijker manier vergeleken kan worden met de gemiddelde leerwinst van andere scholen of van een school in de buurt.

In het kader van het toezicht door de Inspectie van het Onderwijs is een dergelijke controle ook belangrijk. Het is wettelijk vereist dat bij de berekening van de opbrengsten van scholen rekening wordt gehouden met kenmerken van de leerlingpopulatie van scholen. Dit is noodzakelijk om de beoordeling van scholen rechtvaardig te laten zijn. Voor een voorbeeld van onderzoek in het VO waarin is nagegaan hoe de verschillende toegevoegde waarde maten zich tot elkaar verhouden en welke maat voor toegevoegde waarde dan het meest adequaat is voor toezichtdoeleinden, wordt verwezen naar Timmermans, Doolaard, & de Wolf (2011).

2.3 Toegevoegde waardebepaling voor toezichtdoeleinden

Voor de Inspectie van het Onderwijs is het ontwikkelen van een goede maat voor toegevoegde waarde van VO-scholen zeer relevant. Weliswaar is in de toekomstig de eerste vraag die men stelt of leerlingen van een VO-school voldoen aan bepaalde referentieniveaus, maar direct daarna volgt de tweede vraag of de VO-school wel het maximaal haalbare uit de leerlingen heeft gehaald op de diverse opbrengstindicatoren. Om deze tweede vraag te kunnen beantwoorden is inzicht in het beginniveau van leerlingen noodzakelijk. Voor het onderbouwrendement wordt het advies van de basisschool beschouwd als een goede indicator van het beginniveau van leerlingen. Voor het bovenbouwrendement en het eindexamen zijn in het huidige opbrengstenmodel echter alleen proxies voor het beginniveau voorhanden, zoals woonachtig zijn in een armoedeprobleemcumulatiegebied (apcg), indicatie voor leerwegondersteunend onderwijs (lwoo) en percentage leerlingen dat in leerjaar 3 is ingestroomd vanuit een andere school (instroom in derde leerjaar).

Het is essentieel dat toegevoegde waarde modellen rekening houden met de geneste structuur van de dataset (leerlingen zijn gegroepeerd in scholen). Daarom moeten meerniveau regressiemodellen aan de berekeningen ten grondslag liggen. Dergelijke modellen zijn meer

valide dan multiple lineaire regressie modellen, omdat ze rekening houden met de afhankelijkheid van de metingen van leerlingen binnen een school. Leerlingen binnen een school lijken meer op elkaar wat betreft leerprestaties dan willekeurige leerlingen omdat ze bijvoorbeeld van dezelfde leraar les krijgen. Meerniveau modellen worden ook wel hiërarchische modellen, gemengde (mixed) modellen, multilevel modellen, random effects modellen of random coëfficiënt modellen genoemd. De essentie is dat de leerprestatie bij de eindmeting (posttestscore) van een leerling voorspeld wordt uit zijn/haar beginmeting (pretestscore) en het schoollidmaatschap. Op deze manier kan men schatten in hoeverre scholen meer of minder dan het verwachte eindniveau realiseren bij hun leerlingen. De modellen kunnen op vele verschillende manieren worden aangepast om tot een zo realistisch mogelijke schatting van de toegevoegde waarde van scholen te komen. Voorbeelden hiervan zijn het modelleren van niet-rechthoekige ontwikkelingspatronen (zie Guldemond & Bosker, 2009), het modelleren van andere dan continue uitkomstvariabelen met behulp van gegeneraliseerde meerniveau-modellen (bijvoorbeeld voor de BVE-sector: Timmermans, Bosker, Doolaard, & De Wolf, 2012), en het verdisconteren van de invloed van de toeleverende basisscholen of van schoolwisselingen van leerlingen. Dit laatste is toegepast in een onderzoek naar realistische schattingen van schooleffecten van Timmermans, Snijders en Bosker (2013).

De meerniveau modellen hebben als algemeen nadeel dat naarmate de complexiteit toeneemt, de resultaten nog minder transparant en minder toegankelijk zijn voor een niet-statistisch onderlegd publiek. Men zal de berekeningen gaan beschouwen als een ‘black box’. Daar staat tegenover dat meerniveau modellen methodologisch beter zijn dan de multiple lineaire regressiemodellen die in het huidige opbrengstenmodel worden toegepast. Ook lineair regressiemodellen zijn erg lastig te doorgronden door schoolleiders en inspecteurs. Daarom worden de resultaten alleen maar gepresenteerd in de vorm van bolletjes in het huidige opbrengstenmodel of in de vorm van absolute normen in het beoogde opbrengstenmodel. Iets dergelijks kan men ook toepassen bij meerniveau modellen.

Internationale toepassing

In de praktijk vinden we diverse toepassingen van de bovengenoemde modellen voor toezichtdoeleinden. Als voorbeeld bespreken we hieronder twee in de toezichtpraktijk gebruikte methoden voor het schatten van de toegevoegde waarde: Tennessee Value-Added Assessment System (TVAAS) en Contextual Value Added (CVA).

Tennessee Value-Added Assessment System (TVAAS)

In veel staten van de VS past men het Tennessee Value-Added Assessment System toe, soms ook wel Education Value-Added Assessment System (EVAAS) genoemd. Het systeem wordt zowel voor het primair als voor het voortgezet onderwijs gebruikt. Het systeem, ontwikkeld door Sanders en Horn (1994), bepaalt de effectiviteit van schoolsystemen, scholen en

leerkrachten op basis van de leerwinst van hun leerlingen over meerdere jaren. Hiervoor wordt jaarlijks informatie verzameld over de leerprestaties. Met behulp van statistische modellen wordt een voorspelling van de leerprestaties gedaan waarbij gecorrigeerd wordt voor diverse leerling- en schoolkenmerken. Uit hun analyses blijkt dat de controlevariabelen etniciteit, sociaal-economische status, klassegrootte en heterogeniteit in de klas niet significant zijn als voorspellers van de leerwinst van de leerlingen, in tegenstelling tot de effectiviteit van de leerkracht. Amrein-Beardsley (2008) plaatste diverse methodologische kanttekeningen bij TVAAS. Zij vindt het bedenkelijk dat TVAAS zo breed als model voor het meten van toegevoegde waarde is geaccepteerd, terwijl er nog zoveel onduidelijkheden zijn. Er zijn slechts gebrekkige validiteitsstudies uitgevoerd, het model is niet gebruikersvriendelijk, er is een gebrek aan onafhankelijke reviews en belangrijke methodologische problemen zoals hoe om te gaan met ontbrekende gegevens, regressie naar het gemiddelde zijn tot dusverre nog niet bevredigend opgelost.

Contextual Value Added (CVA)

Tot 2011 heeft de Ofsted in de UK voor Key-stage 3-4 scholen (vergelijkbaar met het Nederlandse voortgezet onderwijs, leeftijd 11 tot 16 jaar) gewerkt met Contextual Value Added (CVA) (Ray, 2006). Hierbij is de toegevoegde waarde over een periode van 4 jaar bepaald, waarbij de verbetering in leerprestaties van een leerling vergeleken is met die van andere Britse leerlingen in vergelijkbare omstandigheden. Hiervoor is een nationale database met informatie van 600.000 leerlingen per jaar gebruikt. Als maat voor het beginniveau van leerlingen zijn de individuele prestaties op een centrale toets aan het einde van Key Stage 2 (vergelijkbaar met groep 8 in Nederland) genomen. De prestaties aan het einde van Key Stage 4 (centrale examens) zijn gebruikt als eindniveau van de leerlingen. In het CVA model zijn diverse op leerlingniveau gemeten correctiefactoren toegepast. Voorbeelden hiervan zijn: geslacht, indicatie voor extra zorg (*in care*), het hebben van een rugzak (*special educational needs*), in aanmerking komen voor een gratis schoolmaaltijd (proxy voor SES), eerste taal, schoolwisseling, leeftijd en wonend in achterstandsbuurt (*income deprivation affecting children index*, tevens proxy voor SES). Daarnaast zijn ook kenmerken van de leerlingpopulatie op het niveau van de school meegenomen, bijvoorbeeld het gemiddelde beginniveau.

Sinds 2011 worden in de League Tables van Ofsted alleen nog maar de resultaten van de VA-bepaling van Key stage 4 scholen gepresenteerd. In het VA-model wordt de toegevoegde waarde van een school bepaald door alleen het beginniveau van de leerlingen mee te nemen en geen andere kenmerken van de leerlingen of leerlingpopulatie. De officiële redenen voor het verwijderen van CVA-bepaling is gelegen in de grote complexiteit van de modellen; ouders begrijpen het niet.

3 Deel I: Validering van beoogde opbrengstenmodel

3.1 Inleiding

In 2012 is de Inspectie van het Onderwijs gestart met het vernieuwen van het opbrengstenmodel voor het VO (zie notitie: Prototype: Opbrengsten Model VO, september 2012). Er is gezocht naar bruikbare indicatoren voor de toegevoegde waardebeoordeling van VO-scholen. Daarbij wordt niet langer uitgegaan van gegevens verzameld op school- of afdelingsniveau – bijvoorbeeld gemiddeld centraal examencijfer van de havo-afdeling van een VO-school –, maar van gegevens van individuele leerlingen. Verder zullen de opbrengstindicatoren worden voorzien van absolute normen. Afhankelijk van de samenstelling van de leerlingpopulatie kan de norm van een indicator naar beneden worden bijgesteld.

Het doel van dit deelproject is het valideren van het nieuwe opbrengstenmodel VO. De databestanden van het grootschalige onderzoek Voortgezet Onderwijs Cohort Leerlingen '99 (VOCL'99) vormen een goed uitgangspunt voor een valideringsonderzoek. Met gedetailleerde data van individuele VOCL'99-leerlingen kan nauwkeurig het beginniveau van leerlingen worden vastgesteld. Op basis hiervan en van relevante achtergrondkenmerken van leerlingen kan een zo optimaal mogelijke bepaling van toegevoegde waarde van een school verkregen worden. Deze toegevoegde waarde bepaling kan vervolgens dienen als 'gouden standaard' voor het vergelijken van de opbrengstindicatoren van de huidige en beoogde berekeningswijzen van de Inspectie.

Om met data van een cohortonderzoek zoals VOCL'99 zo goed mogelijk de bijdrage van een school aan de schoolprestaties zichtbaar te maken (de toegevoegde waarde), is het zaak zoveel mogelijk relevante correctiefactoren in de analyses te betrekken (covariaten). Daarbij is het niet uit te sluiten dat belangrijke – maar onbekende of niet te meten – individuele of schoolcontextkenmerken buiten beschouwing zijn gelaten. Voor deze niet-schoole kenmerken wordt dan niet gecorrigeerd. Hierdoor zal een deel van de verbetering in leerprestaties van leerlingen ten onrechte aan de school worden toegeschreven. Alleen met een grootschalig zuiver experimenteel onderzoek zou men nauwkeurig de toegevoegde waarde van een school kunnen vaststellen, maar hieraan kleven grote praktische en ethische bezwaren. Het schatten van de toegevoegde waarde van een school met data uit een cohortonderzoek als VOCL'99 of landelijke Bron- of 1-cijfer-VO-bestanden zijn daarom hoogstens second-best oplossingen. De toegevoegde waarde van de school geschat met dergelijke databestanden wordt daarom door sommigen strikt het *netto* schooleffect genoemd;

het is niet meer dan een schatting van de bijdrage van een school aan de leerprestaties van leerlingen, nadat gecorrigeerd is voor individuele covariaten¹.

Opbrengstenkaart VO

De Inspectie publiceert jaarlijks de Opbrengstenkaart voor het VO (voorheen de Kwaliteitskaart VO). Op deze Opbrengstenkaart wordt op vestigingsniveau een aantal opbrengstindicatoren gepresenteerd. Twee daarvan zijn ‘*Van 3^e leerjaar naar diploma zonder zittenblijven*’ en het ‘*Gemiddeld cijfer centraal examen*’. In de navolgende tekst worden deze opbrengstindicatoren kortweg bovenbouwrendement en gemiddeld centraal examencijfer genoemd. Op deze twee indicatoren zal het valideringsonderzoek zich richten.

Het is niet zozeer het ‘ruwe’ percentage doorstroom in de bovenbouw of het ‘ruwe’ gemiddeld centraal examencijfer dat hier telt, maar de normering van deze indicatoren. In het huidige opbrengstenmodel wordt bij de normering rekening gehouden met de samenstelling van de leerlingpopulatie van de school. Dit is gedaan door een regressie analyse op schoolniveau uit te voeren met correctie voor het percentage van de schoolpopulatie dat afkomstig is uit achterstandswijken (apcg), het percentage leerlingen dat recht heeft op leerwegondersteunend onderwijs (lwoo) en het percentage instromers in het derde leerjaar. Deze correctiefactoren zijn te beschouwen als globale indicatoren voor het beginniveau van leerlingen en voor de niet-schoolse invloeden op de leerprestaties in de bovenbouw. Aan de hand hiervan wordt bepaald of een school – gegeven de samenstelling van de leerlingpopulatie – naar verwachting, boven verwachting dan wel beneden verwachting heeft gepresteerd.

In het huidige opbrengstenmodel wordt er enkel gewerkt met gegevens op vestigings- of afdelingsniveau om het gemiddelde centraal examencijfer en de doorstroom in de bovenbouw van een VO-vestiging te bepalen. Dit geldt ook voor de drie genoemde correctiefactoren. De reden hiervoor is dat gegevens op individueel niveau niet beschikbaar waren. Inmiddels is dit wel het geval. Daarom wil de Inspectie van het Onderwijs het huidige model gaan vervangen door een beoogd model met dezelfde correctiefactoren als in het huidige model, maar dan gebaseerd op gegevens van individuele leerlingen. Vanuit een methodologische invalshoek is dit beoogde model te verkiezen boven het huidige model, omdat de correctie op individueel niveau meer recht doet aan de realiteit. Maar een model waarin het instroomniveau van leerlingen zo nauwkeurig mogelijk geschat wordt, gegeven de mogelijkheden die VOCL’99 daarvoor biedt, zou weer te verkiezen zijn boven het beoogde model. Met een dergelijk model wordt beter inzicht verkregen in de bijdrage van de school aan de leerprestaties van leerlingen dan met het huidige of beoogde model. Dit model wordt in dit deelproject ook kortweg ‘het optimale model’ genoemd. Hoewel de Inspectie waarschijnlijk niet kan beschikken over

¹ Het betreft hier een optelsom van meerdere niet nadere gespecificeerde effecten, inclusief effecten die niet of nauwelijks aan de school kunnen worden toegeschreven. Bij het netto schooleffect kan alleen met zekerheid worden gesteld dat het gevonden effect niet veroorzaakt kan zijn door verschillen in instroomkenmerken van de leerlingpopulatie, want daar is voor gecorrigeerd.

dergelijke gedetailleerde informatie van leerlingen, kan het optimale model wel duidelijk maken wat de afstand is van het huidige en beoogde model tot het zo optimaal mogelijke toegevoegde waarde model. Om het spectrum compleet te krijgen is ook nog een vierde model van belang. Dat is het lege model, het model waarin geen enkele correctiefactor wordt meegenomen. Het schooleffect dat met een het lege model wordt verkregen, wordt ook wel *bruto* schooleffect genoemd.

De centrale vraag in dit deelproject is in hoeverre de hiervoor genoemde modellen leiden tot verschillen in de normering van het bovenbouwrendement en gemiddeld centraal examencijfer van een VO-vestiging. Waar in het hierboven geschetste spectrum staat het beoogde opbrengstenmodel? Komen de resultaten van het beoogde model in de buurt van die van het optimale model, of wijken de resultaten van het beoogde model maar weinig af van die van het huidige model of het lege model? De onderzoeksvragen van dit deelproject zijn als volgt geformuleerd:

1. *Leidt het toepassen van modellen die verschillen in de manier en mate van correctie voor het instroomniveau van leerlingen tot verschillen in de opbrengstindicator rendement bovenbouw van een VO-vestiging?*
2. *Leidt het toepassen van deze modellen tot verschillen in de opbrengstindicator gemiddeld centraal examencijfer voor alle vakken van een VO-vestiging?*

De tweede onderzoeksvraag heeft betrekking op de vraag naar de validiteit van de berekeningswijze van het gemiddelde eindexamencijfer van een VO-afdeling zoals die in het huidige en het beoogde model wordt berekend. En omdat de vraag naar de validiteit raakt aan de betrouwbaarheid van het indicator, zal in dit deelproject ook nog aandacht worden besteed aan de interne consistentie van deze indicator. Daarbij gaat het om de vraag of het gemiddelde centraal examencijfer van alle vakken een weerspiegeling is van de onderliggende centraal examencijfers per vak of vakkencluster (interne consistentie). De bijbehorende aanvullende onderzoeksvraag is:

- *In hoeverre vormt het gemiddelde centraal examencijfer voor alle vakken een goede afspiegeling van de onderliggende centraal examencijfers per vak of vakkencluster?*

Het is denkbaar dat de vier modellen onderling verschillen in de interne consistentie van het eindexamencijfers. Dit is waarschijnlijk terug te voeren op verschillen in de mate van correctie voor het instroomniveau van leerlingen. Verschillen in examencijfers tussen vakken hangen deels samen met leerlingkenmerken, zoals sekse en etniciteit. Allochtone leerlingen behalen bijvoorbeeld in het algemeen lagere eindexamencijfers voor Nederlands dan autochtone leerlingen. De achtergrondkenmerken van leerlingen hangen ook samen met het instroomniveau. Bij het optimale model waarin een nauwkeurige correctie voor instroomniveau wordt uitgevoerd, zal verwachting de interne consistentie het hoogst zijn. De correctie voor etniciteit bijvoorbeeld zal ertoe leiden dat de examencijfers voor Nederlands

meer gaat lijken op de gemiddeld centraal examencijfers; de invloed van etniciteit is er immers uitgefilterd. In hoeverre verschillen in interne consistentie een rol van betekenis spelen in dit deelproject is nader onderzocht. De bijbehorende aanvullende onderzoeksvraag is:

- *Is de interne consistentie verschillend voor de toegepaste modellen?*

3.2 Methode

3.2.1 Beschrijving van de dataset

Voor de analyses wordt gebruik gemaakt van de databestanden van het VOCL'99 cohort. Deze bevatten gegevens over de gehele onderwijsloopbaan inclusief het eindexamen van leerlingen die in 1999 in de eerste klas van het VO zaten, alsmede gegevens op school-, en leerkrachtniveau en alle informatie die noodzakelijk is voor een zinvolle vergelijking van genoemde modellen. (Kuyper, Lubbers & van der Werf, 2003)

Het is van belang dat de correctiefactoren die toegepast worden bij de modellen zo goed mogelijk overeenkomen met die de Inspectie van het Onderwijs hiervoor gebruikt. Het gaat daarbij met name om de correctiefactoren lwoo, apcg (achterstandswijken) en instroom in het derde leerjaar. Er zijn diverse variabelen geschikt om nauwkeurig het beginniveau van leerlingen mee te bepalen of om als relevante achtergrondkenmerken te dienen. Er kan een selectie gemaakt worden uit de variabelen Cito-eindtoetscore, Cito-entreetoets (een verkorte versie van de Cito-eindtoets basisonderwijs), intelligentie, sociale afkomst (SES), geslacht, advies basisschool. Maar het sterke punt van het VOCL'99 bestanden namelijk diverse belangrijke correctiefactoren beschikbaar op individueel niveau, is tevens de achilleshiel van de database. Zoals gebruikelijk in grootschalig cohortonderzoek zijn (deels) ontbrekende data een serieus probleem. Zo is slechts van iets meer dan helft van de cohortleerlingen het IQ (zoals gemeten met de GIVO-toets) bekend. Nog minder rooskleurig is het gesteld met de Cito-eindtoets basisonderwijs. Het draait hier helaas om twee belangrijke individuele correctiefactoren voor de berekeningen van een optimaal model. Omdat de Inspectie in de toekomst niet op grote schaal zal kunnen beschikken over de IQ van leerlingen, zal dit kenmerk bij het optimale model grotendeels buiten beschouwing worden gelaten².

De variabelen die in zowel het huidige als het beoogde model worden gebruikt – gemiddeld centraal examencijfer alle vakken, percentage instroom in leerjaar 3, lwoo en apcg – zijn in de VOCL'99 bestanden aanwezig. Het huidige model kan met behulp van het VOCL'99

² De invloed van IQ is in eerste instantie wel meegenomen in het optimale model, maar op verzoek van de Inspectie is dit later weer uit dit model verwijderd. Bij vergelijking tussen een optimaal model mét IQ en een optimaal model zónder IQ bleek dat het IQ van leerlingen weinig additionele waarde had bovenop de Cito-eindtoetscore van leerlingen.

bestanden exact worden gereproduceerd. Voor wat betreft het beoogde model lukt dit niet voor de volle honderd procent. Het probleem wordt gevormd door de groep instromers in leerjaar 3. Binnen VOCL'99 is weliswaar vast te stellen hoe groot de groep instromers in leerjaar 3 is, maar van deze groep zijn geen verdere gegevens bekend: niet voor wat betreft de rendement bovenbouw of centraal examencijfers, maar ook niet voor wat betreft achtergrondkenmerken of beginniveau. Om deze reden moet er in het beoogde model een aanpassing worden aangebracht. De correctiefactor 'percentage instroom in leerjaar 3' kan in het meerniveau model alleen op schoolniveau een plaats krijgen. De correctie impliceert dat de centraal examencijfers of bovenbouwrendement van leerlingen worden gecorrigeerd voor de mogelijke 'last' die ze ondervinden van een hoog percentage instromende klasgenoten.

De overige variabelen die in het beoogde model worden opgenomen zijn etniciteit, sociale afkomst (SES), sekse, en Cito-eindtoetsscore. Deze zijn als volgt geoperationaliseerd:

Etniciteit: De etnische afkomst van leerlingen is gebaseerd op het geboorteland van de leerling zelf en de ouders/verzorgers. Het is gemeten in zes categorieën, namelijk Nederlands, Turks, Marokkaans, ASA (Antilliaans, Surinaams of Arubaans), overig en onbekend. In de modellen zijn deze categorieën als dummy-variabele meegenomen, met de categorie 'onbekend' als referentiecategorie.

Sociale afkomst (SES): De sociaaleconomische status van leerlingen is vastgesteld aan de hand van het niveau van de hoogste door de ouders/verzorgers voltooide opleiding. Er worden zes niveaus onderscheiden: 1=maximaal basisonderwijs/lagere school, 2=voortgezet onderwijs, eerste trap, 3=voortgezet onderwijs, tweede trap, 4=hoger onderwijs, propedeuse, 5=hoger onderwijs, doctoraal, 6=doctoraat. SES is als ordinale variabele in de modellen opgenomen. Voor details wordt verwezen naar het technisch rapport over het eerste jaar van VOCL '99 (Kuyper, Lubbers & van der Werf, 2003).

Sekse: De sekse van de leerling is een dichotome variabele, met 1 voor meisje en 0 voor jongen.

Cito-eindtoetsscore: De score van de leerlingen op de Cito-eindtoets in groep 8 is variabele gemeten op een intervalniveau. De score varieert van minimaal 500 tot maximaal 550.

3.2.2 Berekening bovenbouwrendement

De huidige berekeningswijze van de opbrengstindicator rendement bovenbouw is gebaseerd op de proporties leerlingen die per leerjaar in de bovenbouw worden bevorderd en de proportie leerlingen dat slaagt voor het examen. In het vmbo heeft de berekening betrekking op de overgang van het derde leerjaar naar het vierde leerjaar, alsmede op de proportie geslaagde leerlingen. Voor de havo heeft het bovenbouwrendement betrekking op twee overgangen en een examenresultaat; voor het vwo is de berekening gebaseerd op drie overgangen en het examenresultaat. De proporties bevorderingen en de proportie geslaagden

voor het eindexamen worden met elkaar vermenigvuldigd en het eindresultaat van deze berekening is dan een fractie (een getal tussen 0 en 1) die de kans weergeeft dat een leerling op de betreffende VO-vestiging onvertraagd de bovenbouw doorloopt. Deze indicator wordt in het huidige model berekend op vestigingsniveau. Voor een voorbeeldberekening wordt verwezen naar technische toelichting van de Opbrengstenkaart 2012.

Bij het beoogde opbrengstenmodel wordt overgestapt naar een berekeningswijze op leerlingniveau. Vanaf het derde leerjaar wordt jaarlijks vastgesteld of een leerling een succesvolle overgang heeft gemaakt of niet. Een leerling is succesvol als hij over is gegaan naar een volgend leerjaar of verder is gegaan in een hoger schoolsoort op hetzelfde leerjaar of een diploma heeft gehaald. Alle andere overgangsvormen worden beschouwd als niet succesvol. Het bovenbouwrendement heeft dan twee mogelijke waarden voor een leerling: een leerling is in de bovenbouw steeds succesvol geweest (score 1) of niet (score 0). Dit maakt dat de modellen waarmee de analyses zullen worden gedaan logistische meerniveau modellen betreffen. Met behulp van deze analyses kan voor elke VO-vestiging de ‘kans’ op een onvertraagde schoolloopbaan in de bovenbouw worden berekend.

Figuur 3.1 Overzicht van de modellen van deelproject 1 die met elkaar vergeleken worden.

Lege model:	meerniveau model zonder correctiefactoren
Huidige model:	<p>lineaire regressie model met als correctiefactoren:</p> <ul style="list-style-type: none"> - percentage instromende leerlingen in leerjaar 3 - percentage leerlingen uit armoedeprobleemcumulatiegebied - percentage lwoo-leerlingen <p>Deze correctiefactoren zijn proxies voor het beginniveau van leerlingen</p>
Beoogde model:	<p>meerniveau model met als correctiefactoren:</p> <ul style="list-style-type: none"> - percentage instromende leerlingen in leerjaar 3 - leerling uit armoedeprobleemcumulatiegebied - leerling heeft lwoo <p>Deze correctiefactoren zijn proxies voor het beginniveau van leerlingen</p>
Optimale model:	<p>meerniveau model met als correctiefactoren:</p> <ul style="list-style-type: none"> - Cito-eindtoetsscore van een leerling (beginniveau meting) - SES van een leerling - Etniciteit van een leerling - Sekse van een leerling <p>De laatste drie factoren zijn niet-schoolse kenmerken die de leerprestaties kunnen beïnvloeden.</p>

3.2.3 Beschrijving van modellen en series analyses

In dit deelproject zullen viertal modellen worden toegepast (zie Figuur 3.1). In de eerste plaats is dat het *lege model*. Bij dit meerniveau model wordt een leerling- en schoolniveau in de

variantie onderscheiden, maar er vindt geen enkele correctie voor kenmerken van leerlingen of scholen plaats. Dit model markeert de ene kant van het spectrum dat in project beschouwd wordt. Het geeft weer wat de het bovenbouwrendement en gemiddeld centraal examencijfer van een VO-vestiging is zonder rekening te houden met het beginniveau van leerlingen. Dit wordt ook wel het bruto model genoemd. Aan de andere kant van het spectrum ligt het *optimale model*. Dit is het meerniveau model met ook een leerling- en vestigingsniveau waarbij de opbrengstindicator wordt gecorrigeerd voor kenmerken die samen zo nauwkeurig mogelijk beginniveau van leerlingen weergeven. Hiermee kan zo goed mogelijk de toegevoegde waarde van een school bepaald worden³; het begin- en eindniveau van leerlingen wordt methodologisch zo optimaal mogelijk in de modellen meegenomen. Het is voor de hand liggend om hiervoor kenmerken te nemen die informatie verstrekken over het gerealiseerde niveau van schoolprestaties in het basisonderwijs, alsmede het potentiële leervermogen van de leerlingen in het eerste leerjaar van het VO. De selectie van variabelen die zijn meegenomen in het optimale model is afgestemd op wat (toekomstig) haalbaar zou kunnen zijn voor de Inspectie van het Onderwijs. De twee modellen van de Inspectie – het *huidige model* en het *beoogde model*⁴ – kunnen vervolgens geplaatst worden ergens op de denkbeeldige lijn tussen het lege model aan de ene kant en het optimale model aan de andere kant. Het huidige model is geheel berekend volgens de operationalisering zoals die door de Inspectie wordt gehanteerd. Het betreft hier een multiple lineaire regressie analyse met naar schoolniveau geaggregeerde variabelen. De toegepaste correctiefactoren zijn het percentage van de schoolpopulatie dat afkomstig is uit achterstandswijken (apcg), het percentage leerlingen dat recht heeft op leerwegondersteuning (lwoo), en het percentage instromers in het derde leerjaar. Het beoogde model is daarentegen weer een meerniveau model met variantie op leerling- en schoolniveau. De toegepaste correctie factoren zijn apcg en lwoo, die op leerlingniveau in de modellen zijn meegenomen, en het percentage instromers in het derde leerjaar dat op schoolniveau is meegenomen. Als de twee modellen van de Inspectie sterk lijken op het lege model en in het

³ Het gaat hier feitelijk om een zo optimaal mogelijk model. Het kader van wat mogelijk is wordt bepaald door de dataverzameling van het VOCL'99 onderzoek die al heeft plaatsgevonden. Binnen VOCL'99 is voor tal van relevante kenmerken informatie verzameld, maar dit is nooit uitputtend. Ook is de in- en uitstroom van leerlingen en de klassensamenstelling niet meegenomen in het optimale model. Dit maakt dat het optimale model *niet* het model is waarmee volledig zuiver en nauwkeurig de bijdrage van een VO-vestiging aan de leerprestaties van leerlingen kan worden bepaald. Een dergelijk model valt buiten het spectrum van dit project, maar is in zijn algemeenheid met de huidige analysemethoden en beschikbare data een onbereikbaar doel.

⁴ Op verzoek van de Inspectie bij eerste bespreking van de resultaten is ook nog een kleine aanpassing op het beoogde model gedaan en doorgerekend: In dit *aangepaste beoogde model* is instroom in derde leerjaar op vestigingsniveau, en zijn apcg en lwoo op zowel leerling- als op vestigingsniveau zijn meegenomen. Over de belangrijkste resultaten van dit aangepaste beoogde model zal hooguit kort worden gerapporteerd in dit rapport.

geheel niet op het optimale model dan zou geconcludeerd moeten worden dat de betreffende modellen niet voldoen aan de verwachtingen; ze corrigeren onvoldoende voor het instroomniveau en samenstelling van de leerlingpopulatie en zijn daarom niet te beschouwen als toegevoegde waarde modellen.

De Inspectie van het Onderwijs onderscheidt vijf verschillen schoolsoorten bij de huidige Opbrengstenkaart VO. Van hoog naar laag zijn dit 1) vwo, 2) havo, 3) vmbo gemengde leerweg (gl) en theoretische leerweg (tl), 4) vmbo kaderberoepsgerichte leerweg (kl), en 5) vmbo basisberoepsgerichte leerweg (bl). De normering van de opbrengstindicatoren bovenbouwrendement en gemiddeld centraal examencijfer wordt gedaan binnen elk van de genoemde schoolsoorten (gestratificeerd). In dit deelproject wordt de werkwijze van de Inspectie van het Onderwijs zo nauwgezet mogelijk gevolgd. Dit betekent dat de vier te vergelijken modellen per opbrengstindicator steeds binnen elke schoolsoortgroep worden geanalyseerd. In totaal zijn 40 modellen geanalyseerd (5 schoolsoorten * 4 modellen * 2 opbrengstindicatoren).

Alvorens de resultaten van de vier modellen uitgebreid onderling te gaan vergelijken, is globaal onderzocht in hoeverre het ontbreken van waarden op de correctiefactoren van invloed op de resultaten. Omwille van de eenvoud is dit alleen gedaan voor het gemiddelde centraal examencijfer en wordt het huidige model alleen vergeleken met het lege en optimale model. Er zijn twee serie analyses uitgevoerd. Bij *serie 1* wordt een zo groot mogelijk aantal scholen betrokken. Hierbij zijn concessies gedaan voor wat betreft de keuze van individuele correctiefactoren voor het optimale model. Naast etniciteit, sociaal economische status (SES) en sekse, zullen de drie VOCL'99 entreetoetsen (afname in het eerste leerjaar VO; de entreetoetsen zijn te beschouwen als een verkorte versie van de CITO-eindtoets basisonderwijs) als correctiefactoren meegenomen worden in het model. Het is feitelijk een *suboptimaal* model omdat de keuze van correctiefactoren op leerlingniveau beperkt is. Bij *serie 2* komt het meenemen van de (meest belangrijke) correctiefactoren op de eerste plaats. Deze restrictie zal tot gevolg hebben dat een kleiner aantal scholen in de analyses van serie 2 zal worden betrokken. In het optimale model is naast etniciteit, SES en sekse, de score op de Cito-eindtoets en IQ in het model betrokken. Een vergelijking tussen de twee analysesreeksen zal uitwijzen of de analyses met een wisseling in het aantal scholen consequenties heeft voor de uitkomsten. Als blijkt dat de resultaten van serie 2 grote overeenkomsten vertonen met die van serie 1, dan zal in de verder analyses uitgegaan worden van de Cito-eindtoetsscore als kenmerk in het optimale model.

3.2.4 *Vergelijking tussen de modellen*

De analyses zullen voor elk van de vier modellen – lege, huidige, beoogde en optimale – steeds op exact dezelfde scholen én exact dezelfde groep leerlingen worden uitgevoerd. Dit

impliceert dat de modellen binnen een serie goed met elkaar vergeleken kunnen worden. Het uitgangspunt voor de analyses is steeds exact hetzelfde basismateriaal op leerlingniveau.

De vier modellen resulteren in vier verschillende schoolresiduen per opbrengstindicator per VO-vestiging. Het schoolresidu is de afstand van bijvoorbeeld het bovenbouwrendement van een VO-vestiging tot het gemiddelde bovenbouwrendement van alle VO-vestigingen die in analyses zijn betrokken (overall-gemiddelde). Op dezelfde manier worden ook vier schoolresiduen per VO-vestiging verkregen voor het gemiddelde centraal examencijfer. Schoolresiduen kunnen worden gebruikt voor de normering van VO-vestigingen. Het geeft weer of een VO-vestiging boven of onder verwachting heeft gepresteerd. Verandering in de waarde van een schoolresidu door toepassing van een ander model geeft inzicht in hoeverre de gekozen berekeningswijze van invloed is op de normering van een VO-vestiging. De mate van samenhang tussen de schoolresiduen van de vier modellen zal worden uitgedrukt door middel van Pearson's correlaties coëfficiënten (r).

3.2.5 *Consistentie van centraal examencijfers voor de vakkenclusters*

Bij dit valideringsonderzoek is het van belang om ook de betrouwbaarheid van de opbrengstindicatoren niet uit het oog te verliezen. Betrouwbaarheid is namelijk een noodzakelijke voorwaarde voor validiteit. Het gaat hier speciaal om de opbrengstindicator gemiddeld centraal examencijfer. De vraag is of een gemiddeld examencijfer van een VO-vestigingen een goede afspiegeling vormt van de onderliggende examencijfers per vak of vakkencluster. Dit wordt de interne consistentie genoemd. Een veelgebruikte maat hiervoor is Cronbach's alpha. Gelet op het doel van de opbrengstindicator zou deze minimaal een waarde van 0,70 moeten hebben. De vuistregel is dat voor 'low stakes testing' de interne consistentie –minimaal 0,70 moet zijn, bij 'high stakes testing' geldt de ondergrens van 0,90. Welke van type 'testing' hier van toepassing is staat ter discussie, maar gelet op het openbaar maken van de opbrengstindicatoren, zou een interne consistentie van minimaal 0,90 wenselijk zijn.

3.3 Resultaten

3.3.1 *Invloed van ontbrekende data op modellen*

De analyses worden uitgevoerd op de databestanden van het VOCL'99 cohort. Hoewel deze databestanden een bijzondere rijke schat aan informatie heeft van een grote groep VO-leerlingen op individueel niveau, zijn de ontbrekende waarden op de bepaalde kenmerken een beperking als men deze als covariaat in een meerniveau analyse wil opnemen.

Tabel 3.1 Beschrijvende statistieken van individuele correctiefactoren in VOCL'99

	Aantal leerlingen	Gemiddelde	Standaard Deviatie	Minimum	Maximum
IQ	9445	102.79	13.29	59	145
Etniciteit	13279	1.52	1.29	.00	5.00
SES	13279	2.77	1.38	.00	5.00
Sekse	13279	1.51	0.50	1.00	2.00
Entreetoets taal	12593	12.64	3.73	1	20
Entreetoets rekenen	12576	12.47	4.37	1	20
Entreetoets informatie verwerking	12374	12.04	4.11	1	20
Cito-eindtoetsscore	7465	536.30	8.89	505	550
Geldig aantal leerlingen	4803				

Om enigszins aan het probleem van ontbrekende data tegemoet te komen zijn er twee series analyses uitgevoerd voor de indicator gemiddeld centraal examencijfer. Bij serie 1 is het aantal VO-vestigingen gemaximaliseerd ($n = 279$) en worden zes covariaten meegenomen in het *suboptimale* model (etniciteit, SES en sekse en de drie entreetoetsen). Bij serie 2 is maximaal gecorrigeerd voor individuele verschillen. Het gaat hier om 180 vestigingen en er worden vijf covariaten in het optimale model betrokken (etniciteit, SES en sekse, IQ en Cito-eindtoetsscore). Bij serie 2 zijn dus de scores op de Entreetoetsen, rekenen, taal en informatieverwerking uit het model van serie 1 vervangen door de score op de Cito-eindtoets en is IQ eraan toegevoegd.

Serie 1

Bij serie 1 - waarbij een zo groot mogelijk aantal scholen is betrokken - zijn de uitkomsten van het lege model en het huidige model voor de VO-vestigingen met elkaar gekruist. Daartoe zijn de schoolresiduen gerangordend en toegewezen aan een van de vijf categorieën volgens het 10%-15%-50%-15%-10% systeem (de 'bolletjes'- systematiek). Tabel 3.2 toont hiervan de resultaten.

 Tabel 3.2 Het huidige model gekruist met het lege model voor het gemiddelde centraal examencijfer, *serie 1*

Percentielen model	huidige	Percentielen lege model					Totaal
		<10	10 - 25	25-75	75 - 90	>90	
<10		19	6	1	0	0	26
10 - 25		5	26	10	0	0	41
25 - 75		2	9	120	10	0	141
75 - 90		0	0	8	26	8	42
>90		0	0	2	6	21	29
Totaal		26	41	141	42	29	279

Er blijkt een grote mate van overeenkomst tussen de twee berekeningswijze te bestaan. De sterke overeenkomst manifesteert zich op de hoofddiagonaal (grijs) en blijkt ook uit de bijbehorende Cohen's kappa: 0,65. De volgende stap is het vergelijken van het huidige model met het *suboptimale* model. Tabel 3.3 toont de resulterende categorieverdeling van deze twee modellen.

Tabel 3.3 Het huidige model gekruist met het *suboptimale* model voor het gemiddelde centraal examencijfer, *serie 1*

Percentielen huidige model	Percentielen optimale model					Totaal
	<10	10 - 25	25-75	75 - 90	>90	
<10	15	8	2	0	1	26
10 - 25	9	18	14	0	0	41
25 - 75	2	15	111	9	4	141
75 - 90	0	0	10	26	6	42
>90	0	0	4	7	18	29
Totaal	26	41	141	42	29	279

Ook bij een vergelijking van het huidige model met het suboptimale model is nog steeds een royale overeenstemming te zien. De meeste VO-vestigingen bevinden zich op de hoofddiagonaal. De Cohen's kappa is 0,52. Gerekend over de gehele hoofddiagonaal blijkt dat het huidige model op twee derde deel van de VO-vestigingen volledig in overeenstemming is met het suboptimale model. Echter, één op de drie VO-vestigingen komt in een andere categorie terecht.

Serie 2

Bij serie 2 - waarbij zoveel mogelijk correctiefactoren worden meegenomen - zijn schoolresidu verdelingen van de 180 VO-vestigingen van het lege model en het huidige model met elkaar gekruist (Tabel 3.4).

Tabel 3.4 Huidige model gekruist met het optimale model voor het gemiddelde centraal examencijfer, *serie 2*

Percentielen huidige model	Percentielen optimale model					Totaal
	<10	10 - 25	25-75	75 - 90	>90	
<10	5	7	4	0	0	16
10 - 25	8	10	9	0	0	27
25 - 75	3	10	64	12	2	91
75 - 90	0	0	11	8	6	25
>90	0	0	3	5	13	21
Totaal	16	27	91	25	21	180

Wat opvalt in Tabel 3.4 is dat de mate van overeenstemming in de indeling van scholen in de vijf beoordelingscategorieën vrij sterk terugloopt in vergelijking met Tabel 3.3. Uitgedrukt in Cohen's kappa is de mate van overeenstemming tussen de beide berekeningswijzen 0,35. Overigens is de mate van overeenstemming tussen het lege model en het huidige model wel tamelijk sterk (Cohen's kappa=0,65). Het huidige model lijkt dus veel meer op het lege model dan op het optimale model.

Uit het voorgaande kan worden opgemaakt dat het in feite weinig uitmaakt of men analyses uitvoert op de steekproef met 279 VO-vestigingen (*serie 1*) dan wel op de steekproef met 180 VO-vestigingen (*serie 2*). In beide gevallen is de overeenkomst tussen het lege model en het huidige model uit te drukken in termen van Cohen's kappa ter grootte van 0,65. Vanwege de grote mate van overeenkomst zijn de overige analyses in dit deelonderzoek uitgevoerd op de steekproef met 180 VO-vestigingen. Het grote voordeel is dat dan alle denkbare modellen kunnen worden vergeleken met het optimale model. Op verzoek van de Inspectie van het Onderwijs zal het IQ van de leerlingen niet langer in het optimale model betrokken worden omdat dit nooit op grote schaal beschikbaar zal zijn voor leerlingen in het VO.

Het optimale model vormt een vast ijkpunt voor de uitkomsten van de overige drie modellen:- het lege model, het huidige model en het beoogde model. Vergelijking van de schoolresiduen tussen de modellen maakt inzichtelijk wat de positie van het lege, huidige en beoogde model is ten opzichte van het optimale model. Hieronder zullen eerst de modellen voor het bovenbouwrendement besproken worden (3.3.2), gevolgd door de modellen voor het gemiddelde centraal examencijfer (3.3.3).

3.3.2 *Vergelijking van modellen voor bovenbouwrendement*

Om de positie van het beoogde en huidige model in het spectrum van enerzijds het lege model en anderzijds het optimale model te bepalen, zijn meerniveau analyses uitgevoerd met de kleine steekproef als uitgangspunt (*serie 2*). De resultaten van de drie meerniveau modellen staan in Tabel 3.5. Het vierde model – het huidige model – betreft een lineaire regressie model met alleen variabelen op schoolniveau. De resultaten van dit model zijn hier niet gepresenteerd omdat vergelijking van een dergelijk model met de meerniveau modellen niet zinvol is.

De drie modellen zijn steeds berekend met data van 1.815 leerlingen op 36 verschillende VO-vestigingen (Tabel 3.5). Van de proxies voor het beginniveau in het beoogde model is alleen apcg significant gerelateerd aan het bovenbouwrendement van havo-leerlingen. Er is sprake van een negatief verband: havo-leerlingen uit een armoedeprobleemcumulatiegebied hebben in het algemeen een lager bovenbouwrendement dan havo-leerlingen die niet in een dergelijk gebied wonen. Het beoogde model wijkt weinig af van het lege model voor het havo; apcg is het enige kenmerk dat het verschil maakt. Voor het optimale model geldt dat de kenmerken Cito-eindtoetsscore, SES en sekse geen onafhankelijk significant effect hebben op het

bovenbouwrendement van leerlingen. Gelet op de geringe verschillen tussen de drie modellen, kan gesteld worden dat het bovenbouwrendement van havo-leerlingen niet af nauwelijks afhankelijk is van het beginniveau, zowel in het beoogde als in optimale model. Een vergelijking van het beoogde model met het lege of optimale model in termen van verklaarde variantie is niet mogelijk omdat het hier logistische regressie modellen betreft.

Tabel 3.5 Resultaten van de meerniveau modellen voor het bovenbouwrendement op havo afdelingen.

	Lege model		Beoogde model		Optimale model	
	Coëfficiënt	SE	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed part:#</i>						
Constante	1,174	0,079**	1,122	0,114**	7,364	5,482
% Instroom in 3 ^o leerjaar			1,145	0,663		
Apcg			-0,326	0,134*		
Cito-eindtoetscore					-0,012	0,010
SES					0,053	0,043
Meisje					0,115	0,112
<i>Random part:</i>						
Tussen VO-vestiging verschillen	0,098	0,050	0,075	0,044	0,096	0,050
Aantal VO-vestigingen	36		36		36	
Aantal leerlingen	1815		1815		1815	

* $p < 0.05$; ** $p < 0,01$ # lwoo ontbreekt in het beoogde model omdat op het havo geen lwoo-leerlingen zitten. Etniciteit ontbreekt in het optimale model als correctiefactor in verband met convergentie problemen bij het schatten van het model.

Soortgelijke tabellen voor het vwo, vmbo-tl/gl, vmbo-bl en vmbo kl staan in de Appendix bij hoofdstuk 3, Tabellen 1 tot en met 4. lwoo is als correctiefactor alleen op het vmbo van toepassing.

In Tabel 3.6 worden de correlatie coëfficiënten getoond voor de samenhang tussen de schoolresiduen van de vier modellen voor het bovenbouwrendement. Hierbij zijn alle schoolsoorten weer samengenomen.

Tabel 3.6 Correlaties tussen het bovenbouwrendement op vestigingsniveau op basis van de vier onderscheiden modellen (n = 180 VO-vestigingen)

	Lege model	Huidige model	Beoogde model	Optimale model
Lege model	1			
Huidige model	0,83**	1		
Beoogde model	0,89**	0,79**	1	
Optimale model#	0,99**	0,80**	0,89**	1

** $p < 0,01$ (twee-zijdig)

Tabel 3.6 toont dat de vier modellen voor wat betreft de berekening van het bovenbouwrendement onderling nagenoeg uitwisselbaar zijn. Opvallend daarbij is de grote overeenkomst tussen het lege model en het optimale model ($r=0,99$). Een mogelijke verklaring hiervoor is dat de correctiefactoren in het optimale model vrijwel nooit significant zijn (zie Appendix bij hoofdstuk 3, Tabellen 1 tot en met 4). Op verzoek van de Inspectie van het Onderwijs is ook nog een aangepast versie van het beoogde model doorgerekend (maar niet gepresenteerd in Tabel 3.6). Ook het *aangepaste* beoogde model, waarbij de correctiefactoren instroom in derde leerjaar op vestigingsniveau, en apcg en lwoo op zowel leerling- als op vestigingsniveau zijn meegenomen, resulteert in hoge correlaties met het beoogde model ($r=0,98$), lege model ($r=0,92$) en optimale model ($r=0,91$). Het enige model voor bovenbouwrendement die lichtelijk uit de pas loopt, is het huidige model. Maar ook hier geldt dat de correlatie met de overige modellen van ongeveer 0,80 als sterk te beschouwen is.

Nu is het in principe mogelijk dat bij het ene model het bovenbouwrendement van een VO-vestiging qua schatting veel lager uitvalt dan bij een ander model, terwijl de correlatie tussen de schattingen van het bovenbouwrendement van beide modellen nagenoeg perfecte samenhang aanduiden. Er is dan sprake van systematische verschillen tussen de modellen. Of dit ook hier het geval is is nader onderzocht. Tabel 3.7 laat de resultaten zien.

