

Schooleffectiviteit en prestatieniveau natuuronderwijs in groep 6: secundaire analyses op TIMSS-2007 data

M.R.M. Meelissen en H. Luyten

Samenvatting

In internationaal perspectief zijn Nederlandse groep 6 leerlingen aanmerkelijk minder goed in natuuronderwijs dan in rekenen-wiskunde. Dit blijkt uit *Trends in International Mathematics and Science Study (TIMSS)-2007*, waar in 36 landen een internationale toets voor rekenen-wiskunde en natuuronderwijs onder 10-jarige leerlingen is afgenomen. Bovendien is Nederland één van de zeven TIMSS-landen waar jongens significant beter in natuuronderwijs presteren dan meisjes. Aan de hand van het schooleffectiviteitsmodel van Scheerens (2008) is met meerniveau-regressie-analyses op de data van bijna 3300 leerlingen, geëxploreerd in hoeverre de in TIMSS gemeten context-, input-, en procesfactoren samenhangen met toetsprestaties in natuuronderwijs in Nederland. Gecontroleerd voor een aantal leerlingkenmerken (*seks*e, *ethniciteit*, *aantal boeken thuis* en *wel/geen vertraagde leerling*) blijken vooral contextkenmerken (*gemiddeld leerlinggewicht van de school* en *denominatie*) voorspellend te zijn van prestaties in natuuronderwijs. Er is tevens een positief verband tussen de perceptie van de schoolleider over het *schoolklimaat* en de prestaties van leerlingen op de toets. Leerlingen lijken verder beter te presteren op de TIMSS-toets naarmate de leerkracht zich beter toegeest acht om les te geven in de verschillende leerstofonderdelen van natuuronderwijs. Ook het gemiddelde oordeel van de leerlingen over het *cognitieve en sociale klimaat in de klas* hangt positief samen met prestaties in natuuronderwijs.

1 Aanleiding

In het kader van de Lissabondoelstellingen van 2000 is in het afgelopen decennium meer aandacht gekomen voor het onderwijs in de natuurwetenschappelijke vakken op de basisschool (Wijnen, 2005). Zo heeft het Platform

Bèta Techniek sinds 2003 verschillende projecten geïnitieerd ter bevordering van de exacte wetenschappen en techniek in het basisonderwijs en ter vergroting van de interesse van meisjes voor deze vakgebieden (Platform Bèta Techniek, 2009). TIMSS-2007 (*Trends in International Mathematics and Science Study*) laat echter zien dat in internationaal perspectief, het Nederlandse basisonderwijs relatief weinig aandacht aan de natuurwetenschappelijke vakken besteedt en dat Nederlandse groep 6 leerlingen niet uitblinken in hun kennis hiervan. Bovendien is Nederland een van de weinige landen waar 10-jarige meisjes minder goed in natuuronderwijs presteren dan jongens.

In TIMSS-2007 hebben in het voorjaar van 2007 zo'n 4400 Nederlandse leerlingen in groep 6 van het basisonderwijs een internationale toets voor rekenen-wiskunde en natuuronderwijs gemaakt. De toets is in 36 landen afgenomen (Martin, Mullis, & Foy, 2008). Het voornaamste doel van TIMSS is een vergelijking tussen landen in onderwijsniveau- en context. Er wordt een onderscheid gemaakt in drie curriculumniveaus: het beoogde curriculum (wat onderwezen zou moeten worden), het uitgevoerde curriculum (wat onderwezen wordt) en het gerealiseerde curriculum ofwel leerlingprestaties (Mullis, Martin, Ruddock, O'Sullivan, Arora, & Erberer, 2005). Het gerealiseerde curriculum wordt gemeten met de internationale TIMSS-toets. Met vragenlijsten wordt informatie verzameld over het beoogde en uitgevoerde curriculum en over factoren waarvan wordt aangenomen dat zij op deze drie niveaus van invloed zijn. Omdat TIMSS sinds 1995 om de vier jaar herhaald wordt, biedt de studie ook de mogelijkheid om op landniveau ontwikkelingen in de tijd te volgen.

In internationaal perspectief blijken Nederlandse groep 6 leerlingen aanmerkelijk minder goed in natuuronderwijs te zijn dan in rekenen-wiskunde. De gemiddelde toetscore van de groep 6 leerlingen voor natuuronder-

wijs in TIMSS-2007, eindigde op de 17e plaats van de internationale ranglijst (Martin et al., 2008). Tien van de 36 landen haalden een significant hogere gemiddelde toetscore dan Nederland. Nederland is sinds 1995 voor natuuronderwijs ingehaald door landen zoals Singapore, Hongkong, Engeland en Hongarije. Rekenen-wiskunde laat weliswaar een kleine, significante daling in gemiddelde toetscore zien, maar de Nederlandse leerlingen behoren nog steeds tot de internationale reken top tien. Alleen de leerlingen uit vier Aziatische landen hebben gemiddeld een significant hogere toetscore voor rekenen gehaald dan de Nederlandse leerlingen.

In verreweg de meeste landen wordt meer tijd aan natuuronderwijs besteed dan in Nederland (Meelissen & Drent, 2008). Tussen Nederlandse scholen zijn er bovendien behoorlijke verschillen in het natuuronderwijsaanbod. Zo is op bijna 30% van de scholen natuuronderwijs geïntegreerd in andere vakken. Op de overige scholen wordt natuuronderwijs als een apart vak aangeboden. Op deze scholen wordt gemiddeld meer tijd aan natuuronderwijs besteed (53 minuten per week) dan op scholen waar het vak geïntegreerd is in andere vakken (43 minuten per week). In vergelijking tot veel andere TIMSS-landen, komen Nederlandse leerlingen ook minder in aanraking met proefjes en experimenten. Ook hierin zijn tussen de Nederlandse basisscholen onderling duidelijke verschillen (Meelissen & Drent, 2008).

Ten slotte blijkt uit TIMSS-2007 dat in zeven van de 36 landen, waaronder Nederland, meisjes een significant lagere score op de natuuronderwijstoets hebben behaald dan jongens (Martin et al., 2008). In de overige landen zijn er geen sekseverschillen of presteren meisjes beter dan jongens. In Nederland hebben jongens op alle drie de inhoudsgebieden (biologie, natuur- en scheikunde en fysische aardrijkskunde) beter gepresteerd dan meisjes. In TIMSS-1995 en -2003 waren jongens ook beter dan meisjes in natuur- en scheikunde en fysische aardrijkskunde, maar in biologie presteerden jongens en meisjes nog gelijk (Knuver & Doolaard, 1997; Meelissen & Doornekamp, 2004).

2 Onderzoeksvragen

De tegenvallende resultaten voor natuuronderwijs in TIMSS voor Nederland zou mogelijk te maken kunnen hebben met de wijze waarop natuuronderwijs in Nederland wordt aangeboden. Hierin blijkt behoorlijk wat variantie te zijn tussen de scholen. Een identificatie van factoren die in de Nederlandse situatie samenhangen met goede prestaties in natuuronderwijs lijkt daarom relevant. Gebruik makend van de TIMSS-data van 2007, is in deze studie geëxploreerd in hoeverre school-, klas- en instructiekenmerken in Nederland samenhangen met verschillen in leerprestaties in natuuronderwijs, gecontroleerd voor de in TIMSS gemeten achtergrondkenmerken van de groep 6 leerlingen. De eerste onderzoeksvraag van deze studie luidt: In hoeverre hangen school-, klas- en instructiekenmerken samen met de prestaties van groep 6 leerlingen op de TIMSS-2007 natuuronderwijstoets, gecontroleerd voor leerlingkenmerken?

