

Begrijpend lezen van het vierde naar het zesde leerjaar

>> Herhalingsmeting van PIRLS in 2018 vergeleken met PIRLS 2016



**Centrum voor
Onderwijseffectiviteit
en -evaluatie
KU Leuven**

Jonas Dockx
Georges Van Landeghem
Koen Aesaert
Jan Van Damme
Bieke De Fraine

Oktober 2019

Inhoudsopgave

Beleidssamenvatting	6
Inleiding	13
Introductie	13
Wat is begrijpend lezen?	13
Wat is PIRLS?	14
Een herhalingsmeting in 2018?	14
Structuur van dit rapport	15
1 Voorbereiding herhalingsmeting 2018	17
1.1 Organisatie afname en verwerking herhalingsmeting	17
1.2 Ontwikkeling toetsboekjes	18
1.2.1 Validiteit en betrouwbaarheid	18
1.2.2 Selectie van teksten en items	19
1.3 Ander toetsmateriaal	21
2 Steekproef	22
2.1 Steekproeftrekking PIRLS 2016	22
2.2 Deelname steekproef 2018	23
2.2.1 Deelname scholen 2018	24
2.2.2 Deelname leerlingen in 2018	27
2.2.3 Conclusie	29
2.3 Vergelijking steekproef herhalingsmeting PIRLS in 2018 en Peilingsonderzoek lezen 2018	29
2.4 Schatting standaardfout	32
2.4.1 PIRLS 2016	32
2.4.2 Herhalingsmeting PIRLS 2018	33
3 IRT-analyse herhalingsmeting 2018	34
3.1 Item respons theorie	34
3.2 IRT-modellen	34
3.3 Toetsen op één meetschaal plaatsen	36
3.4 De PIRLS-meetschaal	37
3.4.1 Parameters van PIRLS 2016	37
3.4.2 Schattingsprocedure plausible values	42
3.4.3 Resultaten	43
3.5 Meetschaal Peilingsonderzoek	43
3.5.1 Parameters van het Peilingsonderzoek	43
3.5.2 Schattingsprocedure WLE-scores	45
3.5.3 Resultaten	45
3.6 Factorscores attitudes tegenover lezen op PIRLS-meetschaal	46
3.6.1 Parameters van PIRLS 2016	46
3.6.2 Schattingsprocedure WLE-scores	48
3.7 Is er een plafondefect bij de herhalingsmeting in 2018?	48

4	Resultaten herhalingsmeting 2018	66
4.1	Het gemiddelde prestatieniveau van leerlingen	66
4.2	Leeftijdverdeling	68
4.3	De internationale prestatieniveaus	68
4.4	Verschillen tussen leerlinggroepen	70
4.4.1	Geslacht	70
4.4.2	Socio-economische status (aantal boeken thuis)	71
4.4.3	De houding van leerlingen tegenover lezen	73
4.4.4	De zelfzekerheid van leerlingen over hun leesvaardigheid	74
4.5	Eigenschappen van de school en leerkracht	75
4.5.1	Gerichtheid op academisch succes	75
4.5.2	Tevredenheid van leerkrachten	76
4.5.3	Onderwijservaring van leerkrachten	78
4.5.4	Nascholing begrijpend lezen van leerkrachten	78
4.5.5	Gebruikte werkvormen	81
4.6	Conclusie	81
5	Leerwinst tussen vierde en zesde leerjaar	84
5.1	Inleiding	84
5.1.1	Gemiddelde leerwinst	84
5.1.2	Leerwinst volgens geslacht	87
5.1.3	Leerwinst volgens geboortejaar	88
5.1.4	Leerwinst volgens thuistaal	89
5.1.5	Leerwinst volgens socio-economische status	90
5.1.6	Leerwinst naarmate de houding van ouders tegenover lezen	91
5.1.7	Leerwinst naarmate leerlingen betrokken zijn bij lezen	92
5.1.8	Leerwinst naarmate de houding van leerlingen tegenover lezen	93
5.1.9	Leerwinst naarmate leerlingen zelfzeker zijn in hun leesvaardigheid	94
5.2	Conclusie	95
6	Evolutie in attitudes tussen vierde en zesde leerjaar	96
6.1	Inleiding	96
6.2	Evolutie van zelfzekerheid van leerlingen over hun leesvaardigheid tussen 4de en 6de leerjaar	96
6.2.1	Gemiddelde evolutie	97
6.2.2	Evolutie volgens geslacht	97
6.2.3	Evolutie volgens geboortejaar	98
6.2.4	Evolutie volgens thuistaal	99
6.2.5	Evolutie volgens benchmarks 2016	100
6.2.6	Evolutie volgens socio-economische status	101
6.3	Evolutie van betrokkenheid bij lezen tussen 4de en 6de leerjaar	102
6.3.1	Gemiddelde evolutie	102
6.3.2	Evolutie volgens geslacht	103
6.3.3	Evolutie volgens geboortejaar	104
6.3.4	Evolutie volgens thuistaal	105
6.3.5	Evolutie volgens benchmarks 2016	106
6.3.6	Evolutie volgens socio-economische status	107
6.4	Evolutie van houding van leerlingen tegenover lezen tussen 4de en 6de leerjaar	108
6.4.1	Gemiddelde evolutie	108
6.4.2	Evolutie volgens geslacht	109
6.4.3	Evolutie volgens geboortejaar	110
6.4.4	Evolutie volgens thuistaal	111
6.4.5	Evolutie volgens benchmarks 2016	112
6.4.6	Evolutie volgens socio-economische status	113
6.5	Conclusie	114

7	Unidimensionaliteit toetsen herhalingsmeting 2018	115
7.1	Inleiding	
7.2	Unidimensionaliteit of multidimensionaliteit?	115
7.3	Unidimensionaliteit PIRLS 2016 en het Peilingsonderzoek lezen 2018	115
7.4	Gelijkenissen en verschillen tussen PIRLS en peiling volgens Vlaamse experts	116
7.4.1	Lengte en moeilijkheidsgraad van de teksten	117
7.4.2	Tekstsoorten	117
7.4.3	Vraagsoorten	117
7.4.4	Samenhang tussen PIRLS en peiling	118
7.5	Onderzoeksvragen en analyseplan unidimensionaliteit	118
7.5.1	Exploratieve factoranalyses	119
7.5.2	Vergelijking 1-factor, 2-factor en bifactor IRT-modellen	119
7.6	Resultaten	121
7.6.1	Exploratieve factoranalyses	121
7.6.2	Vergelijking 1-factor, 2-factor en bifactor IRT-modellen	124
7.6.3	Predictieve validiteit volgens 2-factor IRT-model	138
7.7	Algemene conclusie	141
7.8	PIRLS 2016 op de meetschaal van het Peilingsonderzoek	142
8	Algemeen besluit	144
8.1	Zesde leerjaar in Vlaanderen op niveau vierde leerjaar in toppresterende landen	144
8.2	Vlaanderen maakt geen inhaalbeweging tussen het vierde en zesde leerjaar, integendeel	144
8.3	Attitudes van leerlingen ten opzichte van lezen worden doorgaans negatiever	145
8.4	PIRLS meet begrijpend lezen op een manier die relevant is voor Vlaanderen	145
	Dankwoord	146
	Referenties	147
	Appendices	149
	Appendix 1: Syntax model IRT met parameters van PIRLS in MIRT	149
	Appendix 2: Syntax model IRT met parameters van het Peilingsonderzoek in MIRT	163
	Appendix 3: Syntax model IRT met parameters van PIRLS in MIRT voor attitudes	176
	Appendix 4: Syntax voor exploratieve factoranalyse in Mplus voor toetsboekje 1	179
	Appendix 5: Syntax modellen IRT voor onderzoek unidimensionaliteit toetsboekje 1 in MIRT	180
	Appendix 6: Syntax modellen IRT voor onderzoek unidimensionaliteit toetsboekje 1 in Mplus	182

Beleidsamenvatting

>> In dit rapport beschrijven we de resultaten van de herhalingsmeting van PIRLS in 2018. Deze herhalingsmeting werd uitgevoerd omwille van de teleurstellende resultaten voor Vlaanderen in PIRLS 2016. PIRLS 2016, een internationaal vergelijkende studie, toonde namelijk dat het niveau voor begrijpend lezen van onze leerlingen in het vierde leerjaar van het basisonderwijs tussen 2006 en 2016 opmerkelijk gedaald is. Het doel van de herhalingsmeting is dus om te onderzoeken welk niveau deze leerlingen twee jaar later in het zesde leerjaar bereiken. Deze samenvatting geeft een toegankelijk overzicht van de onderzoeksvragen, voornaamste resultaten en conclusies van dit rapport.

Van de 141 lagere scholen die deelnamen aan PIRLS 2016 waren er 126 lagere scholen die vrijwillig terug deelnamen. Deze 126 scholen hadden 4679 leerlingen die in het vierde leerjaar deelnamen aan PIRLS 2016, en 4615 leerlingen die in het zesde leerjaar deelnamen aan de herhalingsmeting in 2018. 4046 van deze leerlingen namen zowel deel aan PIRLS 2016 als de herhalingsmeting in 2018. De verschillende leerlingenaantallen worden veroorzaakt door veranderingen in de leerlingenpopulatie tussen het vierde en zesde leerjaar. Zo zijn er leerlingen die blijven zitten, van school veranderen, naar het buitengewoon onderwijs gaan of naar het secundair onderwijs gaan. Deze veranderingen worden in de tekst uitvoerig besproken, omdat deze informatie noodzakelijk is om de resultaten correct te interpreteren.

Het rapport bestaat uit zeven hoofdstukken, waarvan drie technische en vier inhoudelijke hoofdstukken. Ieder van de vier inhoudelijke hoofdstukken behandelt één van de algemene onderzoeksvragen:

- Welk niveau van begrijpend lezen hebben Vlaamse leerlingen op het einde van het zesde leerjaar? Dit werd onderzocht in [Hoofdstuk 4](#).
- Hoe groot is de leerwinst van Vlaamse leerlingen tussen het vierde en zesde leerjaar? Dit werd onderzocht in [Hoofdstuk 5](#).
- Hoe evolueren Vlaamse leerlingen in hun attitudes over lezen tussen het vierde en zesde leerjaar? Dit werd onderzocht in [Hoofdstuk 6](#).
- Meten de toetsen van PIRLS en de toetsen van het Peilingsonderzoek begrijpend lezen dezelfde onderliggende vaardigheid? Dit werd onderzocht in [Hoofdstuk 7](#).

Voorafgaand aan deze vier inhoudelijke hoofdstukken zijn er drie hoofdstukken die van technische aard zijn. [Hoofdstuk 1](#) beschrijft de praktische organisatie van de herhalingsmeting, waaronder de samenstelling van de toetsen. [Hoofdstuk 2](#) beschrijft de steekproeftrekking. De 126 lagere scholen die vrijwillig deelnamen aan de herhalingsmeting in 2018 verschillen niet van de 141 lagere scholen die deelnamen aan PIRLS 2016 voor prestaties of SES. De 126 scholen mogen dan ook als een afspiegeling van alle Vlaamse lagere scholen beschouwd worden. [Hoofdstuk 3](#) beschrijft de uitvoering van de IRT-analyses. Dit zijn de statistische analyses om de toetsgegevens op de meetschaal van PIRLS 2016 en de meetschaal van het Peilingsonderzoek te plaatsen.

In wat volgt geven we een samenvatting van de voornaamste resultaten en conclusies voor ieder van de vier algemene onderzoeksvragen. We sluiten af met een oproep om het onderwijs in begrijpend lezen te versterken.

Niveau voor begrijpend lezen in het zesde leerjaar

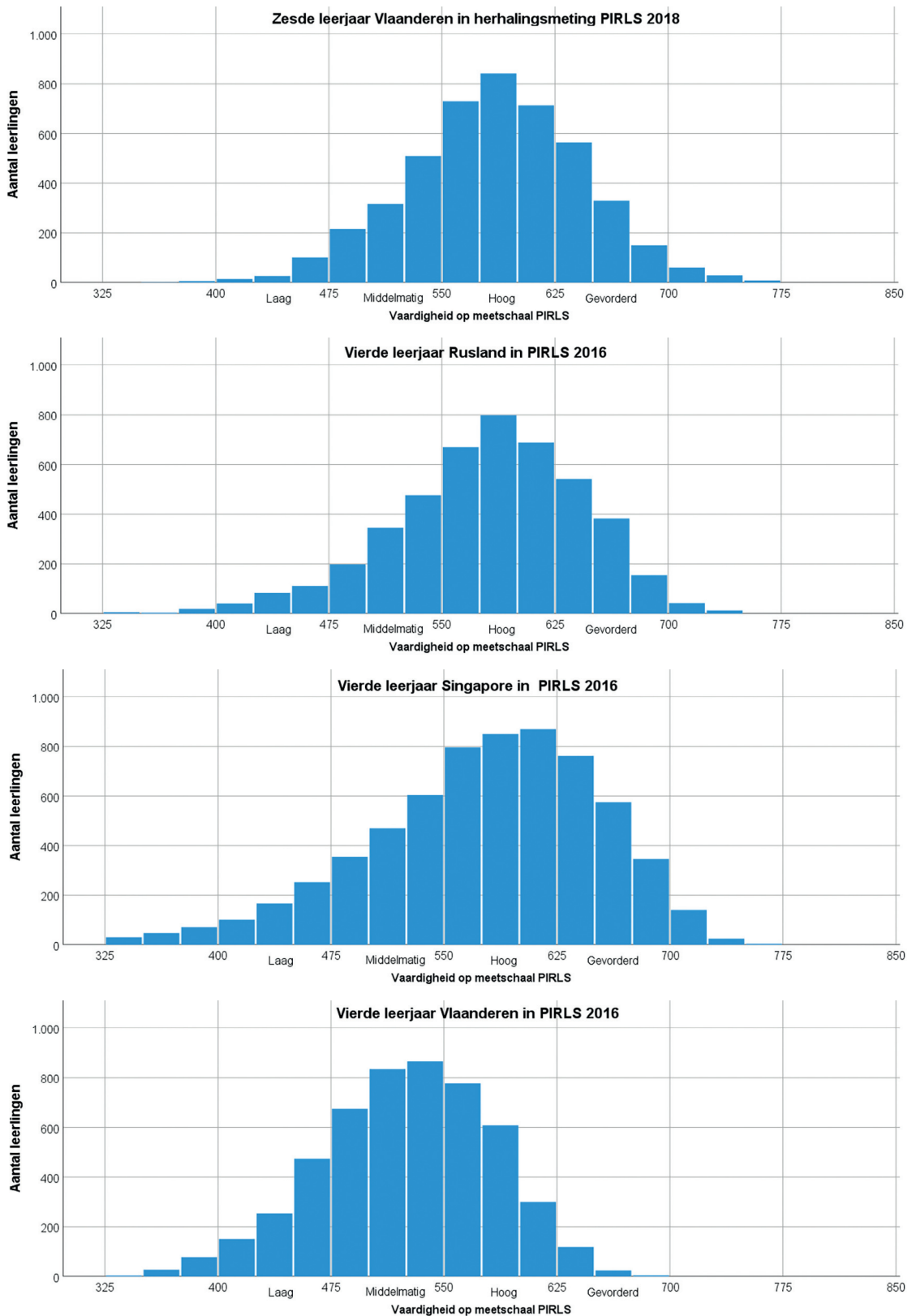
We onderzochten welk niveau Vlaamse leerlingen behalen voor begrijpend lezen op het einde van het zesde leerjaar. Hiervoor vergeleken we hun gemiddelde niveau met de leerlingen in het vierde leerjaar van de landen die deelnamen in PIRLS 2016. We bekeken ook hoe de spreiding van vaardigheid in begrijpend lezen binnen Vlaanderen was. Tenslotte onderzochten we hoe leerlingen naargelang geslacht of andere achtergrondkenmerken verschillen in het zesde leerjaar.

De steekproef om deze onderzoeksvraag te beantwoorden bestond uit 4615 leerlingen in 126 scholen. Dit waren leerlingen die in het zesde leerjaar zaten en op een correcte manier hadden deelgenomen aan de toets begrijpend lezen van de herhalingsmeting. Leerlingen die nieuw waren ingestroomd tussen het vierde en zesde leerjaar in één van de 126 scholen werden opgenomen in de analyses.

We stelden vast dat de Vlaamse leerlingen in het zesde leerjaar in deze steekproef een hoger niveau bereikten dan de leerlingen in het vierde leerjaar van alle andere landen en regio's. De Vlaamse leerlingen in het zesde leerjaar scoren echter niet merkbaar beter dan de hoogst presterende landen in het vierde leerjaar, zoals blijkt uit [Figuur 1](#). Vlaanderen doet er dus twee jaar langer over om het niveau van de sterk presterende landen in PIRLS te bereiken. Verder bereiken 24% van de Vlaamse leerlingen het gevorderd niveau (*advanced benchmark*) in het zesde leerjaar, terwijl dit maar 4% was in het vierde leerjaar. Echter, het relatieve aantal gevorderde leerlingen in het Vlaamse zesde leerjaar is daarmee lager dan in het vierde leerjaar in Singapore (29%) en Rusland (26%).

Wat betreft verschillen tussen leerlingengroepen vonden we dat meisjes hoger scoren dan jongens voor begrijpend lezen, zowel in het vierde als het zesde leerjaar. Bovendien is het verschil tussen jongens en meisjes groter in het zesde leerjaar dan in het vierde leerjaar. We stellen ook vast dat leerlingen met een hogere SES gemiddeld beter scoren; het verschil in prestaties naargelang SES was wel enigszins afgenomen tussen 2016 en 2018.

Figuur 1: vergelijking histogrammen van vaardigheden het 6de leerjaar Vlaanderen met het 4de leerjaar twee topresterende landen



Leerwinst tussen vierde en zesde leerjaar

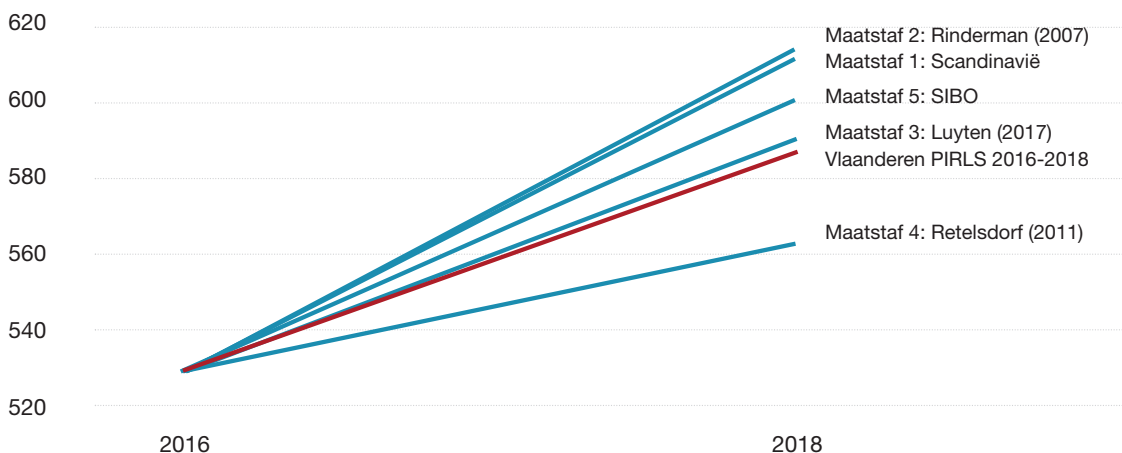
We onderzochten welke leerwinst Vlaamse leerlingen maken in begrijpend lezen tussen het einde van het vierde leerjaar en het einde van het zesde leerjaar. Hiervoor gebruikten we vijf maatstaven die beschrijven hoeveel leerwinst we verwachten van de Vlaamse leerlingen tussen het vierde en het zesde leerjaar. We gingen na in welke mate Vlaamse leerlingen aan deze maatstaven voldoen. Zo konden we ook een uitspraak doen of er sprake is van een ‘inhaalbeweging’ voor begrijpend lezen in de derde graad.