Tabel 3.7 Beschrijvende statistiek van het bovenbouwrendement berekend op basis van de vier modellen, uitgedrukt in proporties.

	Aantal VO-vestigingen	Bovenbouwrendement		
		Gemiddeld	Minimum	Maximum
Lege model	180	0,90	0,43	0,97
Huidige model	179	0,86	0,60	1,00
Beoogde model	180	0,90	0,52	0,97
Optimale model #	180	0,87	0,41	0,95

** $p < 0,01$ (twee-zijdig), # exclusief IQ

Het gemiddelde bovenbouwrendement varieert tussen de 87% en 90% voor de vier modellen. De minimale en maximale waarden per model zijn ook vrijwel gelijk. Voor het aangepaste beoogde model zijn vergelijkbare waarden voor het gemiddelde (90%) en range gevonden (minimum 56%; maximum 97%). Er is geen sprake van systematische verschillen tussen de meerniveau modellen. Hoewel het huidige model ook hier weer lichtelijk uit de pas loopt, kan gesteld worden dat de vier verschillende rendementsberekeningen elkaar nauwelijks ontlopen voor de opbrengstindicator bovenbouwrendement.

3.3.3 Vergelijking van modellen voor centraal examencijfer

De volgende stap in de analyses is het vergelijken van de positie van het beoogde en huidige model ten opzichte van het lege model en het optimale model voor het gemiddelde centraal examencijfer voor alle vakken.

De resultaten van de drie meerniveau modellen staan in Tabel 3.8. De drie modellen zijn berekend met data van 1.108 leerlingen op 46 verschillende VO-vestigingen. De fit van het beoogde model en optimale model is beter dan die van het lege model (beide, $p < 0.001$). Van de proxies voor het beginniveau in het beoogde model heeft geen van de twee kenmerken een eigen onafhankelijk effect op het gemiddelde centraal examencijfer van havo-leerlingen. Samen zorgen ze ervoor dat het model toch beter bij de data past dan het lege model. Voor het optimale model geldt dat alleen de Cito-eindtoetsscore en sekse een eigen significant effect hebben op het gemiddelde centraal examencijfer. Havo-leerlingen met hoge Cito-eindtoetsscore behalen een hoger gemiddeld centraal examencijfer dan havo-leerlingen met een lage Cito-eindtoetsscore. En havo-jongens behalen in het algemeen een hoger gemiddeld centraal examencijfer dan havo-meisjes. Het beoogde model beter past bij de data dan het lege model, maar het beoogde model lijkt iets minder goed te fitten dan het optimale model. Soortgelijke tabellen voor vwo, vmbo-tl/gl, vmbo-bl en vmbo kl staan in de Appendix bij hoofdstuk 3 (Tabellen 5 tot en met 8).

Tabel 3.8 Resultaten van de meerniveau modellen voor het centraal examencijfer havo

	Lege model		Beoogde model		Optimale model	
	Coëfficiënt	SE	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed part:</i>						
Constante	6,367	0,031	6,376	0,048	-9,067	2,116
% instroom in 3 ^e leerjaar			-0,038	0,284		
apcg*			0,000	0,049		
Cito-eindtoetsscore					0,029	0,004
Etniciteit: #						
Nederlands					0,062	0,233
Turks					-0,25	0,293
Marokkaans					-0,086	0,276
Antiliaans, Surinaams, Arubaans					-0,194	0,257
Overig					0,048	0,239
SES					0,014	0,015
Meisje					-0,15	0,037
<i>Random Part:</i>						
Tussen VO-vestiging verschillen	0,023	0,009	0,022	0,009	0,02	0,008
Tussen leerlingen verschillen	0,381	0,016	0,385	0,017	0,353	0,015
<i>Model fit (-2* logLikelihood):</i>	2112,194		2032,656		2024,535	
Aantal VO-vestigingen	46		46		46	
Aantal leerlingen	1108		1108		1108	

* apcg is dichotome variabele (1 = woonachtig in een armoedeprobleemcumulatiegebied, 0 = niet woonachtig in een armoedeprobleemcumulatiegebied); # etniciteit onbekend is referentie-categorie.

Meer uitsluitsel over welke model het beste is wordt verkregen door de schoolresiduen van de vier modellen onderling te vergelijken. In Tabel 3.9 wordt de samenhang tussen de

schoolresiduen van de vier modellen getoond in de vorm van correlatie coëfficiënten. Hierbij zijn de schoolresiduen van de vijf schoolsoorten samengenomen. De aangepaste versie van het beoogde model die op verzoek van de Inspectie is doorgerekend is niet in Tabel 3.9 weergegeven. Dit model lijkt sterk om het beoogde model ($r=0,98$), maar vertoont beduidend minder samenhang met het lege model ($r=0,36$) en het huidige model ($r=0,37$).

Tabel 3.9 Correlaties tussen het gemiddelde centraal examencijfer op vestigingsniveau op basis van de vier onderscheiden modellen (n = 180 VO-vestigingen).

	Lege model	Huidige model	Beoogde model	Optimale model
Lege model	1			
Huidige model	0,86**	1		
Beoogde model	0,46**	0,48**	1	
Optimale model	0,002	0,01	-0,10	1

** p<0,01 (twee-zijdig)

Uit Tabel 3.9 wordt duidelijk dat de modellen voor het gemiddelde centraal examencijfer uiteenvallen in een tweetal groepen, waarbij de ene groep uit slechts één model bestaat: het optimale model. De uitkomsten van het optimale model correleren hooguit zwak met de overige modellen: de bijbehorend correlatie coëfficiënten variëren van $r=0,002$ tot $r=0,11$ (richting van verband buiten beschouwing gelaten). De overige modellen – lege, huidige en (aangepast) beoogde model – vormen samen de tweede groep. De onderlinge samenhang varieert van matig (bijvoorbeeld aangepast beoogd vs. leeg: $r=0,46$) tot zeer sterk (beoogd vs. aangepast beoogd: $r=0,96$). Met betrekking tot het laatste resultaat kan worden gesteld dat het niet uitmaakt of het beoogde model in zijn oorspronkelijke vorm wordt gebruikt (met alleen covariaten op leerlingniveau, behalve % instroom leerjaar 3), dan wel in aangepaste vorm (met covariaten op zowel leerling- als vestigingsniveau).

3.3.4 Consistentie van de centraal examencijfers

In het voorgaande is aandacht besteed aan verschillen tussen VO-vestigingen met betrekking tot het gemiddelde centraal examencijfer voor alle vakken. Dit cijfer is opgebouwd uit de centraal examencijfers op uiteenlopende vakken. Als het gemiddelde centraal examencijfer voor alle vakken een goede afspiegeling is van de centraal examencijfers op de onderliggende vakken, dan mag verwacht worden dat er een positieve correlatie is tussen deze onderliggende centraal examencijfers. Tabel 3.10 laat de correlatiematrix zien waarbij – op vestigingsniveau – alle examenvakken onderling worden gecorreleerd. Zo wordt de consistentie zichtbaar in de mate waarin de vakken onderling correleren.

Tabel 3.7 laat zien dat er sprake van een geringe consistentie; een hoog gemiddeld centraal examencijfer op het ene vakkencluster is geen goede voorspeller van het gemiddelde centraal examencijfer op een ander vakkencluster. Onder het gemiddelde centraal examencijfer voor alle vakken van een VO-vestiging liggen vrij sterk uiteenlopende gemiddelde examencijfers op de vakkenclusters. De sterkste samenhang is gevonden tussen de bètavakken en economie: $r=0,36$.

Tabel 3.7 Correlatie coëfficiënten tussen vakspecifieke centraal examencijfers op schoolniveau, (n = 180 VO-vestigingen)

	Nederlands	Engels, Frans, Duits	Aardrijkskunde, geschiedenis	Economie	Wis-, natuur- scheikunde, biologie
Nederlands	1				
Engels, Frans, Duits	0,25**	1			
Aardrijkskunde, geschiedenis	0,10	0,23*	1		
Economie	0,11	0,17	0,30**	1	
Wis-, natuur-, scheikunde, biologie	0,24**	0,29**	0,19	0,36**	1

***) Correlatie is significant bij $\alpha=0.01$ (2-zijdig).

*) Correlatie is significant bij $\alpha=0.05$ (2-zijdig).

Of de mate van consistentie tussen de gemiddelde vakkenclustercijfers samenhangt met het model dat is toegepast, is nader onderzocht. Daartoe is de Cronbach's alpha berekend voor het lege, huidige, beoogde en optimale model. Het datamateriaal dat is gebruikt voor het berekenen van de correlaties zijn de gemiddeld centraal examencijfers op schoolniveau. Tabel 3.8 toont de resultaten.

Tabel 3.8 Cronbach's alpha en correlatie voor vijf indicatoren van het gemiddelde centraal examencijfer van een vestiging: het vestigingsgemiddelde van het centraal examencijfer voor Nederlands, moderne vreemde talen; aardrijkskunde en geschiedenis, economie en bètavakken

	Cronbach's alpha	Gemiddelde correlatie tussen indicatoren
Lege model	0,63	0,25
Huidige model	0,66	0,28
Beoogde model	0,65	0,28
Optimale model	0,89	0,63

De interne consistentie van het lege, huidige en beoogde model zijn dusdanig laag dat niet gesproken kan worden van een goede afspiegeling (Tabel 3.8). Echter, het optimale model laat een ander beeld zien. In dit model is het gemiddelde centraal examencijfer wel een goede afspiegeling van de onderliggende centraal examencijfers op de vijf vakkenclusters. De correlatie tussen centraal examencijfer van de vijf vakkenclusters is 0,63. De bijbehorende Cronbach's alpha is 0,89. Dit is maar een fractie lager dan 0,90, de waarde die als vuistregel geldt voor de ondergrens voor betrouwbaarheid van 'high stake' meetinstrumenten.

4 Deel II: Onderbouwrendement

4.1 Inleiding

Eén van de manieren waarop de Inspectie van het Onderwijs de opbrengsten van VO-vestigingen beoordeelt is door naar het rendement in de onderbouw te kijken. Het onderbouwrendement is een maat waarin informatie met betrekking tot doubleren, onderwijsposities (combinatie leerjaar en schoolsoort) en adviezen van de basisschool wordt gecombineerd. Het geeft de voortgang van schoolloopbaan van leerlingen weer nadat de eerste twee jaar van het VO zijn voltooid. Een leerling zit dan – als hij voldoet aan de verwachting aan het eind van de basisschool en niet is blijven zitten – in het begin van het 3^e leerjaar van de schoolsoort die overeenkomt met zijn advies. Zijn onderbouwrendement is 100. Maar een leerling kan in die twee jaar ook doubleren of afstromen naar een lagere schoolsoort. Hij doet het slechter dan verwacht; zijn onderbouwrendement is lager dan 100. Een leerling kan het natuurlijk ook beter doen dan verwacht door op te stromen naar een hogere schoolsoort of een klas over te slaan. Zijn onderbouwrendement is dan hoger dan 100. Een onderbouwrendement van 0 wordt verkregen als een leerling twee schoolsoorten is afgestroomd – bijvoorbeeld een havo advies heeft gekregen op de basisschool en na 3 jaar op het vmbo kaderberoepsgericht leerweg zit - en intussen niet is blijven zitten.

Het rendement van leerlingen in de onderbouw kan gemiddeld worden over de VO-vestiging. Dit wordt door de Inspectie *onderbouwrendement* genoemd en is één van de opbrengstindicatoren van de Opbrengstenkaart VO. Omdat bij het onderbouwrendement een begin- en een eindniveau van leerlingen aan elkaar gerelateerd worden, is deze maat te beschouwen als een indicator voor de toegevoegde waarde.

In het huidige opbrengstenmodel van de Inspectie van het Onderwijs wordt het onderbouwrendement retrospectief en op vestigingsniveau berekend. Per groep leerlingen in een bepaalde schoolsoort in leerjaar 3 wordt teruggekeken met welke adviezen ze het VO binnenkwamen, en welk deel van hen doubleerde in de eerste en tweede klas van het VO. Een nadeel van deze werkwijze is dat er aannames over verdelingen van adviezen en doublures op groepsniveau aan ten grondslag liggen. Aannames op groepsniveau hoeven niet overeen te komen met de werkelijkheid op individueel niveau. Dit kan schijnverbanden tot gevolg hebben.

Door invoering van het onderwijsnummer is inmiddels naar schatting van de Inspectie voor circa 65% van de leerlingen in het VO voldoende data beschikbaar om het onderbouwrendement te berekenen met individuele schoolloopbaan-, advies- en eindtoetsgegevens. De gegevens liggen opgeslagen in de Bron-bestanden voor het primair en voorgezet onderwijs. Ze kunnen ook worden gebruikt om het onderbouwrendement

prospectief te berekenen. De groep leerlingen die nieuw instroomt in het eerste leerjaar van het VO vormt dan de te selecteren groep voor wie het onderbouwrendement berekend wordt. De onderwijsposities van deze groep leerlingen wordt vastgesteld na twee overgangsmomenten: na één jaar en na twee jaar vanaf het moment van start in het VO. Op- en afstroom van leerlingen, evenals zittenblijven, is hierdoor direct zichtbaar in de onderwijsposities van de leerlingen na twee jaar. Een prospectieve benadering doet meer recht aan de intuïtieve betekenis van het onderbouwrendement dan een retrospectieve benadering. Het beoogt iets te zeggen over hoe efficiënt leerlingen die starten in VO-1 de eerste twee jaren van het VO doorlopen.

De onderzoeksvraag die op verzoek van de Inspectie van het Onderwijs voorligt betreft de mate van overeenkomst tussen een prospectieve en een retrospectieve benadering van het onderbouwrendement. De eerste onderzoeksvraag voor deelproject II luidt als volgt:

- 1. In welke mate wijkt de huidige retrospectieve bepaling van het onderbouwrendement af van een prospectieve bepaling van het onderbouwrendement met individuele data?*

In het huidige opbrengstenmodel VO wordt geen rekening gehouden met de basisschool waar de leerlingen van afkomstig zijn. Het ligt echter voor de hand dat het onderbouwrendement beïnvloed kan zijn door de toeleverende basisscholen, aangezien men mag verwachten dat eventuele lange termijn effecten van basisscholen voornamelijk in de eerste jaren van het VO een rol kunnen spelen. Verder is de waarde van het onderbouwrendement van een VO-school direct afhankelijk van de kwaliteit van het advies van de toeleverende basisscholen. Het is mogelijk dat VO-vestigingen ‘last’ hebben van PO-vestigingen die inadequate adviezen afgeven. Als een VO-school te maken heeft met toeleverende PO-vestigingen die stelselmatig te hoge adviezen geven – overadviseren –, dan heeft dit zeer waarschijnlijk gevolgen voor hun onderbouwrendement. De verwachtingen van de leerlingen van dergelijke PO-vestigingen zijn te hoog gespannen. Ze hebben grotere kans om af te stromen naar een lagere schoolsoort of te doubleren. Voor de VO-vestigingen uit dit zich in een lager rendement in de onderbouw. Zo worden ze de dupe van de kwaliteit van de adviezen van de toeleverende PO-vestigingen. Bij de ontwikkeling van een nieuw opbrengstenmodel VO is het daarom zinvol om de lange termijn invloed van PO-vestigingen en de invloed van inadequate adviezen nader te onderzoeken. De bijbehorende tweede en derde onderzoeksvraag van deelproject II zijn:

- 2. In hoeverre hebben de toeleverende PO-vestigingen invloed op de waarde van het onderbouwrendement van een VO-vestiging?*
- 3. Welke rol spelen over- en onderadvisering hierbij?*

Op verzoek van de Inspectie van het Onderwijs zal daarnaast ook nog speciale aandacht uitgaan naar de volgende aanvullende vraagstellingen.

- *Is het wenselijk om VO-vestigingen met lwoo-leerlingen te compenseren voor doubleren?*
- *Hoe kan het beste omgegaan worden met de diverse combinaties van schoolsoorten in de onderbouw van VO-vestigingen ?*
- *Hoe kan het beste omgegaan worden met instromende leerlingen in leerjaar 1 en 2? Aan welke VO-vestiging moeten deze leerlingen worden toegekend?*
- *In welke mate is het advies van leerlingen afhankelijk van het onderwijsaanbod op de ontvangende VO-vestigingen?*

De laatste drie aanvullende vragen vallen buiten de analyses die voor de hoofdvraagstellingen nodig zijn. Ze zijn te beschouwen als discussievragen. Daarom worden deze vragen niet in dit hoofdstuk beantwoord, maar komen ze terug bij algemene discussie in Hoofdstuk 7.

4.2 Methode

4.2.1 Huidige berekeningswijze onderbouwrendement

In de Technische Toelichting van de Opbrengstenkaart 2012 (Inspectie van het Onderwijs, 2012) wordt uitgelegd hoe het onderbouwrendement voor de huidige opbrengstkaarten VO is berekend. Daarbij wordt onderscheid gemaakt in het al dan niet hebben van leerwegondersteunend onderwijs (lwoo). Voor leerlingen zonder lwoo geldt de volgende formule:

$$\text{Onderbouwrendement} = 100 * ((\text{schoolsoort jaar 3} - \text{advies VO} - \text{aantal doublures} + 2) / 2)$$

Voor leerlingen met lwoo geldt een aangepaste formule:

$$\text{Onderbouwrendement} = 100 * ((\text{schoolsoort jaar 3} - \text{advies VO} - \text{aantal doublures} + 2,25) / 2)$$

Deze aangepaste formule impliceert dat 1 op de 4 leerlingen met lwoo een jaar mag doubleren zonder dat dat de school wordt aangerekend. Informatie over lwoo is afkomstig uit het bron VO bestand van 2010-2011. Lwoo is een binaire variabele die aangeeft of een leerling in dat schooljaar in aanmerking kwam voor extra leerweg ondersteuning. Een leerling kan in aanmerking komen voor lwoo als hij:

- een leerachterstand heeft op ten minste twee van de vier domeinen: inzichtelijk rekenen, begrijpend lezen, technisch lezen en spellen;
- een IQ heeft dat ligt tussen de waarden 75 tot en met 90;
- een IQ heeft dat ligt tussen de waarden 91 tot en met 120 in combinatie met sociaal-emotionele problematiek.

De uitleg over berekeningswijze van het huidige onderbouwrendement gaat uit van data van individuele leerlingen omdat op deze manier de gebruikers van de Opbrengstenkaart beter zouden begrijpen hoe het onderbouwrendement berekend wordt. In werkelijkheid werkt de Inspectie met data op groepsniveau - om precies te zijn op het niveau van een schoolsoort binnen een VO-vestiging -, en gaat men terug in de tijd voor leerlingen die in de derde klas VO zitten (retrospectief). De werkwijze is als volgt:

Jaarlijks vraagt de Inspectie de benodigde gegevens op bij de VO-vestigingen door middel van het elektronische Adviesformulier. De vestigingen moeten daarin voor alle leerlingen die zijn overgegaan naar leerjaar 3 aangeven wat hun adviezen VO zijn en of ze gedoubleerd zijn. Dit gaat per aangeboden schoolsoort in leerjaar 3. Bijvoorbeeld een brede VO-vestiging van school A biedt in het derde leerjaar de schoolsoorten vmbo-tl, havo en vwo aan. In schooljaar 2012-2013 gaat het om bijvoorbeeld 140 vmbo-tl-, 100 havo-; 50 vwo-leerlingen. Per schoolsoort moet worden aangegeven wat de adviesverdeling is: van de 150 vmbo-tl-leerlingen hebben bijvoorbeeld 15 leerlingen een vmbo-bl/kl, 90 een vmbo-tl, 35 een havo en 10 een vwo-advies. Ook wordt per schoolsoort aangegeven hoeveel leerlingen in leerjaar 1 en leerjaar 2 zijn blijven zitten en hoeveel *niet* zijn blijven zitten in deze twee leerjaren. Bijvoorbeeld van de 150 vmbo-tl-leerlingen zijn in het eerste leerjaar 5 leerlingen en in het tweede leerjaar 15 leerlingen blijven zitten, 130 leerlingen zijn niet blijven zitten. Voor de schoolsoorten vmbo-bl, kl en gl/tl moet ook worden opgegeven hoeveel leerlingen wel en niet lwoo hebben. Op het vmbo-tl van de voorbeeldschool hebben 30 leerlingen wel en 120 leerlingen geen lwoo. Het zo verkregen databestand vormt de basis van de huidige berekening van het onderbouwrendement door de Inspectie. Daarbij worden de volgende stappen gezet:

- De schoolsoort in leerjaar 3 wordt omgezet naar een getal van 1 voor vmbo-bl tot 5 voor vwo (*schoolsoortcijfer*⁵). Het schoolsoortcijfer voor bijvoorbeeld vmbo-tl is 3.
- De adviezen VO krijgen dezelfde getallen (*adviescijfer*) en lopen dus ook op van 1 voor vmbo-bl tot 5 voor vwo. Dubbel-adviezen krijgen het tussenliggende halve getal. Een havo-vwo advies heeft het adviescijfer 4,5.
- De proportie leerlingen met lwoo wordt vastgesteld voor de schoolsoorten vmbo-bl, kl en gl/tl. Lwoo-leerlingen mogen een kwart jaar langer over de onderbouw doen. In het vmbo-tl voorbeeld is de proportie lwoo-leerlingen $30/150=0,20$. Deze proportie wordt vermenigvuldigd met 0,25 en opgeteld bij het schoolsoortcijfer voor vmbo-tl: $0,20*0,25+3=3,05$ (*schoolsoortcijfer met lwoo-correctie*).
- Het onderbouwrendement per schoolsoort in leerjaar 3 wordt gesteld op 1 (=100%).
- Hiervan wordt afgetrokken de som van het aantal doublures in leerjaar 1 en 2 samen, gedeeld door twee keer het totaal aantal leerlingen waarvan de doublurestatus bekend

⁵ Door Inspectie onderwijspositie genoemd

is. Dit is gemiddelde doublurekans in de onderbouw. In het voorbeeld over vmbo-tl: $1 - (20/(2*150)) = 1 - 0,067 = 0,933$ (*onderbouwrendement met doublure-correctie*).

- Dan wordt de afstand tussen de schoolsoort in leerjaar 3 en het advies VO verrekend. Is het advies hoger dan de schoolsoort in leerjaar 3, dan wordt het onderbouwrendement naar beneden bijgesteld. Is het advies lager dan de schoolsoort in leerjaar 3, dan wordt het onderbouwrendement naar boven bijgesteld. Om precies te zijn is de berekening als volgt. Er wordt uitgegaan van het onderbouwrendement met doublure correctie. Bij dit getal wordt opgeteld het verschil tussen het schoolsoortcijfer met lwoo-correctie en adviescijfer, vermenigvuldigd met het aantal leerlingen dat dit adviescijfer heeft. Dit wordt gedaan voor alle voorkomende adviescijfer en uiteindelijk gedeeld door 2* het aantal leerlingen met een geldig advies.

In het vmbo-tl voorbeeld:

Onderbouwrendement vmbo-tl, School A =

$$100*(0,933 + (((3,05-2,5)*15 + (3,05-3)*90 + (3,05-4)*35 + (3,05-5)*10))/2*150) = 78,8$$

Vervolgens wordt het gewogen gemiddelde berekend van het onderbouwrendement van de schoolsoorten die in het derde jaar van een vestiging aanwezig zijn. Op school A gaat het om het onderbouwrendement van het vmbo-tl, havo en vwo. Het gewogen gemiddelde van de drie schoolsoorten op school A is het onderbouwrendement van school A. De Inspectie heeft in aanvulling hierop twee voorwaarden gesteld waaraan de berekening van het gewogen onderbouwrendement van een vestiging moet voldoen:

- tenminste 60% van de leerlingen op een vestiging heeft een bruikbaar advies;
- het aantal leerlingen op een vestiging met een bruikbaar advies is minimaal 10.

Het gewogen onderbouwrendement wordt niet toegewezen aan de vestiging waar de VO-3 leerlingen van afkomstig zijn, maar aan de vestiging waar deze leerlingen in VO-2 zaten. Hierdoor ontstaan geen toewijzingsproblemen voor junior-colleges met alleen de eerste twee leerjaren van het VO. Vanaf VO-3 gaan leerlingen van junior-colleges naar een andere VO-vestiging, meestal wel van dezelfde brin.

Normering

Het onderbouwrendement van een vestiging wordt genormeerd door te bepalen in hoeverre het onderbouwrendement afwijkend is van het gemiddelde onderbouwrendement van VO-vestigingen die vergelijkbare schoolsoorten aanbieden. Welke schoolsoorten dat zijn wordt afgeleid uit de inschrijfgegevens van leerjaar 1 uit het meest recente bron-bestand. De Inspectie hanteert zeven vergelijkingsgroepen:

1. Catorale vwo-vestigingen (aanduiding vwo);
2. Havo-vwo-vestigingen (aanduiding havo-vwo);

3. Categorale vmbo-tl/gl, mavo en avo-vestigingen (aanduiding vmbo gl/tl);
4. Vmbo-gl/tl tot havo, categorale havo en vmbo gl-tl-havo-vwo-vestigingen (aanduiding vmbo-gl/tl-havo-vwo);
5. Vestigingen met alleen vmbo-bl en/of vmbo-kl (aanduiding vmbo-bl/kl)
6. Vmbo-breed-vestigingen en vmbo-breed tot havo-vestigingen (aanduiding vmbo-breed-havo)
7. Vestigingen met breed aanbod van vmbo-bl tot en met vwo (aanduiding vmbo-breed-havo-vwo).

4.2.2 *Retrospectief onderbouwrendement met individuele data*

Om de huidige retrospectieve bepaling van het onderbouwrendement goed te kunnen vergelijken met een prospectieve bepalingswijze (4.2.3) wordt de huidige werkwijze van de Inspectie (4.2.1) zo nauwkeurig mogelijk nagedaan met individuele leerlinggegevens afkomstig uit het meest recente schooljaar: Bron VO 2012-2013 bestand. Schooljaar 2012-2013 is het eerste schooljaar waarin voldoende adviezen van leerlingen in het derde leerjaar VO uit het Bron PO bestanden gekoppeld kunnen worden (Bron PO 2008-2009).

In het Bron VO 2012-2013 bestand zitten 197.754 leerlingen die in schooljaar 2012-2013 in het derde leerjaar VO zitten. Een deel van deze leerlingen doet VO-3 (n = 13.131) echter voor de tweede keer. Ze zijn niet meegenomen in de verdere analyses van dit project omdat de Inspectie deze groep leerlingen ook niet betreft in haar huidige berekeningen. Er blijven 184.623 leerlingen over die op 1144 verschillende VO-vestigingen zitten. Ze vormen het uitgangspunt voor de retrospectieve berekeningen van het onderbouwrendement en worden in de navolgende tekst aangeduid met *VO-3 leerlingen*.

Door het ontbreken van data zullen de berekeningen niet steeds op alle VO-3 leerlingen gebaseerd zijn. Hieronder volgt een overzicht van de databewerkingen die zijn uitgevoerd om retrospectief het onderbouwrendement van een vestiging te kunnen berekenen, met als doel deze opbrengstindicator te baseren op zo veel mogelijk leerlingen en vestigingen en zoveel mogelijk volgens de huidige werkwijze van de Inspectie. Praktijkonderwijs-leerlingen komen niet voor in de Bron VO bestanden. Wel leerlingen met een Pro-advies die toch naar een vmbo-opleiding gaan. Pro-vestigingen hebben niet de verplichting om data aan te leveren over de onderwijsposities van hun leerlingen. Ze vallen buiten de berekeningen van het rendementonderbouw.

Advies VO: Het bestand Bron PO 2009-2010 bevat de adviezen die de VO-3 leerlingen hebben gekregen toen ze in groep 8 van de basisschool zaten. Na koppeling via het onderwijsnummer blijkt van 129.691 leerlingen (70%) een advies beschikbaar te zijn. Hierbij zitten 665 leerlingen die een advies hebben van die tenminste vier schoolsoorten breed (vmbo-bl t/m havo, vmbo-kl t/m vwo, vmbo-bl t/m vwo). In overeenstemming met de huidige

werkwijze van de Inspectie worden deze adviezen te breed bevonden en op missing gezet. De bruikbare adviezen VO worden omgezet in een adviescijfer op de volgende manier:

- Praktijk onderwijs (Pro) = 0
- Vmbo-bl = 1
- Vmbo-bl-kl = 1,5
- Vmbo-kl = 2
- Vmbo-kl-tl/gl = 2,5
- Vmbo-tl/gl of avo = 3
- Vmbo-tl/gl-havo = 3,5
- Havo = 4
- Havo-vwo = 4,5
- Vwo/atheneum/gymnasium = 5.

Aantal doublures: Van de VO-3 leerlingen wordt vastgesteld of ze gedoubleerd zijn in leerjaar 1 of leerjaar 2. Door de Bron-bestanden VO 2010-2011 en 2011-2012 te koppelen aan het Bron VO 2012-2013 bestand, kan de onderwijspositie van de leerlingen en daarmee ook het leerjaar van zittenblijven worden vastgesteld. Voor 176 leerlingen kon geen onderwijspositie worden vastgesteld, ondanks een aantal aannames bij onlogische schoolloopbaan patronen en missing data. Het aantal doublures van deze leerlingen is als missing gemarkeerd.

Schoolsoort in het derde leerjaar VO: VO-3 leerlingen met een onbekende of een te brede schoolsoort - tenminste vier schoolsoorten breed - in het derde leerjaar vormen een probleem voor de berekening van het onderbouwrendement; ze worden op missing gezet. Er zijn 6.083 leerlingen met een te brede schoolsoort (brugjaar), leerlingen met een onbekende schoolsoort komen niet voor. De bruikbare schoolsoorten in leerjaar 3 worden omgezet in een schoolsoortcijfer op de volgende manier:

- Vmbo-bl = 1
- Vmbo-kl = 2
- Vmbo-tl/gl/avo = 3
- Havo = 4
- Vwo/atheneum/gymnasium = 5.

De volgende stap in de databewerking is het aggregeren van de individuele data van VO-3 leerlingen naar het niveau van de aangeboden schoolsoort in leerjaar 3 op een VO-vestiging. Dit is het niveau waarop de Inspectie de data van de vestigingen ontvangt. Het gaat bij deze aggregatie om de sommatie van aantal leerlingen. Op 11 vestigingen is het derde leerjaar voor alle leerlingen een brugjaar. Dit is te breed voor het berekenen van het onderbouwrendement. Deze vestigingen worden verwijderd. Per aangeboden schoolsoort in VO-3 van een vestiging

is weergegeven het aantal leerlingen met een bepaald advies(cijfer), het aantal leerlingen dat is blijven zitten in leerjaar 1 of 2 of niet is blijven zitten, en het aantal leerlingen dat wel en geen lwoo heeft. Het databestand komt nu qua opbouw overeen met het bestand dat voortvloeit uit het elektronische Adviesformulier. Met deze data is het onderbouwrendement berekend per aangeboden schoolsoort in VO-3 van een vestiging volgens de bij 4.2.1 beschreven werkwijze. Conform de werkwijze van de Inspectie zijn twee voorwaarden gesteld voor het presenteren van het onderbouwrendement van een VO-vestiging: 1) op een vestiging moet tenminste 60% van de leerlingen een bruikbaar advies hebben, en 2) het aantal leerlingen op een vestiging met voldoende bruikbare adviezen is minimaal 10. Aan de eerste voorwaarde voldeden 292 vestingen niet, en aan de tweede nog eens 6 vestingen niet. Daarmee komt het aantal vestingen met een gewogen onderbouwrendement uit op 835.

Het onderbouwrendement wordt niet toegewezen aan de vestiging waar de VO-3 leerlingen in schooljaar 2012-2013 zitten, maar aan de vestiging waar deze leerlingen in schooljaar 2011-2012 zaten; VO-2 voor alle leerlingen. De VO-3 leerlingen zaten in hun twee leerjaar op 1148 verschillende vestigingen. Koppeling van het gewogen onderbouwrendement is succesvol voor 802 vestigingen. Uiteindelijk is het mogelijk gebleken om op 70% van de vestigingen waar de VO-3 leerlingen in hun tweede leerjaar zaten het gewogen gemiddelde onderbouwrendement te berekenen. Dit wordt in de navolgende tekst *retrospectief onderbouwrendement* van een vestiging genoemd.

Normering

Voor het normeren van het prospectieve onderbouwrendement zijn dezelfde zeven vergelijkingsgroepen gemaakt als die de Inspectie ook hanteert. De 802 vestigingen zijn toegewezen aan één van de zeven vergelijkingsgroepen op grond van de schoolsoorten die in VO-1 van schooljaar 2012-2013 op de vestiging zijn aangeboden. In hoeverre het onderbouwrendement van een vestiging afwijkend is van het gemiddelde onderbouwrendement van soortgelijke vestigingen wordt bepaald door het verschil te berekenen. De relatieve positie van een vestiging ten opzichte van het vergelijkingsgroeps-gemiddelde wordt zo duidelijk. Een positieve waarde betekent dat het onderbouwrendement van de VO-vestiging boven verwachting is, een negatieve waarde betekent het omgekeerde.

4.2.3 Prospectief onderbouwrendement met individuele data

Voor het beantwoorden van *onderzoekvraag 1* zal het prospectieve onderbouwrendement van een VO-vestiging worden vergeleken met een prospectieve berekeningswijze van het onderbouwrendement. Bij een prospectieve benadering wordt uitgegaan van leerlingen die starten in VO-1 en van wie vervolgens wordt vastgesteld hoe efficiënt ze de eerste twee jaar van het VO doorlopen. Het basisbestand voor de berekening van het prospectieve onderbouwrendement is het Bron-bestand VO 2010-2011. Het bestand bevat alle 194.251 leerlingen die in schooljaar 2010-2011 in het eerste leerjaar VO zitten (VO-1). De leerlingen

verblijven inmiddels 3 jaar in het VO (schooljaar 2012-2013). Voor hen is het mogelijk om het onderbouwrendement prospectief te berekenen. Een deel van de leerlingen doet VO-1 in 2010-2011 voor de tweede keer. Welke leerlingen dat zijn is te achterhalen door het Bronbestand PO 2009-2010 te koppelen via het (versleutelde) onderwijsnummer van de leerlingen. Het gaat om 14.144 leerlingen; ze zijn niet meegenomen in de verdere analyses. Er blijven 180.107 leerlingen over. Ze vormen het uitgangspunt voor de prospectieve berekening van onderbouwrendement. Door het ontbreken van data lukt dat echter voor een deel van de leerlingen niet. Hieronder volgt een overzicht van de databewerkingen die zijn uitgevoerd om tot een dataset te komen waarin van alle VO-1 leerlingen alle benodigde gegevens beschikbaar zijn.

Advies VO: Het bestand Bron-bestand PO 2009-2010 bevat de adviezen VO die de leerlingen hebben gekregen in groep 8 van de basisschool. Na koppeling via het onderwijsnummer blijkt van 138.910 leerlingen een advies beschikbaar te zijn. Dat is 77% van de leerlingen die voor het eerst VO-1 doen en hoger dan aanvankelijk verwacht werd. Verder zijn er 722 leerlingen met (te) brede adviezen (Vmbo-bl t/m havo, vmbo-kl t/m vwo, vmbo-bl t/m vwo). In overeenstemming met de huidige werkwijze van de Inspectie worden de leerlingen met deze brede adviezen uitgesloten van de verdere berekeningen. De bruikbare adviezen worden omgezet in een adviescijfer op dezelfde manier als beschreven is in 4.2.2.

Aantal doublures: De volgende stap is het vaststellen of leerlingen gedoubleerd zijn in het eerste of tweede jaar na start in VO-1. Door de Bron-bestanden VO 2011-2012 en 2012-2013 te koppelen aan het Bron-bestand VO 2010-2011, kan de onderwijspositie van de leerlingen worden vastgesteld in deze schooljaren en daarmee ook hun doublurestatus. Voor 778 leerlingen lukt dit niet, ondanks een aantal aannames over onlogische patronen en missing waarden. Ze zijn uit de dataset verwijderd.

Schoolsoort jaar 3: Conform de huidige werkwijze van de Inspectie worden leerlingen met een te brede of onbekende schoolsoort drie jaar na start in het VO niet meegenomen. Van een te brede schoolsoort is sprake als een leerling in schooljaar 2012-2013 in een brugjaar zit ($n = 5.895$). Het merendeel van hen zit in leerjaar 3 (72%), 28% zit in leerjaar 2 en is dus blijven zitten. Verder zijn er nog 3.894 leerlingen met een onbekende schoolsoort in schooljaar 2012-2013. Hierdoor vallen in totaal weer 9.789 leerlingen af. De bruikbare schoolsoorten in leerjaar 3 zijn omgezet in een schoolsoortcijfer (zie 4.2.2)

Uiteindelijk blijven er nog 127.621 VO-1 leerlingen over. Dat is 71% van het oorspronkelijke aantal leerlingen. Ze zullen in de rest van dit rapport met *VO-1 leerlingen* worden aangeduid. Dit zijn dus de leerlingen die in schooljaar 2010-2011 voor de eerste keer in VO-1 zaten, van wie het advies bekend en niet te breed is, van wie de schoolsoort in schooljaar 2012-2013 bekend en niet te breed is, en van wie de doublurestatus bekend is. De wijze waarop het prospectieve onderbouwrendement is berekend voor de individuele VO-1 leerling volgt exact

de werkwijze beschreven in de ‘Technische Toelichting’ van de Opbrengstenkaart 2012. De daarbij gebruikte formules zijn in 4.2.1 gepresenteerd.

Vestigingsgemiddelde

Om het prospectieve onderbouwrendement van een vestiging vast te stellen wordt het gemiddelde genomen van alle beschikbare individuele onderbouwrendementen op de betreffende VO-vestiging. Bij deze aggregatie naar vestigingsniveau wordt het onderbouwrendement toegekend aan de vestiging waar een leerling in schooljaar 2011-2012 zat, veelal VO-2 voor de leerling. De VO-1 leerlingen zaten in schooljaar 2011-2012 op 1132 verschillende vestigingen. Het gemiddeld aantal leerlingen per vestiging is 113, variërend van 1 tot 522 leerlingen.

Bij huidige werkwijze van de Inspectie voor het berekenen van het onderbouwrendement per vestiging geldt de volgende voorwaarde: van tenminste 60% van de leerlingen op een vestiging is het advies VO beschikbaar en het aantal leerling met een bruikbaar advies is minimaal 10. Als niet aan deze voorwaarden wordt voldaan wordt het onderbouwrendement van een vestiging niet weergegeven op de Opbrengstenkaart VO. Een bruikbaar advies van een leerling houdt niet automatisch in dat het onderbouwrendement van de leerling prospectief berekend kan worden. De onderwijspositie na drie jaar blijkt voor veel leerlingen toch niet beschikbaar of bruikbaar te zijn in de bestanden (zie hierboven). Daarom wordt voorgesteld om de huidige voorwaarden te vervangen door de volgende twee voorwaarden:

- Het aantal leerlingen met een onderbouwrendementsbepaling per vestiging is minimaal 10;
- Van tenminste 60% van de leerlingen op een vestiging moet een onderbouwrendement berekend kunnen worden.

Ten gevolge van de eerste nieuwe voorwaarde blijven nog 1103 van de 1132 vestigingen over met 127.435 leerlingen met bekend onderbouwrendement. Als ook de tweede voorwaarde wordt toegepast resteren nog 854 vestigingen met 110.719 leerlingen met bekend onderbouwrendement. Dat is 75% van het oorspronkelijke aantal vestigingen. Van een kwart van de vestigingen is dus onvoldoende data beschikbaar voor een valide berekening van het onderbouwrendement. In de navolgende tekst wordt dit het prospectieve onderbouwrendement van een vestiging genoemd.

Normering

De normering van het vestigingsgemiddelde van het prospectieve onderbouwrendement is uitgevoerd op dezelfde manier als bij het prospectieve onderbouwrendement (4.2.2). De 854 vestigingen met voldoende data worden vergeleken met één van de zeven vergelijkingsgroepen op grond van de schoolsoorten die op de vestiging werden aangeboden

in VO-1 van schooljaar 2010-2011. Het verschil tussen onderbouwrendement van een VO-vestiging en het gemiddelde onderbouwrendement van soortgelijke vestigingen is de maat die de relatieve positie van een VO-vestiging ten opzichte van het vergelijkingsgroepsgemiddelde weergeeft. Een positieve waarde voor het verschil betekent dat het onderbouwrendement van de VO-vestiging beter is dan verwacht, een negatieve waarde betekent het omgekeerde.

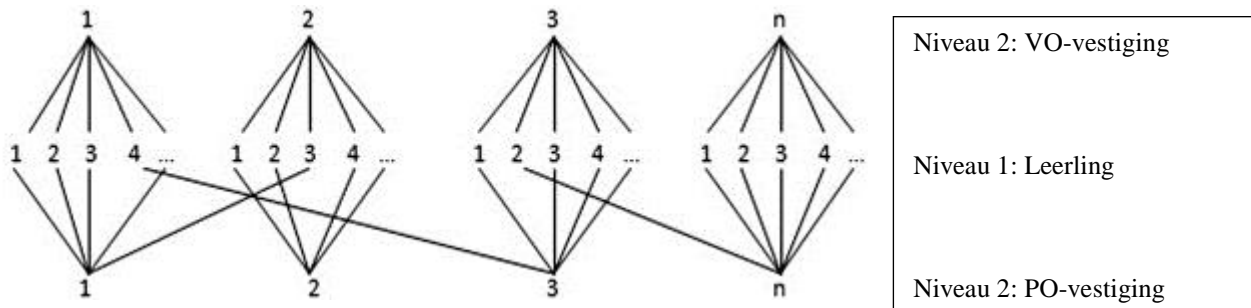
4.2.4 Invloed van PO-school op het prospectieve onderbouwrendement

Twee aspecten van de invloed van de PO-school op het prospectieve rendement zullen in dit deelonderzoek nader onderzocht worden. Ten eerste is dat het algemene lange termijn effect van het onderwijs op de basisschool (*onderzoeksvraag 2*). Naar verwacht zal dit doorwerken in de hoogte van onderbouwrendement van een VO-vestiging. VO-vestiging profiteren van basisscholen die hun leerlingen goed hebben voorbereid op het VO of die het maximale uit een leerlingen hebben weten te halen. Dergelijke leerlingen hebben een voorsprong op hun klasgenoten. Ten tweede is het onderbouwrendement van een VO-vestiging direct afhankelijk is van de kwaliteit van de adviezen van de toeleverende basisscholen (*onderzoeksvraag 3*). Structurele over- en onderadvisering leiden tot vertekening van het onderbouwrendement; dit wordt respectievelijk over- of onderschat.

Kruisklassificatie modellen

Met behulp van meerniveau kruisklassificatie modellen is het mogelijk om de lange termijn invloed van de PO-school op het prospectieve onderbouwrendement te schatten (Snijders & Bosker, 2012). In dit type meerniveau modellen zijn de leerlingen niet alleen genest in VO-vestigingen (niveau 2) maar ook in PO-vestigingen (niveau 1). Er wordt dus twee verschillende niveaus 2 onderscheiden. Figuur 4.1 laat een voorbeeld van de structuur van een kruisklassificatie model zien.

Figuur 4.1 Structuur van een kruisklassificatie model (gebaseerd op Figuur 13.1 in Snijders & Bosker, 2012)



De individuele leerlingen zijn geplaatst in het midden van Figuur 4.1, de diverse VO-vestigingen staan bovenaan. Dit stuk van de figuur toont de basale structuur van een meerniveau model. Van links af gezien zijn de eerste vier leerlingen genest binnen VO-vestiging nummer 1. Met schoolwisselingen wordt geen rekening gehouden. Aangenomen wordt dat alle leerlingen één VO-vestiging bezoeken. De totale variantie in onderbouwendement wordt verdeeld in een stuk variantie tussen leerlingen en een stuk variantie tussen VO-vestigingen. De aanname hierbij is dat niet alleen leerlingen maar ook VO-vestigingen onderling verschillen in het prospectieve onderbouwendement.

Onderaan in Figuur 4.1 staan de PO-vestigingen. Niet alle leerlingen van een VO-school zijn afkomstig van één PO-vestiging. Bijvoorbeeld, leerlingen 1 tot en met 3 van VO-vestiging 1 zijn afkomstig van PO-vestiging 1, maar leerling 4 is afkomstig van PO-vestiging 3. Leerlingen van een PO-vestiging gaan normaliter naar meerdere VO-vestigingen. Bij kruisklassificatie modellen wordt met deze mogelijkheid rekening gehouden. Bij analyses met dergelijke modellen kan de invloed van een PO-vestiging op het onderbouwendement van een VO-vestiging goed geïsoleerd worden en wordt duidelijk in hoeverre de verschillen in onderbouwendement tussen VO-vestigingen te herleiden zijn naar verschillen tussen PO-vestigingen. Bijkomend voordeel van kruisklassificatie modellen is dat relevante kenmerken van PO-vestigingen op het juiste niveau als covariaat kunnen worden meegenomen. In dit hoofdstuk gaat het om het gemiddelde prestatieniveau van de leerlingen in groep 8, geoperationaliseerd door de score op de Cito-eindtoets van de individuele van de PO-vestiging (*Kruisklassificatie model*).

In de navolgende analyses zullen de resultaten van een leeg meerniveau model (zonder covariaten) en kruisklassificatie model (met de covariaten Cito-eindtoetsscore en mate van overadvisering) onderling worden vergeleken. De modellen zijn geschat met behulp van het software pakket MLwin versie 2.27 (Rasbash, Steele, Browne, et al., 2012). Voor de schatting van de coëfficiënten en varianties van deze twee modellen zijn niet de gebruikelijke maximum likelihood schattingsmethoden toegepast, maar is gekozen voor Markov Chain Monte Carlo schattingsmethoden (Browne, 2009). Deze methode is flexibeler voor analyseren van kruisklassificatie modellen (Browne, Goldstein & Rasbash, 2001).

Rekening houden met de invloed van PO-scholen op het prospectieve onderbouwendement zal niet voor alle VO-vestigingen van even groot belang zijn. Om nader te onderzoeken of er een algemene typering gemaakt kan worden van VO-vestigingen waar de grootste verschuivingen optreden, zullen de VO-vestigingen gecategoriseerd worden op basis van de schoolresiduen van het lege meerniveau en kruisklassificatie model. De indeling van VO-vestigingen wordt weergegeven in Tabel 4.1.