In vergelijking met rekenen-wiskunde en leesvaardigheid zijn er tot nu toe in TIMSS-landen weinig secundaire analyses op TIMSS-data uitgevoerd die gericht zijn op de identificatie van factoren die van belang zijn voor het opbrengen van natuuronderwijs in het basisonderwijs (Kaya, 2008). Ook secundaire analyses op de Nederlandse TIMSS-data van 1995 en 2003 hebben zich tot nu toe enkel op rekenen en wiskunde gericht (Bos & Meelissen, 2006; Meelissen & Luyten, 2008). De tweede onderzoeksvraag van deze studie is: In hoeverre zijn er voor de school-, klas- en leerlingkenmerken waarvan bij vraag 1 is gebleken dat zij significant samenhangen met toetsprestaties, verschillen tussen jongens en meisjes in de mate van samenhang?

Vanwege de minder goede prestaties van meisjes ten opzichte van jongens in natuuronderwijs in Nederland, zijn voor de factoren die samenhangen met toetsprestaties ook de differentiële effecten voor sekse berekend. De samenhang tussen de verschillende kenmerken en toetsprestaties hoeft namelijk niet voor alle leerlingen even sterk te zijn. Zo vinden Maeyer en Rymenans (2004) in het voortgezet onderwijs wel verband tussen verschillende schoolkenmerken en de leerpresta-

ties van jongens, maar geen verband tussen deze kenmerken en de leerprestaties van meisjes. Er zijn overigens ook studies waarin geen differentiële effecten voor sekse zijn aangetroffen, maar bijvoorbeeld wel voor sociaal economische status (SES) of etniciteit (Maeyer & Rymenans, 2004).

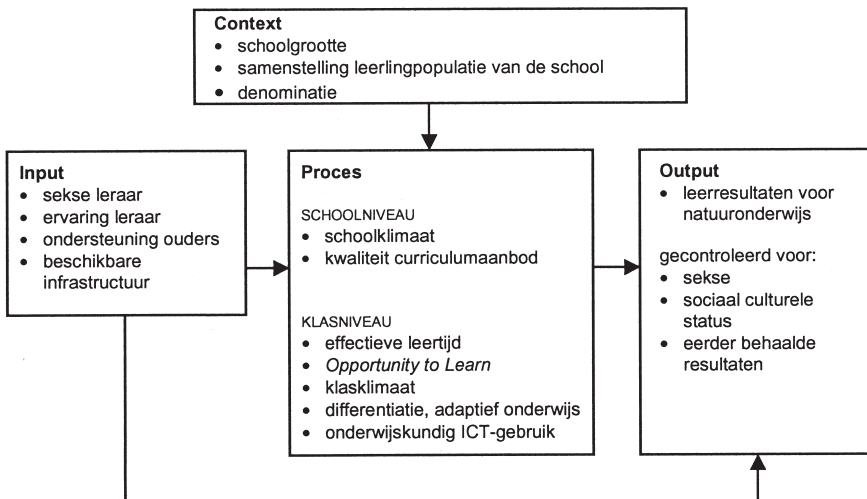
De voorliggende studie is exploratief: het gaat om de identificatie van factoren die samenhangen met toetsprestaties in natuuronderwijs. Het schooleffectiviteitsmodel van Scheerens (1999, 2008) is gebruikt om relevante indicatoren hiervoor in de TIMSS-data te selecteren. Dit model wordt in de volgende paragraaf verder toegelicht.

3 Conceptueel raamwerk en selectie van indicatoren

Omdat het aantal studies naar de effectiviteit van het onderwijs in de natuurwetenschappelijke vakken in het basisonderwijs beperkt is en er geen bruikbaar conceptueel model voor handen lijkt te zijn, is voor de beantwoording van de onderzoeksvragen gebruik gemaakt van het algemene, geïntegreerde model voor schooleffectiviteit van Scheerens (1999, 2008). Dit model is in deze studie gebruikt om relevante factoren in de TIMSS-data te kunnen identificeren en om de volgorde van de analysestappen te bepalen. Het schooleffecti-

viteitsmodel is een conceptueel raamwerk dat voornamelijk gebruikt wordt ter ondersteuning van meta-analyses naar onderwijs-effectiviteit in het basis- en voortgezet onderwijs (Scheerens, Luyten, Steen, & Luyten-de Thouars, 2007). In dit raamwerk, dat gebaseerd is op het klassieke input-proces-outputmodel, vormt de school als organisatie het uitgangspunt. Er wordt een onderscheid gemaakt tussen leerlingkenmerken, context-, input-, proces- en outputfactoren. Procesfactoren kunnen verder onderverdeeld worden naar factoren op schoolniveau, klasniveau en factoren die op beide niveaus van belang kunnen zijn. Het raamwerk staat weergegeven in Figuur 1, evenals de invulling van de factoren met indicatoren die in TIMSS gemeten worden. Deze worden hieronder verder toegelicht.

De afhankelijke variabele in dit onderzoek (outputfactor) is de leerlingsscore op de TIMSS-toets voor natuuronderwijs die in het voorjaar van 2007 in groep 6 is afgenomen. Aan elke leerling zijn vijf *plausible values* toegekend in plaats van één enkele toetsscore. De reden voor het gebruik van *plausible values* heeft te maken met het toetsrotatiesysteem van TIMSS. De primaire doelstelling van internationaal vergelijkende studies zoals TIMSS, is het zo nauwkeurig mogelijk schatten van het gemiddelde prestatieniveau voor nationale populaties. De prioriteit ligt bij



Figuur 1. Conceptueel raamwerk van deze studie, gebaseerd op het schooleffectiviteitsmodel van Scheerens (2008).

het minimaliseren van meetfouten rond het populatiegemiddelde; het minimaliseren van meetfouten op het individuele niveau komt op de tweede plaats. Met het toetsrotatiesysteem, waarbij de toetsitems zijn verdeeld over verschillende toetsboekjes en een leerling één van de random toegewezen toetsboekjes maakt, kan een groter aantal items aan de totale groep leerlingen worden voorgelegd dan wanneer elke leerling dezelfde toets maakt. Het populatiegemiddelde wordt zodoende nauwkeuriger geschat, maar de individuele toetsscores (bijvoorbeeld in percentages correct) zijn in eerste instantie alleen vergelijkbaar tussen leerlingen die hetzelfde toetsboekje en daardoor maar een deel van de toets hebben gemaakt. Berekeningen op basis van de *Item Response Theorie* leveren wel vergelijkbare leerlingsscores op, maar kennen een relatief grote foutmarge. Individuele toetsscores zijn echter in secundaire analyses op TIMSS-data een veel gebruikte outputmaat. In TIMSS wordt daarom de geobserveerde score van een leerling gezien als een indicatie van het bereik waarbinnen iemands ware vaardigheid is gelegen. Er wordt een verdeling van vaardigheidsscores geconstrueerd, waarbij ook rekening wordt gehouden met achtergrondkenmerken van de leerling. *Plausible values* zijn op toevalsbasis getrokken waarden uit deze verdeling (Martin et al., 2008; OECD, 2005). Het toetsrotatiesysteem maakt het mogelijk dat alle onderdelen van het TIMSS-curriculum in voldoende mate in de toets aan de orde komen (Mullis et al., 2005). De geschiktheid van de toets voor Nederland is geëvalueerd aan de hand van de *Test Curriculum Matching Analysis* (TCMA, Meelissen & Drent, 2008). Volgens curriculumexperts is bijna 70% van de natuuronderwijsopgaven passend bij de Nederlandse kerndoelen zoals deze zijn vertaald in de meest gebruikte natuuronderwijsmethoden.

