

Secundaire analyses op de data van TIMSS-2015: een nadere analyse van leerkrachtpercepties en de referentieniveaus voor rekenen en natuuronderwijs

Anja Rebber, Stéphanie van den Berg en Martina Meelissen

Vakgroep Onderzoeksmethodologie, Meetmethoden en Data-analyse (OMD)

Faculteit BMS, Universiteit Twente

Enschede, mei 2017

UNIVERSITEIT TWENTE.

1. Context en onderzoeksvragen

In het voorjaar van 2015 hebben ruim 4500 Nederlandse groep 6 leerlingen een toets gemaakt in rekenen-wiskunde en natuuronderwijs (*science*) in het kader van het internationale project *Trends in International Mathematics and Science Study* (TIMSS). Ten opzichte van eerdere TIMSS-metingen hebben de Nederlandse leerlingen de TIMSS-toets van 2015 minder goed gemaakt (Meelissen & Punter, 2016). Dit geldt zowel voor rekenen als voor natuuronderwijs. Omdat de leerlingprestaties in de meeste andere landen zijn toegenomen of gelijk gebleven, is de internationale positie van Nederland in deze vakgebieden in het basisonderwijs aanmerkelijk verslechterd.

TIMSS fungeert als een monitor voor het landelijk onderwijsniveau in de exacte vakken in groep 6. Het meet elke vier jaar de prestaties van groep 6 leerlingen en verzamelt informatie over de onderwijscontext aan de hand van leerling-, leerkracht-, school- en curriculumvragenlijsten. De informatie over de onderwijscontext in Nederland en in andere landen kan gebruikt worden om op zoek te gaan naar mogelijke verbanden tussen (ontwikkelingen in) deze onderwijsindicatoren en (ontwikkelingen in) leerprestaties. Met een cross-sectioneel onderzoek zoals TIMSS kunnen echter geen verklaringen gegeven worden voor verschillen tussen groepen leerlingen, landen of jaren. Dit geldt ook voor het vinden van mogelijke verklaringen voor de Nederlandse daling in het prestatieniveau in 2015.

De TIMSS-contextdata kunnen weliswaar geen verklaringen bieden voor de tegenvallende leerlingprestaties, maar zij kunnen wel gebruikt worden om secundaire analyses te doen naar onderwijskenmerken die mogelijk samenhangen met leerlingprestaties. Mogelijk verbanden tussen manipuleerbare onderwijskenmerken en toetsprestaties kunnen informatie geven over waar mogelijk voor het Nederlandse onderwijs winst te behalen valt. De gevonden relaties kunnen aanleiding zijn voor vervolgonderzoek waarbij causaliteit wel kan worden vastgesteld.

In opdracht van het Ministerie van Onderwijs, Cultuur en Wetenschap heeft de vakgroep OMD van de Universiteit Twente, die verantwoordelijk was voor het Nederlandse aandeel in TIMSS-2015, secundaire analyses uitgevoerd gebaseerd op drie onderzoeksvragen. De eerste onderzoeksvraag luidt:

1. *In welke mate hangen de percepties van leerkrachten over hun didactische vaardigheden en het leerklimaat op hun school, in Nederland en een aantal relevante vergelijkingslanden, samen met leerlingprestaties in de exacte vakken?*

Als voor Nederland kenmerken van de onderwijscontext vergeleken worden met de andere (omringende) TIMSS-landen dan vallen er een paar zaken op. Hoewel Nederlandse leerkrachten hun beroep en hun school positief beoordelen, ervaren Nederlandse leerkrachten in 2015 een relatieve hoge werkdruk. Ook schatten zij de prestatiegerichtheid van het leerklimaat op hun school vergeleken met veel andere landen laag in en volgen ze weinig bijscholing, met name op het gebied van natuuronderwijs. Leerkrachten voelen zich verder minder toegerust om les te geven in natuuronderwijs dan in rekenen. Nagegaan wordt in hoeverre leerkrachtpercepties van hun eigen vaardigheden en van het schoolklimaat samenhangen met de prestaties van de leerlingen op de TIMSS reken- en natuuronderwijstoets, gegeven de invloed van achtergrondkenmerken op leerling- en schoolniveau. Deze analyses worden niet alleen uitgevoerd op de Nederlandse TIMSS-data, maar ook op de data van Finland, Engeland, Vlaanderen en Zweden.

De tweede onderzoeksvraag die in dit onderzoek wordt beantwoord is:

- 2. In welke mate hangt het uitgevoerde curriculum (percentage onderwezen onderwerpen) in Nederland samen met leerlingprestaties, gegeven relevante achtergrondkenmerken van leerlingen en scholen?*

In de leerkrachtvragenlijst van TIMSS wordt aan leerkrachten per inhoudsdomen een lijst met leerstofgebieden voorgelegd met de vraag in hoeverre dit gebied al behandeld is. In deze studie wordt de mogelijke samenhang tussen dit uitgevoerde curriculum en het gerealiseerde curriculum onderzocht, gegeven de invloed van achtergrondkenmerken op leerling- en schoolniveau. Deze analyses worden alleen voor Nederland uitgevoerd.

De derde onderzoeksvraag luidt als volgt:

- 3. In hoeverre hebben er zich in de periode 2011 en 2015 verschuivingen voorgedaan in de prestaties van (potentieel) excellente leerlingen in rekenen en natuuronderwijs op de verschillende inhouds- en cognitieve domeinen?*

Sinds TIMSS-1995 is het aandeel leerlingen dat het geavanceerde en het hoge niveau haalt in beide vakgebieden geleidelijk afgenomen. Omdat deze afname al in 2011 zichtbaar was, is op basis van de data van PIRLS-2011 en TIMSS-2011 in 2014 een onderzoek verricht naar verschillen in achtergrond- en onderwijskenmerken tussen excellente leerlingen en niet-excellente leerlingen in lezen, rekenen en natuuronderwijs (Meelissen, Netten, Punter, Vermeeren, Drent & Strating, 2014). Leerlingen werden in deze studie als 'excellent' beschouwd als zij tot de 20% best presterende leerlingen in lezen, rekenen of natuuronderwijs behoren op basis van de PIRLS- of TIMSS-toets van 2011. Uit deze studie bleek dat de 20% best presterende Nederlandse 10-jarigen zich vooral van de overige leerlingen in lezen, rekenen en natuuronderwijs onderscheiden in achtergrondkenmerken (geslacht, thuistaal en de sociaaleconomische of culturele status van het gezin) en affectieve leerlingkenmerken (attitude, zelfvertrouwen). De leeromgeving op school van 'excellente' leerlingen in lezen, rekenen of natuuronderwijs verschilde nauwelijks van de leeromgeving van de overige leerlingen. Alleen de leerlingpopulatie van de school leek van belang; 'excellente' leerlingen zitten relatief vaker op een school waarvan de schoolpopulatie gekenmerkt wordt door een laag percentage leerlingen uit economische achterstandsgezinnen.

In dit onderzoek worden de prestaties van de leerlingen die op de verschillende referentieniveaus hebben gepresteerd, nader geëxploreerd. Per inhouds- en cognitief gebied en per toetsjaar wordt de verdeling berekend van leerlingen op het geavanceerde niveau, op het hoge niveau en de overige leerlingen.

In de volgende drie hoofdstukken worden de analysemethodes en de resultaten per onderzoeksvraag beschreven.

2. Leerkrachtpercepties in relatie tot leerprestaties in de exacte vakken

2.1 Variabelen

De analysemodellen zijn opgebouwd met drie soorten factoren: inputfactoren (leerling- en schoolkenmerken), procesfactoren (leerkrachtpercepties) en outputfactoren (toetsprestaties).

Outputfactoren

De outputfactoren zijn in dit geval de leerlingsscores, in de vorm van vijf zogenoemde *plausible values* per leerling op de rekentoets en op de natuuronderwijstoets. De TIMSS-toets bestaat in totaal uit 14 toetsboekjes (versies van de toets). Aan de hand van een toetsrotatiesysteem wordt een leerling één van de 14 toetsboekjes willekeurig toegewezen. De geobserveerde score van een leerling (de scores op de voorgelegde toetsitems) wordt gezien als een indicatie van het bereik waarbinnen de werkelijke vaardigheid van de leerling is gelegen. Voor het bepalen van deze vaardigheid is door de internationale coördinatie van TIMSS aan de hand van IRT-analyses een verdeling van vaardigheidsscores geconstrueerd, gebaseerd op de geobserveerde scores en achtergrondkenmerken van de leerling. *Plausible values* zijn op toevalsbasis getrokken waarden uit deze verdeling. De analyses zijn per vakgebied voor elk van de vijf *plausible values* uitgevoerd en nadien statistisch samengevoegd.

Inputfactoren

Inputfactoren zijn school- en leerlingkenmerken die nauwelijks of moeilijk manipuleerbaar zijn, maar wel een effect kunnen hebben op leerlingprestaties. Op leerlingniveau zijn de achtergrondvariabelen sekse, taal thuis en het aantal boeken thuis (indicator voor sociaal culturele status) meegenomen. Op schoolniveau gaat het om het percentage economische achterstandsl leerlingen, percentage leerlingen op school met een andere thuistaal dan de toetstaal, en urbanisatiegraad.

Procesfactoren

De data over leerkrachtpercepties is verzameld met de leerkrachtvragenlijst. In de analyses zijn de volgende procesfactoren worden meegenomen (voor de beschrijving van deze variabelen zie Meelissen & Punter, 2016, pp. 62-73)

- 1) Perceptie leerkracht werkdruk
- 2) Zelfvertrouwen in didactische vaardigheden in rekenen en natuuronderwijs
- 3) Waargenomen mate van toerusting voor rekenen en natuuronderwijs, per inhoudsdomein
- 4) Frequentie bijscholing in rekenen en natuuronderwijs
- 5) Perceptie prestatiegerichtheid van het schoolklimaat

In TIMSS is het *Partial Credit model* toegepast om samengestelde contextvariabelen te construeren op basis van de antwoordpatronen van een set items die een overkoepelend construct representeren. Voor de factoren toerusting en zelfvertrouwen waren deze variabelen niet beschikbaar in de internationale dataset. Voor deze factoren zijn daarom nieuwe samengestelde variabelen geconstrueerd en is hun betrouwbaarheid in de verschillende landen geanalyseerd aan de hand van Cronbach's alpha. Deze was in alle gevallen voldoende ($\alpha > 0,80$).

