



## UvA-DARE (Digital Academic Repository)

### Verkenning leerwinst als indicator voor onderwijskwaliteit

Roeleveld, J.; van der Veen, I.; Ledoux, G.

**Publication date**

2009

**Document Version**

Final published version

[Link to publication](#)

**Citation for published version (APA):**

Roeleveld, J., van der Veen, I., & Ledoux, G. (2009). *Verkenning leerwinst als indicator voor onderwijskwaliteit*. (SCO-rapport; No. 815). SCO-Kohnstamm Instituut. <http://www.sco-kohnstammstituut.uva.nl/pdf/sco815.pdf>

**General rights**

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), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

# **Verkenning leerwinst als indicator voor onderwijskwaliteit**

Jaap Roeleveld  
Ineke van der Veen  
Guske Ledoux

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Roeleveld, J., Veen I. van der, Ledoux, G.

Verkenning leerwinst als indicator voor onderwijskwaliteit. Jaap Roeleveld, Ineke van der Veen, Guuske Ledoux. Amsterdam: SCO-Kohnstamm Instituut van de Faculteit der Maatschappij- en Gedragswetenschappen, Universiteit van Amsterdam (SCO-rapport nr. 815, projectnummer 40330).

ISBN 978-90-6813-874-0

Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvuldigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen, of op enige manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher.

Uitgave en verspreiding:

SCO-Kohnstamm Instituut

Nieuwe Prinsengracht 130, Postbus 94208, 1090 GE Amsterdam

Tel.: 020-525 1201

[www.sco-kohnstamminstituut.uva.nl](http://www.sco-kohnstamminstituut.uva.nl)

© Copyright SCO-Kohnstamm Instituut, 2008

## Inhoudsopgave

<b>1</b>	<b>Inleiding</b>	1
1.1	Beoordelingen van de opbrengsten van onderwijs	1
1.2	Opbrengsten van scholen	2
1.3	Indicator van het onderwijsstelsel	4
1.4	Opzet van dit rapport	8
1.5	Data	9
<b>2</b>	<b>Leerwinst op Taal en Rekenen in PRIMA</b>	13
2.1	Ontwikkeling van taal tussen groep 4 en groep 8	14
2.2	Ontwikkeling van rekenen/wiskunde tussen groep 4 en groep 8	20
2.3	Enkele conclusies	24
<b>3</b>	<b>Groeimodellen voor Taal en Rekenen in PRIMA</b>	27
3.1	Analyse ontwikkeling taal- en rekenprestaties door middel van groeimodellen	27
3.2	Groeimodellen voor Taal	30
3.3	Groeimodellen voor Rekenen	38
<b>4</b>	<b>Samenvattende terugblik</b>	45
4.1	Opbrengsten bij Taal	45
4.2	Opbrengsten bij Rekenen	47
<b>5</b>	<b>Mogelijkheden en beperkingen van COOL</b>	51
5.1	Algemene opzet	51
5.2	Representativiteit	53
5.3	Uitval	54
5.4	Toetsen	55
<b>6</b>	<b>Slotbeschouwing</b>	59

<b>Literatuur</b>	63
<b>Recent uitgegeven SCO rapporten</b>	65

# 1 Inleiding

In deze verkenning bouwen we voort op de notitie “*Over leerwinst als stelselindicator*” (Jules. L. Peschar, 2007). Daarin wordt ingegaan op verschillende manieren om de opbrengsten van het onderwijs te beoordelen. Als inleidend kader voor de verkenningen in dit rapport zullen we de diverse begrippen, in onze eigen woorden, proberen samen te vatten.

## 1.1 Beoordelingen van de opbrengsten van onderwijs

Uitgangspunt bij de beoordeling van de opbrengsten van onderwijs zijn de **resultaten van individuele leerlingen**. Daar zijn in principe twee varianten van:

- de *prestatie* van een leerling op een bepaald moment (een ‘bruto’ onderwijsresultaat) we kunnen daarbij bijvoorbeeld denken aan:
  - de score op de eindtoets basisonderwijs; of
  - het wel of niet behalen van een referentieniveau voor taal en rekenen (zie Meijerink); of
  - in het voortgezet onderwijs: de behaalde onderwijspositie (volgens de zgn. leer jarenladder) na 3 jaar voortgezet onderwijs;
  - enz.
- de *leerwinst* van een leerling op een bepaald moment (een ‘netto’ onderwijsresultaat) hierbij wordt uitgegaan van een beginmeting en wordt gekeken hoeveel de leerling vooruit is gegaan tussen de beginmeting en de prestatie aan het eind. We kunnen dan denken aan:
  - het verschil in vaardigheidsscore voor taal en rekenen in groep 8 met de score op dezelfde schaal in groep 4 (vgl. paragraaf 2 van dit rapport); of

### *Verkenning leerwinst als indicator voor onderwijskwaliteit*

- in het voortgezet onderwijs: het verschil tussen de behaalde onderwijspositie (volgens de leerjarenladder) minus het advies voor vo van de basisschool (vgl. ORD-paper van de Inspectie: Vreeburg, 2008);
- enz.

Bij de tweede variant wordt er dus rekening mee gehouden dat sommige leerlingen het onderwijs (of precieser: een bepaalde fase van het onderwijs) binnenkomen met meer 'bagage' dan andere leerlingen.

Beide varianten van deze individuele resultaten kunnen vervolgens gebruikt worden ter beoordeling van de *opbrengsten van scholen* én als *indicator van het onderwijsstelsel* als geheel.

## **1.2 Opbrengsten van scholen**

Hierbij gaat het meestal om te bepalen welke scholen het duidelijk 'beter doen' dan gemiddeld en welke scholen het duidelijk 'slechter doen' dan gemiddeld. Het gaat dus om een relatieve waardering.

Uitgaand van de (individuele) *prestaties* van de leerlingen kan een samenvattende maat per school bepaald worden; bijvoorbeeld

- het gemiddelde op de eindtoets basisonderwijs;
- het percentage leerlingen dat het minimum- of het streefniveau voor taal of rekenen behaalt;
- (in vo): de gemiddelde onderwijspositie na 3 jaar voortgezet onderwijs
- enz.

Deze 'output-maten' geven een soort absoluut beeld van de school, zonder dat rekening gehouden wordt met de (soms zeer grote) verschillen tussen scholen in hun leerlingbevolking.

Voor een 'eerlijke' vergelijking van scholen worden daarom 'value added models' toegepast, waarin langs statistische weg wél rekening kan worden gehouden met die verschillen. In feite wordt er dan, op basis van een aantal leerlingkenmerken (zoals opleiding van de ouders, etniciteit, verblijfsduur in Nederland, thuistaal, enz.) een voorspelling gedaan over wat er van een bepaalde leerling verwacht mag worden.

Bij de vergelijking van de scholen wordt dan gekeken naar vragen als:

- is de gemiddelde score op de eindtoets basisonderwijs hoger of lager dan voorspeld kan worden op basis van een aantal kenmerken van de leerlingen in groep 8;
- is het percentage leerlingen dat het minimum- of het streefniveau voor taal of rekenen behaalt hoger of lager dan verwacht kon worden;
- is de gemiddelde onderwijspositie na 3 jaar voortgezet onderwijs hoger of lager dan voorspeld;
- enz.

Uitgaand van de (individuele) *leerwinst* van de leerlingen kan een ook maat per school bepaald worden. Bijvoorbeeld

- de gemiddelde leerwinst voor taal en rekenen tussen groep 4 en groep 8
- de gemiddelde vooruitgang op de leerjarenladder tussen het advies, waarmee de leerlingen het vo binnenkomen en hun onderwijspositie na 3 jaar voortgezet onderwijs;
- enz.

Zoals gezegd is leerwinst een 'netto' onderwijsresultaat, waarbij er rekening wordt gehouden met het feit dat de ene leerling bij binnenkomst al meer kan dan de andere. Maar ook hier kunnen 'value added models' toegepast worden om tot een eerlijker vergelijking van scholen te komen. Want ook de ontwikkeling van een leerling tussen de 'beginmeting' en de 'eindmeting' wordt beïnvloed door meer of minder stimulerende factoren van buiten de school, waardoor het op de ene school makkelijker is om tot een bepaalde leerwinst te komen dan op de andere school.



### *Verkenning leerwinst als indicator voor onderwijskwaliteit*

Bij de vergelijking van de scholen wordt dan gekeken naar vragen als:

- is de gemiddelde leerwinst voor taal en rekenen hoger of lager dan voorspeld kan worden op basis van een aantal kenmerken van de leerlingen;
- is de gemiddelde vooruitgang op de leerjarenladder tussen advies en onderwijspositie na 3 jaar (onderbouwrendement) hoger of lager dan verwacht kon worden;
- enz.

### **1.3 Indicator van het onderwijsstelsel**

De belangrijkste functie van een indicator op stelselniveau is dat deze de mogelijkheid geeft om na te gaan hoe de historische ontwikkeling is: is er sprake van voor- of achteruitgang over de loop der jaren. (Terzijde: een andere functie kan zijn dat de indicator een rol speelt in internationale vergelijkingen.)

Ook op stelselniveau kan een indicator gebaseerd zijn op ('bruto') prestatie of op ('netto') leerwinst; en ook op stelselniveau kan er met rechtstreekse gemiddelden gewerkt worden of met modellen voor toegevoegde waarde.

Uitgaand van de (individuele) *prestaties* van de leerlingen kan een samenvattende maat voor het hele onderwijsstelsel bepaald worden. Bijvoorbeeld:

- het gemiddelde op de eindtoets basisonderwijs;
- het percentage leerlingen dat het minimum- of het streefniveau voor taal of rekenen behaalt;
- (in vo): de gemiddelde onderwijspositie na 3 jaar voortgezet onderwijs
- enz.

Hiermee krijgen we een maat voor de opbrengsten van het Nederlandse onderwijsstelsel op een bepaald moment. Door vergelijking van een aantal opeenvolgende momenten kan een beeld verkregen worden van de historische ontwikkelingen ("stijgt of daalt het percentage leerlingen dat het minimumniveau voor rekenen behaalt?").

Voor sommige (beleids-)vragen kan dit een uitstekende indicator zijn. Bijvoorbeeld als we willen weten of het Nederlandse basisonderwijs er in slaagt om 75% van de leerlingen op een minimumniveau voor rekenen te krijgen, ongeacht de beginsituatie of de (sociale en etnische) herkomst van die leerlingen.

Voor andere vragen is het wel nuttig om rekening te houden met de herkomst van de leerlingen. Wanneer bijvoorbeeld over een (lange) reeks van jaren de prestaties in groep 8 vergeleken worden, dan kan het nodig zijn om rekening te houden met verschuivingen in de leerlingbevolking over die reeks van jaren (zoals bv. een toename van het aandeel leerlingen uit etnische groepen; een stijging van het ouderlijk onderwijsniveau; enz.). Hiervoor kunnen ook ‘value added models’ worden toegepast; nu niet zozeer voor een ‘eerlijke’ vergelijking tussen scholen met een verschillende leerlingpopulatie, maar voor een ‘eerlijke’ vergelijking van leerlingpopulaties door de tijd.

Om een goed beeld te krijgen van wat het onderwijsstelsel presteert kan ook uitgegaan worden van de *leerwinst* van leerlingen. Er kan dan bijvoorbeeld, net als bij de scholen hiervoor maar nu op stelselniveau, gekeken worden naar:

- de gemiddelde leerwinst voor taal en rekenen tussen groep 4 en groep 8
- de gemiddelde vooruitgang op de leerjarenladder tussen het advies, waarmee de leerlingen het vo binnenkomen en hun onderwijspositie na 3 jaar voortgezet onderwijs;
- enz.

Zoals gezegd wordt met leerwinst in elk geval rekening gehouden met het feit dat leerlingen bij het ingaan van een bepaalde fase van het onderwijs verschillen in hun aanvangsniveau. Maar ook hun verdere groei na dat aanvangsniveau wordt beïnvloed door buitenschoolse factoren. Daarom kan het nuttig zijn ook hier met complexere ‘*value added models*’ te werken.

Een extra voordeel daarvan is dat er ook zicht kan komen op de ontwikkeling van specifieke groepen leerlingen (bijvoorbeeld leerlingen van bepaalde etnische her-

*Verkenning leerwinst als indicator voor onderwijskwaliteit*

komst) wat bij kan dragen tot een 'rijkere' stelselindicator dan bijv. een enkel percentage dat een referentieniveau behaalt.

Samenvattend kan, met als uitgangspunt de resultaten van individuele leerlingen ('bruto' prestaties of 'netto' leerwinst), de rechtstreekse output van een onderwijsinstelling, een onderwijssector, of een heel onderwijsstelsel bepaald worden of kan desgewenst de toegevoegde waarde van de instelling, sector of stelsel geschat worden met *value added models*. Daarmee kunnen verschillende typen vragen rond onderwijsopbrengsten beantwoord worden. We hebben geprobeerd de vraagstellingen op de verschillende niveaus in het volgende schema systematisch in beeld te brengen.

Schema 1 Vraagstellingen rond onderwijsopbrengsten

Niveau	uitgangspunt	Type vraagstelling
Individueel	prestaties	wat kan/kent een leerling op een bepaald moment?
	leerwinst	hoeveel heeft een leerling geleerd in een bepaalde periode?
Schoolniveau	prestaties	wat is het gemiddelde niveau van leerlingen op deze school op dit moment? Is dat hoger of lager dan bij andere scholen? <i>value added:</i> zijn de gemiddelde prestaties van leerlingen op deze school hoger of lager dan verwacht kon worden op basis van een aantal kenmerken van de leerlingen?
	leerwinst	hoeveel hebben de leerlingen op deze school gemiddeld geleerd in een bepaalde periode? Is dat meer of minder dan bij andere scholen? <i>value added:</i> hebben de leerlingen op deze school gemiddeld meer of minder geleerd dan verwacht kon worden op basis van een aantal kenmerken van de leerlingen?
Stelselniveau	prestaties	wat is het gemiddelde niveau van Nederlandse leerlingen op een bepaald meetmoment (bv. einde basisonderwijs; op 15-jarige leeftijd; enz.)? Is dat hoger of lager dan in voorafgaande jaren? <i>value added:</i> is het gemiddelde niveau van Nederlandse leerlingen op een bepaald meetmoment hoger of lager dan verwacht kon worden op basis van een aantal kenmerken van de leerlingen (nu en in de voorafgaande jaren)?
	leerwinst	hoeveel leren Nederlandse leerlingen in een bepaalde fase van het onderwijs (bijv. gedurende de basisschool; gedurende de onderbouw vo; enz.)? Is dat meer of minder dan in voorafgaande cohorten? <i>value added:</i> is de leerwinst van Nederlandse leerlingen in een bepaalde fase van het onderwijs meer of minder dan verwacht kon worden op basis van een aantal kenmerken van die leerlingen (nu en in de voorgaande cohorten)?

Peschar (2007) kiest op grond van zijn onderzoek van de (internationale) wetenschappelijke literatuur en OECD- en beleidsstukken voor een benadering waarin, met behulp van longitudinale data, leerwinst gebruikt kan worden als indicator voor het onderwijsstelsel. In zijn notitie komen ook een aantal randvoorwaarden aan de orde. Daaraan zullen we in het verloop van deze verkenning ook aandacht aan besteden.