Achtergrondkenmerken van leerlingen worden in het schooleffectiviteitsmodel expliciet onderscheiden van inputfactoren. Door te controleren voor leerlingkenmerken kan de samenhang vastgesteld worden tussen de overige factoren in het model en het zogenoemde 'nettoprestatieniveau' van de leerlingen (Bosker, 1991). Dit is het prestatieniveau

van de leerlingen gecontroleerd voor hun achtergrond en begaafdheid. De mate waarin in deze studie gecontroleerd kan worden voor deze leerlingkenmerken is echter beperkt. Omdat 10-jarigen moeite hebben om het opleidingsniveau van hun ouders aan te geven, wordt in TIMSS het *aantal boeken thuis* (weergegeven met plaatjes van boekenplanken in de leerlingvragenlijst) als een van de indicatoren voor SES gebruikt. Dit verwijst eerder naar de sociaal culturele status van een leerling dan zijn of haar sociaal economische status. Uit onderzoek blijkt echter dat dit kenmerk zeer sterk samenhangt met andere indicatoren voor SES (Lubienski & Crane, 2010).

Daarnaast wordt begaafdheid van de leerling niet gemeten in TIMSS. Op basis van de geboortedatum is wel berekend of een leerling in de afgelopen zes jaar wel of geen vertraging in zijn of haar schoolcarrière heeft opgelopen. Dit is uiteraard een onnauwkeuriger indicator voor de begaafdheid van de leerling dan bijvoorbeeld een intelligentietoets.

In het schooleffectiviteitsmodel verwijzen de contextfactoren naar structurele land- en schoolkenmerken. In deze studie gaat het om kenmerken van de school, zoals *denominatie* en de samenstelling van de leerlingpopulatie, weergegeven met het *schoolgewicht*. Het *schoolgewicht* is gebaseerd op de oude gewichtenregeling van vóór 2006. Dit betekent dat zowel het opleidingsniveau van de ouders als de etniciteit van de leerling het gewicht bepalen. Hoe hoger het *schoolgewicht* hoe meer leerlingen er op de school zitten met een 1,25 (laag opgeleide autochtone ouders) of 1,9 (laag opgeleide allochtone ouders) indicatie.

In deze studie zijn verder drie inputfactoren meegenomen: kenmerken van de leerkracht (*seks, kennis en vaardigheden*), de *mate waarin ouders het leerproces ondersteunen* en kenmerken van de beschikbare infrastructuur. In dit geval is dit beperkt tot beschikbaarheid van computers (*computer-leerlingratio*).

Het belangrijkste kenmerk van procesfactoren is dat zij manipuleerbaar zijn vanuit het perspectief van de school. De TIMSS-data bevat zeer veel verschillende indicatoren

die tot de procesfactoren gerekend zouden kunnen worden. Voor de selectie van relevante indicatoren is daarom gebruik gemaakt van de lijst met twaalf ‘algemene effectiviteitsverhogende’ procesfactoren die op basis van de meta-analyses van onderwijs-effectiviteitsstudies is samengesteld (Scheerens, 2008). Voor vijf van deze procesfactoren zijn in de TIMSS-data indicatoren aanwezig, voor de overige factoren zijn er geen goede indicatoren beschikbaar. Factoren waaraan in TIMSS-2007 geen aandacht is besteed zijn ondermeer de prestatieverwachtingen van schoolleider en leraren, onderwijskundig leiderschap en het evaluerend vermogen van de school. De procesfactoren en bijbehorende variabelen waarover in TIMSS wel informatie is verzameld, staan verder toegelicht in Tabel 1. Hiervan is het *klimaat in de klas* een zogenoemde nationale optie (wordt alleen in Nederland bevestigd). Het *klimaat in de klas* is gemeten door aan individuele leerlingen een aantal positieve en negatieve stellingen voor te leggen die betrekking hebben op cognitieve veiligheid en sociale veiligheid. Vervolgens is per klas een gemiddelde score berekend.

In het schooleffectiviteitsmodel en ook in deze studie zijn affectieve leerlingfactoren, zoals plezier, zelfvertrouwen of welbevinden buiten beschouwing gelaten omdat ze kunnen worden gezien als een ander type, niet-cognitieve, outputfactor en waarbij bovendien de kans op wederkerige beïnvloeding zeer groot is (Creemers, 1994; Ma & Kishor, 1997; Scheerens, 2008; Wang, Oliver, & Staver, 2008).

4 Methode

4.1 Onderzoekspopulatie

Aan TIMSS-2007 hebben 4444 leerlingen uit groep 6 deelgenomen. Hiervan hebben 110 leerlingen geen leerlingvragenlijst ingevuld. Een random deelsteekproef van 1095 leerlingen heeft extra trendtoetsboekjes gemaakt, maar deze leerlingen hebben geen toetsscore gekregen (zie voor verdere uitleg Meelissen & Drent, 2008). Dit betekent dat de gegevens van 3239 leerlingen zijn geanalyseerd: dit zijn leerlingen die én de geschatte toetsscores

(vijf *plausible values*) hebben gekregen én de leerlingvragenlijst hebben ingevuld. Deze leerlingen zijn afkomstig van 114 scholen en 170 klassen. De data is gewogen met een gestandaardiseerde wegingsfactor, die corrigeert voor de uitval van de getrokken scholen, klassen en leerlingen (voor de berekening van deze wegingsfactor, zie Olson, Martin, & Mullis, 2008).

4.2. Analyseprocedure

De analyses op de TIMSS-data van 2007 zijn uitgevoerd aan de hand van meerniveau-regressieanalyses met het programma MLwiN (2.22). Deze techniek houdt rekening met de verschillende niveaus waarop de data verzameld is, maar ook met de geclusterde steekproeftrekking van TIMSS (steekproefeenheid is school, analyse-eenheid is leerling). In de meeste landen wordt binnen elke getrokken school één klas getrokken. In dat geval zijn er twee niveaus: een leerlingniveau en een hoger niveau. In Nederland zijn alle groepen 6 van de getrokken scholen benaderd voor deelname (Meelissen & Drent, 2008). In de Nederlandse data kunnen daardoor drie niveaus worden onderscheiden: leerling, klas en school.