2.2 Analyseprocedure

De analyses zijn uitgevoerd aan de hand van de multilevel techniek. Deze techniek houdt rekening met de drie niveaus (leerling, leerkracht en school) waarop de data verzameld is, maar ook met de geclusterde steekproeftrekking van TIMSS (Meelissen & Punter, 2016). In Nederland doen de scholen

mee met alle aanwezige groepen 6 (in de andere landen zijn twee klassen getrokken). Om deze reden wordt er in dit onderzoek ook een klasniveau onderscheiden. Eerst is per land het zogenoemde lege model (zonder input- en procesvariabelen) geschat, op basis van de eerste *plausible value*. Hiermee wordt duidelijk hoeveel van de variantie in toetsprestaties toe te wijzen is aan verschillen tussen leerlingen, scholen en klassen.

De relaties tussen input- en procesfactoren op toetsprestaties zijn geanalyseerd aan de hand van twee methoden, om te voorkomen dat ten onrechte variabelen worden uitgesloten vanwege een niet-significant effect. In de eerste methode (*forward stepwise*) zijn eerst de effecten van de inputvariabelen op leerlingniveau in het model opgenomen. Alle niet-significante effecten zijn vervolgens verwijderd en de analyses zijn opnieuw uitgevoerd. Dit proces is herhaald totdat er uitsluitend significante effecten ($\alpha = 0.05$; tweezijdig) resteren. Vervolgens zijn de inputvariabelen op schoolniveau en daarna de procesvariabelen op leerkrachtniveau toegevoegd, telkens volgens bovenstaande procedure, totdat ook hier een model over blijft met uitsluitend significante effecten. In de tweede methode (*backwards stepwise*) zijn alle input- en procesfactoren van zowel het leerling-, leerkracht- als ook het schoolniveau gelijktijdig meegenomen. Vervolgens zijn alle niet-significante variabelen verwijderd ($\alpha = 0.05$; tweezijdig) en is de analyse herhaald. Dit proces is herhaald totdat het model uitsluitend uit significante effecten bestaat. Het definitieve model is op basis van alle vijf *plausible values* vastgesteld. Daarbij wordt een correctie toegepast op de standaardfouten zoals beschreven door Von Davier, Gonzalez, & Mislevy (2009).

In dit rapport worden de eindmodellen gepresenteerd gebaseerd op de *forward stepwise* methode. De *backwards stepwise methode* leverde dezelfde modellen op of modellen met minder variabelen met een significant effect. Omdat het 'percentage leerlingen met andere thuistaal dan Nederlands op school' zeer sterk bleek te correleren met de thuistaal op leerlingniveau is het 'percentage leerlingen met andere thuistaal dan Nederlands op school' in de analyses buiten beschouwing gelaten.

2.3 Resultaten rekenen

Tabel 2.1 geeft een overzicht van de kenmerken van de input- en procesvariabelen die in deze multilevel-analyses zijn meegenomen. In deze paragraaf wordt verder het zogenoemde lege model en de eindmodellen (met alleen significante effecten) voor de verschillende landen voor rekenen weergegeven.

Tabel 2.1

Kenmerken input- en procesfactoren in TIMSS-2015 in Nederland, Vlaanderen, Engeland, Finland en Zweden, gewogen

	Nederland	Vlaanderen	Engeland	Finland	Zweden
Input leerling	%	%	%	%	%
Sekse (meisje)	50	50	51	48	50
Taal thuis (toetstaal)	80	79	83	89	84
Aantal boeken thuis weinig (< 25)	39	34	33	21	30
Aantal boeken thuis gemiddeld (26-100)	39	38	35	42	35
Aantal boeken thuis veel (> 100)	22	27	32	37	35
Input school					
SES (0-10% econ. achterstand)	73	39	28	54	58
SES (11-25% econ. achterstand)	20	38	32	28	24
SES (26-50% econ. achterstand)	7	16	19	14	8
SES (> 50% econ. achterstand)	1	8	21	4	10
	<i>gem. (sd.)</i>	<i>gem. (sd.)</i>	<i>gem. (sd.)</i>	<i>gem. (sd.)</i>	<i>gem. (sd.)</i>
Urbanisatiegraad [1= stedelijk tot 5= platteland]	3,3 (1,2)	3,4 (1,3)	2,5 (1,2)	3,3 (1,2)	3,1 (1,3)
Proces perceptie leerkracht					
Prestatiegerichtheid schoolklimaat*	9,0 (1,2)	9,2 (1,5)	10,7 (2,2)	9,8 (1,4)	9,4 (1,7)
Werkdruk leerkracht**	9,5 (1,5)	10,0 (1,4)	9,2 (2,2)	11,0 (1,3)	9,9 (1,7)
Gem. zelfvertrouwen didactische vaardigh. [1= zeer hoog tot 4= laag]					
- Rekenen	2,1 (0,5)	2,2 (0,4)	1,8 (0,5)	2,0 (0,5)	2,1 (0,5)
- Science	2,6 (0,5)	2,3 (0,5)	2,1 (0,6)	2,0 (0,6)	2,2 (0,5)
Gem. mate van toerusting [1= zeer goed tot 3= onvoldoende]					
- Getallen	1,1 (0,2)	1,0 (0,1)	1,1 (0,2)	1,2 (0,3)	1,2 (0,3)
- Geometrische vormen en meten	1,2 (0,3)	1,1 (0,3)	1,1 (0,3)	1,2 (0,3)	1,2 (0,3)
- Gegevensweergave	1,1 (0,2)	1,1 (0,3)	1,1 (0,3)	1,2 (0,4)	1,2 (0,4)
- Biologie	1,4 (0,5)	1,3 (0,3)	1,4 (0,4)	1,3 (0,3)	1,4 (0,4)
- Natuur- en scheikunde	1,7 (0,6)	1,6 (0,5)	1,4 (0,4)	1,4 (0,4)	1,4 (0,5)
- Fysische aardrijkskunde	1,5 (0,5)	1,3 (0,4)	1,4 (0,4)	1,3 (0,4)	1,3 (0,4)
Frequentie bijscholing [1= 0 tot 5= > 35 uur]					
- Rekenen	2,0 (1,0)	1,8 (0,9)	2,7 (1,0)	1,6 (0,8)	3,2 (1,5)
- Science	1,4 (0,7)	2,0 (1,0)	1,9 (0,8)	1,3 (0,6)	2,0 (1,5)

* internationale TIMSS-schaal, loopt van weinig prestatiegerichtheid naar veel; ** internationale TIMSS-schaal, loopt van veel werkdruk naar weinig.

In het zogenoemde lege model (Tabel 2.2) zijn nog geen potentieel verklarende factoren opgenomen. Het laat zien hoeveel procent van de variantie in rekenprestaties toe te schrijven is aan verschillen tussen leerlingen, verschillen tussen scholen en verschillen tussen klassen in de vijf geselecteerde landen.

Tabel 2.2

Verdeling van de variantie in het lege model rekenprestaties in TIMSS-2015 in Nederland, Vlaanderen, Engeland, Finland en Zweden*

	Variantie		
	% leerling	% school	% klas
Nederland	87,5	9,0	3,5
Vlaanderen	75,8	22,2	2,0
Engeland	67,3	0,6	32,2
Finland	82,4	1,2	16,3
Zweden	75,5	19,0	5,5

* gebaseerd op de eerste plausible value

In Nederland is de variantie op school- en klasniveau samen niet meer dan 12,5%. Dit betekent dat 87,5% van de variantie in toetsprestaties toe te schrijven is aan individuele verschillen tussen leerlingen die niet samenhangen met school en klas. In TIMSS 2011 was dit 82,6% (Meelissen et al., 2014). Vergeleken met de andere vier landen, is de gezamenlijke variantie op school- en klasniveau in Nederland het laagst.

In Engeland en Finland is de variantie tussen klassen groter dan de variantie tussen scholen. Dit zou erop kunnen duiden dat leerlingen in deze landen vaker naar prestatieniveau worden verdeeld over de klassen.

Na het vaststellen van het lege model zijn voor elk land de effecten van de in Tabel 2.1 genoemde input- en procesvariabelen in stappen opgenomen in het model en zijn per stap alle variabelen die geen significant effect lieten zien in het model verwijderd.

In de volgende tabel wordt per land het eindmodel gepresenteerd met alleen significante effecten.