We noemen hier de volgende aandachtspunten:

- a) de opbrengstmaat (toetsen, referentieniveaus, enz.) moet door de jaren heen onveranderd blijven. (of in elk geval door de jaren heen vergelijkbaar zijn)
- b) de ontwikkeling van leerlingen (groei, leerwinst) moet op één schaal uit te drukken zijn; ook deze schaal moet door de jaren heen vergelijkbaar blijven.
- c) om leerwinst te bepalen is een zo vroeg mogelijk beginmeting vereist; toetsen bij jonge kinderen is echter problematisch. (vgl. onderzoek kleutertoets; Van Schooten, Smeets & Driessen, 2007)
- d) ook leerwinst wordt beïnvloed door factoren buiten de school. (zoals aanleg, thuismilieu, financiële omstandigheden, sociaal kapitaal etc.)

## **1.4 Opzet van dit rapport**

In de eerste stap van onze verkenning (paragraaf 2) gaan we uit van een simpele, beschrijvende versie van leerwinst. Daarmee bedoelen we dat we kijken naar het verschil tussen een begin- en een eindmeting. We zullen laten zien dat een dergelijke simpele versie niet zonder meer geschikt is als stelselindicator.

Daarna gaan we complexere groeimodellen bespreken (paragraaf 3). Dit komt dichterbij de bedoeling die Peschar heeft met de stelselindicator. Zoals hij stelt:

”Tegen deze achtergrond valt de keuze positiever uit voor een indicator die gebaseerd is op leerwinst. Wanneer longitudinale leerling-informatie regelmatig beschikbaar komt en de competentiemeting bovendien op een vaste gemeenschappelijke schaal kan worden geplaatst, dan is er een “absolute” maat beschikbaar die het mogelijk maakt te spreken over de leerwinst van leerlingen.

De groei van competenties van leerlingen --leerwinst-- kan zodoende voor het stelsel als geheel zichtbaar worden gemaakt.

Daarbij kunnen interessante vragen worden beantwoord: Is de leerwinst gelijkmatig over de schoolloopbaan verdeeld of zijn er “sprongen”? Zijn er specifieke groepen die meer leerwinst vertonen dan andere? Lopen leerlingen met een achterstand deze gedurende het onderwijs in of juist niet? Op welke groepen moet het beleid zich specifiek richten bij het vergroten van de leerwinst bij degenen die hun achterstand niet inlopen?” (Peschar, 2007, p. 13)

We voeren deze verkenningen uit op de longitudinale data van het PRIMA-cohortonderzoek. Een aantal beperkingen van deze data komen in het verloop van dit rapport aan de orde.

In paragraaf 4 blikken we kort terug op de opbrengsten van de analyses uit de voorgaande paragrafen.

In de toekomst zullen gegevens afkomstig zijn van het sinds schooljaar 2007-2008 in uitvoering zijnde COOL<sup>5-18</sup> cohortonderzoek. In paragraaf 5 onderzoeken we de mogelijkheden en beperkingen die dit nieuwe cohortonderzoek zal gaan krijgen voor het werken met leerwinst als stelselindicator. Daarbij is de informatie verwerkt die wij kregen uit gesprekken met Cito en de Inspectie van het onderwijs.

We sluiten het rapport af met een slotbeschouwing in paragraaf 6.

## **1.5 Data**

Bij het onderzoeken van leerwinst zouden we het liefst de leerwinst bepalen die leerlingen gedurende de *gehele* basisschoolperiode van groep 1 tot en met 8 boeken. Hiervoor is het echter noodzakelijk om meetinstrumenten te gebruiken die in groep 1 en 8 hetzelfde meten en vertaald kunnen worden in scores op één schaal (vergelijk aandachtspunt b hiervoor).

*Verkenning leerwinst als indicator voor onderwijskwaliteit*

Binnen het PRIMA-cohortonderzoek is vanaf groep 2 gemeten. Maar pas vanaf groep 4 staan de taal- en rekenprestaties op dezelfde schaal als de prestaties in hogere groepen. We maken daarom gebruik van gegevens van groep 4 t/m 8. In PRIMA zijn dat gegevens van dezelfde leerlingen uit groep 4, 6 en 8, dus van drie meetmomenten. Daarbij is het mogelijk twee groepen leerlingen van groep 4 t/m 8 te volgen: de leerlingen die tijdens de derde PRIMA-meting in schooljaar 1998/99 in groep 4 zaten en de leerlingen die tijdens de vierde PRIMA-meting in schooljaar 2000/01 in groep 4 zaten (zie Schema 2).

*Schema 2 De onderzochte cohorten PRIMA-leerlingen*

	PRIMA-meting			
cohort:	1998/99	2000/01	2002/03	2004/05
1	4 →	6 →	8	
2		4 →	6 →	8

In deze periode zijn de taal- en rekenoetsen voor deze groepen niet veranderd (vergeleijk aandachtspunt a), terwijl ook de wijze van afnemen binnen PRIMA vanaf 1998/99 niet wezenlijk veranderd is. Hierdoor wordt het berekenen van leerwinst mogelijk.

Voor beide cohorten geldt dat we bij de start in groep 4 uitgaan van de *representatieve* steekproef van de betreffende PRIMA-meting, zodat we een te veralgemeniseren beeld van het Nederlandse basisonderwijs kunnen krijgen. (PRIMA kent ook een aanvullende steekproef, met extra scholen met veel achterstandsleerlingen; deze blijft hier dus buiten beschouwing.)

Een belangrijk probleem bij alle longitudinale (cohort)onderzoekingen is de uitval van onderzoekseenheden tussen de opeenvolgende metingen (*panelmortaliteit*).

Ook in PRIMA is er bij de hiervoor genoemde cohorten sprake van een omvangrijke uitval tussen de start in groep 4 en de eindmeting vier jaar later in groep 8. Beide co-

horten starten in groep 4 met ongeveer 10000 leerlingen. Daarvan zijn er in groep 6 nog ongeveer 5500 over en in groep 8 zo'n 3800.

Deze uitval van rond de 62% over vier schooljaren wordt vooral veroorzaakt doordat scholen besluiten niet meer mee te doen met een volgende ronde van het onderzoek (*schooluitval*); deze is goed voor 50% van de in totaal 62% uitval. De overige 12% uitval betreft uitval van leerlingen op scholen die wel blijven meedoen met het PRIMA-onderzoek (*individuele uitval*). Het gaat dan zowel om leerlingen die van reguliere basisschool veranderen (meestal vanwege verhuizing), als om leerlingen die verwezen worden naar speciaal (basis)onderwijs. Bovendien gaat het om zittenblijvers: in de opzet van PRIMA (met elke twee jaar een meting in de even groepen 2, 4, 6 en 8) zijn zittenblijvers, die tijdens de metingen in een oneven groep terecht zijn gekomen, niet gevolgd.

Er is in het PRIMA-project aanvullend onderzoek gedaan naar de selectiviteit van de uitval (zie Roeleveld & Portengen, 1998; Roeleveld & Vierke, 2003; Vierke & Mulder, 2005). Daaruit komt naar voren dat er bij de uitval van hele scholen geen sprake is van selectiviteit: het "genoeg hebben" van het meedoen met het onderzoek komt voor bij alle soorten scholen: zowel bij kansrijke als bij kansarme scholen, zowel in de grote en middelgrote steden als op het platteland, enz.

Bij de individuele uitval is er natuurlijk wel sprake van selectiviteit. Leerlingen die blijven zitten of verwezen worden naar speciaal (basis)onderwijs zijn vaker minder presterende leerlingen. En dat geldt in mindere mate ook voor leerlingen die van basisschool veranderen.

Bij analyses van de bovengenoemde twee cohorten, waarbij louter gekeken wordt naar de leerlingen die aan alle drie opeenvolgende metingen hebben deelgenomen (zoals hierna in paragraaf 2), hebben we het dus over leerlingen die een reguliere loopbaan door het basisonderwijs hebben doorlopen, zonder zittenblijven of verwijzing.



Om ook rekening te houden met de uitval uit de cohorten moeten voor de uitgevallen leerlingen schattingen gemaakt worden van hun ontbrekende scores. Dat komt nader aan de orde in paragraaf 3.

## 2 Leerwinst op Taal en Rekenen in PRIMA

In deze paragraaf vergelijken we de ontwikkeling van de taal- en rekenscores voor twee groepen leerlingen:

- cohort 1: zit in schooljaar 1998/1999 in groep 4
- cohort 2: zit twee jaar later in 2000/2001 in groep 4

Beide cohorten volgen we in groep 6 en groep 8.

Tegelijk beperken we ons in deze paragraaf tot leerlingen die in *alle* drie opeenvolgende metingen hebben meegedaan aan PRIMA.

Dat betekent dus ‘verlies’ van leerlingen op scholen die niet blijven meedoen met PRIMA. Maar vooral ook: ‘verlies’ van leerlingen die verhuizen, blijven zitten of verwezen worden naar het speciaal basisonderwijs. Zoals in de inleiding vermeld, is het eerste ‘verlies’ niet selectief, maar het tweede wel.

Met moderne schattingstechnieken kan (tot op zekere hoogte) rekening worden gehouden met dit deels selectieve verlies. Een toepassing zien we in paragraaf 3 bij de groei-modellen.

Hier gaan we, ter wille van de eenvoud van het verhaal, alleen uit van leerlingen waarvan we op alle tijdpunten gegevens hebben.

## 2.1 Ontwikkeling van taal tussen groep 4 en groep 8

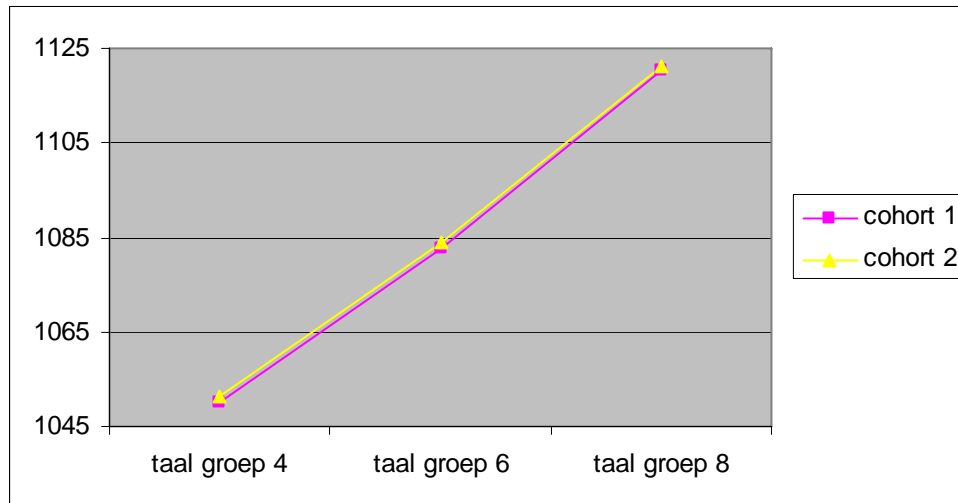
We geven eerst enkele gegevens in een tabel, die we later via grafieken nader zullen bespreken.

Tabel 1 Taalvaardigheidscores in groep 4, 6 en 8 in drie cohorten (gemiddelden)

	cohort 1	cohort 2
groep 4	1050.0	1051.5
groep 6	1082.8	1083.8
groep 8	1120.5	1121.1
leerwinst tussen 4 en 8	70.6	69.6

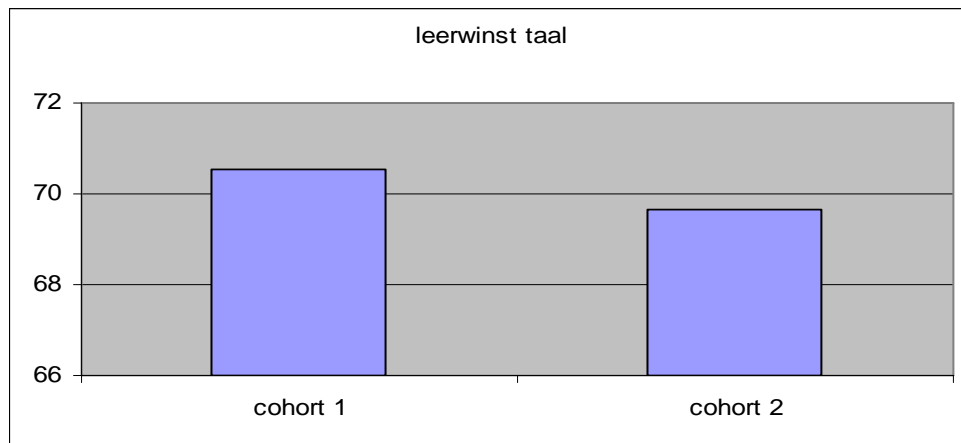
De volgende grafiek brengt deze gemiddelde ontwikkeling van de taalvaardigheidscores van Nederlandse leerlingen in beeld.

Figuur 1 Ontwikkeling taalvaardigheid in twee cohorten

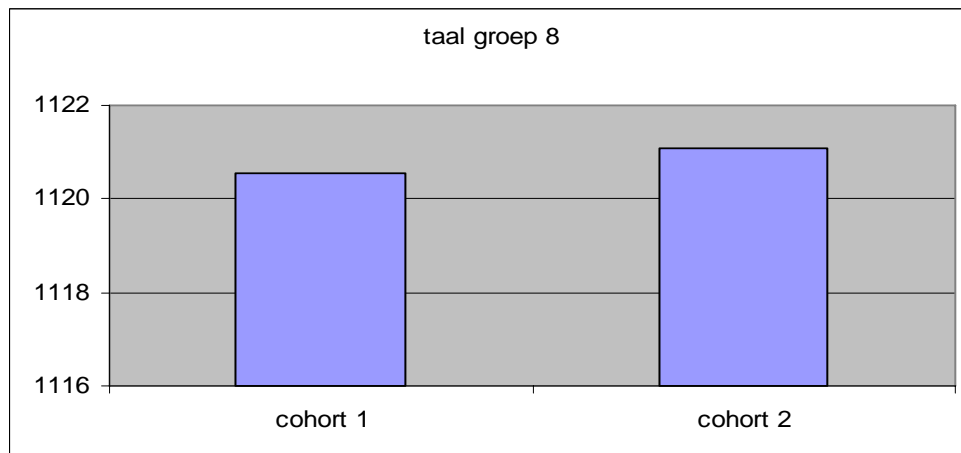


Wanneer we nu wat nader kijken naar de “leerwinsten”, de toename van de vaardigheidsscores voor Taal in de vier jaar onderwijs tussen groep 4 en groep 6, dan krijgen we het volgende beeld.

Figuur 2 Leerwinst taal in twee cohorten



Figuur 3 Taalvaardigheid in groep 8 in twee cohorten



*Verkenning leerwinst als indicator voor onderwijskwaliteit*

We zien dat de leerwinst in cohort 1 groter is dan in cohort 2 (1 punt op de taalvaardigheidschaal). Maar dat wil niet zeggen dat het onderwijsniveau gedaald is. Integendeel, in de volgende grafiek zien we dat het “eindniveau” in groep 8 juist in cohort 2 het hoogst is (ongeveer 0.5 punt hoger).