De steekproef om deze onderzoeksvraag te beantwoorden bestond uit 4046 leerlingen in 126 scholen. Dit waren leerlingen die het zesde leerjaar zaten en op een correcte manier hadden deelgenomen aan zowel de toets begrijpend lezen van de herhalingsmeting in 2018 als aan de toets begrijpend lezen van PIRLS 2016. Leerlingen die in het vijfde leerjaar zaten werden dus niet opgenomen in deze analyses. Leerlingen die nieuw waren ingestroomd tussen het vierde en zesde leerjaar in één van de 126 scholen werden ook niet opgenomen in de analyses.

De vijf maatstaven om te evalueren of de Vlaamse leerlingen veel of weinig leerwinst maken voor begrijpend lezen zijn gebaseerd op voorgaand onderzoek over leerwinst in begrijpend lezen. De voorspellingen volgens deze maatstaven en wat we empirisch vaststelden tussen 2016 en 2018 wordt weergegeven in **Figuur 2**. De eerste twee maatstaven zijn aan de hand van de leerwinst op de meetschaal van internationaal onderzoek. Zo werd er in drie Scandinavische landen een gemiddelde jaarlijkse leerwinst vastgesteld van 41.26 punten in PIRLS. Rindermann had een gelijkaardig conclusie en vond dat één leerjaar onderwijs voor 42 punten leerwinst zorgt. Vlaamse leerlingen maken in de herhalingsmeting slechts 28.05 punten leerwinst op jaarbasis, wat minder is dan deze twee maatstaven. De volgende twee maatstaven zijn aan de hand van de leerwinst uitgedrukt in het aantal standaarddeviaties. Zo voorspellen Luyten, Merell en Tymms (2017) dat Engelse leerlingen op twee jaar tijd 1.03 standaarddeviaties vooruitgaan. Retelsdorf en collega's (2012; 2011) verwachten dan weer dat Duitse leerlingen 0.60 standaarddeviaties vooruitgaan op twee jaar tijd. Vlaanderen bevindt zich met 0.97 standaarddeviaties tussen de beide maatstaven. Ten slotte werd in het SIBO-onderzoek in Vlaanderen tussen het vierde en zesde leerjaar een vooruitgang van 1.26 standaarddeviaties vastgesteld tussen 2007 en 2009, wat meer is dan we nu vaststellen in 2016 en 2018. Voor vier van de vijf maatstaven maken de Vlaamse leerlingen dus minder leerwinst dan verwacht. Er is geen evidentie dat Vlaamse leerlingen een inhaalbeweging maken voor begrijpend lezen tussen het vierde en zesde leerjaar.

Wat betreft verschillen tussen leerlingengroepen vonden we dat meisjes meer leerwinst maken dan jongens. Er is geen relatie tussen SES en de gemaakte leerwinst, terwijl leerlingen die thuis nooit of soms Nederlands spreken wel meer leerwinst maken dan leerlingen die thuis altijd of meestal Nederlands spreken.

Figuur 2: de leerwinst in begrijpend lezen tussen het vierde en zesde leerjaar zoals ze vastgesteld is in PIRLS tussen 2016 en 2018, en zoals ze voorspeld werd door de verschillende maatstaven. De voorspelde curves gebruiken steeds het vastgestelde niveau (gemiddelde en standaarddeviatie) in 2016 als startpunt.



Evolutie attitudes over lezen tussen vierde en zesde leerjaar

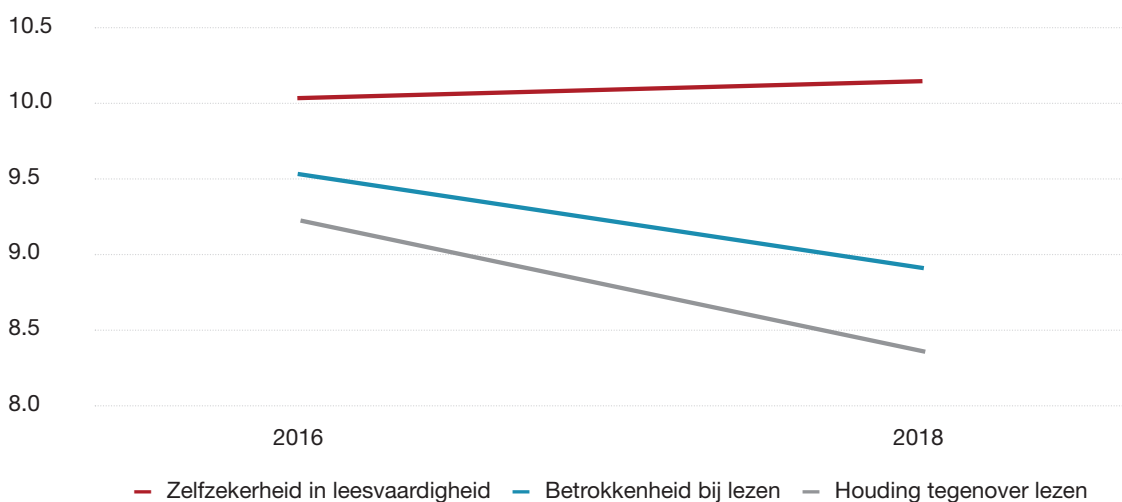
We onderzochten welke evolutie de Vlaamse leerlingen maken voor hun attitudes over lezen tussen het einde van het vierde leerjaar en het einde van het zesde leerjaar. De drie attitudes die onderzocht worden zijn (a) de leerlingen hun zelfzekerheid over hun leesvaardigheid, (b) betrokkenheid bij lezen en (c) de houding van leerlingen tegenover lezen. Om deze resultaten te duiden bekeken we naar hoe de attitudes in het zesde leerjaar internationaal te vergelijken zijn met deze in het vierde leerjaar. Verder bekeken we ook wat de internationale literatuur ons vertelt over de evolutie in deze attitudes. Tenslotte onderzochten we hoe leerlingen naargelang geslacht of andere achtergrondkenmerken verschillen in hun evolutie in attitudes over lezen.

De steekproef om deze onderzoeksvraag te beantwoorden bestond uit 3794 leerlingen in 126 scholen. Dit waren leerlingen die het zesde leerjaar zaten, op een correcte manier hadden deelgenomen aan de toets begrijpend lezen en de leerlingvragenlijst invulden tijdens herhalingsmeting en tijdens PIRLS 2016. Leerlingen die in het vijfde leerjaar zaten werden dus niet opgenomen in deze analyses. Leerlingen die nieuw waren ingestroomd tussen het vierde en zesde leerjaar in één van de 126 scholen werden ook niet opgenomen in de analyses.

De resultaten duiden erop dat de Vlaamse leerlingen al in PIRLS 2016 in het vierde leerjaar relatief negatieve attitudes hadden over lezen wanneer deze vergeleken worden met andere landen. Daarbij zien we nu, zoals weergegeven in **Figuur 3**, dat in het zesde leerjaar de Vlaamse leerlingen zich nog minder betrokken voelen bij lezen en een minder positieve houding hebben. Enkel zelfzekerheid blijft stabiel.

Er is geen noemenswaardig verschil tussen jongens en meisjes voor zelfzekerheid en betrokkenheid in het vierde leerjaar en ook niet in het zesde leerjaar. Voor de houding tegenover lezen zien we dat jongens een grotere daling kennen dan meisjes, terwijl de jongens initieel ook al een minder positieve houding hadden voor lezen. Verder vinden we dat leerlingen die thuis altijd Nederlands spreken meer zelfzeker zijn en een meer positieve houding hebben tegenover lezen dan leerlingen die thuis nooit Nederlands spreken. Beide groepen kennen dezelfde evolutie voor zelfzekerheid, de verschillen blijven dus. Het verschil tussen beide groepen leerlingen in de houding tegenover lezen verdwijnt echter. Leerlingen met hogere prestaties voor begrijpend lezen in 2016 waren meer zelfzeker over hun leesvaardigheid, waren meer betrokken bij lezen en hadden een positievere houding tegenover lezen in 2016. Deze verschillen blijven eerder stabiel tussen het vierde en zesde leerjaar voor zelfzekerheid en betrokkenheid. We vinden dat leerlingen met een hogere SES in 2016 meer zelfzeker waren in lezen, meer betrokken waren lezen en een positievere houding hadden tegenover lezen. Deze verschillen blijven eerder stabiel tussen het vierde en zesde leerjaar.

Figuur 3: de evoluties in zelfzekerheid in leesvaardigheid, betrokkenheid bij lezen en houding tegenover lezen tussen het vierde en zesde leerjaar zoals vastgesteld in PIRLS tussen 2016 en 2018



Meten PIRLS en het Peilingsonderzoek dezelfde vaardigheid voor begrijpend lezen?

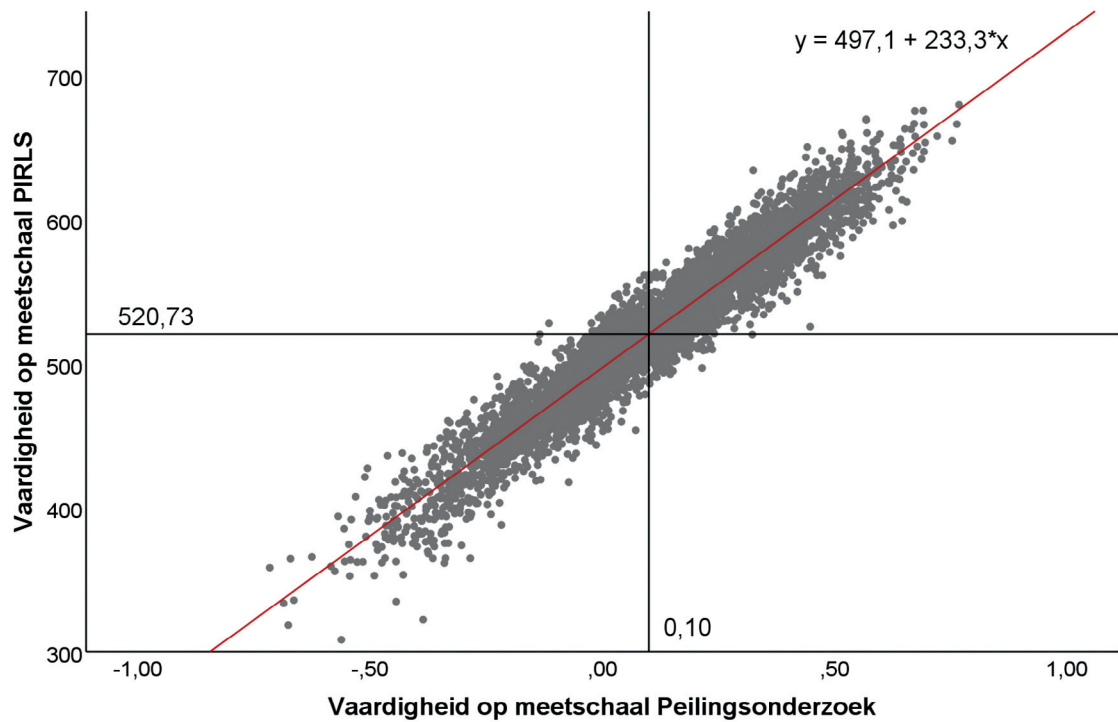
We onderzochten in hoeverre de toetsen van PIRLS (internationaal) en het Peilingsonderzoek (Vlaams) dezelfde onderliggende vaardigheid meten voor begrijpend lezen. Het Vlaamse peilingsonderzoek vertrekt bij de toetsontwikkeling van de Vlaamse eindtermen voor het einde van de lagere school. Het internationale PIRLS-onderzoek vertrekt bij de toetsontwikkeling van een algemene, internationaal overeengekomen definitie van begrijpend lezen. Indien het mogelijk is om de toetsen van het Peilingsonderzoek en PIRLS op één meetschaal te plaatsen, dan kunnen de resultaten van beide studies vergeleken worden op eenzelfde schaal. Dit laat ons toe de vraag te beantwoorden hoe de Vlaamse eindtermen zich verhouden tot de internationale PIRLS-schaal. Tevens kan de afname van PIRLS 2016 bij leerlingen in het vierde leerjaar dan gebruikt worden om te bepalen hoeveel leerlingen in het vierde leerjaar reeds de eindtermen voor begrijpend lezen halen.

De steekproef om deze onderzoeksvraag te beantwoorden bestond uit 4615 leerlingen in 126 scholen. Dit waren leerlingen die het zesde leerjaar zaten en op een correcte manier hadden deelgenomen aan de toets begrijpend lezen van de herhalingsmeting. Leerlingen die in het vijfde leerjaar zaten werden dus niet opgenomen in deze analyses. Leerlingen die nieuw waren ingestroomd tussen het vierde en zesde leerjaar in één van de 126 scholen werden wel opgenomen in de analyses. We maakten in de analyses soms ook gebruik van de oorspronkelijke gegevens van PIRLS 2016.

De resultaten duiden erop dat de samenhang tussen de verschillende toetsvragen voornamelijk naar één onderliggende vaardigheid terug te brengen is. Hoewel de ontwikkeling van de toetsen begrijpend lezen in PIRLS en in het Vlaamse peilingsonderzoek dus volledig los van elkaar gebeurde, bleek toch dat beide soorten items op één onderliggende meetschaal passen. Dit wil zeggen dat het onderscheiden van een aparte vaardigheid voor de items van het Peilingsonderzoek en de items van PIRLS weinig bijdraagt. Een model met één onderliggende vaardigheid verklaart namelijk 26.03% van de variantie in de itemantwoorden, terwijl een model met twee onderliggende vaardigheden 29.70% verklaart van de variantie in de itemantwoorden. Een model met één onderliggende vaardigheid verklaart dus 87.68% van de variantie dat een model met twee aparte vaardigheden verklaart. We besluiten daardoor dat de toetsen van het Peilingsonderzoek lezen en PIRLS hoofdzakelijk dezelfde vaardigheid in begrijpend lezen meten.

Dat beide onderzoeken dezelfde vaardigheid meten laat ons toe, dankzij de gegevens van de herhalingsmeting, te berekenen waar de cesuur van het Peilingsonderzoek lezen zich op de meetschaal van PIRLS bevindt. We vinden dat de cesuur van het Peilingsonderzoek zich op waarde 520.3 van de meetschaal van PIRLS bevindt. Van de 5198 leerlingen die deelnamen aan PIRLS 2016 waren er 53.33% die de cesuur in het vierde leerjaar bereikt hadden, zoals weergegeven in [Figuur 4](#).

Figuur 4: Scatterplot geschatte vaardigheden van leerlingen volgens meetschaal Peilingsonderzoek en meetschaal PIRLS. De groep leerlingen bestaat hier uit de 5198 leerlingen die hadden deelgenomen aan PIRLS 2016. Ook de cesuur van het peilingsonderzoek op beide meetschalen worden door een verticale lijn (0,10) en horizontale lijn (520,73) weergegeven.



Alle hens aan dek voor onderwijskwaliteit

Velen had gehoopt dat de herhaling van het PIRLS-onderzoek in het zesde leerjaar zou aantonen dat onze Vlaamse leerlingen in de derde graad een inhaalbeweging maken qua begrijpend lezen. Maar dat blijkt niet het geval. Er wordt wel leerwinst geboekt in die twee jaren, maar die **voortgang is onvoldoende om bestempeld te worden als 'inhaalbeweging'**.

Het Vlaamse gemiddelde voor PIRLS 2018 bedraagt 586. Dat is hoger dan de scores van de sterk presterende landen in PIRLS 2016 (Rusland: 581 en Singapore: 576). Maar de leerlingen in Rusland en Singapore bereiken dit niveau in het vierde leerjaar en de Vlaamse leerlingen pas in het zesde leerjaar. **Vlaanderen doet er dus twee jaar langer over om het niveau van de sterk presterende landen in PIRLS te bereiken.**

Er klinkt in Vlaanderen vaak kritiek op internationaal vergelijkend onderzoek. We zijn het volmondig eens met de stelling dat die internationale studies niet volstaan om het Vlaamse onderwijs te monitoren. Er is daarnaast nood aan onderzoek en monitoring specifiek voor de Vlaamse context. Maar deze studie toont aan dat de Vlaamse peilingsvragen en de internationale PIRLS-vragen zeer sterk samenhangen. Die samenhang betekent dat het internationale PIRLS-onderzoek relevant is voor het Vlaamse onderwijs. Hoewel PIRLS niet specifiek gericht is op de Vlaamse eindtermen, geeft het dus toch informatie over de vaardigheid van de Vlaamse leerlingen in begrijpend lezen. Anders gezegd: **PIRLS meet begrijpend lezen op een manier die relevant is voor Vlaanderen.**

De conclusie luidt: het Vlaamse lager onderwijs scoort ondermaats voor begrijpend lezen.

Dit onderzoek past in de reeks studies die soms benoemd worden als 'de knipperlichten over de dalende onderwijskwaliteit'. Maar we hopen dat die vele 'knipperlichten' inmiddels een luid signaal vormen, want we merken dat het urgentiebesef bij sommige onderwijsactoren ontbreekt. We roepen alle partners in het onderwijsveld op om het dalende onderwijspeil terug op te krikken. Het ernstig nemen van de PIRLS-resultaten staat niet gelijk aan het reduceren van onderwijskwaliteit tot meetbare leerlinguitkomsten. Wie aandacht heeft voor de PIRLS-resultaten is niet automatisch gekant tégen de brede vorming van jongeren. Maar het is in het belang van onze leerlingen en hun toekomst dat we de PIRLS-resultaten serieus moeten nemen.

De onderzoekers van PIRLS 2016 en PIRLS 2018 hebben drie uitvoerige rapporten geschreven (Tielemans et al., 2017; Tielemans et al., 2019 en het hier voorliggende rapport). Bovendien zijn de resultaten toegelicht en bediscussieerd op vele studiedagen en seminars. We denken dat we daarmee tegemoetkomen aan het advies van de Vlor dat luidt: "Internationaal onderzoek moet steeds goed gekaderd worden binnen het Vlaams onderwijs" (Vlor, 2018).

We willen daarom nogmaals oproepen tot debat, uitwisseling en reflectie over het verbeteren van het niveau van het begrijpend lezen bij onze leerlingen. We willen daar als onderzoekers absoluut toe bijdragen. We pleiten voor een onderbouwd debat, met aandacht voor de complexiteit van oorzaken en dus ook van mogelijke oplossingen. We passen voor een gepolariseerd debat waarin het harde werk van leraren en schoolteams onderuit wordt gehaald. De aantrekkelijkheid van het lerarenberoep is precies dat je als leraar het verschil kan maken in het leren en het leven van je leerlingen.

Vlor. (2018). Onderwijskwaliteit in breed perspectief. Advies over versterken van interne kwaliteitszorg en leerlingenevaluatie (AR-AR-ADV-1819-009). Brussel: VLOR. Retrieved from <https://www.vlor.be/adviezen/onderwijskwaliteit-breed-perspectief>.