Zoals eerder is aangegeven, zijn er in de TIMSS-data vijf *plausible values* voor natuuronderwijs beschikbaar. In deze studie zijn eerst zes modellen geanalyseerd met de eerste *plausible value* als de afhankelijke variabele. In deze analyses is op exploratieve wijze nagegaan welke onafhankelijke variabelen veelbelovend lijken als voorspeller van leerprestaties. Met het eerste, lege model wordt berekend in hoeverre de variantie in toetsscores samenhangt met verschillen tussen scholen (de tussenschoolse variantie), tussen parallelklassen van deze school (tussengroepse variantie) en verschillen tussen individuele leerlingen (binnengroepse variantie). Op basis van het schooleffectiviteitsmodel zijn vervolgens vijf modellen geanalyseerd. In model 1 zijn de beschikbare indicatoren voor de leerlingkenmerken opgenomen. Vervolgens zijn respectievelijk in model 2 tot en met 5 indicatoren voor contextfactoren, inputfactoren, procesfactoren op schoolniveau en procesfactoren op klasniveau opgenomen. Per model zijn de be-

treffende variabelen gelijktijdig aan het vorige model toegevoegd. Vervolgens zijn telkens de variabelen met niet-significante effecten op de eerste *plausible value* (tweezijdig, $\alpha < 0,05$, $|t| < 1,96$) uit het model verwijderd, voordat de variabelen van het volgende model worden toegevoegd. Door de indicatoren die tot dezelfde groep factoren behoren, gelijktijdig in één model op te nemen, zijn de effecten op systematische wijze geëxploreerd, zonder dat de modellen te complex of zelfs instabiel kunnen worden (Veenstra, 1999).

Alle variabelen zijn gestandaardiseerd (gemiddelde is 0, standaarddeviatie is 1). Een effect van een individuele factor op leerprestaties wordt relevant beschouwd bij een regressiecoëfficiënt van circa 0,10 (Scheerens, 2008).

Vervolgens zijn er voor de variabelen met een significant hoofdeffect, interactie-effecten voor sekse in het eindmodel opgenomen om vast te stellen in hoeverre er verschillen zijn tussen jongens en meisjes in samenhang van deze factoren met leerprestaties (onderzoeksvraag 2). Het valt overigens niet uit te sluiten dat variabelen waarvoor er geen hoofdeffect gevonden wordt, toch een interactie-effect hebben met sekse. Bijvoorbeeld omdat het effect voor jongens in even sterke mate positief uitwerkt als dat het voor meisjes negatief uitwerkt. Om het aantal te testen interactie-effecten enigszins beperkt te houden is deze mogelijkheid echter niet onderzocht.

Conform de door de internationale TIMSS-coördinatie aangeraden werkwijze, is ten slotte het model vijf keer getoetst met de *plausible values* voor natuuronderwijs (Von Davier, Gonzalez, & Mislevy, 2009). Tabel 3 toont de resultaten van het eindmodel waarbij de gemiddelde effecten en standaard meetfout (waarin de variantie in effectgroottes tussen de vijf *plausibles* zijn verdisconteerd) van alle vijf *plausible values* zijn weergegeven. Omdat het hier een aanzienlijk aantal regressiecoëfficiënten betreft, is een correctie voor kanskapitalisatie (Bonferroni-correctie) uitgevoerd. Het significantieniveau is gezet op $\alpha < 0,05$, maar omdat er twaalf effecten worden getoetst wordt een effect pas als significant aangemerkt voor $\alpha < 0,0042$

(0,05/12). Dit correspondeert met *t*-waarden die in absolute waarde groter zijn dan 2,86.

Gezien het relatieve grote aantal potentiële relevante factoren, zijn in deze studie geen indirecte relaties die in het schooleffectiviteitsmodel wel worden verondersteld, geëxploreerd. De analyses in deze studie richten zich alleen op het vaststellen van directe effecten van deze factoren op toetsprestaties van groep 6 leerlingen in natuuronderwijs. In Tabel 1 wordt een overzicht gegeven van de verschillende analysemodellen, de factoren en de bijbehorende indicatoren in de TIMSS-2007 data. Van de samengestelde variabelen is tevens het aantal items en Cronbachs alpha weergegeven. Alleen samengestelde variabelen met $\alpha \geq 0,60$ zijn in de analyses meegenomen.

5 Resultaten

5.1 Leeg model

Het lege model laat zien in hoeverre verschillen in toetsprestaties toegeschreven kunnen worden aan verschillen tussen scholen, klassen en leerlingen. Ter vergelijking is in Tabel 2 ook het lege model voor de scores op de TIMSS-2007 rekentoets opgenomen.

De percentages op school- en klasniveau laten zien dat voor natuuronderwijs ruim 13% van de variantie in toetsprestaties schoolgebonden is en bijna 9% klasgebonden is. Het verschil met rekenen-wiskunde is klein; voor beide vakken geldt dat de niet-leerlinggebonden variantie ruim 20% is.

5.2 Leerlingkenmerken en contextfactoren

De effectgroottes van de factoren die significant samenhangen met toetsprestaties worden weergegeven in Tabel 3. Het blijkt dat het *aantal boeken thuis*, de belangrijkste voorspeller is van leerprestaties (0,23). Dit effect is aanmerkelijk groter dan het negatieve effect van *allochtoon* (-0,11) en ook groter dan de voorsprong van jongens (0,09). Leerlingen die de eerste zes jaar van het basisonderwijs één of meer klassen hebben overgedaan, hebben minder goed op de toets gepresteerd dan leerlingen die geen vertering hebben opgelopen. Door de introductie