De grotere leerwinst in cohort 1 betekent dus niet dat de leerlingen daar een hoger eindniveau halen (juist niet), maar heeft meer te maken met het feit dat de beginmeting in groep 4 in dat cohort ongeveer 1.5 punt lager was dan in cohort 2.

In het PRIMA-onderzoek wordt gebruik gemaakt van steekproeven en zowel de niveaus als de leerwinsten uit Tabel 1 en de grafieken kennen daarom natuurlijk ook onzekerheidsmarges. De voorgaande uitkomsten blijken statistisch niet significant te zijn.

Toch hebben we ze hier vermeld, om te illustreren dat een hogere leerwinst in een bepaald cohort niet altijd betekent dat in dat cohort ook het hoogste eindniveau wordt bereikt. In een ander onderzoek hebben we laten zien dat er bij eerdere PRIMA-cohorten wél sprake was van een significant hogere leerwinst in het oudere cohort, terwijl het eindniveau in het jongere cohort significant hoger was. (Dit was het onderzoek naar de ontwikkeling van onderwijsachterstanden tussen 1988 en 2002 (Mulder, Roeleveld, Van der Veen & Vierke, 2005); de betreffende uitkomsten staan in hoofdstuk 3.2: *Longitudinale analyse: leerwinsten.*)

Leerwinst en eindniveau kunnen natuurlijk ook geanalyseerd worden voor verschillende groepen leerlingen. Hierna kijken we naar de taalontwikkeling bij leerlingen onderscheiden naar hun **etnische herkomst**. Daarbij gebruiken we de volgende indeling:

NL	-beide ouders in Nederland geboren
gemengd	-één van beide ouders buiten Nederland geboren
Sur/Ant	-beide ouders van Surinaams/Antilliaanse herkomst
Turks	-beide ouders van Turkse herkomst

*Leerwinst op Taal en Rekenen in PRIMA*

Marok            -beide ouders van Marokkaanse herkomst  
overig           -iedereen die niet onder het voorgaande valt

(N.B. binnen deze laatste groep is helaas geen nader onderscheid Westers vs. Niet-Westers allochtoon mogelijk)

We geven eerst twee grafieken, met achtereenvolgens de leerwinst in beide cohorten voor deze groepen leerlingen en daarna hun eindniveau in groep 8. (zie figuur 4)

Een paar opmerkingen:

Opvallend is de hoge leerwinst van de Turkse leerlingen (en ook de leerlingen van overige herkomst);

Bij sommige groepen is de leerwinst in het eerste cohort het hoogst, maar bij andere die in het tweede cohort; het algemene beeld over het hele cohort verschilt dus tussen de etnische groepen.

Vervolgens geven we de grafiek van het eindniveau. (zie figuur 5)

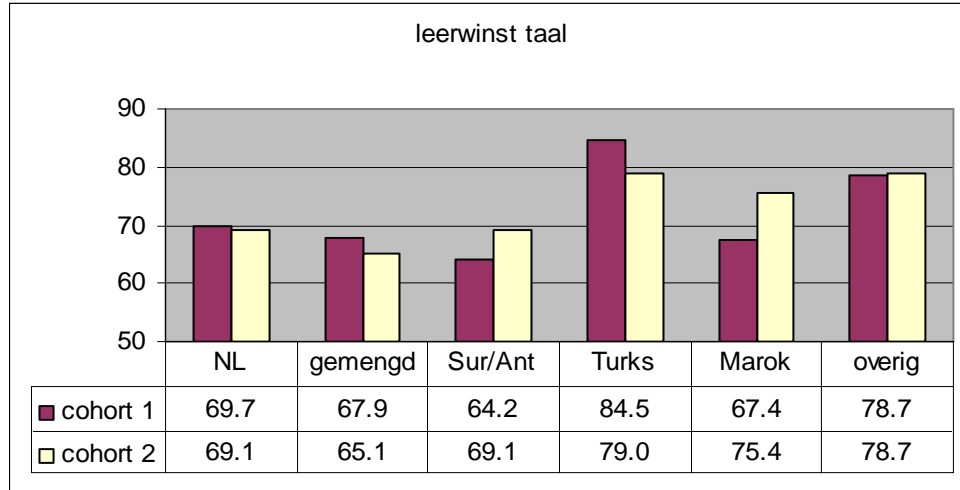
Allereerst valt op dat de Turkse leerlingen, ondanks hun relatief grote leerwinst, toch het laagste eindniveau behalen.

En ook hier: voor sommige groepen zijn de scores uit cohort 1 hoger en voor andere groepen juist de scores uit cohort 2.

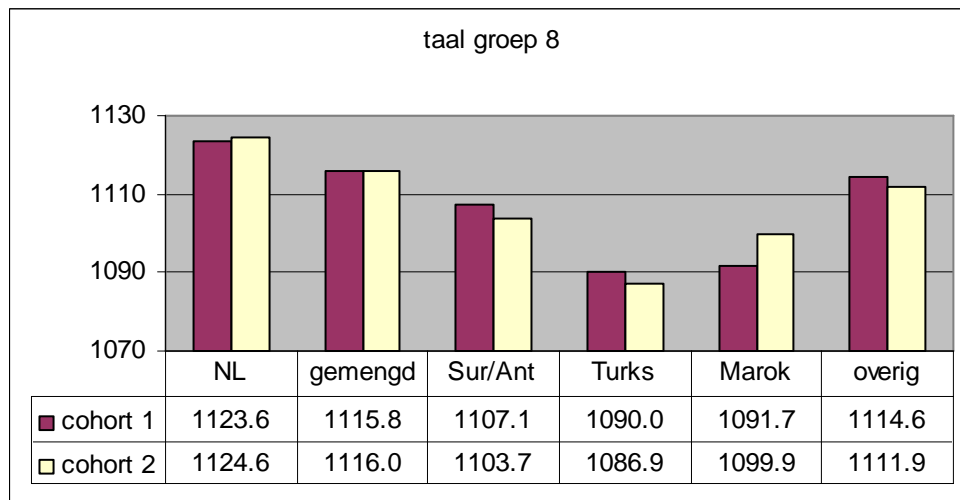
Wat betreft de Turkse leerlingen, zowel de grote leerwinst als het lage eindniveau, wordt het duidelijkst in een grafiekje met de taalontwikkeling van de leerlingen over de jaren. (zie figuur 6 en 7)

Verkenning leerwinst als indicator voor onderwijskwaliteit

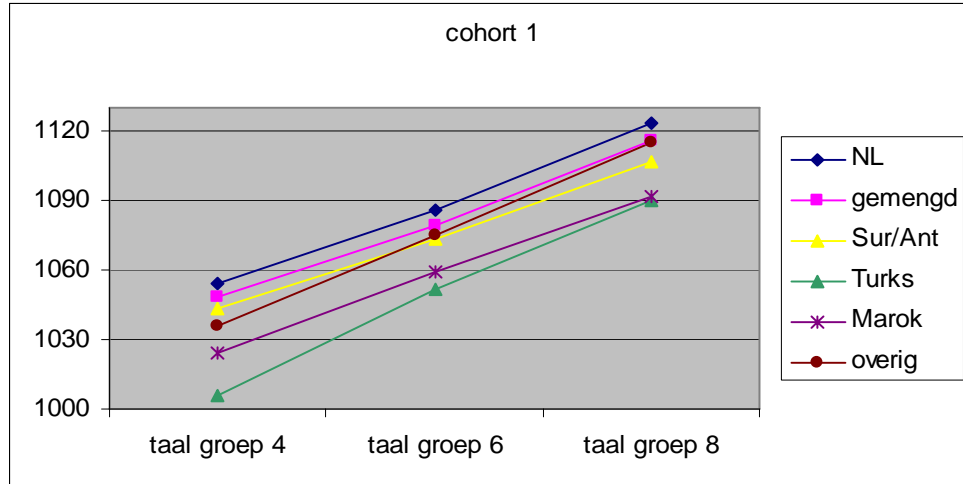
Figuur 4 Leerwinst taal in twee cohorten; naar etnische herkomst



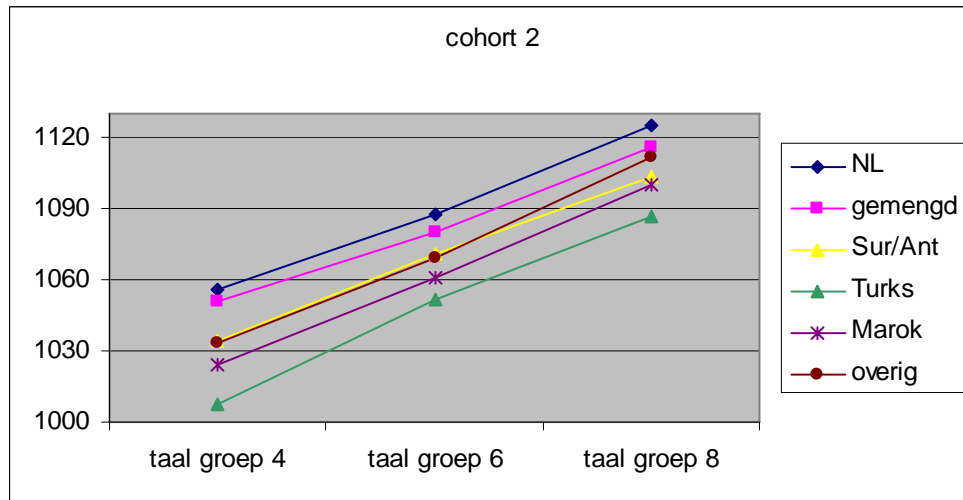
Figuur 5 Taalvaardigheid groep 8 in twee cohorten; naar etnische herkomst



Figuur 6 Ontwikkeling taalvaardigheid in cohort 1; naar etnische herkomst



Figuur 7 Ontwikkeling taalvaardigheid in cohort 2; naar etnische herkomst





## 2.2 Ontwikkeling van rekenen/wiskunde tussen groep 4 en groep 8

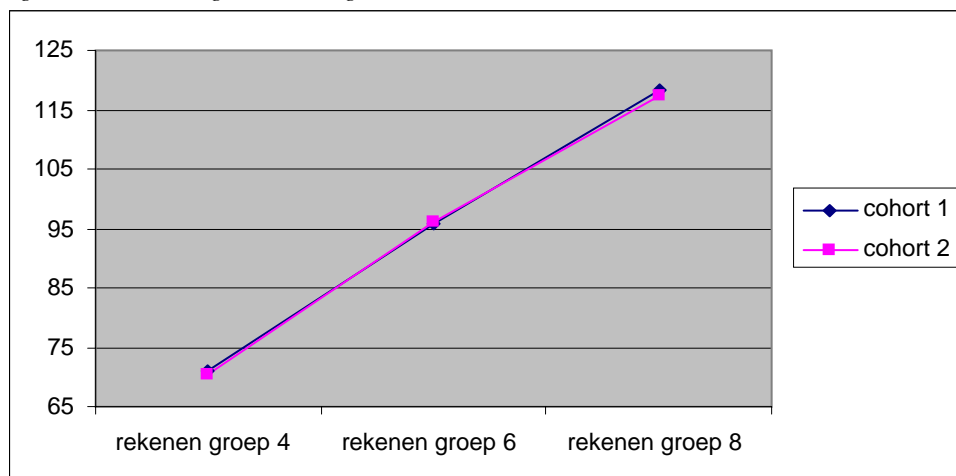
Voor de ontwikkeling van vaardigheden op gebied van rekenen/wiskunde hebben we een vergelijkbare analyse uitgevoerd. We geven eerst weer cijfers in tabelvorm.

Tabel 2 Rekenvaardigheidscores in groep 4, 6 en 8 in drie cohorten (gemiddelden)

	cohort 1	cohort 2
groep 4	70.9	70.5
groep 6	95.9	96.0
groep 8	118.2	117.4
leerwinst tussen 4 en 8	47.3	46.9

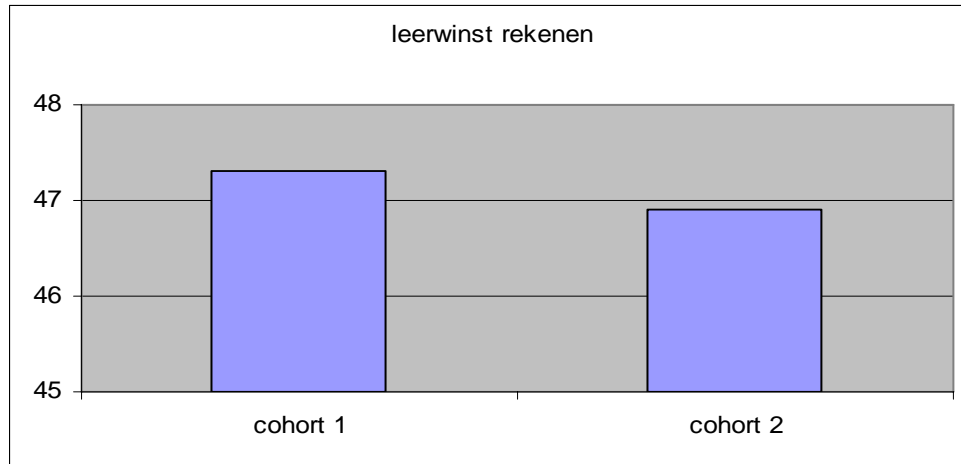
De volgende grafiek brengt de gemiddelde ontwikkeling weer in beeld.

Figuur 8 Ontwikkeling rekenvaardigheid in twee cohorten

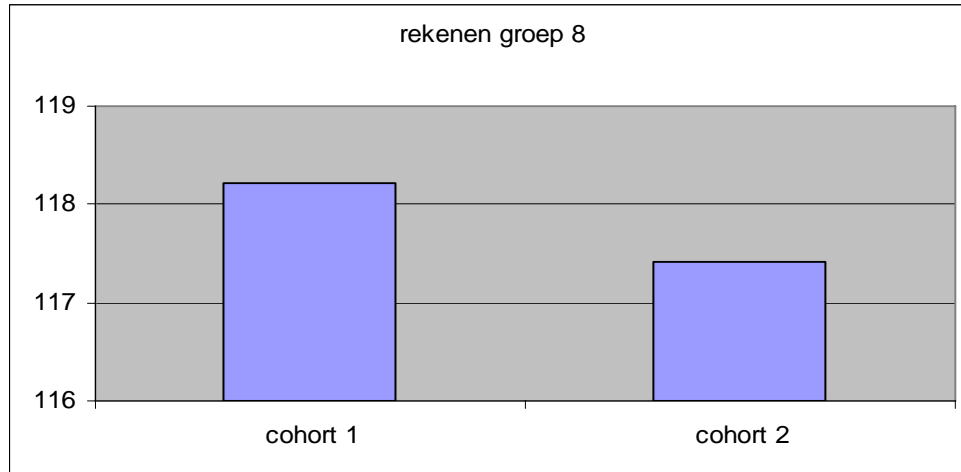


Wanneer we alleen naar de leerwinst tussen groep 4 en groep 8 kijken, dan levert dat het volgende beeld op.

*Figuur 9 Leerwinst rekenen in twee cohorten*



*Figuur 10 Rekenvaardigheid in groep 8 in twee cohorten*



*Verkenning leerwinst als indicator voor onderwijskwaliteit*

Net als bij taal zien we hier in het tweede cohort wat minder leerwinst (ongeveer 1.5 punt op de rekenvaardigheidschaal) dan in het eerste cohort.