Tabel 2.3

Gemiddelde effecten en standard error (se) van input en procesvariabelen op rekenprestaties in TIMSS-2015, eindmodel, voor Nederland, Vlaanderen, Engeland, Finland en Zweden

	Effect (se) [*]				
	Nederland	Vlaanderen	Engeland	Finland	Zweden
Input leerling					
Sekse (meisje)	-7,79 (1,92)	-8,56 (1,62)	-8,21 (2,27)	--	--
Taal thuis (toetstaal)	8,97 (2,57)	17,58 (2,53)	-9,98 (3,52)	17,92 (2,87)	18,35 (2,52)
Aantal boeken thuis weinig (< 25) ^{**}	-30,38 (3,40)	-28,46 (2,96)	-48,33 (2,89)	-46,89 (2,25)	-41,30 (2,27)
Aantal boeken thuis gemiddeld (26-100) ^{**}	-8,41 (2,89)	-10,59 (2,56)	-15,75 (2,90)	-16,31 (1,81)	-19,87 (2,06)
Input school					
SES (0-10% econ. achterstand) ^{***}	--	55,34 (6,99)	46,55 (10,20)	--	38,16 (7,40)
SES (11-25% econ. achterstand) ^{***}	--	43,64 (7,06)	15,83 (9,77)	--	24,23 (8,18)
SES (26-50% econ. achterstand) ^{***}	--	37,29 (7,98)	2,29 (10,65)	--	13,40 (10,68)
Proces: perceptie leerkracht					
Prestatiegerichtheid schoolklimaat	3,69 (1,48)	--	--	--	--
Frequentie bijscholing rekenen	--	-4,02 (1,54)	--	--	--

^{*} op basis van alle 5 plausible values; significant voor $\alpha < 0,05$; ^{**} ten opzichte van aantal boeken thuis veel (meer dan 100 boeken); ^{***} ten opzichte van meer dan 50% economische achterstandsleerlingen

In alle landen zijn de taal thuis (leerling spreekt altijd of bijna altijd thuis de taal van de toets) en aantal boeken thuis (beschouwd als een indicator voor sociaal culturele status) belangrijke voorspellers van rekenprestaties. Na opname van de leerlingkenmerken in het model wordt in totaal 8% van de variantie verklaard, waarvan 5% op leerlingniveau en 3% op schoolniveau (op basis van de eerste *plausible value*).

In tegenstelling tot de andere landen, presteren leerlingen in Engeland die thuis (bijna) altijd Engels praten minder goed dan leerlingen die thuis een andere taal spreken. In Vlaanderen, Finland en Zweden lijkt de taal thuis een grotere rol te spelen dan in Nederland. In Nederland en Vlaanderen lijkt het effect van 'het aantal boeken thuis' kleiner dan in de andere drie landen. In Engeland en Finland hangen daarentegen het aantal boeken thuis en rekenprestaties relatief sterk samen. Een leerling in Engeland met weinig boeken thuis (maximaal één boekenplank, 33% van alle leerlingen) scoort maar liefst 48 punten lager op de rekentoets dan een leerling met veel boeken thuis (minimaal twee boekenkasten, 32% van alle leerlingen, zie Tabel 2.1). Van de schoolkenmerken wordt voor Vlaanderen, Engeland en Zweden een significant negatief effect gevonden voor het percentage economische achterstandsleerlingen op school (inschatting schoolleiders). In Nederland is deze samenhang er niet. Er zijn echter in Nederland relatief weinig scholen die door de schoolleiders bestempeld worden als een school met meer dan een kwart achterstandsleerlingen (zie Tabel 2.1).

Gegeven de invloed van taal thuis en het aantal boeken thuis, blijkt in Nederland van alle meegenomen procesvariabelen, alleen de prestatiegerichtheid van het schoolklimaat positief samen te hangen met rekenprestaties. Leerlingen van een leerkracht die een standaarddeviatie (1,2) boven het gemiddelde op deze schaal scoort, presteren ruim 8 punten hoger dan leerlingen van een leerkracht die een standaarddeviatie onder het gemiddelde op deze schaal scoort. Nederland behoort in TIMSS-2015 tot de landen waar de prestatiegerichtheid relatief laag is (Meelissen & Punter, 2016). In de overige vier landen is er geen verband, dit geldt zowel voor landen met een hoge prestatiegerichtheid (zoals Engeland) als landen waar de prestatiegerichtheid onder het internationaal gemiddelde ligt (zoals Vlaanderen, Mullis et al., 2016). In Vlaanderen heeft de hoeveelheid bijscholing een negatief effect (mogelijk volgen leerkrachten van klassen met zwakke rekenaars meer scholing), maar verder zijn er geen relaties tussen de onderzochte procesfactoren en rekenprestaties in de vijf landen gevonden.

2.4 Resultaten natuuronderwijs

Tabel 2.3

Verdeling van de variantie in het lege model natuuronderwijsprestaties in TIMSS-2015 in Nederland, Vlaanderen, Engeland, Finland en Zweden*

	Variantie		
	<i>% leerling</i>	<i>% school</i>	<i>% klas</i>
Nederland	78,5	15,7	5,9
Vlaanderen	71,3	26,2	2,5
Engeland	64,0	3,6	32,4
Finland	79,3	0,8	19,9
Zweden	71,3	26,2	2,5

gebaseerd op de eerste plausible value

De variantie die in Nederland toegeschreven kan worden aan verschillen tussen scholen, is bij natuuronderwijs hoger dan bij rekenen (21,6% versus 12,5%). Mogelijk wordt dit veroorzaakt omdat er meer spreiding is tussen scholen in mate waarin en de wijze waarop natuuronderwijs onderwezen wordt. Evenals voor rekenen is de variantie tussen klassen in Engeland en Finland het hoogst, hetgeen een structureel andere indeling van leerlingen over de klassen doet vermoeden. In Tabel 2.4 wordt per land het eindmodel gepresenteerd met alleen de significante effecten voor het betreffende land.

Tabel 2.4

Gemiddelde effecten en standard error (se) van input en procesvariabelen op natuuronderwijsprestaties in TIMSS-2015, eindmodel, voor Nederland, Vlaanderen, Engeland, Finland en Zweden

	Effect (se) [*]				
	Nederland	Vlaanderen	Engeland	Finland	Zweden
Input leerling					
Sekse (meisje)	--	--	--	5,92 (1,78)	--
Taal thuis (toetstaal)	13,31 (3,87)	29,30 (2,66)	10,06 (4,19)	26,24 (2,54)	29,15 (2,66)
Aantal boeken thuis weinig (< 25) ^{**}	-29,96 (3,79)	-35,61 (2,48)	-46,40 (2,81)	-45,16 (2,36)	-44,92 (2,71)
Aantal boeken thuis gemiddeld (26-100) ^{**}	-7,81 (3,81)	-15,90 (1,90)	-20,84 (2,82)	-19,33 (1,88)	-21,48 (2,58)
Input school					
SES (0-10% econ. achterstand) ^{***}	--	58,14 (6,64)	47,29 (8,61)	--	54,07 (7,94)
SES (11-25% econ. achterstand) ^{***}	--	48,38 (6,87)	17,83 (8,69)	--	37,04 (8,61)
SES (26-50% econ. achterstand) ^{***}	--	43,08 (7,60)	14,12 (9,86)	--	24,73 (11,48)
Proces: perceptie leerkracht					
Prestatiegerichtheid schoolklimaat	5,15 (1,93)	--	--	3,63 (1,17)	--
Zelfvertrouwen didactische vaardigheden	--	--	--	7,29 (2,95)	--
Mate van toerusting Fysische aardrijkskunde	--	--	-9,87 (4,52)	--	--

^{*} op basis van alle 5 plausible values; significant voor $\alpha < 0,05$; ^{**} ten opzichte van aantal boeken thuis veel (meer dan 100 boeken); ^{***} ten opzichte van meer dan 50% economische achterstandsleerlingen

Het beeld voor natuuronderwijs (science) is in Nederland vergelijkbaar met het beeld voor rekenen. Taal thuis, het aantal boeken thuis (beschouwd als indicator voor sociaal culturele status) en de prestatiegerichtheid van de schoolklimaat hangen significant samen met toetsprestaties. In tegenstelling tot rekenen speelt bij natuuronderwijs het geslacht van de leerling in Nederland geen rol. Door de opname van de leerlingkenmerken in het model wordt in totaal 8% verklaard, waarvan 5% op leerlingniveau, 3% op schoolniveau en 1% op klasniveau (op basis van de eerste *plausible value*).

De achterstand in science van leerlingen die thuis de toetstaal niet of nauwelijks spreken is in de Vlaanderen, Finland en Zweden groter dan in Nederland. In de vorige paragraaf zagen we dat in Engeland thuis Engels sprekende leerlingen iets minder goed presteren in rekenen vergeleken met leerlingen die een thuis een andere taal spreken. Voor science is het effect van thuistaal in Engeland omgekeerd; leerlingen die thuis (bijna) altijd Engels spreken presteren beter in science dan leerlingen die thuis niet de toetstaal spreken.

Het aantal boeken thuis vertoont een significante samenhang met de toetsprestaties in science. Dit effect lijkt relatief groot in Engeland, Finland en Zweden. Evenals bij rekenen wordt er voor Vlaanderen, Engeland en Zweden ook een significant negatief effect gevonden voor het schoolkenmerk 'percentage economische achterstandsleerlingen'.

De prestatiegerichtheid van het schoolklimaat (in de perceptie van de leerkracht) hangt in Nederland positief samen met de prestaties op de natuuronderwijstoets. Leerlingen van een leerkracht die een standaarddeviatie (1,2) boven het gemiddelde op deze schaal scoort, presteren ruim 12 punten hoger dan leerlingen van een leerkracht die een standaarddeviatie onder het gemiddelde op deze schaal scoort. Ook Finland laat een positieve samenhang zien tussen prestatiegerichtheid en de scores voor science.

2.5 Conclusie

In Nederland is de samenhang tussen de leerlingkenmerken 'taal thuis' en 'het aantal boeken thuis' (proxy voor sociaal culturele status van het gezin) en de TIMSS-toetsprestaties minder sterk dan in Vlaanderen, Finland en Zweden. In de Nederlandse en Finse data is voor rekenen en science geen significante samenhang gevonden tussen prestaties en het percentage achterstandsleerlingen op een school. Afgaand op de inschatting van de schoolleiders, zijn er in Finland en in Nederland ook relatief weinig scholen met een groot aandeel economische achterstandsleerlingen.

In Vlaanderen, Engeland en Zweden hangen zowel weinig boeken thuis als een relatief hoog percentage achterstandsleerlingen op school relatief sterk samen met de prestaties in de exacte vakken. In TIMSS is sinds 2011 discussie over het nut van de achtergrondvariabele 'het aantal boeken thuis' vanwege de komst van de e-reader (waardoor er minder gedrukte boeken zou worden gelezen). Uit de resultaten blijkt dat deze inputfactor als voorspeller voor prestaties in 2015 in ieder geval nog steeds relevant is.