Inleiding

Introductie

>> In dit rapport worden de resultaten besproken van de Vlaamse herhalingsmeting van de *Progress in International Reading Literacy Study 2016* (PIRLS 2016) in 2018. PIRLS 2016 was een internationaal vergelijkende studie naar het onderwijs inzake begrijpend lezen met als initiatiefnemer de *International Association for the Evaluation of Educational Achievement* (Mullis & Martin, 2015). De herhalingsmeting van dit onderzoek in 2018 vond enkel plaats binnen Vlaanderen en gebeurde op initiatief van de Vlaamse regering.

Het initiatief voor de herhalingsmeting van PIRLS in 2018 werd genomen op basis van de bezorgdheid in Vlaanderen om het niveau van begrijpend lezen. De resultaten van PIRLS 2016 toonden namelijk dat Vlaanderen sterk gedaald is voor het niveau van begrijpend lezen bij leerlingen in het vierde leerjaar (Mullis, Martin, Foy, & Hooper, 2017; Tielemans, Vandebroeck, Bellens, Van Damme, & De Fraine, 2017; Tielemans, Vanlaar, Van Damme, & De Fraine, 2019). Deze sterke achteruitgang tussen 2006 en 2016 was enigszins onverwacht. Zo gaven de resultaten van de Peilingsonderzoeken naar lezen in 2002, 2007 en 2013 telkens aan dat meer dan 88% van de Vlaamse leerlingen de eindtermen behaalde. Bij de meting van het peilingsonderzoek in 2018 werd wel een daling vastgesteld tot 84%.

De herhalingsmeting van PIRLS in 2018 heeft als doel bij te dragen tot een beter inzicht in het niveau begrijpend lezen in het Vlaamse lager onderwijs. Het is een aanvulling op de nieuwe peiling Nederlands (lezen, luisteren en schrijven) die in mei 2018 bij een representatieve steekproef van leerlingen uit het zesde leerjaar werd afgenomen. Deze herhalingsmeting is echter ook gericht op het beantwoorden van onderzoeksvragen waarover voorgaande onderzoeken geen informatie bieden.

In deze inleiding lichten we toe hoe begrijpend lezen binnen deze studie gedefinieerd is, wat het PIRLS-onderzoek is, en wat de onderzoeksvragen zijn van de herhalingsmeting in 2018.

Wat is begrijpend lezen?

Vooraleer we in dit rapport de resultaten van de herhalingsmeting van PIRLS in 2018 toelichten moeten we weten hoe begrijpend lezen binnen PIRLS gedefinieerd wordt. PIRLS beschrijft begrijpend lezen als een ingewikkeld samenspel van uiteenlopende deelvaardigheden (bijv. fonologische vaardigheden en woordherkenning) en kennisbronnen (bijv. wereldkennis en metacognitieve kennis), een complexe vaardigheid met een grote maatschappelijke relevantie (Stanovich, 1980). PIRLS 2016 hanteert hiervoor expliciet de volgende definitie: "begrijpend lezen is de vaardigheid om geschreven taal te begrijpen en te gebruiken, zoals vereist door de maatschappij en/of het individu. Lezers kunnen betekenis construeren uit teksten van verschillende vormen. Ze lezen om te leren, om deel te nemen aan leesgroepen op school en in het alledaagse leven en voor het plezier" (Mullis & Martin, 2015, p. 12).

Een definitie is echter niet voldoende, want op basis van deze definitie kan begrijpend lezen nog steeds op verschillende wijzen geoperationaliseerd worden. Begrijpend lezen wordt in PIRLS 2016 geoperationaliseerd in termen van leesdoelen en leesprocessen (Mullis & Martin, 2015). De twee leesdoelen, met name leeservaring en informatieverwerving, geven aan waarom leerlingen lezen. Indien leerlingen tekstkenmerken (bijv. thema, personages en genre) verkennen of lezen voor hun plezier, dan staat het doel van leeservaring op de voorgrond. Wanneer leerlingen echter informatie uit een tekst moeten halen om deze nadien actief te gebruiken, dan is informatieverwerving het centrale doel. Naast de twee leesdoelen kunnen er vier leesprocessen onderscheiden worden. Deze processen verwijzen naar hoe leerlingen betekenis construeren. De vier leesprocessen zijn: (1) focus leggen op en vinden van expliciet vermelde informatie, (2) eenvoudige conclusies trekken, (3) tekstinterpretatie en (4) tekstevaluatie.

Bij de herhalingsmeting van PIRLS in 2018 hanteren we dezelfde definitie als PIRLS 2016. De herhalingsmeting van PIRLS in 2018 heeft namelijk als doel om de evolutie in de vaardigheid tot begrijpend lezen tussen 2016 en 2018 te beschrijven. Daarom moeten de toetsen van de herhalingsmeting van PIRLS in 2018 op dezelfde meetschaal liggen als PIRLS 2016. De definitie en operationalisering moeten dus gelijk zijn.

We merken tenslotte op dat de PIRLS-teksten en toetsvragen opgesteld worden voor internationaal vergelijkend onderzoek. Ze zijn niet specifiek gericht op het Vlaamse curriculum of op een ander curriculum, maar zijn in zekere mate curriculumafhankelijk. De PIRLS-vragen zijn dus niet gekoppeld aan de Vlaamse eindtermen (Oliveri, Rutkowski, & Rutkowski, 2018; Tielemans et al., 2019; Vlor, 2018).

Wat is PIRLS?

PIRLS is een studie die de gemiddelde prestaties van leerlingen voor begrijpend lezen in kaart wil brengen (Mullis & Martin, 2015). De nadruk ligt hierbij op een vergelijking tussen verschillende onderwijssystemen. Dit impliceert dat in principe enkel uitspraken kunnen gedaan worden over de onderwijsprestaties van de groep leerlingen van één of meer landen en niet over de scores van individuele leerlingen.

PIRLS wordt sinds 2001, als follow-up van de *Reading Literacy Study* (RLS), om de vijf jaar georganiseerd en gecoördineerd door de IEA. De IEA is een toonaangevende organisatie in de coördinatie en uitbouw van grootschalige internationale dataverzamelingen, waaronder ook de *Trends in International Mathematics and Science Study* (TIMSS), een studie die de wiskunde- en wetenschapsprestaties van leerlingen in kaart brengt.

Het beschrijven en verklaren van verschillen in gemiddelde prestaties in begrijpend lezen tussen landen vereist echter een instrumentarium dat in verschillende landen kan worden afgenomen. Om meer inzicht te krijgen in het begrijpend leesonderwijs, werd bij leerlingen zowel een toets begrijpend lezen als een vragenlijst die peilt naar belangrijke achtergrondkenmerken afgenomen (Mullis & Martin, 2015). Hun ouders, leerkracht en directie vulden eveneens een vragenlijst in om relevante thuis-, klas- en schoolkenmerken die bijdragen aan de onderwijsprestaties in kaart te brengen. Nationale curricula werden geïnventariseerd via een curriculumvragenlijst om inzicht te verwerven in het leerplan van elk deelnemend land. Deze vragenlijst werd in 2016 ingevuld door een onderwijsinspecteur en vertegenwoordigers van de drie grote onderwijsnetten in Vlaanderen (GO!, OVSG en Katholiek Onderwijs Vlaanderen). Een synthese van de door hen aangeleverde informatie (en soms een compromis tussen de diverse informaties) vormde de basis voor de finale curriculuminformatie over het Vlaamse onderwijs. Deze informatie wordt voor alle deelnemende landen gepresenteerd in de PIRLS 2016 Encyclopedie (Mullis, Martin, Goh, & Prendergast, 2017).

De IEA zocht een moment tijdens de schoolloopbanen van leerlingen waarop het optimaal was begrijpend lezen te toetsen. Voor de IEA is dit het moment waarop leerlingen vier jaren formele scholing hebben gehad, wat in Vlaanderen overeenkomt met het vierde leerjaar lager onderwijs. Dit afnamemoment is niet willekeurig gekozen. Het vierde leerjaar vormt immers een belangrijk overgangsmoment waarbij de nadruk verschoven is van technisch naar begrijpend lezen (Mullis & Martin, 2015). Anderzijds is dit nog voor de start in het secundair onderwijs in de deelnemende landen. De keuze voor een bepaald leerjaar betekent wel dat sommige van de deelnemende leerlingen één of zelfs twee jaar vertraging kunnen hebben, terwijl andere leerlingen kunnen versneld zijn. De aanpak van PIRLS en TIMSS is op dit punt anders dan bij het *Programme for International Student Assessment* (PISA), waarbij wordt gekozen voor een steekproef van 15-jarige leerlingen, ongeacht het leerjaar waarin zij zitten.

Met studies als TIMSS en PIRLS wil de IEA één belangrijk doel nastreven: het wereldwijd verbeteren van de kwaliteit van het onderwijs (IEA, 2019). Door het bevorderen van internationale samenwerking en overleg over de leerprestaties voor begrijpend lezen, voorzien zij de beleidsmakers van referentiecriteriën en van feedback over hoe leerlingen in hun onderwijssysteem presteren. Op deze manier vormt de IEA een bron van inspiratie voor curricula en onderwijs-hervormingen.

Een herhalingsmeting in 2018?

De Vlaamse resultaten van PIRLS 2016 duiden erop dat tussen 2006 en 2016 een sterke achteruitgang in het niveau van begrijpend lezen heeft plaatsgevonden (Mullis, Martin, Goh, et al., 2017; Tielemans et al., 2017, 2019). Vlaanderen was de grootste daler in de internationale vergelijking tussen 2006 en 2016 over alle landen. Omdat PIRLS 2016 bij leerlingen van het vierde jaar werd afgenomen, werd de vraag gesteld hoe tijdens het vijfde en zesde leerjaar de prestaties voor begrijpend lezen verder evolueren. Daarom besloot de Vlaamse regering om een herhalingsmeting te organiseren van de PIRLS-toetsen bij de zesdejaars in het 2018. Het Centrum voor Onderwijseffectiviteit en -evaluatie voerde deze herhalingsmeting uit.

De algemene onderzoeksvraag die de herhalingsmeting van PIRLS in 2018 wil beantwoorden is wat het leesniveau is van Vlaamse leerlingen op het einde van het zesde leerjaar. Deze algemene onderzoeksvraag bestaat uit verschillende deelvragen. Om het bereikte leesniveau op het einde van het zesde leerjaar te kaderen is immers ook kennis nodig over: de verschillen tussen verschillende leerlingengroepen, de evolutie van het leesniveau doorheen de tijd, wat deze evolutie mee bepaalt, en hoe begrijpend lezen geoperationaliseerd kan worden.

De volgende vier onderzoeksvragen worden daarom in het rapport van deze herhalingsmeting beantwoord:

1. Wat is het niveau van begrijpend lezen van de Vlaamse populatie van leerlingen op het einde van het zesde leerjaar wanneer deze internationaal vergeleken wordt? Deze vraag kan begrepen worden als een vraag naar de gemiddelde score van Vlaamse leerlingen. Deze vraag heeft echter ook betrekking op de spreiding in resultaten en hoeveel leerlingen in het zesde leerjaar zich in elk van de internationale standaarden bevinden. Verder wordt deze onderzoeksvraag zowel in het algemeen gesteld als voor specifieke deelgroepen. Zo zal worden onderzocht of dit niveau verschilt tussen bepaalde groepen leerlingen. Zo zal er onder andere vergeleken worden tussen jongens en meisjes, naargelang sociaaleconomische status en volgens andere achtergrondkenmerken. Aansluitend wordt ook bekeken welke verschuivingen in prestatieniveaus er zijn tussen 2016 en 2018.
2. Hoe groot is de leerwinst in begrijpend lezen tussen het einde van het vierde leerjaar en het einde van het zesde leerjaar? Leerwinst betekent hier het verschil in vaardigheid tussen het vierde en zesde leerjaar voor dezelfde groep leerlingen. Deze onderzoeksvraag wordt in het algemeen gesteld, maar ook voor specifieke deelgroepen. Zo zal worden onderzocht of de leerwinst verschilt voor bepaalde groepen leerlingen. Zo zal er onder andere vergeleken worden tussen jongens en meisjes, naargelang sociaaleconomische status en naargelang de thuistaal van leerlingen. Tevens zal er nagegaan worden of we kunnen spreken van een 'inhaalbeweging' voor begrijpend lezen tussen het vierde leerjaar en het zesde leerjaar door onze bevindingen te vergelijken met internationale literatuur.
3. Hoe evolueren de attitudes van leerlingen tussen het einde van het vierde leerjaar en het einde van het zesde leerjaar? Deze onderzoeksvraag wordt in het algemeen gesteld, maar ook voor specifieke deelgroepen. Zo zal er onder andere vergeleken worden tussen jongens en meisjes, naargelang sociaaleconomische status en naargelang de thuistaal van leerlingen.
4. Meten de toetsen van PIRLS en het Peilingsonderzoek dezelfde onderliggende vaardigheid voor begrijpend lezen? Er zal dus worden nagegaan of beide onderzoeken op één meetschaal geplaatst kunnen worden. Het al dan niet kunnen plaatsen van de toetsen van beide onderzoeken op één meetschaal is ook relevant om beide onderzoeken te duiden. Indien het mogelijk is om de toetsen van het Peilingsonderzoek en PIRLS op één meetschaal te plaatsen, dan kunnen de resultaten van beide studies vergeleken worden op eenzelfde schaal. Dit laat ons toe de vraag te beantwoorden hoe de Vlaamse eindtermen zich verhouden tot de internationale PIRLS-schaal. Tevens kan de afname van PIRLS 2016 bij leerlingen in het vierde leerjaar dan gebruikt worden om te bepalen hoeveel leerlingen in het vierde leerjaar reeds de eindtermen voor begrijpend lezen halen.

Om de onderzoeksvragen te beantwoorden voldoet de herhalingsmeting van PIRLS in 2018 aan verschillende vereisten. Zo hebben de toetsen van de herhalingsmeting een overlap in teksten en vragen met PIRLS 2016, zo kunnen de toetsen van beide tijdstippen op één meetschaal gelegd worden. Verder hebben de toetsen van de herhalingsmeting een overlap in teksten en vragen met het Peilingsonderzoek lezen 2018, zo kan onderzocht worden of PIRLS en het Peilingsonderzoek dezelfde vaardigheid van begrijpend lezen meten. De scholen die deelnamen aan de herhalingsmeting in 2018 bestaan enkel uit scholen die deelnamen aan PIRLS 2016. De meeste leerlingen in de herhalingsmeting van PIRLS in 2018 namen dus ook deel aan PIRLS 2016. Op deze manier kan de leerwinst van dezelfde scholen en leerlingen onderzocht worden.

Ten slotte merken we op dat, in tegenstelling tot de oorspronkelijke afname van PIRLS 2016 waar 45 landen en tien regio's aan deelnamen, de herhalingsmeting van PIRLS in 2018 enkel in Vlaanderen plaatsvond. Er is in deze studie dus geen ranking van landen en regio's voor begrijpend lezen in het zesde leerjaar. Wanneer een internationale vergelijking in dit rapport gemaakt wordt is dit in vergelijking met leerlingen uit het vierde leerjaar, of, inzake leerwinst, op basis van studies in andere landen.

Structuur van dit rapport

Dit rapport bestaat uit verschillende hoofdstukken die elk handelen over een ander onderdeel van de gegevensverzameling of analyses. Sommige lezers zijn waarschijnlijk enkel geïnteresseerd in bepaalde onderdelen. Daarom geven we hier kort weer wat het onderwerp is van elk hoofdstuk.

In **Hoofdstuk 1** beschrijven we de voorbereiding van de herhalingsmeting. In deze eerste sectie geven we kort weer hoe de herhalingsmeting voor PIRLS in 2018 werd georganiseerd. We beschrijven onder andere hoe de deelname van scholen verworven werd, hoe de afname van toetsen en vragenlijsten verliep, en hoe de verwerking van het testmateriaal gebeurde. In de tweede sectie bespreken we hoe de toetsen van de herhalingsmeting van PIRLS in 2018 werden ontwikkeld. We beschrijven onder andere welke uitgangspunten bij de toetsontwikkeling werden gehanteerd. Verder tonen we hoe items en teksten van PIRLS 2016 werden geselecteerd, en hoe de items en teksten van het Peilingsonderzoek lezen 2018 werden geselecteerd.

In **Hoofdstuk 2** beschrijven we de steekproef van de herhalingsmeting van PIRLS in 2018. Aangezien de herhalingsmeting bestond uit vrijwillig deelnemende scholen van de steekproef van PIRLS 2016, wordt beschreven wat de kenmerken waren van scholen die vrijwillig deelnamen. Omdat de herhalingsmeting van PIRLS in 2018 vergeleken wordt met het Peilingsonderzoek lezen 2018, vergelijken we ook beide steekproeven.

In **Hoofdstuk 3** bespreken we hoe *Item Response Theory* (IRT) werd toegepast op de herhalingsmeting van PIRLS in 2018. Daarom wordt eerst een korte beschrijving gegeven van de algemene principes van IRT. Vervolgens tonen we hoe IRT werd toegepast in PIRLS 2016 en hoe de herhalingsmeting van PIRLS in 2018 op de meetschaal van PIRLS 2016 werd geplaatst. Op een gelijkaardige manier tonen we hoe IRT werd toegepast in het Peilingsonderzoek lezen 2018 en hoe de herhalingsmeting van PIRLS in 2018 op de meetschaal werd geplaatst van het Peilingsonderzoek lezen 2018. Aansluitend tonen we hoe, via de herhalingsmeting van PIRLS in 2018, PIRLS 2016 op de meetschaal van het Peilingsonderzoek lezen 2018 werd gelegd. Als laatste geven we aan hoe IRT werd gebruikt om verschillende attitudeschalen in deze herhalingsmeting op de meetschaal te plaatsen van PIRLS 2016.

Hoofdstuk 4 handelt over de onderzoeksvraag naar het niveau van begrijpend lezen van de Vlaamse leerlingen op het einde van het zesde leerjaar. Dit hoofdstuk beschrijft het niveau voor begrijpend lezen, zowel in het algemeen als voor specifieke leerlingengroepen. In dit hoofdstuk wordt steeds een vergelijking gemaakt tussen 2016 en 2018 voor de scholen die zowel deelnamen aan PIRLS 2016 als aan de herhalingsmeting van 2018. Ook hoeveel leerlingen zich binnen elke van de internationale standaarden bevinden zal besproken worden.

Hoofdstuk 5 handelt over de onderzoeksvraag naar de leerwinst in begrijpend lezen van de Vlaamse leerlingen tussen het vierde leerjaar en het zesde leerjaar. Dit hoofdstuk start met eerder beschrijvende analyses over de leerwinst van leerlingen, zowel in het algemeen als voor specifieke leerlingengroepen. In de analyses van dit hoofdstuk worden enkel leerlingen opgenomen die zowel deelnamen aan PIRLS 2016 als aan de herhalingsmeting van PIRLS in 2018. Tevens zal in dit hoofdstuk internationale literatuur besproken worden zodat we kunnen nagaan of er sprake is van een 'inhaalbeweging' tussen het vierde leerjaar en zesde leerjaar voor begrijpend lezen in Vlaanderen (Zie Vlor, 2018).

Hoofdstuk 6 handelt over de onderzoeksvraag naar de evolutie in de attitudes van Vlaamse leerlingen tussen het vierde leerjaar en het zesde leerjaar. Dit hoofdstuk start met eerder beschrijvende analyses over de evolutie van leerlingen, zowel in het algemeen als voor specifieke leerlingengroepen. In de analyses van dit hoofdstuk worden enkel leerlingen opgenomen die zowel deelnamen aan PIRLS 2016 als aan de herhalingsmeting van PIRLS in 2018.