Maar in tegenstelling tot de scores bij taal blijkt nu ook het eindniveau in het tweede cohort bijna een punt lager.

Bij rekenen is het beeld dus eenduidiger dan bij taal: hier is cohort 1 zowel qua leerwinst als qua eindniveau wat hoger dan cohort 2.

Ook hier weer een opmerking over de significantie van deze verschillen. Het verschil in eindniveau rekenvaardigheid is statistisch significant, maar het verschil in leerwinst niet.

Opnieuw kunnen leerwinst en eindniveau natuurlijk ook weer geanalyseerd worden voor verschillende groepen leerlingen. Voor rekenen kijken we naar de ontwikkeling in twee cohorten tussen leerlingen, onderscheiden naar het **opleidingniveau** van hun ouders:

max lo -de hoogste opleiding van beide ouders is (maximaal) alleen lager onderwijs

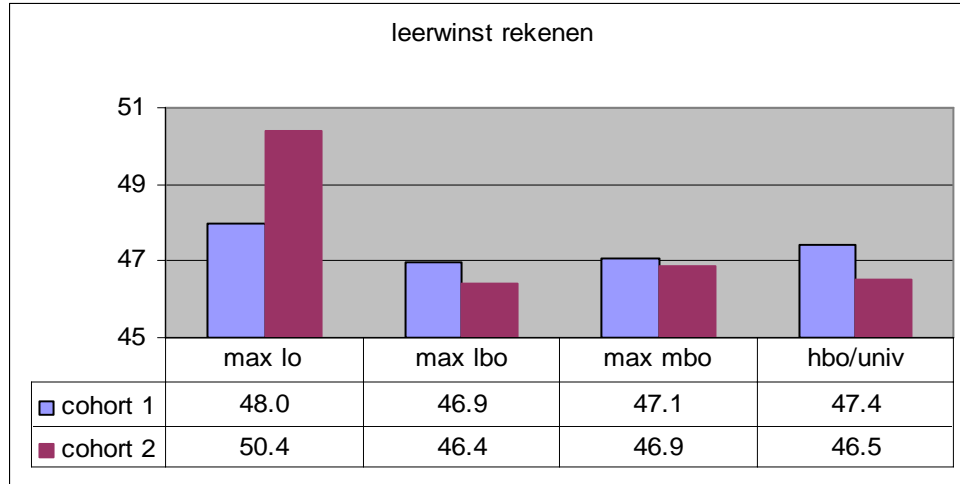
max lbo -de hoogste opleiding van beide ouders ligt op lbo-niveau

max mbo-de hoogste opleiding van beide ouders ligt op mbo-niveau

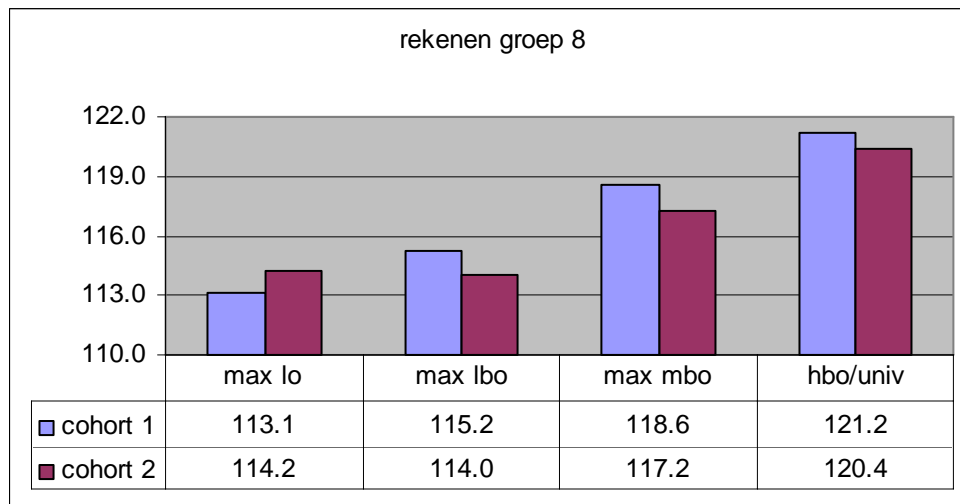
hbo/univ -tenminste één van beide ouders heeft hoger onderwijs gevolgd

Bij figuur 11 en 12 valt de hoge leerwinst op bij de laagst opgeleide groep. Bovendien is bij deze groep de leerwinst in het tweede cohort groter dan in het eerste cohort; bij de drie andere groepen is dat juist andersom.

Figuur 11 Leerwinst rekenen in twee cohorten; naar opleidingsniveau ouders



Figuur 12 Rekenvaardigheid groep 8 in twee cohorten; naar opleidingsniveau ouders



Het eindniveau voor rekenen/wiskunde in groep 8 laat het vertrouwde patroon zien: hoe hoger de ouderlijke opleiding, des te beter de rekenprestaties.

Over het algemeen zijn de prestaties in het tweede cohort lager dan in het eerste. Maar bij de laagst opgeleide groep juist niet: daar zijn de prestaties in het tweede cohort beter en ze komen daardoor op gelijke hoogte als de groep met maximaal lbo.

### **2.3 Enkele conclusies**

De voorgaande analyses maken duidelijk dat het doorgaans niet voldoende is om alleen naar leerwinst te kijken, maar dat ook (tenminste) het eindniveau in beschouwing moet worden genomen.

We hebben hier reeds op gewezen bij de eerder genoemde rapportage over ontwikkelingen bij de verschillende doelgroepen van het Onderwijs Achterstanden Beleid OAB. In dat onderzoek, op basis van eerdere PRIMA-cohorten, was zowel bij taal als bij rekenen sprake van een duidelijke afname van leerwinst in het latere cohort, maar tegelijk ook van een ongeveer gelijkblijvend eindniveau. (Zie Mulder et al., 2005, in het bijzonder hoofdstuk 3.2: *Longitudinale analyse: leerwinsten.*)

Verder hebben we gezien dat het algemene beeld van de ontwikkeling van bepaalde prestaties vaak niet op dezelfde manier terug komt bij verschillende groepen leerlingen. Naast een algemene indicator voor het onderwijsstelsel lijkt het daarom ook van belang die indicator te verbijzonderen naar die verschillende groepen leerlingen.

Dit is ook van belang voor vergelijking door de tijd. Opeenvolgende cohorten kunnen immers (in mindere of meerdere) mate van samenstelling veranderen. We kunnen hier bijvoorbeeld denken aan het nog steeds stijgende onderwijsniveau van de ouders van leerlingen. Maar ook in de samenstelling van de verschillende etnische groepen kunnen in de loop der jaren verschuivingen optreden, die van invloed kunnen zijn op zowel leerwinst als eindniveau.

*Leerwinst op Taal en Rekenen in PRIMA*

Tot slot merken we op dat we gebruik hebben gemaakt van (weliswaar omvangrijke) steekproeven, waardoor er altijd sprake zal zijn van onzekerheidsmarges rondom verschillen (tussen groepen en cohorten) in leerwinst en niveau.



## **3 Groeimodellen voor Taal en Rekenen in PRIMA**

### **3.1 Analyse ontwikkeling taal- en rekenprestaties door middel van groeimodellen**

De taal- en rekenprestaties zijn verder geanalyseerd door middel van latente groeicurve analyse (LGCA). Het voordeel van deze analysetechniek is dat het ontwikkelingspatroon van elke leerling kan worden weergegeven in de vorm van een unieke groeicurve. Er wordt rekening mee gehouden dat leerlingen een verschillend prestatieniveau hebben aan de start in groep 4 en ook dat individuele leerlingen verschillen in de mate waarin hun prestaties meer of minder toenemen. In de analyse worden het niveau van een variabele tijdens de eerste meting (het intercept van de lijn), hier in groep 4, en de mate van groei in een variabele (de helling/richtingscoëfficiënt van de lijn) als latente variabelen beschouwd. Dit type analyse kan worden uitgevoerd met behulp van programma's voor het analyseren van structurele modellen, in ons geval Mplus (Muthén & Muthén, 2004).

Het voordeel van het modelleren van het intercept en de helling als latente variabelen is dat deze variabelen in relatie gebracht kunnen worden met elkaar en met andere latente of controlevariabelen. Een negatieve samenhang tussen het startniveau en de mate van groei betekent bijvoorbeeld dat een lager startniveau overeenkomt met een hoger groeiniveau en andersom. Een ander voorbeeld is een sekse-effect op de latente variabelen. Dit geeft weer in hoeverre jongens en meisjes verschillen in het startniveau en in de mate van groei. We kunnen bijvoorbeeld ook vaststellen in hoeverre Turkse leerlingen meer of minder vooruit zijn gegaan in taalprestaties van groep 4 naar 8 dan Marokkaanse leerlingen.

Voor analyses met groeimodellen is het noodzakelijk om over gegevens te beschikken van ten minste drie meetmomenten. Bij drie meetmomenten kan alleen lineaire groei worden onderzocht, om groeiversnellingen te kunnen onderzoeken zijn ten minste ge-



*Verkenning leerwinst als indicator voor onderwijskwaliteit*

gegevens van vier meetmomenten nodig. Binnen het PRIMA cohort onderzoek wordt vanaf groep 2 gemeten. Alleen de taal- en rekenprestaties vanaf groep 4 staan echter op dezelfde schaal als de prestaties in hogere groepen. We maken daarom gebruik van gegevens van groep 4 t/m 8. In PRIMA zijn dat gegevens van dezelfde leerlingen uit groep 4, 6 en 8, dus van drie meetmomenten.

Toetsgegevens die zijn verzameld op meerdere meetmomenten zijn binnen PRIMA alleen beschikbaar voor leerlingen die onvertraagd zijn doorgestroomd naar de volgende groep én op dezelfde school zijn gebleven. Van leerlingen die zijn blijven zitten, veranderd zijn van school, bijvoorbeeld door verhuizing, of verwezen zijn naar het speciaal (basis)onderwijs hebben we deze gegevens niet. Bij het vergelijken van de leerwinst van verschillende groepen leerlingen zoals hiervoor gepresenteerd in paragraaf 2 wordt er dus geen rekening mee gehouden dat onder bepaalde groepen leerlingen meer kinderen blijven zitten, verwezen worden naar het speciaal onderwijs of van school veranderen dan onder andere groepen leerlingen. Er is sprake van selectieve uitval: alleen van de beter presterende leerlingen houden we gegevens van meerdere meetmomenten over.

Het liefst zouden we een uitspraak doen over hoe de gemiddelde groep 4-leerling zich tot en met groep 8 ontwikkeld heeft, dus inclusief de leerlingen die uitvallen. Om dit toch zo goed mogelijk te kunnen benaderen, is uitgegaan van de representatieve groep 4 leerlingen aan de start en zijn vervolgens de scores van de kinderen waarvan we in latere metingen geen gegevens hebben, geschat. Daarbij is gebruik gemaakt van de (achtergrond-)kenmerken en scores van de latere uitvallers. De missende gegevens zijn geschat met behulp van ‘Full-information maximum likelyhood estimation’ (FIML)<sup>1</sup>.

---

<sup>1</sup> FIML is gebaseerd op de aanname dat missende waarden ‘missing at random’ zijn, wat wil zeggen dat missende waarden voorspeld kunnen worden uit de beschikbare gegevens. Het helemaal verwijderen van leerlingen met missende waarden (listwise deletion) is gebaseerd op de striktere aanname dat missende gegevens ‘completely at random’, dus volledig willekeurig zijn.

De analyses zijn als volgt opgebouwd. Bij de analyse van zowel de taal- als de rekenprestaties zijn drie modellen geschat. Er is bij elk model een multi-groep analyse uitgevoerd, waarbij de groepen de twee cohorten zijn en getoetst is of er verschillen zijn tussen beide cohorten. Bij alle uitgevoerde groeimodellen is rekening gehouden met de geneste structuur van de gegevens: hier leerlingen binnen klassen.

In het **eerste model** is geschat hoe de ontwikkeling in prestaties van de twee cohorten kinderen in zijn algemeenheid verlopen is en of hierin verschillen zijn tussen de twee cohorten. Hiervoor zijn alleen de prestaties in groep 4, 6 en 8 in het model opgenomen. Met deze analyse zijn het startniveau, de prestaties in groep 4, en de groei in prestaties per jaar bepaald. Te verwachten valt dat we hier lagere prestaties in groep 4 vinden dan in de voorgaande paragrafen. De reden hiervoor is dat in de voorgaande paragrafen alleen gemiddelden gepresenteerd zijn voor leerlingen van wie gegevens over zowel groep 4, 6 als 8 beschikbaar zijn. Dit zijn kinderen die onvertraagd zijn doorgestroomd naar groep 8 en op dezelfde basisschool zijn gebleven, kinderen die zijn blijven zitten maken hier dus bijvoorbeeld geen deel van uit. Aangezien er in groeimodelanalyse een schatting plaatsvindt van missende gegevens, ligt het voor de hand dat de schatting van prestaties in groep 4 op basis van het model lager zal zijn. Bij de presentatie van de resultaten wordt hier aandacht aan besteed.

In het **tweede model** is etnische herkomst als covariaat toegevoegd aan model 1. De uitkomsten van dit model laten zien in hoeverre de prestaties van onderscheiden groepen allochtone leerlingen zich van groep 4 naar 8 verschillend ontwikkeld hebben en of hierin verschillen zijn tussen beide cohorten. Bij etnische herkomst is uitgegaan van het geboorteland van de ouders, met als categorieën: autochtoon, gemengd (een autochtone en een allochtone ouder), Surinaams en Antilliaans, Turks, Marokkaans en overig. Autochtone leerlingen vormen de referentiecategorie. De analyse laat bijvoorbeeld zien in hoeverre de onderscheiden groepen allochtone leerlingen in groep 4 een achterstand hadden ten aanzien van autochtone leerlingen en of een eventuele achterstand is ingelopen.

### *Verkenning leerwinst als indicator voor onderwijskwaliteit*

In het **derde model** zijn aan model 2 de covariaten sekse, opleidingsniveau ouders en leeftijd toegevoegd.

De uitkomsten laten zien hoe de prestaties van de onderscheiden etnische groepen kinderen zich ontwikkelen wanneer rekening wordt gehouden met het opleidingsniveau van hun ouders, hun leeftijd (dus eerder opgelopen vertraging in de schoolloopbaan) en hun sekse en in hoeverre dit verschilt tussen beide cohorten. (Of bijvoorbeeld anders geformuleerd: in hoeverre de prestaties van kinderen met een verschillend ouderlijk opleidingsniveau zich anders ontwikkelen wanneer rekening wordt gehouden met hun etnische herkomst, leeftijd en sekse en in hoeverre dit verschilt tussen beide cohorten.)

Met deze analyse kunnen dus vragen worden beantwoord zoals genoemd in de inleiding: “Zijn er specifieke groepen die meer leerwinst vertonen dan andere? Lopen leerlingen met een achterstand deze gedurende het onderwijs in of juist niet?”. Bij opleiding is het hoogst voltooide niveau binnen het gezin genomen: maximaal lager onderwijs, maximaal lager beroepsonderwijs, maximaal middelbaar beroepsonderwijs, en hoger beroeps- en wetenschappelijk onderwijs. Kinderen met ouders met maximaal een mbo-opleiding vormen de referentiecategorie. Uit de schooladministratie komt de informatie over de sekse van het kind en diens leeftijd. De leeftijd van de kinderen is uitgedrukt in jaren.