Tabel 4.1 Typering van het vertekening van onderbouwrendement van VO-vestigingen door wel en niet rekening te houden met de invloed van de PO-vestigingen

		Schoolresidu kruisklassificatie model		
		Onder gemiddeld	Gemiddeld	Boven gemiddeld
Schoolresidu leeg meerniveau model	Onder gemiddeld	<i>Neutraal</i>	<i>Te negatief</i>	<i>Veel te negatief</i>
	Gemiddeld	<i>Te positief</i>	<i>Neutraal</i>	<i>Te negatief</i>
	Boven gemiddeld	<i>Veel te positief</i>	<i>Te positief</i>	<i>Neutraal</i>

De percentielen van de verdeling van de schoolresiduen zijn als uitgangspunt genomen; de p10 markeert de grens van het ‘onder gemiddeld’ en de p90 de grens van het ‘boven gemiddeld’. Door de keuze van de afkappingen p10 en p90 richt de typering van VO-vestigingen op de meeste extreme waarden. Op basis van het schema in Tabel 4.1 zijn drie groepen VO-vestigingen onderscheiden. In de eerste plaats zijn dit VO-vestigingen die profiteren van de invloed van de PO-school. Het prospectieve onderbouwrendement wordt te positief ingeschat als geen rekening wordt gehouden met de invloed van de toeleverende PO-scholen zoals in het lege meerniveau model (*‘(veel) te positief’*). De tweede groep is de neutrale groep. Ze liggen op de diagonaal van Tabel 4.1 (*‘neutraal’*). Het wel of niet meenemen van de PO-vestiging in de analyses niet tot noemenswaardige veranderingen in het onderbouwrendement. De derde groep is de groep VO-vestigingen die last heeft van de kwaliteit van de toeleverende PO-vestigingen (*‘(veel) te negatief’*). Het prospectieve onderbouwrendement wordt te laag geschat bij een leeg meerniveau model. Als wél rekening wordt gehouden met de toeleverende PO-scholen dan komt het prospectieve onderbouwrendement op deze groep gunstiger uit. De drie groepen worden onderling vergeleken wat betreft aanwezig schoolsoorten (*vergelijkingsgroep*), gemiddelde Cito-eindtoetscore, adviezen (*adviescijfer*) en schoolsoort 3 jaar na start in het VO (*schoolsoortcijfer*) van hun leerlingen en het percentage doublures in leerjaar 1 en 2, allochtone leerlingen en lwoo-leerlingen.

Mate van overadvisering

Inventarisatie van vertekening door in adequate advisering begint met het vaststellen van in hoeverre PO-scholen adviezen afgeven die te hoog of te laag zijn gezien de score op de Cito-eindtoets van de leerlingen. Uitgangspunt hierbij zijn de PO-vestigingen die toeleverancier zijn voor de VO-vestigingen met voldoende data om het prospectieve onderbouwrendement te berekenen (854 VO-vestigingen met 110.719 leerlingen). Niet op al deze PO-vestigingen wordt de Cito-eindtoets in groep 8 afgenomen. Een deel van de leerlingen valt om deze reden af. Resterend voor de analyses zijn 69.801 leerlingen die op 3.428 verschillende PO-vestigingen hun Cito-eindtoets hebben gemaakt.

Met behulp van een meerniveau model is de mate van overadvisering van een PO-vestiging vastgesteld, conform een onderzoek van de Boer, Bosker en van de Werf (2010). In dit model wordt de samenhang tussen de score op de Cito-eindtoets (onafhankelijke variabele) en het advies van de leerlingen (*Advies PO*, afhankelijke variabele, zie 4.2.3) bepaald, waarbij rekening wordt gehouden met het feit dat leerlingen i (niveau 1) genest zijn binnen PO-vestigingen j (niveau 2). Het meerniveau model met variantie op leerlingen- en PO-vestigingsniveau heeft de volgende vergelijking:

$$\text{Advies } PO_{ij} = \beta_{0ij} + \beta_1 \text{Cito} - \text{eindtoetsscore}_{ij} + \mu_{0j} + e_{0ij}$$

Het schoolresidu μ_{0j} in dit model geeft aan of de betreffende PO-vestiging gemiddeld genomen adviezen afgeeft die hoger of lager zijn dan op grond van de score op de Cito-eindtoets mag worden verwacht. Een negatief schoolresidu duidt op onderadvisering, een positief schoolresidu op overadvisering. De schoolresiduen worden bewaard als variabele onder de naam 'mate van overadvisering'. Als voor onder- en overadvisering de grenswaarden van respectievelijk -0,5 en +0,5 schoolsoort worden gehanteerd, dan is op 2% van de 3438 PO-vestigingen sprake van onderadvisering en op 1% sprake van overadvisering⁶.

4.2.5 Compensatie voor lwoo-leerlingen bij bepaling van onderbouwrendement

Volgens de huidige berekeningswijze van het (retrospectieve) onderbouwrendement geldt voor leerlingen met leerwegondersteunend onderwijs (lwoo) een aangepaste formule voor het berekenen van het onderbouwrendement (zie paragraaf 4.2.1). De aanpassing impliceert dat men accepteert dat 25% van de lwoo-leerlingen een jaar doubleert in de eerste twee jaar van het VO. De eerste aanvullende onderzoeksvraag heeft hier betrekking op. De vraag is of lwoo-leerlingen daadwerkelijk zo frequent blijven zitten in vergelijking tot niet-lwoo leerlingen.

Dit is onderzocht op basis van het Bron-VO 2010-2011 bestand. Het bestand bevat leerlingen die in schooljaar 2010-2011 in het eerste leerjaar VO zitten (VO-1) en inmiddels drie jaar in het VO verblijven (schooljaar 2012-2013). Een deel van de leerlingen heeft in 2010-2011 VO-1 voor de tweede keer gedaan. Ze zijn verwijderd uit het databestand. Van de resterende leerlingen (n=180.107) is de onderwijspositie na 1 en 2 jaar VO gekoppeld en de doublurestatus bepaald. Bij 1074 leerlingen is dit niet geluk, er blijven 179.033 leerlingen over. Van deze groep heeft 12,1% in VO-1 lwoo gekregen en 5,4% is minimaal 1 keer blijven zitten. De doublurefrequentie van de leerlingen met lwoo is vergeleken met die van leerlingen zonder lwoo. Daarbij is rekening gehouden met de schoolsoort van de leerlingen in VO-3.

⁶ Door rekening te houden met de geneste structuur van de data worden minder extreme waarden voor over- en onderadvisering van PO-vestiging gevonden dan wanneer een lineaire regressie analyse zou worden toegepast op de data. Uit een lineaire regressie analyse met dezelfde grenswaarden volgt dat op 92% van de PO-scholen adviezen worden gegeven die in overeenstemming zijn met de scores op de Cito-eindtoets, overadvisering komt voor op 3% van de PO-vestigingen, onderadvisering op 5%.

4.3 Resultaten

4.3.1 Retrospectief onderbouwrendement met individuele data

De beschrijvende statistiek van het vestigingsgemiddelde van het retrospectieve onderbouwrendement bij de diverse stappen van de databewerking is weergegeven in Tabel 4.2.

Tabel 4.2 Retrospectief onderbouwrendement bij de verschillende voorwaarden voor presentatie op de Opbrengstenkaart VO

	N	Gemiddelde	SD	min	max
Alle vestigingen	1128	95,7	11,9	31,6	197,6
Vestigingen met min 60% geldige adviezen	841	96,3	11,4	44,7	197,6
Vestigingen met min 60% geldige adviezen en minimaal 10 leerlingen met geldig advies	835	96,1	11,0	44,7	197,6
Vestigingen met min 60% geldige adviezen, minimaal 10 leerlingen met geldig advies en toe te wijzen aan VO-2 op vestiging	802	96,3	10,8	57,3	197,6

Op basis van het meest recente Bron-bestand is het mogelijk om het retrospectieve onderbouwrendement te berekenen voor 1128 van de 1144 beschikbare VO-vestigingen. Het overall gemiddelde onderbouwrendement op deze vestigingen is 95,7, met een standaard deviatie van 11,9. Door toepassing van de twee voorwaarden voor presentatie en door het toewijzen van het onderbouwrendement aan de vestiging waarop de VO-3 leerlingen een schooljaar eerder zaten, valt bijna 30% van de vestigingen af ($n = 802$). Er lijkt sprake te zijn van selectieve uitval, gezien de stijging van het overall gemiddelde onderbouwrendement naar 96,3, een kleiner wordende standaard deviatie en een minimale waarde die gestegen is naar 57,3. De ‘slechtst’ presterende scholen vallen af. Hier speelt waarschijnlijk mee dat op kleine VO-scholen het onderbouwrendement relatief instabiel is.

4.3.2 Prospectief onderbouwrendement met individuele data

Om een eerste indruk te krijgen van de groep VO-1 leerlingen van wie op individueel niveau het onderbouwrendement berekend kan worden zijn in Tabel 4.3 en 4.4 enkele beschrijvende statistieken gepresenteerd. De tussen haakjes vermelde getallen in Tabel 4.3 voor adviescijfer en schoolsoortcijfer zijn gebruikt voor de berekening van het individueel onderbouwrendement met de formules beschreven in paragraaf 2.1. De percentages tellen op tot 100%.

Tabel 4.3 Beschrijvende statistieken van VO-1 leerlingen voor wie onderbouwrendement prospectief is berekend (n = 127.621)

Leerlingkenmerk		Procentuele verdeling
Advies VO (adviescijfer)	Pro (0)	0,2
	Vmbo-bl (1)	7,6
	Vmbo-bl-kl (1,5)	4,2
	Vmbo-kl (2)	8,1
	Vmbo-kl-gt/tl (2,5)	3,4
	Vmbo-gl/tl/avo (3)	19,0
	Vmbo-gl/tl-havo (3,5)	10,7
	Havo (4)	15,1
	Havo-vwo (4,5)	15,2
	Vwo/atheneum/gymnasium (5)	16,6
Doublurestatus in eerste 2 jaar VO	0	97,6
	1x	2,4
	2x	0,0
Schoolsoort jaar 3 (schoolsoortcijfer)	Vmbo-bl (1)	10,1
	Vmbo-kl (2)	14,6
	Vmbo-gl/tl/avo (3)	28,4
	Havo (4)	22,4
	Vwo (5)	24,4

Uit Tabel 4.3 volgt dat bijna 98% van de VO-1 leerlingen onvertraagd in VO-3 aankomt. Dit is hoger dan het landelijk gemiddelde percentage. Dit valt te verklaren doordat vertraagde leerlingen in het algemeen meer missing data hebben en voor hen het individuele onderbouwrendement niet berekend kan worden. Ook leerlingen die in de onderbouw VO in- of uitstromen maken geen deel uit van de berekeningen omdat ook zij missing data hebben; het advies, het schoolsoort in 3e leerjaar VO of het aantal doublures is voor deze leerlingen niet vast te stellen.

De beschrijvende statistieken van het prospectieve onderbouwrendement van de VO-1 leerlingen met voldoende bruikbare data zijn gepresenteerd in Tabel 4.4 per schoolsoort. Na twee jaar VO behalen de VO-1 leerlingen een gemiddeld rendement van 96,2. Dit komt goed overeen met het overall vestigingsgemiddelde van het retrospectieve onderbouwrendement (Tabel 4.2: 96,3). In zijn algemeenheid wordt op hogere schoolsoorten een hoger onderbouwrendement gehaald. Het gemiddelde prospectieve onderbouwrendement loopt op van 82,1 op het vmbo-bl tot 113,2 op het vwo. Uitzondering op deze regel vormt de

schoolsoort waar de meeste leerlingen na 2 jaar VO zitten - vmbo-tl/gl of (m)avo -; hier is het prospectieve onderbouwrendement lager dan op het vmbo-kl.

Tabel 4.4 Prospectief onderbouwrendement op leerlingniveau per schoolsoort na 3 jaar VO

Schoolsoort	N	Gemiddelde	SD	min	max
Vmbo-bl	12.947	82,1	38,1	-100	162,5
Vmbo-kl	18.646	92,6	43,2	-50	212,5
Vmbo-gl/tl/avo	36.220	89,8	38,0	-50	262,5
Havo	28.623	95,2	30,7	0	300
Vwo	31.185	113,2	22,3	0	312,5
Totaal	127.621	96,2	35,6	-100	312,5

Prospectieve onderbouwrendement op vestigingsniveau

De beschrijvende statistieken van het vestigingsgemiddelde van prospectieve onderbouwrendement zijn weergegeven in Tabel 4.5 voor alle vestigingen en alleen voor vestigingen met voldoende beschikbare data. Ter vergelijking zijn ook de leerlinggemiddelden gepresenteerd.

Tabel 4.5 Prospectief onderbouwrendement op leerling- en vestigingsniveau voor alle vestigingen en vestigingen met voldoende beschikbare data

	N	Gemiddelde	SD	min	max
<i>Alle vestigingen:</i>					
Leerlingniveau	127.621	96,2	35,6	-100	312,5
Vestigingsniveau	1132	96,6	11,9	36,3	212,5
<i>Vestigingen met voldoende data:</i>					
Leerlingniveau	110.719	96,4	35,2	-100	300
Vestigingsniveau	854	96,8	9,1	53,4	145,3

Als het prospectieve onderbouwrendement op vestigingen met voldoende data wordt vergeleken met de oorspronkelijke VO-1 vestigingen (*alle vestigingen*; zie ook laatste regel Tabel 4.4) is het gemiddelde onderbouwrendement op leerlingniveau vrijwel gelijk gebleven (96,4 vs. 96,2). Hetzelfde geldt voor de standaard deviatie op leerlingniveau (35,2 vs. 35,6). Ook het vestigingsgemiddelde van het prospectieve onderbouwrendement verandert nauwelijks (96,8 vs. 96,6) als alleen vestigingen met voldoende data worden meegenomen. De

spreiding en de range tussen minimum en maximum op vestigingsniveau is wel kleiner geworden. Dit past bij het verwijderen van kleine en daardoor vaak extreme vestigingen uit de berekeningen. Het verwijderen heeft niet geleid tot relevante verandering van het gemiddelde prospectieve onderbouwrendement op leerling- en op vestigingsniveau.

In Tabel 4.6 wordt een overzicht gegeven van het gemiddelde prospectieve onderbouwrendement en spreiding per vergelijkingsgroep. Er zijn 14 vestigingen die niet aan een vergelijkingsgroep konden worden toegewezen (Vergelijklingsgroep ‘Onbekend’). Dit komt doordat de vergelijkingsgroepsindeling gebaseerd is op de meest recente vestigingsinformatie (Bron bestand VO 2012-2013), conform de huidige werkwijze van de Inspectie. Deze informatie is gekoppeld aan de VO-1 leerlingen op basis van het brin en vestigingsnummer van schooljaar 2011-2012. Voor 14 vestigingen lukte de koppeling niet; ze komen wel voor in 2012-2013, maar niet in 2011-2012.

Tabel 4.6 Gemiddelde prospectieve onderbouwrendement per vergelijkingsgroep voor 854 vestigingen met voldoende data

Vergelijklingsgroep	N	Gemiddelde	SD
Vwo	38	98,7	3,5
Havo-vwo	91	100,3	5,1
Vmbo-gl/tl	41	93,5	9,1
Vmbo-gl/tl-havo-vwo	208	95,1	7,4
Vmbo-bl/kl	89	99,8	10,1
Vmbo-breed/havo	195	97,0	11,9
Vmbo-breed-havo-vwo	178	95,4	8,0
Onbekend	14	98,5	12,9

Uit Tabel 4.6 volgt dat het aantal vestigingen per vergelijkingsgroep vrij sterk varieert; van 38 voor categorale vwo-vestigingen tot 208 voor vestigingen die de schoolsoorten vmbo-gl/tl tot en met vwo in VO-1 aanbieden. Dit heeft tot gevolg dat de betrouwbaarheid van de gemiddelde waarde in een groep wisselend is. Daardoor is de normering van het onderbouwrendement van een VO-vestiging afhankelijk van de omvang van de vergelijkingsgroep waartoe een vestiging behoort.

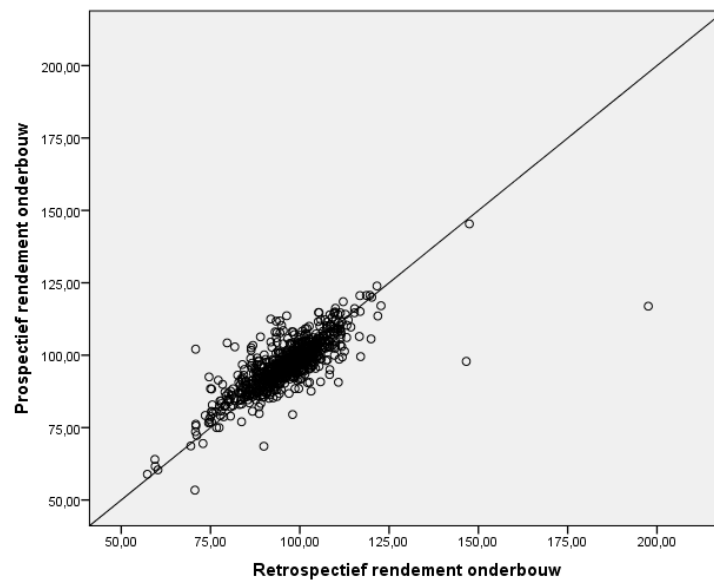
4.3.3 Retrospectief versus prospectief onderbouwrendement met individuele data

Het vergelijken van de retrospectieve met de prospectieve berekeningswijze van het gemiddelde onderbouwrendement van een VO-vestiging kan alleen gedaan worden voor VO-vestigingen waar voldoende data beschikbaar is voor beide type berekeningen. Dit is het geval voor 744 VO-vestigingen. De overeenkomst tussen beide type berekeningswijzen is redelijk

hoog. De Pearson's correlatie coëfficiënt is 0,79. Het gemiddelde verschil is -0,2 (SD = 6,5). Dit betekent dat het retrospectieve onderbouwrendement, dat gebaseerd is op vrijwel alleen niet-doublerende 3^e klassers, gemiddelde genomen iets lager is dan het prospectieve onderbouwrendement, dat gebaseerd is op doublerende en niet doublerende 1^e klassers. Zou van doublerende 3^e klassers ook het advies bekend zijn – hetgeen nu nog een kwestie van tijd is - dan zal naar verwachting het retrospectieve onderbouwrendement nog wat lager uitvallen. Hierdoor zal verschil tussen beide typen onderbouwrendement groter worden.

Figuur 4.2 toont de scatterplot van de twee berekeningswijzen, samen met de lijn $y = x$. Dit is de lijn waarop de punt van een VO-vestiging zou moeten liggen als de prospectieve en retrospectieve berekeningswijze tot exact dezelfde waarde voor het gemiddelde onderbouwrendement zou leiden.

Figuur 4.2 Retrospectief versus prospectief onderbouwrendement voor vestigingen met voldoende data



Er zijn rechtsonder twee uitbijtende vestigingen te zien. Het blijkt hier te gaan om vestigingen waarbij het retrospectieve onderbouwrendement op relatief weinig leerlingen is gebaseerd⁷. Instabiliteit lijkt hier wederom een rol van betekenis te spelen.

⁷ Er zijn twee uitbijtende vestigingen te zien. Het retrospectieve rendement is er aanzienlijk hoger dan het prospectieve rendement. Het gaat om twee vmbo-breed/Havo vestigingen. Het prospectieve onderbouwrendement is 97,8 en 116,9 en is gebaseerd op respectievelijk 82 en 51 leerlingen. Het retrospectieve rendement op deze twee scholen is daarentegen opvallend hoog (respectievelijk 146,6 en 197,6) en gebaseerd op een aanzienlijk kleiner aantal leerlingen (resp. $n = 25$ en $n = 14$).

Onderzocht is wat de algemene kenmerken zijn van de 10% VO-vestigingen met de grootste afstand tot de diagonaal, zowel aan de bovenzijde (p10; prospectief hoger dan retrospectief) als onderzijde (p90; retrospectief hoger dan prospectief). Tabel 4.7 toont hiervan de resultaten. De vergelijkingsgroep ‘Onbekend’ is niet weergegeven.

Tabel 4.7 Vergelijking van schoolkenmerken voor VO-vestigingen met het grootste verschil tussen retrospectief en prospectief onderbouwrendement. Vestigingsgemiddelde (SD) zijn weergegeven tenzij anders aangegeven

	Prospectief hoger dan retrospectief		Retrospectief hoger dan prospectief
	(boven diagonaal) N = 56		Random diagonaal N = 594
Verskil retrospectieve en prospectieve onderbouw rendement	-11,1 (4,7)	-0,3 (2,9)	10,1 (9,9)
Vergelijkingsgroep:			
Vwo	0%	6%	0%
Havo-vwo	0%	12%	22%
Vmbo-gl/tl	3%	5%	7%
Vmbo-gl/tl-havo-vwo	6%	27%	31%
Vmbo-bl/kl	37%	8%	1%
Vmbo-breed-havo	49%	22%	8%
Vmbo-breed-havo-vwo	6%	21%	31%
Totaal	100%	100%	100%
Gemiddelde Cito-eindtoetscore	525,4 (4,6)	534,5 (7,4)	536,9 (5,6)
Advies PO (<i>adviescijfer</i>)	6,2 (0,6)	7,4 (0,9)	7,7 (0,7)
Schoolsoort jaar 3 (<i>schoolsoortcijfer</i>)	6,1 (0,6)	7,3 (1,0)	7,6 (0,8)
Doublures in leerjaar 1 en 2 (%)	0,1 (0,5)	2,0 (3,1)	4,0 (5,6)
Westers en Niet Westers allochtone leerlingen (%)	31,2 (28,1)	20,6 (19,3)	24,3 (25,2)
lwoo-leerlingen (%)	32,2 (23,7)	12,1 (18,7)	6,6 (14,6)

De groep VO-vestigingen die boven de diagonaal gelegen is, heeft een retrospectieve onderbouwrendement die aanzienlijk lager uitvalt dan het prospectieve (-11,1). Verder kenmerken deze groep VO-vestigingen zich doordat het meestal om vmbo-bl/kl en vmbo-breed-havo-vestigingen gaat (samen 86%). De gemiddelde waarde die is gevonden voor de Cito-eindtoetscores, adviezen en onderwijspositie na 3 jaar passen goed bij vmbo-bl/kl en vmbo-breed-havo-vestigingen. Opvallend is daarentegen dat op deze VO-vestigingen de

leerlingen niet of nauwelijks blijven zitten (gemiddeld 0,1%), veel leerlingen lwoo hebben (32%) en het percentage allochtone leerlingen relatief hoog is (gemiddelde 31,2%).

De groep VO-vestigingen die beneden de diagonaal ligt heeft een retrospectieve onderbouwrendement dat aanzienlijk hoger uitvalt dan het prospectieve (10,1). Deze groep bevat vooral vmbo-gl/tl-havo-vwo, vmbo-breed-havo-vwo en havo/vwo-vestigingen gaat (samen 832%). Het gaat hier dus om de VO-vestigingen die de hogere schoolsoorten aanbieden, maar geen categorale vwo-vestigingen zijn. De gemiddelde waarde voor de Cito-eindtoetsscores, adviezen en onderwijspositie na 3 jaar is passend voor VO-vestigingen met de hogere schoolsoorten. Wat wel opmerkelijk is, is dat leerlingen er relatief vaak blijven zitten (gemiddeld 4%) en weinig leerlingen lwoo hebben (6,6%). Het percentage allochtone leerlingen is nu minder opvallend; gemiddelde 24,3%, hetgeen enigszins hoger is dan in de groep scholen die rondom de diagonaal gelegen is.

Bij het normeren van het onderbouwrendement wordt een VO-vestiging vergeleken met andere VO-vestigingen die soortgelijke schoolsoorten aanbieden. Tabel 4.8 toont de resultaten van de vergelijking tussen het retrospectieve en prospectieve onderbouwrendement per vergelijkingsgroep.

Tabel 4.8 Verschil tussen retrospectief en prospectief onderbouwrendement per vergelijkingsgroep

Vergelijkingsgroep	N	Retrospectief - prospectief					
		Retrospectief	Prospectief	Gemiddeld	SD	Min	Max
Vwo	31	101,0	99,0	2,0	1,6	-1,1	5,3
Havo-vwo	86	103,3	100,3	3,0	3,1	-3,3	17,5
Vmbo-gl/tl	38	93,7	93,8	0,1	4,2	-8,7	8,8
Vmbo-gl/ tl-havo-vwo	189	97,2	95,3	1,9	3,6	-9,7	21,4
Vmbo-bl/kl	73	93,4	99,6	-6,0	6,7	-31,5	8,8
Vmbo-breed-havo	170*	93,6	96,7	-3,1	9,1	-18,3	80,7
Vmbo-breed-havo-vwo	153	96,2	95,4	0,8	5,0	-13,0	20,5
Totaal	740	96,5	96,7	-0,2	6,5	-31,3	80,7

Noot *) inclusief de twee uitbijtende scholen.

Opmerkelijk is het relatief grote negatieve verschil tussen retrospectief en prospectief onderbouwrendement in de vmbo-bl/kl (-6,0) en vmbo-breed-havo (-3,1) vergelijkingsgroep. Het retrospectief rendement is voor deze vergelijksgroepen gemiddeld respectievelijk 93,4 en 93,6, het prospectieve rendement gemiddeld 99,6 en 96,7. Deze resultaten zijn in overeenstemming met die in Tabel 4.7; de groep vestigingen die boven de diagonaal liggen en

een extreem negatieve verschilwaarden hebben, zijn vooral vmbo-bl/kl en vmbo-breed-havo-vestigingen. Als doublerende leerlingen beter meegenomen hadden kunnen worden bij het retrospectieve onderbouwrendement berekeningen, dan was het verschil voor vmbo-breed-havo nog groter geweest.

Het is niet geheel duidelijk hoe het grote verschil tussen retrospectieve en prospectieve onderbouwrendement op het vmbo-bl/kl (-6,0) te verklaren is. Er zijn systematische verschillen gevonden in aantallen leerlingen per vestiging voor vestigingen waar de verschillen het grootst zijn. Waarschijnlijk is de oorzaak gelegen in het algemene gegeven dat er meer leerlingen afstromen naar een lagere schoolsoort dan er leerlingen opstromen naar een hogere schoolsoort. Bij de berekening van het prospectieve onderbouwrendement in het vmbo-bl/kl worden instromende leerlingen niet meegenomen, uitstromende leerlingen wel. De uitstromende leerlingen zijn waarschijnlijk de betere leerlingen – met mogelijk te lage adviezen of die te laag geplaatst zijn – die naar vmbo-gl/tl gaan voordat ze in leerjaar 3 zitten. De resterende leerlingen die na drie jaar nog steeds op het vmbo-bl/kl zitten zijn leerlingen met relatief lage adviezen, waaronder praktijkonderwijs, die niet zijn blijven zitten (dit blijkt uit de data). Leerlingen met een praktijkonderwijs advies in het vmbo-bl/kl beïnvloeden het prospectieve onderbouwrendement positief; ze kunnen hoger zitten dan hun advies. Bij de berekening van het retrospectieve onderbouwrendement in het vmbo-bl/kl worden instromende leerlingen wel meegenomen en uitstromende leerlingen niet. De instromers zijn vooral afkomstig van het vmbo-tl/gl, en minder van het praktijkonderwijs. Afstromende vmbo-tl-leerlingen kunnen zijn blijven zitten voordat ze afstromen. De relatief hoge adviezen en het aantal doublures van deze groep afstromende leerlingen beïnvloeden het onderbouwrendement negatief. Prospectief kan alleen een positieve beïnvloeding van het onderbouwrendement een rol spelen⁸, retrospectief zal dit vooral een negatieve beïnvloeding zijn. De resultante van deze twee vormen van beïnvloeding hangt direct samen met de omvang van de instroom en uitstroom op het vmbo-bl/kl. In het algemeen is er vaker sprake van afstromende leerlingen (hier instroom vmbo-tl) dan van opstromende leerlingen (hier uitstroom van de beste vmbo-bl/kl-leerlingen en instroom praktijkonderwijs). Samen maakt dit dat er een systematische verschil is tussen het retrospectieve en prospectieve onderbouwrendement op het vmbo-bl/kl.

4.3.4 *Invloed PO-school op prospectief onderbouwrendement*

De invloed van de PO-school op het onderbouwrendement is onderzocht aan de hand van twee verschillende meerniveau modellen (*onderzoeksvraag 2*). In Tabel 4.9 worden de resultaten van deze modellen gepresenteerd. De afhankelijke variabele in deze modellen is het prospectieve onderbouwrendement van leerlingen. Het gaat hier om VO-leerlingen die twee

⁸ Afstroom naar het praktijkonderwijs kan niet worden meegenomen omdat het praktijkonderwijs niet meegenomen wordt in de Bron VO bestanden.

schooljaren geleden nog in groep 8 van de basisschool zaten, die een Cito-eindtoets hebben gemaakt en die nu op VO-vestiging zitten waarvan voldoende data beschikbaar is om het prospectieve onderbouwrendement te kunnen berekenen (zie Tabel 4.4).

Tabel 4.9 Resultaten van de meerniveau analyses voor het bepalen van de invloed van de PO-school op het prospectieve onderbouwrendement

	Leeg meerniveau model		Kruisklassificatie model	
	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed Part:</i>				
Constante	96,23	(0,31)	95,10	(0,33)
Gemiddelde Cito-eindtoetscore (gecentreerd*)			-0,65	(0,02)
Gemiddelde mate van overadvisering (gecentreerd*)			-24,17	(0,90)
<i>Random Part:</i>				
Tussen VO-vestiging verschillen	54,77	(3,77)	70,61	(4,81)
Tussen leerlingen verschillen	1139,73	(6,02)	1066,14	(5,93)
Tussen PO-vestiging verschillen			40,53	(2,69)
<i>Model fit (DIC):</i>				
Aantal VO-vestigingen	689990,7		686615,9	
Aantal leerlingen	849		849	
Aantal PO-vestigingen	69801		69801	
			3428	

*) gecentreerd rond algemeen gemiddelde

De constante in het lege meerniveau model van Tabel 4.9 geeft het algemeen gemiddelde van het prospectieve onderbouwrendement weer; dat is 96,2. Dit komt goed overeen met de gemiddelde waarde die is gevonden basis van beschrijvende statistieken (Tabel 4.5). Uit de significante variantie term op VO-vestigingsniveau (54,8) volgt dat VO-vestigingen van elkaar verschillen in onderbouwrendement. Verder geeft dit model inzicht in de welke deel van de totale variantie in prospectief onderbouwrendement is gelegen op het niveau van VO-vestigingen en welk deel op het niveau van leerlingen. Dat is respectievelijk 4,6%⁹ en 95,4%.

De invloed van de toeleverende PO-vestigingen is geanalyseerd door de gemiddelde score op de Cito-eindtoets van de leerlingen en daarnaast ook nog de gemiddelde mate van overadvisering van de PO-vestiging als covariaat in een kruisklassificatie model mee te nemen. De fit van het model – blijkend uit de daling in DIC (Spiegelhalter, Best, Carlin, et al., 2002) – is beduidend beter dan een leeg kruisklassificatie model¹⁰. Rekening houden met de

⁹ = 54,8 / (54,8 + 1139,7) * 100

¹⁰ Er is ook een leeg kruisklassificatie model geschat met dezelfde leerlingen en vestigingen. In vergelijking met het lege meerniveau model verbetert de fit aanzienlijk (DIC=688141,1) De variantie die aan PO-scholen wordt toegeschreven is in dit model 62,11 (SE 3,24). In vergelijking met het lege meerniveau model zijn de verschillen tussen VO-scholen 10% kleiner geworden (van 54,8 naar 49,6) en de verschillen tussen leerlingen 5% kleiner

invloed van de PO-scholen lijkt van belang te zijn gezien de significante variantie die aan PO-vestigingen kan worden toegeschreven (40,5). Uit het kruisklassificatie model in Tabel 4.9 is af te leiden dat leerlingen afkomstig van PO-vestigingen met een hoge gemiddelde Cito-eindtoetsscore, in het algemeen een lager onderbouwrendement behalen (-0,65). Leerlingen afkomstig van PO-vestigingen waar relatief vaak te hoge adviezen worden afgegeven hebben gemiddeld genomen een lager onderbouwrendement dan leerlingen die niet van dergelijke PO-vestigingen afkomstig zijn (-24,17). Uit de kruisklassificatie modellen is gebleken dat er VO-vestigingen zijn aan te wijzen die ‘last’ hebben van toeleverende PO-vestigingen die overadviseren. Aan de andere kant zijn er ook VO-vestigingen die (onterecht) profiteren van PO-vestigingen die onderadviseren. Door te lage adviezen van toeleverende PO-vestigingen weten zij een onderbouwrendement te behalen die te hoog is. De vertekening werkt dus twee kanten op.

De volgende stap in de analyses is na te gaan in welke mate het onderbouwrendement van VO-vestigingen vertekend is als de invloed van PO-vestigingen niet wordt meegenomen. Om hier meer inzicht in te krijgen zijn de schoolresiduen van de twee modellen onderling vergeleken. De beschrijvende statistieken van de schoolresiduen zijn weergegeven in Tabel 4.10.

Tabel 4.10 Beschrijvende statistiek van de schoolresiduen van VO-vestigingen verkregen bij de meerniveau modellen

	Aantal	Min.	Max.	Gemiddelde	Standaard deviatie
Schoolresidu leeg meerniveau model	849	-23,33	22,34	,01	6,24
Schoolresidu kruisklassificatie model	849	-28,69	20,20	-,00	7,23

Als rekening wordt gehouden met de geneste structuur van de data, dan blijkt dat de verschillen in gemiddeld prospectief onderbouwrendement tussen VO-vestigingen aanzienlijk kleiner worden dan wanneer dit niet wordt gedaan. Op basis van eenvoudige beschrijvende statistieken (Tabel 4.5) is een spreiding gevonden van 9,1 met een range tussen minimale en maximale waarde van 92. Op grond van het lege meerniveau model is deze spreiding 6,2 en de range 58 (Tabel 4.10). VO-vestigingen verschillen minder van elkaar in onderbouwrendement dan het in eerste instantie leek. Dit pleit ervoor dat het toepassen van

(van 1139,7 naar 1086,6). Van de totale variantie is nu 4,1% gelegen op het niveau van VO-vestigingen, 90,7% op het niveau van leerlingen en 5,1% op het niveau van PO-vestigingen.

een meerniveau analyses op de data essentieel is om de verschillen tussen VO-vestigingen in onderbouwrendement goed vast te stellen.

In vergelijking met het lege meerniveau model is bij het kruisklassificatie model de spreiding van de schoolresiduen toegenomen (Tabel 4.10: respectievelijk, 6,2 en 7,2). Als rekening wordt gehouden met de invloed van de toeleverende PO-scholen gaan VO-vestigingen weer meer van elkaar verschillen in onderbouwrendement. Uit een vergelijking met een leeg kruisklassificatie model blijkt dat dit vooral door de corrigerende werking van de covariaten te verklaren is, en niet door het gebruik van een kruisklassificatie model op zich. Door correctie voor gemiddeld Cito-eindtoetscores en mate van overadvisering worden verschillen tussen VO-vestigingen minder gemaskeerd. In overeenstemming hiermee is de samenhang tussen de schoolresiduen van beide modellen. De correlatie is tussen de schoolresiduen van beide modellen is $r=0,86$ ($n=849$, $p<0,001$), hetgeen lager is dan de correlatie tussen de schoolresiduen van het lege meerniveau en een leeg kruisklassificatie model ($r=0,98$).

Voor een nadere typering van VO-vestigingen waar het rekening houden met de invloed van de toeleverende PO-vestigingen heeft geleid tot relatief grote verschuiving in de waarde van het prospectieve onderbouwrendement, zijn de VO-vestigingen gecategoriseerd (zie Tabel 4.11). Hierbij zijn de p10 en p90 van de verdeling van de schoolresiduen als uitgangspunt genomen. Bij het lege meerniveau model hebben de p10 en de p90 de waarden -8,0 en +7,5, en bij het kruisklassificatie model de waarden -9,5 en +9,2.

Tabel 4.11 Indeling van VO-vestigingen ten gevolge van vertekening van het prospectieve onderbouwrendement als geen rekening wordt gehouden met toeleverende PO-vestigingen

		Schoolresidu kruisklassificatie model		
		Onder gemiddeld	Gemiddeld	Boven gemiddeld
Schoolresidu leeg meerniveau model	Onder gemiddeld	<i>Neutraal</i> $N = 63$	<i>Te negatief</i> $N = 22$	<i>Veel te negatief</i> $N = 0$
	Gemiddeld	<i>Te positief</i> $N = 21$	<i>Neutraal</i> $N = 623$	<i>Te negatief</i> $N = 34$
	Boven gemiddeld	<i>Veel te positief</i> $N = 0$	<i>Te positief</i> $N = 38$	<i>Neutraal</i> $N = 48$

Voor VO-vestigingen die op de diagonaal liggen heeft het rekening houden met de invloed van toeleverende PO-vestigingen geen grote consequenties ('*neutraal*'). Anders is het voor VO-vestigingen waarbij het niet-verdisconteren van de invloed van PO-vestigingen leidt tot over- of onderschatting van het prospectieve onderbouwrendement. In Tabel 4.11 zijn dit de VO-vestigingen waarvan het schoolresidu van het lege meerniveau model respectievelijk

‘(veel)te positief’ (links onder, N = 59) of ‘(veel) te negatief’ is (rechts boven, N = 56). De algemene schoolkenmerken van deze drie groepen VO-vestigingen zijn vergeleken. Tabel 4.12 toont hiervan de resultaten voor 808 VO-vestigingen met volledige data op de te vergelijken kenmerken.

Tabel 4.12 Vergelijking van schoolkenmerken voor VO-vestigingen voor vertekening van het prospectieve onderbouwrendement als geen rekening wordt gehouden met de invloed van PO-vestigingen

	Schoolresidu van leeg meerniveau model te negatief N = 54	Neutraal N = 684	Schoolresidu van leeg meerniveau model is te positief N = 58
Gemiddelde verschil schoolresiduen leeg meerniveau model – kruisklassificatie model	-4,2 (2,5)***	0,0 (3,4)	4,2 (2,4)***
Vergelijkingsgroep:			
Vwo	15%***	4%	0%***
Havo-vwo	26%	11%	2%
Vmbo-gl/ tl	4%	5%	5%
Vmbo-gl/ tl-havo-vwo	37%	27%	3%
Vmbo-bl/kl	0%	10%	24%
Vmbo-breed-havo	2%	21%	55%
Vmbo-breed-havo-vwo	17%	23%	10%
Totaal	100%	100%	100%
Advies PO (<i>adviescijfer</i>)	4,2 (0,6)***	3,3 (0,9)	2,2 (0,6)***
Schoolsoort jaar 3 (<i>schoolsoortcijfer</i>)	4,1 (0,7)***	3,2 (1,0)	2,2 (0,7)***
Doublures in leerjaar 1 en 2 (%)	3,4 (3,6)*	2,3 (3,9)	0,2 (0,9)***
Gemiddelde Cito-eindtoetscore	540,2 (5,0)***	533,9 (7,5)	525,9 (5,1)***
Westers en Niet Westers allochtone leerlingen (%)	26,0 (19,0)	23,2 (22,7)	26,3 (27,7)
lwoo-leerlingen (%)	1,6 (6,1)***	13,1 (19,1)	36,0 (23,8)***

Student t-tests t.o.v. de neutrale groep: * p < 0,05; ** p < 0,01; *** p < 0,001, m.u.v. de vergelijkingsgroep (Pearson Chi² test)

Uit Tabel 4.12 volgt dat VO-vestigingen die te kort wordt gedaan als de invloed van PO-vestigingen niet wordt meegenomen, vooral vestigingen met een vwo-afdeling betreffen (samen 95%) (Pearson Chi² = 38,5, df=6, p<0,001). In overeenstemming hiermee is gevonden dat deze VO-vestigingen zich kenmerken door relatief hoge gemiddelde Cito-eindtoetscores, adviezen en schoolsoort drie jaar na start in het VO en door een laag percentage lwoo-leerlingen in vergelijking tot de groep neutrale VO-vestigingen (alle t-tests p<0,001).

Opvallend is wel dat er vrij veel doublures voorkomen: 3,4% vs. 2,3% ($t=2,065$, $df=747$, $p=0,04$). Het percentage allochtone leerlingen is niet significant verschillend tussen de twee groepen.

VO-vestigingen die profiteren van de invloed van toeleverende PO-vestigingen zijn vestigingen met een vmbo-afdeling. Het gaat met name om vmbo-bl/kl en vmbo-breed-havo-vestigingen (samen 79%) (Pearson $\chi^2=58,2$, $df=6$, $p<0,001$). Logischerwijze hebben de VO-vestigingen in deze groep relatief lage gemiddelde Cito-eindtoetsscores, adviezen en schoolsoortcijfer na drie jaar VO in vergelijking tot de groep neutrale VO-vestigingen (alle t-tests $p<0,001$). Opvallend is echter dat er niet of nauwelijks doublures voorkomen (0,2%) en er veel meer lwoo-leerlingen zijn (36%). Het percentage allochtone leerlingen is niet verschillend tussen beide groepen.

4.3.5 Bodem- en plafondefecten in relatie tot vertekening van het onderbouwrendement

Onduidelijk is waarom juist op vwo en lage vmbo-afdelingen het niet meenemen van de invloed van PO-vestigingen leidt tot relatief grote vertekening van het prospectieve onderbouwrendement. Mogelijk kan dit (groten)deels verklaard worden door het optreden van bodem- en plafondefecten die inherent zijn aan de bepaling van het onderbouwrendement van een leerling. Plafondefecten spelen met name op categorale vwo-vestigingen. Vrijwel alle leerlingen komen hier binnen met het hoogste advies (vwo) en kunnen daarna alleen maar afstromen of doubleren. Daarom zal het onderbouwrendement van categorale vwo-vestigingen in de regel lager zijn dan 100. Bodemeffecten spelen met name op categorale vmbo-bl/kl-vestigingen een rol. Leerlingen op categorale vmbo-bl/kl-vestigingen hebben vrijwel allemaal een laag advies (vmbo-bl/kl of praktijkonderwijs), en kunnen alleen maar opstromen. Als doubleren zelden voorkomt, dan zal het gemiddelde onderbouwrendement van categorale vmbo-bl/kl vestigingen meestal hoger zijn dan 100.

In hoeverre genoemde bodem- en plafondefecten van invloed zijn op de mate van vertekening van het onderbouwrendement is nader onderzocht door in te zoomen op categorale vwo ($n=37$) en categorale vmbo-bl/kl ($n=79$) vestigingen. Binnen deze twee vergelijkingsgroepen is de werking van bodem- en plafondefecten het sterkst.

Tabel 4.13 geeft een overzicht van de resultaten. Bij slechts 22% van alle categorale vwo-vestigingen is gevonden dat het niet meenemen van de invloed van PO-vestigingen leidt tot onderschatting van het onderbouwrendement. Ze behoren tot de groep met een ‘te negatief’ onderbouwrendement (Tabel 4.11). Het gemiddelde verschil tussen de schoolresiduen van de twee modellen is enigszins lager (-6,8) dan dat van de categorale vwo-vestigingen die tot de neutrale groep behoren (-5,1). Kijkend naar de leerlingkenmerken die samen het onderbouwrendement bepalen (advies, doublures en schoolsoort na 3 jaar) dan zijn er geen significante verschillen gevonden tussen de twee groepen. Ook de gemiddelde Cito-eindtoetsscore of etnische herkomst van de leerlingen verschilt niet tussen de groepen.

Tabel 4.13 Vergelijking van schoolkenmerken voor categorale vwo en vmbo-bl/kl-vestigingen voor vertekening van het prospectieve onderbouwrendement als geen rekening wordt gehouden met de invloed van PO-vestigingen

	Schoolresidu van lege meerniveau model te negatief	Neutraal	Schoolresidu van lege meerniveau model is te positief
Vergelijkingsgroep vwo	N = 8	N = 29	N = 0
Gemiddelde verschil schoolresiduen leeg meerniveau model – kruisklassificatie model	-6,8 (1,0)	-5,1 (1,4)**	
Advies PO (<i>adviescijfer</i>)	4,9 (0,1)	4,9 (0,1)	
Schoolsoort jaar 3 (<i>schoolsoortcijfer</i>)	5,0 (0,01)	5,0 (0,03)	
Doublures in leerjaar 1 en 2 (%)	4,9 (4,0)	7,8 (5,1)	
Gemiddelde Cito-eindtoetsscore	547,2 (0,7)	547,0 (0,8)	
Westers en Niet Westers allochtone leerlingen (%)	22,9 (9,7)	17,7 (9,8)	
Vergelijkingsgroep vmbo-bl/kl	N = 0	N = 65	N = 14
Gemiddelde verschil schoolresiduen basaal meerniveau model – kruisklassificatiemodel		3,9 (1,9)	4,7 (1,8)
Advies PO (<i>adviescijfer</i>)		2,0 (0,3)	1,8 (0,6)*
Schoolsoort jaar 3 (<i>schoolsoortcijfer</i>)		1,8 (0,3)	1,7 (0,3)
Doublures in leerjaar 1 en 2 (%)		0,0 (0,0)	0,0 (0,0)
Gemiddelde Cito-eindtoetsscore		523,1 (3,2)	522,5 (3,7)
Westers en Niet Westers allochtone leerlingen (%)		36,4 (27,3)	41,3 (36,9)
Iwoo-leerlingen (%)		37,9 (20,9)	46,5 (22,1)

* $p < 0,05$; ** $p < 0,01$

Binnen de groep categorale vmbo-bl/kl-vestigingen is een klein aantal vestigingen aan te wijzen waar het niet verdisconteren van de invloed van PO-vestigingen leidt tot een overschatting van het onderbouwrendement (Tabel 4.13: 18%). Het gemiddelde verschil tussen de schoolresiduen in de groep met overschatting van het onderbouwrendement is 4,7, tegen 3,9 voor de ‘neutrale’ categorale vmbo-bl/kl-vestigingen. Het iets lagere gemiddelde advies (1,8 vs. 2,0, $t=-2,182$, $df=77$, $p=0,03$) zal een enigszins hogere onderbouwrendement tot gevolg hebben. De overige kenmerken zijn (gemiddelde Cito-eindtoetsscore, het percentage allochtone en lwoo leerlingen) zijn niet significant verschillend.

Op categorale vmbo-bl/kl-vestigingen lijken meer aanwijzingen te vinden zijn voor de beïnvloeding door een bodemeffect (lagere adviezen) dan op categorale vwo vestigingen voor het beïnvloeding door een plafondeffect. Hierbij moet echter bedacht worden dat het vmbo-bl/kl niet het laagste schoolsoort is en ook niet het laagste advies is dat leerlingen kunnen

krijgen. Dat is namelijk het praktijkonderwijs, maar hierover is geen informatie beschikbaar in de bron VO bestanden. Het advies praktijkonderwijs is wel meegenomen in de berekening van het onderbouwrendement. Dit verklaart waarom het gemiddelde advies lager is dan 2,0.

Op basis van deze bevindingen is af te leiden dat het over- of onderschatten van het onderbouwrendement bij vestigingen met een vwo- of lage vmbo-afdeling hooguit voor een klein deel te verklaren is door plafond- of bodemeffecten. Door middel van stratificatie naar schoolsoort wordt voor de invloed van deze effecten gecorrigeerd, maar dan blijkt dat er nog steeds een (klein) verschil tussen de schoolresiduen van de twee meerniveau modellen bestaan. Bovendien is gevonden dat ongeveer 80% van de categorale vwo- en vmbo-bl/kl vestigingen tot de groep ‘neutrale’ vestigingen behoort. Op deze scholen leidt het wel of niet meenemen van de invloed van PO-scholen niet tot noemenswaardige veranderingen in het gemiddelde onderbouwrendement. Bodem- en plafondeffecten spelen wel een rol bij de bepaling van het onderbouwrendement op categorale vwo- en vmbo-bl/kl vestigingen, maar dit werkt niet door in de invloed die PO-scholen kunnen hebben op de waarde van het onderbouwrendement van een VO-vestiging.