Hoofdstuk 7 handelt over de onderzoeksvraag of de toetsen van PIRLS en de toetsen van het Peilingsonderzoek dezelfde onderliggende vaardigheid voor begrijpend lezen meten. In dit hoofdstuk gebruiken we onder andere verschillende tests op basis IRT om na te gaan of er één of twee aparte onderliggende vaardigheden gemeten worden. We onderzoeken dit echter ook binnen het kader van de *structural equation modeling* (SEM). We beantwoorden de vraag hoe de Vlaamse eindtermen zich verhouden tot de internationale PIRLS-schaal. We geven ook aan hoe de operationalisering van begrijpend lezen verschilt tussen het Vlaamse Peilingsonderzoek en het internationale PIRLS-onderzoek.

We sluiten af met een algemene conclusie. Hierin vatten we eerst de conclusies uit de voorgaande hoofdstukken samen. Vervolgens bespreken we verschillende implicaties voor het Vlaamse onderwijs. Ten slotte geven we enkele beleidsadviezen.

1 Voorbereiding herhalingsmeting 2018

1.1 Organisatie afname en verwerking herhalingsmeting

>> In deze sectie lichten we kort toe hoe de herhalingsmeting van PIRLS in 2018 werd georganiseerd.

De scholen die onderdeel waren van de steekproef van de herhalingsmeting van PIRLS in 2018 bestonden exclusief uit scholen die hadden deelgenomen aan PIRLS 2016. Alle scholen van de steekproef van PIRLS 2016, uitgezonderd de BuLO scholen, werden in maart en april 2018 gecontacteerd met de vraag of ze wilden deelnemen aan de herhalingsmeting. De deelname van scholen aan de herhalingsmeting gebeurde op vrijwillige basis. In de sectie over de steekproef worden de kenmerken van scholen die deelnamen aan de herhalingsmeting besproken.

Tijdens de maand maart werden de nodige gegevens opgevraagd bij de scholen die deelnamen aan de herhalingsmeting. Er werden onder andere klaslijsten verzameld van de leerlingen in het zesde leerjaar (en de leerlingen die aan PIRLS 2016 deelnamen, maar zittenblijver waren). Met deze informatie kon het juiste aantal toetsmateriaal voorzien worden en kon het toetsmateriaal ook gepersonaliseerd worden. Daarop volgend gebeurde de inpak van het testmateriaal in de laatste week van april en de eerste week van mei. Het toetsmateriaal werd vervolgens op het einde van de eerste week van mei 2018 per post verstuurd naar de scholen.

De herhalingsmeting van PIRLS in 2018 bestond uit de afname van toetsen begrijpend lezen, een leerlingvragenlijst en een leerkrachtvragenlijst. De toetsen en vragenlijsten werden steeds op papier afgenomen, zo was vergelijkbaarheid met de afname van de toetsen PIRLS 2016 gegarandeerd. Sommige achtergrondkenmerken (zoals SES) werden niet opnieuw bevestigd, omdat deze reeds uitgebreid bevestigd werden in 2016 en doorgaans stabiel blijven. De leerlingvragenlijst bevatte wel vragen naar thuistaal en aantal boeken thuis, zodat we ook voor de nieuw ingestroomde leerlingen enige achtergrondinformatie hebben. Er werden geen vragenlijsten afgenomen van ouders, directeurs of leraren van andere leerjaren.

De afname van de herhalingsmeting verliep op een gelijkaardige wijze als in PIRLS 2016. De scholen mochten de toetsen en vragenlijsten inplannen in de periode tussen 14 mei en 8 juni. De scholen kregen dus 4 weken tijd. Net zoals in 2016 waren tijdens de toetsafname geen toetsassistenten aanwezig, maar kregen de leraren een informatiebundel met duidelijke richtlijnen voor afname die ze strikt dienden op te volgen. Elke leerling kreeg een toetsboekje met twee of drie teksten tijdens een klassikale afname. De afname verliep zoals tijdens PIRLS 2016 in twee delen van telkens 40 minuten die, onderbroken door een korte pauze, onmiddellijk na elkaar werden afgenomen. De leerlingvragenlijst en leerkrachtvragenlijst van de herhalingsmeting bevatten grotendeels dezelfde vragen als in PIRLS 2016. Er werd bijvoorbeeld gevraagd naar het gebruikte handboek en de onderwijspraktijk. Bijkomend werd ook leesplezier bij leraren en de deelname aan leesbevorderingsprojecten bevestigd.

Na de afname bezorgden de scholen ons de ingevulde toetsen en vragenlijsten terug via de post. Hiervoor werden verzendetiketten voorzien. Het toetsmateriaal werd vervolgens door het PIRLS-team verzameld, gecontroleerd en gesorteerd.

Vervolgens kon met de verwerking van het toetsmateriaal gestart worden. De verwerking verliep in drie fases. De eerste fase was de scoring van de antwoorden van de leerlingen op de open toetsvragen. De scoring gebeurde door masterstudenten die hiervoor werden opgeleid. Als garantie op een voldoende interbeoordelaarsbetrouwbaarheid werd een deel van de toetsvragen dubbel gescoord. Na het scoren werden deze scores, samen met de antwoorden van leerlingen op de meerkeuzevragen, ingevoerd in hiervoor speciaal ontworpen invoerschermen. Ook bij de invoer werd een controle uitgevoerd van elke jobstudent door middel van een gedeeltelijke dubbele invoer. Zowel bij de scoring als bij de invoer werden dezelfde procedures gevolgd als bij PIRLS 2016.

In de derde fase van de verwerking werden de vaardigheidsscores van leerlingen geschat. Voor de vaardigheidsscores werd *Item Response Theory* (IRT) gebruikt, wat een statistische methode is om toetsen te analyseren en op basis daarvan vaardigheidsscores te schatten. Het voordeel van deze methode is dat verschillende testen op één gemeenschappelijke meetschaal geplaatst kunnen worden. Enkel zo kan de vooruitgang van leerlingen doorheen de tijd beschreven worden. Het is ook deze methode die door het internationale PIRLS-onderzoek gebruikt wordt.

Het gebruik van IRT staat ons dan ook toe de toetsscores van de herhalingsmeting van PIRLS in 2018 op dezelfde meetschaal te plaatsen als PIRLS 2016. In [Hoofdstuk 3](#) gaan we dieper in op het gebruik van IRT op de toetsen van de herhalingsmeting.

1.2 Ontwikkeling toetsboekjes

In deze sectie geven we weer hoe de toetsen van de herhalingsmeting van PIRLS in 2018 werden ontwikkeld. We beschrijven eerst welke uitgangspunten bij de toetsontwikkeling werden gehanteerd. Vervolgens tonen we hoe items en teksten van PIRLS 2016 werden geselecteerd, en hoe de items en teksten van het Peilingsonderzoek lezen 2018 werden geselecteerd.

1.2.1 Validiteit en betrouwbaarheid

Elk onderzoek naar de vaardigheden van leerlingen wordt steeds met hetzelfde probleem geconfronteerd: ‘Hoe meten we de onderliggende vaardigheid waar we als onderzoeker geïnteresseerd in zijn?’. De constructie van een toets kan op verschillende wijzen gebeuren, maar wordt voornamelijk aan de hand van twee criteria geëvalueerd (Kline, 1986). Het eerste criterium is of de toets de beoogde onderliggende vaardigheid daadwerkelijk meet, dit is de validiteit. Het tweede criterium is of de toets de onderliggende vaardigheid op een voldoende betrouwbare manier meet, dit is de betrouwbaarheid. Voor onderzoek naar de vaardigheden van leerlingen is het dus noodzakelijk dat de gebruikte toetsen valide en betrouwbaar zijn.

De validiteit van een toets kan op verschillende manieren gevrijwaard worden. Zo werden tijdens de toetsconstructie van PIRLS 2016 enkel items geselecteerd waarover experts het eens waren dat deze de vaardigheid voor begrijpend lezen bij een internationaal leerlingenpubliek meten (Mullis & Martin, 2015). Bij het Peilingsonderzoek lezen 2018 werden enkel items gekozen waarover Vlaamse experts het eens waren dat deze de vaardigheid voor begrijpend lezen bij een Vlaams leerlingenpubliek meten (Steunpunt Toetsontwikkeling en Peilingen, 2019).

Enkel een oordeel van experts is echter niet voldoende om de validiteit van een toets te garanderen. Daarom wordt er doorgaans in een verkennende studie onderzocht of de items ook daadwerkelijk dezelfde vaardigheid meten. In eerste instantie wordt er onderzocht of elk item van een toets op zo'n wijze met de andere items samenhangt zodat de samenhang enkel veroorzaakt kan zijn door de onderliggende vaardigheid. Er mag dus geen andere vaardigheid systematisch een rol spelen. Dit is de constructvaliditeit. In de verkennende studie worden daarom toetsen afgenomen bestaande uit de ontwikkelde items. Indien er items zijn die een andere vaardigheid meten, worden zij verwijderd uit de finale toetsen. Zowel PIRLS 2016 als het Peilingsonderzoek lezen 2018 onderzochten de validiteit in zo'n verkennende studie (Mullis, Martin, Foy, et al., 2017; Steunpunt Toetsontwikkeling en Peilingen, 2019). Bij PIRLS 2016 werd ook nagegaan of de toets samenhangt met andere variabelen zoals verwacht kan worden op basis van de definitie en voorgaande informatie over de vaardigheid. Dit heet de predictieve validiteit (Mullis, Martin, Foy, et al., 2017).

Aansluitend op het onderzoek naar de validiteit van de toetsen werd ook de betrouwbaarheid onderzocht. Het verkennend onderzoek geeft namelijk ook zicht op hoe elk item tot de betrouwbaarheid van een toets bijdraagt. De betrouwbaarheid van een toets is hierbij een functie van hoe goed elk item de onderliggende vaardigheid meet, maar ook van de moeilijkheidsgraad van ieder item. Een item kan immers wel de onderliggende vaardigheid meten, maar als het item te moeilijk of te makkelijk is voor de beoogde populatie dan zal het item niet bijdragen tot de betrouwbaarheid van de toets. Bij de samenstelling van de finale toets(en) worden dan ook die items geselecteerd die het meeste bijdragen tot een betrouwbare meting van de onderliggende vaardigheid. Omdat PIRLS 2016 geïnteresseerd was in de vaardigheidsschatting van leerlingen van verschillende vaardigheidsniveaus werden items met verschillende moeilijkheidsgraden opgenomen. Omdat het Peilingsonderzoek lezen 2018 voornamelijk geïnteresseerd was in welke leerlingen al dan niet de eindtermen behalen, werden er enigszins meer items geselecteerd waarvan de moeilijkheidsgraad dicht bij de cesuur lag.

Aangezien de herhalingsmeting van PIRLS in 2018 diende als een herafname van PIRLS 2016, lijkt het hergebruiken van de toetsen van PIRLS 2016 een logische keuze. Immers werden de toetsen van PIRLS 2016 al als valide en betrouwbaar bevonden, zowel internationaal als in Vlaanderen. Verder zou de herhalingsmeting van PIRLS in 2018 daarmee eenvoudig op dezelfde meetschaal te plaatsen zijn als PIRLS 2016. De toetsen van PIRLS 2016 zouden echter te veel gemakkelijke items bevatten voor de steekproef in het zesde leerjaar, wat een lage betrouwbaarheid zou geven. We verwachtten zelfs dat de betrouwbaarheid bij de sterkste leerlingen zo laag zou zijn dat er een plafondeffect zou zijn.

De verschillen in vaardigheden tussen de sterkere leerlingen zou dan niet meer vastgesteld kunnen worden. Daarom werd er gekozen om moeilijkere items op te nemen van het Peilingsonderzoek lezen 2018. Immers zijn de toetsen van dit onderzoek ook valide en betrouwbaar voor begrijpend lezen. We werkten daarom, noodgedwongen, in de veronderstelling dat PIRLS 2016 en het Peilingsonderzoek lezen 2018 exact dezelfde onderliggende vaardigheid voor begrijpend lezen meten. De veronderstelling dat PIRLS en het Peilingsonderzoek lezen 2018 hetzelfde meten konden we enkel na de herhalingsmeting toetsen, zoals beschreven in [Hoofdstuk 7](#).

1.2.2 Selectie van teksten en items

De veronderstelling dat de items van PIRLS 2016 en het Peilingsonderzoek lezen 2018 dezelfde vaardigheid meten vereenvoudigde de selectie van teksten en items. We moesten uit beide onderzoeken slechts die teksten en items selecteren die zouden bijdragen tot een betrouwbare herhalingsmeting. Dit hield in dat we items moesten hebben die de onderliggende vaardigheid goed meten en die niet te gemakkelijk of te moeilijk waren voor leerlingen in het zesde leerjaar.

We moeten hierbij opmerken dat de selectie van items en teksten nog aan andere praktische voorwaarden moest voldoen dan validiteit en betrouwbaarheid. Zo moesten de toetsen binnen twee tijdsloten van 40 minuten afgenomen dienen te worden. Verder moest de gelijke verdeling van verhalende en informatieve teksten binnen PIRLS behouden blijven. Ook moesten we de toetsboekjes zo ontwerpen dat leerlingen niet dezelfde tekst ontvingen als tijdens PIRLS 2016. Anders zou er een hertoetsingseffect kunnen zijn. Verder moest elk item bij voldoende leerlingen (ongeveer 500) worden afgenomen om stabiele IRT-analyses te kunnen uitvoeren.

In een eerste stap maakten we daarom een overzicht van de teksten en items van PIRLS 2016 volgens hun moeilijkheidsgraad. In een tweede stap beslisten we welke teksten van PIRLS 2016 we behielden. In [Tabel 1](#) staat een overzicht van deze teksten met de gemiddelde, minimum en maximum moeilijkheidsgraad van de items per tekst, het aantal items per tekst, wat voor soort tekst het was en of de tekst geselecteerd werd. In het algemeen werden de twee gemakkelijkste teksten niet geselecteerd, terwijl de twee moeilijkste teksten wel geselecteerd werden. De andere teksten werden eerder om praktische redenen al dan niet geselecteerd.

De lezers die geïnteresseerd zijn in de PIRLS-teksten en -vragen, vinden enkele voorbeeldteksten in de twee Vlaamse rapporten: Marie en de rode kip (Tielemans et al., 2017); De reis van de groene zeeschildpad (Tielemans et al., 2017) en Bloemen op het dak (Tielemans et al., 2019).

Tabel 1: Teksten PIRLS 2016

Tekst	Gem.	Min.	Max.	n	Tekstsoort	PIRLS	
						Literacy	Selectie
Bloemen op het dak	-0.364	-1.419	0.642	13	Verhalend		X
Olivier en de griffioen	-0.051	-0.993	1.145	13	Verhalend		X
Schitterend stro	0.008	-1.011	0.574	14	Verhalend		
Pemba Sherpa	-1.234	-2.530	0.318	17	Verhalend	X	
De lege pot	-0.162	-0.957	0.413	17	Verhalend		
Marie en de rode kip	0.221	-0.743	1.644	16	Verhalend		X
Haaien	0.101	-0.891	0.979	12	Informatief		X
IJslandse paarden	0.203	-1.522	0.944	15	Informatief		X
Leonardo da Vinci	0.001	-2.275	0.685	12	Informatief		
Hoe hebben we leren vliegeren?	-1.216	-3.130	0.464	17	Informatief	X	
Waar is de honing?	0.138	-0.999	0.791	13	Informatief		
De reis van de groene zeeschildpad	0.062	-0.973	0.622	16	Informatief		X

Notitie: Gem. = gemiddelde moeilijkheidsgraad; Min. = minimum moeilijkheidsgraad; Max. = maximum moeilijkheidsgraad; n = aantal items per tekst, PIRLS literacy: eerder eenvoudige teksten die in de eerste plaats bedoeld zijn voor ontwikkelingslanden

Aansluitend werden in de derde stap teksten geselecteerd van het Peilingsonderzoek lezen 2018. In **Tabel 2** staat een overzicht van deze teksten met de gemiddelde, minimum en maximum moeilijkheidsgraad van de items per tekst, het aantal items per tekst, wat voor soort tekst het was en of de tekst geselecteerd werd. De teksten werden zo geselecteerd dat er voor ieder toetsboekje een goede aanvulling was van items met andere moeilijkheidsgraden dan de items afkomstig van PIRLS 2016.

Tabel 2: Teksten Peilingsonderzoek lezen 2018

Tekst	Gem.	Min.	Max.	n	Selectie
Verjaardagsfeestje	-0.521	-1.253	-0.111	6	
Pretparken	0.050	-0.125	0.339	3	
Dierentabel	-0.148	-1.111	0.389	4	
Moeilijk gaat ook duel	0.103	-0.252	0.350	5	X
Kids at the movies	0.187	0.072	0.422	5	
Gitaar	0.171	-0.195	0.431	4	
De magische sleutel	-0.051	-0.354	0.254	6	X
Vliegtuig	-0.196	-0.693	0.105	5	X
Postzegel	0.250	0.024	0.644	5	X
Sportclubs	-0.583	-1.468	0.121	4	

Notitie: Gem. = gemiddelde moeilijkheidsgraad; Min. = minimum moeilijkheidsgraad; Max. = maximum moeilijkheidsgraad; n = aantal items in tekst

Het selectieproces van teksten leidde tot 11 teksten (zie **Tabel 3**). In totaal gaat het om 111 items.

Tabel 3: De 11 geselecteerde teksten voor PIRLS2018, zesde leerjaar

Tekst	Bron	Verhalend	Informatief	Ntal items	Afnameduur
1 IJslandse paarden	PIRLS2016		X	15	40'
2 Bloemen op het dak	PIRLS2016	X		14	40'
3 Haaien	PIRLS2016		X	12	40'
4 Olivier en de griffioen	PIRLS2016	X		13	40'
5 Marie en de rode kip	PIRLS2016	X		16	40'
6 De reis van de groene zeeschildpad	PIRLS2016		X	16	40'
7 Moeilijk gaat ook duel	peiling		X	5	15'
8 Postzegel	peiling		X	5	15'
9 De magische sleutel	peiling	X		6	20'
10 Vliegtuig	peiling		X	5	20'
11 Chinese horoscoop	SFB	?	?	5	15' (?)

Deze 11 teksten werden gecombineerd tot tien toetsboekjes met verschillende combinaties van teksten. Elk toetsboekje bestond uit twee of drie teksten, waarbij elke tekst bestond uit een verhaal of artikel en bijbehorende open vragen en/of meerkeuzevragen. Leerlingen kregen twee teksten als de teksten enkel afkomstig waren van PIRLS 2016. Leerlingen kregen drie teksten als de teksten ook afkomstig waren uit het peilingsonderzoek of het Centrum voor Schoolfeedback (SFB-project). Deze teksten waren namelijk korter en hadden minder vragen. Voor de afname kregen leerlingen tweemaal 40 minuten werktijd, met een korte pauze tussenin. Voor een tekst afkomstig van PIRLS 2016 kregen de leerlingen 40 minuten. Er moesten echter twee teksten gelezen worden binnen de 40 minuten wanneer het teksten waren van het Peilingsonderzoek lezen 2018 en/of het SFB-project. De tekst uit het SFB-project was een extra moeilijke tekst die specifiek bedoeld is voor de sterkere groepen leerlingen. Er waren 10 toetsboekjes, waarvan zes met drie teksten en vier met twee teksten. Er waren in totaal 11 teksten, waarvan sommigen in drie toetsboekjes zaten, terwijl anderen in twee toetsboekjes zaten. Dit was volgens een niet volledig overlappend design (**Tabel 4**). Om het werkgeheugen van de leerlingen niet te zwaar te belasten, diende elke leerling slechts één toetsboekje in te vullen. Dit betekent dat niet alle leerlingen van een klas dezelfde teksten en vragen voorgelegd kregen. Bij de toewijzing van de toetsboekjes werd ook voorkomen dat leerlingen in 2018 dezelfde tekst zouden krijgen als in 2016. In **Tabel 4** wordt weergegeven welke teksten in welk van de toetsboekjes werden opgenomen.