## **3.2 Groeimodellen voor Taal**

In het **eerste model** is nagegaan hoe de ontwikkeling in taalprestaties van de twee cohorten in zijn algemeenheid is verlopen en of er significante verschillen zijn tussen beide cohorten. In beide cohorten was het startniveau 1046.9 en de groei in taalprestaties per jaar 17.8 punten. De ontwikkeling in taalprestaties verschilt dus niet significant tussen de twee cohorten. Zoals aangegeven, zijn in deze analyse de missende gegevens geschat. In 2.1 werden alleen de prestaties gerapporteerd van kinderen voor wie gegevens in zowel groep 4, 6 als 8 bekend zijn. Dit zijn, zoals aangegeven, kinderen die niet zijn veranderd van basisschool én onvertraagd zijn doorgestroomd van groep 4 naar 8. Kinderen die zijn blijven zitten of zijn verwezen

*Groeimodellen voor Taal en Rekenen in PRIMA*

naar het speciaal onderwijs zitten hier bijvoorbeeld niet bij. In 2.1 is dus gerapporteerd over een kansrijkere groep. In onderstaande tabel zijn de schattingen op basis van het groeimodel naast de taalscores uit Tabel 1 in paragraaf 2.1 gezet.

*Tabel 3 Taalvaardigheidsscores in groep 4, 6 en 8 in de twee cohorten (gemiddelden) en schattingen op basis van groeimodel (model 1)*

	geobserveerde scores onvertraagde kinderen (Tabel 1)		schatting op basis van groeimodel
	cohort 1	cohort2	cohort 1 & 2
groep 4	1050.0	1051.5	1046.9
groep 6	1082.8	1083.8	1082.4
groep 8	1120.5	1121.1	1117.9
leerwinst tussen 4 en 8	70.6	69.6	71.0

Inderdaad is in de tabel te zien dat de prestaties van de groep 4 kinderen in de hier gerapporteerde analyse wat lager liggen dan die genoemd in paragraaf 2.1.

In Tabel 4 zijn de resultaten weergegeven van het **tweede model** waarin is nagegaan hoe de taalprestaties van onderscheiden etnische groepen leerlingen zich van groep 4 naar 8 ontwikkeld hebben en of hierin verschillen zijn tussen de twee cohorten. Wanneer er geen significant verschil was tussen beide cohorten, zijn de resultaten in de kolom ‘geen verschil tussen de cohorten’ weergegeven, aangezien de getallen voor cohort 1 en 2 dan gelijk zijn.



de taalprestaties van autochtone leerlingen. De overige allochtone leerlingen presteerden in groep 4 5.00 punten hoger.

De leerlingen met een autochtone en een allochtone ouder uit cohort 1 en 2 gingen evenveel vooruit in prestaties per jaar als de autochtone leerlingen. De taalprestaties van de Surinaamse leerlingen uit cohort 1 gingen per jaar 3.97 punten meer vooruit dan de taalprestaties van autochtone leerlingen; voor de Surinaamse en Antilliaanse leerlingen uit cohort 2 was dit wat lager, namelijk 3.19 punten meer per jaar.

De Turkse leerlingen uit cohort 1 hadden dezelfde groei als de autochtone leerlingen, maar in cohort 2 is hun groei 2.47 punt meer. Marokkaanse leerlingen uit cohort 1 en 2 gingen van groep 4 naar 8 per jaar 3.51 punten meer vooruit dan autochtone leerlingen. De overige allochtone leerlingen uit cohort 1 verschilden niet van autochtone leerlingen in de mate waarin hun prestaties per jaar vooruit gingen, de prestaties van de overige allochtone kinderen uit cohort 2 gingen per jaar met 1.39 punten meer vooruit dan de prestaties van autochtone leerlingen.

In Tabel 5 is voor de twee cohorten weergegeven in hoeverre er in groep 8 nog een taalachterstand is voor allochtone leerlingen die in groep 4 ten aanzien van autochtone leerlingen een achterstand hadden.

*Tabel 5 Model 2: verschil in taalprestaties van allochtone ten aanzien van autochtone leerlingen in groep 4 en 8 voor de twee cohorten (alleen voor allochtone groepen met een taalachterstand in groep 4)*

	Achterstand in groep 4	Afname achterstand per jaar	groep 8: verschil met autochtone leerlingen
gemengd NL-all. cohort 1 & 2	-18.56	0.00	-18.56
Surinaams/Antilliaans cohort 1	-48.26	3.97	-32.37
Surinaams/Antilliaans cohort 2	-48.26	3.19	-35.49
Turks cohort 1	-34.66	0.00	-34.66
Turks cohort 2	-34.66	2.47	-24.79
Marokkaans cohort 1 en 2	-23.56	3.51	- 9.52

*Verkenning leerwinst als indicator voor onderwijskwaliteit*

Te zien is dat Marokkaanse kinderen in groep 8 nog 9.52 punten lager presteren dan autochtone leerlingen en daarmee de kleinste achterstand hebben in groep 8. In groep 4 presteerden zij nog 23.56 punten lager. Vergelijken we de prestaties van de allochtone leerlingen in groep 8 met die gepresenteerd in Figuur 5, dan is te zien dat er verschillen zijn met de resultaten in Tabel 5. In Figuur 5 is te zien dat *alle* allochtone groepen leerlingen in groep 8 lager presteren dan autochtone leerlingen en bijvoorbeeld dat de achterstand van Marokkaanse leerlingen uit cohort 2 ten aanzien van autochtone leerlingen 24.7 punten bedraagt.

Vervolgens is het **derde model** geschat waarin is nagegaan wat de relatie is tussen de ontwikkeling in taalprestaties (startniveau en groei) en sekse, etnische herkomst, het opleidingsniveau van de ouders en de leeftijd. Bovendien is nagegaan in hoeverre hierin verschillen zijn tussen de twee cohorten.

In Tabel 6 zijn de resultaten weergegeven. Wanneer er geen significant verschil was tussen beide cohorten, zijn de resultaten in de kolom 'geen verschil tussen de cohorten' weergegeven.

Groeimodellen voor Taal en Rekenen in PRIMA

Tabel 6 Resultaten multigroep groeimodel taal met als groepen de twee cohorten en met de covariaten sekse, etnische herkomst en leeftijd (model 3)

		cohort 1	geen verschil tussen cohorten	cohort 2
	startniveau*groei		ns	
sekse	meisje*startniveau		3.73	
	meisje*groei	ns		-1.12
etnische herkomst (ref = nl)	gemengd NL-all.*startniveau		-14.63	
	Surinaams/Antilliaans*startniveau		-38.55	
	Turks*startniveau		-23.13	
	Marokkaans*startniveau		-20.01	
	overig*startniveau		ns	
	gemengd NL-all.*groei		ns	
	Surinaams/Antilliaans*groei	5.37		3.91
	Turks*groei		2.52	
	Marokkaans*groei		3.95	
	overig*groei		ns	
opleidings- niveau	lagere school*startniveau		-13.47	
	lbo*startniveau		-9.65	
	ouders hoger onderwijs*startniveau		8.43	
(ref = mbo)	lagere school*groei		ns	
	lbo*groei		-0.60	
	hoger onderwijs*groei		0.83	
leeftijd	leeftijd*startniveau		-4.12	
	leeftijd*groei		-2.39	

Opnieuw zien we dat het startniveau en de groei in taalprestaties niet significant samenhangen.



### *Verkenning leerwinst als indicator voor onderwijskwaliteit*

In beide cohorten waren de taalprestaties in groep 4 van *meisjes* gemiddeld 3.73 punten hoger dan die van jongens. Voor cohort 1 was de ontwikkeling in taalprestaties van groep 4 naar 8 voor jongens en meisjes gelijk, voor cohort 2, het meest recente cohort, ontwikkelden de taalprestaties van meisjes zich minder goed dan die van jongens: de taalprestaties van meisjes namen van groep 4 naar 8 gemiddeld met 1.12 punten per jaar, in totaal 4.48 punten, minder toe.

De taalprestaties van de meeste onderscheiden groepen *allochtone* leerlingen in groep 4 waren lager dan die van autochtone leerlingen: die van leerlingen met een autochtone en allochtone ouder 14.6 punten lager, die van Marokkaanse leerlingen 20.0 punten lager, van Turkse leerlingen 23.13 punten lager en van Surinaamse en Antilliaanse leerlingen 38.55 punten lager. De prestaties in groep 4 van overige allochtone leerlingen verschilden niet van die van autochtone leerlingen. De prestaties in groep 4 van de onderscheiden etnische groepen verschilden niet significant tussen beide cohorten.

Wat betreft de relatie tussen de groei in prestaties en etnische herkomst is er alleen voor Surinaamse en Antilliaanse leerlingen een verschil tussen de twee cohorten. De prestaties van Surinaamse en Antilliaanse kinderen in cohort 1 namen per jaar met 5.37 punten meer toe dan de prestaties van autochtone leerlingen (van groep 4 – 8 in totaal 21.48 punten meer); voor cohort 2 was dit wat lager, namelijk 3.91 punten meer per jaar dan autochtone leerlingen (van groep 4 – 8 in totaal 15.64 punten). De taalprestaties van overige allochtone leerlingen en van allochtone leerlingen met een allochtone en een autochtone ouder namen vanaf groep 4 per jaar gemiddeld evenveel toe als de taalprestaties van autochtone leerlingen. Voor Turkse en Marokkaanse leerlingen is er net als voor de Surinaamse en Antilliaanse leerlingen een sterkere toename in prestaties per jaar dan voor autochtone leerlingen. Surinaamse en Antilliaanse, Turkse en Marokkaanse leerlingen maken van groep 4 naar 8 dus een deel van hun achterstand op autochtone leerlingen goed. In onderstaande tabel is voor de twee cohorten weergegeven in hoeverre er in groep 8 nog een achterstand is voor allochtone leerlingen die in groep 4 ten aanzien van autochtone leerlingen een achterstand hadden. De overige allochtone leerlingen worden in de tabel niet genoemd. Zij had-

den in groep 4 even hoge prestaties als autochtone leerlingen en de toename in prestaties per jaar verschilde evenmin van die van autochtone leerlingen.

Tabel 7 Model 3: verschil in taalprestaties van allochtone ten aanzien van autochtone leerlingen in groep 4 en 8 voor de twee cohorten (alleen voor allochtone groepen met een taalachterstand in groep 4)

	achterstand in groep 4	afname achterstand per jaar	groep 8: verschil met autochtone leerlingen
gemengd NL-all. cohort 1 & 2	-14.63	0.00	-14.63
Surinaams/Antilliaans cohort 1	-38.55	5.37	-17.06
Surinaams/Antilliaans cohort 2	-38.55	3.91	-22.93
Turks cohort 1 & 2	-23.13	2.52	-13.05
Marokkaans cohort 1 & 2	-20.01	3.95	-4.22

Te zien is dat leerlingen met een autochtone en een allochtone ouder in groep 4 de kleinste achterstand hadden ten aanzien van autochtone leerlingen, terwijl in groep 8 de achterstand van zowel Turkse als Marokkaanse leerlingen kleiner is dan die achterstand: van de laatstgenoemde groep is deze nog maar 4.22 punten.

Tabel 6 laat zien dat hoe hoger het *opleidingsniveau* van de ouders van leerlingen is, des te hoger de prestaties van hun kinderen zijn. Zo zijn de taalprestaties van kinderen met ouders met ten hoogste een lagere school opleiding in groep 4 13.47 punten lager dan de taalprestaties van kinderen met ouders met ten hoogste middelbaar beroepsonderwijs en zijn de prestaties van kinderen met ouders die hoger onderwijs hebben genoten 8.43 punten hoger dan die van kinderen met middelbaar opgeleide ouders. De groei in taalprestaties per jaar verschilt niet tussen kinderen met ouders met maximaal een lagere schoolopleiding en kinderen met middelbaar opgeleide ouders. De groei in taalprestaties van kinderen met ouders met lager beroepsonderwijs is lager (0.60 punten per jaar) en die van kinderen met hoger opgeleide ouders hoger (0.83 punten per jaar) dan van kinderen met middelbaar opgeleide ouders. Er zijn geen verschillen tus-

### *Verkenning leerwinst als indicator voor onderwijskwaliteit*

sen de twee cohorten in de relatie tussen het opleidingsniveau van de ouders en het startniveau en de groei in taalprestaties.

Naarmate kinderen een jaar *ouder* zijn, zijn hun taalprestaties in groep 4 in beide cohorten 4.12 punten lager en nemen hun taalprestaties met 2.39 punten per jaar minder toe. Doorgaans zullen dit leerlingen zijn die al eerder zijn blijven zitten of die kleuterbouwverlenging hebben gehad.

### **3.3 Groeimodellen voor Rekenen**

Om te bepalen hoe de ontwikkeling in rekenprestaties van de twee cohorten in zijn algemeenheid is verlopen, is net als bij taal een multi-groep groeimodel geschat waarin alleen de rekenprestaties in groep 4, 6 en 8 zijn opgenomen (**model 1**). In beide cohorten was het startniveau 68.7 en de groei in rekenprestaties 12.0 punten per jaar. Net als bij taal hebben we de schattingen op basis van het groeimodel naast de scores gezet van kinderen die onvertraagd zijn doorgestroomd en niet van basisschool zijn veranderd (zie Tabel 8).

*Tabel 8 Rekenvaardigheidsscores in groep 4, 6 en 8 in de twee cohorten (gemiddelden) en schattingen op basis van groeimodel*

	geobserveerde scores onvertraagde kinderen (Tabel 2)		schatting op basis van groeimodel
	cohort 1	cohort2	cohort 1 & 2
groep 4	70.9	70.5	68.7
groep 6	95.9	96.0	92.8
groep 8	118.2	117.4	116.8
leerwinst tussen 4 en 8	47.3	46.9	48.1

*Groeimodellen voor Taal en Rekenen in PRIMA*

De tabel laat zien dat de prestaties in groep 4 die zijn geschat op basis van het groei-model waarin missende gegevens zijn geschat, zoals verwacht lager zijn dan de rekenprestaties van onvertraagde kinderen in Tabel 2 in paragraaf 2.2.

Vervolgens hebben we het **tweede model** geschat waarin is nagegaan hoe de rekenprestaties van onderscheiden etnische groepen leerlingen zich van groep 4 naar 8 ontwikkeld hebben en of hierin verschillen zijn tussen de twee cohorten. De resultaten zijn weergegeven in Tabel 9. Wanneer er geen significant verschil was tussen beide cohorten, zijn de resultaten in de kolom ‘geen verschil tussen de cohorten’ weergegeven. Te zien is dat er bij rekenen veel meer verschillen tussen de twee cohorten zijn dan bij taal.