4.3.6 *Doubleren van lwoo-leerlingen*

Bij de bepaling van het onderbouwrendement worden VO-vestigingen gecompenseerd voor lwoo-leerlingen. Een op de vier lwoo-leerlingen mag een jaar doubleren zonder dat dit negatieve gevolgen heeft voor het onderbouwrendement van de betreffende vestiging. Onderzocht is hoe vaak lwoo-leerlingen doubleren in vergelijking met niet-lwoo leerlingen (*eerste aanvullende onderzoeksvraag*).

Tabel 4.14 toont voor alle schoolsoorten het aantal keer dat leerlingen blijven zitten in de eerste twee jaar van het VO. Daarbij is een uitsplitsing gemaakt naar leerlingen die wel en geen lwoo hebben gehad in het eerste leerjaar VO.

Tabel 4.14. Percentage doublures in eerste twee jaar van het VO voor leerlingen die in VO-1 wel en geen lwoo hebben gehad

Leerwegondersteuning	Aantal doublures in VO-1 en VO-2		
	0	1	2
Ja	20.407	1304	17
	93,9%	6,0%	0,1%
Nee	148.817	8446	42
	94,6%	5,4%	0,0%
Totaal	169.224	9750	59
	94,5%	5,4%	0,0%

Er is nauwelijks verschil tussen de doublure-frequentie van lwoo-leerlingen (6,1%) en niet lwoo-leerlingen (5,4%). Mogelijk zijn er per schoolsoort wel duidelijker verschillen aantoonbaar. Daarom is ook een uitsplitsing naar schoolsoort in leerjaar 3 gemaakt. De resultaten hiervan staan in Tabel 4.15.

Op de twee laagste schoolsoorten waar lwoo gegeven kan worden - vmbo-bl en vmbo-kl – zijn wel veel lwoo-leerlingen (resp. 64% en 26%), maar de leerlingen op deze schoolsoorten blijven niet zitten (Tabel 4.15). Lwoo-leerlingen op het vmbo-tl/gl - een relatief kleine groep (5%) - blijven in het algemeen juist iets minder vaak zitten dan niet-lwoo-leerlingen op deze schoolsoort ($p < 0.001$). Verder blijkt dat op de schoolsoorten die niet mee genomen (kunnen) worden in de berekeningen van het onderbouwrendement ('Brugjaar' en 'Onbekend') ook lwoo-leerlingen te vinden zijn. Met name op het schoolsoort 'Onbekend' komen vrij veel lwoo-leerlingen voor, maar ze blijven even vaak zitten als niet-lwoo-leerlingen in de eerste twee jaar VO (69%).

Tabel 4.15. Percentage doublures in eerste twee jaar van het VO voor leerlingen die in VO-1 wel en geen lwoo hebben gehad, per schoolsoort

Leerwegondersteuning	Aantal doublures in VO-1 en Vo-2		
	0	1	2
<i>Vmbo-bl:</i>			
Ja	100%	0,0%	0,0%
Nee	100%	0,0%	0,0%
<i>VMBO kl:</i>			
Ja	100%	0,0%	0,0%
Nee	100%	0,0%	0,0%
<i>Vmbo-tl gl:</i>			
Ja	98,2%	1,8%	0,0%
Nee	97,3%	2,7%	0,0%
<i>VO3-brugjaar:</i>			
Ja (n = 50)	50,0%	50,0%	0,0%
Nee (n = 7.774)	71,5%	28,4%	0,1%
<i>Onbekend:</i>			
Ja (n = 1.824)	31,1%	67,9%	0,9%
Nee (n = 7.774)	31,1%	68,3%	0,6%

5 Deel III: toegevoegde waarde van vmbo-tl naar havo

5.1 Inleiding

Het valide schatten van de toegevoegde waarde van een vestiging of school is afhankelijk van een aantal voorwaarden. In de eerste plaats dienen er longitudinale gegevens beschikbaar te zijn op het niveau van de leerlingen (Goldstein, 1997). Daarnaast moet er zowel informatie zijn met betrekking tot het beginniveau en het eindniveau van de leerlingen voor een vastgestelde periode onderwijs (Aitkin & Longford, 1986; Nash, 2003; Thrupp et al., 2002; Willms, 1992). Idealiter zijn er daarnaast nog een aantal andere achtergrondgegevens van de leerlingen bekend. Tot slot moet er gebruik gemaakt worden van statistische modellen (meerniveau) waarmee op een adequate wijze omgegaan wordt met de hiërarchische structuur van de data (Aitkin & Longford, 1986; Goldstein, 1997; Snijders & Bosker, 2012).

Gegeven de huidige beschikbaarheid van gegevens bij de Inspectie van het Onderwijs kan de toegevoegde waarde van vestigingen in het voortgezet onderwijs (VO) voor een aantal specifieke groepen leerlingen worden bepaald. Eén van deze groepen betreft de leerlingen die na het behalen van het vmbo-tl-diploma doorstromen naar havo. Voor deze groep leerlingen – hier *doorstromers* genoemd - zijn de gegevens aanwezig die voldoen aan de eerder genoemde voorwaarden. De periode waarover de toegevoegde waarde in dat geval kan worden bepaald is de tijd tussen het behalen van het diploma vmbo-tl en het voor het eerst deelnemen aan het havo centraal examen. Voor deze groep leerlingen geldt dan dat de prestaties op het centraal examen vmbo-tl gezien kunnen worden als een indicatie voor het beginniveau en de prestaties van de leerlingen op het centraal examen havo kunnen worden gezien als een indicatie voor het eindniveau.

Dit deelproject III is een verkenning naar de mogelijkheden van toegevoegde waarde bepalingen van havo-afdelingen voor de groep leerlingen die van vmbo-tl doorstromen naar havo. In dit hoofdstuk komen twee specifieke onderzoeksvragen aan de orde komen, namelijk:

1. *Wat is de toegevoegde waarde van een VO-vestiging op het gemiddelde centraal examencijfer havo voor vmbo-tl gediplomeerde leerlingen die doorgestroomd zijn naar havo?*
2. *Verschildt de toegevoegde waarde van havo-afdelingen van VO-vestigingen tussen de verplichte vakken Nederlands, Engels en wiskunde?*

Daarnaast is er in overleg met de opdrachtgever voor gekozen om ook te kijken naar de relatie tussen de toegevoegde waarde van havo-afdelingen en het rendement van deze afdelingen voor vmbo-tl gediplomeerde leerlingen die doorgestroomd zijn naar havo. De reden voor het

simultaan onderzoeken van rendementen en toegevoegde waarde op basis van examencijfers is mogelijk strategisch gedrag van scholen. Er zijn voor scholen verschillende manieren om de toegevoegde waarde artificieel te doen stijgen (Timmermans, Bosker, De Wolf, Doolgaard & Van der Werf, *in press*), zoals het oefenen voor examens, leerlingen laten uitstromen of zittenblijven, etc. Echter, rendementsindicatoren bieden scholen de tegengestelde “incentives”, namelijk om leerlingen zo snel mogelijk door het onderwijs te laten stromen en op een zo hoog mogelijk niveau. De combinatie van toegevoegde waarde en het rendement kunnen daardoor een meer volledig beeld geven van de opbrengsten van een vestiging. De kracht van opbrengstbepalingen in onderwijstoezicht met betrekking tot strategisch gedrag van scholen ligt in de juiste keuze voor een combinatie van indicatoren (Koretz, 2003). Hiervoor is een onderzoeksvraag aan het onderzoek toegevoegd, namelijk:

- *Hoe verhoudt de toegevoegde waarde van een VO-vestiging op het gemiddelde centraal examencijfer zich tot het rendement voor vmbo-tl gediplomeerde leerlingen die doorgestroomd zijn naar havo?*

In de volgende paragraaf wordt de methode van het onderzoek en de gebruikte dataset nader beschreven. Daarna volgen de resultaten van doorstroom en examenresultaten op leerlingniveau, gevolgd door de resultaten van het in kaart brengen van de toegevoegde waarde van de havo-afdelingen voor het gemiddelde centraal examencijfer alsmede voor de centraal examencijfers van de verplichte vakken. Tot slot wordt het doorstroomrendement van de havo-afdelingen in kaart gebracht en wordt onderzocht hoe alle indicatoren aan elkaar gerelateerd zijn.

5.2 Methode

5.2.1 Beschrijving van de dataset

Voor het analyseren van de toegevoegde waarde van vestigingen voor de leerlingen die vanuit vmbo-tl doorstromen naar de havo zijn verschillende gegevensbestanden gecombineerd. Een eerste databron is het bestand met examengegevens van leerlingen uit vmbo-tl. De eindexamencijfers van de leerlingen kunnen worden gebruikt als een indicatie voor het beginniveau van de leerlingen voorafgaand aan het havo-traject dat de leerlingen zullen doorlopen. In het huidige onderzoek is gebruik gemaakt van examengegevens van vmbo-tl uit het jaar 2009. Het onderzoek betreft derhalve leerlingen die in de zomer van 2009 de overstap maakten naar havo. Uit deze examenbestanden zijn de centraal examencijfers van de leerlingen gehaald voor de vakken Nederlands, Engels en wiskunde, alsmede het gemiddelde cijfer over alle vakken waarin de leerlingen een centraal examen hebben afgelegd.

Een tweede bron van gegevens die wordt gebruikt zijn de inschrijvingsgegevens in het voortgezet onderwijs. Deze databestanden bevatten twee belangrijke typen van informatie. In

de eerste plaats zijn dat achtergrondkenmerken van leerlingen, zoals de sekse, etnische achtergrond (onder andere cumi), leerwegondersteuning (lwoo) en of een leerling woonachtig is in een armoedeprobleemcumulatiegebied (apcg). Achtergrondgegevens kunnen worden gebruikt bij de schatting van de toegevoegde waarde als covariaten om tot een zo eerlijk mogelijke vergelijking van vestigingen te komen. In de tweede plaats bevatten de inschrijvingsgegevens informatie over de doorstroom van de leerling gedurende het havo-traject. Het betreft een selectie van 8.236 leerlingen die in schooljaar 2008-2009 in de examenklas van vmbo-tl zat en die doorgestroomd zijn naar havo. Op basis hiervan kan worden vastgesteld in hoeverre de leerlingen in de nominale tijd het havo-examen succesvol hebben afgerond of dat er sprake is van zittenblijven of uitval (zie ook beschrijving doorstroomrendement).

Een derde bron van gegevens betreffen de gegevens van de havo-examens. De examencijfers van de leerlingen in de havo kunnen worden gebruikt als de afhankelijke variabele in de schatting van de toegevoegde waarde van vestigingen. Ze vormen een indicatie voor het eindniveau van leerlingen. Omdat in dit onderzoek gebruikt wordt gemaakt van de groep leerlingen die in de zomer van 2009 de overstap naar havo maakten, zijn er verschillende bestanden voor de havo-examen gebruikt. Dit betreft de examengegevens van de jaren 2010, 2011 en 2012. Hierdoor zijn examengegevens beschikbaar voor leerlingen die versneld zijn doorgestroomd (examengegevens 2010), die nominaal zijn doorgestroomd (examengegevens 2011) en die er een jaar langer over hebben gedaan om het havo-examen te bereiken (examengegevens 2012). Wederom zijn uit deze examenbestanden de volgende gegevens van de leerlingen gehaald: centraal examencijfers voor de vakken Nederlands, Engels en wiskunde alsmede het gemiddelde centraal examencijfers over alle vakken waarin de leerling eindexamen heeft gedaan. Hierbij moet worden opgemerkt dat sommige leerlingen voor verschillende wiskunde vakken examencijfers hebben. Voor deze leerlingen is (in overleg met de opdrachtgever) gebruik gemaakt van het gemiddelde wiskunde cijfer. Het gaat hier om een zeer kleine groep leerlingen, namelijk 4 leerlingen voor het examen in 2010, 0 leerlingen voor het examen in 2011 en 8 leerlingen voor het examen in 2012. Tevens zijn voor sommige leerlingen de eindexamencijfers van meerdere havo-examens in opeenvolgende jaren beschikbaar. Voor deze leerlingen zullen de examencijfers gebruikt worden van het eerste jaar dat ze deelgenomen aan het havo-examen. Het betreft 276 leerlingen die zowel in 2011 (nominaal) als in 2012 eindexamen hebben gedaan.

5.2.2 *Toegevoegde waarde*

Voor het huidige onderzoek zullen vier toegevoegde waarde indicatoren geschat worden voor vestigingen op basis van leerlingen die doorstromen vanuit vmbo-tl naar havo. Dat zijn toegevoegde waarde indicatoren op basis van het gemiddelde examencijfer en tevens losse indicatoren voor de kernvakken Nederlands, Engels en wiskunde.

De toegevoegde waarde van vestigingen zullen worden geschat op basis van meerniveau regressie modellen. Voor het schatten van deze modellen is gebruik gemaakt van de software MLwiN 2.28 (Rasbash et al., 2009). In de modellen zullen twee niveaus onderscheiden worden, namelijk leerlingen (niveau 1) en VO-vestigingen (niveau 2). VO-vestigingen zijn hier geoperationaliseerd als een combinatie van het brinnummer en het vestigingsnummer. De vestiging waarop de leerling zit in het eerste jaar na de overgang naar havo zal worden gebruikt. Dit betekent dat in dit specifieke onderzoek de vestiging van schooljaar 2009-2010 wordt gebruikt.

In de meerniveau modellen zijn de cijfers van de leerlingen op het havo centraal examen de afhankelijke variabele. De cijfers van de leerlingen op het vmbo-tl centraal examen en de achtergrondkenmerken cumi, apcg en lwoo zullen in de analyses worden meegenomen als voorspellers van de havo centraal examencijfers. Voor elke afhankelijke variabele zijn drie meerniveau modellen geschat, waarvan twee varianten gezien kunnen worden als toegevoegde waarde. Het eerste model betreft een leeg model en geeft de verschillen tussen havo-afdelingen weer voor de bruto centraal examencijfers. In dit model is nog geen sprake van correctie voor kenmerken van de leerlingen bij aanvang en kan daardoor nog niet gezien worden als toegevoegde waarde of als een “eerlijke” vergelijking van de opbrengsten van havo-afdelingen. In het tweede model wordt het beginniveau van de leerlingen als covariaat toegevoegd (Type AA), waardoor een eerste simpele toegevoegde waarde model ontstaat (Timmermans, Doolaard & De Wolf, 2011). Bij het derde model worden naast het beginniveau ook verschillende achtergrondkenmerken toegevoegd. Hierdoor ontstaat een schatting van het Type A schooleffect. De achtergrondkenmerken die in het derde model worden meegenomen zijn:

Leerwegondersteuning (lwoo): Dit is een binaire variabele die aangeeft of een leerling gedurende het voortgezet onderwijs in aanmerking kwam voor leerwegondersteuning of niet. De operationalisatie is gelijk aan die in hoofdstuk 4. De leerling krijgt lwoo als er sprake is van leerachterstand op ten minste twee van de vier domeinen: inzichtelijk rekenen, begrijpend lezen, technisch lezen en spellen en een IQ dat ligt tussen de waarden 75 tot en met 90, of van een IQ dat ligt tussen de waarden 91 tot en met 120 in combinatie met sociaal-emotionele problematiek.

Etnische achtergrond (cumi): De etnische achtergrond van leerlingen is gemeten in drie categorieën, te weten autochtone leerlingen, Westers allochtone leerlingen en niet Westers allochtone leerlingen.

Armoedeprobleemcumulatiegebied (apcg): De variabele armoedeprobleemcumulatiegebied is niet zozeer een kenmerk van de leerling, maar vooral een kenmerk van de buurt/wijk waarin de leerling woont. Op basis van de vier cijferige postcode is vastgesteld in hoeverre er in een buurt sprake is van veel inwoners met een laag inkomen, veel inwoners met een uitkering

en/of veel niet Westers allochtone inwoners. Op basis van relatieve aantallen veel inwoners met een laag inkomen, een uitkering en/of niet Westers allochtone inwoners worden acht categorieën van buurten onderscheiden.

Uit de meerniveau modellen worden de residuen op vestigingsniveau gehaald. Met uitzondering van het lege model zijn deze residuen een schatting van de toegevoegde waarde van een vestiging. Het residu van een vestiging kan worden geïnterpreteerd als het verschil in behaalde examencijfers (op het havo-examen) tussen deze specifieke vestiging en de overige vestigingen in dezelfde steekproef, waarbij een correctie heeft plaatsgevonden voor aanvangs- en achtergrondverschillen. Een positief residu geeft een indicatie dat de betreffende vestiging hogere havo-examencijfers realiseert dan verwacht, terwijl een negatief residu een indicatie is dat de betreffende vestiging lagere examencijfers realiseert dan verwacht.

5.2.3 Doorstroomrendement

Naast de indicator voor toegevoegde waarde zal ook een indicator voor het rendement van de vestigingen worden onderzocht voor vmbo-tl gediplomeerde leerlingen die opgestroomd zijn naar havo. De grondslag hiervoor is dat vestigingen mogelijk de toegevoegde waarde voor deze specifieke groep leerlingen kunnen beïnvloeden door leerlingen te laten zitten waardoor ze een langere voorbereiding hebben naar het examen of door uitstroom (onderzoeksvraag 3). Het rendement van een vestiging wordt vastgesteld op basis van inschrijvings- en examengegevens en zal worden uitgedrukt in het percentage leerlingen dat nominaal het havo-examen succesvol afgelegd heeft. Het wordt als volgt berekend:

$$\text{Rendement doorstromende leerlingen} = 100 * \frac{\text{aantal nominale leerlingen}}{\text{totaal aantal leerlingen}}$$

In deze berekening wordt uitgegaan van het totale aantal leerlingen die in 2009 beginnen aan de overstap van vmbo-tl naar havo. Van de totale groep leerlingen wordt bepaald welk percentage leerlingen binnen de nominale tijd voor het havo-examen slagen. Vergelijkbaar met het bovenbouwrendement wordt ook het resultaat van het havo-examen in de bepaling van het doorstroomrendement betrokken (Inspectie van het Onderwijs, 2011).

Voor de grootste groep leerlingen die deze overstap maakten in 2009 wordt het rendement bepaald op basis van twee gegevens, namelijk 1) de overgang van havo-4 naar havo-5 in 2010/2011 en 2) het resultaat van het havo-eindexamen in 2011. Leerlingen die sneller dan nominaal het havo-examen bereiken worden hierin natuurlijk ook meegerekend.

Een leerling wordt niet gerekend als nominaal wanneer de leerling blijft zitten bij de overgang van havo-4 naar havo-5 en daardoor pas in 2012 deelneemt aan het havo-examen, wanneer de leerling uitstroomt, wanneer de leerling zakt voor het havo-examen of wanneer er geen havo-examengegevens van de leerling bekend zijn. Een kleine groep leerlingen (31) maakt na één jaar havo direct de overstap van havo naar het vwo. Deze leerlingen zullen niet worden

meegerekend in de bepaling van het doorstroomrendement evenals in de bepaling van de toegevoegde waarde van een vestiging.

5.3 Resultaten

5.3.1 Samenhang tussen beginniveau, achtergrondkenmerken en examencijfers havo

Correlaties tussen de examencijfers van de schoolsoorten vmbo-tl en havo zijn gepresenteerd in Tabel 5.1 (links boven). Tussen haakjes staat het aantal leerlingen waarop de samenhang is gebaseerd. Dit is het aantal leerlingen waarvoor beide cijfers beschikbaar zijn. In de licht grijsgekleurde delen staan de correlaties tussen de verschillende examencijfers binnen het vmbo-tl-examen ofwel binnen het havo-examen. Hierbij valt op dat het gemiddelde centraal examencijfer matig samenhangt met de losse centraal examencijfers voor Nederlands, Engels en wiskunde. Dit beeld komt zowel bij de vmbo-tl-examencijfers als bij de havo-examencijfers naar voren. De correlatie tussen de losse examencijfers van Nederlands, Engels en wiskunde onderling is maar zeer beperkt binnen elke schoolsoort.

Tabel 5.1 Correlaties tussen centraal examencijfers vmbo-tl en havo (aantal)

	Vmbo gemiddeld cijfer alle vakken	Vmbo cijfer Neder- lands	Vmbo cijfer Engels	Vmbo cijfer wiskunde	Havo gemiddeld cijfer alle vakken	Havo cijfer Neder- lands	Havo cijfer Engels
Vmbo cijfer Nederlands	,487** (8230)						
Vmbo cijfer Engels	,576** (8230)	,229** (8228)					
Vmbo cijfer wiskunde	,612** (7271)	,119** (7270)	,104** (7270)				
Havo gemiddeld cijfer alle vakken	,590** (6166)	,276** (6165)	,374** (6164)	,324** (5488)			
Havo cijfer Nederlands	,248** (5896)	,277** (5895)	,180** (5894)	,050** (5281)	,485** (5898)		
Havo cijfer Engels	,462** (5898)	,197** (5897)	,689** (5896)	,083** (5283)	,546** (5900)	,219** (5896)	
Havo cijfer wiskunde	,252** (5148)	,113** (5147)	,000 (5146)	,337** (4970)	,627** (5149)	,178** (4932)	,052** (4934)

**p<.001

Voor toegevoegde waarde schattingen in dit deelproject zijn de donker grijs gekleurde cellen in Tabel 5.1 het meest interessant. Dit betreft de samenhang tussen het cijfer op het vmbo-tl examen (beginniveau) en het corresponderende cijfer op het havo-examen (eindniveau). Zowel voor het gemiddelde cijfer alle vakken ($r=0,59$) als voor Engels ($r=0,69$) is er een relatief sterke samenhang tussen de prestaties van de leerlingen op het vmbo-tl en havo-examen. In de toegevoegde waarde modellen voor deze twee cijfers, zal de voormeting (vmbo-tl-examen) een relatief goede voorspeller zijn van de nameting (havo-examen). De samenhang is aanzienlijk zwakker voor wiskunde ($r=0,34$) en Nederlands ($r=0,28$).

In Tabel 5.2 worden verschillen in havo centraal examencijfers gepresenteerd voor verschillende groepen leerlingen op basis van de achtergrondkenmerken. Wanneer er een relatie bestaat tussen de achtergrondkenmerken van de leerlingen en de prestaties op het havo-examen en wanneer havo-afdelingen verschillen in leerlingpopulatie, is het zinvol om deze achtergrondkenmerken op te nemen in de bepaling van de toegevoegde waarde.

Tabel 5.2 Verschillen in eindexamencijfers havo voor subgroepen leerlingen op basis van achtergrond kenmerken.

Categorieën		Gemiddelde cijfer alle vakken	Nederlands	Engels	Wiskunde
Sekse	Jongen	6,2**	5,8**	6,4**	6,2*
	Meisje	6,0	6,1	5,9	6,1
Cumi	Autochtoon	6,2**	6,0**	6,2**	6,2**
	Niet Westerse allochtoon	5,8	5,7	6,0	5,7
	Westerse allochtoon	6,2	6,0	6,5	6,0
	Onbekend	6,1	5,7	7,2	5,7
apcg	Geen overschrijding	6,2**	5,9	6,1*	6,2**
	Laag inkomen	6,0	5,9	6,2	5,8
	Niet Westerse allochtoon	6,1	5,9	6,3	6,0
	Uitkering	6,2	6,0	6,2	6,1
	Laag inkomen + uitkering	6,2	6,0	6,4	6,2
	Laag inkomen + Niet Westerse allochtoon	6,1	5,9	6,5	6,0
	Uitkering + Niet Westerse allochtoon	6,1	5,8	6,1	6,1
	Cumulatie gebied	6,0	5,9	6,2	5,9
lwoo	Ja	5,9*	5,6*	5,8*	5,7*
	Nee	6,1	5,9	6,2	6,1

**p<.001, *p<.01

Op het havo-examen presteren de jongens over het algemeen beter dan de meisjes. Wanneer de etnische achtergrond van leerlingen in ogenschouw worden genomen, blijkt dat met name de niet Westerse allochtone leerlingen lager scoren op het havo-examen. Wanneer gekeken wordt naar de buurt waar de leerlingen wonen, dan valt op dat er kleine verschillen tussen de typen buurten zijn in de examencijfers op het havo-examen. Tot slot is er een verschil in prestaties tussen de doorgestroomde leerlingen die gedurende het vmbo wel of geen leerwegondersteuning hebben gehad: leerlingen met twee scoren gemiddeld genomen lager op het havo-examen.

5.3.2 *Toegevoegde waarde van havo-afdelingen voor gemiddeld cijfer van doorstromers*

In Tabel 5.3 zijn de coëfficiënten weergegeven van de meerniveau modellen voor het gemiddelde centraal examencijfer van alle vakken. Hierin zijn alleen leerlingen opgenomen van wie alle informatie beschikbaar is en die zitten op een vestiging waarvoor gegevens beschikbaar zijn van tenminste 10 leerlingen. Voor het gemiddeld cijfer alle vakken leidt dit tot een selectie van 5.117 leerlingen in 273 vestigingen. Slechts 2 leerlingen vallen buiten de selectie doordat van hen geen gegevens beschikbaar waren van het vmbo-examen. De overige leerlingen vallen buiten de selectie omdat zij zitten op een havo-vestiging waarvan voor minder dan 10 leerlingen gegevens beschikbaar zijn. Al met al zijn het meer dan 200 vestigingen die niet voldoen aan het criterium van volledige gegevens van tenminste 10 leerlingen.

Uit het lege model in Tabel 5.3 komt naar voren dat deze selectie van leerlingen gemiddeld een 6,14 scoort op het eindexamen havo. Dit komt goed overeen met de totale steekproef leerlingen waarvan de havo-examencijfers beschikbaar zijn (Appendix bij hoofdstuk 5, Tabel 3). Wanneer naar de verdeling van variantie gekeken wordt, valt op dat het merendeel van de variantie is gebonden aan het niveau van de leerlingen (tussen leerlingen verschillen) en dat slechts een beperkt deel van de variantie kan worden toegeschreven aan havo-vestigingen. Van de totale variantie is 9% gebonden aan het vestigingsniveau (tussen VO-vestiging verschillen).

In het tweede model – het model met alleen het beginniveau – worden de examencijfers van de leerlingen op het vmbo-tl-examen toegevoegd als controle variabele (Tabel 5.3). Er is een significante positieve relatie tussen de examencijfers op het vmbo-tl-examen en het daaropvolgende havo-examen. Het meenemen van het beginniveau van de leerlingen leidt tot een afname van de variantie op zowel het niveau van de leerlingen als het niveau van de vestigingen. De afname van de variantie tussen vestigingen is een indicatie dat de havo-vestigingen verschillen in hun leerlingpopulatie met betrekking tot het beginniveau.

Tabel 5.3 Coëfficiënten van meerniveau modellen voor gemiddeld centraal examencijfer voor alle vakken

	Leeg model		Beginniveau model		Beginniveau & achtergrondkenmerken model	
	Coëfficiënt	SE	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed Part:</i>						
Constante	6,139*	0,016	6,138*	0,012	6,171*	0,014
Gemiddeld cijfer vmbo			0,739*	0,014	0,721*	0,014
Cumi: #						
Niet Westerse allochtoon					-0,162*	0,025
Westerse allochtoon					-0,027	0,033
Onbekend					0,183	0,396
Apcg:						
Laag inkomen					-0,099	0,056
Niet W-allochtoon					-0,038	0,028
Uitkering					0,001	0,041
Laaginkomen + uitkering					0,007	0,047
Laag inkomen + niet W-allochtoon					-0,040	0,080
Uitkering + niet W-allochtoon					-0,005	0,044
Cumulatiegebied					0,029	0,029
Lwoo: Ja					-0,105*	0,053
<i>Random Part:</i>						
Tussen VO-vestiging verschillen	0,047	0,006	0,022	0,003	0,020	0,003
Tussen leerlingen verschillen	0,461	0,009	0,306	0,006	0,304	0,006
<i>Model fit: -2*loglikelihood:</i>						
Aantal VO-vestigingen	273		273		273	
Aantal leerlingen	5117		5117		5117	

autochtoon is referentiegroep; * p < 0,05

In het derde model van Tabel 5.3 worden naast het beginniveau van leerlingen ook nog verschillende achtergrondkenmerken van deze leerlingen toegevoegd. Deze toevoeging leidt niet tot een grote afname van de residuele variantie op het niveau van de leerlingen of het niveau van de vestigingen. Toch zijn er enkele significante verschillen tussen groepen waarneembaar. In de eerste plaats presteren niet Westerse allochtone leerlingen lager op het havo-examen dan de autochtone leerlingen (wanneer alle andere variabelen gelijk zijn). Leerlingen die leerwegondersteuning hebben gehad op het vmbo lijken gemiddeld genomen

iets lagere cijfers te halen op het havo-eindexamen. Daarnaast zijn er geen opvallende verschillen tussen leerlingen die wonen in buurten met overschrijdingen met betrekking tot laag inkomen, uitkeringen of niet Westerse allochtonen.

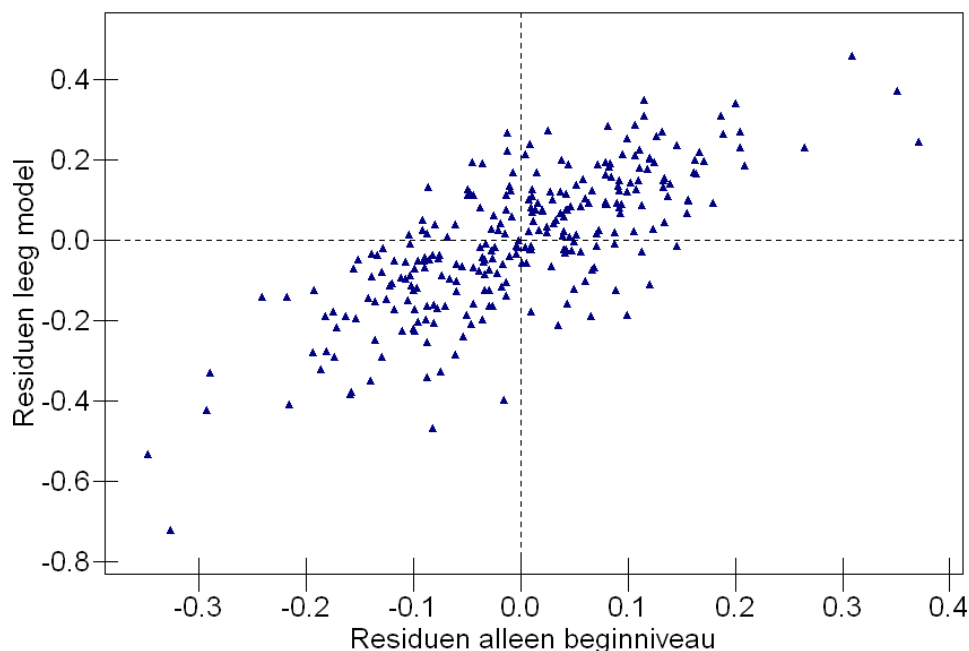
De correlaties tussen de vestigingsresiduen afkomstig uit de drie meerniveau modellen zijn gepresenteerd in Tabel 5.4.

Tabel 5.4 Correlaties tussen vestigingsresiduen voor gemiddeld centraal examencijfer voor alle vakken

	Leeg model	Beginniveau model	Beginniveau & achtergrondkenmerken model
Leeg model	1,00		
Beginniveau model	0,76	1,00	
Beginniveau & achtergrondkenmerken model	0,74	0,99	1,00

De correlatie tussen het lege model en het model met alleen beginniveau is relatief sterk ($r=0,76$). Echter, de correlatie is verre van perfect wat er op duidt dat het toevoegen van het beginniveau aan het model leidt tot aanzienlijke wijzigingen. Zie ook Figuur 5.1.

Figuur 5.1 Scatterplot van vestigingsresiduen voor leeg en beginniveau model gemiddeld centraal examencijfer voor alle vakken



Dit betekent ook dat er een aanzienlijk verschil zit tussen de bruto en netto examenresultaten van havo-vestigingen als doorstroom-vestigingen. Wat verder nog opvalt is de zeer sterke correlatie tussen het de vestigings-residuen uit het model met alleen beginniveau en het model waarin daarnaast ook achtergrondkenmerken van de leerlingen zijn opgenomen ($r=0,99$). Dit duidt erop dat de toegevoegde waarde van havo-vestigingen voor doorstromende vmbo-tl-leerlingen nauwelijks verandert als voor kenmerken van de leerlingpopulatie wordt gecorrigeerd.

5.3.3 *Toegevoegde waarde van havo-afdelingen voor Nederlands van doorstromers*

Naast de toegevoegde waarde voor het gemiddelde centraal examencijfer voor alle vakken kan de toegevoegde waarde ook bepaald worden voor de losse cijfers voor Nederlands, Engels en wiskunde. Voor elk van deze afhankelijke variabelen worden dezelfde selectie criteria gebruikt voor leerlingen en vestigingen als bij het gemiddelde centraal examencijfer.

In Tabel 5.5 worden de drie modellen gepresenteerd voor het schatten van de toegevoegde waarde voor het vak Nederlands. Voor Nederlands geldt dat er een selectie overblijft van 4.987 leerlingen in 259 vestigingen die voldoen aan de gestelde selectie criteria. Het lege model laat zien dat de selectie van havo-leerlingen gemiddeld een 5,9 scoort op het centraal examen Nederlands. Dit komt wederom goed overeen met de prestaties van alle doorstromende leerlingen op het Nederlands examen havo (Appendix bij hoofdstuk 5, Tabel 3). Wanneer naar de verdeling van de variantie wordt gekeken valt op dat de verschillen tussen leerlingen groot en verschillen tussen de vestigingen maar zeer beperkt zijn. Van de totale variantie wordt slechts 5,5% gebonden aan het niveau van de vestigingen. Dit geeft een indicatie dat op vestigingsniveau de verschillen in examencijfers Nederlands niet zo heel groot zijn.

In het tweede model is het beginniveau van de leerlingen toegevoegd als voorspeller. In dit geval betreft dit de cijfers van de leerlingen op het Nederlands centraal examen vmbo-tl. Zoals te verwachten, blijkt dat er een positieve relatie is tussen de examencijfers van de leerlingen op het vmbo-tl en havo-examen. Ook voor Nederlands leidt het toevoegen van het beginniveau aan het model tot een afname van de variantie op leerling- en vestigingsniveau, echter, deze afname is niet zo sterk als voor het gemiddelde eindexamencijfer. Dit is niet verwonderlijk aangezien de correlatie (Tabel 5.1) tussen de examencijfers van Nederlands aanzienlijk minder sterk was.

De achtergrondkenmerken van de leerlingen worden aanvullend nog toegevoegd in het derde model (Tabel 5.5). Hieruit komt een vergelijkbaar beeld naar voren als bij het gemiddelde centraal examencijfer. Het toevoegen leidt tot een kleine afname van de residuele variantie op het niveau van leerlingen en vestigingen. Wanneer alle andere kenmerken van de leerling gelijk zijn, scoren niet Westerse allochtone leerlingen gemiddeld genomen lager dan de autochtone leerlingen (referentiegroep). Tevens behalen de leerlingen in een cumulatiegebied

(apcg) gemiddeld genomen lagere cijfers op het Nederlands eindexamen havo dan de leerlingen die woonachtig zijn in een gebied zonder overschrijdingen. Tot slot blijkt dat ook voor het Nederlands eindexamen de leerlingen die gedurende het vmbo leerwegondersteuning hebben gehad lager scoren dan de leerlingen die geen leerwegondersteuning hebben gehad.

Tabel 5.5 Coëfficiënten van toegevoegde waarde modellen voor centraal examencijfer Nederlands

	Leeg model		Beginniveau model		Beginniveau en achtergrond-kenmerken model	
	Coëfficiënt	SE	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed Part:</i>						
Constante	5,913*	0,019	5,913*	0,018	5,946*	0,021
Cijfer Nederlands vmbo			0,338*	0,017	0,322*	0,017
Cumi: #						
Niet-Westerse allochtoon					-0,223*	0,040
Westerse allochtoon					-0,038	0,055
Onbekend					0,177	0,636
Apcg:						
Laag inkomen					0,033	0,092
Niet W-allochtoon					-0,018	0,045
Uitkering					0,039	0,067
Laaginkomen + uitkering					0,002	0,076
Laag inkomen + niet W-allochtoon					0,072	0,132
Uitkering + niet W-allochtoon					-0,101	0,071
Cumulatiegebied					0,098*	0,048
Lwoo: Ja					-0,181*	0,087
<i>Random Part:</i>						
Tussen VO-vestiging verschillen	0,050	0,009	0,041	0,007	0,038	0,007
Tussen leerlingen verschillen	0,850	0,018	0,790	0,016	0,784	0,016
<i>Model fit: -2*loglikelihood:</i>						
Aantal VO-vestigingen	259		259		259	
Aantal leerlingen	4897		4897		4897	

autochtoon is referentiegroep; * $p < 0,05$

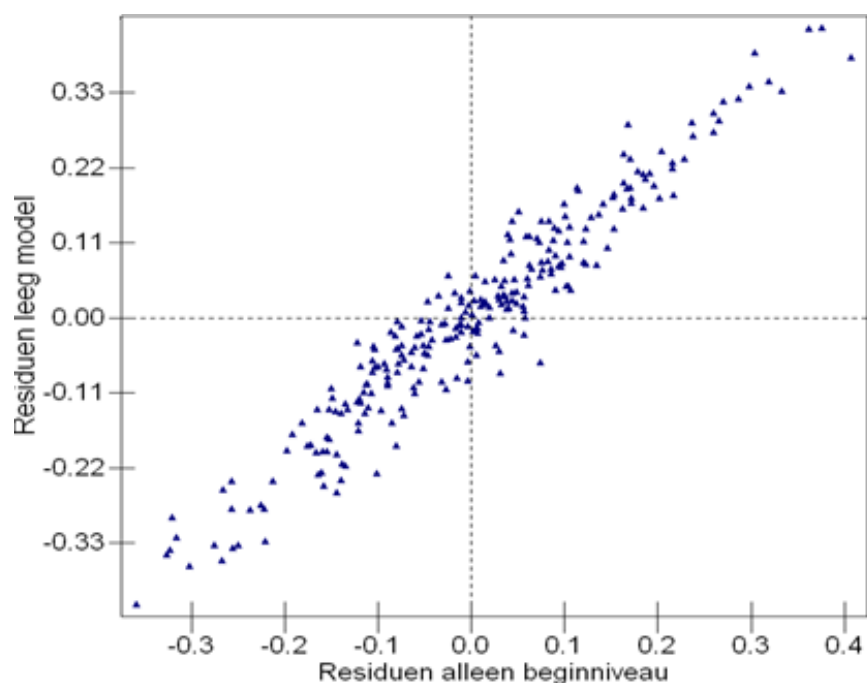
De correlaties tussen de residuen op het niveau van de vestigingen afkomstig uit de drie meerniveau modellen voor Nederlands worden gepresenteerd in Tabel 5.6.

Tabel 5.6. Correlaties tussen vestigingsresiduen voor centraal examencijfer Nederlands

	Leeg model	Beginniveau model	Beginniveau en achtergrondkenmerken model
Leeg model	1,00		
Beginniveau model	0,96	1,00	
Beginniveau & achtergrondkenmerken model	0,95	0,99	1,00

Er zijn sterke correlaties gevonden tussen de vestigingsresiduen uit alle drie modellen, sterker dan bij de modellen voor het gemiddelde centraal examencijfer voor alle vakken. De correlatie tussen het lege model en het model met alleen beginniveau bedraagt 0,96. Dit indiceert dat er wel een paar verschuivingen tussen vestigingen plaatsvinden, maar dat dit beperkt is. Om dit verder zichtbaar te maken zijn de vestigingsresiduen uit het lege model en het model met alleen beginniveau weergegeven in Figuur 5.2.

Figuur 5.2 Scatterplot van vestigingsresiduen voor leeg en beginniveau model centraal examencijfer Nederlands



Wanneer dit omgezet zou worden naar rangordes, blijkt dat de sterk presterende vestigingen amper in rangorde verschuiven door het toevoegen van het beginniveau in de schatting van de toegevoegde waarde. De grootste verschuivingen in de rangorde vinden plaats bij de gemiddelde en zwakkere vestigingen. Wederom is de correlatie tussen het de vestigingsresiduen uit het model met alleen beginniveau en het model waarin daarnaast ook achtergrondkenmerken zijn opgenomen het sterkst ($r=0,99$). Dit duidt erop dat de toegevoegde waarde voor vestigingen voor deze specifieke groep leerlingen niet wezenlijk verandert door het toevoegen van achtergrondkenmerken van leerlingen.

5.3.4 Toegevoegde waarde van havo-afdelingen voor Engels van doorstromers

De resultaten voor de coëfficiënten van de toegevoegde waarde modellen voor Engels zijn gepresenteerd in Tabel 5.7. Voor Engels bestaat de selectie van leerlingen uit 4.895 leerlingen afkomstig van 259 vestigingen. Deze selectie van leerlingen heeft een gemiddeld cijfer van 6,2 behaald op het eindexamen havo Engels. Dit komt goed overeen met de prestaties van de totale groep doorstromende leerlingen (Appendix bij hoofdstuk 5, Tabel 3).

In het lege model zijn de verschillen tussen vestigingen weer kleiner dan de verschillen tussen de leerlingen. Van de totale variantie in eindexamencijfers voor Engels kan slechts 3,4% gebonden worden aan het niveau van vestigingen. Uit het tweede model blijkt dat er wederom sprake is van een significante positieve relatie tussen de centraal examencijfers Engels voor vmbo-tl en havo. Vergelijkbaar met het gemiddelde examencijfer voor alle vakken leidt voor Engels de toevoeging van het beginniveau tot een behoorlijke daling van de variantie op zowel leerling als vestigingsniveau. De afname van de variantie op vestigingsniveau duidt op het bestaan van verschillen in leerlingpopulatie tussen vestigingen in het beginniveau voor het vak Engels. De variantie in het derde model is amper afgenomen door het toevoegen van de achtergrondkenmerken van leerlingen, hoewel er her en der wel wat statistisch significante verschillen tussen groepen leerlingen zijn. Opvallend is dat er geen significante verschillen zijn tussen de prestaties van de leerlingen op het havo-examen Engels voor leerlingen met een verschillende etnische achtergrond (Cumi) en voor leerlingen die gedurende het vmbo-tl wel of geen leerwegondersteuning hebben gehad (lwoo).

Tabel 5.7 Coëfficiënten van toegevoegde waarde modellen voor centraal examencijfer Engels

	Leeg model		Beginniveau model		Beginniveau en achtergrond-kenmerken model	
	Coëfficiënt	SE	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed Part:</i>						
Constante	6,165*	0,024	6,164*	0,017	6,124*	0,021
Cijfer Engels vmbo			0,821*	0,012	0,816*	0,012
Cumi: #						
Niet Westerse allochtoon					-0,025	0,041
Westerse allochtoon					0,087	0,056
Onbekend					1,266	0,656
Apcg:						
Laag inkomen					0,164	0,094
Niet W-allochtoon					0,167*	0,045
Uitkering					-0,029	0,069
Laaginkomen + uitkering					0,208*	0,078
Laag inkomen + niet W-allochtoon					0,188	0,136
Uitkering + niet W-allochtoon					0,094	0,073
Cumulatiegebied					0,063	0,049
Lwoo: Ja					-0,139	0,090
<i>Random Part:</i>						
Tussen VO-vestiging verschillen	0,057	0,013	0,027	0,006	0,027	0,006
Tussen leerlingen verschillen	1,612	0,033	0,849	0,018	0,843	0,017
<i>Model fit: -2*loglikelihood:</i>						
	16359,6		13206,0		13173,5	
Aantal VO-vestigingen	259		259		259	
Aantal leerlingen	4895		4895		4895	

autochtoon is referentiegroep; * p < 0,05

De correlaties tussen de residuen op het niveau van de vestigingen afkomstig uit de drie meerniveau modellen voor Engels worden gepresenteerd in Tabel 5.8. In de eerste plaats valt het op dat de correlatie tussen het lege model en het model met alleen beginniveau vrij sterk is ($r=0,69$). Echter, de correlatie is verre van perfect. Het toevoegen van het beginniveau aan de schatting van de toegevoegde waarde van een vestiging leidt tot aanzienlijke wijzigingen. Dit duidt erop dat de verschillen tussen bruto en netto prestaties van vestigingen nog meer van elkaar verschillen voor Engels dan voor het gemiddelde eindexamencijfer. Wat wederom

opvalt is de zeer sterke correlatie tussen het de vestigingsresiduen uit het model met alleen beginniveau en het model waarin daarnaast ook achtergrondkenmerken zijn opgenomen ($r=0,99$). De toegevoegde waarde van vestigingen voor het vak Engels voor deze specifieke groep leerlingen verandert nauwelijks door het toevoegen van achtergrondkenmerken van de leerlingen.

Tabel 5.8 Correlaties tussen vestigingsresiduen voor centraal examencijfer Engels.

	Lege model	Beginniveau model	Beginniveau en achtergrondkenmerken model
Lege model	1,00		
Beginniveau model	0,69	1,00	
Beginniveau & achtergrondkenmerken model	0,67	0,99	1,00

5.3.5 Toegevoegde waarde van havo-afdelingen voor wiskunde van doorstromers

De resultaten voor de coëfficiënten van de toegevoegde waarde modellen voor wiskunde zijn gepresenteerd in Tabel 5.9. Voor wiskunde bestaat de selectie van leerlingen uit 3.742 leerlingen afkomstig van 217 vestigingen. Het hanteren van deze selectie criteria leidt ertoe dat voor wiskunde de toegevoegde waarde voor minder vestigingen bepaald kan worden.

De selectie van doorstromende vmbo-tl-leerlingen in de meerniveau analyses heeft een gemiddeld cijfer van 6,1 behaald op het centraal examen havo voor het vak wiskunde. Dit komt goed overeen met de prestaties van de totale groep doorstromende leerlingen (Tabel 5.3). In totaal kan 8,3% van de totale variantie in wiskunde cijfers worden gebonden aan het niveau van vestigingen. Vergelijkbaar met de resultaten van Nederlands leidt het toevoegen van een beginniveau wel tot een afname van de varianties op de beide niveaus, maar is deze afname niet zo sterk als voor Engels en het gemiddelde cijfer alle vakken. Ook hier kan dit worden verklaard vanuit de relatief zwakke relatie die gevonden is tussen vmbo-tl en havo-examencijfers voor wiskunde (Tabel 5.4). Voor wiskunde worden dan ook hoge correlaties gevonden tussen alle modellen (Tabel 5.10). Deze zijn vergelijkbaar met de correlaties die eerder ook voor Nederlands zijn gevonden.