Tabel 4: Geselecteerde teksten per toetsboekje

Toetsboekje	Deel 1 tekst(en)		Deel 2 tekst(en)	
1	Marie en de rode kip		Chinese horoscoop	Postzegel
2	Haaien		Vliegtuig	Postzegel
3	De magische sleutel	Moeilijk gaat ook duel	De reis van de groene zeeschildpad	
4	Vliegtuig	Postzegel	Bloemen op het dak	
5	IJslandse paarden		Marie en de rode kip	
6	Olivier en de griffioen		Haaien	
7	De reis van de groene zeeschildpad		Olivier en de griffioen	
8	Bloemen op het dak		IJslandse paarden	
9	De magische sleutel	Chinese horoscoop	IJslandse paarden	
10	Moeilijk gaat ook duel	Chinese horoscoop	Olivier en de griffioen	

Het combineren van internationale vragen en Vlaamse vragen in één toets is redelijk uniek, maar helpt de Vlaamse beleidsmakers om te kunnen inschatten wat de waarde is van het PIRLS-onderzoek voor het Vlaamse basisonderwijs (Oliveri et al., 2018).

In 2016 waren er 16 verschillende toetsboekjes (Tielemans et al., 2017) en in 2018 waren er 10 verschillende toetsboekjes. Hieronder ziet u hoeveel leerlingen elke combinatie van toetsboekje in 2016 en 2018 invulden. Bepaalde combinaties komen niet voor, omdat de toetsboekjes dezelfde teksten bevatten.

Tabel 5: Bestaande combinaties toetsboekjes vierde leerjaar en zesde leerjaar

Toetsboekje 2018 L6												
Toetsboekje											Totaal	
2016 L4	1	2	3	4	5	6	7	8	9	10	/	
1	42	31	31			44	49			37	59	293
2	38	24	34	27		38	43			36	55	295
3	18	20	25	25	23	23	27	28	26	21	54	290
4	27		23	26	38		34	31	29	21	54	283
5	23	20	18	21	31	22	24	23	21	23	60	286
6	45		32	30	46			32	42		59	286
7	40	28	25	24	35			39	36		62	289
8	18	18	23	27	24	28	31	22	18	19	61	289
9	25	30	27		39	20	37		26	26	57	287
10	17	20	30	26	27	23	21	20	21	19	63	287
11	30		30	22	35		31	23	31	27	58	287
12	39	23	28	29	39			28	33		69	288
13	31	27	21	1	29	30	35		25	27	68	294
14	32	35	39	21		35	40			35	57	294
15	15	16	24	23	27	19	24	23	19	26	75	291
16		121		110		123		132	96	98	179	859
/	50	58	50	40	71	47	75	58	62	57		568
Totaal	490	471	460	452	464	452	471	459	485	472	1090	5766

1.3 Ander toetsmateriaal

Naast de toetsboekjes voor Nederlands begrijpend lezen werden ook de leerlingvragenlijst en een leerkrachtvragenlijst van PIRLS 2016 opnieuw afgenomen. De vragenlijsten van PIRLS 2016 dienen om inzicht te verwerven in de contexten waarin leerlingen leren. De leerlingvragenlijst handelt over verschillende aspecten van leerlingen hun leven thuis en op school waaronder biografische informatie, hun thuisomgeving, het schoolklimaat waarin ze leren en hun attitudes ten opzichte van lezen. De leerkrachtvragenlijst vraagt naar leerkrachten hun opleiding, professionele ontwikkeling en ervaring in het leesonderwijs. Verder wordt er ook gevraagd naar hoe de leerkrachten lesgeven in de klas, op welke manier ze instructie geven en de strategieën die ze implementeren.

2 Steekproef

>> De steekproef van de herhalingsmeting in 2018 bestond uit de scholen die reeds deelnamen aan PIRLS 2016 en vrijwillig een tweede maal deelnamen. In deze sectie beschrijven hoe de PIRLS 2016 steekproef representatief was in het vierde leerjaar, welke scholen en leerlingen deelnamen in 2018, en wat dit betekent naar representativiteit van de steekproef voor de Vlaamse populatie van leerlingen het zesde leerjaar.

2.1 Steekproeftrekking PIRLS 2016

Een eerste stap binnen de steekproeftrekking van PIRLS 2016 was het bepalen van de populatie. Deze populatie werd bepaald als alle leerlingen die op het moment van de testafname in het vierde leerjaar zaten of de leerstof van het vierde leerjaar onderwezen kregen indien zij in het BuLO zaten. Voor het BuLO bestond de populatie uit leerlingen in type 3 (emotionele of gedragsstoornis) en de toenmalige types 1 (lichte verstandelijke beperking) en 8 (ernstige leerstoornissen). Leerlingen uit andere types en leerlingen uit de types 1, 3 en 8 die niet de leerstof van het vierde leerjaar onderwezen kregen, werden beschouwd als niet behorende tot de onderzoekspopulatie. Franstalige basisscholen en kleine scholen (d.i. scholen met minder dan vijf leerlingen in het vierde leerjaar) waren ook geen onderdeel van de populatie (Tielemans et al., 2017).

In de tweede stap werd informatie verzameld over deze populatie via de administratieve databank die door de overheid ter beschikking werd gesteld. Volgens deze gegevens ging om een totaal van 69733 leerlingen verspreid over 2404 scholen, waarvan 2665 leerlingen in het BuLO (182 scholen). Hierin zaten echter nog negen Franstalige scholen met 304 leerlingen, dewelke werden verwijderd omdat ze niet tot de doelpopulatie behoorden. Er werden ook 69 kleine scholen (186 leerlingen) verwijderd omdat zulke kleine scholen volgens de IEA ongeschikt zijn voor analyses. In totaal werden zo 0,70% van de leerlingen verwijderd.

In de derde stap werd een representatieve steekproef op basis van deze gegevens getrokken. De steekproeftrekking bestond uit twee trappen. In de eerste trap werden de scholen geselecteerd en in de tweede trap werden leerlingen in de scholen geselecteerd. De eerste trap was een gestratificeerde steekproeftrekking, waarbij scholen volgens de strata van bepaalde criteria geselecteerd werden. Een eerste criterium was de provincie waarin de scholen gelegen waren. Ten tweede werd er gekeken naar het net waartoe deze scholen behoorden. Ten derde werd er ook rekening gehouden met het leerlingenpubliek: er werd gekeken naar het percentage indicatorleerlingen. Dit was het aantal leerlingen dat aantikt op de SES-kenmerken, wat maakte dat deze leerlingen in aanmerking komen voor extra financiering. Voor alle geselecteerde scholen waren ook vervangscholen voorzien die tot dezelfde strata behoorden van de criteria. Zo werden er 150 scholen geselecteerd en 2 vervangscholen per geselecteerde school voorzien. Naast de 150 scholen voor regulier basisonderwijs werden in PIRLS 2016 ook 10 scholen voor buitengewoon lager onderwijs (BuLO) geselecteerd. In een tweede stap van de steekproeftrekking werden binnen de geselecteerde scholen alle vierdejaarsklassen van de geselecteerde scholen opgenomen.

In PIRLS 2016 waren 119 van de 150 eerste selectiescholen regulier lager onderwijs bereid tot deelname. Daarnaast namen er nog 22 vervangscholen deel. Van de tien scholen BuLO waren er vijf eerste selectiescholen bereid tot deelname en namen er nog twee vervangscholen deel. Er werd deze scholen gevraagd alle leerlingen die de leerstof van het vierde leerjaar kregen aangeboden, op te nemen in de steekproef.

In totaal werden er in 2016 gegevens verzameld van 148 scholen, waarvan zeven scholen voor BuLO. Dit komt overeen met een totaal van 5378 leerlingen, van wie 139 leerlingen uit het BuLO. Waar de populatie 3,82% BuLO-leerlingen bevatte, zagen we dus dat de steekproef slechts 2,58% BuLO-leerlingen telde. Om te corrigeren voor verschillen tussen populatie en uiteindelijke steekproefaantallen werd in internationaal vergelijkende analyses gewerkt met gewichten. Concreet werd deze 'onderrepresentatie' van BuLO-leerlingen in de steekproef ten opzichte van de populatie opgevangen door de BuLO-leerlingen een hoger gewicht toe te kennen in de analyses. Zo bedroeg hun uiteindelijke aandeel in de analyses het geschatte populatieaandeel, met name 3,82%. [Tabel 6](#) bevat een overzicht van de oorspronkelijke steekproef en de uiteindelijk deelnemende scholen en leerlingen.

Tabel 6: Steekproef PIRLS 2016

	Totaal	Regulier lager onderwijs	BuLO
Deelnemende scholen	148	141	7
Aantal leerlingen	5198	5069	129

Aan de leerkrachten van de deelnemende klassen werd gevraagd een klassenlijst in te vullen en hierbij aan te geven of er leerlingen volgens hen officieel uitgesloten moesten worden omwille van (1) een ernstige fysieke beperking, (2) ernstige leer- en of leesproblemen en/of (3) minder dan één jaar in Vlaanderen en onvoldoende beheersing van de Nederlandse taal. Aangezien deze uitsluitingsnorm in alle deelnemende landen werd toegepast, werden deze leerlingen niet mee opgenomen in de internationale steekproef. In Vlaanderen gaat het om een totaal van 28 leerlingen, wat gelijk is aan 0,52% van de getrokken steekproef.

In PIRLS 2016 zijn de uitsluitingen op school- en individueel niveau samen goed voor 1,22%. Rekening houdend met de internationale uitsluitingsnormen en de drie criteria (provincie, onderwijsnet en SES) werd een steekproef bekomen die voldoende representatief is voor het Vlaams lager onderwijs in het vierde leerjaar (zie ook Tielemans et al., 2017).

2.2 Deelname steekproef 2018

De herhalingsmeting van PIRLS in 2018 gebeurde op vrijwillige basis, scholen die deelnamen in 2016 werd vrijblijvend opgebeld en gemaïld met de vraag of zij terug wilden deelnemen in 2018. De 7 BuLO-scholen in PIRLS 2016 werden in 2018 niet gevraagd om opnieuw deel te nemen. Uiteindelijk beslisten 127 van de 141 scholen (90,1%) die in 2016 deelnamen om mee te doen aan de herhalingsmeting in 2018. Eén school van de vrijwillig deelnemende scholen had de toetsafname echter incorrect uitgevoerd. Deze school wordt niet opgenomen in de analyses, wat resulteert in 126 vrijwillig en correct deelnemende scholen (89,4%). De meeste scholen die deelnamen in 2016 namen dus ook deel in 2018. Het is echter mogelijk dat de groep van scholen die deelnam aan de herhalingsmeting in 2018 verschilde van de groep van scholen die deelnam in 2016. Anders gezegd: de 14 scholen die beslisten om niet deel te nemen en de ene school die incorrect deelnam (10,6%) verschillen mogelijks van de 126 scholen die beslisten om wel deel te nemen.

Binnen de 126 scholen waren er 4902 leerlingen. 4860 van deze leerlingen bevonden zich in het 6de leerjaar, terwijl 42 van deze leerlingen zich in het 5de leerjaar bevonden¹. Van deze 4860 leerlingen in het zesde leerjaar waren er 4625 leerlingen die hadden deelgenomen aan de toets Nederlands begrijpend lezen in de herhalingsmeting. 10 van deze leerlingen voldeden echter aan de uitsluitingscriteria van PIRLS, en zij werden niet opgenomen in de analyses. De resterende 4615 leerlingen worden in deze studie gebruikt om het niveau voor begrijpend lezen in het zesde leerjaar te beschrijven in [Hoofdstuk 4](#).

Van de 4615 leerlingen waren er 4046 leerlingen die ook hadden deelgenomen aan de toets begrijpend lezen in PIRLS 2016. Deze leerlingen worden gebruikt in [Hoofdstuk 5](#) om de leerwinst tussen het vierde leerjaar (2016) en zesde leerjaar (2018) te beschrijven

Zowel in de steekproef van PIRLS 2016 als de steekproef van de herhalingsmeting in 2018 namen zo goed als alle leerlingen in het vierde of zesde leerjaar deel. Non-respons van leerlingen binnen scholen kon de resultaten dus niet op een betekenisvolle wijze vertekenen. Toch zagen we dat, binnen scholen die deelnemen aan de herhalingsmeting in 2018, veel leerlingen die deelnamen in 2016 niet meer deelnamen in 2018. De oorzaak hiervan was niet specifiek gebonden aan deze steekproef, maar omdat binnen de Vlaamse populatie van leerlingen veel leerlingen nog van school veranderen tussen het vierde en zesde leerjaar, of het zesde leerjaar overslaan bijvoorbeeld wanneer ze 'op leeftijd' starten in 1B.

Hoe de steekproef van scholen verschilt tussen PIRLS 2016 en PIRLS 2018 onderzochten we aan de hand van verschillende schoolkenmerken. Welke leerlingen tussen het 4de en 6de leerjaar verdwenen in de deelnemende scholen werd onderzocht aan de hand van verschillende leerlingkenmerken.

¹ De 42 leerlingen in het vijfde leerjaar waren zittenblijvers. Het gaat om leerlingen die ofwel het vierde ofwel het vijfde leerjaar overdeden in dezelfde school.

2.2.1 Deelname scholen 2018

Om te onderzoeken hoe de scholen die deelnamen aan de herhalingsmeting 2018 verschilden van de groep scholen die deelnam aan PIRLS 2016, vergeleken we de kenmerken van beide groepen scholen. We keken daarbij ook hoeveel leerlingen in het schooljaar 2015-2016 in de scholen aanwezig waren. De schoolkenmerken die we onderzochten waren:

- De provincie. Een school kon gelegen zijn de in de provincie Antwerpen, het Brussels Hoofdstedelijk gewest, Limburg, Oost-Vlaanderen, Vlaams-Brabant of West-Vlaanderen.
- Het onderwijsnet. Een school kon tot het gesubsidieerde vrij onderwijs behoren, het gemeenschapsonderwijs of het gesubsidieerd officieel onderwijs. In de beschrijving van de steekproef onderscheiden we enkel het vrij onderwijs en het officieel onderwijs.
- Gemiddelde prestaties begrijpend lezen 2016. Wegens de deelname aan PIRLS 2016 kenden we van elke school de gemiddelde prestaties voor begrijpend lezen in 2016. Daarom werd ook nagegaan of de deelname aan de herhalingsmeting in 2018 samenhang met de gemiddelde prestatie voor begrijpend lezen in 2016. We verdeelden de scholen in vijf groepen van gelijke grootte volgens hun gemiddelde vaardigheid. De eerste groep had de kleinste gemiddelde vaardigheid in begrijpend lezen in 2016, de vijfde groep had de grootste gemiddelde vaardigheid in begrijpend lezen in 2016.
- Percentage leerlingen met een laagopgeleide moeder (OKI). Van elke school was informatie beschikbaar over hoeveel leerlingen aantikten op elk van de vier kenmerken van de onderwijskansarmoede-indicator. Dit omvat onder andere het percentage leerlingen waarvan de moeder maximaal het lager secundair onderwijs heeft afgerond. We verdeelden de scholen in vijf groepen volgens het percentage leerlingen dat aantikte op deze indicator. In de eerste groep tikt minder dan 10% aan (zeer laag), in de tweede groep tikt 10% tot 20% aan (laag), in de derde groep tikt 20% tot 30% aan (gemiddeld), in de vierde groep tikt 30% tot 50% aan (hoog), en in de vijfde groep tikt meer dan 50% aan (zeer hoog).
- Percentage leerlingen met een schooltoelage (OKI). Van elke school was informatie beschikbaar over hoeveel leerlingen aantikten op elk van de vier kenmerken van de onderwijskansarmoede-indicator. Dit omvat onder andere het percentage leerlingen waarvan het gezin in aanmerking komt voor een schooltoelage en deze aangevraagd heeft. We verdeelden de scholen in vijf groepen volgens het percentage leerlingen dat aantikte op deze indicator. In de eerste groep tikt minder dan 10% aan (zeer laag), in de tweede groep tikt 10% tot 20% aan (laag), in de derde groep tikt 20% tot 30% aan (gemiddeld), in de vierde groep tikt 30% tot 50% aan (hoog), en in de vijfde groep tikt meer dan 50% aan (zeer hoog).
- Percentage leerlingen met een gezin角度 anders dan het Nederlands (OKI). Van elke school was informatie beschikbaar over hoeveel leerlingen aantikten op elk van de vier kenmerken van de onderwijskansarmoede-indicator. Dit omvat onder andere het percentage leerlingen waarvan de gezin角度 niet Nederlands is. We verdeelden de scholen in vijf groepen volgens het percentage leerlingen dat aantikte op deze indicator. In de eerste groep tikt minder dan 10% aan (zeer laag), in de tweede groep tikt 10% tot 20% aan (laag), in de derde groep tikt 20% tot 30% aan (gemiddeld), in de vierde groep tikt 30% tot 50% aan (hoog), en in de vijfde groep tikt meer dan 50% aan (zeer hoog).

2.2.1.1 Provincie

Eerst bekeken we hoeveel scholen en leerlingen per provincie deelnamen aan PIRLS 2016, en hoeveel van deze scholen en leerlingen deelnamen aan de herhalingsmeting in 2018. **Tabel 7** toont voor beide afnamemomenten het aantal scholen per provincie, het percentage van het aantal scholen per provincie, het aantal leerlingen per provincie en het percentage van het aantal leerlingen per provincie. We vonden geen grote verschuivingen in de vertegenwoordiging van de provincies in de steekproef wanneer we de herhalingsmeting in 2018 vergeleken met PIRLS 2016.

Tabel 7: Deelname aan herhalingsmeting volgens provincie

Provincie	Steekproef 2016				Herhalingsmeting 2018			
	Scholen		Leerlingen		Scholen		Leerlingen	
	n	%	n	%	n	%	n	%
Antwerpen	38	27.0	1520	30.0	32	25.4	1362	29.1
Brussels H.G.	7	5.0	175	3.5	6	4.8	158	3.4
Limburg	18	12.8	656	12.9	17	13.5	630	13.5
Oost-Vlaanderen	31	22.0	1074	21.2	30	23.8	1056	22.6
Vlaams-Brabant	23	16.3	867	17.1	19	15.1	731	15.6
West-Vlaanderen	24	17.0	777	15.3	22	17.5	742	15.9
Totaal	141	100.0	5069	100.0	126	100.0	4679	100.0

2.2.1.2 Onderwijsnetten

Vervolgens bekeken we hoeveel scholen en leerlingen van elk onderwijsnet deelnamen aan PIRLS 2016 en hoeveel van deze scholen en leerlingen van elk onderwijsnet deelnamen aan de herhalingsmeting in 2018 (zie [Tabel 8](#)). We vonden geen grote verschuivingen in de vertegenwoordiging van de onderwijsnetten wanneer we de herhalingsmeting in 2018 vergeleken met PIRLS 2016 (zie [Tabel 8](#)).