De taal die leerlingen thuis spreken speelt in Engeland mogelijk een andere rol in relatie tot de toetsscores dan in de overige vier landen. In tegenstelling tot de andere landen, rekenen leerlingen die thuis (bijna) altijd Engels praten minder goed dan leerlingen die thuis een andere taal dan de toetstaal spreken. Voor science is de relatie omgekeerd.

Gegeven het effect van taal thuis en het aantal boeken thuis blijkt uit deze analyses dat in Nederland de prestatiegerichtheid van het schoolklimaat positief samenhangt met prestaties in de exacte vakken. In de andere landen vinden we deze samenhang alleen terug bij Finland voor science (natuuronderwijs). Met de TIMSS-data kan slechts een samenhang en geen causaliteit worden vastgesteld. Meer prestatiegerichtheid van de school zou een positief effect kunnen hebben op leerprestaties, maar het is ook goed mogelijk dat op scholen met relatief hoog presterende leerlingen, de verwachtingen van leerkrachten, ouders en schoolleiders over de prestaties van de leerlingen hoger liggen. Met andere woorden: of leerlingen presteren beter bij hogere verwachtingen en/of beter presterende leerlingen zorgen voor hogere verwachtingen. Gezien de relatieve minder hoge prestatiegerichtheid van Nederland en opzichte van de overige TIMSS-landen (Mullis et al., 2013), zou nader onderzoek naar de relatie tussen de prestatiegerichtheid van scholen en leerprestaties desalniettemin voor Nederland relevant zijn.

Uit de internationale resultaten van TIMSS-2011 bleek dat leerlingen op scholen die in staat waren goede werkcondities voor hun leerkrachten te realiseren, beter presteerden dan leerlingen op scholen

waar de werkcondities door de leerkrachten minder gewaardeerd werden (Mullis et al., 2013). De gevonden (mogelijk wederkerige) relatie tussen werkcondities en leerprestaties was de voornaamste reden om in de TIMSS-leerkrachtvragenlijst van 2015 meer aandacht aan dit onderwerp te besteden. Zo is er een uitgebreide een vraag over werkdruk van leerkrachten (*Challenges Facing Teachers*) opgenomen (Mullis et al., 2016). Uit TIMSS-2015 blijkt dat de werkdruk van Nederlandse leerkrachten hoger ligt dan internationaal gemiddeld (Mullis et al., 2016). De analyses tonen in de vijf onderzochte landen echter geen (directe) relatie aan tussen werkdruk van de leerkracht en leerprestaties. In Nederland, Vlaanderen en Zweden geldt dit ook voor het zelfvertrouwen in het didactisch handelen en de gepercipieerde toerusting in de TIMSS-leerstofonderdelen van rekenen en natuuronderwijs.

3. Uitgevoerde curriculum in relatie tot leerprestaties

3.1 Variabelen

Voor dit onderzoek is per inhoudsdomein nagegaan in hoeverre er een samenhang is tussen de mate waarin het inhoudsdomein behandeld is en de prestaties van de leerling, gegeven een aantal achtergrondkenmerken van de leerling en de school. In TIMSS worden namelijk voor elke leerling niet alleen vijf *plausible values* voor het gehele vakgebied (rekenen of natuuronderwijs) berekend maar ook per inhoudsdomein. Zoals eerder is aangegeven, wordt in TIMSS aan leerkrachten per inhoudsdomein een lijst met leerstofgebieden voorgelegd met de vraag in hoeverre deze gebieden behandeld zijn (vorig schooljaar, dit schooljaar en (nog) niet behandeld; zie ook Meelissen & Punter, 2016). Voor de analyses zijn de categorieën 'vorig schooljaar' en 'dit schooljaar' samengenomen.

3.2 Analysemethode

Voor de exploratie van de samenhang tussen toetsprestaties en wat er onderwezen is, is gebruik gemaakt van dezelfde multilevel techniek en dezelfde analysemethoden als beschreven in paragraaf 2.2. De analyses zijn voor elk inhoudsdomein apart uitgevoerd. De volgende variabelen zijn in de analyses meegenomen: sekse, taal thuis, aantal boeken thuis, percentage achterstandsl leerlingen op school, urbanisatiegraad school en de mate waarin de leerstofgebieden binnen het inhoudsdomein behandeld zijn (uitgevoerd curriculum).

3.3 Resultaten rekenen

In Tabel 3.1 wordt per inhoudsgebied het zogenoemde lege model weergegeven. Het model laat per inhoudsdomein zien hoeveel procent van de variantie in prestaties in de inhoudsdomeinen toe te schrijven is aan verschillen tussen leerlingen, verschillen tussen scholen en verschillen tussen klassen.

Tabel 3.1

*Verdeling van de variantie in het lege model rekenprestaties in TIMSS-2015**

	Variantie		
	% leerling	% school	% klas
Rekenen	87,5	9,0	3,5
Getallen	88,6	8,2	3,3
Geometrische vormen en meten	87,0	10,0	3,0
Gegevensweergave	89,3	8,5	2,2

* gebaseerd op de eerste plausible value

In het vorige hoofdstuk bleek dat de variantie op school en klasniveau voor rekenen samen niet meer dan 12,5% is. Dit betekent dat 87% van de variantie in toetsprestaties toe te schrijven is aan verschillen tussen leerlingen die niet samenhangen met school- en klaskenmerken. De verschillen tussen de domeinen zijn zeer klein. Het kleinste 'schooleffect' is er voor Getallen, de tussenschoolse variantie is het grootst voor Geometrische vormen en meten. Van de drie inhoudsgebieden past dit laatste domein het minst goed bij het Nederlandse beoogde en uitgevoerde curriculum (Meelissen & Punter, 2016). Hierdoor is er waarschijnlijk ook iets meer spreiding tussen scholen in de mate waarin ze hier aandacht aan besteden.

Tabel 3.2 geeft het eindmodel (met alleen variabelen met significante effecten) voor de verschillende inhoudsdomeinen weer.

Tabel 3.2

Gemiddelde effecten en standard error (se) van leerling-, school- en uitgevoerd curriculum op rekenprestaties in TIMSS-2015, eindmodel

	Effect (se)*			
	Rekenen	Getallen	Geometrische vormen en meten	Gegevensweergave
Input leerling				
Sekse (meisje)	-7,85 (1,63)	-10,12 (2,48)	-6,22 (2,00)	--
Taal thuis (Nederlands)	11,53 (2,29)	11,54 (2,86)	11,45 (4,13)	18,36 (6,23)
Aantal boeken thuis weinig (< 25)**	-31,01 (2,91)	-31,83 (2,55)	-28,12 (4,37)	-37,42 (6,79)
Aantal boeken thuis gemiddeld (26-100)**	-10,36 (2,62)	-11,78 (2,74)	-7,23 (3,80)	-7,94 (4,55)

* op basis van alle 5 plausible values; significant voor $\alpha < 0,05$; ** ten opzichte van aantal boeken thuis veel (meer dan 100 boeken); *** geen significante effecten voor schoolfactoren (% achterstandsleerlingen, urbanisatiegraad) en uitgevoerd curriculum (% onderwezen onderwerpen per inhoudsdomein)

De mate waarin de leerstof van elk inhoudsdomein voor rekenen is behandeld laat noch een samenhang zien met de toetsprestaties in rekenen noch in de afzonderlijke rekendomeinen. Alleen de leerlingkenmerken sekse, taal thuis en het aantal boeken thuis hebben een significant effect op de toetsprestaties in de verschillende domeinen. De relaties tussen thuistaal en boeken thuis met prestaties in Gegevensweergave zijn relatief het sterkst. Leerlingen die aangeven dat er thuis weinig

boeken zijn (maximaal één boekenplank) scoren 37 punten lager in Gegevensweergave dan leerlingen die aangeven dat er thuis veel boeken zijn (minimaal twee boekenkasten).

3.4 Resultaten natuuronderwijs

De resultaten voor natuuronderwijs staan in Tabel 3.3 (lege model) en Tabel 3.4 (eindmodel).

Tabel 3.3

*Verdeling van de variantie in het lege model natuuronderwijsprestaties in TIMSS-2015**

	Variantie		
	% leerling	% school	% klas
Natuuronderwijs	78,5	15,7	5,9
Biologie	77,8	16,6	5,6
Natuur- en scheikunde	79,4	14,1	6,5
Fysische aardrijkskunde	83,5	11,7	4,9

* gebaseerd op de eerste plausible value

De tussenschoolse variantie en de variantie tussen klassen is het laagst voor Fysische aardrijkskunde. Bijna 84% van alle variantie in toetsprestaties op dit inhoudsgebied kan toegeschreven worden aan individuele leerlingverschillen. De volgende tabel geeft het eindmodel weer.

Tabel 3.4

Gemiddelde effecten en standard error (se) van leerling-, school- en uitgevoerd curriculum op natuuronderwijsprestaties in TIMSS-2015, eindmodel

	Effect (se)*			
	Natuur- onderwijs	Biologie	Natuur- en scheikunde	Fysische aardrijkskunde
Input leerling				
Sekse (meisje)	--	9,70 (2,19)	--	-13,17 (3,60)
Taal thuis (Nederlands)	14,90 (3,10)	13,36 (2,69)	17,76 (3,74)	11,49 (3,73)
Aantal boeken thuis weinig (< 25)**	-29,93 (3,1)	-29,19 (2,60)	-25,09 (4,40)	-46,21 (5,83)
Aantal boeken thuis gemiddeld (26-100)**	-9,10 (2,93)	-10,29 (3,56)	-2,86 (3,35)	-23,74 (6,16)

* op basis van alle 5 plausible values; significant voor $\alpha < 0,05$; ** ten opzichte van aantal boeken thuis veel (meer dan 100 boeken); *** geen significante effecten voor schoolfactoren (% achterstandsleerlingen, urbanisatiegraad) en uitgevoerd curriculum (% onderwezen onderwerpen per inhoudsgebied)

Evenals voor rekenen, geldt voor natuuronderwijs dat het gemiddeld percentage onderwezen onderwerpen per inhoudsgebied noch effect heeft op de prestaties op het betreffende inhoudsgebied, noch op de gehele natuuronderwijstoets.