Tabel 5.9 Coëfficiënten van toegevoegde waarde modellen voor centraal examencijfer wiskunde

	Leeg model		Beginniveau model		Beginniveau en achtergrondkenmerken model	
	Coëfficiënt	SE	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed Part:</i>						
Constante	6,111*	0,032	6,111*	0,030	6,197*	0,034
Cijfer wiskunde vmbo			0,382*	0,018	0,363*	0,018
Cumi: #						
Niet Westerse allochtoon					-0,285*	0,061
Westerse allochtoon					-0,002	0,086
Onbekend					0,000	0,000
Apcg:						
Laag inkomen					-0,194	0,137
Niet W-allochtoon					-0,213*	0,067
Uitkering					0,008	0,102
Laaginkomen + uitkering					-0,051	0,122
Laag inkomen + niet W-allochtoon					0,103	0,230
Uitkering + niet W-allochtoon					-0,011	0,107
Cumulatiegebied					-0,061	0,075
Lwoo: Ja					-0,257	0,133
<i>Random Part:</i>						
Tussen VO-vestiging verschillen	0,134	0,022	0,114	0,019	0,104	0,018
Tussen leerlingen verschillen	1,480	0,035	1,321	0,031	1,309	0,031
<i>Model fit: -2*loglikelihood:</i>						
Aantal VO-vestigingen	217		217		217	
Aantal leerlingen	3742		3742		3742	

autochtoon is referentiegroep; * p < 0,05

Tabel 5.10 Correlaties tussen vestigingsresiduen voor centraal examencijfer wiskunde

	Leeg model		Beginniveau model		Beginniveau en achtergrondkenmerken model	
Leeg model		1,00				
Beginniveau model		0,95		1,00		
Beginniveau & achtergrondkenmerken model		0,93		0,99		1,00

Wanneer de voorgaande resultaten van de meerniveau modellen in ogenschouw worden genomen voor alle uitkomstmaten, kan worden gesteld dat voor Engels en het gemiddelde centraal examencijfer alle vakken relatief grote verschillen tussen havo-vestigingen zijn gevonden wat betreft het wel of niet corrigeren voor het beginniveau. Deze verschillen zijn maar zeer beperkt voor Nederlands en wiskunde. Voor alle vier uitkomstmaten komt naar voren dat meenemen van achtergrondkenmerken in de schatting van de toegevoegde waarde van havo-vestigingen niet leidt tot substantiële veranderingen in de schattingen van de effectiviteit. Op basis hiervan kan eventueel gekozen worden voor een toegevoegde waarde model waarin alleen het beginniveau van leerlingen is opgenomen. In de volgende paragraaf zal daarom verder worden gegaan met dit eenvoudige toegevoegde waarde model.

5.3.6 *Samenhang tussen toegevoegde waarde gemiddeld cijfer, Nederlands, Engels en wiskunde van havo-afdelingen*

De tweede onderzoeksvraag betreft de relatie tussen de toegevoegde waarde schattingen van havo-afdelingen voor de verschillende afhankelijke variabelen. In de voorgaande paragraaf zijn voor 217 tot 273 vestigingen schattingen gemaakt voor de toegevoegde waarde op basis van het gemiddelde eindexamencijfer, Nederlands, Engels en wiskunde. In Tabel 5.11 zijn de correlaties weergegeven tussen de schattingen van de toegevoegde waarde van havo-afdelingen voor de vier afhankelijke variabelen.

Tabel 5.11 Correlaties tussen vestigingsresiduen voor de vier toegevoegde waarde schattingen

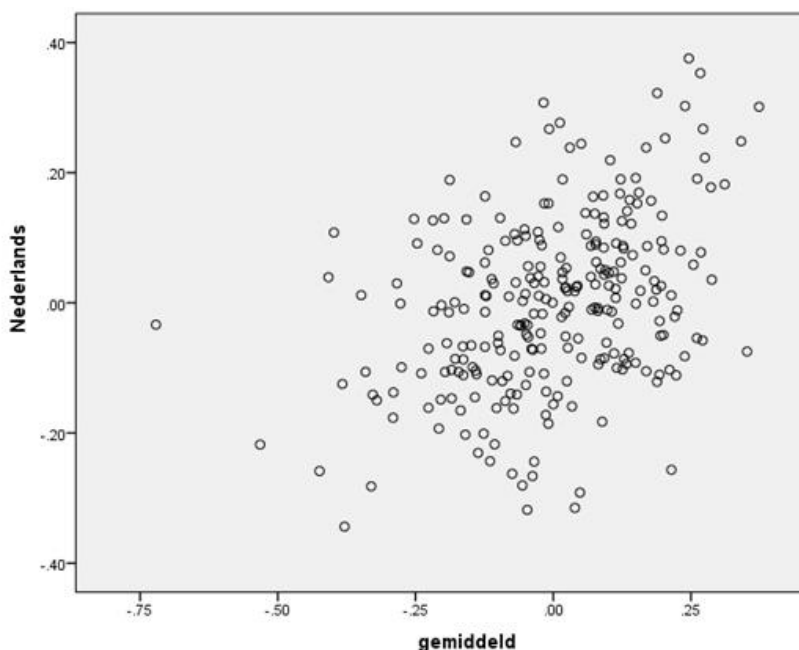
	Gemiddeld cijfer alle vakken	Nederlands	Engels
Nederlands	0,385** N=259		
Engels	0,452** N=259	0,135* N=259	
Wiskunde	0,582** N=217	0,254** N=209	0,224** N=209

** . Correlation is significant at the 0.01 level (2-tailed). * . Correlation is significant at the 0.05 level (2-tailed).

Tabel 5.11 toont dat de correlaties tussen de toegevoegde waarde schattingen positief maar relatief laag zijn. Gezien de eerder gevonden correlaties tussen de examencijfers op het havo-examen (Tabel 5.1) is dit resultaat van een relatief zwakke associaties tussen de toegevoegde waarde schattingen niet verwonderlijk. De positieve associaties betekenen dat over het algemeen een hogere schatting van de toegevoegde waarde op de ene afhankelijke variabele samengaat met een hoge schatting op de toegevoegde waarde voor een andere afhankelijke

variabele. Bijvoorbeeld een havo-afdeling met een hoge toegevoegde waarde voor het gemiddelde centraal examencijfer voor alle vakken heeft waarschijnlijk ook een hoge toegevoegde waarde voor Nederlands, Engels of wiskunde. Echter, doordat de correlaties relatief laag zijn betekent dit resultaat ook dat er een aanzienlijk aantal vestigingen is waarbij dit niet het geval is. Om dit verder zichtbaar te maken is als voorbeeld een scatterplot gemaakt waarin de toegevoegde waarde voor het gemiddelde cijfer alle vakken is afgezet tegen de toegevoegde waarde voor Nederlands (Figuur 5.3).

Figuur 5.3 Scatterplot van vestigingsresiduen voor gemiddeld centraal examencijfer alle vakken en Nederlands



Figuur 5.3 laat zien dat ondanks dat er een zwakke positieve relatie is, scholen met een vergelijkbare toegevoegde waarde voor het gemiddelde centraal examencijfer voor alle vakken veel van elkaar verschillen in de toegevoegde waarde voor het vak Nederlands. Wanneer naar specifieke combinaties van toegevoegde waarde schattingen wordt gekeken, valt op dat de toegevoegde waarde voor wiskunde en het gemiddelde cijfer alle vakken het sterkst met elkaar samenhangen. Iets zwakkere correlaties worden gevonden tussen de toegevoegde waarde op basis van het gemiddelde cijfer alle vakken en toegevoegde waarde op basis van Nederlands en Engels. De correlaties zijn het laagst voor de samenhang van de toegevoegde waarde tussen de afzonderlijke kernvakken.

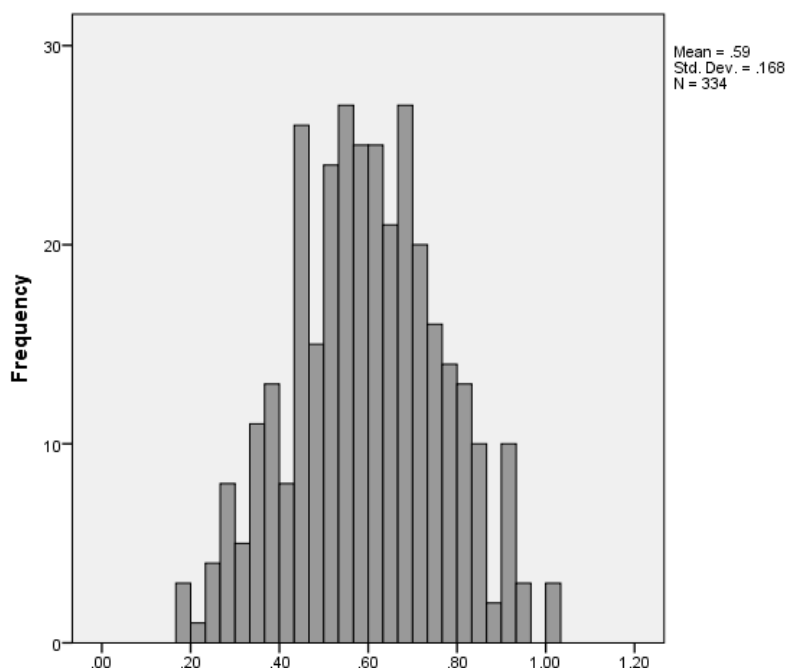
5.3.7 Doorstroomrendement van havo-afdelingen

Het nominaal doorlopen van het havo-traject en het behalen van een havo-diploma kan bepaald worden voor 8.205 leerlingen. Van deze leerlingen slaagt 57,8% (n=4.742) binnen de

nominale tijd voor het havo-examen. De overige leerlingen (n=3.463) stromen uit, zijn gezakt voor het havo-examen of zijn bij de overgang van havo-4 naar havo-5 blijven zitten. De 31 leerlingen die na 1 jaar havo-4 direct doorstromen naar vwo zijn niet in deze bepaling opgenomen.

Wanneer wederom alleen de vestigingen waarvan tenminste 10 leerlingen gegevens hebben voor het rendement worden geselecteerd, kan voor 334 havo-afdelingen een doorstroomrendement worden bepaald. Dit aantal is groter dan bij de toegevoegde waarde bepalingen, omdat ook uitstromende leerlingen in het rendement zijn betrokken. Het percentage leerlingen dat nominaal slaagt voor het havo-examen verschilt aanzienlijk tussen de havo-afdelingen; van minimaal 17% tot maximaal 100% nominale doorstroom. Gemiddeld realiseren de 334 havo-afdelingen een doorstroomrendement van 56%. De verdeling van de rendementen van havo-afdelingen is gepresenteerd in Figuur 5.4.

Figuur 5.4 Histogram van de doorstroomrendementen van havo-afdelingen



5.3.8 Samenhang tussen toegevoegde waarde en doorstroomrendement van havo-afdelingen

Het doorstroomrendement van havo-afdelingen is bepaald omdat scholen hun toegevoegde waarde artificieel zouden kunnen laten stijgen door leerlingen vroegtijdig uit te laten stromen of te laten zitten, waardoor de sterkere leerlingen overblijven en de voorbereidingstijd voor het havo-examen langer wordt. In Tabel 5.12 zijn de correlaties gepresenteerd tussen het doorstroomrendement van de havo-afdelingen en de vier toegevoegde waarde indicatoren.

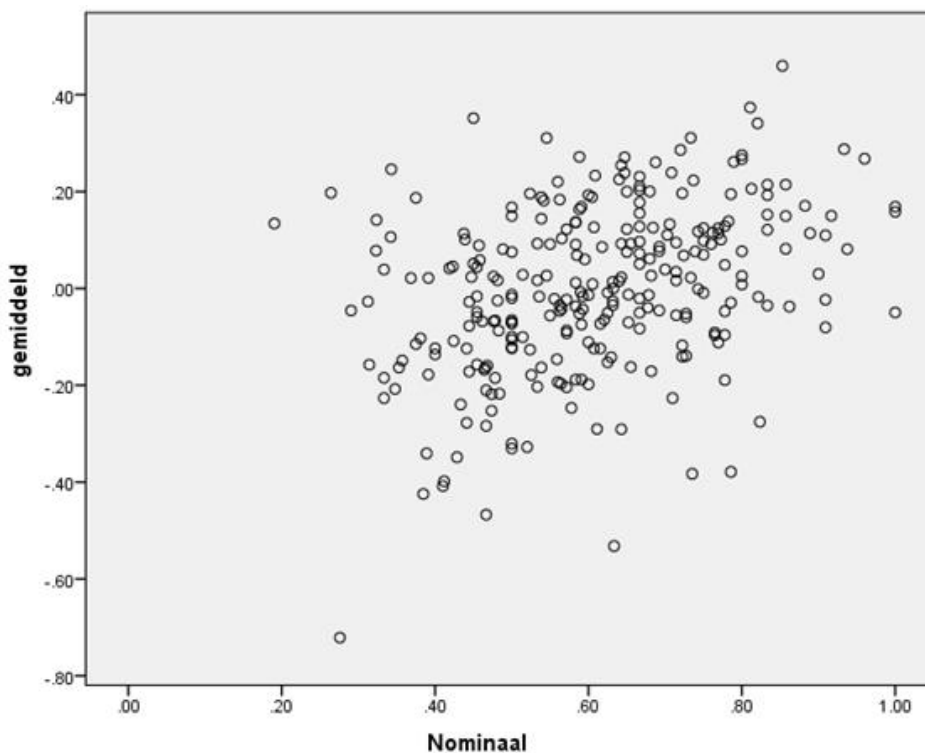
Tabel 5.12 Correlaties tussen doorstroomrendement en de vier toegevoegde waarde indicatoren

	Gemiddeld cijfer alle vakken	Nederlands	Engels	wiskunde
Doorstroomrendement	0,364*	0,087	0,074	0,125

* $p < 0,001$

Tussen het doorstroomrendement en de toegevoegde waarde op basis van het gemiddelde centraal examencijfer voor alle vakken wordt een zwakke positieve associatie gevonden ($r=0,36$). Dit betekent dat over het algemeen een hoog rendement voor doorstromers op een havo-afdeling samengaat met een hoge toegevoegde waarde (gemiddelde cijfer) voor deze groep leerlingen. Echter de zwakke samenhang indiceert ook dat er relatief veel scholen zijn die van deze relatie afwijken. Ter illustratie is in Figuur 5.5 een scatterplot te zien waarin het doorstroomrendement is afgezet tegen de toegevoegde waarde op basis van het gemiddelde cijfer alle vakken. In deze figuur staat het doorstroomrendement op de x-as (nominaal) en de toegevoegde waarde op de y-as (gemiddeld).

Figuur 5.5 Scatterplot van doorstroomrendement en toegevoegde waarde van het gemiddelde centraal examencijfer alle vakken



In Figuur 5.5 is te zien dat er een groep havo-afdelingen is die een laag doorstroomrendement heeft in combinatie met een relatief hoge score op de toegevoegde waarde voor het gemiddelde centraal examencijfer voor alle vakken (links boven in de figuur). Hier zou strategisch gedrag een rol kunnen spelen. Daarnaast is er een groep scholen die relatief hoge doorstroomrendementen behaalt, maar daarbij slechts een beperkte toegevoegde waarde heeft voor het gemiddelde centraal examencijfer voor alle vakken (rechts onder in de figuur). Voor de toegevoegde waarde indicatoren die gebaseerd zijn op de afzonderlijke examencijfers van Nederlands, Engels en wiskunde geldt dat er geen statistisch significante associatie is gevonden met het doorstroomrendement van de havo-afdelingen.

De positieve samenhang tussen rendement en toegevoegde waarde voor het gemiddeld examencijfer alle vakken maakt duidelijk dat de meerderheid van scholen waarschijnlijk geen strategisch gedrag vertoont om goede eindexamenresultaten te behalen voor doorstromende vmbo-tl-leerlingen. De zwakte van dit verband toont aan dat er ook uitzonderingen op deze regel kunnen zijn. Dit indiceert dat alleen het bepalen van een toegevoegde waarde indicator onvoldoende is voor het in kaart brengen van de opbrengsten van een havo-afdeling voor de doorstromende leerlingen van vmbo-tl naar havo.

6 Deel IV: Extern rendement vmbo-kl en bl

6.1 Inleiding

In de huidige opbrengstbeoordeling gebruikt de Inspectie van het Onderwijs verschillende manieren om naar de leeropbrengsten van leerlingen in scholen te kijken. Voorbeelden hiervan zijn de hoogte van de examencijfers van de leerlingen, maar ook de mate waarin leerlingen op- en afstromen of blijven zitten gedurende het voortgezet onderwijs (VO). Daar waar mogelijk worden deze gegevens gecorrigeerd voor verschillende leerlingkenmerken om tot een zo eerlijk mogelijke vergelijking van scholen te kunnen komen. Deze gebruikte manieren kunnen worden gezien als de opbrengsten van scholen op de korte termijn (oftewel het interne rendement). Een aanvulling hierop kan zijn om te kijken hoe de onderwijsloopbanen van leerlingen van een VO-vestiging eruit zien na het verlaten van de school. De indicator die hiervoor door de Inspectie van het Onderwijs is ontwikkeld is het externe rendement.

Het externe rendement is met name belangrijk voor die sectoren van het funderend onderwijs die niet als eindpunt beschouwd kunnen worden, zoals het voortgezet onderwijs. Een belangrijk doel van het VO is het realiseren van een passende doorstroom naar een vervolgopleiding in het hoger of middelbaar beroepsonderwijs. Dit geldt in het bijzonder voor de leerlingen die een vmbo-opleiding volgen, aangezien het afronden van een vmbo-opleiding niet gelijk staat aan het behalen van een startkwalificatie. Tevens bestaat beleid dat een groep vmbo-leerlingen uit de basisberoepsgerichte leerweg niet deelnemen aan het examen, maar ongediplomeerd kunnen doorstromen naar een mbo-opleiding¹¹. Deze groep leerlingen komt bij de huidige opbrengstindicatoren van de Inspectie nog onvoldoende naar voren, aangezien van deze leerlingen geen prestaties bekend zijn aan het einde van het voortgezet onderwijs en daardoor niet opgenomen kunnen worden in de indicatoren voor het centraal examen. Het externe rendement kan door de Inspectie van het Onderwijs gebruikt worden om ook deze leerlingen in de opbrengstbepaling van een VO-vestiging mee te nemen. Het externe rendement van het vmbo is dan een indicator voor de mate waarin VO-vestigingen erin slagen om hun leerlingen succesvol te plaatsen in het vervolgonderwijs. Voor vmbo basisberoepsgerichte (bl) of kaderberoepsgerichte leerweg (kl) kan dat zijn het vervolgen in een hogere leerweg in het vmbo of plaatsing in een mbo-opleiding op niveau 2 of hoger (zie ook 6.2.2.).

¹¹ In actieplan “Focus op Vakmanschap 2011 – 2015” wordt voorgesteld dat de drempelloze doorstroom van leerlingen uit het vmbo naar niveau 2 opleidingen in het mbo wordt afgeschaft vanaf schooljaar 2013-2014. Echter, de mogelijkheden tot samenwerking tussen VO- en mbo-instellingen wordt vergroot door het verlengen van het VM2 experiment.

In de notitie ‘Eindrapportage onderzoek indicator extern rendement vmbo’ worden drie mogelijke normen besproken voor het externe rendement. Op basis van deze normen zou kunnen worden bepaald in hoeverre het externe rendement van een VO-vestiging voldoende is. De drie mogelijke normen zijn 80, 85 en 90 procent succesvol geplaatste leerlingen. De effecten van deze voorgestelde normering zijn relevant om mee te nemen in dit onderzoek.

De onderzoeksvragen die centraal staan in dit hoofdstuk zijn:

1. *Wat is het externe rendement na één en twee jaar mbo van vmbo basis- of kaderberoepsgerichte leerweg afdelingen van VO-vestigingen gegeven het examenklas niveau van hun leerlingen?*
2. *Wat zijn de effecten van de voorgestelde normering voor percentage succesvolle plaatsingen?*

In overleg met de opdrachtgever is ervoor gekozen om daarnaast nog specifiek te onderzoeken in hoeverre het externe rendement van VO-vestigingen wordt beïnvloed of vertekend door invloeden vanuit het mbo. Hierbij kunnen de volgende deelvragen worden geformuleerd:

3. *In hoeverre is de succesvolle plaatsing van leerlingen in het eerste jaar na het verlaten van een vmbo basis- of kaderberoepsgerichte leerweg gerelateerd aan de VO-vestiging enerzijds en mbo-instelling anderzijds?*
4. *In hoeverre leidt het corrigeren voor invloeden van mbo-instellingen tot veranderingen van de schattingen van het externe rendement voor VO-vestigingen?*

In de volgende paragraaf zal een korte schets worden gegeven van mogelijke invloeden die leiden tot vertekening van het externe rendement, een voorbeeld van een andere context waar een vergelijkbaar probleem speelt en ten slotte de methodologie die veelal wordt gebruikt om dergelijke invloeden te onderzoeken.

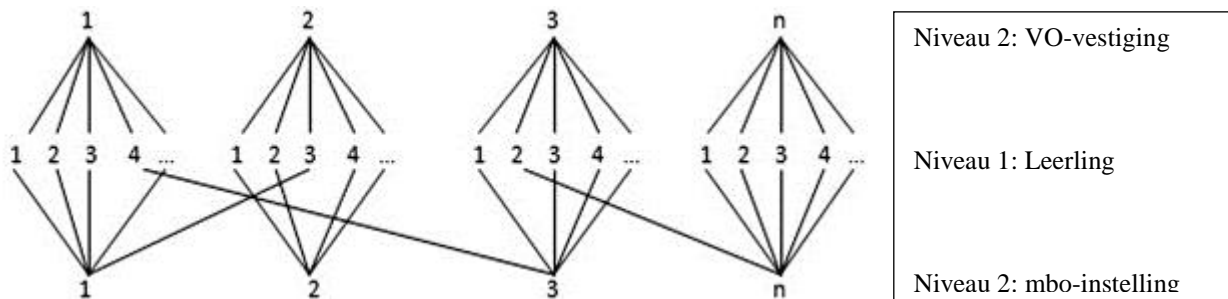
6.1.1 Onderzoeken van lange termijn effecten

Het externe rendement van een vmbo-bl of -kl afdeling geeft aan welk deel van de leerlingen op deze schoolsoorten een of twee jaar na afsluiting nog onderwijs volgt op een mbo-opleiding op niveau 2 of hoger of op een hogere leerweg van vmbo. Dit zijn dan de succesvol geplaatste leerlingen. Hoe meer succesvolle plaatsingen een VO-vestiging realiseert, hoe hoger het externe rendement en hoe beter de opbrengsten van vmbo-bl en kl-afdelingen op de betreffende VO-vestiging zijn. Echter, een succesvolle plaatsing is van meer factoren afhankelijk dan alleen de VO-vestiging. Met name voor plaatsing in het beroepsonderwijs geldt dat de beschikbaarheid van opleidingen op de verschillende niveaus en plaatsingsbeleid van mbo-instellingen mogelijke factoren kunnen zijn waardoor leerlingen wel of niet succesvol geplaatst worden. Deze invloeden vanuit de mbo-instelling kunnen voor VO-vestigingen een verschillende impact hebben op het externe rendement. Deze effecten zijn nog niet in deze hoedanigheid onderzocht, echter de Eindnotitie Indicator extern rendement laat

zien dat er grote verschillen in succesvolle plaatsing zijn tussen sectoren in het mbo (Van der Heijden & Vreeburg, 2010). In het externe rendement zou men echter het liefst alleen de unieke bijdrage van een VO-vestiging willen vangen, om zo tot een zo eerlijk mogelijke vergelijking tussen vestigingen te komen. Hiervoor is het noodzakelijk dat onderzocht wordt in hoeverre het mogelijk is om de eventuele invloeden van mbo-instellingen uit het externe rendement te filteren.

Een vergelijkbaar probleem speelt bij het analyseren van eventuele lange termijn effecten van basisscholen (Timmermans, Sniijders & Bosker, 2013). Hierbij wordt getracht te onderzoeken of leerlingen van bepaalde basisscholen meer vooruitgang boeken in het VO dan leerlingen van andere basisscholen. Maar ook hier staat de vooruitgang niet alleen onder invloed van een eventueel lange termijn effect van een basisschool, maar ook van de effectiviteit van de bezochte school voor VO. Omdat er sprake is van een vergelijkbare problematiek als in dit deelproject kan voor het onderzoeken van het externe rendement van het vmbo gebruikt gemaakt worden van dezelfde leidraden voor analyse.

Figuur 6.1 Structuur van een kruisklassificatie model



Om zicht te krijgen op de unieke bijdrage van de VO-vestiging aan het externe rendement zal getracht moeten worden de invloeden van de mbo-instelling hieruit te halen. Dit wordt gedaan door middel van het schatten van een meerniveau model waarbij leerlingen (niveau 1) zowel zijn genest in VO-vestigingen (niveau 2) als in mbo-instellingen (niveau 2). Deze modellen worden ook wel kruisklassificatie modellen genoemd en zijn ook in Hoofdstuk 4 toegepast. Dergelijke modellen kennen de structuur zoals deze weergegeven is in Figuur 6.1. In het midden van de figuur zijn de leerlingen zichtbaar. Deze leerlingen zijn genest binnen VO-vestigingen, wat zichtbaar is in het bovenste deel van de Figuur. Tevens zijn de leerlingen genest in mbo-instellingen, zichtbaar in het onderste deel van de figuur. Echter, niet alle leerlingen van dezelfde VO-vestiging zullen daarna geplaatst worden in dezelfde mbo-instelling, maar kunnen zich verspreiden over meerdere mbo-instellingen. Bijvoorbeeld, de leerlingen van VO-vestiging 1 gaan hoofdzakelijk naar mbo-instelling 1, maar leerling 4 gaat naar mbo-instelling 3. Daarnaast zijn de leerlingen van mbo-instelling afkomstig van

meerdere VO-vestigingen. De leerlingen van mbo-instelling 1 zijn zoals gezegd grotendeels afkomstig van VO-vestiging 1, maar ook uit VO-vestiging 2. Het analyseren van een dergelijk model leidt tot het isoleren van de afzonderlijke effecten van VO-vestigingen en mbo-instellingen op de plaatsing van de leerlingen in een mbo niveau 1, 2, 3, of 4 opleiding.

Het analyseren van een dergelijk model kan daardoor zicht geven op de invloeden van mbo-instellingen op de succesvolle plaatsing van VO-leerlingen en ook op de vertekeningen die ontstaan op de schattingen van het externe rendement van een VO-vestiging. In de volgende empirische analyse zal uitgegaan worden van dit model als conceptueel sterkste model voor het analyseren van het externe rendement. De huidige indicator en een simpeler meerniveau model zullen met dit conceptueel sterkste kruisklassificatie model worden vergeleken.

6.2 Methode

6.2.1 Beschrijving van de dataset

De totale dataset bestaat uit 101.616 cases van leerlingen in de vierde (of vijfde klas) van het vmbo. Het betreft de dataset van alle Nederlandse leerlingen die in schooljaar 2010-2011 in de examenklas van het vmbo hebben gezeten (Bron-VO bestanden). Voor het externe rendement kijken we vervolgens naar de plaatsing van de leerlingen in het mbo in het schooljaar 2011-2012 en het schooljaar 2012-2013.

De geleverde dataset bevat de resultaten en doorstroom van leerlingen uit alle leerwegen. Voor dit onderzoek kijken we alleen naar het externe rendement van de leerlingen uit de leerwegen basisberoepsgericht (23.720 leerlingen) en kaderberoepsgericht (26.931 leerlingen). De leerlingen in de basisberoepsgericht leerweg (bl) zitten verspreid op 507 vestigingen van VO-scholen (combinatie van brinnummer en vestigingsnummer). Per vestiging zijn er gemiddeld 47 leerlingen in de basis leerweg, met een minimum van 1 en een maximum van 234 leerlingen. Ook de leerlingen in de kaderberoepsgericht leerweg zijn verdeeld over 507 VO-vestigingen. Per vestiging zijn er gemiddeld 53 leerlingen in de kader leerweg, met een minimum van 1 en een maximum van 176 leerlingen.

6.2.2 Extern rendement

Voor het schatten van het externe rendement wordt gebruik gemaakt van de beslisregels zoals deze op dit moment al door de Inspectie van het Onderwijs worden toegepast. Deze beslisregels zijn beschreven in de notitie “Toelichting Plaatsing na vierde jaar vmbo” beschikbaar via de website van de Inspectie van het Onderwijs en in de Notitie “Eindrapportage onderzoek indicator extern rendement vmbo”. Op basis van deze beslisregels wordt voor iedere leerling bepaald of er na het examenjaar van het vmbo sprake is van een succesvolle plaatsing in het mbo of een hogere leerweg in het VO. De globale beslisregels voor het bepalen van een succesvolle plaatsing van een leerlingen zijn weergegeven in Tabel

1. De variabele die de succesvolle plaatsing van leerlingen weergeeft vormt de kern van de indicator extern rendement.

Tabel 6.1 Welke doorstroom na het vierde jaar is succesvol voor vmbo basis- en kaderberoepsgerichte leerweg?*

School	Vsv	Vmbo-bl/kl	Vmbo-gl/tl of hoger	Mbo 1	Mbo 2	Mbo 3	Mbo 4
Gediplomeerd	Nee	Nvt	Ja	Nee	Ja	Ja	Ja
Ongediplomeerd	Nee	Nvt	Ja	Nee	Ja	Ja	Ja

Vsv= voortijdig school verlaten.

* Tabel overgenomen uit notitie “Eindrapportage onderzoek indicator extern rendement vmbo”

In het externe rendement worden een aantal groepen leerlingen buiten beschouwing gelaten. In de eerste plaats betreft dit de leerlingen die hun onderwijsloopbaan vervolgen in dezelfde of een lagere leerweg, aangezien het hier om leerlingen gaat die de opleiding niet verlaten. Een leerling die een opleiding vmbo-basis heeft gedaan en daarna doorstroomt naar kader wordt dus niet in het externe rendement meegenomen. In Tabel 6.1 zijn deze leerlingen zichtbaar in de kolom (Vmbo-bl/kl). In de tweede plaats blijven ook de leerlingen buiten beschouwing die hun onderwijsloopbaan buiten VO, VAVO of mbo vervolgen. Een mogelijke reden hiervoor is dat er voor deze groep leerlingen maar zeer beperkt zicht is op hun onderwijsloopbanen, waardoor slecht kan worden vastgesteld of de leerlingen een succesvol vervolg van de onderwijsloopbaan ondergaan.

Voor de overige leerlingen geldt dat er wordt gekeken naar de onderwijsposities in het jaar na het verlaten van de VO-instelling. Op basis van deze onderwijsposities wordt vastgesteld of er sprake is geweest van een succesvolle plaatsing of niet. Leerlingen zijn succesvol geplaatst wanneer zij binnen het VO of VAVO-onderwijs volgen in een hogere leerweg of schoolsoort (Vmbo-gt/tl of hoger in Tabel 6.1). Daarnaast geldt voor de bl- en kl-leerlingen dat er sprake is van een succesvolle plaatsing wanneer ze onderwijs volgen in een mbo-niveau 2 opleiding of hoger. Niet succesvol geplaatst zijn de leerlingen die voortijdig schoolverlater zijn (vsv in tabel 1) of die hun onderwijs vervolgen in een mbo-niveau 1 opleiding. Opgemerkt moet worden dat er geen onderscheid gemaakt wordt tussen verschillende groepen voortijdig schoolverlaters (met of zonder startkwalificatie of vmbo/mbo-1-diploma). In een aantal gevallen kan de door de leerling gevolgde onderwijsloopbaan niet worden ondergebracht in één van de kolommen van Tabel 6.1. In die gevallen is succesvolle plaatsing gebaseerd op coderingen zoals ze staan beschreven in de Notitie “Eindrapportage onderzoek indicator extern rendement vmbo”.

Op basis van deze door de Inspectie van het Onderwijs geformuleerde regels wordt een binaire variabele aangemaakt die aangeeft of er sprake is van een succesvolle plaatsing (0=niet succesvol, 1=succesvol). Deze variabele is de basis voor alle varianten van het externe rendement die in dit hoofdstuk beschreven zullen worden. Nadat op basis van de voorgaande beslisregels wordt bepaald welke leerlingen in het externe rendement worden opgenomen en welke leerlingen succesvol of niet-succesvol zijn geplaatst, kan de huidige indicator voor het externe rendement worden bepaald aan de hand van de volgende formule:

$$\text{Extern rendement} = \frac{\text{Aantal succesvol geplaatste leerlingen}}{\text{Totaal aantal leerlingen in extern rendement}} * 100\%$$

Op deze manier wordt antwoord gegeven op de eerste hoofdvraag van dit deelproject naar het externe rendement een en twee jaar na het verlaten van het vmbo.

De Inspectie overweegt om voor de normering van het externe rendement de grenswaarden van 80, 85 of 90 procent succesvol geplaatste leerlingen toe te passen. De gevolgen voor de classificatie van VO-vestigingen bij hantering van de verschillende grenswaarden worden inzichtelijk gemaakt door per grenswaarde het percentage VO-vestigingen dat onder of boven de norm valt te berekenen. Dit beantwoordt de tweede hoofdvraag van deelproject IV.

6.2.3 *Multilevel en kruisklassificatie modellen voor extern rendement*

Voor het isoleren van de unieke bijdrage van VO-vestigingen aan het externe rendement zullen zogenoemde kruisklassificatie modellen worden geschat. Dit zijn een specifiek type meerniveau modellen. Hiervoor is gebruikt gemaakt van de MLwiN 2.28 software (Rasbash, Steele, Browne, Goldstein, 2009; Browne, 2009). De analyses zullen afzonderlijk uitgevoerd worden voor vestigingen van de basisberoepsgerichte leerweg en voor de kaderberoepsgerichte leerweg. Tevens zullen afzonderlijke analyses uitgevoerd worden voor het externe rendement 1 jaar na de examenklas vmbo en 2 jaar na de examenklas vmbo. In totaal betekent dit dat voor bijna alle VO-vestigingen vier schattingen beschikbaar zullen zijn van het externe rendement, namelijk hun externe rendement voor de vmbo-basis-leerlingen na één jaar en na twee jaar en het externe rendement voor hun vmbo-kader-leerlingen na één jaar en na twee jaar.

Bij elke afzonderlijke analyse zullen twee modellen met elkaar worden vergeleken, namelijk 1) een meerniveau model waarbij de leerlingen zijn genest in een vmbo-vestiging, en 2) een kruisklassificatie meerniveau model waarbij leerlingen zowel genest zijn in vmbo-vestigingen als in mbo-instellingen. Het eerste meerniveau model maakt gebruik van dezelfde data als de huidige indicator voor het externe rendement. Daarvan verwachten we dat de resultaten ook grotendeels overeen zullen komen. In het kruisklassificatie model wordt van meer informatie gebruik gemaakt, aangezien daar ook de inschrijving in de mbo-instelling wordt meegenomen. De coëfficiënten uit het kruisklassificatie model geven antwoord op de eerste twee deelvragen.

In elk van de meerniveau modellen zal de variabele succesvolle plaatsing van een leerling gebruikt worden als afhankelijke variabele. Omdat dit een variabele is die twee waarden kent, namelijk 0 (niet succesvol) en 1 (wel succesvol) zijn logistische regressiemodellen geschat. Aan de twee modellen worden geen onafhankelijke of controle variabelen toegevoegd, aangezien dit in de huidige indicator voor het externe rendement ook niet wordt gedaan. Het gaat dus om zogenaamde ‘lege’ modellen. Voor het schatten van de beide typen meerniveau modellen zullen niet de gebruikelijke maximum likelihood schattingsmethoden worden gebruikt, maar is gekozen voor Markov Chain Monte Carlo schattingsmethoden. Deze complexe Bayesiaanse schattings-methode wordt om twee redenen gekozen. In de eerste plaats is deze methode meer flexibel voor analyseren van kruisklassificatie modellen (Browne, Goldstein & Rasbash, 2001). In de tweede plaats leidt deze schattingsmethode in logistische regressie modellen tot betere schattingen van de verschillen tussen VO-vestigingen en mbo-instellingen (Browne, & Draper, 2006).

De vergelijking tussen het meerniveau model en het kruisklassificatie model geeft zicht op de vertekeningen van het externe rendement door invloeden van de mbo-instellingen en op basis daarvan kan de derde deelvraag beantwoord worden. De vergelijking tussen de twee modellen zal in de eerste plaats gemaakt worden door het berekenen van de correlatie van de externe rendementen van vestigingen tussen de twee modellen. Er is ook gekeken naar verschuivingen volgens de ‘bolletjes’-systematiek, maar de resultaten hiervan zijn beschreven in de Appendix bij hoofdstuk 5. Het kruisklassificatie model kan alleen geschat worden voor de leerlingen die na de examenklas vmbo doorstromen naar het mbo. Voor de overige leerlingen, bijvoorbeeld de voortijdig schoolverlaters of leerlingen die naar hogere leerwegen doorstromen, is er geen identificatie variabele beschikbaar voor de mbo-instelling. Dit betekent dat er een zeer selecte groep leerlingen wordt opgenomen in deze analyse. Een groot deel van de leerlingen met een niet-succesvolle plaatsing (vsv) zijn niet in deze analyses betrokken. Een vergelijkbare selectie van leerlingen is gemaakt in de Notitie “Eindrapportage onderzoek indicator extern rendement vmbo” (Van der Heijden & Vreeburg, 2010) voor het analyseren van verschillen in succesvolle plaatsen naar mbo sector- en instellingsverschillen. Het spreekt voor zich dat bij de vergelijking van het meerniveau model met het kruisklassificatie model dezelfde steekproef van leerlingen gebruikt zal worden.

6.2.4 *Correctie voor controle variabelen*

In de inschrijvingsbestanden die gebruikt worden voor het bepalen van het externe rendement van VO-vestigingen zijn een aantal achtergrondkenmerken van leerlingen beschikbaar die eventueel meegenomen kunnen worden in het externe rendement. Door het statistisch controleren voor de achtergrondkenmerken wordt mogelijk een eerlijkere vergelijking tussen vestigingen gerealiseerd. Richtlijnen voor de keuze van de controle variabelen in schattingen van school effecten zijn: 1) de variabelen moeten samenhangen met succesvolle plaatsing en 2) het betreft kenmerken van leerlingen die buiten de controle van de school liggen. Drie

achtergrondkenmerken uit het inschrijvingsbestand komen mogelijk in aanmerking om meegenomen te worden in het externe rendement. Het gaat om:

Leerwegondersteuning (lwoo): Dit is een binaire variabele die aangeeft of een leerling gedurende het voortgezet onderwijs in aanmerking kwam voor extra leerweg ondersteuning. De operationalisatie is gelijk aan die in Hoofdstuk 4 en 5.

Etnische achtergrond (cumi): De etnische achtergrond van leerlingen is gemeten in drie categorieën, te weten autochtone leerlingen, Westers allochtone leerlingen en niet Westers allochtone leerlingen. Deze variabele is gelijk aan die in Hoofdstuk 5.

Armoedeprobleemcumulatiegebied (apcg): De variabele armoedeprobleemcumulatiegebied is niet zozeer een kenmerk van de leerling, maar vooral een kenmerk van de buurt/wijk waarin de leerling woont. Op basis van relatieve aantallen van veel inwoners met een laag inkomen, een uitkering en/of niet Westers allochtone inwoners worden acht categorieën van buurten onderscheiden. Deze variabele komt geheel overeen met die in Hoofdstuk 5.

De invloed van bovenstaande achtergrondkenmerken op de schattingen van het externe rendement zal worden onderzocht door middel van een meerniveau model. In dit model is de succesvolle plaatsing van leerlingen de afhankelijke variabele. Hiervoor wordt een twee niveau structuur geanalyseerd waarbij leerlingen (niveau 1) zijn genest binnen VO-vestigingen (niveau 2). Voor deze analyse wordt niet gebruik gemaakt van de beperkte selectie van leerlingen (alleen doorstroom naar mbo), maar is gebruik gemaakt van de volledige groep leerlingen waarvoor succesvolle plaatsing bepaald kan worden. Deze keuze is gemaakt om de vergelijkbaarheid met de huidige indicator voor het externe rendement zo groot mogelijk te houden.

De achtergrondkenmerken zullen worden meegenomen als verklarende variabelen op het niveau van de leerlingen. Om zicht te krijgen op de invloed van het meenemen van de achtergrondkenmerken zullen residuen (schattingen van het externe rendement) bewaard worden voor zowel een meerniveau model zonder achtergrondkenmerken als voor een meerniveau model met de drie achtergrondkenmerken.

6.3 Resultaten

6.3.1 *Extern rendement volgens huidige rekenregels een jaar na examenklas vmbo-bl of kl*

In Tabel 6.2 worden enkele beschrijvende statistieken weergegeven met betrekking tot de aantallen leerlingen per vestiging waarover het externe rendement is bepaald. Het gaat hier om het externe rendement één jaar nadat de leerlingen in de examenklas van vmbo-bl of kl zaten.

Tabel 6.2 Beschrijvende informatie over aantallen leerlingen per vestiging voor extern rendement na 1 jaar

Variabele	Basisberoepsgerichte leerweg (N=507)			Kaderberoepsgerichte leerweg (N=507)		
	Gemiddelde	SD	Range	Gemiddelde	SD	Range
Succesvolle plaatsingen	37,7	25,4	0 - 92	48,9	30,3	0 - 167
Niet succesvolle plaatsingen	5,8	7,4	0 - 187	2,5	2,4	0 - 16
Leerlingen opgenomen in extern rendement	43,6	29,9	0 - 219	51,3	31,4	1 - 175
Leerlingen niet in extern rendement	3,2	6,7	0 - 73	1,8	2,2	0 - 12

In de eerste plaats laat Tabel 6.2 zien dat er tussen vestigingen aanzienlijke verschillen in de aantallen leerlingen zijn waarover het externe rendement kan worden bepaald. Voor vmbo-bl varieert dit van geen enkele leerling tot 219 leerlingen per vestiging. Ook bij de vmbo-kl zijn de verschillen aanzienlijk; van minimaal 1 tot maximaal 175 leerlingen. Op een vergelijkbare wijze zijn er met name in het vmbo-bl aanzienlijke verschillen in de aantallen leerlingen die *niet* meegenomen worden in het externe rendement. Met betrekking tot de aantallen niet succesvolle plaatsingen is de spreiding tussen vestigingen voor de basisberoepsgerichte leerweg (7,4) aanzienlijk groter dan voor de kaderberoepsgerichte leerweg (2,4). Daarentegen zijn de verschillen tussen vestigingen in de aantallen leerlingen met een succesvolle plaatsing groter voor de kaderberoepsgerichte leerweg (30,4 vs. 25,4).

Gemiddeld halen VO-vestigingen waar de basisberoepsgerichte leerweg wordt aangeboden een extern rendement van 86,8%. Dit gemiddelde is vergelijkbaar met de resultaten gepresenteerd in de Notitie “Eindrapportage onderzoek indicator extern rendement vmbo”. De verschillen tussen vestigingen zijn groot aangezien de range van het externe rendement ligt tussen 0 en 100 procent. Vestigingen waar de kaderberoepsgerichte leerweg wordt aangeboden halen gemiddeld een extern rendement 94,5%.

Door de Inspectie zijn drie mogelijke normen voorgesteld voor het externe rendement van een VO-vestiging. Op basis van deze normen zou kunnen worden bepaald in hoeverre het externe rendement van een VO-vestiging voldoende is. De drie mogelijke normen zijn 80, 85 en 90 procent succesvol geplaatste leerlingen. In Tabel 6.3 zijn de percentages VO-vestigingen weergegeven die onder of boven deze normen vallen op basis van deze dataset. Deze tabel geeft het overzicht voor zowel voor vestigingen met tenminste 10 leerlingen als voor de complete groep vestigingen. In deze laatste groep zit ook een aantal vestigingen met een zeer klein aantal leerlingen, die betrokken zijn in het bepalen van het externe rendement. Wanneer alleen naar de VO-vestigingen wordt gekeken met tenminste 10 leerlingen waarvoor succesvolle plaatsing kan worden vastgesteld loopt het aantal vestigingen terug van 506 naar 463 voor de vmbo-bl en van 507 naar 486 vestigingen voor de vmbo-kl.

Tabel 6.3 Resultaten van extern rendement voor VO-vestigingen voor bl en kl voor de plaatsing van leerlingen 1 jaar na vmbo examenklas

Alle vestigingen:

Norm	Basisberoepsgerichte leerweg (N=506)*				Kaderberoepsgerichte leerweg (N=507)			
	Boven de norm		Onder de norm		Boven de norm		Onder de norm	
	N	%	N	%	N	%	N	%
80%	422	83,2	84	16,6	493	97,2	14	2,8
85%	343	67,7	163	32,1	481	94,9	26	5,1
90%	233	46,0	273	53,8	432	85,2	75	14,8

Vestigingen met tenminste 10 leerlingen in extern rendement:

Norm	Basisberoepsgerichte leerweg (N=463)*				Kaderberoepsgerichte leerweg (N=486)			
	Boven de norm		Onder de norm		Boven de norm		Onder de norm	
	N	%	N	%	N	%	N	%
80%	393	84,9	70	15,1	480	98,8	6	1,2
85%	315	68,0	148	32,0	469	96,5	17	3,5
90%	209	45,1	254	54,9	421	86,6	65	13,4

* Voor één vestiging kan het externe rendement niet worden bepaald.

Tabel 6.3 laat zien dat het externe rendement van vestigingen voor de kaderberoepsgerichte leerweg hoger zijn dan voor de basisberoepsgerichte leerweg. Wanneer naar de norm van 90% succesvolle plaatsingen wordt gekeken blijkt dat 85,2% van alle vestigingen voor kl daaraan voldoet. Voor de bl voldoet slechts 46,0% van de vestigingen aan deze norm van 90% succesvolle plaatsingen. Een verklaring voor het verschil tussen kl en bl kan liggen in de beslisregels voor het bepalen van een succesvolle plaatsing. Zowel voor leerlingen afkomstig uit kl als vanuit bl wordt gesteld dat een plaatsing in een mbo opleiding op niveau 2 succesvol is. Dit is een relatief soepele beslisregel voor leerlingen uit de kaderberoepsgerichte leerweg¹². Eén van de achterliggende redenen voor deze keuze is dat bij een strengere beslisregel (minimaal mbo-niveau 3 voor vmbo-kl-leerlingen) de verschillen in succesvolle plaatsing tussen VO-instellingen groter werd.

¹² De achtergrond voor de soepele beslisregel voor vmbo-kl-leerlingen was dat vooral bij deze opleiding het mbo-effect sterk speelt. Afhankelijk van het ROC en de sector van het ROC kunnen deze leerlingen geplaatst worden in niveau 2, 3 of 4 van het mbo. De laagst mogelijke zinvolle norm is gekozen om met dit effect rekening te houden.

6.3.2 Extern rendement volgens huidige rekenregels twee jaar na examenklas vmbo-bl of kl

In Tabel 6.4 worden enkele beschrijvende statistieken weergegeven met betrekking tot de aantallen leerlingen per vestiging waarover het externe rendement is bepaald wanneer het gaat om succesvolle plaatsingen twee jaar nadat de leerlingen in de examenklas van vmbo-bl of kl zaten. Hierbij moet worden opgemerkt dat alleen leerlingen zijn opgenomen in deze bepaling die in 2010-2011 in de examenklas vmbo-basis of kader zaten en na dat examenjaar de opleiding hebben verlaten. Al deze leerlingen waren ook opgenomen in het externe rendement na één jaar.