Tabel 8: Deelname aan herhalingsmeting volgens onderwijsnet

Onderwijsnet	Steekproef 2016				Herhalingsmeting 2018			
	Scholen		Leerlingen		Scholen		Leerlingen	
	n	%	n	%	n	%	n	%
Officieel	53	37.6	1819	35.9	44	34.9	1573	33.6
Vrij	88	62.4	3250	64.1	82	65.1	3106	66.4

2.2.1.3 Gemiddelde vaardigheid begrijpend lezen 2016

Vervolgens bekeken we of de deelname van scholen en leerlingen aan de herhalingsmeting in 2018 verschilde naargelang de vijf categorieën van het schoolgemiddelde voor de vaardigheid in begrijpend lezen in PIRLS 2016. [Tabel 9](#) toont voor beide afnamemomenten het aantal scholen per categorie, het percentage scholen per categorie, het aantal leerlingen per categorie, en het percentage leerlingen per categorie. We vonden dat het schoolgemiddelde weinig invloed heeft op de deelname van scholen aan de herhalingsmeting in 2018. De 14 scholen die beslisten om in 2018 niet opnieuw deel te nemen aan PIRLS behoren tot de categorieën 'zeer laag' (n=4), laag (n=4), gemiddeld (n=3), en hoog (n=3).

Tabel 9: Deelname aan herhalingsmeting volgens schoolgemiddelde vaardigheid begrijpend lezen in PIRLS 2016

Gemiddelde vaardigheid 2016	Steekproef 2016				Herhalingsmeting 2018			
	Scholen		Leerlingen		Scholen		Leerlingen	
	n	%	n	%	n	%	n	%
Zeer laag	28	19.9	951	18.8	24	19.0	818	17.5
Laag	28	19.9	1042	20.6	24	19.0	978	20.9
Gemiddeld	29	20.6	1191	23.5	26	20.6	1097	23.4
Hoog	28	19.9	1020	20.1	24	19.0	921	19.7
Zeer hoog	28	19.9	865	17.1	28	22.2	865	18.5

2.2.1.4 Percentage leerlingen met laagopgeleide moeder

Vervolgens bekeken we of de deelname van scholen en leerlingen aan de herhalingsmeting in 2018 verschilde naargelang de vijf categorieën van het percentage leerlingen met een laagopgeleide moeder. We vonden dat het percentage leerlingen met een laagopgeleide moeder weinig uitmaakt voor de deelname aan de herhalingsmeting in 2018 (zie [Tabel 10](#)).

Tabel 10: Deelname aan herhalingsmeting volgens percentage leerlingen met laagopgeleide moeder

Categorie	Steekproef 2016				Herhalingsmeting 2018			
	Scholen		Leerlingen		Scholen		Leerlingen	
	n	%	n	%	n	%	n	%
Zeer laag	40	28.4	1552	30.6	39	31.0	1535	32.8
Laag	53	37.6	1876	37.0	45	35.7	1657	35.4
Gemiddeld	23	16.3	839	16.6	21	16.7	778	16.6
Hoog	17	12.1	551	10.9	15	11.9	517	11.0
Zeer Hoog	8	5.7	251	5.0	6	4.8	192	4.1

2.2.1.5 Percentage leerlingen met schooltoelage

Vervolgens bekeken we of de deelname van scholen en leerlingen aan de herhalingsmeting in 2018 verschilde naargelang de vijf categorieën van het percentage leerlingen met recht op een schooltoelage. [Tabel 11](#) toont voor beide afnamemomenten het aantal scholen per categorie, het percentage scholen per categorie, het aantal leerlingen per categorie, en het percentage per categorie. We vonden dat het percentage leerlingen met recht op een schooltoelage van een school weinig uitmaakt voor de deelname aan de herhalingsmeting in 2018.

Tabel 11: Deelname aan herhalingsmeting volgens percentage leerlingen met recht op een schooltoelage

Categorie	Steekproef 2016				Herhalingsmeting 2018			
	Scholen		Leerlingen		Scholen		Leerlingen	
	n	%	n	%	n	%	n	%
Zeer laag	25	17.7	841	16.6	23	18.1	792	16.9
Laag	54	38.3	1973	38.9	51	40.2	1858	39.7
Gemiddeld	29	20.6	1018	20.1	24	18.9	895	19.1
Hoog	21	14.9	787	15.5	19	15.0	748	16.0
Zeer Hoog	12	8.5	450	8.9	10	7.9	386	8.2

2.2.1.6 Percentage leerlingen met andere thuistaal

Vervolgens bekeken we of de deelname van scholen en leerlingen aan de herhalingsmeting in 2018 verschilde naargelang de vijf categorieën van het percentage leerlingen die een andere thuistaal hebben dan de onderwijstaal per school. [Tabel 12](#) toont voor beide afnamemomenten het aantal scholen per categorie, het percentage van het aantal scholen per categorie, het aantal leerlingen per categorie, en het percentage van het aantal leerlingen per categorie. We vonden dat het percentage leerlingen met een andere thuistaal van een school weinig uitmaakt voor de deelname aan de herhalingsmeting in 2018.

Tabel 12: Deelname aan herhalingsmeting volgens percentage leerlingen met een andere thuistaal

Categorie	Steekproef 2016				Herhalingsmeting 2018			
	Scholen		Leerlingen		Scholen		Leerlingen	
	n	%	n	%	n	%	n	%
Zeer laag	66	46.8	2354	46.4	62	49.2	2253	48.2
Laag	28	19.9	1024	20.2	23	18.3	904	19.3
Gemiddeld	12	8.5	461	9.1	11	8.7	437	9.3
Hoog	18	12.8	703	13.9	17	13.5	690	14.7
Zeer Hoog	17	12.1	527	10.4	13	10.3	395	8.4

2.2.2 Deelname leerlingen in 2018

In de vorige sectie onderzochten we welke scholen vrijwillig deelnamen aan de herhalingsmeting van PIRLS in 2018 aan de hand van verschillende schoolkenmerken. In deze sectie bespreken we welke leerlingen verdwenen in de deelnemende scholen tussen PIRLS 2016 en de herhalingsmeting in 2018. Hiervoor vergeleken we leerlingkenmerken van beide groepen. De leerlingen die hier onderzocht werden (n=4679) zaten in de 126 scholen die deelnamen aan de herhalingsmeting in 2018. Van die 4679 leerlingen (die in mei 2016 in de 126 scholen zaten), namen er 4046 opnieuw deel in mei 2018 toen ze in het zesde leerjaar zaten. In deze paragraaf bekijken we dus in hoeverre de uitval van 633 leerlingen binnen de deelnemende scholen al dan niet selectief is.

De leerlingkenmerken die we onderzochten waren:

- Prestaties begrijpend lezen 2016. Wegens de deelname aan PIRLS 2016 kenden we van elke leerling zijn of haar prestaties voor begrijpend lezen in 2016. Daarom werd ook nagegaan of de deelname aan de herhalingsmeting in 2018 samenhang met de schoolse prestaties voor begrijpend lezen in 2016. We verdeelden de leerlingen in vijf groepen van gelijke grootte volgens hun vaardigheid. De eerste groep had de kleinste vaardigheid in begrijpend lezen in 2016, de vijfde groep had de grootste vaardigheid in begrijpend lezen in 2016.
- Geboortjaar. Van elke leerling wisten we in welk jaar hij of zij geboren was.
- Geslacht. De leerling was een jongen of een meisje.
- Boeken thuis. Wegens de deelname aan PIRLS 2016 wisten we van elke leerling of er veel boeken zijn in het huis van de leerling. Dit was immers een vraag in de vragenlijst van PIRLS 2016. Deze variabele wordt vaak gebruikt als een indicatie van sociaaleconomische status.
- Thuis taal. Wegens de deelname aan PIRLS 2016 wisten we van elke leerling welke taal zij thuis spraken. Dit was immers een vraag in de vragenlijst van PIRLS 2016.

2.2.2.1 Vaardigheid begrijpend lezen 2016

Eerst bekeken we of de deelname van leerlingen aan de herhalingsmeting in 2018 versilde naargelang hun prestaties voor begrijpend lezen in PIRLS 2016. **Tabel 13** toont voor beide afnamemomenten het aantal leerlingen naargelang de vijf categorieën van vaardigheid en het percentage leerlingen per categorie in de steekproef van de herhalingsmeting. We vonden dat leerlingen met hogere vaardigheid in PIRLS 2016 enigszins vaker deelnamen aan de herhalingsmeting in 2018 dan leerlingen met een lagere vaardigheid. Dat komt hoogstwaarschijnlijk omdat binnen de 126 scholen de laagvaardige leerlingen vaker uitvielen tussen het vierde en het zesde leerjaar (door onder andere zittenblijven, overgaan op leeftijd naar 1B en schoolveranderen). Deze atypische loopbanen worden verder in dit rapport nog toegelicht.

Tabel 13: Deelname aan herhalingsmeting vaardigheid begrijpend lezen leerlingen in PIRLS 2016

Categorie	Steekproef 2016		Herhalingsmeting 2018	
	n	%	n	%
Zeer laag	937	20.0	715	17.7
Laag	938	20.0	800	19.8
Gemiddeld	935	20.0	826	20.4
Hoog	932	19.9	848	21.0
Zeer Hoog	937	20.0	857	21.2
Totaal	4679	100.0	4046	100.0

2.2.2.2 Geboortejaar

Vervolgens bekeken we hoeveel leerlingen van verschillende geboortejaren deelnamen aan PIRLS 2016 en hoeveel van deze leerlingen deelnamen aan de herhalingsmeting in 2018. **Tabel 14** toont voor beide afnamemomenten het aantal leerlingen per geboortejaar en het percentage leerlingen per geboortejaar in de steekproef van de herhalingsmeting. We vonden dat leerlingen van het geboortejaar 2006 (leerlingen 'op leeftijd') vaker deelnamen aan de herhalingsmeting dan leerlingen van eerdere geboortejaren.

Tabel 14: Deelname aan herhalingsmeting volgens geboortejaar leerlingen

Categorie	Steekproef 2016		Herhalingsmeting 2018	
	n	%	n	%
2003	4	0.1	0	0.0
2004	54	1.2	13	0.3
2005	528	11.3	355	8.8
2006	4030	86.1	3629	89.7
2007	62	1.3	49	1.2
2008	1	0.0	0	0.0

2.2.2.3 Geslacht

Vervolgens bekeken we hoeveel jongens en meisjes deelnamen aan PIRLS 2016 en hoeveel jongens en meisjes deelnamen aan de herhalingsmeting in 2018. **Tabel 15** toont voor beide afnamemomenten het aantal jongens en meisjes en het percentage van het aantal jongens en meisjes in de steekproef van de herhalingsmeting. We vonden dat er geen noemenswaardige verschuiving plaatsvond in de verhouding tussen jongens en meisjes over beide afnamemomenten.

Tabel 15: Deelname aan herhalingsmeting volgens geslacht leerlingen

Geslacht	Steekproef 2016		Herhalingsmeting 2018	
	n	%	n	%
Meisje	2388	51.0	2081	51.4
Jongen	2291	49.0	1965	48.6

2.2.2.4 Selectie volgens aantal boeken thuis

Vervolgens bekeken we of de deelname van leerlingen aan de herhalingsmeting in 2018 verschilde naargelang de vijf categorieën van het aantal boeken thuis. **Tabel 16** toont voor beide afnamemomenten het aantal leerlingen per categorie en het percentage leerlingen per categorie in de steekproef van de herhalingsmeting. We vonden geen noemenswaardige verschuiving in de verhoudingen tussen de verschillende groepen tussen PIRLS 2016 en de herhalingsmeting in 2018.

Tabel 16: Deelname aan herhalingsmeting volgens aantal boeken thuis

Categorie	Steekproef 2016		Herhalingsmeting 2018	
	n	%	n	%
Zeer laag	377	8.1	277	6.8
Laag	1087	23.2	947	23.4
Gemiddeld	1799	38.4	1588	39.2
Hoog	770	16.5	695	17.2
Zeer Hoog	573	12.2	490	12.1

2.2.2.5 Thuistaal

Vervolgens bekeken we of de deelname van leerlingen aan de herhalingsmeting in 2018 verschilde naargelang de thuistaal². **Tabel 17** toont voor beide afnamemomenten het aantal leerlingen per categorie en het percentage van het aantal leerlingen per categorie in de steekproef van de herhalingsmeting. We vonden geen noemenswaardige verschuiving in de verhouding tussen de verschillende groepen tussen PIRLS 2016 en de herhalingsmeting in 2018.

Tabel 17: Deelname aan herhalingsmeting volgens thuistaal leerlingen (ASBG03)

Categorie	Steekproef 2016		Herhalingsmeting 2018	
	n	%	n	%
Altijd Nederlands	2986	63.8	2672	66.0
Meestal Nederlands	477	10.2	400	9.9
Soms Nederlands	939	20.1	767	19.0
Nooit Nederlands	209	4.5	162	4.0

2.2.3 Conclusie

In deze sectie onderzochten we of de vrijwillige deelname van scholen aan de herhalingsmeting in 2018 selectief is naargelang verschillende schoolkenmerken. Daarbij vonden we dat er geen selectie-effect is voor de verschillende schoolkenmerken. We onderzochten vervolgens welke leerlingen tussen het vierde en zesde leerjaar verdwijnen uit de deelnemende scholen aan de hand van verschillende leerlingkenmerken. Daarbij vonden we dat er geen selectie-effect is voor geslacht, boeken thuis of thuistaal. Leerlingen met lagere vaardigheid voor begrijpend lezen in PIRLS 2016 en een ander geboortjaar dan de normaalvorderende leerlingen verdwijnen vaker uit het onderzoek.

2.3 Vergelijking steekproef herhalingsmeting PIRLS in 2018 en Peilingsonderzoek lezen 2018

Voor elke steekproef is het primaire doel dat ze representatief is voor de populatie waaruit ze getrokken is. Als de steekproef representatief is, dan zijn de uitkomsten op basis van de steekproef ook geldig voor de populatie binnen een zekere foutenmarge. Bij de herhalingsmeting van PIRLS in 2018 richtten we ons op de populatie van leerlingen in het zesde leerjaar. PIRLS 2016 was gericht op een representatieve steekproef voor leerlingen in het vierde leerjaar. Aangezien de scholen van PIRLS2016 op vrijwillige basis deelnamen aan de herhalingsmeting in 2018, was het onze hoop dat de herhalingsmeting representatief was voor het zesde jaar. In de vorige sectie vonden we dat de niet-deelnemende scholen aan de herhalingsmeting van 2018 enigszins verschilden van de deelnemende scholen. Voor de vergelijking tussen de herhalingsmeting van PIRLS in 2018 en het Peilingsonderzoek lezen 2018 was het echter noodzakelijk dat beide studies dezelfde populatie, leerlingen in het zesde leerjaar, beschreven. Daarom onderzochten we hoe de steekproeven van deze herhalingsmeting in 2018 en het Peilingsonderzoek lezen 2018 verschilden.

Aan de herhalingsmeting van PIRLS in 2018 namen 126 lagere scholen correct deel. Aan het Peilingsonderzoek lezen 2018 namen 119 lagere scholen deel. We merken daarbij op dat negen scholen aan zowel de herhalingsmeting van PIRLS in 2018 als het Peilingsonderzoek lezen 2018 deelnamen. We gaan na wat de gelijkenissen en verschillen zijn tussen beide steekproeven aan de hand van verschillende schoolkenmerken. Beide studies namen immers een steekproef van scholen, terwijl alle leerlingen binnen het zesde leerjaar opgenomen werden. Het volstond dus enkel de steekproeven te vergelijken op schoolniveau. We vergeleken daarom de schoolkenmerken van beide steekproeven. We keken daarbij ook hoeveel leerlingen in het schooljaar 2017-2018 in de scholen aanwezig waren.

2 Deze informatie is afkomstig van de leerlingvragenlijst (ASBG03) die werd afgenomen bij PIRLS 2016. De leerling kreeg een meerkeuzevraag "Hoe vaak spreek je thuis Nederlands of Vlaams? Vul slechts één bolletje in: O Ik spreek thuis altijd Nederlands of Vlaams O Ik spreek thuis bijna altijd Nederlands of Vlaams, O Ik spreek thuis soms Nederlands of Vlaams en soms een andere taal, en O Ik spreek thuis nooit Nederlands of Vlaams.

De schoolkenmerken die we onderzochten waren:

- De provincie. Een school kon gelegen zijn de in de provincie Antwerpen, het Brussels Hoofdstedelijk gewest, Limburg, Oost-Vlaanderen, Vlaams-Brabant of West-Vlaanderen.
- Het onderwijsnet. Een school kon tot het gesubsidieerde vrij onderwijs behoren, het gemeenschapsonderwijs of het gesubsidieerd officieel onderwijs. In de beschrijving van de steekproef onderscheiden we enkel het vrij onderwijs en het officieel onderwijs.
- Percentage leerlingen met een laag opgeleide moeder (OKI). We verdeelden de scholen in vijf groepen volgens het percentage leerlingen dat aantikte op deze indicator. In de eerste groep tikt minder dan 10% aan (zeer laag), in de tweede groep tikt 10% tot 20% aan (laag), in de derde groep tikt 20% tot 30% aan (gemiddeld), in de vierde groep tikt 30% tot 50% aan (hoog), en in de vijfde groep tikt meer dan 50% aan (zeer hoog).
- Percentage leerlingen met een schooltoelage (OKI). We verdeelden de scholen in vijf groepen volgens het percentage leerlingen dat aantikte op deze indicator. In de eerste groep tikt minder dan 10% aan (zeer laag), in de tweede groep tikt 10% tot 20% aan (laag), in de derde groep tikt 20% tot 30% aan (gemiddeld), in de vierde groep tikt 30% tot 50,0% aan (hoog), en in de vijfde groep tikt meer dan 50% aan (zeer hoog).
- Percentage leerlingen met een gezinstaal anders dan het Nederlands (OKI). We verdeelden de scholen in vijf groepen volgens het percentage leerlingen dat aantikte op deze indicator. In de eerste groep tikt minder dan 10% aan (zeer laag), in de tweede groep tikt 10% tot 20% aan (laag), in de derde groep tikt 20% tot 30% aan (gemiddeld), in de vierde groep tikt 30% tot 50% aan (hoog), en in de vijfde groep tikt meer dan 50% aan (zeer hoog).

2.3.1.1 Provincie

Eerst bekeken we hoeveel scholen en leerlingen per provincie deelnamen aan de herhalingsmeting van PIRLS in 2018 en aan het Peilingsonderzoek lezen in 2018. **Tabel 18** toont voor beide studies het aantal scholen per provincie, het percentage scholen per provincie, het aantal leerlingen per provincie en het percentage leerlingen per provincie. We vonden dat in de herhalingsmeting van PIRLS in 2018 leerlingen uit Antwerpen meer vertegenwoordigd waren dan in het Peilingsonderzoek lezen 2018.