*Tabel 9 Resultaten model 2: multi-groep groeimodel rekenen met als groepen de twee cohorten en met als covariaat etnische herkomst*

		cohort	geen verschil tus- sen cohorten	cohort 2
	startniveau*groei		-8.57	
etnische	gemengd NL-all.*startniveau	-6.37		-7.11
herkomst	Surinaams/Antilliaans*startniveau	-8.58		-8.30
(ref = nl)	Turks*startniveau	-8.16		-9.58
	Marokkaans*startniveau	-3.93		-4.24
	overig*startniveau	1.44		3.36
	gemengd NL-all.*groei	1.02		ns
	Surinaams/Antilliaans*groei	0.74		0.78
	Turks*groei	ns		1.12
	Marokkaans*groei	0.82		0.71
	overig*groei		ns	

Naarmate de rekenprestaties in groep 4 lager zijn, is er sprake van een hogere groei per jaar en naarmate de rekenprestaties in groep 4 hoger zijn, van een lagere groei. Dit is gelijk voor beide cohorten.

*Verkenning leerwinst als indicator voor onderwijskwaliteit*

In tegenstelling tot bij taal worden bij rekenen dus achterstanden in groep 4 (geheel of gedeeltelijk) gecompenseerd in de latere jaren.

Overige *allochtone* leerlingen hadden in groep 4 hogere prestaties dan autochtone leerlingen (cohort 1 1.44 en cohort 2 3.36 punten hoger). De andere groepen allochtone leerlingen hadden in groep 4 lagere rekenprestaties dan autochtone leerlingen. De rekenprestaties van de overige allochtone leerlingen uit cohort 1 en 2, van leerlingen met een autochtone en een allochtone ouder uit cohort 2, en van Turkse leerlingen uit cohort 1, namen per jaar evenveel toe als de rekenprestaties van autochtone leerlingen. Van de andere onderscheiden groepen allochtone leerlingen namen de rekenprestaties van groep 4 naar 8 meer toe dan die van autochtone leerlingen. Zo namen de rekenprestaties van Marokkaanse kinderen uit cohort 1 met 0.82 punten per jaar en van Marokkaanse kinderen uit cohort 2 met 0.71 punten per jaar meer toe dan de rekenprestaties van autochtone kinderen.

In Tabel 10 is voor de twee cohorten weergegeven in hoeverre er in groep 8 nog een rekenachterstand is voor allochtone leerlingen die in groep 4 ten aanzien van autochtone leerlingen een achterstand hadden.

Tabel 10 Model 2: verschil in rekenprestaties van allochtone ten aanzien van autochtone leerlingen in groep 4 en 8 voor de twee cohorten (alleen voor allochtone groepen met een rekenachterstand in groep 4)

	achterstand in groep 4	afname achterstand per jaar	groep 8: ver- schil met autochtone leerlingen
cohort 1:			
gemengd NL-all.	-6.37	1.02	-2.28
Surinaams/Antilliaans	-8.58	0.74	-5.62
Turks	-8.16	0.00	-8.16
Marokkaans	-3.93	0.82	-0.66
cohort 2:			
gemengd NL-all.	-7.11	0.00	-7.11
Surinaams/Antilliaans	-8.30	0.78	-5.18
Turks	-9.58	1.12	-5.10
Marokkaans	-4.24	0.71	-1.39

Van de groepen allochtone kinderen met een rekenachterstand in groep 4 hebben de Marokkaanse kinderen in groep 8 de kleinste achterstand. Ook in groep 4 was hun achterstand kleiner dan die van Turkse, Surinaamse en Antilliaanse leerlingen en van leerlingen met een allochtone en een autochtone ouder. Van de meeste allochtone groepen is de achterstand in groep 8 kleiner geworden. Alleen van Turkse leerlingen uit cohort 1 en van leerlingen met een allochtone en een autochtone ouder uit cohort 2 is de achterstand ten aanzien van autochtone leerlingen in groep 4 gelijk aan die in groep 8.

Vervolgens is nagegaan wat de relatie is tussen de ontwikkeling in rekenprestaties (startniveau en groei) en sekse, etnische herkomst, het opleidingsniveau van de ouders en de leeftijd en is nagegaan in hoeverre hierin verschillen zijn tussen de twee cohorten (**model 3**). De resultaten zijn weergegeven in Tabel 11.

*Verkenning leerwinst als indicator voor onderwijskwaliteit*

*Tabel 11 Resultaten multigroep groeimodel rekenen met als groepen de twee cohorten en met de covariaten sekse, etnische herkomst en leeftijd (model 3)*

		cohort 1	geen verschil tussen cohorten	cohort 2
	startniveau*groei		-8.77	
sekse	meisje*startniveau	-3.24		-4.45
	meisje*groei	ns		0.37
etnische herkomst (ref = nl)	gemengd NL-all.*startniveau		-5.25	
	Surinaams/Antilliaans*startniveau	-5.73		-5.19
	Turks*startniveau		-5.37	
	Marokkaans*startniveau	-3.10		-2.96
	overig*startniveau	ns		1.91
	gemengd NL-all.*groei	1.07		ns
	Surinaams/Antilliaans*groei		0.87	
	Turks*groei	0.52		0.99
	Marokkaans*groei	0.88		0.80
	overig*groei		ns	
	opleidingsniveau	lagere school*startniveau	-4.24	
lbo*startniveau		-3.16		-2.93
ouders (ref = mbo)	hoger onderwijs*startniveau	2.77		3.58
	lagere school*groei	ns		0.41
	lbo*groei		ns	
	hoger onderwijs*groei		ns	
leeftijd	leeftijd*startniveau	-1.37		-1.31
	leeftijd*groei	-0.70		-0.84

Ook hier is te zien dat voor beide cohorten geldt dat naarmate de rekenprestaties in groep 4 lager zijn, er sprake is van een hogere groei per jaar en naarmate de rekenprestaties in groep 4 hoger zijn, er sprake is van een lagere groei.

*Meisjes* uit beide cohorten hebben in groep 4 lagere prestaties dan jongens (in cohort 1 3.24 en in cohort 2 4.45 punten lager). Voor meisjes uit cohort 1 verandert deze achterstand niet, voor meisjes uit cohort 2 neemt de achterstand met 0.37 punten per jaar af.

Overige *allochtone* leerlingen uit cohort 1 hebben even hoge prestaties in groep 4 en overige *allochtone* leerlingen uit cohort 2 iets hogere rekenprestaties (1.91 punten hoger) dan *autochtone* leerlingen. Hun groei in rekenprestaties per jaar is even hoog als die van *autochtone* leerlingen. Net als voor de taalprestaties hebben we de rekenprestaties van *allochtone* leerlingen met lagere prestaties dan *autochtone* leerlingen in groep 4 in een aparte tabel weergegeven. Op basis van de gegevens in deze tabel is gemakkelijker te zien in hoeverre een achterstand wordt ingelopen, of zelfs wordt omgezet in een voorsprong.

Tabel 12 Model 3: verschil in rekenprestaties van *allochtone* ten aanzien van *autochtone* leerlingen in groep 4 en 8 voor de twee cohorten (alleen voor *allochtone* groepen met een rekenachterstand in groep 4)

	achterstand in groep 4	Afname achter- stand per jaar	groep 8: verschil met <i>autochtone</i> leerlingen
<i>cohort 1:</i>			
gemengd NL-all.	-5.25	1.07	-0.97
Surinaams/Antilliaans	-5.73	0.87	-2.26
Turks	-5.37	0.52	-3.28
Marokkaans	-3.10	0.88	+0.40
<i>cohort 2:</i>			
gemengd NL-all.	-5.25	0.00	-5.25
Surinaams/Antilliaans	-5.19	0.87	-1.72
Turks	-5.37	0.99	-1.43
Marokkaans	-2.96	0.80	+0.22

Van de *allochtone* groepen met een achterstand ten aanzien van *autochtone* kinderen in groep 4 hebben de Marokkaanse kinderen in zowel cohort 1 als 2 de minste achter-



*Verkenning leerwinst als indicator voor onderwijskwaliteit*

stand. In groep 8 is hun achterstand ten aanzien van autochtone leerlingen omgezet in een lichte voorsprong. Deze is voor cohort 2 wat minder dan voor cohort 1. Binnen cohort 1 hebben Turkse kinderen in groep 8 de grootste achterstand ten aanzien van autochtone leerlingen, terwijl dit binnen cohort 2 voor kinderen met een allochtone en een autochtone ouder het geval is.

Net als bij taal zijn de rekenprestaties van kinderen in groep 4 hoger naarmate het *opleidingsniveau* van hun ouders hoger is. De achterstand van kinderen met lager opgeleide ouders ten aanzien van kinderen met middelbaar opgeleide ouders is binnen cohort 2 iets kleiner dan binnen cohort 1 (ouders lagere school: cohort 1 4.24 punten lager, cohort 2 3.89 punten lager, ouders lbo: cohort 1 3.16 punten lager, cohort 2 2.93 lager). Kinderen uit cohort 2 met hoger opgeleide ouders hebben in groep 4 een grotere voorsprong op rekenen (3.58 punten) dan kinderen met hoger opgeleide ouders uit cohort 1 (2.77 punten). Wat betreft de groei in rekenprestaties vonden we één significant verschil: van de leerlingen met ouders met maximaal een lagere schoolopleiding uit cohort 2 namen de rekenprestaties meer toe per jaar dan van kinderen met middelbaar opgeleide ouders.

Net als bij taal, presteren kinderen in groep 4 naarmate zij *ouder* zijn lager en nemen hun rekenprestaties minder toe. Er is hierin een verschil tussen de cohorten: kinderen uit cohort 2 die een jaar ouder zijn, hebben in groep 4 weliswaar iets minder lage prestaties dan kinderen uit cohort 1, per jaar neemt hun achterstand meer toe dan die van de kinderen uit cohort 1: bij deze kinderen uit cohort 1 is er sprake van 0.70 punten minder groei per jaar, bij deze kinderen in cohort 2 van 0.84 minder groei per jaar.

## 4 Samenvattende terugblik

Er kan op allerlei manieren naar de opbrengsten van onderwijs gekeken worden. We geven hier allereerst een beknopte opsomming van hoe er in deze verkenningen naar de opbrengst op het gebied van **taal** is gekeken.

### 4.1 Opbrengsten bij Taal

#### *prestatieniveau in groep 8*

Allereerst kan er natuurlijk eenvoudigweg gekeken worden naar het vaardigheidsniveau van leerlingen aan het eind van het basisonderwijs. Bij de laatste twee PRIMA metingen was de gemiddelde PRIMA-taalvaardigheidsscore in groep 8 in de referentiesteekproeven van schooljaar 2002/2003 en 2004/2005 respectievelijk 1119.1 en 1118.3. Zo te zien is er dus sprake van een (lichte) teruggang in taalvaardigheid.

Maar omdat het hier niet om populatiegegevens gaat maar om steekproeven, zal er natuurlijk altijd sprake zijn van een onzekerheidsmarge. Het verschil tussen beide gemiddelden is dan ook statistisch niet significant (ook al is de steekproefgrootte in beide jaren rond de 9000 leerlingen!).

#### *leerwinst*

In paragraaf 2 hebben we enkele analyses gepresenteerd met een eenvoudige versie van leerwinst. Voor leerlingen die we (in twee opeenvolgende cohorten) hebben kunnen volgen van groep 4 in het basisonderwijs tot en met groep 8 zijn we nagegaan hoeveel deze leerlingen zijn vooruitgegaan op de PRIMA-taalvaardigheidschaal (die over deze groepen heen vergelijkbaar is).

De *leerwinst* bedroeg in de twee cohorten respectievelijk 70.6 en 69.6 punt (Tabel 1, par. 2). Zo bezien is er in het tweede cohort sprake van een (lichte) teruggang in leerwinst.

### *Verkenning leerwinst als indicator voor onderwijskwaliteit*

Tegelijk hebben we laten zien dat het *prestatieniveau in groep 8* van deze twee cohorten leerlingen juist omhoog is gegaan. Het gemiddelde in groep 8 in schooljaar 2002/2003 (cohort 1) was 1120.5 en dat in schooljaar 2004/2005 was 1121.1 (Tabel 1, par. 2).

Onze conclusie was dan ook dat alleen deze eenvoudige versie van leerwinst onvoldoende is als stelselindicator. Leerwinst én niveau zullen beide, in samenhang, onderdeel van zo'n indicator moeten zijn.

(Overigens gaat het ook hier natuurlijk weer om steekproeven en blijken ook hier alle gevonden verschillen statistisch niet significant. Maar het ging ons hier niet om toetsing, maar om het meer principiële probleem dat leerwinst niet los van niveau mag worden beschouwd.)

Verder merken we op dat het gemiddelde prestatieniveau in groep 8 voor de cohortleerlingen (die we hebben gevolgd door groep 4 – groep 6 – groep 8) hoger ligt dan voor het hiervoor genoemde prestatieniveau voor alle leerlingen in groep 8. Zoals we hebben aangegeven komt dit door de selectieve uitval van zwakkere leerlingen uit deze cohorten.

### *groeimodellen*

In een groeimodel wordt de samenhang van groei en (begin-)niveau expliciet gemodelleerd. Bovendien kunnen er binnen dezelfde analyse schattingen verkregen worden voor uitgevallen leerlingen. In paragraaf 3 hebben we een aantal analyses gepresenteerd met dergelijke groeimodellen.

#### *'leeg' groeimodel (model 1)*

Allereerst is een model onderzocht waarin verder geen specifieke leerlingkenmerken werden opgenomen, maar alleen de taalscores in de groepen 4, 6 en 8. De conclusies bij dit model waren dat de ontwikkeling van de leerlingen in beide cohorten niet wezenlijk van elkaar verschilt. Er is sprake van een gelijk beginniveau en een gelijke groei. En het beginniveau en de groei hangen niet samen: bij een hoger beginniveau zien we dus niet meer (of minder) groei dan een laag beginniveau.

Het eindniveau in groep 8 wordt voor beide cohorten geschat op 1117.9 en de leerwinst over 4 schooljaren op 71.0 punten op de PRIMA-taalvaardigheidschaal (Tabel 3, par 3).

We merken nog op dat hier het eindniveau lager wordt ingeschat dan bij de directe prestatiemeting in groep 8 of de ‘eenvoudige’ leerwinst. Dat heeft rechtstreeks te maken met het meeschatten van de zwakkere, uitgevallen leerlingen in dit model.

*groeimodellen voor groepen leerlingen (model 2 en 3)*

Het groeimodel maakt het ook mogelijk om voor verschillende groepen leerlingen na te gaan hoe bij hen de samenhang van ontwikkeling en niveau is en of er sprake is van verschillen in niveau en ontwikkeling: verschillen tussen groepen leerlingen, maar ook verschillen in de tijd, tussen cohorten.

Uit het model voor de verschillende etnische groepen (model 2) kwam naar voren dat deze groepen in groep 4 beginnen met een achterstand op de autochtone leerlingen, maar dat tegelijk hun groei doorgaans groter is. Hierdoor lopen zij tijdens hun loopbaan naar groep 8 een deel van hun eerdere achterstand in (vgl. Tabel 5, par. 3).

Het groeimodel kan vervolgens nog verder uitgebreid worden met andere relevante leerlingkenmerken, waardoor het beeld van de groei vollediger (maar ook complexer) kan worden. Het hier geanalyseerde model 3 (Tabel 6) liet aanvullend bijvoorbeeld zien dat kinderen van ouders met hoger onderwijs beginnen met een hoger startniveau en vervolgens ook meer groei laten zien; hun voorsprong dus vergroten. Oudere leerlingen (veelal zittenblijvers) starten lager en zien daarna hun achterstand alleen maar toenemen.