Tabel 1

Factoren en bijbehorende indicatoren in TIMSS-2007, per analysemodel

Factor	Indicator(en) in TIMSS-2007
MODEL 0: OUTPUT	toetsscore natuuronderwijs (vijf <i>plausible values</i>)
MODEL 1: LEERLINGKENMERKEN	
sekse	meisje/jongen
'sociaal culturele status'	aantal boeken thuis
autochtoon/allochtoon	geboorteland ouders
eerder behaalde resultaten	niet/wel vertraagd
MODEL 2: CONTEXT	
schoolgrootte	aantal leerlingen op school
samenstelling leerlingpopulatie	schoolgewicht (gewichtregeling van vóór 2006: hoe hoger het schoolgewicht hoe meer leerlingen er op de school zitten met een 1,25 (laag opgeleide autochtone ouders) of 1,9 (laag opgeleide allochtone ouders)
denominatie	niet/wel openbaar, niet/wel protestant-christelijk, niet/wel rooms-katholiek
MODEL 3: INPUT	
sekse leraar	vrouw/man
onderwijservaring leraar	onderwijservaring, geen/wel specialisatie natuuronderwijs, door leerkracht gepercipieerde toerusting natuuronderwijs (22 items (leerstofonderdelen), $\alpha = 0,95$)
ondersteuning ouders	de door leerling gepercipieerde ondersteuning natuuronderwijs (3 items, $\alpha = 0,68$)
beschikbare infrastructuur	computer-leerlingratio voor groep 6
MODEL 4A: PROCES SCHOOL	
schoolklimaat	door schoolleider gepercipieerd schoolklimaat (oordeel over consensus en verwachtingen van het leerkrachtenteam, betrokkenheid van ouders en gedrag van leerlingen (8 items, $\alpha = 0,68$); door leerkracht gepercipieerd schoolklimaat (zelfde 8 items, $\alpha = 0,76$); door leerkracht ervaren belemmeringen in onderwijs door probleemgedrag leerlingen (5 items, $\alpha = 0,83$); schoolgemiddelde leerlingperceptie onveiligheid op school (pesten, fysiek geweld, uitsluiting, (3 items, $\alpha = 0,62$))
kwaliteit curriculumaanbod	wel/geen aanbod verrijkingsstof natuuronderwijs op school
MODEL 4B: PROCES KLAS	
effectieve leertijd	minuten natuuronderwijs per week, omvang huiswerk natuuronderwijs, verdeling tijd over inhoudsdomen, wel/geen integratie natuuronderwijs in andere vakken
<i>Opportunity to Learn</i>	aantal leerstofonderdelen van de TIMSS-toets dat volgens de leerkracht onderwezen is, klasgemiddelde tijd aan proefjes (4 items, $\alpha = 0,71$), klasgemiddelde tijd aan observeren natuurlijke fenomenen
klimaat in de klas *	klasgemiddelde leerlingperceptie veiligheid klas: positieve en negatieve stellingen over cognitieve veiligheid (bijvoorbeeld: "ik ben tijdens de les bang om uitgelachen te worden") en sociale veiligheid (bijvoorbeeld: "mijn klasgenootjes vinden mij aardig", in totaal 14 items, $\alpha = 0,76$)
differentiatie *	frequentie leerkracht-gestuurde onderwijskenmerken (6 items, $\alpha = 0,78$); frequentie leerling-georiënteerde onderwijskenmerken (12 items, $\alpha = 0,80$)
ICT-gebruik	intensiteit computergebruik bij natuuronderwijs

Noot: * nationale optie: vraag alleen in Nederlandse instrumenten opgenomen.

Tabel 2

Verdeling van de variantie in het lege model voor leerprestaties natuuronderwijs en rekenen-wiskunde, TIMSS-2007

Variantie component	Natuuronderwijs	Rekenen-wiskunde
Schoolniveau	13,3%	9,4%
Klasniveau	8,6%	11,7%
Leerlingniveau	78,1%	78,9%

Noot: Eerste *plausible value* voor natuuronderwijs en rekenen-wiskunde

Tabel 3

Gemiddelde effecten en standard error (se) van leerling-, school- en klasfactoren op natuuronderwijsprestaties in TIMSS-2007, gewogen

	Effect	s.e.
Intercept	-0,02	0,03
<i>Leerlingenkenmerken</i>		
Sekse leerling (jongen)	0,09 *	0,02
Geboorteland ouder(s) (allochtoon)	-0,11 *	0,03
Aantal boeken thuis	0,23 *	0,02
Vertraagd	-0,13 *	0,02
<i>Context</i>		
Schoolgewicht	-0,17 *	0,04
Denominatie (Protestant-christelijk)	-0,16 *	0,03
<i>Input</i>		
Perceptie leerkracht toerusting natuuronderwijs	0,09 *	0,03
Perceptie leerling ondersteuning ouders	-0,08 *	0,02
Computer-leerlingratio	-0,06	0,03
<i>Proces: school</i>		
Door schoolleider gepercipieerd schoolklimaat	0,11 *	0,03
<i>Proces: klas</i>		
Klasgemiddelde leerlingperceptie sociale en cognitieve veiligheid in de klas	0,09 *	0,03
Klasgemiddelde intensiteit proefjes natuuronderwijs	0,06	0,03

Noot: *significant voor $\alpha < 0,05$ (tweezijdig, na Bonferroni-correctie; $t > 2,86$ of $t < -2,86$)

van alleen leerlingenkenmerken in het model wordt 16,6% van de totale variantie verklaard.

Door contextfactoren met een significant effect aan Model 1 toe te voegen, neemt het percentage verklaarde variantie met 6,4% toe. Deze toename is toe te schrijven aan het negatieve effect van *schoolgewicht* (-0,17) en *denominatie* (-0,16). Leerlingen op Protestants-christelijke (PC) scholen blijken minder goed op de natuuronderwijstoets te hebben gepresteerd dan leerlingen op scholen van een andere denominatie (rooms-katholiek, openbaar en algemeen bijzonder).

5.3 Input- en procesfactoren

In het Nederlandse TIMSS-2007 rapport is geconcludeerd dat leerkrachten zich gemiddeld meer dan voldoende toegerust voelen om les te geven in natuuronderwijs. Alleen in het inhoudsdomen natuur- en scheikunde voelen de leerkrachten zich gemiddeld slechts 'voldoende' toegerust (Meelissen & Drent, 2008). Tabel 3 laat zien dat *gepercipieerde toerusting van leerkrachten in natuuronderwijs* in beperkte mate samenhangt met de leerprestaties van leerlingen (0,09).

Het lijkt er verder op dat zwakkere leerlingen meer aanmoediging van hun ouders ervaren dan beter presterende leerlingen. De *door de leerling gepercipieerde aanmoediging door ouders* laat namelijk een licht negatief verband (-0,09) met toetsprestaties in natuuronderwijs zien. Het aantal computers dat op school voor groep 6 leerlingen beschikbaar is, hangt licht positief samen met toetsprestaties. Dit effect is echter niet significant voor $< 0,05$, indien men corrigeert voor kanskapitalisatie. De effectgrootte (-0,06) is negatief omdat het hier gaat om het aantal leerlingen dat gebruik kan maken van één computer. De beschikbaarheid van computers zegt echter nog niets over mogelijke effecten van computergebruik in het onderwijs. Uit de analyses blijkt dat de *intensiteit van computergebruik* tijdens de les geen directe samenhang laat zien met natuuronderwijsprestaties. De effecten van de inputvariabelen zijn klein; het percentage verklaarde variantie van model 2 (leerlingenkenmerken en context) neemt in model 3 door de toevoeging van deze inputfactoren met slechts 1% toe.

De TIMSS-indicator voor het *schoolklimaat* blijkt een direct (beperkt) positief effect

op toetsprestaties te hebben (0,11). Zoals in Tabel 1 is aangegeven verwijst het *schoolklimaat* in TIMSS naar het oordeel van de schoolleider over consensus en verwachtingen van het leerkrachtenteam, over de betrokkenheid van ouders en over het gedrag van leerlingen. Aan de leerkrachten zijn dezelfde stellingen over het *schoolklimaat* voorgelegd als aan de schoolleider. Leerkrachten zijn over het algemeen iets minder positief over het *schoolklimaat* dan schoolleiders (Meelissen & Drent, 2008). Het oordeel van de leerkracht hangt echter niet samen met de toetsprestaties van de leerlingen.