De tabel laat wel een andere opmerkelijke uitkomst zien; leerlingen met weinig boeken thuis (indicator van sociaal culturele status) scoren ruim 46 punten lager op het domein Fysische aardrijkskunde dan leerlingen met veel boeken thuis. Ook leerlingen met een gemiddeld aantal boeken thuis scoren fors lager in Fysische aardrijkskunde dan leerlingen met meer dan 100 boeken thuis. Het verschil tussen deze leerlinggroepen is ook in de andere inhoudsgebieden significant, maar niet zo groot als voor Fysische

aardrijkskunde. Taal thuis heeft het grootste effect voor Natuur- en scheikunde. Leerlingen die thuis bijna nooit Nederlands spreken, scoren bijna 18 punten lager op de natuur- en scheikundeopgaven.

3.5 Conclusie

Het doel van deze analyse was om na te gaan in hoeverre er een verband is tussen wat een leerling onderwezen heeft gekregen (het uitgevoerde curriculum) en de toetscore van de leerling (het gerealiseerde curriculum). Voor beide vakgebieden is er geen relatie tussen de mate waarin de leerstofonderdelen van de TIMSS-toets onderwezen zijn en de toetsscores. Dit resultaat komt overeen met een studie waarin secundaire analyses zijn uitgevoerd op TIMSS-2011 data; ook hieruit bleek dat er geen relatie is tussen het uitgevoerde en gerealiseerde curriculum (Luyten, 2017). Een mogelijke verklaring voor het ontbreken van deze samenhang zou te maken kunnen hebben met de wijze waarop het uitgevoerde curriculum in TIMSS bevraagd wordt (en bevraagd kan worden). Leerkrachten geven per leerstofonderdeel van een inhoudsdomen aan of dit onderdeel al aan bod is geweest. Het geeft geen inzicht in de hoeveel tijd die aan een onderwerp is besteed, hoe diepgaand een leerstofgebied behandeld is en hoe effectief de instructie is geweest.

De resultaten laten wel een sterke samenhang zien tussen het aantal boeken thuis (indicator voor sociaal culturele status) en de prestaties in Fysische aardrijkskunde. Dit zou erop kunnen duiden dat leerlingen uit een gezin met een hoge sociaal culturele status thuis meer meekrijgen over aardrijkskundeonderwerpen dan de overige leerlingen. De relatie tussen het aantal boeken thuis en de prestaties in de andere inhoudsgebieden is minder sterk dan voor Fysische aardrijkskunde, maar evengoed nog behoorlijk. Het zou daarom interessant zijn om bijvoorbeeld in toekomstige TIMSS-studies, meer informatie te verzamelen over de sociaal culturele status, zoals museumbezoek en buitenschoolse activiteiten van de leerlingen. Het ontwikkelen van aanvullende indicatoren voor het meten van sociaal culturele status is ook relevant in de context van de eerdere genoemde discussie over het aantal boeken thuis en de beschikbaarheid van e-readers.

De achterstand in toetsprestaties van leerlingen die thuis weinig Nederlands spreken, is het grootst in natuur- en scheikunde. Nader onderzoek zal moeten uitwijzen in hoeverre de taligheid van de TIMSS-opgaven (en daarmee de leesvaardigheid) binnen de verschillende inhoudsdomen een rol speelt bij de prestaties van leerlingen.

4. Prestaties leerlingen naar referentieniveau

4.1 Variabelen en analysemethode

In TIMSS worden vier zogenoemde referentiepunten of *benchmarks* onderscheiden. Een referentiepunt wordt gebruikt om aan te geven hoeveel procent van de leerlingen in een land een bepaalde standaard haalt. In TIMSS worden vier referentiepunten onderscheiden (Mullis, Martin, Foy & Hooper, 2016):

- Het *geavanceerde niveau* gerelateerd aan een toetsscore van minimaal 625
- Het *hoge niveau* gerelateerd aan een toetsscore van 550 tot 625
- Het *middenniveau* gerelateerd aan een toetsscore van 475 tot 550
- Het *basisniveau* gerelateerd aan een toetsscore van 400 tot 475
- *Onder basisniveau* gerelateerd aan een toetsscore lager dan 400

Voor elk inhouds- en cognitief domein is nagegaan hoe de leerlingen verdeeld zijn over de verschillende referentieniveaus. Hiervoor zijn de leerlingen drie groepen verdeeld: het geavanceerde niveau (de zogenoemde excellente leerlingen), het hoge niveau en het midden/lage niveau (lager dan 550). Voor deze analyses is gebruik gemaakt van de *IEA's IDB-analyzer*. Het betreft hier een exploratie, verschillen tussen de jaren zijn niet op significantie getoetst.

4.2 Resultaten rekenen

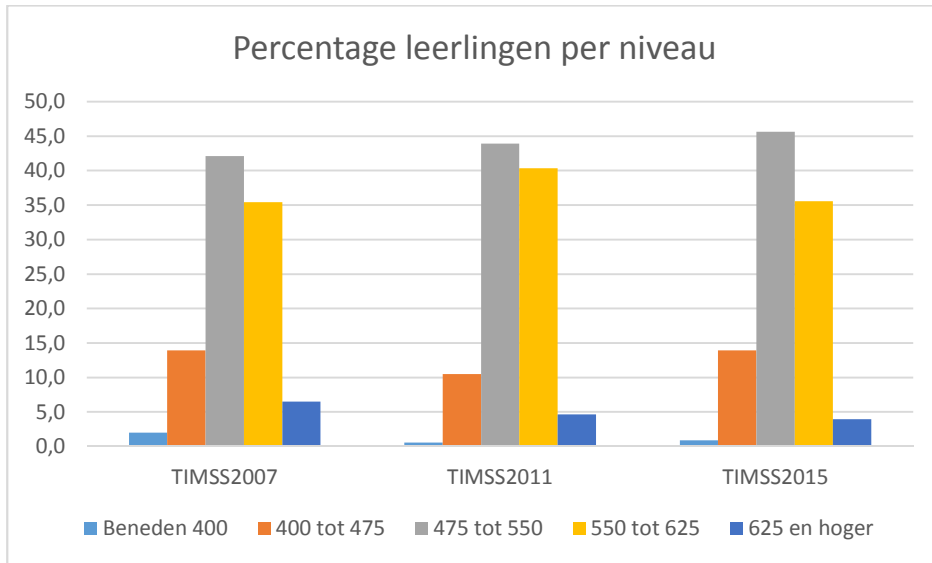
Rekenen

Ter achtergrondinformatie staan in Tabel 4.1 de aantallen leerlingen die in de periode 2007-2015 een bepaald referentieniveau voor rekenen hebben behaald en in Figuur 4.1 de verdeling in percentages. De aantallen in met name de laagste categorie (een score onder de 400) zijn zeer klein. In alle jaren heeft slechts 1% van alle leerlingen het basisniveau voor rekenen niet gehaald.

Tabel 4.1

Verdeling van de leerlingen over de verschillende referentieniveaus rekenen per TIMSS-meting, in absolute aantallen

	TIMSS-2007	TIMSS-2011	TIMSS-2015
Beneden 400	67	18	42
400 tot 475	468	340	631
475 tot 550	1410	1418	2059
550 tot 625	1186	1302	1605
625 en hoger	217	150	179
Totaal	3349	3229	4515

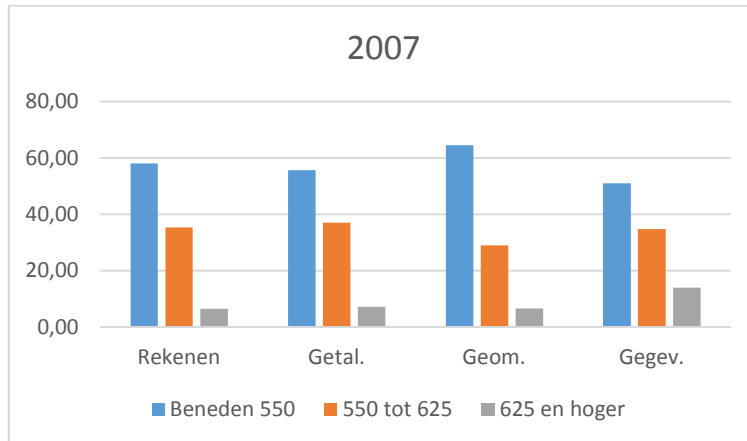


Figuur 4.1: Verdeling van de leerlingen over de verschillende referentieniveaus rekenen per TIMSS-meting, in percentages

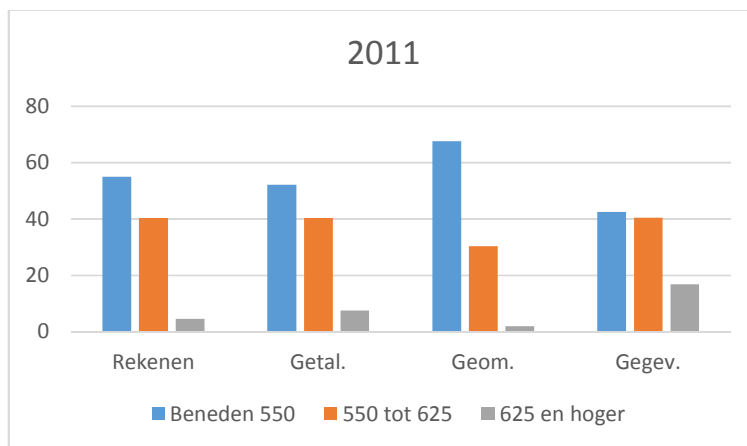
Hoewel de aantallen leerlingen die de TIMSS-toets hebben gemaakt per jaar verschilt (met name in 2015 ligt deze hoger dan in 2007 en 2011) is verdelingstrend over de verschillende niveaus nagenoeg gelijk gebleven. De verdelingen van TIMSS-2007 en TIMSS-2015 lijken meer op elkaar dan die van TIMSS-2011. In 2011 waren er namelijk relatief meer leerlingen die op het hoge niveau presteerden en minder leerlingen die niet verder kwamen dan het basisniveau. Het percentage leerlingen op geavanceerd niveau daalt sinds 2007. Het percentage leerlingen op het middenniveau stijgt.