## **4.2 Opbrengsten bij Rekenen**

Ook voor rekenen geven we een –nog beknopter– overzicht van hoe er in de voorgaande verkenningen naar opbrengsten is gekeken.

*Verkenning leerwinst als indicator voor onderwijskwaliteit*

*prestatieniveau in groep 8*

Bij de laatste twee PRIMA metingen was de gemiddelde PRIMA-taalvaardigheid-score in de referentiesteekproeven van schooljaar 2002/2003 en 2004/2005 respectievelijk 117.7 en 117.1.

Net als bij taal zien we dus een (heel lichte) teruggang in rekenvaardigheid.

En in tegenstelling tot taal is hier het verschil tussen beide gemiddelden statistisch significant.

*leerwinst*

Ook voor rekenen hebben we de eenvoudige versie van leerwinst bepaald, uitgaande van leerlingen die zowel in groep 4, als in groep 6 en 8, hebben deelgenomen aan PRIMA.

De *leerwinst* in Cito-rekenvaardigheid bedroeg in de twee cohorten achtereenvolgens 47.3 en 46.9 punten (Tabel 2, par. 2), een lichte achteruitgang dus.

En bij rekenen is er ook bij het *prestatieniveau in groep 8* sprake van achteruitgang: van 118.2 naar 117.4 op de rekenvaardigheidschaal.

Voor rekenen wijzen leerwinst en niveau dus in dezelfde richting: een lichte achteruitgang.

(Het verschil in leerwinst tussen beide cohorten is niet significant; het lagere niveau in groep 8 wel.)

*'leeg' groeimodel (model 1)*

Bij de rekenvaardigheid vinden we, net als bij taal, dat startniveau en groei in beide cohorten niet significant van elkaar verschillen. Het eindniveau in groep 8 wordt voor beide cohorten geschat op 116.8 en de leerwinst over 4 schooljaren op 48.1 punten op de Cito-rekenvaardigheidschaal (Tabel 8, par 3).

In tegenstelling tot bij taal vinden we hier dat er een negatieve samenhang is tussen startniveau en groei: naarmate het startniveau hoger is, is er sprake van minder groei. En andersom: leerlingen die laag beginnen groeien meer en lopen daardoor in de volgende vier jaar achterstand in.

*groeimodellen voor groepen leerlingen (model 2 en 3)*

Een model voor de groei van rekenvaardigheid van leerlingen uit verschillende etnische groepen (model 2; Tabel 9) laat zien dat deze groepen doorgaans een lager startniveau van rekenvaardigheid in groep 4 hebben dan autochtone leerlingen, maar soms ook meer groei per jaar kennen. Er treden hier meer verschillen op tussen de cohorten dan bij de vergelijkbare analyses van de taalontwikkeling.

Verdere uitbreiding van het groeimodel met andere relevante leerlingkenmerken (model 3) laat zien dat hoe hoger het opleidingsniveau van de ouders, des te hoger de startprestaties voor rekenen in groep 4, maar er is verder vrijwel geen verschil in groeitempo. Wat betreft leeftijd geldt ook hier, net als bij taal, dat oudere leerlingen lager starten én minder groeien. Interessant is nog dat, rekening houdend met de andere leerlingkenmerken in het model, de Marokkaanse leerlingen hun aanvankelijke achterstand in rekenen volledig inlopen.



## 5 Mogelijkheden en beperkingen van COOL

### 5.1 Algemene opzet

Het COOL<sup>5-18</sup> cohortonderzoek is opgezet om leerlingen te kunnen volgen in hun gehele loopbaan door het (funderend) onderwijs. Daartoe worden er metingen gedaan in groep 2, groep 5 en groep 8 van het basisonderwijs, in het derde leerjaar van het voortgezet onderwijs en (volgens planning) in de eindexamenklassen van havo en vwo en in het mbo (in het 6<sup>e</sup> jaar voortgezet onderwijs).

In de opzet is voorts voorzien in metingen van geletterdheid en gecijferdheid door de hele onderwijsloopbaan. Verder worden ook een aantal motivationele en sociaal-emotionele variabelen (schalen) gemeten, die vanaf groep 5 door de tijd gevolgd kunnen worden.

In principe zou met de gegevens uit het COOL<sup>5-18</sup> onderzoek dus elke 3 jaar een “stand van zaken” van het onderwijs gemaakt kunnen worden.

Dat kan allereerst door het maken van ‘**dwarsdoorsnedes**’: de vergelijking van een prestatieniveau in een bepaalde fase van de onderwijsloopbaan (bv. groep 8) met dat prestatieniveau in het verleden. Indien de afgenomen toetsen te verbinden zijn met de referentieniveaus van de commissie Meijerink, dan kan dat ook informatie opleveren voor de vraag of over de jaren meer of minder leerlingen het fundamentele of het streefniveau voor taal en rekenen beheersen.

Met de vrijwel afgeronde eerste meting van COOL zijn zulke dwarsdoorsnede vergelijkingen mogelijk voor het prestatieniveau in groep 2 en dat in groep 8, omdat uit beide groepen vergelijkbare gegevens zijn verzameld in het eerdere PRIMA-onderzoek. (Een vergelijking van het derde leerjaar voortgezet onderwijs met het VOCL-cohort zal, door verandering van instrumenten, problematischer zijn.)



*Verkenning leerwinst als indicator voor onderwijskwaliteit*

In volgende COOL-metingen wordt het aantal mogelijk vergelijkingen natuurlijk uitgebreid tot groep 2, 5 en 8 van het basisonderwijs en leerjaar 3 van voortgezet onderwijs.

Daarnaast kunnen ook longitudinale cohortgegevens geanalyseerd worden, die informatie geven in termen van **leerwinst** en (later) **groeimodellen**.

Daarbij kan worden nagegaan hoe geletterdheid en gecijferdheid zich ontwikkelen tijdens de onderwijsloopbaan en of daar verschillen in optreden tussen cohorten. Zoals bekend is de opbouw van zulke longitudinale cohorten een kwestie van lange adem, omdat individuele leerlingen gevolgd moeten (blijven) worden door hun feitelijke onderwijsloopbaan.

Met de vrijwel afgeronde eerste meting van COOL komt in principe longitudinale informatie beschikbaar voor twee cohorten leerlingen:

PRIMA-groep 2 (schooljaar 2004/2005)	→	COOL groep 5
PRIMA-groep 8 (schooljaar 2004/2005)	→	COOL jaar 3 vo

Bij het laatste cohort is daarnaast (voor een deel van de leerlingen; zie ook verderop bij uitval) ook nog informatie uit eerdere PRIMA-metingen beschikbaar.

Bij de volgende COOL-meting in schooljaar 2010/2011 worden de mogelijkheden van longitudinale analyses verder uitgebreid:

PRIMA-groep 2 →	cool groep 5 →	cool groep 8
	cool groep 2 →	cool groep 5
	cool groep 8 →	cool jaar 3vo
PRIMA-groep 8 →	cool jaar 3vo →	cool jaar 6vo

Hier komen meer longitudinale gegevens beschikbaar met drie meetmomenten, waarmee analyses met behulp van groeimodellen mogelijk worden. Bij toekomstige COOL-metingen worden deze mogelijkheden nog verder vergroot.

## **5.2 Representativiteit**

Net als bij eerdere cohortonderzoekingen PRIMA en VOCL zijn en worden de gegevens voor COOL verzameld bij **steekproeven** van scholen voor basis- en voortgezet onderwijs.

In het basisonderwijs gaat het om een steekproef van ongeveer 550 scholen. Daarvan vormen 420 scholen de referentiesteekproef, die als representatief voor het Nederlandse basisonderwijs mag worden beschouwd. De overige scholen vormen de zogenaamde aanvullende steekproef, waarin een oversampling heeft plaatsgevonden van scholen met veel achterstandsleerlingen en van traditionele vernieuwingscholen. De representativiteit van de referentiesteekproef wordt afgemeten aan:

- samenstelling van de leerlingbevolking naar leerlinggewicht (op basis van Cf-gegevens);
- schoolgrootte;
- denominatie;
- stedelijkheid;
- spreiding over het land (landsdeel/provincie).

Op deze punten is het bij PRIMA altijd gelukt een goede steekproef van scholen te verkrijgen. Ook bij de eerste meting van COOL is dit gelukt.

Er zijn natuurlijk basisscholen, die om principiële redenen geen gebruik maken van methode-onafhankelijke landelijk genormeerde toetsen; deze zullen ook niet aan het COOL-onderzoek (willen) meedoen. Volgens de Onderwijsinspectie gaat het hier echter om “een zeer klein deel van de Nederlandse basisscholen” (Inspectie van het Onderwijs, 2008, p. 47).

De steekproeftrekking in het voortgezet onderwijs is aanzienlijker complexer dan die in het basisonderwijs, doordat er veel verschillende (combinaties van) onderwijstypen

bestaan en doordat er vaak sprake is van (zeer) grote scholen, die verdeeld zijn over zelfstandige lokaties.

Wij hebben op dit moment nog geen preciese gegevens over de eerste COOL-steekproef in het voortgezet onderwijs. Duidelijk is dat er de nodige aandacht zal worden gegeven aan de representativiteit van die steekproef en dat er zonodig een wegging zal plaatsvinden om die representativiteit te optimaliseren.

### **5.3 Uitval**

Een belangrijk probleem bij alle longitudinale onderzoeken is de uitval van onderzoekseenheden tussen de opeenvolgende metingen (*panelmortaliteit*).

Zowel bij PRIMA als bij VOCL is steeds sprake geweest van een omvangrijke uitval. Bij het ontwerp van COOL is gewezen op de nieuwe mogelijkheden die de invoer van het onderwijsnummer biedt om uitgevallen leerlingen op te sporen en te blijven volgen.

Hoewel deze mogelijkheden zich nog wel in de praktijk moeten bewijzen (er kunnen zich nog veel problemen voordoen rond de koppeling van identificerende gegevens als namen/geboortedata uit de onderzoeksdata en de gegevens uit de onderwijsnummerbestanden) levert dit wel aantrekkelijke perspectieven.

Voorals de sterk selectieve ‘individuele’ uitval (door zittenblijven, verwijzing naar speciaal onderwijs) en de minder selectieve uitval van leerlingen die verhuizen zal kunnen verminderen. Uit het onderwijsnummerbestand kan achterhaald worden op welke school een uitgevallen leerling vervolgens zit en op die school kunnen van die leerling weer scores op de (in de praktijk veel gebruikte) Cito-toetsen opgevraagd worden. Mogelijk zullen ook de andere instrumenten van COOL bij die leerlingen worden afgenomen.

Naast deze ‘individuele’ uitval is er echter ook sprake van een veel omvangrijker uitval van hele scholen, die niet meer aan het onderzoek willen deelnemen. Ook van leerlingen op zulke scholen kan uit het onderwijsnummerbestand informatie worden

verkregen over hun onderwijspositie (type onderwijs, groep of klas) en over een eventuele Cito-eindtoets voor het basisonderwijs en eindexamenresultaten voor het voortgezet onderwijs.

Naar verwachting zal het specifieke COOL-instrumentarium echter niet voor al deze leerlingen gebruikt kunnen worden, zodat gegevens over een doorlopende schaal van geletterdheid en gecijferdheid zullen ontbreken.

Zoals eerder gesteld is deze uitval van hele scholen echter tot nu toe niet systematisch gebleken.

Een laatste punt hierbij is dat ouders kunnen weigeren dat gegevens uit het onderwijsnummerbestand gekoppeld gaan worden. Bij het basisonderwijs is wat dit betreft een procedure van *informed consent* gebruikt: ouders zijn op de hoogte gesteld van opzet en doel van het onderzoek, inclusief de koppeling aan gegevens van het onderwijsnummerbestand. Daarbij zijn zij gewezen op de mogelijkheid om hun kind niet mee te laten doen aan het onderzoek, door dit aan de school door te geven of rechtstreeks aan de onderzoekers.

Er is op dit moment nog geen precies overzicht van de aantallen ouders die van deze mogelijkheid gebruik hebben gemaakt. (Een eerste schatting is: minder dan 1%.) Ook over de selectiviteit van deze specifieke vorm van uitval is op dit moment nog geen informatie beschikbaar.

## **5.4 Toetsen**

Bij het COOL-onderzoek wordt, voor de cognitieve vaardigheden, uitgegaan van bestaande toetsen uit het Cito leerlingvolgsysteem (LOVS) in het basisonderwijs en het Cito Volg en Adviesstelsel (VAS) in het voortgezet onderwijs. Bij het gebruik van deze toetsen voor het analyseren door de tijd van leerwinsten spelen in principe twee onderscheiden problemen:

*Verkenning leerwinst als indicator voor onderwijskwaliteit*

*- de toetsscores moeten door de tijd vergelijkbaar zijn en blijven*

Bij gebruik van de toetsen zoals Cito die heeft opgenomen in LOVS en VAS speelt het probleem dat deze toetsen na verloop van tijd vervangen worden door nieuwe versies, omdat de oude toetsen, naar het oordeel van Cito, te bekend worden en daarmee minder geschikt om de leervorderingen van leerlingen te meten. Concreet speelt dat op dit moment bijvoorbeeld bij de toetsen Rekenen/Wiskunde en Begrijpend Lezen in het basisonderwijs, maar ook bij andere toetsen valt te verwachten dat 'oude' versies gaandeweg worden opgevolgd door 'nieuwe' versies.

Om de toetsscores toch door de tijd te kunnen vergelijken wordt door Cito steeds getracht om een omrekening te maken waarmee een score op een 'oude' toets vergelijkbaar wordt met die op een 'nieuwe' toets. Voor Begrijpend Lezen is dat, volgens Cito, goed mogelijk. Voor Rekenen/Wiskunde moet daar nog nader onderzoek naar gedaan worden.

We merken wel op dat, ook als een omrekening mogelijk (d.w.z. psychometrisch verantwoord) is, er natuurlijk altijd een grotere foutenmarge rondom een 'omgerekende' toetsscore zal bestaan dan rond de oorspronkelijke toetsscore.

*- de toetsscores moeten de bepaling van leerwinst mogelijk maken*

Hiervoor is het nodig dat scores op de diverse toetsen, die op verschillende leeftijden (groep 2, groep 5, groep 8, 3<sup>e</sup> jaar vo en 6<sup>e</sup> jaar vo) worden afgenomen, vertaald kunnen worden naar één onderliggende schaal. In het COOL-onderzoek wordt gesproken van schalen voor geletterdheid en gecijferdheid.

Bij de toetsen uit het Cito-LOVS (basisonderwijs) is dit voor de hogere groepen 5 en 8 het geval. Er is voor de toetsen die in deze groepen worden afgenomen sprake van een doorlopende vaardigheidsschaal (bijv. voor de binnen COOL gebruikte toetsen Rekenen/Wiskunde, Begrijpend Lezen en Leeswoordenschat).

Bij de verbinding tussen deze vaardigheidsschalen naar 'onderen' (de kleutergroepen) en naar 'boven' (de VAS-toetsen in het voortgezet onderwijs) zijn er beduidend meer problemen. Dit wordt door Peschar (2007) ook al aangegeven in zijn paradox 1:

“Het instrument moet gedurende de gehele levensloop van 5 tot 18 jaar goed toepasbaar zijn.(.....)