Van de zes procesvariabelen op klasniveau is er alleen sprake van een significante samenhang tussen toetsprestaties en *klimaat in de klas* (0,09). Het blijkt dat een leerling uit een klas waar de leerlingen zich gemiddeld genomen prettig en veilig voelen een hogere score haalt op de natuuronderwijstoets in vergelijking tot een leerling uit een klas met een gemiddeld minder positief klasklimaat. De samenhang tussen toetsprestaties en de mate waarin de klas in aanraking komt met natuurkundige proefjes en experimenten (0,06) is niet langer significant als gecorrigeerd wordt voor kanskapitalisatie.

De door Scheerens (2008) genoemde effectiviteitsverhogende procesfactoren 'effectieve leertijd' (zoals *aantal minuten natuuronderwijs per week*) en *Opportunity to Learn* (OTL; aantal leerstofonderdelen van de TIMSS-toets dat volgens de leerkrachten onderwezen is), laten geen relatie met toetsprestaties zien. Ook de verschillende TIMSS-indicatoren die verwijzen naar een mate van *differentiatie* blijken eveneens niet samen te hangen met het prestatieniveau van de leerlingen. Het gaat dan ondermeer om de mate waarin leerkracht-gestuurde of leerling-georiënteerde kenmerken voorkomen in de dagelijkse onderwijspraktijk (Meelissen & Drent, 2008).

De leerlingkenmerken en de context-, input- en procesfactoren met een significant effect, verklaren samen 26,2 % van de totale variantie in leerlingcores op de natuuronderwijstoets. Als laatste stap zijn vervolgens de differentiële effecten voor sekse in het eindmodel geanalyseerd. Er blijken echter geen significante interactie-effecten met sekse te

zijn voor de variabelen die een significant hoofdeffect hebben laten zien. Dit betekent ondermeer dat de samenhang tussen toetsprestaties en de *toerusting van de leerkracht* en het *klimaat in de klas*, niet verschilt tussen jongens en meisjes.

6 Conclusie en discussie

Het voornaamste doel van deze studie is om inzicht te krijgen in mate waarin school-, klas- en instructiefactoren samenhangen met de Nederlandse toetsprestaties in natuuronderwijs in TIMSS-2007 (onderzoeksvraag 1). De directe aanleiding vormde de voor Nederland tegenvallende resultaten van TIMSS-2007 en de sekseverschillen in natuuronderwijsprestaties.

Voor de identificatie van de indicatoren in de TIMSS-data alsmede voor het bepalen van de volgorde van de meerniveau-analysemodellen, is gebruik gemaakt van het school-effectiviteitsmodel van Scheerens (2008) en de resultaten van meta-analyses naar school-effectiviteit (Scheerens et al., 2007). Het schooleffectiviteitsmodel maakt een onderscheid in leerlingkenmerken, context-, input- en procesfactoren.

Ondanks de verschillen tussen de Nederlandse scholen in de wijze waarop natuuronderwijs wordt aangeboden, blijkt dat de variantie in toetsprestaties op leerlingniveau veel groter is dan de tussenschoolse en tussen-groepse variantie. De verschillen in de toetscores tussen leerlingen in natuuronderwijs kunnen voor ongeveer een vijfde toegeschreven worden aan verschillen tussen scholen en klassen. Individuele leerlingkenmerken zijn het meest bepalend voor het prestatieniveau. Dit geldt zowel voor natuuronderwijs als voor rekenen.

Van de contextfactoren op schoolniveau is de gevonden relatie tussen *denominatie* en toetsprestaties het meest opmerkelijk. Leerlingen op PC-scholen hebben minder goed op de natuuronderwijstoets gepresteerd dan leerlingen op niet PC-scholen. Over een mogelijk verband tussen denominatie van de school op leerprestaties is tot nu toe weinig bekend. In het Onderwijsverslag van 2009 wordt geconcludeerd dat zowel PC- als

Rooms-katholieke (RK) scholen relatief minder vaak als 'zwakke scholen' worden aangemerkt in vergelijking tot basisscholen met een andere denominatie (Inspectie van het Onderwijs, 2009a). Een ander onderzoek van de Inspectie van het Onderwijs laat echter zien dat PC-scholen en openbare scholen minder vaak bovengemiddelde opbrengsten hebben in vergelijking tot RK- en bijzondere scholen (Inspectie van het Onderwijs, 2009b). Mogelijke verklaringen voor deze verschillen tussen de denominaties worden er echter niet gegeven.

De mate waarin leerkrachten zich toegerust voelen om les te geven in de verschillende leerstofonderdelen hangt positief samen met de prestaties van leerlingen op de TIMSS-natuuronderwijstoets. Ten opzichte van andere landen is het gemiddelde percentage Nederlandse leerlingen dat les van een leerkracht krijgt die zegt zeer goed toegerust te zijn in alle getoetste natuuronderwijs-onderwerpen, zeer laag (27% ten opzichte van 54% internationaal gemiddeld), terwijl de geschiktheid van de toets voor Nederland relatief hoog is (Martin et al., 2008). Mogelijk is in deze toerusting van de leerkrachten voor natuuronderwijs voor Nederland winst te behalen.

Van de procesfactoren die in deze studie zijn meegenomen, is er alleen een duidelijk positief verband met prestaties gevonden voor het *schoolklimaat* gepercipieerd door de schoolleider en het *klasklimaat* gepercipieerd door de leerlingen. Andere effectiviteitsverhogende factoren zoals *instructietijd*, blijken in deze studie geen direct effect te hebben op prestaties. In de vijf-landenstudie van Kaya en Rice (2009) op basis van de data van TIMSS-2003, wordt eveneens in geen van de landen (VS, Singapore, Japan, Australië en Scotland) een direct effect aangetroffen van het aantal minuten natuuronderwijs per week op toetsprestaties in groep 6. Deze variabele zegt echter niets over wijze waarop deze instructie wordt aangeboden en de kwaliteit ervan. Hierover wordt in TIMSS maar zeer beperkt informatie verzameld. De in deze studie gevonden samenhang tussen de *toerusting van de leerkracht in natuuronderwijs* en leerlingprestaties, zou erop kunnen wijzen dat de kwaliteit van de instructie in natuur-

onderwijs wel van belang is. Overigens zou het ontbreken van een directe samenhang tussen *instructietijd* en prestaties in de Nederlandse situatie zou ook te maken kunnen hebben met de beperkte tijd die überhaupt aan natuuronderwijs besteed wordt. Tussen scholen is hierin er weliswaar variantie, maar de meeste beter presterende landen besteden aanmerkelijk meer tijd aan natuuronderwijs dan Nederland. Mogelijk kan een verhoging naar buitenlands niveau wel een positieve invloed hebben op de prestaties van de Nederlandse leerlingen. Bovendien zou de hoeveelheid tijd die aan natuurwetenschappelijke vakken wordt besteed een indicatie kunnen zijn voor de aandacht die er in deze landen voor deze vakken in het primair onderwijs is.