Inhoudsdomeinen

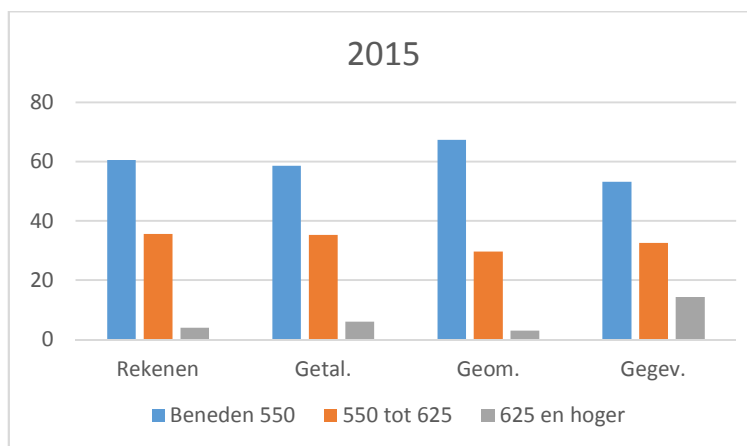
De volgende figuren tonen de verdelingen van leerlingen naar referentieniveau per inhoudsdomein en per TIMSS-meting. Tabel 4.2 geeft per domein de percentages leerlingen weer die in de drie TIMSS-metingen het geavanceerde rekenniveau hebben gehaald. Als achtergrondinformatie staat in Tabel 4.3 tevens informatie over de geschiktheid van de TIMSS-toets voor het uitgevoerde curriculum (zie hoofdstuk 3), eveneens per TIMSS-meting.



Figuur 4.2: Verdeling van de leerlingen over de verschillende referentieniveaus rekenen per inhoudsdomoien TIMSS-2007, in percentages



Figuur 4.3: Verdeling van de leerlingen over de verschillende referentieniveaus rekenen per inhoudsdomoien TIMSS-2011, in percentages



Figuur 4.4: Verdeling van de leerlingen over de verschillende referentieniveaus rekenen per inhoudsdomoien TIMSS-2015, in percentages

Tabel 4.2

Percentage leerlingen dat het **geavanceerde** referentieniveau haalt voor rekenen (score 625 en hoger), per inhoudsdomein, in TIMSS-2007, TIMSS-2011 en TIMSS-2015

	% TIMSS-2007	% TIMSS-2011	% TIMSS-2015
Getallen	7,2	7,5	6,1
Geometrische vormen en meten	6,6	2,0	3,0
Gegevensweergave	14,1	17,0	14,3
Totaal rekenen	6,5	4,7	4,0

Tabel 4.3

Percentage leerlingen dat de inhoudelijke domeinen voor rekenen in groep 6 of eerder onderwezen heeft gekregen, TIMSS-2007, TIMSS-2011 en TIMSS-2015

	% onderwezen onderwerpen in groep 6		
	TIMSS-2007	TIMSS-2011	TIMSS-2015
Getallen	64	63	70
Geometrische vormen en meten	45	43	51
Gegevensweergave	71	84	84
Totaal rekenen	60	63	64

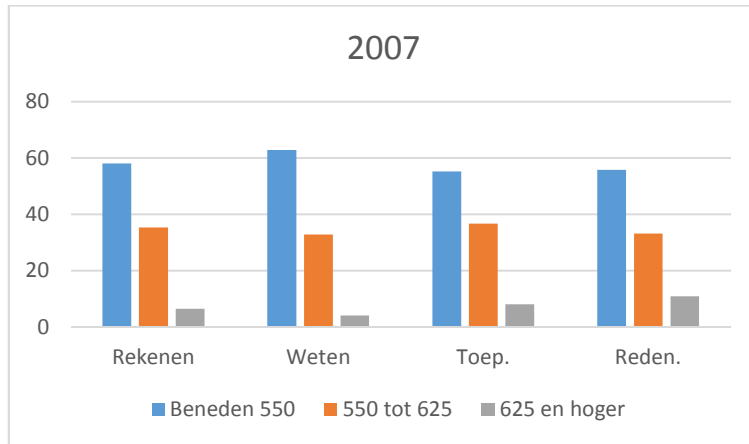
Vergeleken met de andere domeinen, past het inhoudsgebied Gegevensweergave het beste bij het Nederlandse uitgevoerde curriculum. Het aandeel excellente leerlingen (toetscore 625 en hoger) in dit gebied is sinds 2007 op een vergelijkbaar niveau gebleven, met een uitschieter naar boven in 2011. Ten opzichte van de andere inhoudsdomeinen is het aandeel leerlingen met een score van 625 en hoger voor Gegevensweergave aanmerkelijk groter. Vergeleken met 2007 zijn in 2011 en 2015 gemiddeld meer onderwerpen over Gegevensweergave in groep 6 aan bod gekomen, dit heeft echter in 2015 niet geresulteerd in een groter aandeel excellente leerlingen op dit gebied.

De percentages excellente leerlingen in het inhoudsgebied Geometrische vormen en meten is in TIMSS het laagst: 7% in 2007, 2% in 2011 en 3% in 2015. In 2015 zijn gemiddeld meer onderwerpen binnen Geometrische vormen en meten van de TIMSS-toets (51%) onderwezen dan in de eerdere jaren. Dit gaat echter niet gepaard met een stijging van het aandeel leerlingen dat het geavanceerde niveau haalt.

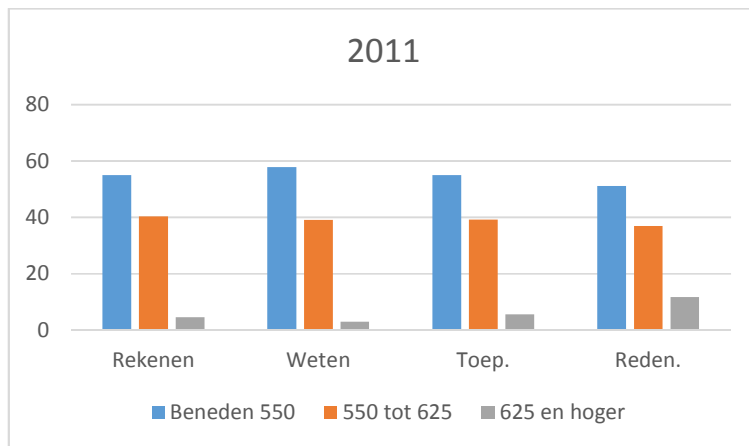
In 2015 zijn ook meer TIMSS-onderwerpen in het domein Getallen in groep 6 onderwezen dan in eerdere TIMSS-metingen. Ook voor Getallen geldt dat deze toename niet samen lijkt te gaan met een toename van de leerlingen die het meest geavanceerde niveau halen.

Cognitieve domeinen

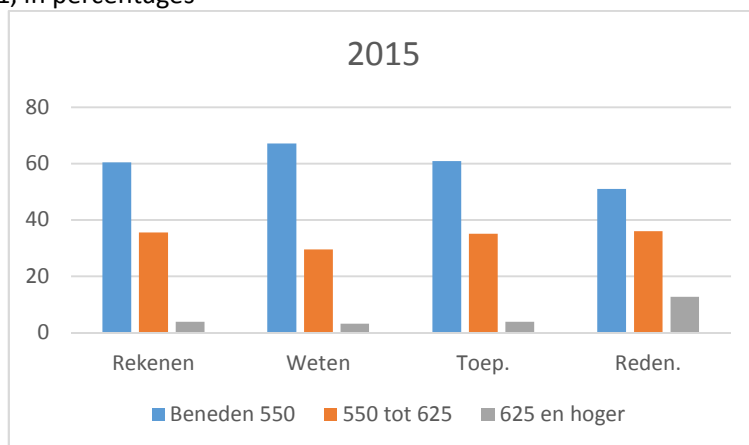
De volgende figuren tonen de verdelingen van leerlingen naar referentieniveaus per cognitief domein en per TIMSS-meting. Tabel 4.4 toont het percentage leerlingen dat voor elk cognitief domein het geavanceerde niveau haalt (score 625 en hoger) in de drie TIMSS-metingen.



Figuur 4.5 Verdeling van de leerlingen over de verschillende referentieniveaus rekenen per cognitief domein TIMSS-2007, in percentages



Figuur 4.6: Verdeling van de leerlingen over de verschillende referentieniveaus rekenen per cognitief domein TIMSS-2011, in percentages



Figuur 4.7: Verdeling van de leerlingen over de verschillende referentieniveaus rekenen per cognitief domein TIMSS-2015, in percentages

Tabel 4.4

Percentage leerlingen dat het **geavanceerde** referentieniveau haalt voor rekenen (score 625 en hoger), per cognitief domein, in TIMSS-2007, TIMSS-2011 en TIMSS-2015

	% TIMSS-2007	% TIMSS-2011	% TIMSS-2015
Weten	4,3	3,0	3,7
Toepassen	8,1	5,7	3,9
Redeneren	10,9	11,9	12,8
Totaal rekenen	6,5	4,7	4,0

In alle drie de TIMSS-metingen is het aandeel leerlingen dat het geavanceerde niveau haalt het grootst voor Redeneren. Dit aandeel is in de afgelopen acht jaar zelfs licht gestegen. In deze periode hebben maar weinig leerlingen (rond de 4%) met de TIMSS-weetopgaven een score van 625 of hoger gehaald. De relatief grootste daling in het percentage excellerende leerlingen is te vinden in het domein Toepassen, dit percentage is in 2015 gehalveerd ten opzichte van 2007.