Tabel 18: Deelname aan herhalingsmeting volgens provincie

Provincie	Peiling 2018				PIRLS 2018				Populatie
	Scholen		Leerlingen		Scholen		Leerlingen		Leerlingen
	n	%	n	%	n	%	n	%	%
Antwerpen	32	26.9	729	23.4	32	25.4	1341	29.1	28.3
Brussels H.G.	7	5.9	130	4.2	6	4.8	134	2.9	4.4
Limburg	16	13.4	492	15.8	17	13.5	600	13.0	12.3
Oost-Vlaanderen	23	19.3	724	23.3	30	23.8	1064	23.1	22.6
Vlaams-Brabant	15	12.6	429	13.8	19	15.1	690	15.0	16.2
West-Vlaanderen	26	21.8	607	19.5	22	17.5	786	17.0	16.2
Totaal	119	100	3111	100	126	100.0	4615	100.0	100

2.3.1.2 Netten

Vervolgens bekeken we hoeveel scholen en leerlingen per onderwijsnet deelnamen aan de herhalingsmeting van PIRLS in 2018 en aan het Peilingsonderzoek lezen 2018. **Tabel 19** toont voor beide studies het aantal scholen per onderwijsnet, het percentage scholen per onderwijsnet, het aantal leerlingen per onderwijsnet en het percentage leerlingen per onderwijsnet. We vonden dat in de herhalingsmeting van PIRLS in 2018 leerlingen uit het vrije onderwijsnet meer vertegenwoordigd waren.

Tabel 19: Deelname aan herhalingsmeting volgens onderwijsnet

Onderwijsnet	Peiling 2018				PIRLS 2018				Populatie
	Scholen		Leerlingen		Scholen		Leerlingen		Leerlingen
	n	%	n	%	n	%	n	%	%
Officieel	50	42.0	1210	38.9	44	34.9	1422	30.8	38.50
Vrij	69	58.0	1901	61.1	82	65.1	3193	69.2	61.50

2.3.1.3 Percentage leerlingen met laagopgeleide moeder

Vervolgens bekeken we of de deelname van scholen en leerlingen verschilde tussen beide studies naargelang het percentage leerlingen met een laagopgeleide moeder (Tabel 20). We vonden dat in de herhalingsmeting van PIRLS in 2018 leerlingen in de zeer lage categorie meer vertegenwoordigd waren dan in het Peilingsonderzoek lezen 2018. De leerlingen in de zeer hoge categorie waren dan weer minder vertegenwoordigd dan in het Peilingsonderzoek lezen 2018.

Tabel 20: Deelname aan herhalingsmeting volgens percentage leerlingen met laagopgeleide moeder

Categorie	Peiling 2018				PIRLS 2018				Populatie
	Scholen		Leerlingen		Scholen		Leerlingen		Leerlingen
	n	%	n	%	n	%	n	%	%
Zeer laag	29	24.4	725	23.3	44	34.9	1749	37.9	29.6
Laag	52	43.7	1539	49.5	41	32.5	1561	33.8	34.7
Gemiddeld	17	14.3	391	12.6	20	15.9	647	14.0	15.4
Hoog	14	11.8	313	10.1	16	12.7	544	11.8	12.2
Zeer Hoog	7	5.9	143	4.6	5	4.0	114	2.5	8.2

2.3.1.4 Percentage leerlingen met schooltoelage

Vervolgens bekeken we of de deelname van scholen en leerlingen verschilde tussen beide studies naargelang het percentage leerlingen met recht op een schooltoelage (Tabel 21). We vonden dat in de herhalingsmeting van PIRLS in 2018 leerlingen in de hoge categorie meer vertegenwoordigd waren dan in het Peilingsonderzoek lezen 2018. Leerlingen in de gemiddelde categorie waren dan weer minder vertegenwoordigd dan in het Peilingsonderzoek lezen 2018.

Tabel 21: Deelname aan herhalingsmeting volgens percentage leerlingen met recht op een schooltoelage

Categorie	Peiling 2018				PIRLS 2018				Populatie
	Scholen		Leerlingen		Scholen		Leerlingen		Leerlingen
	n	%	n	%	n	%	n	%	%
Zeer laag	23	19.3	633	20.3	26	20.6	1007	21.8	21.5
Laag	49	41.2	1356	43.6	52	41.3	1945	42.1	40.2
Gemiddeld	24	20.2	713	22.9	22	17.5	757	16.4	14.8
Hoog	12	10.1	175	5.6	18	14.3	571	12.4	14.0
Zeer Hoog	11	9.2	234	7.5	8	6.3	335	7.3	9.6

2.3.1.5 Percentage leerlingen met andere thuistaal

Vervolgens bekeken we of de deelname van scholen en leerlingen verschilde tussen beide studies naargelang het percentage leerlingen met een andere thuistaal (Tabel 22). Tussen beide studies vonden we geen grote verschillen in het percentage scholen per categorie en het percentage leerlingen per categorie. We vonden dat in de herhalingsmeting van PIRLS in 2018 leerlingen in de lage categorie minder vertegenwoordigd waren dan in het Peilingsonderzoek lezen 2018.

Tabel 22: Deelname aan herhalingsmeting volgens percentage leerlingen met een andere thuistaal

Categorie	Peiling 2018				PIRLS 2018				Populatie
	Scholen		Leerlingen		Scholen		Leerlingen		Leerlingen
	n	%	n	%	n	%	n	%	%
Zeer laag	61	51.3	1587	51.0	61	48.4	2230	48.3	47.7
Laag	19	16.0	473	15.2	24	19.0	963	20.9	18.9
Gemiddeld	14	11.8	391	12.6	12	9.5	500	10.8	9.3
Hoog	14	11.8	392	12.6	13	10.3	445	9.6	10.8
Zeer Hoog	11	9.2	268	8.6	16	12.7	477	10.3	13.3

2.3.1.6 Conclusie

In deze sectie vergeleken we de steekproef van de herhalingsmeting van PIRLS in 2018 en de steekproef van het Peilingonderzoek Nederlands 2018 met de populatiegegevens. Omdat beiden een steekproef van scholen zijn, alle leerlingen van het zesde leerjaar werden steeds opgenomen, werd de representativiteit aan de hand van verschillende schoolkenmerken onderzocht. We vonden dat beide steekproeven voor verschillende schoolkenmerken beperkte verschillen hebben ten opzichte van elkaar en van de populatie.

2.4 Schatting standaardfout

2.4.1 PIRLS 2016

Omdat PIRLS 2016 gebruik maakt van een steekproef om de resultaten voor het begrijpend lezen in verschillende landen te beschrijven is er een standaardfout (Mullis & Martin, 2015). De standaardfout geeft aan hoe onzeker men is over hoe goed de steekproef de populatie van leerlingen in een land beschrijft. Deze onzekerheid is er omdat de volledige populatie van leerlingen in dat land niet gemeten wordt. Om de standaardfout te berekenen moeten er geweten zijn hoe de steekproeftrekking verlopen is.

In Vlaanderen gebruikt PIRLS 2016 een steekproef van scholen. Er worden namelijk scholen uit de populatie van scholen getrokken die, zo goed als mogelijk, representatief zijn voor alle scholen in Vlaanderen. De steekproeftrekking gebeurt daarbij op gestratificeerde wijze. Dit wil zeggen dat er op populatieniveau geweten was hoeveel leerlingen in elke provincie wonen, hoeveel leerlingen behoren tot verschillende categorieën van sociaaleconomische achtergrond, en hoeveel jongens en meisjes er zijn. De populatie van scholen wordt daarom ingedeeld in verschillende groepen naargelang de gemiddelde leerlingkenmerken van scholen. Deze groepen worden strata genoemd. Binnen elk van deze strata worden dan een steekproef op toeval getrokken. De scholen worden echter zo getrokken dat elk stratum even goed vertegenwoordigd is in de steekproef als in de populatie. Op deze manier zijn de op voorhand gekende kenmerken van de Vlaamse populatie van scholen onvertekend vertegenwoordigd in de steekproef. Er is echter nog steeds een standaardfout omdat binnen elk stratum de steekproeftrekking nog op toeval berust.

We merken op dat in de rapporten van PIRLS vaak gesproken wordt over “*multistage sampling*”, omdat de steekproeftrekking in veel landen in twee ‘*stages*’ (stappen) verliep. Zo werd er in sommige landen in een eerste stap een steekproeftrekking van scholen gemaakt, terwijl er in een tweede stap een steekproeftrekking was van leerlingen in het vierde leerjaar binnen scholen. In de Vlaamse steekproef was er echter maar één stap, de steekproeftrekking van scholen, want alle leerlingen van het vierde leerjaar binnen een geselecteerde school werden steeds opgenomen.

Om de standaardfout te berekenen wordt de methode van *Jackknife Repeated Replications* (JRR) gebruikt. De basis van dit idee is dat de verschillende onderscheiden strata van scholen bepalen hoe groot de standaardfout is. Daarom worden paren van de meest vergelijkbare scholen in één ‘zone’ geplaatst. Vervolgens wordt berekend hoe hard het gemiddelde van de uitkomst verandert als één school uit de zone verwijderd wordt en de andere school in die zone een dubbel gewicht krijgt. De school die verwijderd wordt krijgt dus een replicatiegewicht van 0 en de school die overblijft krijgt een replicatiegewicht van 2. Dit wordt voor iedere zone gedaan. Op basis hiervan kan geschat worden hoe gevoelig een statistiek is voor het verwijderen van een school van de verschillende zones. Dit geeft dan ook de standaardfout weer. De formule voor het toekennen van een replicatiegewicht k_{hj} voor elke school binnen een zone is dus relatief eenvoudig:

$$\begin{aligned} k_{hj} &= 2 \text{ voor leerlingen in school } j \text{ van zone } h \\ &= 0 \text{ voor leerlingen in de andere school van zone } h \\ &= 1 \text{ voor leerlingen in een andere zone} \end{aligned}$$

Dit wordt 150 keer gedaan, want elk van de twee scholen binnen een zone wordt eenmaal een replicatiegewicht van 0 gegeven. Zo worden 150 sets van *replicate sampling weights* geschat:

$$W_{hji} = k_{hj} * W_{0i}$$

De term W_{0i} is het algemene steekproefgewicht van leerling i . De term W_{hji} is het resulterende *replicate sampling weight* van leerling i in zone h wanneer school j in de steekproef een replicatiegewicht van 2 krijgt en de andere school van het paar een replicatiegewicht van 0 krijgt. Voor elke leerling zijn er 150 van zo'n *replicate sampling weights*, want er zijn 150 scholen die verwijderd kunnen worden.

Voor elke statistiek die berekend wordt worden deze 150 gewichten gebruikt. Zo wordt elke statistiek 150 keer geschat, waarbij elk *replicate sampling weight* éénmaal gebruikt wordt. De variantie tussen de waarde van de statistiek geeft vervolgens de standaardfout weer. Dit gebeurt volgens de volgende formule:

$$Var_{jrr}(t_0) = \frac{1}{2} \sum_{h=1}^{75} \sum_{j=1}^2 (t_{hj} - t_0)^2$$

De term t_0 geeft de statistiek weer zoals ze geschat is met de algemene steekproefgewichten W_{0i} . De term t_{hj} geeft dezelfde statistiek weer zoals ze geschat is op basis van de set van *replicate sampling weights* W_{hij} zoals ze bekomen is in de zone h ($h=1, \dots, 75$), waar de j de school (eerste of tweede) in de zone zit en de andere verwijderd is. Deze formule is de kern van de *Jackknife Repeated replications* methode en $Var_{jrr}(t_0)$ geeft daarbij weer wat de variantie is tussen de schattingen. De vierkantswortel hiervan is gelijk aan de standaardfout.

2.4.2 Herhalingsmeting PIRLS 2018

In de herhalingsmeting van PIRLS in 2018 kan de oorspronkelijke structuur van *zones* gebaseerd op de *strata* niet meer gebruikt worden. Hierdoor kunnen de replicate sampling weights ook niet meer gebruikt worden. Aangezien 32 van de 158 scholen van de oorspronkelijke steekproef niet meer deelnemen is de indeling volgens zones niet meer correct. 53 zones kunnen behouden blijven zoals ze oorspronkelijk waren gebruikt in PIRLS 2016. Vervolgens zijn er 20 zones waar slechts één school overblijft en twee zones waar geen school overblijft, voor deze zones kunnen dus geen replicatiegewichten toegekend worden. Daarom is het noodzakelijk de zones terug in te delen. Om het opzet van PIRLS te behouden wordt ervoor gekozen om zones met slechts één school te fuseren. Hierbij worden de zones gefuseerd volgens de achtergrondkenmerken van de scholen, de meest vergelijkbare scholen werden zo in één zone geplaatst. Of scholen vergelijkbaar zijn wordt bepaald op basis de sociaaleconomische kenmerken die PIRLS 2016 gebruikte bij de steekproeftrekking. Het resultaat hiervan zijn 63 zones (met twee scholen per zone) op basis waarvan de *replicate sampling weights* berekend worden. Op basis hiervan wordt ook de standaardfout berekend. Deze manier om de standaardfout te schatten gebruiken we in [Hoofdstuk 5](#) en [Hoofdstuk 6](#).

3 IRT-analyse herhalingsmeting 2018

3.1 Item respons theorie

>> Vooralere we IRT gebruiken om de herhalingsmeting van PIRLS in 2018 op de meetschaal van PIRLS 2016 of de meetschaal van het Peilingsonderzoek te leggen is het noodzakelijk te weten wat IRT is. IRT kan begrepen worden als een familie van statistische modellen die beschrijven hoe één of meer onderliggende vaardigheden gerelateerd zijn aan de antwoorden van leerlingen op verschillende items. In de meest simpele vorm is er één onderliggende vaardigheid die bepaalt wat de kans is dat het antwoord van een leerling op een item juist is. Hoe een onderliggende vaardigheid gerelateerd is aan de antwoordrespons is dus het voornaamste voorwerp van onderzoek bij IRT-analyses.

De IRT-modellen laten ons ook toe te beschrijven hoe betrouwbaar de items van een toets de onderliggende vaardigheid meten. Deze betrouwbaarheid is daarbij een functie van hoe goed alle items een onderliggende vaardigheid meten, en ook hoe moeilijk deze items zijn. De IRT-modellen laten ons ook toe te beschrijven hoe goed de antwoordresponsen te herleiden zijn tot één of meer onderliggende vaardigheden. Dit is een belangrijk aspect bij het onderzoek naar validiteit, meet de toets wat we willen meten.

We beschrijven eerst hoe IRT de betrouwbaarheid en validiteit van meetinstrumenten operationaliseert. Vervolgens beschrijven we de verschillende IRT-modellen die van belang zijn voor de herhalingsmeting van PIRLS in 2018. Nadien beschrijven we hoe we IRT-modellen kunnen toepassen om de herhalingsmeting van PIRLS in 2018 op de meetschaal van PIRLS 2016 te leggen. Aansluitend beschrijven we hoe we IRT-modellen kunnen toepassen om de herhalingsmeting van PIRLS in 2018 op de meetschaal van het Peilingsonderzoek Nederlands begrijpend lezen 2018 te leggen. Afsluitend beschrijven we hoe we de factorscores van verschillende concepten die werden gemeten bij de herhalingsmeting van PIRLS in 2018 op de meetschaal van PIRLS 2016 plaatsen. Een factorscore is het geschatte niveau, uitgedrukt in een numerieke score, dat een leerling heeft voor bepaalde attitude.

3.2 IRT-modellen

IRT is gericht op hoe de vaardigheid van leerlingen gerelateerd is aan de kans op het geven van een juist(er) antwoord op een item (Embretson & Reise, 2000). De kans om een item juist(er) te beantwoorden neemt daarbij toe naargelang de vaardigheid van een leerling toeneemt. Binnen IRT zijn er verschillende mogelijke modellen om de vaardigheid en de kans op een juist(er) antwoord op een item te relateren. In de volgende paragrafen beschrijven we de volgende modellen: het 2-parameter model, het 3-parameter model en het *generalized partial credit* model. In de volgende paragrafen bespreken we deze drie modellen zoals ze zijn toegepast in deze studie.

Het 2-parameter model omvat twee parameters per item: de interceptparameter en de discriminatieparameter. De interceptparameter van een item geeft aan hoe gemakkelijk het item is. Hoe hoger de waarde van deze parameter, hoe makkelijk het item. In dit model geldt dat hoe groter de vaardigheid van een respondent, hoe groter de kans op een juist antwoord. Anderzijds, hoe lager de vaardigheid van een respondent, hoe lager de kans op een juist antwoord. De discriminatieparameter geeft de sterkte van het verband tussen de kans op een juist antwoord en de onderliggende vaardigheid weer. Deze parameter geeft met andere woorden aan hoe goed het item discrimineert tussen personen met een verschillend vaardigheidsniveau. Zo maakt een item met een hoge discriminatieparameter een duidelijk onderscheid tussen leerlingen met een hogere dan wel lagere vaardigheid. Daardoor geeft de interceptparameter in dit model slechts gemiddeld genomen aan hoe gemakkelijk het item is. We merken op dat in andere parameterisaties van het 2-parameter model de negatieve interceptparameter gedeeld wordt door de discriminatieparameter ($-b/a$). De resulterende parameter noemt men dan de moeilijkheidsparameter.

Hieronder staat het 2-parameter model dat we in dit onderzoek gebruikten:

$$P(x_i = 1 | \theta_j, a_i, b_i) = \frac{1}{1 + \exp(-a_i * \theta_j + b_i)}$$

x_i = of het antwoord op item i juist (1) of fout (0) is

θ_j = de vaardigheid van de leerling op schaal j

a_i = de discriminatieparameter van item i

b_i = het intercept van het item i

Het 3-parameter model kan beschouwd worden als een uitbreiding van het 2-parameter model, met drie parameters per item. Aan dit model wordt een gokparameter toegevoegd bij meerkeuzevragen. De gokparameter dient als een correctie voor het gokgedrag van leerlingen. De voorgaande modellen stellen immers dat bij een lage vaardigheid de kans op een juist antwoord nul benadert. Dit is waarschijnlijk incorrect bij meerkeuzevragen waar leerlingen kunnen gokken. De gokparameter beschrijft dus de kans dat een leerling een vraag juist beantwoordt wanneer deze leerling te weinig vaardig is om het juiste antwoord te kennen.

$$P(x_i = 1 | \theta_j, a_i, b_i, y_i) = y_i + \frac{(1 - y_i)}{1 + \exp(-a_i * \theta_j + b_i)}$$

x_i = of het antwoord op item i juist (1) of fout (0) is

θ_j = de vaardigheid van de leerling op schaal j

a_i = de discriminatieparameter van item i

b_i = het intercept van het item i

y_i = de gokparameter van het item i

Het *generalized partial credit* model kan beschouwd worden als een uitbreiding van het 2-parameter model. Zo is het 2-parameter model gericht op het analyseren van items die juist of fout kunnen zijn en respectievelijk een waarde 1 of 0 hebben. Het *generalized partial credit* model is gericht op items die verschillende scores kunnen hebben. In de toetsen die hier geanalyseerd worden zijn er items waarvoor leerlingen een score 2, 1 of 0 kunnen krijgen en items waarvoor leerlingen een score 3, 2, 1 of 0 kunnen krijgen. Hoe hoger de score, hoe juister het antwoord op het item. Het *generalized partial credit* model beschrijft hoe elk van de drie of vier antwoordcategorieën samenhangen met de onderliggende vaardigheid. Voor een item met vier antwoordcategorieën zijn er vier parameters. Er is een parameter die beschrijft vanaf welke onderliggende vaardigheid de kans kleiner is dan 50% om een waarde 3 te hebben, er is een parameter die beschrijft vanaf welke onderliggende vaardigheid de kans kleiner is dan 50% om een waarde 2 te hebben, en er is een parameter die beschrijft vanaf welke onderliggende vaardigheid de kans kleiner is dan 50% om een waarde 1 te hebben. Er is ook een discriminatieparameter die aangeeft wat de sterkte is van het verband tussen de antwoordcategorieën en de onderliggende vaardigheid. Deze parameter geeft met andere woorden aan hoe goed het item discrimineert tussen personen met een verschillende vaardigheidsniveau. Zo maakt een item met een hoge discriminatieparameter een duidelijk onderscheid tussen leerlingen met een hogere dan wel lagere vaardigheid.

$$P(x_i = k | \theta_j, a_i, d_{i,0}, \dots, d_{i,m_i-1}, k_{i,0}, \dots, k_{i,m_i-1}) = \frac{\exp(a_i k_{i,k} * (a_i * \theta_j) + d_{i,k})}{\sum_{k=0}^{m_i-1} \exp(a_i k_{i,k} * (a_i * \theta_j) + d_{i,k})}$$

x_i = of de antwoordcategorie k is op het item i . Bij 3 antwoordcategorieën is k 0, 1 of 2, waarbij $m_i = 3$ het totale aantal antwoordcategorieën beschrijft. Bij 4 antwoordcategorieën is k 0, 1, 2 of 3, waarbij $m_i = 4$ het totale aantal antwoordcategorieën beschrijft.