Naar de samenhang van *Opportunity to Learn* met toetsprestaties is vooral voor wiskunde in leerjaar 2 (grade 8) in de verschillende TIMSS-landen onderzoek verricht (e.g. Dumay & Dupriez, 2007; Pong & Pallas, 2001). Deze studies laten wel een positieve relatie zien. Mogelijk leren leerlingen, meer dan voor rekenen, ook buiten school over onderwerpen die te maken hebben met biologie, natuur-, schei- en aardrijkskunde. De vergelijking tussen het Nederlandse beoogde en gerealiseerde curriculum wijst eveneens in deze richting (Meelissen & Drent, 2008). Van de 180 TIMSS-opgaven voor natuuronderwijs pasten 51 opgaven inhoudelijk niet bij het Nederlandse onderwijs. Van deze niet-passende (meerkeuze- en open) opgaven wordt 41% toch door meer dan de helft van de leerlingen goed beantwoord (Meelissen & Drent, 2008). Dit wijst erop dat zij een deel van hun kennis ergens anders dan tijdens de lessen hebben opgedaan.

De tweede onderzoeksvraag is beantwoord aan de hand van het berekenen van interactie-effecten voor de sekse van de leerling in het eindmodel voor de variabelen met een significant hoofdeffect. Er zijn in deze studie geen differentiële effecten hiervoor aangetroffen. Dit betekent dat voor deze onderzochte indicatoren, de samenhang met natuuronderwijsprestaties niet verschilt tussen jongens en meisjes.

7 Beperkingen

TIMSS biedt landen een rijke databron met informatie over de onderwijscontext op school-, klas-, en leerlingniveau en de mogelijkheid om factoren te identificeren die mogelijk van belang zijn voor effectief onderwijs binnen een land of onderwijssysteem. De data kent echter ook een aantal beperkingen.

Voor het bepalen van het netto prestatieniveau zijn de leerlingkenmerken achtergrond en begaafdheid van belang. In het TIMSS-instrumentarium ontbreekt een goede indicator voor het meten van de begaafdheid. In TIMSS-1999 heeft Vlaanderen in het voortgezet onderwijs bij 14-jarigen leerlingen naast de toets als nationale optie een intelligentietest afgenomen (Van den Broeck, Opendakker, & Van Damme, 2005). Deze variabele verklaarde veel variantie: 22% op leerlingniveau en zelfs 60% op klasniveau. De afname van een extra intelligentietest betekent echter ook een extra belasting voor de leerkrachten en leerlingen die aan TIMSS mee doen. Dit zal ten koste gaan van de deelnamebereidheid, waardoor de kans op uitsluiting van Nederland in de internationale vergelijking groter wordt. Nederland is namelijk een van de weinige TIMSS-landen waar het voldoen aan internationale responsen een zeer moeizaam proces is (Martin, Mullis, Gonzalez, & Chrostowski, 2004).

Naast het netto prestatieniveau wordt in schooleffectiviteitsonderzoek ook wel de 'leerwinst' als effectmaat gehanteerd. Leerwinst is de vooruitgang in leerlingprestaties tussen meerdere meetmomenten, zoals tussen meerdere leerjaren. Met de TIMSS-data kan ook deze leerwinst niet worden vastgesteld, omdat in elke ronde een nieuwe steekproef getrokken wordt en vervolgens op de geselecteerde scholen in één leerjaar (groep 6, leerjaar 2 of eindexamenjaar) eenmalig een toets wordt afgenomen.

In deze studie is het schooleffectiviteitsmodel van Scheerens (2008) als raamwerk gebruikt om factoren te identificeren voor opname in de analysemodellen. In dit generieke model zijn factoren opgenomen waarvan op basis van metastudies is aangetoond dat zij in bepaalde onderwijssystemen of lan-

den van belang kunnen zijn voor de opbrengsten van onderwijs. Zoals ook uit de vergelijking tussen deze studie en andere studies duidelijk is geworden, kan de mate waarin input- en procesfactoren van belang zijn voor leerprestaties, per onderwijssysteem, per vakgebied en per leerlingpopulatie verschillen. In deze exploratie is voor Nederland uiteindelijk maar een beperkt aantal input- en procesfactoren gevonden dat samenhangt met toetsprestaties in natuuronderwijs. Voor een deel kan dit het gevolg zijn van het feit dat in deze exploratie alleen naar directe effecten is gekeken. Daarnaast kan het echter ook te maken hebben met de beschikbaarheid van relevante indicatoren in de TIMSS-data. Over een aantal van de empirisch onderbouwde 'effectiviteitverhogende' procesfactoren op school- en klasniveau was in het instrumentarium van TIMSS-2007 geen indicator aanwezig. In het lopende onderzoek, TIMSS-2011, zijn de instrumenten echter verder aangescherpt op basis van empirische onderbouwing, hetgeen ondermeer geresulteerd heeft in de toevoeging van effectiviteitsbevorderende factoren op schoolniveau, namelijk de mate waarin de school de betrokkenheid van ouders stimuleert en onderwijskundig leiderschap (Mullis, Martin, Rudock, O'Sullivan, & Preuschoff, 2009).

Tenslotte worden in het schooleffectiviteitsmodel indirecte en causale verbanden tussen de verschillende groepen factoren verondersteld. Causaliteit kan met deze explorerende, relationele analyses waarbij prestaties op één moment worden gemeten, niet worden vastgesteld. Ditzelfde geldt voor de effecten van mediators op veronderstelde relaties of wederkerige relaties. De resultaten van deze studie kunnen echter wel een aanknopingspunt zijn voor vervolgstudies naar bijvoorbeeld de effecten van naschoolse trainingsprogramma's voor leerkrachten of meer aandacht voor natuuronderwijs in de initiële lerarenopleiding. Daarnaast kan meer de diepte worden ingegaan door voor een beperkt aantal factoren indirecte of mediërende effecten (pad-analyse) te schatten waarbij tevens rekening wordt gehouden met hiërarchische data aan de hand van meer geavanceerde programmatuur (zoals Mplus, Muthén & Muthén, 2010). Een voorbeeld hiervan zou

kunnen zijn om ook te kijken naar de indirecte relatie tussen toerusting van de leerkracht en toetsprestaties, die bijvoorbeeld via het klasklimaat zou kunnen lopen. Overigens kan ook met het meest geavanceerde en uitgewerkte statistische model geen zekerheid worden verkregen over causaliteit, omdat er altijd ongemeten variabelen zullen zijn die echte causale verbanden vertekenen (Gustafsson, 2008). In de reguliere TIMSS-studies, waarbij alleen schriftelijke data verzameld wordt en geen lessen worden geobserveerd, zullen deze ongemeten variabelen (naast begaafdheid) vooral liggen op het vlak van de kenmerken en de kwaliteit van instructie en de interactie tussen leerkracht en leerling.