4.3 Resultaten natuuronderwijs

Natuuronderwijs

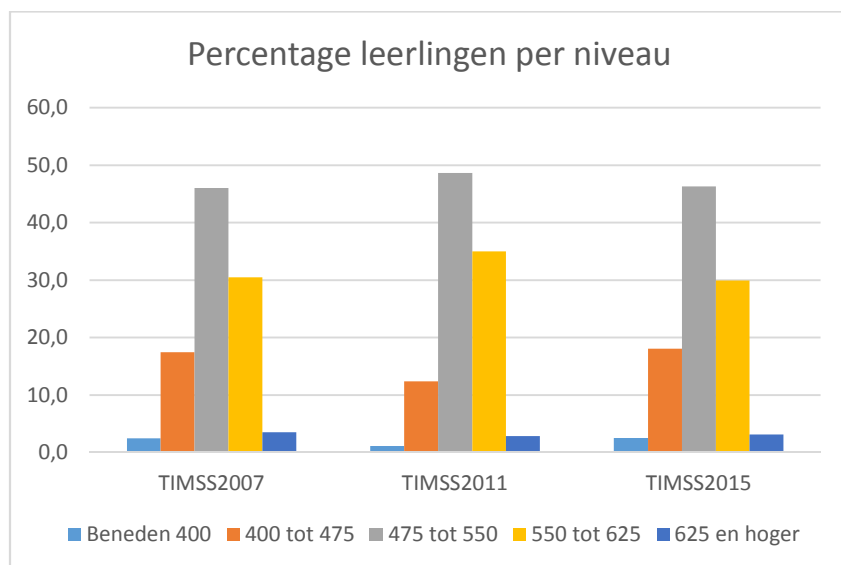
In Tabel 4.5 staan de aantallen leerlingen die in de periode 2007-2015 voor natuuronderwijs een bepaald referentieniveau hebben behaald.

Tabel 4.5

Verdeling van de leerlingen over de verschillende referentieniveaus natuuronderwijs per TIMSS-meting, in absolute aantallen

	TIMSS-2007	TIMSS-2011	TIMSS-2015
Beneden 400	84	37	115
400 tot 475	585	400	816
475 tot 550	1541	1570	2091
550 tot 625	1021	1129	1352
625 en hoger	117	93	142
Totaal	3349	3229	4515

De aantallen in de laagste categorie zijn zeer klein (een score onder de 400). In alle jaren heeft slechts 1% tot 3% van alle leerlingen het basisniveau niet gehaald. Figuur 4.2 geeft de verdeling weer van leerlingen die in de periode 2007-2015 voor natuuronderwijs een bepaalde referentieniveau hebben behaald.

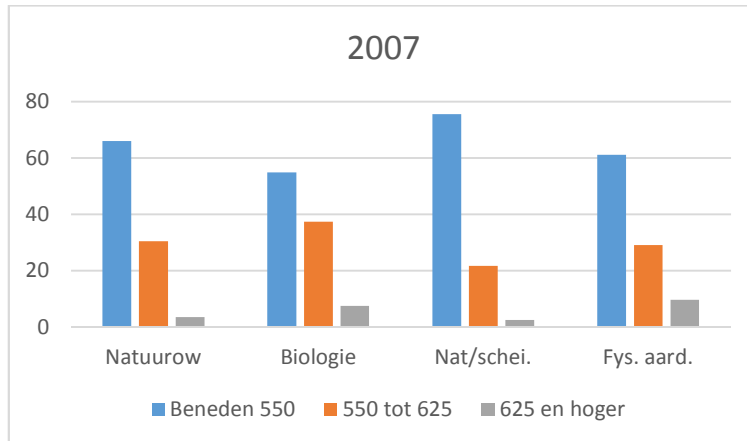


Figuur 4.8: Verdeling van de leerlingen over de verschillende referentieniveaus natuuronderwijs per TIMSS-meting, in percentages

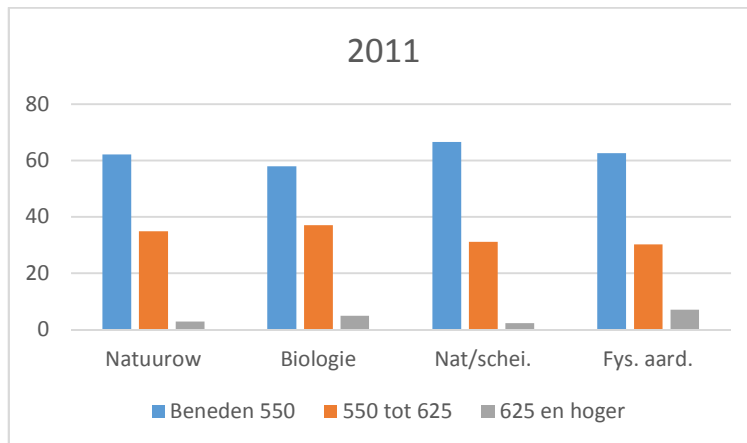
Evenals voor rekenen is de verdelingstrend over de verschillende niveaus is nagenoeg gelijk gebleven. Het jaar 2011, het jaar waarop er voor natuuronderwijs een duidelijk prestatietoename te zien was in Nederland, vertoont een licht afwijkend patroon vergeleken met TIMSS-2007 en TIMSS-2015. In 2011 waren er namelijk duidelijk meer leerlingen die op het hoge niveau presteerden en minder leerlingen die niet verder kwamen dan het basisniveau.

Inhoudsdomeinen

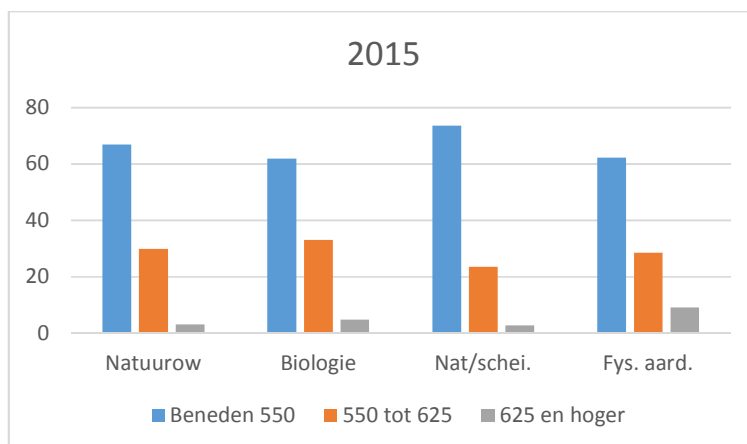
De volgende figuren laten voor 2007, 2011 en 2015 de verdeling over de referentieniveaus zien per inhoudsdomein. In Tabel 4.6 wordt per inhoudsdomein het percentage leerlingen weergegeven dat in de afgelopen acht jaar het geavanceerde referentieniveau heeft behaald. Tabel 4.7 toont per TIMSS-jaar de geschiktheid van de toets voor het uitgevoerde curriculum van natuuronderwijs.



Figuur 4.9: Verdeling van de leerlingen over de verschillende referentieniveaus natuuronderwijs per inhoudsdomein TIMSS-2007, in percentages



Figuur 4.10: Verdeling van de leerlingen over de verschillende referentieniveaus natuuronderwijs per inhoudsdomein TIMSS-2011, in percentages



Figuur 4.11: Verdeling van de leerlingen over de verschillende referentieniveaus natuuronderwijs per inhoudsdomein TIMSS-2015, in percentages

Tabel 4.6

Percentage leerlingen dat het **geavanceerde** referentieniveau haalt voor natuuronderwijs (score 625 en hoger), per inhoudsdomain, in TIMSS-2007, TIMSS-2011 en TIMSS-2015

	% TIMSS-2007	% TIMSS-2011	% TIMSS-2015
Biologie	7,6	5,0	4,9
Natuur- en scheikunde	2,6	2,3	2,8
Fysische aardrijkskunde	9,7	7,1	9,2
Totaal natuuronderwijs	3,5	2,9	3,1

Tabel 4.7

Percentage leerlingen dat de inhoudelijke domeinen voor natuuronderwijs in groep 6 of eerder onderwezen hebben gekregen, TIMSS-2007, TIMSS-2011 en TIMSS-2015

	% onderwezen onderwerpen in groep 6		
	TIMSS-2007	TIMSS-2011	TIMSS-2015
Biologie	61	60	58
Natuur- en Scheikunde	34	32	38
Fysische aardrijkskunde	50	54	59
Totaal natuuronderwijs	49	49	51

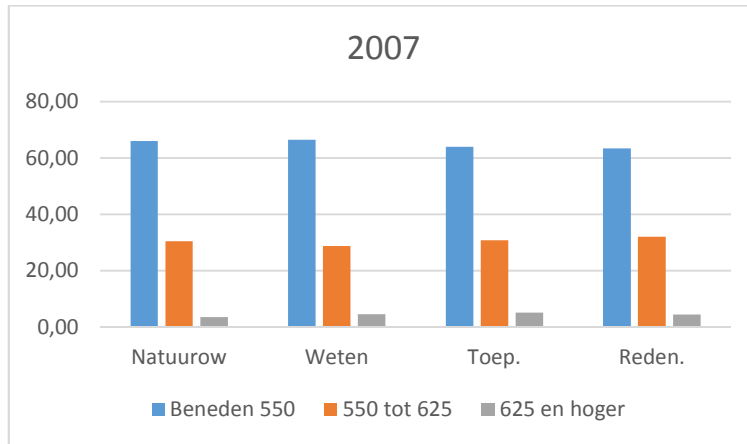
In TIMSS-2015 is de geschiktheid van de TIMSS-toets voor Natuur- en scheikunde en Fysische aardrijkskunde iets groter dan in de jaren daarvoor. Evenals voor de rekendomeinen het geval was, vertaalt de toegenomen geschiktheid zich niet naar hogere percentages leerlingen die op het geavanceerde niveau presteren.