**Paradox 1** De eerste paradox ontstaat doordat de competenties van een zesjarige leerling (nog) maar een beperkt bereik hebben in vergelijking met die van een achttienjarige student. Per definitie is er geen gemeenschappelijk instrument dat alle competenties voor beide leerlingen compleet in beeld kan brengen. Er zijn vaardigheden die een zesjarige nog niet kan hebben ontwikkeld (in vergelijking met kinderen op latere leeftijd) terwijl er omgekeerd vaardigheden van jongere leerlingen zijn die niet meer in het repertoire van oudere leerlingen voorkomen. Er is dus geen perfecte oplossing. Wel is er een keus, namelijk om de meting in te perken tot die competenties die redelijk in elkaars verlengde liggen. Meestal komen taal- en rekenvaardigheden daarvoor in aanmerking als doorsnee van de competenties van beide leeftijdsgroepen.” (p.12)

Terwijl er dus geen perfecte oplossing is wordt door Cito wel steeds gewerkt aan zo goed mogelijke benaderingen voor doorlopende vaardigheidsschalen. Zo zijn er goede perspectieven om de toets Taal voor Kleuters uit groep 2 te verbinden met Woordenschattoetsen uit de hogere groepen. Ook de koppeling van de taaltoetsen uit VAS aan de vaardigheidsschalen zoals die in het basisonderwijs gebruikt worden lijkt mogelijk.

Bij ‘gecijferdheid’ (doorlopende schalen voor rekenen/wiskunde) zijn de problemen groter. Aan de onderkant gaat het dan om de vraag of scores op de (kleuter-)toets Ordenen wel opgevat kunnen worden als een bepaalde mate van cijfervaardigheid (cijfers zijn in die toets nog niet aan de orde). En aan de bovenkant speelt het probleem dat de ontwikkeling in het basisonderwijs vooral de reken-vaardigheid betreft, terwijl het in het voortgezet onderwijs vooral om de ontwikkeling van de wiskunde-vaardigheid gaat. (Doordat in het voortgezet onderwijs veelvuldig de rekenmachine wordt gebruikt is het zelfs goed mogelijk dat de reken-vaardigheid daar achteruitgaat, terwijl de meer abstracte wiskunde-vaardigheid toeneemt.)

*Verkenning leerwinst als indicator voor onderwijskwaliteit*

Uit het gesprek met Cito kwam naar voren dat er voor de twee bovengenoemde problemen geen perfecte oplossing bestaat. In termen van Peschar: een paradox laat geen perfecte oplossing toe. Tegelijk lijken voor beide problemen redelijke benaderingen van oplossingen mogelijk, die analyses op COOL in termen van leerwinst in principe mogelijk maken. Wel zullen de onzekerheidsmarges bij dergelijke benaderingen groter zijn dan bij de ideaaltypische situatie dat éénzelfde, onveranderde toets door alle jaren en bij alle leeftijden afgenomen wordt.

## 6 Slotbeschouwing

1. Doorslaggevend in de keuze voor een bepaalde ‘opbrengstmaat’ van het onderwijs is de vraag die daarmee beantwoord zou moeten worden.

Als die vraag is: “wat is het onderwijsniveau, bijvoorbeeld aan het eind van het basis-onderwijs of na de eerste fase van voortgezet onderwijs?” dan ligt het voor de hand met een of meerdere geschikte instrumenten (toetsen) de competenties van leerlingen op dat moment van hun loopbaan te meten.

Een voorbeeld hiervan zijn de referentieniveaus, zoals die voorgesteld worden door de commissie Meijerink (*Over de drempels met taal en rekenen*, 2008). Maar ook internationale onderzoeken als Pisa geven zo’n ‘momentopname’ van wat Nederlandse leerlingen op een bepaalde leeftijd kunnen en kennen.

Maar meestal is de (beleids-)vraag specifiek: “wat is de bijdrage van het onderwijs aan dat onderwijsniveau?”. Dan gaat het om de vraag of de bijdrage van het onderwijs valt te isoleren van meer of minder stimulerende overige factoren, waar het onderwijs geen controle over heeft (zoals aanleg, thuismilieu, financiële omstandigheden, sociaal kapitaal, peer group etc.)

Eén poging daartoe is om uit te gaan van **leerwinst**. Daarbij is er sprake van een ‘begin’ meting en een eind meting. Het eindniveau van een leerling wordt dan in elk geval gecorrigeerd voor een sterkere of zwakkere startpositie.

Vaak wordt er, daarnaast of daarbovenop, ook gewerkt met **value added models**, waarin de toegevoegde waarde van het onderwijs wordt geschat. In zulke modellen wordt langs statistische weg gecorrigeerd voor zoveel mogelijk van de hiervoor genoemde meer of minder stimulerende factoren. De meting van die factoren gaat vaak via benaderingen (proxy’s); bijvoorbeeld het opleidingsniveau van ouders als benadering van onderwijsondersteunend thuismilieu en sociaal kapitaal.



### *Verkenning leerwinst als indicator voor onderwijskwaliteit*

De value added models zijn vooral bekend van het “zo eerlijk mogelijk” onderling vergelijken van de opbrengsten van scholen, door rekening te houden met de verschillende leerlingpopulaties op die scholen.

De modellen leveren daarmee relatieve informatie op: welke school presteert, na correctie voor de leerlingsamenstelling, beter dan gemiddeld en welke school minder dan gemiddeld. Zoals Peschar (2007) aangeeft zijn dergelijke relatieve indicatoren minder geschikt om te gebruiken als stelselindicator.

Bij een stelselindicator is de achterliggende vraag wat de bijdrage van (een bepaalde fase van) het onderwijsstelsel is aan het onderwijsniveau dat leerlingen in de betreffende fase van het onderwijs bereiken. Daarbij gaat het dan dus niet alleen om het eindniveau, maar ook om de groei, de leerwinst, van de leerlingen in die fase.

Zoals we betoogd hebben is het echter niet voldoende om louter naar die winst te kijken, maar moeten leerwinst en niveau in samenhang worden beschouwd. De verkenningen op PRIMA met groeimodellen laten zien dat daarmee die samenhang kan worden gemodelleerd, terwijl ook de ontwikkeling van specifieke groepen leerlingen nader kan worden geanalyseerd.

2. Naast dit principiële punt (welke vraag willen we beantwoord hebben) kunnen er ook andere, vaak meer praktische overwegingen zijn bij de keuze voor een indicator.

Dat betreft bijvoorbeeld de termijn waarop gegevens beschikbaar kunnen komen.

Bij een indicator die uitgaat van het onderwijsniveau op een bepaald moment zijn de benodigde gegevens snel beschikbaar: er hoeft maar één momentopname gemaakt te worden. Voor een indicator die uitgaat van leerwinst is die termijn altijd (veel) langer. Er zijn tenminste metingen op twee opeenvolgende tijdstippen nodig en bij voorkeur nog meer dan twee metingen.

In het COOL<sup>5-18</sup> onderzoek is voorzien in metingen om de drie jaar. Daarmee zal het geruime tijd duren voordat leerwinstmetingen of zelfs groeimodellen (waarbij tenminste drie metingen nodig zijn) op basis van de representatieve steekproef uit dat cohortonderzoek gebruikt kunnen worden als stelselindicatoren.

De enige mogelijkheid om sneller leerwinst-gegevens vanuit COOL te kunnen gebruiken is een verkorting van de drie-jaars-cyclus van het onderzoek.

Zo is het in principe mogelijk om van de leerlingen in het onderzoek ook tussentijds, jaarlijks tussen de onderzoeksjaren in, toetscores bij de deelnemende basisscholen op te vragen. In de onderzoeksopzet van het COOL-onderzoek zelf is dit niet voorzien en daar zou dus extra dataverzameling en financiering voor nodig zijn.

In het voortgezet onderwijs kunnen met weinig moeite jaarlijkse onderwijspositiegegevens (onderwijstype én leerjaar) verkregen worden uit het reeds operationele onderwijsnummerbestand. Daarmee kan de ‘groei’ in die onderwijsposities geanalyseerd worden. Voor de groei in (cognitieve) vaardigheden zouden ook in het voortgezet onderwijs echter extra tussentijdse metingen nodig zijn.

Een andere overweging is of een stelselindicator haaks zou kunnen staan op indicatoren voor toegevoegde waarde (op basis van leerwinst), zoals die bijvoorbeeld door de Onderwijsinspectie zullen gaan worden gebruikt. Uit ons gesprek met de Inspectie kwam echter naar voren dat men daar nog volop bezig is met het verkennen van mogelijkheden daartoe en dat nog allerminst duidelijk is hoe (school-)indicatoren op basis van leerwinst er uit zullen komen te zien.

Wel is duidelijk dat bij zulke (school-)indicatoren verschillende relevante leerlingkenmerken een rol zullen moeten gaan spelen. In die zin is niet te verwachten dat stelsel-indicatoren, die ontwikkelingen bij verschillende groepen leerlingen nader specificeren, haaks zullen komen te staan op de Inspectie-indicatoren.

Verder is het, zoals in de notitie van Peschar ook wordt gesteld, van belang dat een stelselindicator ook min of meer “intuïtief” begrepen kan worden door een groter publiek dan alleen onderzoekers. In deze verkenningen hebben wij ons geconcentreerd op de “methodologie” van de opzet van indicatoren voor het onderwijsstelsel en zijn de uitkomsten op basis van de PRIMA-data, zoals bijvoorbeeld de verschillen naar etnische herkomst, meer als illustratie van die methodologie gebruikt.

Er zal zeker de nodige moeite gedaan moeten worden om van de parameterschattingen in de statistische modellen naar meer direct begrijpbare uitspraken over

(ontwikkelingen in) het onderwijsstelsel als geheel te komen. Maar het moet mogelijk zijn om dat te doen in termen van bijvoorbeeld groei (leerwinst) over een bepaalde periode en het al dan niet inlopen van achterstanden door specifieke groepen leerlingen. En dat lijken ons begrippen die goed uit te leggen en te begrijpen zijn, ook voor een breder publiek.

3. Tot slot keren we terug naar het begin van deze slotbeschouwing: welke vraag willen we beantwoord hebben als we de opbrengsten van het onderwijsstelsel gaan bepalen?; en welke indicatoren zijn dan nuttig en nodig?

Voor de relatief eenvoudige vraag ‘wat kunnen en kennen Nederlandse leerlingen op een bepaald moment?’ kunnen referentieniveaus, PPO-metingen of indicatoren uit internationale onderzoeken als PISA en TIMSS gebruikt worden.

Voor meer complexe vragen als de bijdrage van het onderwijs aan de ontwikkeling van leerlingen zullen ook complexere dataverzamelingen en analysemethoden gebruikt moeten worden. Op basis van deze verkenning pleiten we er voor longitudinale cohortdata te gebruiken, zoals die bijvoorbeeld in COOL<sup>5-18</sup> verzameld worden. Daarmee kan, op kortere of langere termijn, de groei van alle of van onderscheiden groepen leerlingen in het onderwijs worden gemodelleerd en zo getracht worden om de bijdrage van het onderwijs aan die groei zo goed mogelijk te onderscheiden van allerlei buitenschoolse factoren.

## Literatuur

Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2008). Over de drempels met taal en rekenen.

Inspectie van het Onderwijs (2008). Analyse en waardering van opbrengsten. Primair Onderwijs.

Utrecht: Inspectie van het Onderwijs

Mulder, L., Roeleveld, J., Veen, I. van der & Vierke, H. (2005) Onderwijsachterstanden tussen 1988 en 2002: ontwikkelingen in basis- en voortgezet onderwijs.

Nijmegen/Amsterdam: ITS/SCO-Kohnstamm Instituut

Peschar, J.L. (2007) Over leerwinst als stelselindicator. Haren

Roeleveld, J & R. Portengen (1998). Uitval en instroom bij het PRIMA-cohortonderzoek

Amsterdam/Nijmegen: SCO-Kohnstamm Instituut / ITS

Roeleveld, J. & Vierke, H. (2003). Uitval en instroom bij de derde meting van het PRIMA-cohortonderzoek. Intern onderzoeksrapport PRIMA. Amsterdam/Nijmegen: SCO-Kohnstamm Instituut/ITS

Schooten, E. van, Smeets, E. & Driessen, G. (2007). Taaltoetsen voor taalmazke leerlingen in het primair onderwijs. Amsterdam, SCO-Kohnstamm Instituut

Vierke & Mulder (2005). Migratie van leerlingen binnen en tussen basisscholen. Nijmegen: ITS

Vreeburg, B. (2008) Absoluut versus relatief beoordelen van de opbrengsten van scholen. Paper Onderwijs Researchdagen 2008.

Wolf, I. de & Vreeburg, B. (2008) Validiteit van opbrengstmaten: leerwinst op scholen. Paper Onderwijs Researchdagen 2008.



## Recent uitgegeven SCO rapporten

- 810 Derriks, M., Kat, E. de.,  
Maatjesaanpak in een Jenaplanschool
- 809 Jong, U. de, Ledoux, G., Emmelot, Y., Roeleveld, J.  
Opleiden in de school
- 808 Daalen, M. van, Vrieze, G., Karsten, S., Kessel, N. van  
Onbevoegd lesgeven: een noodoplossing.
- 805 Fukkink, R.  
Evaluatie van het SHare In Trust-project.
- 804 Peetsma, T., Veen, I. van der  
Een tweede onderzoek naar de beïnvloeding van motivatie bij vmbo-  
leerlingen.
- 803 Veen, A., Roeleveld, J., Daalen, M. van  
Implementatie en effecten van Voor- en Vroegschoolse Educatie in Rot-  
terdam. Samenvatting.
- 802 Blok, H., Karsten, S.  
Vervangend onderwijs aan kinderen met een richtingbezwaaar.
- 801 Karstanje, P.N., Glaude, M.T., Ledoux, G., Verbeek, A.E.  
Beleidsvoerend vermogen van basisscholen: stand van zaken en verkla-  
rende factoren.
- 799 Derriks, M., Kat, E. de, Voncken, E.  
Iedereen bij de les!
- 797 Roede, E., Derriks, M., Boogaard, M.  
Sociale Competentie in het Rotterdamse Onderwijs. Eindrapport.
- 796 Veen, I. van der,  
Amsterdamse Schakelklassen in het schooljaar 2006/2007.
- 795 Voncken, E., Derriks, M., Ledoux, G.  
Vernieuwen zonder blauwdruk, de plussen en minnen.
- 794 Voncken, E., Derriks, M., Ledoux, G., Slegers, P., Kock, J. de  
Een hele toer. Ervaringen van schoolleiders en docenten met de vernieu-  
wing van de onderbouw VO.
- 793 Roede, E., Derriks, M.  
Zijn Montessori scholen toekomstbestendig ingericht?
- ITS'07 Smeets, E., Veen, I. van der, Derriks, M., Roeleveld, J.  
Zorgleerlingen en leerlingenzorg op de basisschool.
- 791 Verbeek, F., Felix, C.  
Resultaten van de startersopleiding van de Academie voor Wetgeving.

Deze rapporten zijn te bestellen via:

<http://www.sco-kohnstamminstituut.uva.nl/webwinkel/bestellen.htm>