Tabel 6.4 Beschrijvende informatie over aantallen leerlingen per vestiging voor extern rendement na twee jaar

Variabele	Basisberoepsgerichte leerweg (N=507)			Kaderberoepsgerichte leerweg (N=507)		
	Gemiddelde	SD	Range	Gemiddelde	SD	Range
Succesvolle plaatsingen	36,7	25,1	0 – 174	47,5	29,5	0 – 160
Niet succesvolle plaatsingen	6,9	7,5	0 – 92	3,8	3,3	0 – 18
Leerlingen opgenomen in extern rendement	43,5	29,9	0 – 219	51,3	31,4	1 – 175
Leerlingen niet in extern rendement	3,3	6,8	0 – 73	1,8	2,2	0 – 12

Wanneer naar de aantallen leerlingen per vestiging wordt gekeken komt grotendeels hetzelfde beeld naar voren als bij het externe rendement één jaar na de examenklas. Er zijn aanzienlijke verschillen tussen vestigingen in de aantallen leerlingen waarover het externe rendement na twee jaar kan worden bepaald. Op een vergelijkbare wijze zijn er met name in de basisberoepsgerichte leerweg aanzienlijke verschillen in de aantallen leerlingen die niet meegenomen worden in het externe rendement (standaard deviatie van 6,8). Met betrekking tot de aantallen niet succesvolle plaatsingen zijn de verschillen tussen vestigingen voor basisberoepsgerichte leerweg aanzienlijk groter dan voor kaderberoepsgerichte leerweg (bl: standaard deviatie 7,7; kl: standaard deviatie 3,3). In vergelijking met de resultaten één jaar na de examenklas neemt het gemiddelde aantal niet succesvolle plaatsingen per vestiging toe. Gemiddeld zijn er op vestigingen bijna 37 succesvolle plaatsingen van basis-leerlingen en bijna 48 succesvolle plaatsingen van kader-leerlingen.

In Tabel 6.5 worden de percentages vestigingen weergegeven die onder of boven de normen van 80, 85 of 90 procent succesvolle plaatsingen liggen. Deze tabel geeft het overzicht voor zowel alle vestigingen als voor een selectie van vestigingen met tenminste 10 leerlingen met een externe rendementsbepaling.

Tabel 6.5 Resultaten van extern rendement voor VO-vestigingen voor bl en kl voor de plaatsing van leerlingen twee jaar na vmbo examenklas

Alle vestigingen:

Norm	Basisberoepsgerichte leerweg (N=506)*				Basisberoepsgerichte leerweg (N=507)			
	Boven de norm		Onder de norm		Boven de norm		Onder de norm	
	N	%	N	%	N	%	N	%
80%	375	74,0	131	25,8	485	95,7	22	4,3
85%	278	54,8	228	45,0	443	87,4	64	12,6
90%	161	31,8	345	68,0	352	69,4	155	30,6

Vestigingen met tenminste 10 leerlingen in extern rendement:

Norm	Basisberoepsgerichte leerweg (N=463)*				Kaderberoepsgerichte leerweg (N=486)			
	Boven de norm		Onder de norm		Boven de norm		Onder de norm	
	N	%	N	%	N	%	N	%
80%	346	74,7	117	25,3	472	97,1	14	2,9
85%	251	54,2	212	45,8	431	88,7	55	11,3
90%	138	29,8	325	70,2	342	70,4	144	29,6

* Voor één vestiging kan het externe rendement niet worden bepaald.

Gemiddeld halen vestigingen waar de basisberoepsgerichte leerweg wordt aangeboden een extern rendement van 84,4%. Hiermee is het gemiddelde externe rendement na twee jaar iets lager dan het gemiddelde rendement van vestigingen na één jaar. De verschillen tussen vestigingen voor bl zijn groot aangezien de range van het externe rendement ligt tussen 20,6 en 100 procent. Vestigingen waar de kaderberoepsgerichte leerweg wordt aangeboden halen gemiddeld een extern rendement van 92,4%. Ook dit is iets lager dan het externe rendement na één jaar (94,5%). Vergelijkbaar met de gemiddelden laat ook Tabel 6.5 zien dat het externe rendement van vestigingen voor de kaderberoepsgerichte leerweg hoger is dan voor de basisberoepsgerichte leerweg. Wanneer de huidige norm voor het externe rendement wordt gehanteerd - 90% succesvolle plaatsingen -, voldoet 69% van alle vestigingen daaraan voor kl. Voor de bl voldoet slechts 32% van de vestigingen aan de norm van 90%.

6.3.3 Isoleren van VO-vestigingseffect in extern rendement een jaar na vmbo-bl of kl

Een van de problemen van een indicator voor het externe rendement van vmbo-bl of kl afdelingen van VO-vestigingen kan de mogelijke vertekening door mbo-instellingen zijn. Dit is getracht te analyseren door middel van een meerniveau kruisklassificatie model. De resultaten van het meerniveau model en het kruisklassificatie model voor het externe rendement van vmbo-vestigingen wordt gepresenteerd in Tabel 6.6. Het gaat hier om de plaatsing van de leerlingen in het eerste jaar na de examenklas vmbo. In de analyses zijn

alleen de leerlingen opgenomen die aan een aantal voorwaarden voldoen, namelijk 1) alleen leerlingen die doorstromen naar een mbo opleiding, 2) alleen leerlingen van een VO-vestiging waarvan tenminste 10 leerlingen met gegevens over succesvolle plaatsing beschikbaar zijn, en 3) alleen leerlingen in mbo-instellingen waarvan tenminste 10 leerlingen met gegevens over succesvolle plaatsing beschikbaar zijn.

Tabel 6.6 Resultaten van de meerniveau analyses voor het bepalen van de unieke bijdrage VO-vestiging aan extern rendement voor succesvolle plaatsing één jaar na examenklas vmbo-bl of kl

	Basisberoepsgerichte leerweg		Kaderberoepsgerichte leerweg	
	Multilevel model	Kruisklassificatie model	Multilevel model	Kruisklassificatie model
<i>Fixed part:</i>				
Constante	2,95 (0,06)*	3,32 (0,16)*	4,63 (0,10)*	4,94 (0,16)*
<i>Random part:</i>				
Tussen VO-vestiging verschillen	0,54 (0,07)	0,45 (0,07)	0,61 (0,16)	0,44 (0,15)
Tussen mbo-instelling verschillen		0,77 (0,25)		0,64 (0,25)
<i>Model fit (DIC):</i>				
Aantal VO-vestigingen	458	458	484	484
Aantal mbo-instellingen		64		64
Aantal leerlingen	20.097	20.097	24.818	24.818

Uit het lege meerniveau model voor de leerlingen afkomstig uit de examenklas bl blijkt dat er relevante verschillen zijn in succesvolle plaatsing tussen vmbo-vestigingen. Van de verschillen in succesvolle plaatsing tussen leerlingen is 14,1%¹³ gerelateerd aan de vmbo-vestigingen waar zij op hebben gezeten. Voor de kader-leerlingen lijken verschillen tussen de vestigingen iets groter. Van de verschillen in succesvolle plaatsing is 15,6% gerelateerd aan de vmbo-vestigingen waar deze leerlingen op hebben gezeten.

Wanneer voor de bl-leerlingen naar het kruisklassificatie model wordt gekeken valt op dat de fit van het model beter wordt. Dit blijkt uit een kleinere waarde voor de Deviance Information

¹³ Het percentage variantie op het niveau van de VO-vestigingen is bepaald door de Intraclass correlatie (ICC) te berekenen (Snijders & Bosker, 2012). In een logistisch model wordt geen variantie gegeven op leerlingniveau, maar hiervoor kan 3,29 worden ingevuld. De totale variantie in succesvolle plaatsing is $0,54+3,29=3,83$. Daarvan is 14% (0,54) gerelateerd aan het niveau van VO-vestigingen.

Criteria¹⁴ (DIC) indicator. Een vergelijkbaar maar kleiner verschil in DIC wordt gevonden voor leerlingen afkomstig uit een examenklas kl. Dit betekent voor beide groepen leerlingen dat het model waarbij leerlingen zowel genest zijn in een vmbo-vestiging als in een mbo-instelling een betere weerspiegeling van de data geeft. In praktische zin kan dit opgevat worden dat zowel de vmbo-vestiging als de mbo-instelling een rol spelen bij de eventuele succesvolle plaatsing van de leerlingen één jaar nadat ze in de examenklas hebben gezeten.

De resultaten gepresenteerd in Tabel 6.6 laten zien dat de verschillen in succesvolle plaatsing tussen basis-leerlingen sterker gerelateerd zijn aan de mbo-instelling waar de leerlingen naar toe gaan, dan dat deze verschillen gerelateerd zijn aan de vmbo-vestiging waar de leerlingen van afkomstig zijn. De variantie tussen mbo-instellingen (0,77) is namelijk groter dan de variantie tussen VO-vestigingen (0,45). Van alle verschillen tussen leerlingen in succesvolle plaatsing wordt 17,1%¹⁵ (ICC mbo) gebonden aan de mbo-instelling waar de leerlingen naartoe zijn gegaan en 10,0% (ICC VO) aan de VO-vestiging waar de leerlingen vandaan zijn gekomen. Voor de kader-leerlingen is een vergelijkbaar beeld gevonden. De variantie tussen VO-vestigingen neemt af door het opnemen van de mbo-instellingen in het kruisklassificatie model. Ten opzichte van de verschillen tussen VO-vestigingen (0,44) zijn de verschillen tussen mbo-instellingen groter (0,64). Voor de kl-leerlingen geldt dat van alle verschillen tussen leerlingen in succesvolle plaatsing 14,6% (ICC mbo) wordt gebonden aan de mbo-instelling waar de leerlingen naartoe zijn gegaan en 10,1% (ICC VO) wordt gebonden aan de VO-vestiging waar de leerlingen vandaan zijn gekomen.

6.3.4 *Isoleren van VO-vestigingseffect in extern rendement twee jaar na vmbo-bl of kl*

De resultaten van het meerniveau model en het kruisklassificatie model voor het externe rendement van vmbo-vestigingen twee jaar na de examenklas wordt gepresenteerd in Tabel 6.7. Wederom is de analyse uitgevoerd op een selectie van leerlingen. De selectiecriteria van leerlingen die meegenomen worden in het externe rendement na twee jaar zijn dezelfde als voor het externe rendement één jaar na de examenklas.

¹⁴ Dit is een indicator voor de fit van een model die gebruikt wordt voor het vergelijken van multilevel modellen die geschat zijn op basis van MCMC schattingsmethoden. De DIC is een combinatie van de fit van het model bij de data en de complexiteit van het model. Hoe kleiner de DIC waarde hoe beter het model. Een afname van 5 punten in de DIC waarde kan worden gezien als een significante verbetering van de fit van een model (Spiegelhalter, Best, Carlin, Van der Linde, 2002).

¹⁵ De Intraclass correlaties in een kruisklassificatie model kunnen worden geschat op basis van de formules uit Snijder & Bosker, 2012; pagina 109. De intraclass correlatie voor mbo (ICC mbo) kan ook worden geïnterpreteerd als de correlatie tussen de succesvolle plaatsingen van leerlingen die naar dezelfde mbo-instelling gaan maar afkomstig zijn van een andere VO-vestiging. Vergelijkbaar kan de intraclass correlatie voor VO (ICC VO) worden geïnterpreteerd als de correlatie tussen de succesvolle plaatsingen van leerlingen die van dezelfde VO-vestiging afkomstig zijn, maar naar een andere mbo-instelling zijn gegaan.

Tabel 6.7 Resultaten van de meerniveau analyses voor het bepalen van de unieke bijdrage VO-vestiging aan extern rendement voor succesvolle plaatsing twee jaar na examenklas vmbo-bl of kl

	Basisberoepsgerichte leerweg		Kaderberoepsgerichte leerweg	
	Multilevel model	Kruisklassificatie model	Multilevel model	Kruisklassificatie model
<i>Fixed part:</i>				
Constante	3,85 (0,07)*	4,04 (0,11)*	5,65 (0,17)*	5,80 (0,22)*
<i>Random part:</i>				
Tussen vestiging verschillen	0,49 (0,09)	0,45 (0,11)	0,56 (0,30)	0,38 (0,29)
Tussen mbo-instelling verschillen		0,27 (0,11)		0,42 (0,24)
<i>Model fit (DIC):</i>				
Aantal VO-vestigingen	451	451	483	483
Aantal mbo-instellingen		67		68
Aantal leerlingen	18.438	18.438	23.827	23.827

Het lege meerniveau model voor de leerlingen afkomstig uit de examenklas bl laat wederom zien dat er relevante verschillen zijn in succesvolle plaatsing tussen vmbo-vestigingen (variantie vestigingsniveau 0,49). Van de verschillen in succesvolle plaatsing tussen basisleerlingen is 13,0% gerelateerd aan de vmbo-vestigingen waar zij op hebben gezeten. Dit percentage ligt een fractie lager dan het percentage na één jaar. Voor de kader-leerlingen lijken verschillen tussen de vestigingen wederom iets groter (0,56) dan voor basisleerlingen, maar ook voor kl is de variantie op vestigingsniveau iets lager dan na één jaar. Van de verschillen in succesvolle plaatsing twee jaar na de examenklas voor kl-leerlingen is 14,5% gerelateerd aan de vmbo-vestigingen waar deze leerlingen op hebben gezeten.

Wanneer voor de bl-leerlingen naar het kruisklassificatie model wordt gekeken, valt op dat de fit van het model beter wordt (Tabel 6.7). Dit blijkt uit een kleinere waarde voor de DIC indicator. Een vergelijkbaar maar wederom kleiner verschil in DIC wordt gevonden voor leerlingen afkomstig uit een examenklas kl. Dit betekent voor beide groepen leerlingen dat het model waarbij leerlingen zowel genest zijn in een vmbo-vestiging als in een mbo-instelling een betere weerspiegeling van de data geeft. Dus ook twee jaar na de examenklas kan worden gesteld dat zowel de vmbo-vestiging als de mbo-instelling hierin een rol spelen bij de succesvolle plaatsing.

Wanneer naar de grootte van de random effecten (varianties) wordt gekeken valt op dat de verhouding in de variantie tussen VO-vestigingen en mbo-instellingen anders is geworden. Twee jaar na de examenklas is er relatief meer variantie in plaatsing tussen VO-vestigingen (11,2%) dan tussen mbo-instellingen (6,7%) voor de basisleerlingen. Hier lijkt geen

eenvoudige verklaring voor te zijn. Voor de kader-leerlingen lijken verschillen in succesvolle plaatsing ongeveer even sterk gerelateerd te zijn aan de mbo-instelling (10,3%) als aan de vmbo-vestiging (9.3%).

6.3.5 *Vergelijking van extern rendement van VO-vestigingen voor de drie varianten van extern rendement*

In de voorgaande paragrafen zijn de resultaten van drie verschillende indicatoren voor het extern beschreven. Deze drie indicatoren verschillen in de mate waarin ze rekening houden met eventuele vertekeningen door plaatsingsbeleid van mbo-instellingen en de hiërarchische structuur van de data. Elk van deze indicatoren leidt mogelijk tot een andere schatting van het extern rendement van een VO-vestiging. Om de drie externe rendement indicatoren onderling te vergelijken zijn correlaties bepaald op basis van de geschatte externe rendementen op het niveau van de vestigingen. De correlaties zijn weergegeven in Tabel 6.8. Uit Tabel 6.8 kan zowel de vergelijking tussen de indicatoren voor het externe rendement worden gehaald, als een indicatie van de stabiliteit van het externe rendement bij de vergelijking van de indicator één en twee jaar na de examenklas. De cellen die licht grijs gekleurd zijn in de tabel geven de correlatie weer van het externe rendement één en twee jaar na de examenklas waarbij dezelfde variant van rendementsbepaling is gebruikt.

In de eerste plaats valt in Tabel 6.8 op dat er een vrij sterke samenhang is tussen de huidige indicator voor extern rendement en de indicator die is gebaseerd op een strikt hiërarchisch meerniveau model¹⁶ (bl: $r=0,77$; $p<0,001$; kl: $r=0,56$; $p<0,001$). Wederom zijn verschillen hier toe te schrijven aan verschillen in bepaling en verschillen in selectie van leerlingen. Ondanks de relatief sterke samenhang tussen deze twee indicatoren blijkt uit dat de beide indicatoren toch tot een aanzienlijk andere indeling van vestigingen komen in termen van bolletjes (Zie appendix bij Hoofdstuk 6).

In de tweede plaats valt op dat de samenhang met de indicator voor extern rendement gebaseerd op een kruisklassificatie model zeer gering is. Dit laatste geldt zowel voor de huidige indicator model (bl: $r=0,10$; $p<0,05$; kl: $r=-0,07$; ns) als voor het externe rendement gebaseerd op een strikt hiërarchisch meerniveau model (bl: $r =0,09$; $p<0,05$; kl: $r= -0,10$; $p<0,05$). Dit betekent dat de succesvolle plaatsing van leerlingen na één jaar voor een aanzienlijk deel toe te schrijven is aan de mbo-instelling waar de leerling naar vertrekt. Het betekent tevens dat een eenvoudige indicator gebaseerd op het percentage succesvol geplaatste leerlingen of een relatief simpel meerniveau model onvoldoende in staat zijn om de bijdrage van de VO-vestigingen op de succesvolle plaatsing van leerlingen in kaart te brengen. Vergelijkbare resultaten worden gevonden wanneer naar de samenhang tussen het

¹⁶ Hier wordt een leeg meerniveau model bedoeld waarin alleen een leerling- en vestigingsniveau wordt onderscheiden. Dit in tegenstelling tot het kruisklassificatie model.

kruisklassificatie model en de andere modellen wordt gekeken voor het externe rendement twee jaar na de examenklas.

Tabel 6.8 Correlaties tussen extern rendement uit huidige, meerniveau en kruisklassificatie model voor één en twee jaar na examenklas vmbo-bl of kl

Basisberoepsgerichte leerweg	Huidig na 1 jaar	Meerniveau na 1 jaar	Kruisklassificatie na 1 jaar	Huidig na 2 jaar	Meerniveau na 2 jaar
Meerniveau na 1 jaar	0,767*** 458				
Kruisklassificatie na 1 jaar	0,102* 458	0,093* 458			
Huidig na 2 jaar	0,631*** 506	0,451*** 458	0,024 458		
Meerniveau na 2 jaar	0,539*** 451	0,553*** 448	0,023 448	0,570*** 451	
Kruisklassificatie na 2 jaar	0,029 451	0,004 448	0,530*** 448	0,028 451	0,060 451
Kaderberoepsgerichte leerweg	Huidig na 1 jaar	Meerniveau na 1 jaar	Kruisklassificatie na 1 jaar	Huidig na 2 jaar	Meerniveau na 2 jaar
Meerniveau na 1 jaar	0,564*** 484				
Kruisklassificatie na 1 jaar	-0,070 484	-0,104* 484			
Huidig na 2 jaar	0,731*** 507	0,246*** 484	-0,013 484		
Meerniveau na 2 jaar	0,282*** 483	0,292*** 482	-0,054 482	0,309*** 483	
Kruisklassificatie na 2 jaar	-0,011 483	-0,018 482	0,275*** 482	0,045 483	-0,028 483

*p < 0,05; ** p < 0,01; *** p < 0,001

Een derde opvallende bevinding in de Tabel 6.8 is de relatieve stabiliteit van de indicatoren voor het externe rendement voor de basisberoepsgerichte leerweg (licht grijze cellen). Voor alle drie de indicatoren wordt een matige samenhang gezien tussen het externe rendement één jaar na de examenklas en het externe rendement twee jaar na de examenklas (huidig model:

$r=0,63$; $p<0,001$; meerniveau: $r=0,55$; $p<0,001$; kruisklassificatie: $r=0,53$; $p<0,001$). De sterkste samenhang wordt echter gevonden voor de huidige indicator. Het beeld met betrekking tot de stabiliteit van het externe rendement is anders wanneer naar de kaderberoepsgerichte leerweg wordt gekeken. Wederom laat het huidige model een sterke samenhang zien tussen het externe rendement één en twee jaar na de examenklas ($r=0,731$; $p<0,001$). De beide meerniveau modellen laten echter een minder grote stabiliteit zijn. Een mogelijke oorzaak voor de verschillen tussen de beide indicatoren kan liggen in het zeer grote aantal leerlingen in de kl-leerweg die succesvol worden geplaatst (op basis van de bestaande beslissingsregels). Een aanzienlijk deel van de VO-vestigingen behaald een extern rendement van 100%.

Naast een vergelijking tussen de onderliggende modellen die gebruikt zijn voor het bepalen van het externe rendement kunnen de modellen ook inzicht geven in de mate waarin verschillende afdelingen binnen één vestiging een vergelijkbaar extern rendement behalen. Hiertoe zijn correlaties berekend tussen het externe rendement van vestigingen waarvoor zowel een schatting van het externe rendement beschikbaar is voor zowel bl als kl. Deze correlaties worden gepresenteerd in Tabel 6.9.

Tabel 6.9 Correlaties tussen extern rendement van vmbo-bl en kl afdelingen binnen VO-vestigingen

	Basis- of Kaderberoepsgerichte leerweg één jaar na examenklas	Basis- of Kaderberoepsgerichte leerweg twee jaar na examenklas
Huidig	0,448*** (N=493)	0,357*** (N=493)
Multilevel	0,241*** (N=445)	0,237*** (N=441)
Kruisklassificatie	0,196*** (N=445)	0,169*** (N=441)

* $p<.05$; ** $p<.01$; *** $p<.001$

Alle correlaties, zowel één als twee jaar na de examenklas zijn positief. Dit geeft een indicatie dat er een trend is dat een hoog extern rendement van een bl-afdeling in een vestiging samengaat met een hoog extern rendement van de kl-afdeling binnen dezelfde vestiging. De gevonden correlaties zijn niet heel erg hoog wat betekent dat er ook vele vestigingen zijn waarbij de bovenstaande trend niet opgaat. Een ander duidelijk patroon is dat de samenhang tussen het externe rendement van bl en kl-afdelingen binnen dezelfde vestiging afneemt naarmate gekeken wordt naar succesvolle plaatsing die in de tijd verder af staat van de examenklas. De correlatie twee jaar na de examenklas zijn minder sterk dan de correlaties één jaar na de examenklas. Tot slot is gevonden dat bij de huidige indicator de sterkste correlaties tussen bl en kl-afdelingen binnen een vestiging gevonden worden.

6.3.6 *Vergelijking van huidige indicator met meerniveau modellen op basis van percentiel scores*

Op verzoek van de Inspectie van het Onderwijs is ook een tweede wijze van vergelijken van de modellen toegepast. Met deze vergelijkingen wordt getracht een beeld te schetsen van de validiteit van de huidige indicator. Hierbij is het uitgangspunt dat het kruisklassificatie model het conceptueel sterkste model is, waartegen de huidige indicator voor extern rendement kan worden afgezet. Naast het maken van een vergelijking op basis van correlaties tussen de drie varianten van het externe rendement is ook getracht om een vergelijking te maken op basis van de percentiel scores. Hiervoor is gebruik gemaakt van de norm die gebruikt wordt bij het huidige model (ten minste 90% succesvolle plaatsingen). Aan de hand daarvan is vastgesteld hoeveel VO-vestigingen onder deze norm presteren (zie ook Tabel 6.3 en 6.5 bij de norm van 90%). Vervolgens zijn de gevonden percentages gebruikt als cesuur bij de meerniveau en kruisklassificatie modellen.

In Tabel 6.10 worden de resultaten weergegeven van de vergelijking tussen de indicatoren voor het externe rendement één jaar na de examenklas. In deze vergelijking zijn alleen de scholen opgenomen waarvan voor tenminste 10 leerlingen gegevens beschikbaar zijn.

Tabel 6.10 Vergelijking tussen indicatoren voor extern rendement door leggen van cesuur van huidige model op de meerniveau modellen één jaar na de examenklas

		Multilevel model (N=458)		Kruisklassificatie model (N=458)	
		Boven de norm	Onder de norm	Boven de norm	Onder de norm
Huidige indicator	Boven de norm	158 (34,5%)	50 (10,9%)	99 (21,6%)	108 (23,6%)
	Onder de norm	49 (10,7%)	201 (43,9%)	109 (23,8%)	142 (31,0%)
<i>Kaderberoepsgerichte leerweg</i>					
		Multilevel model(N=484)		Kruisklassificatie model (N=484)	
		Boven de norm	Onder de norm	Boven de norm	Onder de norm
Huidige indicator	Boven de norm	385 (79,5%)	34 (7,0%)	364 (75,2%)	55 (11,4%)
	Onder de norm	35 (7,2)	30 (6,2%)	56 (11,6)	9 (1,9%)

Bij een norm van tenminste 90% succesvolle plaatsingen bij de huidige indicator scoort 45,1% van de basisberoepsgerichte leerweg afdelingen van VO-vestigingen boven de norm en 54,9% onder de norm (zie ook Tabel 6.3). In Tabel 6.10 zijn tussen haakjes de percentages vestigingen in de cellen weergegeven. Wanneer de norm van 90% ook bij de andere modellen wordt gebruikt, wordt een relatief grote mate van overeenstemming tussen het huidige model

en het meerniveau model gevonden voor vmbo-basis. In totaal presteren 78,4% van de scholen op beide indicatoren vergelijkbaar, namelijk in beide gevallen boven de norm of in beide gevallen onder de norm. De gevonden samenhang voor de indeling boven of onder de norm tussen het huidige model en het meerniveau model is vrij sterk ($\kappa=0,55$; $p<0,001$). Voor de vergelijking tussen het huidige model en het kruisklassificatie model wordt een minder grote overeenstemming gevonden. In totaal presteren 52,6% van de vestigingen op beide indicatoren vergelijkbaar, namelijk in beide gevallen boven de norm of in beide gevallen onder de norm (Tabel 6.10). Gegeven de lagere correlatie die eerder werd gevonden is deze bevinding niet verwonderlijk. De gevonden samenhang voor de indeling boven of onder de norm tussen het huidige model en het kruisklassificatie model is dan ook veel zwakker ($\kappa=0,04$; $p=0,347$).

Bij het hanteren van een norm van tenminste 90% succesvolle plaatsingen één jaar na de examenklas van de kaderberoepsgerichte leerweg wordt gevonden dat 87% van de vestigingen boven en slechts 13% onder de norm scoort (zie Tabel 6.3). Uit Tabel 6.10 is af te leiden dat bij een vergelijking tussen het huidige en meerniveau model is er overeenstemming in de indeling boven of onder de norm voor 86% van de VO-vestigingen. Hoewel het percentage overeenstemming hoog is, is de gevonden samenhang relatief laag ($\kappa=0,115$; $p<0,001$). Dit heeft te maken met het zeer grote aantal vestigingen wat boven de norm presteert in het huidige model. Bij een vergelijking tussen het huidige model en het kruisklassificatie model is er overeenstemming in de indeling boven of onder de norm voor 77% van de VO-vestigingen. Het leggen van de cesuur uit het huidige model voor extern rendement op de resultaten van het kruisklassificatie model leidt dus tot een aanzienlijke overeenstemming. Echter, ook hier is de gevonden samenhang (wanneer gecorrigeerd wordt voor toeval) zeer beperkt ($\kappa=0,002$; $p=0,873$). Deze lage samenhang is gegeven de eerder gevonden correlaties niet verwonderlijk. De grote mate van overeenstemming is hier het gevolg van het relatief grote aantal scholen dat boven de norm presteert.

In Tabel 6.11 worden de resultaten weergegeven van de vergelijking tussen de indicatoren voor het externe rendement twee jaar na de examenklas. De resultaten van deze vergelijking van de modellen zijn erg vergelijkbaar met de resultaten één jaar na de examenklas. In alle gevallen is er sprake van een relatief grote overeenstemming tussen onder en boven de norm. Zowel voor de basis- als kaderberoepsgerichte leerweg is de overeenstemming tussen het huidige en het meerniveau model groter dan de overeenstemming tussen het huidige en het kruisklassificatie model. In alle gevallen worden slechts beperkte samenhangen gevonden ondanks het hoge percentage overeenstemming. Dit duidt erop dat de gevonden overeenstemming grotendeels te wijten zijn aan de gebruikte cesuur die bepaald is aan de hand van het percentage scholen dat volgens de huidige indicator onder de norm scoorde.

Tabel 6.11 Vergelijking tussen indicatoren voor extern rendement door leggen van cesuur van huidige model op de meerniveau modellen twee jaar na de examenklas

		Multilevel model (N=451)		Kruisklassificatie model (N=451)	
		Boven de norm	Onder de norm	Boven de norm	Onder de norm
Huidige indicator	Boven de norm	67 (14,9%)	71 (15,7%)	48 (10,6%)	90 (20,0%)
	Onder de norm	68 (15,1%)	245 (54,3%)	87 (19,3%)	226 (50,1%)
Kappa		0,197***		0,051	

		Multilevel model (N=483)		Kruisklassificatie model (N=483)	
		Boven de norm	Onder de norm	Boven de norm	Onder de norm
Huidige indicator	Boven de norm	257 (53,2%)	85 (17,6%)	247 (51,1%)	95 (19,7%)
	Onder de norm	84 (17,4%)	57 (11,8%)	94 (19,5%)	47 (9,7%)
Kappa		0,110 ***		0,039	

*** p<0,001

6.3.7 Effecten van gebruik van controle variabelen op extern rendement

In de huidige bepaling van het externe rendement worden geen achtergrondkenmerken meegenomen om eventueel te controleren voor de invloed van de achtergrond van leerlingen op de plaatsing in een hogere leerweg of het mbo. In de methodologie van de bestaande indicator is het lastig om na te gaan in hoeverre de achtergrond van leerlingen gerelateerd is aan een succesvolle plaatsing. In een meerniveau model kan relatief eenvoudig worden onderzocht wat de unieke bijdrage is van achtergrondkenmerken van leerlingen op de succesvolle plaatsing van leerlingen. Om dit te onderzoeken is een vergelijkbaar meerniveau model geanalyseerd waarbij leerwegondersteuning (lwoo), etnische achtergrond (cumi) en armoedeprobleemcumulatiegebied (apcg) zijn meegenomen als mogelijke voorspellers. De analyse is uitgevoerd voor het externe rendement één jaar na de examenklas voor de afdeling basisberoepsgerichte leerweg. Alle leerlingen die in 2010 ingeschreven waren in een vestiging waarbij voor tenminste 10 leerlingen gegevens bekend zijn voor het externe rendement, zijn in de analyse opgenomen. Dit zijn dus ook leerlingen die voortijdig schoolverlaten of doorstromen naar een hogere leerweg. Dit betekent dat het een andere selectie van leerlingen betreft dan de voorgaande meerniveau analyses. Deze ruimere selectie is gemaakt om een betere vergelijkbaarheid te realiseren met de huidige indicator voor het externe rendement. De resultaten van deze analyse zijn gepresenteerd in Tabel 6.12.

Tabel 6.12 Resultaten van voorspellen van een succesvolle plaatsing één jaar na examenklas op basis van achtergrondkenmerken van vmbo-bl leerlingen

	Leeg meerniveau model		Meerniveau model met achtergrondkenmerken	
	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed Part:</i>				
Constante	2,06	0,04*	2,31	0,06*
<i>Cumi: #</i>				
Niet Westers allochtoon			-0,003	0,06
Westers allochtoon			-0,16	0,09
Onbekend			-3,28	0,36*
<i>Apcg:</i>				
Laag inkomen			-0,04	0,19
Niet W-allochtoon			-0,14	0,07
Uitkering			-0,26	0,11*
Laag inkomen + uitkering			-0,50	0,12*
Laag inkomen + niet W-allochtoon			-0,15	0,21
Uitkering + niet W-allochtoon			-0,30	0,10*
Cumulatiegebied			-0,50	0,06*
Lwoo: Ja			-0,11	0,05
<i>Random Part:</i>				
Tussen VO-vestiging verschillen	0,40	0,04	0,32	0,04
<i>Model fit (DIC)</i>		16407,2	16280,8	
Aantal VO-vestigingen	463		463	
Aantal leerlingen	21.874		21.874	

*p<0,05

Het model met achtergrondkenmerken past als geheel beter bij de data dan het lege model. Dit wordt duidelijk uit een kleinere waarde voor de DIC. Het model waarin achtergrondkenmerken als mogelijke voorspellers van een succesvolle plaatsing van leerlingen zijn opgenomen resulteert in een aantal unieke significantie relaties. Een eerste unieke relatie betreft leerlingen die een indicatie hebben voor leerwegondersteuning (lwoo). Gecontroleerd voor alle andere achtergrondkenmerken van de leerlingen is gevonden dat lwoo-leerlingen een kleinere kans hebben om succesvol te worden geplaatst dan leerlingen zonder lwoo (referentiegroep). Leerlingen met een onbekende etnische achtergrond hebben een kleinere kans dan autochtone leerlingen om succesvol te worden geplaatst. Daarnaast zijn er verschillen tussen leerlingen woonachtig in typen buurten met overschrijdingen (apcg) waar

de kans op een succesvolle plaatsing kleiner is in vergelijking met leerlingen woonachtig in wijken waar geen overschrijding is (referentie groep). Een grotere kans op een niet succesvolle plaatsing is gevonden voor leerlingen woonachtig in een cumulatiegebied, in een wijk met overschrijdingen van uitkering, in een wijk met uitkering en laaginkomen en tenslotte in een wijk met een overschrijding van uitkering en niet Westers allochtonen.

Uit Tabel 6.12 blijkt dat een aantal achtergrondkenmerken van vmbo-bl-leerlingen samenhangen met de succesvolle plaatsing een jaar na de examenklas. Aangezien leerlingpopulaties van scholen verschillen kan het daarom verantwoordt worden om ook voor het externe rendement achtergrondkenmerken mee te nemen in de bepaling. Dit leidt vervolgens tot de vraag in hoeverre het meenemen van deze achtergrondkenmerken daadwerkelijk leidt tot veranderingen van de schattingen van het externe rendement van VO-vestigingen. De correlatie tussen het externe rendement op basis van een leeg meerniveau model en het externe rendement op basis van een meerniveau model met controle variabelen is hier 0,97. Dit impliceert dat het meenemen van de achtergrondkenmerken van de vmbo-bl-leerlingen in de bepaling van het externe rendement na één jaar beperkt is.

7 Conclusie en discussie

Op basis van de door Inspectie van het Onderwijs geformuleerde doelen 1 en 2 is het onderzoek opgesplitst in vier deelprojecten. Het eerste deelproject heeft betrekking op het valideren van het beoogde opbrengstenmodel met VOCL'99 data. Het tweede deelproject richtte zich op het vergelijken van een prospectieve berekeningswijze van het rendement in de onderbouw VO met de huidige retrospectieve berekeningswijze. In het derde deelproject gaat over de toegevoegde waardebeoordeling voor een specifieke groep leerlingen namelijk vmbo-tl-leerlingen die doorstromen naar het havo. Van hen is een meting van het beginniveau (eindexamen vmbo-tl) en eindniveau (eindexamen havo) beschikbaar. Het vierde deelproject heeft een verkennend karakter. De mogelijkheden van het bepalen van het externe rendement van vmbo-bl en kl leerlingen die hun opleiding voortzetten in een hogere leerweg van vmbo of op het mbo zijn onderzocht.

Hieronder wordt bij elk deelproject eerst een overzicht van de vraagstellingen gegeven, waarna een bespreking van de conclusies volgt en de resultaten worden bediscussieerd. Ook komen in dit hoofdstuk de aanvullende discussievragen over onderbouwendement van deelproject 2, de algemene implementatie- en evaluatievragen van de opdrachtgever en het derde doel van dit onderzoek aan de orde.

Deelproject I: Validering van beoogd opbrengstenmodel

Leidt het toepassen van modellen die verschillen in de manier en mate van correctie voor het instroomniveau van leerlingen tot verschillen in de opbrengstindicator rendement bovenbouw van een VO-vestiging?

Leidt het toepassen van deze modellen tot verschillen in de opbrengstindicator gemiddeld centraal examencijfer van alle vakken van een VO-vestiging?

In hoeverre vormt de indicator gemiddeld centraal examencijfer voor alle vakken een goede afspiegeling van de onderliggende centraal examencijfers per vak of vakkencluster?

Is de interne consistentie van de indicator gemiddeld centraal examencijfer voor alle vakken verschillend voor de toegepaste modellen?

In deelproject 1 zijn vier modellen met elkaar vergeleken. Hieronder volgt een overzicht.

Lege model:	meerniveau model zonder correctiefactoren
Huidige model:	<p>lineaire regressie model met als correctiefactoren:</p> <ul style="list-style-type: none"> - percentage instromende leerlingen in leerjaar 3 - percentage leerlingen uit armoedeprobleemcumulatiegebied - percentage lwoo-leerlingen <p>Deze correctiefactoren zijn proxies voor het beginniveau van leerlingen</p>
Beoogde model:	<p>meerniveau model met als correctiefactoren:</p> <ul style="list-style-type: none"> - percentage instromende leerlingen in leerjaar 3 - leerling uit armoedeprobleemcumulatiegebied - leerling heeft lwoo <p>Deze correctiefactoren zijn proxies voor het beginniveau van leerlingen</p>
Optimale model:	<p>meerniveau model met als correctiefactoren:</p> <ul style="list-style-type: none"> - Cito-eindtoetscore van een leerling; is een beginniveau-meting. - SES van een leerling - etniciteit van een leerling - sekse van een leerling <p>De laatste drie factoren zijn niet-schoolse kenmerken die de leerprestaties kunnen beïnvloeden.</p>

In deelproject I zijn het huidige en beoogde opbrengstenmodel voor de opbrengstindicatoren rendement bovenbouw en gemiddeld centraal examencijfer voor alle vakken geplaatst in het spectrum van enerzijds een meerniveau opbrengstenmodel zonder enige vorm van correctie voor beginniveau van leerlingen (lege model) en anderzijds een meerniveau opbrengstenmodel waarin het instroomniveau van individuele leerlingen zo goed mogelijk wordt verdisconteerd (optimale model). Het huidige en beoogde model verschillen van elkaar in de manier waarop dezelfde set van correctiefactoren wordt opgenomen. In het huidige model wordt het aanvangsniveau van leerlingen benaderd door armoedeprobleemcumulatiegebied (apcg), leerwegondersteuning (lwoo) en percentage instroom in derde leerjaar op vestigingsniveau in een lineair regressie model op te nemen. Bij het beoogde model zijn apcg en lwoo op leerlingniveau en percentage instroom in derde leerjaar op vestigingsniveau in een meerniveau model opgenomen.

Voor de bepaling van het bovenbouwrendement lijken de resultaten van de vier modellen sterk op elkaar. De correlaties tussen schoolresiduen – de mate waarin een VO-vestiging beter of slechter presteert dan gemiddeld - van de vier modellen variëren tussen 0,79 en 0,99. Het lijkt feitelijk weinig uit te maken welke model gehanteerd wordt. De enige uitzondering wordt wellicht gevormd door het huidige model. De berekeningswijze van het huidige model is methodologisch minder sterk (geen meerniveau model) en correlaties van de schoolresiduen van dit model met de andere modellen zijn het laagst (0,79-0,83). Ook is de gevonden minimumwaarde voor het bovenbouwrendement van een vestiging bij het huidige model 60%,

hetgeen beduidend hoger is dan voor de andere modellen met minimumwaarden die variëren van 45% tot 56%.

Voor het gemiddelde centraal examencijfer voor alle vakken zijn wel duidelijke verschillen tussen de modellen gevonden. In het spectrum liggen het lege en optimale model verder uiteen dan bij het bovenbouwrendement. Op grond van de gevonden correlaties tussen de schoolresiduen kan gesteld worden dat het huidige model sterk lijkt op het lege model ($r=0,86$) en zeker niet op het optimale model ($r=0,01$). Het beoogde model neemt een positie in tussen het huidige en optimale model, maar lijkt toch aanzienlijk meer op het huidige model ($r=0,48$) en lege model ($r=0,46$), dan op het optimale model ($r=-0,10$).

De sterke samenhang tussen het huidige model en lege model voor beide opbrengstindicatoren is in lijn met het lage percentage verklaarde variantie van het huidige model zoals dit is gepresenteerd in de technische toelichting op de Opbrengstenkaart 2012 (Inspectie van het Onderwijs, 2012). De verklaringskracht van de vestigingsgemiddelde correctiefactoren (apcg, lwoo en instroom 3^e leerjaar) in de regressieanalyses is maar zeer bescheiden. Dit houdt dan automatisch in dat het gecorrigeerde bovenbouwrendement van een VO-vestiging slechts marginaal zal afwijken van de ongecorrigeerde bovenbouwrendement. Hetzelfde geldt voor het gemiddelde centraal examencijfer voor alle vakken.

Niet alleen de samenhang tussen schoolresiduen is van belang om de keuze voor een model goed te kunnen onderbouwen. Ook de betrouwbaarheid van de opbrengstindicator speelt een rol. Daarom is voor het gemiddelde centraal examencijfer voor alle vakken nagegaan of dit cijfer een goede afspiegeling vormt van de onderliggende centraal examencijfers per vak of vakkencluster, en of dit afhankelijk is van het toegepaste model. De opvallende bevinding is, dat alleen bij het optimale model de cijfers op een vijftal (clusters van) vakken onderling hoog correleren ($r=0,63$). Deze onderlinge samenhang is voor het huidige, beoogde en lege model veel lager ($r=0,25-0,27$). De hoge interne consistentie bij het optimale model geeft aan, dat met dit model opbrengstverschillen tussen vestigingen voor het gemiddelde centraal examencijfer betrouwbaar gemeten zouden kunnen worden. In het verlengde daarvan moet ook geconcludeerd worden, dat zowel het huidige als beoogde model vrij ver afstaan van een model waarmee betrouwbaar en zo nauwkeurig mogelijk de toegevoegde waarde van een VO-vestiging voor het gemiddelde centraal examencijfer kan worden bepaald.

Op basis van bovenstaande kan geconcludeerd worden dat het beoogde model een betere keuze is dan het huidige model als methode voor de Inspectie van het Onderwijs om de opbrengsten van VO-vestigingen onderling te vergelijken. Dit geldt zowel voor het bovenbouwrendement als het gemiddelde centraal examencijfer voor alle vakken. Er zijn nog wel twee kanttekeningen bij te plaatsen. Bij het bovenbouwrendement lijkt het beoogde model wel het sterkst op het optimale model, maar de verschillen tussen de vier modellen zijn maar klein. Voor het gemiddelde centraal examencijfer voor alle vakken zijn wel grote verschillen

tussen de modellen aangetoond. Bij de keuze voor het beoogde model moet men zich goed realiseren, dat met dit model hooguit een matige indicatie van de toegevoegde waarde van een vestiging in het gemiddeld eindexamencijfer wordt verkregen.

Het streven naar toepassing van het optimale model in de toezichtpraktijk lijkt vooralsnog niet aanbevelingswaardig voor het bovenbouwrendement. De geringe winst in nauwkeurigheid van de toegevoegde waarde indicator weegt hier niet op tegen de hoge kosten en inspanningen van de extra dataverzameling. Als men echter zo goed mogelijk de toegevoegde waarde van een school in het gemiddelde centraal examencijfer voor alle vakken wil bepalen, dan kan men niet om het optimale model heen. Dan moeten belangrijke indicatoren voor individuele leerprestaties bij aanvang wel meegenomen worden in het model. Gebleken is dat het potentiële leervermogen (IQ) buiten beschouwing kan worden gelaten. Wat dit betreft kan met de score op Cito-eindtoets basisonderwijs worden volstaan.

Deelproject II: Rendement onderbouw VO

In welke mate wijkt de huidige retrospectieve bepaling van het onderbouwrendement af van een prospectieve bepaling van het onderbouwrendement met individuele data?

In hoeverre hebben de toeleverende PO-vestigingen invloed op de waarde van het onderbouwrendement van een VO-vestiging?

Welke rol spelen over- en onderadvisering hierbij?

Is het wenselijk om VO-vestigingen met lwoo-leerlingen te compenseren voor doubleren?

Hoe kan het beste omgegaan worden met de diverse combinaties van schoolsoorten in de onderbouw van VO-vestigingen?

Hoe kan het beste omgegaan worden met instromende leerlingen in leerjaar 1 en 2? Aan welke VO-vestiging moeten deze leerlingen worden toegekend?

In welke mate is het advies van leerlingen afhankelijk van het onderwijsaanbod op de ontvangende VO-vestigingen?

Vergelijking retrospectieve met prospectieve onderbouwrendement

In dit deelproject is op twee manieren het onderbouwrendement van een vestiging berekend met individuele data van de meeste recente Bron-bestanden. De eerste variant volgt zo veel mogelijk de huidige werkwijze van de Inspectie. Deze is retrospectief van aard met data op groepsniveau (schoolsoort per vestiging). De Inspectie vraagt jaarlijks van leerlingen die overgegaan zijn naar leerjaar 3 van het VO de adviesverdeling en het aantal zittenblijvers in leerjaar 1 en 2 op bij de VO-vestigingen en berekent hiermee het onderbouwrendement. De

eerste variant onderbouwrendement die in dit project is doorgerekend is ook retrospectief van aard, maar dan met data van individuele leerlingen die in schooljaar 2012-2013 voor de eerste keer in leerjaar 3 zitten. Een belangrijke beperking hierbij is dat een deel van deze leerlingen is blijven zitten in de afgelopen twee jaar en dat van hen (nu nog) geen adviezen beschikbaar zijn. Het retrospectieve onderbouwrendement heeft in dit project daarom alleen betrekking op (vrijwel allemaal) onvertraagde leerlingen.

De tweede variant van het onderbouwrendement is een werkwijze die de Inspectie overweegt om in de toekomst te gaan gebruiken: prospectief met wederom individuele data. Voor leerlingen die drie jaar geleden (schooljaar 2010-2011) voor de eerste keer in leerjaar 1 VO zaten is nagegaan hoe efficiënt ze door het VO gestroomd zijn. Zonder doublures zouden ze in schooljaar 2012-2013 in VO-3 moeten zitten. Koppeling van het advies per leerling is van deze groep redelijk succesvol (77%). Het aantal doublures en de schoolsoort in leerjaar 3 is voor vrijwel alle leerlingen goed vast te stellen op basis van de Bron-bestanden. Uiteindelijk is voor 71% van de deze VO-1 leerlingen het individuele prospectief rendement onderbouw berekend.

Het onderbouwrendement van een VO-vestiging is verkregen door steeds het gewogen gemiddelde te nemen van rendement onderbouw per schoolsoort (variant 1) of van het individuele rendement onderbouw (variant 2). Daarbij zijn de regels gehanteerd dat tenminste 60% van de leerlingen een bruikbaar advies/onderbouwrendement-bepaling heeft en het aantal leerlingen met een onderbouwrendement-bepaling per vestiging minimaal 10 is.

Met de retrospectieve werkwijze (variant 1) is het mogelijk om voor 70% van de vestigingen het onderbouwrendement voldoende betrouwbaar te berekenen; het gemiddelde onderbouwrendement van alle VO-vestigingen is 96,3. Met de prospectieve werkwijze met individuele data (variant 2) kan voor een hoger percentage van de vestigingen (75%) het onderbouwrendement berekend worden. Het overall gemiddelde prospectieve onderbouwrendement is met 96,8 iets hoger dan de retrospectieve variant. De samenhang tussen de twee varianten van rendement onderbouw is vrij hoog ($r=0,79$). De uitbijters blijken vestigingen te zijn met relatief een gering aantal leerlingen met een onderbouwrendements bepaling.

Vergelijking tussen de twee varianten per schoolsoort leert dat er in geringe mate sprake is van systematische verschillen. Dit is met name het geval op het vmbo basisberoepsgerichte leerweg (bl) en kaderberoepsgerichte leerweg (kl). Het verschil tussen retrospectief en prospectief onderbouwrendement is opvallend groot (-6,0). Het retrospectieve onderbouwrendement onderbouw is hier lager dan het prospectieve (93,4 versus 99,6). Verschillen in omvang en type van in- en uitstromende leerlingen liggen hier waarschijnlijk aan ten grondslag. Op andere schoolsoorten zijn de verschillen tussen de twee varianten kleiner. Op schoolsoorten met vwo als hoogste schoolsoort is het verschil +2,0, op de

schoolsoort vmbo-tl is het verschil gemiddeld +0,1 en op de schoolsoort vmbo-breed met havo als hoogste schoolsoort is het verschil -3,1.