In de drie TIMSS-metingen haalt telkens relatief de grootste groep leerlingen het geavanceerde niveau in Fysische aardrijkskunde. Dit gebied komt echter minder aan bod in groep 6 dan Biologie.

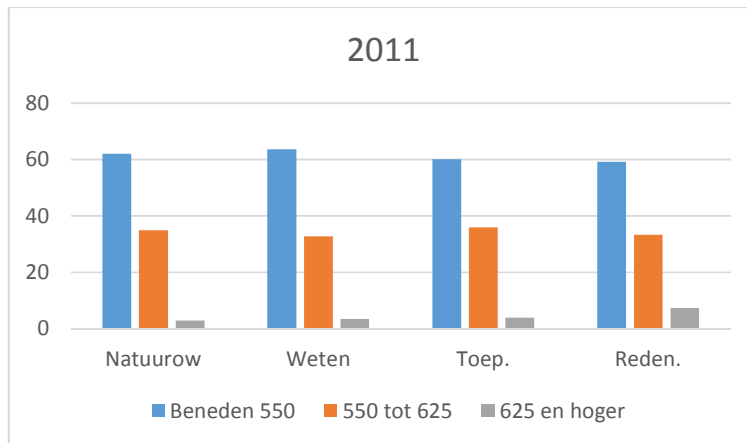
Het domein Natuur- en scheikunde krijgt in alle drie de toetsjaren de minste aandacht in groep 6. Het percentage getoetste onderwerpen binnen dit domein is in Nederland ook aanmerkelijk lager dan internationaal gemiddeld (Meelissen & Punter, 2016). Natuur- en scheikunde wordt ook gekenmerkt door het laagste aandeel excellente leerlingen; slechts 2% tot 3% van de leerlingen heeft hierin een score gehaald van 625 of hoger. Ook het percentage leerlingen dat het hoge niveau haalt voor Natuur- en scheikunde is relatief laag, 22% in 2007, 31% in 2011 en 24% in 2015.

Cognitieve domeinen

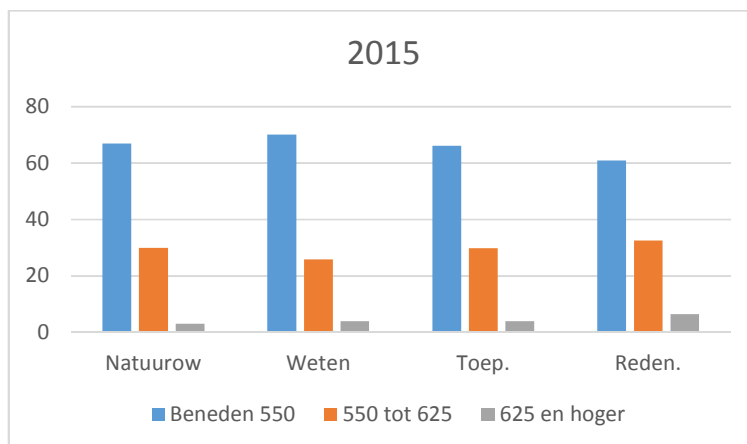
De volgende figuren tonen de verdelingen van leerlingen naar referentieniveaus per cognitief domein en per TIMSS-meting. Tabel 4.8 toont het percentage leerlingen dat voor elk cognitief domein op het geavanceerde niveau (625 en hoger) heeft gepresteerd.



Figuur 4.12: Verdeling van de leerlingen over de verschillende referentieniveaus natuuronderwijs per cognitief domein TIMSS-2007, in percentages



Figuur 4.13: Verdeling van de leerlingen over de verschillende referentieniveaus natuuronderwijs per cognitief domein TIMSS-2011, in percentages



Figuur 4.14: Verdeling van de leerlingen over de verschillende referentieniveaus natuuronderwijs per cognitief domein TIMSS-2015, in percentages

Tabel 4.8

Percentage leerlingen dat het **geavanceerde** referentieniveau haalt voor natuuronderwijs (score 625 en hoger), per cognitief domein, in TIMSS-2007, TIMSS-2011 en TIMSS-2015

	% TIMSS-2007	% TIMSS-2011	% TIMSS-2015
Weten	4,7	3,6	4,0
Toepassen	5,2	4,0	3,9
Redeneren	4,5	7,5	6,4
Totaal natuuronderwijs	3,5	2,9	3,1

Ten opzichte van 2007 haalt een hoger percentage leerlingen in 2011 en 2015 het geavanceerde niveau voor Redeneren. Het aandeel van excellerende leerlingen in overige domeinen blijft in deze periode op een vergelijkbaar niveau.

Aan rekenen wordt in Nederland ongeveer 4,5 keer meer onderwijstijd besteed dan aan natuuronderwijs (Meelissen & Punter, 2016). In de drie TIMSS-metingen haalt echter in beide vakgebieden rond de 4% van de groep 6 leerlingen het geavanceerde niveau in Weten. De percentages leerlingen die sinds 2007 excelleren in de redeneeropgaven voor rekenen (zie tabel 4.4), liggen wel aanmerkelijk hoger dan voor natuuronderwijs.

4.4 Conclusie

Vergeleken met de andere inhouds- en cognitieve rekendomeinen, is het aandeel leerlingen dat op geavanceerd niveau presteert het grootst voor het inhoudsdomen Gegevensweergave en voor het cognitieve rekendomein Redeneren (boven de 10%). Dit geldt voor alle drie TIMSS-metingen. De percentages excellerende leerlingen in het domein Redeneren zijn sinds 2007 zelfs licht gestegen.

Binnen de exacte vakken zijn Geometrische vormen en meten en Natuur- en scheikunde de zwakste inhoudsdomen. De gemiddelde toetsscores zijn lager dan in de andere domeinen en ook relatief laag ten opzichte van andere landen (Martin et al., 2016; Mullis et al., 2016). Uit bovenstaande vergelijking tussen de drie TIMSS-metingen blijkt dat in de afgelopen 8 jaar rond 3% van de Nederlandse groep 6 leerlingen in deze domeinen excelleert.

Het is echter de vraag in hoeverre deze ‘achterstand’ ingehaald moet worden. Beide inhoudsdomen passen ook het minst goed bij het Nederlandse beoogde curriculum. Bovendien blijkt voor de verschillende domeinen dat een toename in geschiktheid van de toets voor het Nederlandse uitgevoerde onderwijs (ofwel er worden meer TIMSS-onderwerpen behandeld in de les) niet gepaard hoeft te gaan met hogere gemiddelde toetsscores of een groter aandeel excellent presterende leerlingen. Bij deze conclusie past dezelfde relativering als bij de conclusie in Paragraaf 3.5; leerkrachten geven in TIMSS alleen aan of een onderdeel al aan bod is geweest. TIMSS-data geven geen inzicht in de hoeveelheid tijd die aan het onderwerp is besteed, hoe diepgaand een leerstofgebied is behandeld en hoe effectief de instructie is geweest. In TIMSS-2015 is bijvoorbeeld ook gebleken dat leerkrachten minder zelfvertrouwen hebben in hun didactische vaardigheden voor natuuronderwijs dan voor rekenen (Meelissen & Punter, 2016). Dit geldt vooral voor instructie aan talentvolle leerlingen. De eerste stap naar meer excellentie in natuuronderwijs zou nader onderzoek kunnen zijn naar de effectiviteit van deze instructie en de mate waarin leerkrachten bijvoorbeeld op de PABO en in de methoden voor Natuur en & Techniek, gestimuleerd dan wel ondersteund worden in het bedienen van talentvolle bèta-leerlingen.

Tot slot blijkt dat voor zowel rekenen als natuuronderwijs er relatief weinig leerlingen zijn (rond de 4%) die op geavanceerd niveau presteren op het cognitieve domein Weten en dat de percentages excellente leerlingen in Toepassen (licht) zijn gedaald. Het is hierbij eveneens de vraag of dit in een andere aanpak moet resulteren is of dat het beter kunnen redeneren dan weten of toepassen, past bij het Nederlandse beoogde curriculum.

5. Gebruikte literatuur

Luyten, H. (2017) Predictive Power of OTL Measures in TIMSS and PISA. In: Jaap Scheerens (Ed.), *Opportunity to Learn, Curriculum Alignment and Test Preparation: A Research Review*. SpringerBriefs in education. Springer, pp. 103 - 119.

Martin, M. O., Mullis, I. V. S., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International Results in Science*. Verkregen via <http://timssandpirls.bc.edu/timss2015/international-results/>.

Meelissen, M. R. M., Netten, A., Punter, R.A., Vermeeren, Ch., Drent, M. & Strating H. (2014). *TIPI-talenten: een studie naar excellente leerlingen in TIMSS en PIRLS. Eindrapportage*. Enschede: University of Twente, Nijmegen: Radboud University

Meelissen, M. R. M. & Punter, R. A. (2016). *Twintig jaar TIMSS. Ontwikkelingen in leerlingprestaties in de exacte vakken in het basisonderwijs 1995-2015*. Enschede: Universiteit Twente.

Mullis, I. V. S. & Martin, M. O. (Eds.) (2013). *TIMSS 2015 Assessment Frameworks*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International Results in Mathematics*. Verkregen via <http://timssandpirls.bc.edu/timss2015/international-results/>.

Von Davier, M., Gonzalez, E., & Mislevy, R. (2009). What are plausible values and why are they useful? *IERI. Monograph series. Issues and methodologies in large-scale assessments*, 2, 9-36.