Literatuur

- Bos K. Tj., & Meelissen, M. R. M. (2006). Exploring influencing factors on mathematics achievement in grade 8 in the Netherlands, Belgium Flanders and Germany: Results of secondary analysis on TIMSS data. In S. Howie & Tj. Plomp (Eds.), *Contexts of learning: Lessons learned from TIMSS* (pp. 195-210). New York: Routledge.
- Bosker, R. J. (1991). De consistentie van school-effecten in het basisonderwijs. *Tijdschrift voor Onderwijsresearch*, 16 (4), 206-218.
- Broeck, A., van den., Opdenakker, M. C., Damme, J. van. (2005). The effects of student characteristics on mathematics achievement in Flemish TIMSS 1999 data [Elektronische versie]. *Educational Research and Evaluation*, 11 (2), 107-121.
- Creemers, B. P. M. (1994). *The effective classroom*. London: Cassell.
- Dumay, X., & Dupriez, V. (2007). Accounting for class effect using the timss 2003 eighth-grade database: Net effect of group composition, net effect of class process, and joint effect [Elektronische versie]. *School Effectiveness and School Improvement*, 18 (4), 383-408.
- Gustafsson, J. E. (2008). Effects of international comparative studies on educational quality of educational research [Elektronische versie]. *European Educational Research Journal*, 7 (1), 1-17.
- Inspectie van het Onderwijs. (2009a). *De staat van het onderwijs. Onderwijsverslag 2007/2008*. Utrecht, Nederland: Inspectie van het Onderwijs.
- Inspectie van het Onderwijs. (2009b). *De sterke basisschool: Definitie en kenmerken*. Utrecht, Nederland: Inspectie van het Onderwijs.
- Kaya, S. (2008). *The effects of student-level and classroom-level factors on elementary students' science achievement in five countries*. Dissertation. Florida State University, Florida
- Kaya, S., & Rice, D. C. (2009). Multilevel effects of student and classroom factors on elementary science achievement in five countries [Elektronische versie]. *International Journal of Science Education*. 32(10), 1337-1363.
- Knuver, J. W. M., & Doolaard, S. (1997). *Rekenen-wiskunde en natuuronderwijs op de basisschool. Nederlandse aandeel in TIMSS populatie 1*. Enschede, Nederland: OCTO, Universiteit Twente.
- Lubienski, S. T., & Crane, C. C. (2010). Beyond free lunch: which family background measures matter? [Elektronische versie]. *Education policy analysis archives*, 18 (11), 1-43.
- Ma, X., & Kishor, N. (1997). Attitude toward self, social factors, and achievement in mathematics: A meta-analytic review [Elektronische versie]. *Educational Psychology Review*, 9, 89-120.
- Martin, M. O., Mullis, I. V. S., Gonzalez, E. J., & Chrostowski, S. J. (2004). *TIMSS 2003 International Science Report. Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Boston, MA: Boston College, TIMSS & PIRLS International Study Center.
- Martin, M. O., Mullis, I. V. S., & Foy, P. (2008). *TIMSS 2007 International Science Report. Findings from IEA's Trends in International Mathematics and Science Study at the Fourth and Eighth Grades*. Boston, MA: Boston College, TIMSS & PIRLS International Study Center.
- Maeyer, S. A. J. H., & Rymenans, R. M. C. (2004). *Onderzoek naar kenmerken van effectieve scholen*. Dissertatie. Universiteit Utrecht, Utrecht, Nederland.
- Meelissen, M. R. M., & Doornekamp, B. G. (2004). *TIMSS-2003 Nederland: leerprestaties in exacte vakken in het basisonderwijs*. Enschede, Nederland: Universiteit Twente.
- Meelissen, M. R. M., & Drent, M. (2008). *TIMSS-2007 Nederland: trends in leerprestaties in*

exacte vakken in het basisonderwijs. Enschede, Nederland: Universiteit Twente.

Meelissen, M. R. M., & Luyten, H. (2008). The Dutch gender gap in mathematics: small for achievement, substantial for beliefs and attitudes [Elektronische versie]. *Studies in Educational Evaluation*, 34 (2), 82-93.

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, & Preuschoff, C. (2009). *TIMSS 2011 Assessment Frameworks*. Boston, MA: Boston College, TIMSS & PIRLS International Study Center.

Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C., Arora A., & Erberber, E. (2005). *TIMSS 2007 Assessment Frameworks*. Boston, MA: Boston College, TIMSS & PIRLS International Study Center.

Muthén, L. K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthen.

OECD (2005). *PISA 2003 data analysis manual, SPSS users*. Paris: OECD.

Olson, J. F., Martin, M. O., & Mullis, I. V. S. (2008). *TIMSS 2007 Technical report*. Boston, MA: Boston College, TIMSS & PIRLS International Study Center.

Platform Bèta Techniek (2009). *Bètatechniek Agenda 2011-2016. Blijven investeren in de toekomst van Nederland*. Den Haag, Nederland: Platform Bèta Techniek.

Pong, S., & Pallas, A. (2001). Class size and eighth-grade math achievement in the United States and abroad [Elektronische versie]. *Educational Evaluation and Policy Analysis*, 23(3), 251-273.

Scheerens, J. (2008). *Een overzichtsstudie naar school- en instructie-effectiviteit*. Enschede, Nederland: Universiteit Twente.

Scheerens, J. (1999). School effectiveness concepts and theories. In A.J. Visscher (Ed.), *Managing schools towards high performance. Linking school management theory to the school effectiveness knowledge base*. (pp. 37-70). Lisse, Nederland: Swets & Zeitlinger.

Scheerens, J., Luyten, H., Steen, R., & Luyten-de Thouars, Y. (2007). *Review and meta-analyses of school and teaching effectiveness*. Enschede, Nederland: Universiteit Twente.

Veenstra, R. (1999). *Leerlingen-klassen-scholen*. Dissertatie. Amsterdam: Thela Thesis.

Von Davier, M., Gonzalez, E. & Mislevy, R. J (2009). What are plausible values and why

are they useful? In M. Von Davier & D. Hastedt (Eds.), *IERI. Monograph series. Issues and methodologies in large-scale assessments* (pp. 9-36). Hamburg, Duitsland: IER Institute.

Wang, J., Oliver, J. S., & Staver, J. R. (2008). Self-concept and science achievement: investigation a reciprocal relation model across the gender classification in a crosscultural context [Elektronische versie]. *Journal of Research in Science Teaching*, 45 (6), 711-725.

Wijnen, W. (2005). *Visie op wetenschap en techniek in het basisonderwijs*. Amsterdam: Expertgroep Wetenschap en Techniek Basisonderwijs.

Manuscript aanvaard: 18 mei 2011

Auteurs

M.R.M. Meelissen is universitair docent/projectleider TIMSS-Nederland sinds TIMSS-2003, Vakgroep Onderwijsorganisatie en -management, Faculteit Gedragswetenschappen (Onderwijskunde). **H. Luyten** is als universitair hoofd-docent verbonden aan deze vakgroep.

Correspondentieadres: M.R.M. Meelissen, Universiteit Twente, Postbus 217, 7500 AE Enschede, m.r.m.meelissen@utwente.nl.

Abstract

School effectiveness and science achievement in grade 4: secondary analyses on TIMSS-2007 data

In TIMSS-2007, Dutch fourth grade students performed considerable lower in science compared to their performance in mathematics. Furthermore, the Netherlands was one of the few countries where girls were outperformed by boys in science. Guided by the model of School Effectiveness (Scheerens, 2008), this study examined the relation between malleable school and class factors and science achievement, controlled for student characteristics within different school contexts. Multi-level analyses on TIMSS data of 3300 fourth grade students show that, controlled for students characteristics (gender, minority, number of books at home and repeating class), two

contextual school characteristics are strongly associated with science achievement; the average social economic status of the school population and the religious ideology of the school. The results further suggest that teachers' ability in teaching science and a positive climate in both school and class are relevant for raising achievement.