Het prospectieve onderbouwrendement heeft in dit deelproject betrekking op zowel onvertraagde als vertraagde leerlingen en geeft daarmee een volledig beeld van de schoolloopbanen van leerlingen dan het retrospectieve onderbouwrendement. Alle leerlingen die drie jaar in het VO verblijven, ongeacht welke leerjaar of schoolsoort, maken in principe deel uit van de prospectieve berekeningswijze. Zou van doublerende 3^e klassers ook het advies bekend zijn – hetgeen slechts een kwestie van tijd is, want ook deze gegevens komen beschikbaar – dan zal naar verwachting het retrospectieve onderbouwrendement nog wat lager uitvallen. Hierdoor zal het verschil tussen beide varianten groter worden.

Geconcludeerd kan worden dat een prospectieve berekening van het onderbouwrendement te verkiezen is boven een retrospectieve berekening. Hiervoor zijn een aantal redenen te noemen. Ten eerste leidt een prospectieve berekening tot een hoger percentage vestigingen waarvoor dit met voldoende zekerheid bepaald kan worden. Ten tweede voldoet een prospectieve benadering beter aan de intuïtieve betekening van de term rendement onderbouw: hoe efficiënt doorlopen leerlingen die starten in het VO de onderbouw. Ten derde is gebleken dat er sprake is van systematische verschillen tussen beide varianten. Dit is niet voor alle schoolsoorten even sterk. Het speelt met name op het vmbo basis- en kaderberoepsgerichte leerweg. Er zijn duidelijke indicaties gevonden dat een retrospectieve berekeningswijze van het onderbouwrendement leidt tot een negatieve vertekening. Dit komt op het vmbo-bl/kl vooral door de relatief grote instroom van vmbo-tl-leerlingen in vergelijking met de instroom vanuit het praktijkonderwijs. Het retrospectieve onderbouwrendement van een vmbo-bl/kl-vestiging krijgt hierdoor een lagere waarde. Dit zou leiden tot het ongewenst effect dat vmbo-bl/kl-vestigingen worden ‘afgerekend’ op het feit dat ze afstromende vmbo-tl-leerlingen een plek geven op hun school.

Invloed basisschool

In het huidige opbrengstenmodel VO wordt geen rekening gehouden met de basisschool waar de leerlingen vandaan komen. Het is heel goed mogelijk dat het onderbouwrendement beïnvloed wordt door de toeleverende basisscholen. Lange termijn effecten van basisscholen zullen voornamelijk in de eerste jaren van het VO een rol van betekenis kunnen spelen. Verder is het denkbaar dat VO-scholen ‘last’ of ‘profijt’ hebben van basisscholen die structureel respectievelijk te hoge of te lage adviezen afgeven. Het eerste aspect is nader onderzocht door een leeg meerniveau model met alleen een leerling- en VO-vestigingsniveau te vergelijken met een kruisklassificatie model met leerling-, VO- en PO-vestigingsniveau. Van de totale variantie in prospectief onderbouwrendement is 91% toe te schrijven aan verschillen tussen leerlingen, 4% aan verschillen tussen VO-vestigingen en 5% aan verschillen tussen toeleverende PO-vestigingen. Daarmee is het aandeel van de PO-vestigingen van even groot belang als het aandeel van de VO-vestigingen.

Het tweede aspect - vertekening van het onderbouwrendement door de invloed van PO-vestigingen - is nader onderzocht door de gemiddelde Cito-eindtoetsscore en de mate van overadvisering van de toeleverende PO-vestigingen mee te nemen als correctiefactoren in een kruisklassificatie model. VO-vestigingen die relatief veel leerlingen ontvangen van PO-vestigingen met een gemiddeld hoog prestatieniveau in groep 8 behalen in het algemeen een lager onderbouwrendement. Ook geldt dat VO-vestigingen die relatief veel leerlingen ontvangen van PO-scholen die structureel te hoge adviezen afgeven in het algemeen een lager onderbouwrendement behalen. Deze VO-vestigingen worden dus benadeeld door dergelijke basisscholen. Aan de andere kant zijn er ook VO-vestigingen die profijt hebben van toeleverende PO-vestigingen die te lage adviezen afgeven. Deze VO-vestigingen kunnen relatief makkelijk een hoog onderbouwrendement behalen.

Om meer zicht te krijgen op welke VO-vestigingen last en profijt hebben van de toeleverende PO-vestigingen zijn deze nader gekarakteriseerd. Benadeelde VO-vestigingen hebben in ieder geval een vwo-afdeling en er komen vrij veel doublures voor. Profiterende VO-vestigingen blijken steeds een vmbo afdeling te hebben; het gaat vooral om vmbo-bl/kl en vmbo-breedhavo-vestigingen. Opvallend is hier verder dat er niet of nauwelijks doublures voorkomen en er relatief veel lwoo-leerlingen zijn. Omdat het hier vooral om vwo of lage vmbo afdelingen lijkt te gaan kan het zijn dat hier bodem- of plafondeffecten van invloed zijn. Op categorale vwo-vestigingen en categorale vmbo-bl/kl-vestigingen zou deze vorm van vertekening de grootste effecten moeten hebben. Dit is onderzocht door een gestratificeerde analyse per schoolsoort uit te voeren. Gevonden is echter dat ongeveer 78% van de categorale vwo-vestigingen en 82% van categorale vmbo-bl/kl-vestigingen behoort tot de categorie 'neutraal'; het maak op deze vestigingen niet of nauwelijks uit of de invloed van de PO-vestigingen wordt verdisconteerd. Geconcludeerd is dat bodem- en plafondeffecten wel spelen bij de berekening van het onderbouwrendement, zeker op categorale vwo- en vmbo-bl/kl-vestigingen, maar dat dit niet of nauwelijks doorwerkt in de invloed die een PO-vestiging kan hebben op onderbouwrendement van een VO-vestiging.

Normering met vergelijkingsgroepen

Bij de normering van het onderbouwrendement wordt het gemiddelde onderbouwrendement van een VO-vestiging vergeleken met het gemiddelde in de vergelijkingsgroep waartoe de betreffende VO-vestiging behoort. De Inspectie heeft daarvoor zeven verschillende vergelijkingsgroepen gemaakt. Uit de analyses blijkt dat het aantal vestigingen per vergelijkingsgroep sterk varieert; van 38 voor categorale vwo-vestigingen tot 208 voor vestigingen die de schoolsoorten vmbo-gl/tl tot en met vwo in het eerste leerjaar aanbieden. Dit heeft tot gevolg dat de betrouwbaarheid van de normering van een VO-vestiging nogal wisselend is. Voor brede VO-vestigingen kan veel nauwkeurig de relatieve positie worden vastgesteld dan voor categorale vwo-vestigingen. Een ander punt van zorg is dat een klein aantal VO-vestigingen niet aan een vergelijkingsgroep kan worden toegewezen. Het betreffen

nieuwe VO-vestigingen. Omdat voor de berekeningen data uit het meest recente schooljaar gekoppeld moet worden aan data uit een schooljaar eerder, ontstaan er koppelingsproblemen. Dit is echter een probleem van alle tijden. Als er nieuwe VO-vestigingen worden opgericht zal het altijd enige tijd duren voordat ze meegenomen kunnen worden in een evaluatieonderzoek.

Gelet op de bodem- en plafondeffecten die voortdurend een rol spelen bij de bepaling van het onderbouwrendement en de invloed die toeleverende PO-scholen hebben op juist de hoogste en laagste schoolsoorten is het aanbevelingswaardig om vergelijksgroepen te blijven hanteren bij de normering van het onderbouwrendement. Overwogen dient te worden of het niet minder dan zeven groepen zouden moeten zijn. Daarbij is het van belang dat het aantal VO-vestigingen per vergelijkingsgroep ongeveer even groot is.

Instroom in leerjaar 1 en 2

Bij de berekeningen van het prospectieve onderbouwrendement wordt geen rekening gehouden met leerlingen die in de loop van leerjaar 1 of 2 instromen. Uitstromende leerlingen worden wel meegenomen. Het is in analyse-technisch mogelijk om tegelijkertijd de invloed van de PO-school en het effect van wisseling van VO-vestiging te verdisconteren in de berekeningen. Dit zou kunnen met een meerniveau multiple membership kruisklassificatie model. De verblijfsduur op een VO-vestiging wordt daarbij als gewicht meegenomen.

In het proefschrift van Timmermans (2013) is dit type meerniveau analyse toegepast op VOCL'99 data van havo-leerlingen. Het doel was om de invloed van complexe schooleffecten (de invloed van PO-school én de invloed van wisseling van VO-school) op de eindexamencijfers te ontvlechten. Het percentage leerlingen dat meer dan één VO-school bezocht sinds de start in leerjaar 1 VO was 6%. Gevonden is dat de schattingen voor de varianties en de coëfficiënten van covariaten nauwelijks veranderen, als ook de invloed van instromende en uitstromende leerlingen werd meegenomen in de modellen.

Voor het huidige project is het aantal leerlingen dat wisselt van VO-vestiging gedurende de eerste twee leerjaar gelijk aan 6,7%. Op basis van dit lage percentage is te verwachten dat de geschatte waarden en schoolresiduen van de modellen nauwelijks zullen veranderen als meerniveau multiple membership kruisklassificatie model worden toegepast. In combinatie met de toename van de complexiteit van de modellen lijkt het daarom niet zinvol om naast de invloed van de PO-school ook nog de VO-vestigingswisselingen mee te nemen bij de bepaling van het prospectieve onderbouwrendement. Het is relevanter om te focussen op de kwaliteit van de data in de Bron-bestanden. Door het ontbreken van voldoende (goede) adviezen van leerlingen kon namelijk voor 25% van de VO-vestigingen geen prospectieve onderbouwrendement worden bepaald.

Compensatie voor leerlingen met leerwegondersteuning

Bij de berekening van het rendement onderbouw worden VO-vestigingen met lwoo-leerlingen gecompenseerd voor de veronderstelde grotere kans op doubleren van deze leerlingen. In dit

deelproject is onderzocht of lwoo-leerlingen daadwerkelijk vaker blijven zitten dan niet-lwoo-leerlingen. Hierbij zijn uitsplitsingen naar schoolsoort gemaakt. De resultaten tonen aan dat dit niet het geval is. Op de twee laagste schoolsoorten - vmbo-bl en vmbo kl – blijven leerlingen niet zitten, lwoo-leerlingen dus ook niet. Op het vmbo-tl/gl blijven lwoo-leerlingen juist iets minder vaak zitten dan niet-lwoo-leerlingen op het vmbo-tl/gl. Vmbo-vestigingen met relatief veel lwoo-leerlingen profiteren hier van. Verondersteld wordt bij de huidige berekeningsmethoden dat 1 op de 4 lwoo-leerlingen blijft zitten in de onderbouw van het VO. Dit is zeker niet het geval. Het is niet verstandig om VO-vestigingen te blijven compenseren voor lwoo-leerlingen op de manier waarop dat tot nu toe gedaan is. Maar hier moeten wel twee dingen uit elkaar gehouden worden. Als lwoo-leerlingen niet doubleren dankzij de extra ondersteuning die ze krijgen, dan is de financiële compensatie voor lwoo-leerlingen effectief gebleken. Als vestigingen bij de berekening van het onderbouwrendement gecompenseerd worden voor het feit dat ze lwoo-leerlingen hebben, dan is dit op grond van de resultaten van dit onderzoek feitelijk een overcompensatie. Het zou afgeschaft kunnen worden. De vestigingen hebben hier geen last van en de vergelijking tussen vestigingen wordt eerlijker.

Samenhang advies en onderwijsaanbod van ontvangende VO-vestigingen

Het rendement onderbouw van een VO-vestiging wordt in belangrijke mate bepaald door het kwaliteit van het advies dat de basisschool afgeeft voor een leerling. Met name brede adviezen – meer dan twee schoolsoorten breed – zijn daarbij problematisch: ze zorgen voor een grotere mate van onnauwkeurigheid wat betreft het startniveau van de leerling. Hoewel PO-scholen worden gewezen op het feit dat ze niet te brede schooladviezen mogen afgeven, lijkt dit nog steeds vrij frequent voor te komen. De veronderstelling is dat leerlingen die uitstromen naar brede brugklassen in VO-1, ook brede adviezen krijgen; vraag en aanbod worden op elkaar afgestemd. Op basis van de Bron-bestanden VO-2010-2011 en PO-2009-2010 is deze veronderstelling nader onderzocht¹⁷. Gevonden is dat het percentage brede adviezen gering is: 2.5% van het totaal. Er is een zwak positief verband tussen de breedte van het advies van de leerling en de breedte van de brugklas waar deze leerling een jaar later naar toe gaat (VO-1) ($r=0,13$). Het advies dat een leerling krijgt hangt dus maar in geringe mate samen met het onderwijsaanbod op de ontvangende VO-vestiging. Te brede adviezen komen weinig voor. Er is daarom niet sprake van een problematische situatie.

Deelproject III: Toegevoegde waarde van bovenbouw VO

Wat is de toegevoegde waarde van een VO-vestiging op het gemiddelde centraal examencijfer havo voor vmbo-tl gediplomeerde leerlingen die opgestroomd zijn naar het havo?

¹⁷ De aanname daarbij is dat vmbo-tl en vmbo gl één schoolsoort zijn: de breedte is gelijk aan 1. De breedte is maximaal 4 (schoolsoorten vmbo-bl tot en met vwo).

Verschilt de toegevoegde waarde van havo-afdelingen van VO-vestigingen tussen de verplichte vakken Nederlands, Engels en wiskunde?

Hoe verhoudt de toegevoegde waarde van een VO-vestiging op het gemiddelde centraal examencijfer zich tot het rendement voor vmbo-tl gediplomeerde leerlingen die doorgestroomd zijn naar het havo?

In dit deelproject zijn de mogelijkheden van het gebruik van toegevoegde waarde voor het gemiddelde centraal examencijfer havo van gediplomeerde vmbo-tl-leerlingen geïnventariseerd. Een eerste belangrijke bevinding is dat er sprake is van een zeer beperkte groep leerlingen en vestigingen waarvoor voldoende gegevens beschikbaar zijn. Hoewel de overstap naar de havo gemaakt werd door 8.236 vmbo-tl-leerlingen op 477 vestigingen, kan de toegevoegde waarde slechts bepaald worden voor 5.117 van deze leerlingen (62%) op 273 vestigingen (57%). Dit is een aanzienlijke afname. Er zijn twee belangrijke redenen aan te wijzen. In de eerste plaats is er een aanzienlijke groep leerlingen (20% na het eerste jaar) die het doorstroomtraject niet afmaakt en waarvoor dus geen gegevens beschikbaar zijn met betrekking tot het havo-examen. En in de tweede plaats betreft dit het criterium dat alleen vestigingen in de analyse worden opgenomen waarvan voor tenminste tien leerlingen alle nodige informatie beschikbaar is. Aangezien het hier om een specifieke groep leerlingen gaat leidt dit criterium tot een aanzienlijke afname van leerlingen en vestigingen. Wanneer naar losse vakken wordt gekeken loopt dit aantal nog verder terug.

Een tweede bevinding is dat de havo-afdelingen verschillen in hun toegevoegde waarde voor deze selecte groep leerlingen. Voor elke gebruikte uitkomstmaat (gemiddeld centraal examencijfer voor alle vakken, en centraal examencijfer voor Nederlands, Engels en wiskunde) is voldoende variantie gevonden tussen havo-afdelingen: de havo-afdelingen verschillen dus in dit opzicht. Voor elke uitkomstmaat zijn drie meerniveau modellen geschat die variëren in aantal achtergrondkenmerken (geen, alleen beginniveau en beginniveau met achtergrondkenmerken). Met name voor het gemiddelde centraal examencijfer voor alle vakken en het cijfer voor Engels leidt het meenemen van het beginniveau in de schatting van de toegevoegde waarde tot aanmerkelijke verschuivingen van vestigingen ten opzichte van het lege model. Voor Nederlands en wiskunde zijn deze verschuivingen aanzienlijk kleiner. Voor alle uitkomstmaten geldt dat het toevoegen van achtergrondkenmerken van leerlingen, zoals apcg, lwoo en cumi niet leidt tot een grote verschuiving van scholen. Dat betekent dat voor deze specifieke groep leerlingen (Vmbo-tl naar havo) gekozen kan worden voor een relatief simpele toegevoegde waarde bepaling met een model waarin alleen het beginniveau van de leerlingen is opgenomen.

Ten derde wordt een beperkte consistentie gevonden van de toegevoegde waarde van havo-afdelingen voor het gemiddelde centraal examencijfer voor alle vakken, Nederlands, Engels

en wiskunde. In alle gevallen is er sprake van een relatief zwakke, maar positieve samenhang. Dit betekent dat over het algemeen een hogere schatting van de toegevoegde waarde van een havo-afdeling op de ene uitkomstmaat samengaat met een hoge schatting op de toegevoegde waarde van een havo-afdeling voor een andere uitkomstmaat. Bijvoorbeeld een havo-afdeling met een hoge toegevoegde waarde voor het gemiddelde eindexamencijfer heeft waarschijnlijk ook een hoge toegevoegde waarde voor Nederlands, Engels of wiskunde. Echter, doordat de correlaties relatief laag zijn, betekent dit resultaat ook dat er een aanzienlijk aantal scholen is waarbij dit minder of niet het geval is. De sterkste samenhang werd gevonden tussen de toegevoegde waarde voor gemiddeld centraal examencijfer en de toegevoegde waarde voor wiskunde. De zwakke samenhang van de toegevoegde waarde indicatoren onderling kan mogelijk worden verklaard uit verschillen in inhouden tussen het vmbo-tl en havo-examen, maar kan bijvoorbeeld ook voortvloeien uit onbetrouwbaarheid van examens.

Doorstroomrendement op het havo van gediplomeerde vmbo-tl-leerlingen

In het kader van mogelijk strategisch gedrag is er in dit onderzoek ook specifieke aandacht geweest voor het rendement van de havo-afdelingen voor gediplomeerde vmbo-tl-leerlingen en voor de relatie tussen het doorstroomrendement en de toegevoegde waarde. Het doorstroomrendement kon worden bepaald voor een grotere groep vestigingen (334) aangezien ook uitstromende leerlingen in het doorstroomrendement zijn opgenomen. De vestigingen bleken aanzienlijk te verschillen in hun rendementen die ze voor deze groep leerlingen hebben behaald. Het doorstroomrendement kende een range van 17% tot 100%, met een gemiddelde van 56%. Wanneer gekeken wordt naar de associatie tussen het doorstroomrendement en toegevoegde waarde, valt op dat er alleen voor het gemiddelde centraal examencijfer alle vakken sprake is van een positieve correlatie. Dit betekent dat vestigingen die een hoog doorstroomrendement realiseren over het algemeen ook een hoge toegevoegde waarde voor het gemiddelde centraal examencijfer realiseren. Echter, gezien de relatief lage correlatie is het onvoldoende om de opbrengsten van een havo-afdeling als doorstroom-afdeling te bepalen op basis van één van de twee opbrengstindicatoren. Zowel het doorstroomrendement als de toegevoegde waarde dienen in ogenschouw te worden genomen om valide uitspraken te kunnen doen over opbrengsten van scholen in de bovenbouw.

Het huidige onderzoek roept ook nog een aantal extra vragen op. Voorbeeld hiervan is de vraag naar mogelijke verschillen in opbrengstindicatoren van havo-afdelingen voor de doorgestroomde vmbo-tl-leerlingen en voor de overige havo-leerlingen. Op basis van de voor dit project beschikbare data kan dat niet verder onderzocht worden. Men zou verwachten dat vestigingen met hoge opbrengsten voor de doorstromende leerlingen ook hoge opbrengsten realiseren voor de reguliere leerlingen. Bij het onderzoeken van de opbrengsten van scholen voor doorstromende en reguliere havo-leerlingen spelen een aantal moeilijkheden, zoals de periode waarover de toegevoegde waarde wordt bepaald, wisselingen van instellingen, verschillen in maten voor het beginniveau.

Een tweede vraag voor vervolgonderzoek is of er verschillen zijn in toegevoegde waarde en doorstroomrendement voor VO-vestigingen met veel doorstroom binnenshuis (binnen de brin) of veel doorstroom van buiten. Hoewel de doorstroom van leerlingen van vmbo-tl naar havo voornamelijk binnen een onderwijsinstelling plaatsvindt (78,2%), zijn er aanzienlijke verschillen tussen instellingen in het relatief aantal leerlingen dat doorstroomt binnenshuis of vanuit andere instellingen.

Deelproject IV: Extern rendement van VMBO basis- en kaderberoepsgerichte leerweg

Wat is het externe rendement na een en twee jaar mbo van vmbo basis- en kaderberoepsgerichte leerweg afdelingen van VO-vestigingen gegeven het examenklasniveau van hun leerlingen?

Wat zijn de effecten van de voorgestelde normering voor het percentage succesvolle plaatsingen?

In hoeverre is de succesvolle plaatsing van leerlingen in het eerste jaar na het verlaten van een vmbo basis- of kaderberoepsgerichte leerweg gerelateerd aan de VO-vestiging enerzijds en mbo-instelling anderzijds?

In hoeverre leidt het corrigeren voor invloeden van mbo-instellingen tot veranderingen van de schattingen van het externe rendement voor VO-vestigingen?

In dit deelproject worden de mogelijkheden verkent voor bepalen van het externe rendement van VO-vestigingen voor de basis- en kaderberoepsgerichte leerwegen van het vmbo. Het externe rendement betreft hier de mate waarin VO-vestigingen deze groep leerlingen succesvol weten te plaatsen in het mbo. Ook de effecten van de voorgestelde normering van 80%, 85% of 90% succesvolle plaatsingen zijn in dit deelproject aan de orde gekomen. Daarnaast zal speciale aandacht uitgaan naar de invloed die de mbo-instelling heeft op de bepaling van het externe rendement van VO-vestigingen.

Het onderzoek heeft zich gericht op VO-vestigingen met tenminste 10 leerlingen waarvoor de succesvolle plaatsing kan worden vastgesteld op basis van de beslisregels van de Inspectie van het Onderwijs. Deze selectie van vestigingen leidt ertoe dat voor een deel van de vestigingen geen extern rendement berekend kan worden. Voor de basisberoepsgerichte leerweg betreft dit ongeveer 10% van de vestigingen en voor de kaderberoepsgerichte leerweg ongeveer 5% van de vestigingen. Deze richtlijn van minimaal 10 leerlingen per school wordt ook wel door de Inspectie van het Onderwijs gehanteerd om een relatief betrouwbaar beeld te krijgen van een vestiging, met de bijkomende beperking dat niet alle vestigingen meegenomen kunnen

worden. Als de voorgestelde normering van 90% succesvolle plaatsingen na 1 jaar wordt gehanteerd dan zit slechts 46% van alle basis-afdelingen boven de norm, en 85% van alle kader-afdelingen. Bij een minder strenge norm van 85% of 80% succesvolle plaatsingen na 1 jaar, voldoet respectievelijk 68% en 83% van alle basis-afdelingen aan de eis, en 95% en 97% van alle kader-afdelingen.

Invloed van mbo-instelling op extern rendement

Het externe rendement van vestigingen vmbo basis- en kaderberoepsgericht is deels afhankelijk van factoren die samenhangen met de mbo-instellingen waar de leerlingen naar toe zijn gegaan, zoals het plaatsingsbeleid of de beschikbaarheid van opleidingen op de verschillende niveaus. Bij het externe rendement van een VO-vestiging is het zaak om de uniek bijdrage van de VO-vestiging vast te stellen. Hierdoor kunnen VO-vestigingen onderling op een eerlijke manier vergeleken worden. In dit deelproject zijn daarom meerniveau kruisklassificatie modellen toegepast om het effect van de mbo-instellingen uit het externe rendement te filteren. De modellen laten in de eerste plaats zien dat succesvolle plaatsing van leerlingen in een vervolgopleiding na de examenklas bl of kl zowel gerelateerd is aan de VO-vestigingen waar de leerlingen op hebben gezeten als aan de mbo-instellingen waar de leerlingen naar toe zijn gegaan. Daarnaast valt op dat de verschillen in succesvolle plaatsing van basis- en kaderleerlingen één jaar na de examenklas sterker gerelateerd zijn aan de mbo-instelling waar de leerlingen naar toe gaan dan aan de vmbo-vestiging waar de leerlingen van afkomstig zijn. Echter, ook na het uitfilteren van de invloed van mbo-instellingen zijn er nog verschillen waarneembaar tussen VO-vestigingen. In praktische zin betekent dit dat zowel de vmbo-vestiging als de mbo-instelling een rol spelen bij de eventuele succesvolle plaatsing van de vmbo-bl en kl-leerlingen één jaar nadat ze in de examenklas hebben gezeten. De resterende verschillen tussen VO-vestigingen in het kruisklassificatie model zijn een belangrijke indicatie dat scholen in het VO verschillen in hun extern rendement en dat het daarom zinvol is om dit verder te onderzoeken en mogelijk ook om indicatoren te ontwikkelen die in onderwijstoezicht gebruikt kunnen worden.

Een andere belangrijke bevinding is dat het meenemen van mbo-instellingen in de berekeningen van het externe rendement leidt tot zeer grote veranderingen van het geschatte extern rendement van VO-vestigingen. De gevonden overeenstemming was zeer beperkt en ook de correlatie tussen de schoolresiduen was laag. Dit indiceert dat het bepalen van het externe rendement van vmbo-bl en kl-afdelingen van VO-vestigingen zonder daarbij de invloed van mbo-instellingen mee te nemen, leidt tot inaccurate schattingen van het externe rendement van VO-vestigingen. Hierbij is het uitgangspunt dat het kruisklassificatie model conceptueel het beste model is, waar de andere modellen tegen afgezet kunnen worden. Ook het huidige model dat op dit moment door de Inspectie wordt gehanteerd kan derhalve nog niet als een valide indicator worden gezien. Het feit dat de mbo-instelling een grote impact

heeft op de schattingen van het externe rendement van een VO-vestiging maakt dat een valide indicator relatief complex zal zijn en niet zomaar narekenbaar voor scholen.

De Eindnotitie Indicator extern rendement laat zien dat er grote verschillen in succesvolle plaatsing zijn tussen sectoren in het mbo (Van der Heijden & Vreeburg, 2010). Dit is ook de aanleiding geweest om in dit onderzoek te kijken naar de vertekeningen van het externe rendement door statistisch te controleren voor de mbo-instellingen waar de leerlingen naar toe gaan. Echter, dit onderzoek kan worden gezien als een eerste verkenning van deze indicator, en er zijn nog vele vragen onbeantwoord.

In de eerste plaats betreft dit de stabiliteit van de indicator voor extern rendement voor meerdere opeenvolgende uitstroomcohorten van leerlingen. In dit onderzoek is alleen gekeken naar het uitstroomcohort leerlingen die in schooljaar 2010/2011 in de examenklas hebben gezeten. Het zou voor de context van het onderwijstoezicht interessant zijn om na te gaan in hoeverre het externe rendement stabiel blijft over opeenvolgende jaren. Dit kan een indicatie geven van de betrouwbaarheid van de indicator. Maar daarnaast kan ook gekeken worden naar de sectoren op opleidingen waar de leerlingen naar toe gaan en in hoeverre succesvolle plaatsingen per opleiding of sector verschillen.

In de tweede plaats betreft dit de verschillende onderliggende assumpties van de modellen die ten grondslag liggen aan de bepaling van het externe rendement van VO-vestiging. Dit zal nader onderzocht moeten worden. Een voorbeeld hiervan is dat het kruisklassificatie model als assumptie heeft dat er geen interactie is tussen de VO-vestigingen waarvan de leerlingen afkomstig zijn en de mbo-instellingen waar de leerlingen naar toe gaan. Door samenwerkingsprojecten of afspraken tussen VO-vestigingen en mbo-instellingen kan het zijn dat er specifieke combinaties zijn van VO-vestigingen en mbo-instellingen waarbij het percentage succesvolle plaatsingen erg hoog of erg laag zijn. In de lopende VM2-experimenten wordt getracht de aansluiting tussen vmbo-basis en mbo niveau 2 opleidingen te versterken door een intensieve samenwerking. Hierdoor zou een situatie kunnen ontstaan dat door deze samenwerking de aansluiting verbeterd en in het samenwerkingsverband veel succesvolle plaatsingen gerealiseerd kunnen worden. Dergelijke voorbeelden van een ‘goed huwelijk’ tussen een VO-vestiging en een mbo-opleiding of –instelling passen niet bij de onderliggende assumptie van het kruisklassificatie model, namelijk geen interactie. In hoeverre deze assumptie houdbaar is kan verder onderzocht worden.

Derde doelstelling: verkenning van meer differentiatie in toezicht

Het derde doel van dit project betreft het verkennen van de mogelijkheden hoe de verkregen informatie kan worden ingezet voor de differentiatie in het toezicht, waarbij onderscheid wordt gemaakt tussen scholen met zwakke opbrengsten, scholen met voldoende opbrengsten maar met mogelijkheden tot verbetering en scholen met goede opbrengsten.

De data-analyses van de vier deelprojecten hebben in alle gevallen geleid tot opbrengstindicatoren die ingezet kunnen worden om scholen onderling te vergelijken. Dit kan door gebruikt te maken van de schoolresiduen uit de statistische modellen. Het schoolresidu geeft de positie van een VO-vesting weer ten opzichte van het (corrigeerde) gemiddelde van alle scholen in de steekproef. De schoolresiduen van alle modellen kunnen eenvoudig bewaard worden in de databestanden waarmee de analyses worden gedaan. De waarden van de verkregen schoolresiduen volgen een bepaalde verdeling. De vorm van de verdeling kan verschillen per opbrengstindicator, maar het is altijd mogelijk om op een of andere manier drie groepen van VO-vestigingen te onderscheiden. Daarbij moeten afkapgrenzen gekozen worden. Waar de grens tussen zwakke scholen, scholen met voldoende opbrengsten maar met verbeteringsmogelijkheden en scholen met goede opbrengsten moet komen te liggen is in zekere zin arbitrair, maar welke keuze ook gemaakt, het zal discussie oproepen. De Inspectie van het Onderwijs zal uiteindelijk deze keuze moet maken. Het is belangrijk om de keuzes goed te onderbouwen en voldoende draagvlak in het onderwijsveld te verkrijgen. In de huidige toezichtpraktijk is de bolletjes-systematiek een geaccepteerde manier om dit te doen. Daarbij worden afkapgrenzen op basis van percentielen in de verdeling gehanteerd. Het is ook mogelijk om bij normale verdeelde opbrengstindicatoren over te gaan op afkapgrenzen die gebaseerd zijn op de spreiding rondom het gemiddelde.

Een recent voorbeeld van gebruik van schoolresiduen in combinatie met afkapgrenzen die gebaseerd zijn op de spreiding zijn de leerwinstbepalingen op schoolniveau in de Pilot leerwinst en toegevoegde waarde PO (Janssens, e.a., 2014). In de praktijk zijn hier dusdanig goede ervaringen mee opgedaan, dat er nu concrete plannen zijn om dit in te gaan bouwen in leerlingvolgsystemen voor PO-scholen (zie brief van Staatsecretaris Dekkers aan de Tweede Kamer van 13 juni 2014: Dekkers, 2014). Dit betekent evenwel dat scholen in de nabije toekomst gaan werken met schoolresiduen zonder dat hierbij de lastige term schoolresidu gebezigd zal worden in de communicatie met de scholen.

Literatuur

- Aitkin, M., & Longford, N. (1986). Statistical modelling issues in school effectiveness studies. *Royal Statistical Society*, 149, 1-43. doi:0035/9238/86/1490001
- Amrein-Beardsley, A. (2008). Methodological concerns about the Education Value-Added Assessment System. *Educational Researcher*, 37, 65-75.
- Betebenner, D. W. (2007). *Estimation of student growth percentiles for the Colorado student assessment program*.
http://www.cde.state.co.us/cdedocs/Research/PDF/technicalsgppaper_betebenner.pdf
- Bosker, R., Lam, J.F., Luyten, H., Steen, R., & Vos, H. de. (1998). *Het vergelijken van scholen*. Enschede: OCTO.
- Bosker, R. J., Béquin, A., & Rekers-Mombarg, L.T.M. (2001). Hoe meten we de prestatie van een school? In: Dijkstra, A.B., Karsten, S., Veenstra, R., & Visscher, A. (Red.) *Het oog der natie: scholen op rapport, standaarden voor de publicatie van schoolprestaties*.(121-135). Assen, Van Gorcum.
- Bosker, R.J. (2012) De toegevoegde waarde van een school: begripsbepaling, meting en causale attributie. In: Dijkstra, A.B., & Janssens, J.G. (Red.) *Om de kwaliteit van het onderwijs. Kwaliteitsbepaling en kwaliteitsbevordering*. (93-104). Assen, Van Gorcum.
- Browne, W.J. (2009). *MCMC estimation in MLwiN v2.1*. Bristol: Centre for Multilevel Modelling, University of Bristol
- Browne, W.J., Goldstein, H., Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling*, 1, 103-124.
- Castellano, K.E. & Ho, A.D. (2013). *A Practitioner's Guide to Growth Models*. The Council of Chief State School Officers.
http://scholar.harvard.edu/files/andrewho/files/a_practitioners_guide_to_growth_models.pdf
- De Boer, H., Bosker, R.J., & Van der Werf, M.P.C. (2010). Sustainability of teacher expectation bias effects on long-term student performance. *Journal of Educational Psychology*, 102, 168-179.
- Dekker, S. (2014). *Nut en waarde van toetsing in het funderend onderwijs*. Brief aan de Tweede Kamer.
http://www.tweedekamer.nl/kamerstukken/brieven_regering/detail.jsp?id=2014Z11041&did=2014D22160

- Dijkstra, A. B., Karsten, S., Veenstra, R., & Visscher, A. (Red.) (2001). *Het oog der natie: scholen op rapport, standaarden voor de publicatie van schoolprestaties*. Assen, Van Gorcum.
- Goldstein, H. (1997). Methods in school effectiveness research. *School Effectiveness and School Improvement*, 8, 369-395. doi: 0924-3453/97/0804-0369
- Guldemon, H., & Bosker, R. J. (2009). School effects on students' progress – a dynamic perspective. *School Effectiveness and School Improvement*, 20 (2), 255-268.
- Harris, D. N. (2011). *Value-added measures in education: What every educator needs to know*. Cambridge, MA: Harvard Education Press.
- Hofman, R. H., & Spijkerboer, A. W. (2009). *Ervaren deficiënties door havo- en mbo-opleidingen in de basisbagage van vmbo'ers*. Groningen: GION.
- Janssens, F.J.G., Rekers-Mombarg, L., & Lacor, E. (2014). *Leerwinst en toegevoegde waarde in het primair onderwijs*. Den Haag, Ministerie van OCW.
<http://www.rijksoverheid.nl/documenten-en-publicaties/rapporten/2014/01/30/leerwinst-en-toegevoegde-waarde-in-het-primair-onderwijs.html>
- Koretz, D.M. (2003) Using multiple measures to address perverse incentives and score inflation. *Educational Measurement: Issues and Practice*, 22(2), 18–26.
- Kuyper, H., Lubbers, M.J., & Werf, M.P.C. van der (2003). VOCL'99-1: Technisch rapport. Groningen: GION.
- McCaffrey, D. F., Lockwood, J. R., Koretz, D., Louis, T. A., & Hamilton, L. (2004). Models for value-added modeling of teacher effects. *Journal of Educational and Behavioral Statistics*, 29(1): 67–101.
- Nash, R. (2003). Is the school composition effect real? A discussion with evidence from the UK PISA data. *School Effectiveness and School Improvement*, 14, 441 – 457. doi: 10.1076/sesi.14.4.441.17153
- Inspectie van het Onderwijs. (2012). *Opbrengstenkaart 2012, technische toelichting*. Utrecht: Inspectie van het Onderwijs.
- Onderwijsraad (2002). *Wat scholen toevoegen*. Den Haag: Onderwijsraad.
- OECD (2008). *Measuring improvements in learning outcomes, best practices to assess the value added of schools*. Paris; OECD publications.
- Rasbash, J., Steele, F., Browne, W. J., & Goldstein, H. (2012). *A user's guide to MLwiN*, 2.26. Bristol: Centre for Multilevel Modelling, University of Bristol.

- Ray, A. (2006). *School value added measures in England: A paper for the OECD project on the development of value-added models in education systems*.
<https://consumption.education.gov.uk/publications/eOrderingDownload/RW85.pdf>
- Raudenbush, S.W., & Bryk, A.S. (1986). A hierarchical model for studying school effects. *Sociology of Education*, 59, 1–17.
- Sammons, P., Thomas, S., & Mortimore, P. (1997). *Forging links: effective schools and effective department*. London: Paul Chapman.
- Sanders, W. L., & Horn, S. P. (1994). The Tennessee Value-Added Assessment System (TVAAS): Mixed-model methodology in educational assessment. *Journal of Personnel Evaluation in Education*, 8, 299–311.
- Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon.
- Snijders, T. A. B., & Bosker, R.J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Los Angeles / London / New Delhi / Singapore / Washington DC: Sage Publications.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B*, 64, 583 – 639.
- Thomas, S., Peng, W-J, Gray, J., (2007). Modelling patterns of improvement over time: value added trends in English secondary school performance across ten cohorts: *Oxford Review of Education*. 33, 261 – 295.
- Thrupp, M., Lauder, H., & Robinson, T. (2002). School composition and peer effects. *International Journal of Educational Research*, 37, 483 – 504. doi: 10.1016/S0883-0355(3)0016-8
- Timmermans, A. C., Doolaard, S., & De Wolf, I. F. (2011). Conceptual and empirical differences among various value added models. *School Effectiveness and School Improvement*, 22, 393 – 413. doi: 10.1080/09243453.2011.590704
- Timmermans, A.C. (2012). *Value added in educational accountability: possible, fair and useful?* Groningen: GION, Rijksuniversiteit Groningen.
- Timmermans, A.C., Bosker, R. J., Doolaard, S., & De Wolf, I. (2012). Value added as an indicator of educational effectiveness in Dutch senior secondary vocational education. *Journal of Vocational Education and Training*, 64, 417-432. DOI: 10.1080/13636820.2012.727849

Timmermans, A. C., Snijders, T. A. B., & Bosker, R. J. (2013). In search of value added in case of complex school effects. *Journal of Educational and Psychological Measurement*, 73, 210 - 228. DOI: 10.1177/0013164412460392

Van Bijsterveldt, M. (2011). *Actieplan "Focus op vakmanschap 2011 – 2015"*. Brief aan de Tweede Kamer.

Van der Heijden, M., & Vreeburg, B. (2010). *Eindrapportage onderzoek indicator extern rendement vmbo*. Utrecht: Inspectie van het Onderwijs.

Appendix bij hoofdstuk 3

Tabel 1. Resultaten van de meerniveau modellen voor het bovenbouwrendement op vmbo basisberoepsgerichte leerweg

	Lege model		Beoogde model		Optimale model	
	Coëfficiënt	SE	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed part:</i>						
Constante	2,911	0,175	3,616	0,357	-25,680	12,360
% Instroom in 3e leerjaar			-0,905	1,731		
Apcg			-0,875	0,376		
Lwoo			-0,632	0,376		
Cito-eindtoetscore					0,054	0,024
SES					0,276	0,152
Meisje					-0,060	0,357
<i>Random part:</i>						
Tussen VO-vestiging verschillen	0,000	0,000	0,000	0,000	0,000	0,000
Aantal VO-vestigingen	22		22		22	
Aantal leerlingen	659		659		659	

etniciteit ontbreekt als correctiefactor omdat er problemen bij het schatten van het model ontstonden (geen convergentie).

Tabel 2. Resultaten van de meerniveau modellen voor het bovenbouwrendement op vmbo kaderberoepsgerichte leerweg

	Lege model		Beoogde model		Optimale model	
	Coëfficiënt	SE	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed part:</i>						
Constante	3,067	0,248	2,680	0,310	-1,397	13,770
% Instroom in 3e leerjaar			3,119	1,787		
Apcg			0,244	0,406		
Lwoo			-0,232	1,039		
Cito-eindtoetscore					0,008	0,026
SES					0,084	0,148
Meisje					0,582	0,380
<i>Random part:</i>						
Tussen VO-vestiging verschillen	0,526	0,395	0,278	0,311	0,433	0,366
Aantal VO-vestigingen	25		25		25	
Aantal leerlingen	705		705		705	

etniciteit ontbreekt als correctiefactor omdat er problemen bij het schatten van het model ontstonden (geen convergentie).

Tabel 3. Resultaten van de meerniveau modellen voor het bovenbouwrendement op vmbo gemengde en theoretische leerweg.

	Lege model		Beoogde model		Optimale model	
	Coëfficiënt	SE	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed part:</i>						
Constante	2,514	0,130	2,345	0,168	-25,600	5,249
% Instroom in 3e leerjaar			3,068	0,792		
Apcg			-0,516	0,161		
Lwoo			-1,471	0,400		
Cito-eindtoetscore					0,052	0,010
SES					0,127	0,052
Meisje					0,263	0,138
<i>Random part:</i>						
Tussen VO-vestiging verschillen	0,706	0,186	0,504	0,153	0,624	0,172
Aantal VO-vestigingen	63		63		63	
Aantal leerlingen	3253		3253		3253	
# etniciteit ontbreekt als correctiefactor omdat er problemen bij het schatten van het model ontstonden (geen convergentie).						

De tabel voor het bovenbouwrendement van het havo staat in Hoofdstuk 3, Tabel 3.5.

Tabel 4. Resultaten van de meerniveau modellen voor het bovenbouwrendement op vwo

	Lege model		Beoogde model		Optimale model	
	Coëfficiënt	SE	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed part:</i>						
Constante	1,915	0,149	1,765	0,191	-13,290	6,150
% Instroom in 3e leerjaar			2,437	1,150		
Apcg			-0,342	0,182		
Lwoo			0,000	0,000		
Cito-eindtoetscore					0,028	0,011
SES					0,054	0,051
Meisje					-0,028	0,140
<i>Random part:</i>						
Tussen VO-vestiging verschillen	0,582	0,184	0,476	0,160	0,547	0,176
Aantal VO-vestigingen	35		35		35	
Aantal leerlingen	1900		1900		1900	
# etniciteit ontbreekt als correctiefactor omdat er problemen bij het schatten van het model ontstonden (geen convergentie).						

Tabel 5 Resultaten van de meerniveau modellen voor het eindexamencijfer vmbo basisberoepsgerichte leerweg

	Lege model		Beoogde model		Optimale model	
	Coëfficiënt	SE	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed part:</i>						
Constante	6,921	0,060	7,248	0,061	-14,495	1,984
% instroom in 3 ^e leerjaar			-1,001	0,412		
Apcg*			-0,212	0,065		
Lwoo			-0,625	0,067		
Cito-eindtoetscore					0,042	0,004
Etniciteit: #						
Nederlands					-0,553	0,315
Turks					-0,835	0,336
Marokkaans					-0,692	0,337
Antiliaans, Surinaams, Arubaans					-0,419	0,391
Overig					-0,67	0,321
SES					0,064	0,023
Meisje					-0,021	0,053
<i>Random Part:</i>						
Tussen VO-vestiging verschillen	0,075	0,028	0,022	0,012	0,016	0,01
Tussen leerlingen verschillen	0,464	0,027	0,404	0,024	0,382	0,022
<i>Model fit (-2* logLikelihood):</i>	1305,411		1121,064		1165,213	
Aantal VO-vestigingen	33		33		33	
Aantal leerlingen	612		612		612	

* apcg is dichotome variabele (1=woonachtig in een armoedeprobleemcumulatiegebied , 0=niet woonachtig in een armoedeprobleemcumulatiegebied); # etniciteit onbekend is referentie-categorie.

Tabel 6 Resultaten van de meerniveau modellen voor het eindexamencijfer vmbo kaderberoepsgerichte leerweg

	Lege model		Beoogde model		Optimale model	
	Coëfficiënt	SE	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed part:</i>						
Constante	6,506	0,051	6,515	0,077	-6,551	2,365
% instroom in 3 ^e leerjaar			-0,027	0,366		
Apcg*			0,004	0,069		
Lwoo			-0,173	0,165		
Cito-eindtoetscore					0,025	0,004
Etniciteit: #						
Nederlands					-0,003	0,303
Turks					-0,562	0,344
Marokkaans					-0,209	0,337
Antiliaans, Surinaams, Arubaans					0,18	0,359
Overig					-0,059	0,314
SES					-0,02	0,023
Meisje					-0,123	0,054
<i>Random Part:</i>						
Tussen VO-vestiging verschillen	0,051	0,021	0,059	0,024	0,040	0,017
Tussen leerlingen verschillen	0,458	0,026	0,466	0,028	0,422	0,024
<hr/>						
<i>Model fit (-2* logLikelihood):</i>	1326,727		1232,441		1272,206	
Aantal VO-vestigingen	36		36		36	
Aantal leerlingen	629		629		629	

* apcg is dichotome variabele (1=woonachtig in een armoedeprobleemcumulatiegebied , 0=niet woonachtig in een armoedeprobleemcumulatiegebied); # etniciteit onbekend is referentie-categorie.

Tabel 7 Resultaten van de meerniveau modellen voor het eindexamencijfer vmbo gemengde en theoretische leerweg

	Lege model		Beoogde model		Optimale model	
	Coëfficiënt	SE	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed part:</i>						
Constante	6,371	0,033	6,413	0,043	-8,936	1,219
% instroom in 3 ^e leerjaar			-0,237	0,180		
Apcg*			-0,095	0,034		
Lwoo			-0,016	0,118		
Cito-eindtoetscore					0,029	0,002
Etniciteit: #						
Nederlands					0,027	0,186
Turks					-0,482	0,205
Marokkaans					-0,195	0,203
Antiliaans, Surinaams, Arubaans					-0,053	0,205
Overig					-0,003	0,190
SES					0,013	0,011
Meisje					-0,090	0,026
<i>Random Part:</i>						
Tussen VO-vestiging verschillen	0,057	0,013	0,047	0,011	0,037	0,009
Tussen leerlingen verschillen	0,431	0,013	0,419	0,012	0,393	0,011
<hr/>						
<i>Model fit (-2* logLikelihood):</i>	4941,747		4743,775		4703,206	
Aantal VO-vestigingen	78		78		78	
Aantal leerlingen	2421		2421		2421	

* apcg is dichotome variabele (1=woonachtig in een armoedeprobleemcumulatiegebied , 0=niet woonachtig in een armoedeprobleemcumulatiegebied); # etniciteit onbekend is referentie-categorie.

De tabel voor het eindexamencijfer havo staat in Hoofdstuk 3, Tabel 3.8.

Tabel 8 Resultaten van de meerniveau modellen voor het eindexamencijfer vwo

	Lege model		Beoogde model		Optimale model	
	Coëfficiënt	SE	Coëfficiënt	SE	Coëfficiënt	SE
<i>Fixed part:</i>						
Constante	6,503	0,035	6,499	0,050	-32,268	2,818
% instroom in 3 ^e leerjaar			0,057	0,297		
Apcg*			0,005	0,051		
Lwoo			0,000	0,000		
Cito-eindtoetsscore					0,071	0,005
Etniciteit: #						
Nederlands					-0,070	0,348
Turks					-0,183	0,423
Marokkaans					-0,044	0,523
Antiliaans, Surinaams, Arubaans					-0,180	0,372
Overig					-0,047	0,352
SES					0,060	0,014
Meisje					-0,076	0,036
<i>Random Part:</i>						
Tussen VO-vestiging verschillen	0,030	0,011	0,031	0,011	0,024	0,009
Tussen leerlingen verschillen	0,530	0,020	0,528	0,020	0,459	0,017
<i>Model fit (-2* logLikelihood):</i>	3276,408		3206,655		3062,646	
Aantal VO-vestigingen	41		41		41	
Aantal leerlingen	1469		1469		1469	

* apcg is dichotome variabele (1=woonachtig in een armoedeprobleemcumulatiegebied , 0=niet woonachtig in een armoedeprobleemcumulatiegebied); # etniciteit onbekend is referentie-categorie.

Appendix bij hoofdstuk 5

Beschrijvende statistieken van de gebruikte dataset

Uit de van de Inspectie van het Onderwijs gekregen inschrijvingsgegevens komt naar voren dat een groep van 8.236 leerlingen die in 2008/2009 geslaagd zijn in de theoretische leerweg doorstromen naar havo. Deze 8.236 leerlingen beginnen in 2009/2010 aan het havo-traject in 477 VO-vestigingen. Het aantal leerlingen per vestiging varieert tussen 1 en 68 leerlingen. Gemiddeld gaat het per vestiging om 17,3 leerlingen. Bij de doorstroom van leerlingen van vmbo-tl naar havo lijkt het vooral te gaan om leerlingen die doorstromen binnen de eigen onderwijs instelling (brin), aangezien 78,2% van de leerlingen binnen eigen instelling blijft en 21,8% van de leerlingen naar een andere instelling gaat. Eenzelfde beeld wordt zichtbaar op het niveau van de onderwijsinstellingen. Gemiddeld is 26% van de doorgestroomde leerlingen binnen een instelling afkomstig uit een andere instelling. Echter, hierin zijn grote verschillen waarneembaar tussen instellingen, variërend van 0% (alleen doorstroom in huis) tot 100% (alleen doorstroom uit andere instelling).

In Tabel 1 zijn enkele beschrijvende statistieken van deze groep doorstromende leerlingen weergegeven voor zowel hun achtergrond als hun prestaties op het examen vmbo-tl. Opvallend is dat er relatief veel jongens zijn (53%) die de doorstroom naar havo maken. Daarnaast is er toch een groep van bijna 200 leerlingen (2,4%) die gedurende het vmbo leerwegondersteuning heeft gehad en die de doorstroom naar havo maken. Dit zijn leerlingen die bij aanvang of gedurende het vmbo aanzienlijke achterstanden hadden op een aantal belangrijke cognitieve domein eventueel gecombineerd met een sociaal-emotionele problematiek. Van bijna alle leerlingen die doorstromen zijn gegevens bekend van het eindexamen vmbo-tl. Voor vier leerlingen zijn geen examencijfers bekend. Over het algemeen behalen de leerlingen een ruim voldoende gemiddelde eindcijfer. De verschillen tussen leerlingen in cijfers op het centraal examen zijn echter groot. Voor het gemiddelde eindexamencijfer is de range van gevonden cijfers 3,9 tot 8,9. In totaal zijn er 130 studenten die gemiddeld lager dan een 5,5 hebben gescoord op het centraal examen theoretische leerweg. Gemiddeld genomen presteerden de leerlingen het beste op het eindexamen Engels en laagste op het centraal examen Nederlands. Voor alle losse examens geldt dat er een aanzienlijke spreiding is in de examencijfers. Het is opvallend dat er voor een relatief grote groep leerlingen geen cijfer bekend is voor wiskunde.

Tabel 1 Beschrijvende statistieken

Variabele	N	Categorieën	Gemiddelde	Standaard deviatie	%
Sekse	8236	Jongen			53.0
		Meisje			47.0
Cumi	8236	Autochtoon			76.8
		Niet Westerse allochtoon			16.5
		Westerse allochtoon			6.6
		Onbekend			0.1
Generatie allochtoon	8236	Autochtoon			76.8
		1 ^{ste} generatie allochtoon			5.3
		2 ^{de} generatie allochtoon			17.9
		Onbekend			0.1
Apcg	8236	Geen overschrijding			59.9
		Laag inkomen			2.3
		Niet Westerse allochtoon			12.5
		Uitkering			4.1
		Laag inkomen + uitkering			3.0
		Laag inkomen + Niet Westerse allochtoon			1.0
		Uitkering + Niet Westerse allochtoon			3.8
		Cumulatie gebied			12.7
		Onbekend			0.7
Lwoo	8236	Ja			2.4
		Nee			97.6
Leeftijd	8236		15,23	0,50	
Examencijfer	8232	Gemiddeld	6.78	0.59	
	8230	Nederlands	6.53	0.78	
	8230	Engels	7.01	1.09	
	7271	Wiskunde	6.87	1.19	

Naast de achtergrondkenmerken en examencijfers van de leerlingen kunnen ook de onderwijsloopbanen in kaart worden gebracht. In Tabel 2 zijn de onderwijsposities van de leerlingen weergegeven vanaf schooljaar 2009/2010.

Tabel 2 Doorstroom van leerlingen in havo

School- jaar	Onderwijs- positie	N	%	Geslaagd	Afge- wezen	Terug- getrok- ken	Certi- ficaat	Gespreid	N.v.t.
2009/	Havo-3	1	0,0						1
2010	Havo-4	8227	99,9						8227
	Havo-5	8	0,1	3		1			4
2010/	Havo-4	635	7,7						635
2011	Havo-5	5742	69,7	4739	900	59		1	43
	Vwo-4	1	0,0						1
	Vwo-5	30	0,4						30
	Havo naar vavo-4	28	0,3						28
	Havo naar vavo-5	137	1,7	24	8	17	64		24
	n.v.t.	1663	20,2						
2011/	Havo-4	1	0,0						1
2012	Havo-5	835	10,1	635	167	21			12
	Vwo-5	71	0,9						71
	Vwo-6	21	0,3	18	3				
	Havo naar vavo-4	1	0,0						1
	Havo naar vavo-5	474	5,8	305	54	20	69	2	24
	Vwo naar vavo-5	1	0,0						1
	Vwo naar vavo-6	14	0,2	3	1	1	8		1
	n.v.t.	6818	82,8						

Tabel 2 toont dat de meerderheid van de naar havo doorstromende leerlingen te vinden in de vierde klas. Slechts een paar leerlingen lijken over te stappen naar de havo-examenklas. Van de acht leerlingen die doorstromen naar de havo-examenklas zijn er 3 leerlingen direct geslaagd voor het havo-examen, 1 leerling is teruggetrokken en de overige leerlingen hebben niet aan het examen meegedaan.

Wanneer naar de overgang naar schooljaar 2010/2011 wordt gekeken dan valt het relatieve grote aantal leerlingen op waarvoor geen inschrijvingsgegevens in het voortgezet onderwijs meer van bekend zijn. Dit zijn leerlingen die niet meer in het bekostigd voortgezet onderwijs staan ingeschreven. Hieronder vallen de drie leerlingen die in 2009/2010 al examen gedaan hebben, maar ook een groot aantal leerlingen die het VO hebben verlaten (1.663; 20,2%). Op

basis van de beschikbare gegevens is het niet mogelijk om te achterhalen of deze leerlingen alsnog besloten hebben om naar het mbo te gaan of welke andere onderwijsloopbanen de leerlingen volgen. Zoals verwacht mag worden kan het merendeel van de leerlingen in dit schooljaar worden gevonden in de vijfde klas van de havo (5.742 leerlingen; 69,7%). Van de leerlingen die in de examenklas zitten hebben de meeste leerlingen deelgenomen aan het centraal examen en is de grootste groep van de deelnemende leerlingen geslaagd (4.739). Echter, de groep leerlingen die niet zijn geslaagd voor het examen is aanzienlijk (900 leerlingen afgewezen). De groep leerlingen die is blijven zitten en ook in schooljaar 2010/2011 in havo-4 zit is aanzienlijk, namelijk 635 leerlingen (7,7%). Tot slot is er ook nog een kleine groep leerlingen die andere wegen kiest, namelijk via het vwo (totaal 31 leerlingen, 0,4%) of via havo naar vavo (totaal 165, 2,0%).

Wanneer de onderwijsposities van de doorstromende leerlingen in 2011/2012 worden bekeken komt naar voren dat voor meer dan 80% van de leerlingen geen inschrijvingsgegevens in het voortgezet onderwijs meer bekend zijn. Dit is niet verwonderlijk gezien het relatief grote aantal leerlingen wat binnen de nominale tijd het eindexamen heeft gehaald (4.739) en het aantal leerlingen die na het eerste jaar al waren uitgestroomd (1.663). In de inschrijvingsgegevens van dit schooljaar zijn er een aantal zaken die opvallen. In de eerste plaats is dit het relatief grote aantal leerlingen die doorstromen naar het vwo (92) en van havo naar vavo (475). Opnieuw doen bijna alle leerlingen in de examenklas mee aan het examen, en opnieuw slaagt het merendeel van de leerlingen. Deze leerlingen worden gerekend als vertraagd, aangezien ze er een jaar langer dan nominaal over hebben gedaan. Echter, zowel voor de havo als van havo naar vavo is er een aanzienlijke groep leerlingen afgewezen (respectievelijk 167 en 54).

Slechts een zeer beperkt deel van de leerlingen is nog te vinden in de inschrijvingsgegevens van 2012/2013, derhalve zijn is deze groep niet opgenomen in Tabel 2. In totaal gaat het om 182 leerlingen, waarvan de helft (90 leerlingen) zich in leerjaar 5 van havo naar vavo bevindt en de overige leerlingen zijn verspreid over havo, vwo en vwo naar vavo.

De cijfers van de leerlingen op het centraal examen havo worden gebruikt als afhankelijke variabele in de bepaling van de toegevoegde waarde van de havo-afdelingen. Vergelijkbaar met de examencijfers van het vmbo is ook hier een gemiddelde examencijfer bepaald, waarin alle vakken zijn opgenomen waarin de leerling het centraal eindexamen heeft afgelegd. En daarnaast worden de losse centraal examencijfers van de leerlingen op de vakken Nederlands, Engels en wiskunde gebruikt. In Tabel 3 worden beschrijvende statistieken gegeven van de scores van de leerlingen op het centraal examen havo. Hierbij moet worden opgemerkt dat het voor de meeste leerlingen gaat om de cijfers van het examen in 2011 en voor een beperkt aantal leerlingen voor het examen van 2010 (versneld) of 2012 (vertraagd). Hierbij wordt de assumptie gemaakt dat de cijfers van de examens in opeenvolgende jaren stabiel zijn. Dit houdt in dat eenzelfde prestatie in de verschillende examens met eenzelfde cijfers beloond

wordt. Met andere woorden, een cijfer 6 in 2010 weerspiegelt eenzelfde prestatie als een 6 in 2011.

Tabel 3 Beschrijvende statistieken van de cijfers van het centraal examen havo van naar havo doorstromende vmbo-tl gediplomeerde leerlingen

	N	Minimum	Maximum	Gemiddelde	Standaard Deviatie
Gemiddeld cijfer over alle vakken	6168	3,0	8,8	6,1	0,7
Cijfer Nederlands	5898	1,0	8,8	5,9	1,0
Cijfer Engels	5900	2,0	9,6	6,2	1,3
Cijfer Wiskunde	5149	1,0	10,0	6,1	1,3

Het gemiddeld examencijfer op één van de examens tussen 2010 en 2012 is bekend voor 6.168 naar havo doorstromende leerlingen. Dit zijn zowel leerlingen die geslaagd zijn als de leerlingen die voor het examen gezakt zijn. Voor de losse vakken zijn minder gegevens beschikbaar variërend tussen 5.149 observaties voor wiskunde en 5.900 observatie voor Engels. Zowel voor het gemiddelde cijfer op het centraal examen als de losse cijfer voor de vakken valt de grote spreiding op tussen het hoogste en laagst behaalde cijfer. Echter, voor alle vier de cijfers ligt het gemiddelde rond de 6. Net als bij de vmbo-examen cijfers ligt het gemiddelde cijfer voor Nederlands iets lager dan de cijfers voor Engels, wiskunde en het gemiddelde over alle vakken.

Appendix bij hoofdstuk 6

Beschrijvende statistieken van de gebruikte dataset

De gebruikte dataset voor het onderzoeken van het externe rendement vmbo basisberoepsgerichte leerweg (bl) en kaderberoepsgerichte leerweg (kl) bevatten respectievelijk 23.720 en 26.931 leerlingen in 507 vestigingen van VO-scholen die in schooljaar 2010-2011 in het laatste leerjaar van hun opleiding zaten. In Tabel 1 zijn de examenresultaten van de leerlingen weergegeven. Voor beide leerwegen geldt dat de meerderheid van de leerlingen slaagt. In de beslissingsregels voor het externe rendement is het examenresultaat niet specifiek opgenomen, maar wordt een succesvolle plaatsing slechts gebaseerd op de gevolgde leerweg. Echter in de gegevens die op internet beschikbaar zijn voor scholen wordt wel een onderscheid gemaakt tussen de succesvolle plaatsing van gediplomeerde en ongediplomeerde leerlingen. Opvallend in het databestand is de grote groep leerlingen uit de basisberoepsgerichte leerweg met de codering examenklas. Dit zijn leerlingen die wel in de examenklas hebben gezeten, maar niet hebben deelgenomen aan het examen.

Tabel 1 Examenresultaat van vmbo-bl en kl leerlingen

				Examenresultaat 2010-2011			
				Slaag	Zak	Geen examen gedaan	Totaal
Leerweg	vmbo	2010- 2011	Basisberoepsgerichte leerweg	20.197	982	2.541	23.720
			Kaderberoepsgerichte leerweg	24.840	1.486	605	26.931

Naast gegevens over de examenresultaten van de leerlingen zijn er drie variabelen beschikbaar die mogelijk als controlevariabelen kunnen dienen. In principe worden deze niet meegenomen, aangezien deze ook niet beschreven staan in de huidige beslissingsregels. Echter, er kan verkennend worden gekeken in hoeverre deze controlevariabelen impact hebben op de schattingen van het externe rendement van VO-scholen (vestigingen). De beschikbare variabelen zijn leerwegondersteuning (lwoo), armoedeprobleemcumulatiegebied (apcg) en etnische achtergrond (cumi). De verdeling van achtergrond kenmerken voor de bl en kl-leerlingen zijn in Tabel 2 weergegeven. Met name het verschil tussen de basisberoepsgerichte leerweg en de kaderberoepsgerichte leerweg in het aantal leerlingen met leerwegondersteuning is opvallend. Meer dan de helft van de leerlingen in bl heeft leerwegondersteuning terwijl de aantallen in kl aanzienlijk kleiner zijn.

Tabel 2 Achtergrondgegevens van vmbo-bl en kl leerlingen

Variabele		Basisberoeps- gerichte leerweg	Kaderberoeps- gerichte leerweg
Leerwegondersteuning (lwoo)	Wel	13.487	7.035
	Niet	10.233	19.896
Etnische achtergrond leerling	Autochtoon	16.236	20.585
	Niet Westerse allochtoon	6.119	4.981
	Westerse allochtoon	1.259	1.336
	Onbekend	106	29
Armoedeprobleemcumulatiegebied (Apcg)	Geen overschrijding	11.828	15.397
	Laag inkomen	426	523
	Niet Westerse allochtonen	3.032	3.484
	Uitkering	963	1.035
	Laag inkomen + niet Westerse allochtonen	257	223
	Laag inkomen + uitkering	862	829
	Niet Westerse allochtonen + uitkering	1.001	1.047
	Cumulatiegebied	5.240	4.240
	Onbekend	111	153

Doorstroom van leerlingen uit Vmbo-bl en kl naar leerjaar 2011-2012

De doorstroom van de leerlingen naar de verschillende schoolsoorten in het schooljaar 2011-2012 wordt in Tabel 3 weergegeven. Dit is het eerste jaar nadat de leerlingen in de examenklas hebben gezeten. Uit de onderstaande tabel blijkt dat de uitstroom van leerlingen uit zowel bl als kl zeer divers is, hoewel voor beide leerwegen geldt dat verreweg de grootste groep leerlingen hun onderwijsloopbaan vervolgt in het mbo. Daarnaast valt een groep leerlingen op die voortijdig schoolverlaten, namelijk 6,5% van de bl-leerlingen en 3,3% van de kl-leerlingen. Tenslotte zijn er een aantal leerlingen met opvallende doorstroom gegevens, namelijk naar het hoger onderwijs of schoolverlater met startkwalificatie. Voor deze beide bijzondere doorstroomgroepen blijkt uit de tabellen in de notitie "Eindrapportage onderzoek indicator extern rendement vmbo" dat deze groepen worden gecodeerd als leerlingen met een succesvolle plaatsing. Deze codering is in dit onderzoek overgenomen. Deze eindrapportage biedt geen duidelijkheid over de codering van plaatsing voor de leerlingen naar het speciaal onderwijs (SO). In overleg met de opdrachtgever is er voor gekozen om deze groep leerlingen te coderen als niet succesvol geplaatst.

Zoals beschreven is de doorstroom van bl- en kl-leerlingen succesvol wanneer ze onderwijs gaan volgen in een mbo-opleiding van niveau 2 of hoger. In Tabel 4 wordt een verder onderscheid gemaakt naar het niveau van de vervolgopleiding in het mbo. Uit deze tabel blijkt dat voor beide leerwegen geldt dat het grootste deel van de leerlingen succesvol in het mbo wordt geplaatst, hoewel er voor de basisberoepsgerichte leerweg een aanzienlijk deel van de leerlingen (5,4%) naar een mbo-niveau 1 opleiding gaat.

Tabel 3 Doorstroom van leerlingen naar onderwijssectoren in schooljaar 2011-2012

Sector vervolg schoolloopbaan 2011-2012	Basisberoepsgerichte leerweg		Kaderberoepsgerichte leerweg	
	Aantal	Percentage	Aantal	Percentage
Speciaal Onderwijs	131	0,6	50	0,2
Voortgezet Onderwijs	1.697	7,2	1.024	3,8
Middelbaar Beroeps Onderwijs	20.347	85,8	24.970	92,7
Volwassenen Educatie	5	0,0	6	0,0
Hoger Onderwijs	1	0,0		
Schoolverlater met start kwalificatie	30	0,1	3	0,0
Vsv met vmbo diploma	948	4,0	702	2,6
Vsv met eerder vmbo diploma			1	0,0
Vsv met mbo1	85	0,4	2	0,0
Vsv	476	2,0	173	0,6
Totaal	23.720		26.931	

Tabel 4 Doorstroom naar opleidingsniveaus in het mbo

Niveau mbo opleiding 2011-2012	Basisberoepsgerichte leerweg		Kaderberoepsgerichte leerweg		Plaatsing
	Aantal	Percentage	Aantal	Percentage	
Assistentenopleiding (niveau 1)	1.285	5,4	321	1,2	Niet succesvol
Basisberoepsopleiding (niveau 2)	18.118	76,4	5.390	20,0	Succesvol
Vakopleiding (niveau 3)	805	3,4	10.269	38,1	Succesvol
Middenkader/specialist (niveau 4)	139	0,6	8.990	33,4	Succesvol
Totaal	20.347	85,8	24.970	92,7	

Eventuele verschillen in de mate van succesvolle plaatsing tussen groepen leerlingen met een verschillende achtergrond geeft mogelijk inzicht in de mate waarin het externe rendement beïnvloed wordt door de specifieke leerlingpopulatie van een school. In Tabel 5 is weergegeven hoeveel leerlingen in het eerste jaar na de examenklas succesvol geplaatst zijn, niet succesvol geplaatst zijn of voor de bepaling van het externe rendement buiten beschouwing worden gelaten. Tevens wordt in de tabel een overzicht gegeven van de succesvolle plaatsing naar de achtergrond van de leerlingen op basis van apcg, lwoo, cumi en examenresultaat.

Opvallend is dat bl-leerlingen met leerwegondersteuning iets vaker succesvol geplaatst lijken te worden dan bl-leerlingen zonder leerwegondersteuning ($\chi^2=424,3$; $df=2$; $p<0,001$). Een dergelijk verschil is veel minder sterk gevonden voor kl-leerlingen ($\chi^2=18,8$; $df=2$; $p<0,001$). Tevens worden zowel bij basis ($\chi^2=572,1$; $df=6$; $p < 0,001$) als bij kader ($\chi^2=216,3$; $df=6$; $p<0,001$) de autochtone leerlingen vaker succesvol geplaatst dan allochtone leerlingen. Voor beide leerwegen geldt dat zowel de Westers allochtone en de niet Westers allochtone leerlingen relatief minder vaak succesvol worden geplaatst dan autochtone leerlingen. De verschillen in succesvolle plaatsing op basis van de variabele apcg laat veel grotere verschillen zien voor leerlingen afkomstig uit de basis-leerweg ($\chi^2=423,2$; $df=16$; $p<0,001$) dan voor leerlingen uit de kader-leerweg ($\chi^2=177,7$; $df=16$; $p<0,001$). De algemene trend lijkt dat er minder succesvolle plaatsingen zijn wanneer een leerling afkomstig is uit een buurt met meerdere overschrijdingen. De grootste verschillen in succesvolle plaatsing worden echter gevonden tussen leerlingen die wel of niet geslaagd zijn voor hun eindexamen vmbo. De kans op een succesvolle plaatsing is vele malen groter voor geslaagde leerlingen dan voor gezakte leerlingen of leerlingen die geen examen hebben gedaan. Deze grote verschillen tussen geslaagde en gezakte leerlingen worden zowel gevonden voor leerlingen uit de basis-leerweg ($\chi^2=6378,6$; $df=4$; $p< 0,001$) als voor leerlingen uit de kader-leerweg ($\chi^2=14.450,1$; $df=4$; $p< 0,001$).

De gevonden verschillen in de succesvolle plaatsing van leerlingen geven een mogelijke indicatie dat er naast eventuele effecten van plaatsingsbeleid van mbo-instellingen mogelijk ook andere factoren een rol spelen bij de succesvolle plaatsing van leerlingen in een vervolgopleiding (hogere leerweg of mbo2 opleiding), namelijk de achtergrond van de leerlingen. Daarom lijkt het zinvol om te verkennen in hoeverre er verschillen zijn in een indicator extern rendement met en zonder controlevariabelen. De resultaten van deze vergelijkingen zijn beschreven in 6.3.7.

Tabel 5 Overzicht van succesvol geplaatste, niet succesvol geplaatste leerlingen en leerlingen die buiten beschouwing worden gelaten één jaar na de examenklas (2011/2012) voor de leerwegen vmbo-bl en kl

Variabele		Basisberoepsgerichte leerweg			Kaderberoepsgerichte leerweg		
		Succesvol geplaatst	Niet succesvol geplaatst	Niet in extern rendement	Succesvol geplaatst	Niet succesvol geplaatst	Niet in extern rendement
		N=19.120	N=2.958	N=1.642	N=24.775	N=1.249	N=907
		%	%	%	%	%	%
Totaal		80,6	12,5	6,9	92,0	4,6	3,4
Leerweg- ondersteu- ning	Wel	83,0	13,0	4,0	91,1	4,7	4,2
	Niet	77,4	11,8	10,8	92,3	4,6	3,1
Etnische achtergrond	Autochtoon	83,0	11,1	5,9	93,1	4,0	2,9
	Niet Westerse allochtoon	75,9	14,6	9,5	88,4	6,4	5,2
	Westerse allochtoon	78,6	14,1	7,3	89,1	7,0	4,0
	Onbekend	10,4	75,5	14,2	58,6	37,9	3,4
Armoede probleem cumulatie- gebied (Apcg)	Geen overschrijding	84,9	10,0	5,1	93,5	3,5	3,0
	Laag inkomen	82,9	8,9	8,2	93,9	3,4	2,7
	Niet Westerse allochtonen	80,1	12,2	7,7	90,8	5,5	3,7
	Uitkering	80,6	12,8	6,6	92,7	4,0	3,4
	Laag inkomen + niet Westerse allochtonen	77,8	12,5	9,7	87,9	7,6	4,5
	Laag inkomen + uitkering	72,6	21,2	6,1	91,9	5,9	2,2
	Niet Westerse allochtonen + uitkering	78,9	13,7	7,4	89,3	6,4	4,3
	Cumulatie- gebied	72,8	16,8	10,4	88,1	7,4	4,5
	Onbekend	8,2	13,5	4,5	90,8	5,2	3,9
Examen- resultaat	Geslaagd	89,1	7,1	3,8	96,8	3,1	0,0
	Gezakt	24,6	45,1	30,2	33,8	15,3	50,9
	Geen examen gedaan	34,6	42,6	22,8	36,4	40,0	23,6

Doorstroom van leerlingen uit Vmbo-bl en kl naar leerjaar 2012-2013

De doorstroom van de leerlingen uit examenjaar 2010-2011 naar de verschillende schoolsoorten in het schooljaar 2012-2013 wordt in Tabel 6 weergegeven. Dit betreft het tweede schooljaar nadat de leerlingen in de examenklas hebben gezeten. Uit de tabel blijkt dat de uitstroom van zowel basis- als kader-leerlingen twee jaar nadat ze in de examenklas hebben gezeten wederom zeer divers is. Voor beide leerwegen geldt echter dat verreweg de grootste groep leerlingen hun onderwijsloopbaan vervolgt in het mbo. Dit geldt voor 83,5% van de bl-leerlingen en zelfs 92,1% van de kl-leerlingen. Deze percentages liggen iets lager dan de percentages leerlingen die in het beroepsonderwijs waren ingeschreven één jaar na de examenklas.

Het aantal leerlingen dat aangemerkt wordt als voortijdig schoolverlater is twee jaar na de examenklas opgelopen. Van de bl-leerlingen werd één jaar na het examenjaar 6,5% aangemerkt als voortijdig schoolverlater, terwijl twee jaar na de examenklas dit is opgelopen 14,6%. Ook het aantal schoolverlaters onder de kl-leerlingen twee jaar na de examenklas (7,4%) lijkt verdubbeld ten opzichte van één jaar na de examenklas (3,3%). Tenslotte zijn er wederom een aantal leerlingen met opvallende doorstroomgegevens, namelijk naar het hoger onderwijs of schoolverlater met startkwalificatie

Tabel 6 Doorstroom van vmbo-bl/kl-leerlingen naar onderwijssectoren in schooljaar 2012-2013

Sector vervolg schoolloopbaan 2012-2013	Basisberoepsgerichte leerweg		Kaderberoepsgerichte leerweg	
	N	%	N	%
Speciaal Onderwijs	116	0,5	51	0,2
Voortgezet Onderwijs	303	1,3	61	0,2
Middelbaar Beroeps Onderwijs	19.801	83,5	24.791	92,1
Volwassenen Educatie	28	0,1	39	0,1
Hoger Onderwijs	1	0,0	1	0,0
Schoolverlater met start kwalificatie	384	1,6	123	0,5
Vsv met vmbo diploma	2.088	8,8	1.567	5,8
Vsv met eerder vmbo diploma	3	0,0	2	0,0
Vsv met mbo1	275	1,2	32	0,1
Vsv	721	3,0	264	1,0
Totaal	23.720		26.931	

Zoals eerder beschreven is de doorstroom van vmbo-bl/kl-leerlingen succesvol wanneer ze onderwijs gaan volgen in een mbo-opleiding van niveau 2 of hoger. Daarnaast is ook een inschrijving in een hogere leerweg in het vmbo aangemerkt als een succesvolle plaatsing. Uit de vorige tabel bleek al dat de grootste groepen leerlingen afkomstig uit een basis- of kaderleerweg hun onderwijsloopbaan voortzetten in het mbo. In Tabel 7 wordt daarom een verder onderscheid gemaakt naar het niveau van de vervolgopleiding in het mbo.

Tabel 7 Doorstroom van vmbo-bl/kl-leerlingen naar opleidingsniveaus in het mbo

Niveau mbo opleiding 2011-2012	Basisberoepsgerichte leerweg		Kaderberoepsgerichte leerweg		Plaatsing
	N	%	N	%	
Assistentenopleiding (niveau 1)	578	2,4	145	0,5	Niet succesvol
Basisberoepsopleiding (niveau 2)	16.754	70,6	6.069	22,5	Succesvol
Vakopleiding (niveau 3)	1.988	8,4	9.890	36,7	Succesvol
Middenkader/specialist (niveau 4)	481	2,0	8.687	32,3	Succesvol
Totaal	19.801	83,5	24.791	92,1	

Uit deze tabel blijkt dat voor beide leerwegen geldt dat het grootste deel van de leerlingen succesvol in het mbo wordt geplaatst. Voor de bl-leerlingen geldt dat twee jaar na de examenklas de meerderheid van de leerlingen ingeschreven staat in een niveau 2 mbo-opleiding (70,6%). Dit is iets minder dan één jaar na de examenklas, maar door de gevonden toename van het aantal voortijdig schoolverlaters is dit te verwachten. De grootste afname is te zien in het aantal leerlingen in een niveau-1-opleiding. Eén jaar na de examenklas betrof deze groep 5,4 % van alle leerlingen en twee jaar na de examenklas betreft dit nog maar 2,4% van de leerlingen. Op basis van deze tabellen kan niet vastgesteld worden waardoor deze afname veroorzaakt wordt. Hoewel het voor de hand ligt dat het een combinatie betreft van voortijdig schoolverlaten en doorstroom naar opleidingen op niveau 2. Voor de kaderleerlingen zijn er twee jaar na de examenklas geen grote verschuivingen opgetreden ten opzichte van één jaar na de examenklas.

In Tabel 8 is weergegeven hoeveel leerlingen succesvol geplaatst zijn, niet succesvol geplaatst zijn of voor de bepaling van het externe rendement buiten beschouwing worden gelaten voor de overgang naar het schooljaar 2012-2013. Tevens geeft de tabel een overzicht van de succesvolle plaatsing van subgroepen van leerlingen op basis van hun achtergrond.

Tabel 8 Overzicht van succesvol geplaatste, niet succesvol geplaatste leerlingen en leerlingen die buiten beschouwing worden gelaten twee jaar na de examenklas (2012/2013) voor de leerwegen vmbo-bl en kl

Variabele		Basisberoepsgerichte leerweg			Kaderberoepsgerichte leerweg		
		Succesvol geplaatst	Niet succesvol geplaatst	Niet in extern rendement	Succesvol geplaatst	Niet succesvol geplaatst	Niet in extern rendement
		N=18.600	N=3.471	N=1.649	N=24.097	N=1.924	N=910
		%	%	%	%	%	%
Totaal		78,4	14,6	7,0	89,5	7,1	3,4
Leerweg- ondersteu- ning	Wel	81,1	14,9	4,0	89,2	6,7	4,2
	Niet	74,9	14,2	10,8	89,6	7,3	3,1
Etnische achtergrond	Autochtoon	80,6	13,5	5,9	90,8	6,3	2,9
	Niet Westerse allochtoon	74,5	16,0	9,5	85,3	9,4	5,2
	Westerse allochtoon	75,2	17,3	7,5	85,1	10,9	4,0
	Onbekend	11,3	73,6	15,1	51,7	44,8	3,4
Armoede probleem cululatie- gebied (Apcg)	Geen overschrijding	83,1	11,7	5,2	91,3	5,7	3,0
	Laag inkomen	82,9	8,9	8,2	93,9	3,4	2,7
	Niet Westerse allochtonen	76,2	16,1	7,7	87,8	8,5	3,7
	Uitkering	77,7	15,7	6,6	89,7	7,0	3,4
	Laag inkomen + niet Westerse allochtonen	71,6	18,7	9,7	84,8	10,8	4,5
	Laag inkomen + Uitkering	70,1	23,7	6,3	89,0	8,8	2,2
	Niet Westerse allochtonen + uitkering	75,5	17,0	7,5	85,6	10,0	4,4
	Cumulatie- gebied	71,1	18,4	10,4	85,1	10,4	4,5
	Onbekend	80,2	15,3	4,5	85,0	11,1	3,9
Examen- resultaat	Geslaagd	85,5	10,7	3,8	93,8	6,2	0,0
	Gezakt	37,0	32,6	30,4	37,4	11,6	51,0
	Examenklas	38,2	38,9	22,9	40,8	35,4	23,8

Op basis van de achtergrond van leerlingen zijn er een aantal opvallende verschillen in de plaatsing van leerlingen. Deze verschillen komen grotendeels overeen met de verschillen in succesvolle plaatsing die gevonden zijn tussen groepen leerlingen op basis van hun achtergrond één jaar na de examenklas. Wederom worden bl-leerlingen met leerwegondersteuning iets vaker succesvol geplaatst dan bl-leerlingen zonder leerwegondersteuning ($\chi^2=422,9$; $df=2$; $p<0,001$). Voor kader geldt dat de verschillen in succesvolle plaatsing tussen leerlingen met en zonder leerwegondersteuning aanzienlijk kleiner zijn ($\chi^2=22,0$; $df=2$; $p<0,001$). Echter, door de grootte van de steekproef is ook hier sprake van een statistisch significant verschil in succesvolle plaatsing. Ook vergelijkbaar met de plaatsing na één jaar zijn de verschillen in succesvolle plaatsing van leerlingen op basis van cumi. Zowel bij de bl ($\chi^2=457,7$; $df=6$; $p<0,001$) als bij de kl ($\chi^2=228,7$; $df=6$; $p<0,001$) worden de autochtone leerlingen vaker succesvol geplaatst dan allochtone leerlingen. Het blijkt dat zowel de Westerse als de niet Westerse allochtone leerlingen minder vaak succesvol worden geplaatst dan de autochtone leerlingen.

Evenals de plaatsing van leerlingen één jaar na de examenklas zijn de verschillen in succesvolle plaatsing twee jaar na de examenklas op basis van de variabele apcg veel groter voor leerlingen afkomstig uit bl ($\chi^2=425,5$; $df=16$; $p<0,001$) dan voor leerlingen uit kl ($\chi^2=201,2$; $df=16$; $p<0,001$). Naarmate er meer overschrijdingen zijn in een buurt neemt het relatieve aantal succesvolle plaatsingen af. Wederom worden de grootste verschillen in succesvolle plaatsing gevonden tussen leerlingen die wel of niet geslaagd zijn voor hun eindexamen vmbo. De kans op een succesvolle plaatsing is vele malen groter voor geslaagde leerlingen dan voor gezakte leerlingen of leerlingen die geen examen hebben gedaan. Deze grote verschillen worden zowel gevonden voor leerlingen uit de basis-leerweg ($\chi^2=4320,8$; $df=4$; $p<0,001$) als uit de kader-leerweg ($\chi^2=12,955,8$; $df=4$; $p<0,001$).

Vergelijken van verschillende indicatoren van extern rendement

In hoofdstuk 6 zijn drie verschillende indicatoren voor het externe rendement van vestigingen in het VO beschreven. De drie indicatoren verschillen in de mate waarin ze rekening houden met vertekeningen door plaatsingsbeleid van mbo-instellingen en de hiërarchische structuur van de data. Elke indicator kan leiden tot een andere schatting van het externe rendement van een vestiging. In deze paragraaf zullen op vestigingsniveau de verschillen tussen de indicatoren onderzocht worden door te kijken naar verschuivingen in percentielcategorieën volgens de bolletjes-systematiek. In 6.3.5 zijn door middel van correlaties de drie indicatoren onderling vergeleken.

Een eerste vergelijking die kan worden gemaakt is een vergelijking tussen de huidige indicator voor extern rendement en een indicator op basis van meerniveau modellen. In Tabel 9 is weergegeven hoe deze twee indicatoren zich tot elkaar verhouden, zowel voor basis- als kaderleerwegen en zowel voor het externe rendement één jaar en twee jaar na de examenklas. De cellen op de diagonaal (licht grijs) geven overeenstemming weer in de indeling van

vestigingen in bolletjes tussen de twee indicatoren voor het externe rendement. Vestigingen in deze cellen krijgen voor beide indicatoren een identieke beoordeling in termen van de bolletjes. In alle cellen buiten de diagonaal is er geen overeenstemming tussen de indicatoren in de indeling van vestigingen in bolletjes.

Tabel 9 Vergelijking van de schattingen van het externe rendement van VO-vestigingen op basis van het huidige en meerniveau model

Basisberoepsgerichte leerweg 2011/2012 (N=458)						Kaderberoepsgerichte leerweg 2011/2012 (N=484)					
Percentielen meerniveau model						Percentielen meerniveau model					
Percentielen huidige model	<10	10 - 25	25- 75	75 - 90	>90	Percentielen huidige model	<10	10 - 25	25- 75	75 - 90	>90
<10	24	7	2	0	0	<10	15	14	12	0	0
10 – 25	20	24	25	2	0	10 - 25	19	22	30	2	1
25 – 75	1	38	161	41	11	25 - 75	14	36	123	47	33
75 – 90	0	0	33	14	28	>75 *	0	0	77	24	15
>90	0	0	8	12	7						
Basisberoepsgerichte leerweg 2012/2013 (N=451)						Kaderberoepsgerichte leerweg 2012/2013 (N=483)					
Percentielen meerniveau model						Percentielen meerniveau model					
Percentielen huidige model	<10	10 - 25	25- 75	75 - 90	>90	Percentielen huidige model	<10	10 - 25	25- 75	75 - 90	>90
<10	16	6	10	1	1	<10	11	6	19	1	1
10 – 25	16	17	30	6	0	10 - 25	14	8	44	6	5
25 – 75	13	39	131	34	28	25 - 75	23	42	106	51	30
75 – 90	0	5	40	16	14	75 - 90	0	3	31	11	12
>90	0	0	15	10	10	>90	0	13	42	3	1

*Meer dan 10% van de VO-vestigingen behaalt een extern rendement van 100%

Voor het externe rendement één jaar na de examenklas van leerlingen afkomstig uit de bl-leerweg is de overeenstemming in bolletjes 50,2% tussen de huidige indicator en een indicator op basis van een meerniveau model. De associatie van de indelingen in bolletjes tussen de beide indicatoren is zwak maar statistisch significant ($\kappa=0,25$; $n=458$; $p<0,001$). De overeenstemming is iets minder groot voor het externe rendement één jaar na de examenklas voor de kader-leerweg (38.0%; $\kappa=0,08$; $n=484$; $p=0,003$) en voor het externe rendement twee

jaar na de examenklas voor de basis-leerweg (42.1%; $\kappa=0,11$; $n=451$; $p<0,001$). De laagste overeenstemming in indeling in bolletjes wordt gevonden voor het externe rendement twee jaar na de examenklas voor de kader-leerweg (28,4%; $\kappa=-0,058$; $n=483$; $p=0,023$). In dit laatste geval is er zelfs sprake van een negatieve waarde voor Kappa. Dit betekent dat de gevonden overeenstemming in de indeling in bolletjes tussen de beide indicatoren lager is dan wat op basis van toeval verwacht mag worden. Deze gevonden percentages overeenstemmingen tussen de modellen zijn relatief laag. Daarbij moet worden opgemerkt dat het aantal vestigingen die meer dan 1 bolletje verschillen tussen de beide indicatoren zeer beperkt is.

De verschillen tussen de twee indicatoren zijn te herleiden naar een aantal verschillen. In de eerste plaats betreft dit de manier van vaststellen van het externe rendement. De huidige indicator betreft het percentage leerlingen dat succesvol is uitgestroomd. In het meerniveau model wordt geschat wat de kans op een succesvolle uitstroom is voor leerlingen in een bepaalde vestiging, waarbij rekening gehouden wordt met de hiërarchische structuur van de data. Een tweede verschil zit in de selectie van leerlingen die gebruikt zijn voor de bepaling van de beide indicatoren. In de indicator op basis van het meerniveau model zijn alleen leerlingen meegenomen die doorgestroomd zijn naar een mbo-opleiding. Deze selectie is toegepast om een valide vergelijking te kunnen maken tussen het meerniveau model en het kruisklassificatie model. Hierdoor zijn leerlingen die een succesvolle overstap naar een hogere leerweg in het VO of HO niet in de bepaling meegenomen en ook zijn de leerlingen die voortijdig schoolverlater zijn niet in de analyse opgenomen. Ook dit kan tot verschillen in toekenning van bolletjes leiden tussen de twee indicatoren.

Een tweede vergelijking die kan worden gemaakt tussen de extern rendement indicatoren is de vergelijking tussen het externe rendement uit een strikt hiërarchisch meerniveau model (leerlingen binnen VO-vestigingen) en het externe rendement uit een kruisklassificatie model (leerlingen binnen zowel VO-vestiging als mbo-instelling). De resultaten van deze tweede vergelijking zijn gepresenteerd in Tabel 10. Voor deze vergelijking geldt dat de beide indicatoren voor het externe rendement gebaseerd zijn op precies dezelfde selectie van leerlingen. Verschillen in toekenning tussen de beide indicatoren kunnen daardoor direct worden toegeschreven aan verschillen in de specificatie van het model, namelijk het al dan niet meenemen van de mbo-instellingen en eventuele invloeden daarvan op succesvolle plaatsing.

In z'n algemeenheid kan gesteld worden dat de overeenstemming tussen het meerniveau model en het kruisklassificatie model in de indeling in bolletjes minder sterk is dan de overeenstemming die eerder werd gevonden tussen de huidige indicator en het meerniveau model (ondanks de verschillen in selectie). Dit geldt voor alle vergelijkingen in Tabel 10 met uitzondering van het externe rendement twee jaar na de examenklas voor de kader-leerweg (29.8%; $\kappa=-0,026$; $n=483$; $p=0,315$). Ondanks dat de overeenstemming een fractie sterker is,

is de bijbehorende waarde voor Kappa nog steeds negatief en niet significant afwijkend van nul. Voor de overige vergelijkingen worden de volgende overeenstemmingen gevonden: bl na één jaar 33,4% ($\kappa=0,028$; $n=458$; $p=0,291$), bl na twee jaar 32,4% ($\kappa=0,002$; $n=451$; $p=0,935$), kl na één jaar 32,0% ($\kappa=0,008$; $n=484$; $p=0,761$).

Tabel 10 Vergelijking van de schattingen van het externe rendement van VO-vestigingen op basis van het meerniveau en kruisklassificatie model

Basisberoepsgerichte leerweg 2011/2012 (N=458)						Kaderberoepsgerichte leerweg 2011/2012 (N=484)					
Percentielen						Percentielen					
kruisklassificatie model						kruisklassificatie model					
Percentielen meerniveau model	<10	10 - 25	25- 75	75 - 90	>90	Percentielen meerniveau model	<10	10 - 25	25- 75	75 - 90	>90
<10	11	4	18	8	4	<10	4	6	25	6	7
10 – 25	5	12	37	6	9	10 - 25	5	12	33	14	8
25 – 75	20	34	118	36	21	25 - 75	26	34	123	35	24
75 – 90	5	9	37	9	3	75 - 90	9	8	40	11	5
>90	4	10	19	10	3	>90	4	12	21	7	5
Basisberoepsgerichte leerweg 2012/2013 (N=451)						Kaderberoepsgerichte leerweg 2012/2013 (N=483)					
Percentielen						Percentielen					
kruisklassificatie model						kruisklassificatie model					
Percentielen meerniveau model	<10	10 - 25	25- 75	75 - 90	>90	Percentielen meerniveau model	<10	10 - 25	25- 75	75 - 90	>90
<10	4	10	23	3	5	<10	5	5	26	7	5
10 – 25	7	13	29	12	6	10 - 25	9	6	41	9	7
25 – 75	24	26	114	40	22	25 - 75	26	42	116	35	23
75 – 90	3	13	36	7	8	75 - 90	5	11	39	10	7
>90	7	5	24	5	5	>90	3	8	20	11	7

Wederom zijn de gevonden overeenstemming maar zeer beperkt en in geen van de gevallen is er sprake van een statistisch significante associatie in indeling in bolletjes. De gevonden overeenstemming tussen de indeling in bolletjes van de beide indicatoren is mogelijk een artefact van de grote groep vestigingen in het middelste bolletje (50% van de vestigingen). De gevonden percentages zijn amper hoger dan je op basis van toeval zou verwachten. Uit de onderstaande kruistabellen kan worden opgemaakt dat bij de vergelijking tussen extern

rendement op basis van het meerniveau model en het kruisklassificatie model er veel meer vestigingen worden gevonden waarbij de indeling meer dan 1 bolletje verschilt.

Een derde vergelijking die kan worden gemaakt betreft de vergelijking tussen de huidige indicator voor extern rendement en de indicator op basis van een kruisklassificatie mode. Ook bij deze vergelijking moet worden opgemerkt dat er zowel een verschil is in de bepaling van het externe rendement als een verschil in de gebruikte selectie van leerlingen. De overeenstemming in de indeling van bolletjes tussen deze twee indicatoren is weergegeven in Tabel 11.

Tabel 11 Vergelijking van de schattingen van het externe rendement van VO-vestigingen op basis van het huidige en kruisklassificatie model

Basisberoepsgerichte leerweg 2011/2012 (N=458)						Kaderberoepsgerichte leerweg 2011/2012 (N=484)					
Percentielen						Percentielen					
kruisklassificatie model						kruisklassificatie model					
Percentielen huidige model	<10	10 - 25	25- 75	75 - 90	>90	Percentielen huidige model	<10	10 – 25	25- 75	75 - 90	>90
<10	8	3	12	6	4	<10	7	5	19	4	6
10 - 25	3	13	35	13	7	10 - 25	7	9	31	18	9
25 - 75	25	39	127	33	28	25 - 75	19	39	139	38	18
75 - 90	6	9	41	13	6	>75*	15	19	53	13	16
>90	3	5	14	4	1						
Basisberoepsgerichte leerweg 2012/2013 (N=451)						Kaderberoepsgerichte leerweg 2012/2013 (N=483)					
Percentielen						Percentielen					
kruisklassificatie model						kruisklassificatie model					
Percentielen huidige model	<10	10 - 25	25- 75	75 - 90	>90	Percentielen huidige model	<10	10 – 25	25- 75	75 - 90	>90
<10	2	7	18	4	3	<10	4	6	21	5	2
10 - 25	8	12	33	10	6	10 - 25	5	12	42	11	7
25 - 75	23	33	134	35	20	25 - 75	31	31	123	38	29
75 - 90	8	14	31	11	11	75 - 90	4	12	22	11	8
>90	4	1	10	7	6	>90	4	11	34	7	3

*Meer dan 10% van de VO-vestigingen behaalt een extern rendement van 100%

Ook voor deze vergelijking geldt dat de overeenstemming in het algemeen lager ligt dan de overeenstemming tussen de huidige indicator en het meerniveau model. Dit geldt voor alle vergelijkingen in de onderstaande tabel met uitzondering van het externe rendement twee jaar na de examenklas voor de kader-leerweg (31,7%; $\kappa=-0,009$; $n=483$; $p=0,719$). Voor de overige vergelijkingen worden de volgende overeenstemmingen gevonden: bl na één jaar 35,4% ($\kappa=0,026$; $n=458$; $p=0,320$), kl na één jaar 34,7% ($\kappa=0,027$; $n=484$; $p=0,290$), bl na twee jaar 36,6% ($\kappa=0,49$; $n=451$; $p=0,070$). Wederom zijn de gevonden overeenstemming maar zeer beperkt en in geen van de gevallen is er sprake van een statistisch significante associatie in indeling in bolletjes.