θ_j = de vaardigheid van de leerling op schaal j

a_i = de discriminatieparameter van item i

$b_{i,v}$ = de locatieparameter van categorie v van het item i . $d_{(i,0)} = 0$ voor identificatie van het model. $a_i * (d_{i,v-1} - d_{i,v})$ beschrijft voor welke θ_j de curve van categorie $v-1$ de curve van categorie v snijdt.

$k_{i,v}$ = de volgorde van categorie v van het item i . Bij 3 antwoordcategorieën ($m_i = 3$) is $ak_0 = 0$, $ak_1 = 1$ en $ak_2 = 2$. Bij 4 antwoordcategorieën ($m_i = 4$) is $ak_0 = 0$, $ak_1 = 1$, $ak_2 = 2$ en $ak_3 = 3$

Het 2-parameter model, 3-parameter model en *generalized partial credit* model zoals ze hier besproken zijn veronderstellen dat er slechts één onderliggende vaardigheid verantwoordelijk is voor de antwoorden van leerlingen. Dit is een veronderstelling van unidimensionaliteit. Het is echter mogelijk dat meerdere onderliggende vaardigheden de kans op het juist invullen van een item veroorzaken. Indien dit het geval is zijn multidimensionele modellen aangewezen. Indien er meerdere onderliggende vaardigheden zijn, dan zal een IRT-model met één onderliggende vaardigheid geen goede beschrijving geven van de testresultaten. Er is echter veel evidentie dat de items binnen PIRLS 2016 één onderliggende vaardigheid meten en dat de items binnen het Peilingsonderzoek lezen 2018 één onderliggende vaardigheid meten. We weten echter niet of beide studies dezelfde vaardigheid in begrijpend lezen meten. Het is mogelijk dat er multidimensionaliteit is wanneer de items van beide studies in één toets gebruikt worden. In [Hoofdstuk 7](#) wordt deze vraag onderzocht. Voorlopig veronderstellen we dat de items van zowel PIRLS 2016 als het Peilingsonderzoek lezen 2018 dezelfde onderliggende vaardigheid meten.

Samenvattend stellen we dat IRT veel mogelijkheden biedt om de relatie tussen de onderliggende vaardigheid en de kans op een juist antwoord op een item te beschrijven. De keuze voor een bepaald model is daarbij deels afhankelijk van hoe de toets en items ontworpen zijn.

3.3 Toetsen op één meetschaal plaatsen

In deze sectie beschrijven we hoe verschillende toetsen op één meetschaal geplaatst kunnen worden. Dit is noodzakelijk om te begrijpen hoe de herhalingsmeting van PIRLS in 2018 zowel op de meetschaal van PIRLS 2016 als op de meetschaal van het Peilingsonderzoek lezen 2018 geplaatst werd.

Er zijn meerdere manieren om verschillende toetsen op één meetschaal te plaatsen (voor een overzicht zie Dorans, Moses, & Eignor, 2010). In deze tekst concentreren we ons op het gebruik van zogenaamde ankeritems. Dit zijn items die in meerdere toetsen gebruikt worden. Door deze overlap van items kunnen toetsen op één meetschaal geplaatst worden. Bij de samenstelling van de toetsen van de herhalingsmeting van PIRLS in 2018 werden voornamelijk items opgenomen van PIRLS 2016. Deze items kunnen dus als ankeritems gebruikt worden om de herhalingsmeting van PIRLS in 2018 en PIRLS 2016 op één meetschaal te plaatsen. Er werden echter ook items opgenomen van het Peilingsonderzoek lezen 2018. Deze items kunnen dus als ankeritems gebruikt worden om de herhalingsmeting van PIRLS in 2018 en het Peilingsonderzoek lezen 2018 op één meetschaal te plaatsen.

Ankeritems moeten aan verschillende voorwaarden voldoen vooraleer twee of meer toetsen op een betrouwbare wijze op één meetschaal geplaatst worden. Zo wordt aangeraden dat minstens 20% van het totaal aantal items per toets ankeritems zijn (Kolen & Brennan, 2014 p. 288). Daarnaast moeten deze ankeritems de verschillende inhouden van de onderliggende vaardigheid voldoende meten (Li, Jiang, & von Davier, 2012). Vaak wordt gesteld dat de ankeritems een 'kleine test' op zich moeten vormen. De gekozen ankeritems mogen ook niet te moeilijk of te gemakkelijk zijn voor de leerlinggroepen die de verschillende toetsen invullen. Zowel de overlap tussen de herhalingsmeting van PIRLS in 2018 met PIRLS 2016, of de overlap tussen de herhalingsmeting van PIRLS in 2018 met het Peilingsonderzoek lezen 2018 voldoet aan deze voorwaarden.

Een bijkomende veronderstelling, vaak is deze niet expliciet vermeld, is dat beide toetsen ook dezelfde onderliggende vaardigheid meten. Als beide toetsen niet dezelfde onderliggende vaardigheid meten, dan kunnen beide toetsen niet op een betekenisvolle manier op dezelfde meetschaal gelegd worden. Om succesvol op één meetschaal geplaatst te worden moeten beide toetsen dezelfde vaardigheid meten. In [Hoofdstuk 7](#) wordt deze vraag onderzocht.

Door de aanwezigheid van ankeritems in de herhalingsmeting van PIRLS in 2018 kunnen we de testresultaten op dezelfde meetschaal plaatsen als deze van PIRLS 2016 en het Peilingsonderzoek lezen 2018. Om de meetschaal van PIRLS 2016 te hanteren moeten de oorspronkelijke parameterwaarden voor de ankeritems van het PIRLS 2016 gebruikt worden. Om de meetschaal van het Peilingsonderzoek lezen 2018 te hanteren moeten de oorspronkelijke parameterwaarden voor de ankeritems van het Peilingsonderzoek lezen 2018 gebruikt worden. De volgende secties tonen hoe de oorspronkelijke parameterwaarden gebruikt werden om de herhalingsmeting van PIRLS in 2018 te plaatsen op de meetschaal van PIRLS 2016 en op de meetschaal van het Peilingsonderzoek lezen 2018.

3.4 De PIRLS-meetschaal

3.4.1 Parameters van PIRLS 2016

In deze sectie beschrijven we hoe de vaardigheidsschattingen van de leerlingen die deelnamen aan de herhalingsmeting van PIRLS in 2018 op de internationale PIRLS-schaal werden geplaatst. Dit is de meetschaal waarop zowel PIRLS 2006 als PIRLS 2016 (Mullis & Martin, 2015) gesitueerd zijn. In Appendix 1 wordt de syntax weergegeven van de analyses in de gebruikte statistische software, MIRT 1.30 (Chalmers, 2012) in R 3.6.0 (R Core Team, 2018).

Het schatten van de vaardigheidsscores voor de herhalingsmeting startte met het schatten van een IRT-model op basis van de testgegevens van de testgroep. Doorgaans worden in zo'n IRT-model de itemparameters vrij geschat, terwijl het gemiddelde en de standaarddeviatie van de testgroep respectievelijk op 0 en 1 vastgezet worden. Omdat we echter op basis van de bestaande PIRLS-schaal werkten werden de itemparameters vastgezet volgens de parameterwaarden van de internationale PIRLS-schaal. Het gemiddelde en de standaarddeviatie van de steekproef werden dus vrij geschat op basis van de itemparameters en de testgegevens. De waarden van de itemparameters van de items die afkomstig zijn van het Peilingsonderzoek lezen 2018 werden wel vrij geschat.

Merk op dat de itemparameters van de internationale PIRLS-schaal niet direct toepasbaar waren binnen onze statistische software. Hoewel PIRLS de 2-parameter, 3-parameter- en *generalized partial credit* modellen gebruikt zoals beschreven in onze sectie over IRT, is de parameterisatie verschillend. Zo wordt binnen het PIRLS-onderzoek het 2-parameter model op de volgende wijze beschreven:

$$P(x_i = 1 | \theta_j, a_i, b_i) = \frac{1}{1 + \exp(-1.7 * a_i * (\theta_j - b_i))}$$

Er zijn twee verschillen tussen het 2-parameter model zoals het gedefinieerd wordt in PIRLS en het 2-parameter model zoals het gedefinieerd wordt in deze studie. Zo wordt binnen PIRLS een schaalconstante gebruikt met een waarde van "1.7" waarmee de discriminatieparameter vermenigvuldigd wordt. De reden hiervoor is dat de IEA de parameters van het logistische model zou kunnen vergelijken met de parameters van het probit model. Verder wordt er binnen PIRLS gewerkt met " $-a_i * (\theta_j - b_i)$ " in de plaats van " $-a_i * \theta_j + b_i$ " waardoor de waarde van de interceptparameter afhankelijk wordt van de waarde van discriminatieparameter tijdens de schattingsprocedure. Bijgevolg moeten de parameters van de PIRLS-schaal voor het 2-parameter model eerst getransformeerd worden voor onze parameterisatie van het 2-parameter model. Om de discriminatieparameter " a_i " voor onze parameterisatie te bekomen moeten we $a_i * 1.7$ berekenen volgens de PIRLS parameterisatie. Om het intercept b_i voor onze parameterisatie te bekomen moeten we " $a_i * 1.7 * (-b_i)$ " berekenen volgens de PIRLS parameterisatie. De gokparameter " y_i " is gelijk in beide parameterisaties.

Bij het PIRLS-onderzoek wordt het 3-parameter model op de volgende wijze beschreven:

$$P(x_i = 1 | \theta_j, a_i, b_i, y_i) = y_i + \frac{(1 - y_i)}{1 + \exp(-1.7 * a_i * (\theta_j - b_i))}$$

Er zijn twee verschillen tussen het 3-parameter model zoals het gedefinieerd wordt in PIRLS en het 3-parameter model zoals het gedefinieerd wordt in deze studie. Zo wordt binnen PIRLS een schaalconstante gebruikt met een waarde van "1.7" waarmee de discriminatieparameter vermenigvuldigd wordt. De reden hiervoor is dat de IEA de parameters van het logistische model wil vergelijken met de parameters van het probit model. Verder wordt er binnen PIRLS gewerkt met " $-a_i * (\theta_j - b_i)$ " in de plaats van " $-a_i * \theta_j + b_i$ " waardoor de waarde van de interceptparameter afhankelijk wordt van de waarde van discriminatieparameter tijdens de schattingsprocedure. Bijgevolg moeten de parameters van de PIRLS-schaal voor het 3-parameter model eerst getransformeerd worden voor onze parameterisatie van het 3-parameter model. Om de discriminatieparameter " a_i " voor onze parameterisatie te bekomen moeten we $a_i * 1.7$ berekenen volgens de PIRLS parameterisatie. Om het intercept b_i voor onze parameterisatie te bekomen moeten we " $a_i * 1.7 * (-b_i)$ " berekenen volgens de PIRLS parameterisatie. De gokparameter " y_i " is gelijk over beide parameterisaties.

Binnen het PIRLS-onderzoek wordt het *generalized partial credit* model voor de toetsgegevens op de volgende wijze beschreven:

$$P(x_i = k | \theta_i, a_i, d_{i,1}, \dots, d_{i,m_i-1}, k_{i,0}, \dots, k_{i,m_i-1}) = \frac{\exp(\sum_{k=0}^y 1.7 * a_i(\theta_i - b_i + d_{i,k}))}{\sum_{g=0}^{m_i-1} \exp(\sum_{k=0}^g 1.7 * a_i(\theta_i - b_i + d_{i,k}))}$$

Er zijn meerdere verschillen tussen het *generalized partial credit* model zoals het gedefinieerd wordt in PIRLS en het *generalized partial credit* model zoals het gedefinieerd wordt in deze studie. Zo wordt binnen PIRLS een schaalconstante gebruikt met een waarde van “1.7” waarmee de discriminatieparameter vermenigvuldigd wordt. De reden hiervoor is dat de IEA de parameters van het logistische model wil vergelijken met de parameters van het probit model.

Verder wordt er binnen PIRLS gewerkt met de natuurlijke exponent van een som van deelfuncties, “ $\sum_{k=0}^g 1.7 * a_i(\theta_i - b_i + d_{i,k})$,” terwijl er in onze studie wordt gewerkt met de natuurlijke exponent van een deelfunctie, “ $\exp(a_i k_{i,k} * (a_i * \theta_i) + d_{i,k})$.” Bijgevolg moeten de parameters van de PIRLS-schaal voor het *generalized partial credit* model eerst getransformeerd worden voor onze parameterisatie van het *generalized partial credit* model. Om de discriminatieparameter “ a_i ” voor onze parameterisatie te bekomen moeten we $a_i * 1.7$ berekenen volgens de PIRLS parameterisatie. De locatieparameter $d_{0,k}$ heeft de waarde 0 over beide parameterisaties. Om de locatieparameter $d_{i,1}$ voor onze parameterisatie te bekomen moeten we “ $-1.7a_i * (b_i - d_{i,1})$ ” berekenen volgens de PIRLS parameterisatie. Om de locatieparameter $d_{i,2}$ voor onze parameterisatie te bekomen moeten we “ $-1.7a_i * ((b_i - d_{i,1}) + (b_i - d_{i,2}))$ ” berekenen volgens de PIRLS parameterisatie. Om de locatieparameter $d_{i,3}$ voor onze parameterisatie te bekomen moeten we “ $-1.7a_i * ((b_i - d_{i,1}) + (b_i - d_{i,2}) + (b_i - d_{i,3}))$ ” berekenen volgens de PIRLS parameterisatie.

Tabel 23 toont de waarden van de itemparameters zoals ze beschreven zijn in het internationale PIRLS-onderzoek. Vervolgens worden de waarden getoond van de itemparameters na de transformatie voor het gebruik in MIRT. Dit zijn de waarden die vastgezet zijn in het IRT-model dat diende om de vaardigheidsscores van de herhalingsmeting van PIRLS in 2018 op de internationale PIRLS schaal te schatten.

Tabel 23: Itemparameters zoals gerapporteerd door het PIRLS-onderzoek en getransformeerd voor MIRT

Item	PIRLS					MIRT					
	a_i	b_i	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$	a_i	b_i	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$	y_i
L021B01C	0.677	-2.530				1.151	2.912				
L021B02M	0.958	-2.353				1.629	3.832				0.101
L021B03M	0.721	-2.406				1.226	2.949				0.250
L021B04C	0.706	-1.260				1.200	1.512				
L021B05M	0.883	-1.315				1.501	1.974				0.223
L021B06M	0.440	-1.847				0.748	1.382				0.250
L021B07C	0.745	-1.160				1.267	1.469				
L021B08C	0.840	-1.359				1.428	1.941				
L021B09C	0.855	-1.140				1.454	1.657				
L021B10M	0.660	-0.338				1.122	0.379				0.192
L021B11M	0.979	-1.224				1.664	2.037				0.186
L021B12M	0.738	-2.202				1.255	2.763				0.127
L021B13C	0.542	-0.967	0.693	-0.693		0.921		1.530	1.782		
L021B14M	1.075	-0.597				1.828	1.091				0.149
L021B15C	0.419	0.318				0.712	-0.227				
L021B16C	0.435	-0.185	0.092	-0.092		0.740		0.205	0.274		
L021B17C	0.705	-0.414	0.684	-0.684		1.199		1.316	0.992		
L021E01C	0.796	-3.130				1.353	4.236				
L021E02C	0.771	-2.116				1.311	2.773				
L021E03M	1.343	-0.521				2.283	1.189				0.270
L021E04M	1.106	-0.378				1.880	0.711				0.262
L021E05M	1.118	-1.568				1.901	2.980				0.242
L021E06M	0.936	-2.465				1.591	3.922				0.304
L021E07C	0.463	-0.810	-0.018	0.018		0.787		0.623	1.275		
L021E08M	1.023	-0.749				1.739	1.303				0.201
L021E09M	0.620	0.464				1.054	-0.489				0.340
L021E10C	0.859	-2.716				1.460	3.966				

Item	PIRLS					MIRT					
	a_i	b_i	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$	a_i	b_i	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$	y_i
L021E11M	1.205	-0.817				2.049	1.674				0.226
L021E12C	1.116	-2.136				1.897	4.052				
L021E13C	0.528	-0.275				0.898	0.247				
L021E14C	0.493	0.186	0.852	-0.852		0.838		0.558	-0.312		
L021E15C	0.795	-2.219				1.352	2.999				
L021E16C	0.706	-0.959				1.200	1.151				
L021E17M	1.047	-0.467				1.780	0.831				0.171
R011F01M	1.334	-0.627				2.268	1.422				0.148
R011F02M	0.666	-0.848				1.132	0.960				0.243
R011F03M	0.920	-0.666				1.564	1.042				0.157
R011F04M	1.307	-0.831				2.222	1.846				0.228
R011F05M	0.940	-0.255				1.598	0.407				0.217
R011F06C	0.776	-0.152				1.319	0.201				
R011F07C	0.503	0.375	-0.896	0.896		0.855		-1.087	-0.641		
R011F08C	1.149	-0.328				1.953	0.641				
R011F09C	1.011	-0.627	0.074	-0.074		1.719		1.205	2.155		
R011F10C	0.846	-1.419				1.438	2.041				
R011F11M	0.739	0.208				1.256	-0.261				0.192
R011F12C	0.618	0.642	-0.471	0.471		1.051		-1.169	-1.349		
R011F13M	1.124	-0.199				1.911	0.380				0.270
R011L01M	0.532	-2.275				0.904	2.058				0.146
R011L02M	0.749	0.685				1.273	-0.872				0.241
R011L03C	0.616	-0.474				1.047	0.496				
R011L04C	0.667	0.418	1.643	-1.051	-0.592	1.134		1.389	-0.277	-1.422	
R011L05M	1.186	0.352				2.016	-0.710				0.206
R011L06C	0.656	0.191				1.115	-0.213				
R011L07M	0.772	0.474				1.312	-0.622				0.154
R011L08C	0.801	0.612	0.703	-0.703		1.362		0.124	-1.667		
R011L09M	0.963	-0.809				1.637	1.324				0.226
R011L10C	0.732	0.681	0.231	-0.231		1.244		-0.560	-1.695		
R011L11M	0.912	-0.354				1.550	0.549				0.189
R011L12C	0.735	0.509	0.810	-0.810		1.250		0.376	-1.272		
R021K01C	0.422	-0.891	0.186	-0.186		0.717		0.773	1.278		
R021K02C	0.807	-0.559				1.372	0.767				
R021K03M	1.004	0.081				1.707	-0.138				0.184
R021K04M	1.062	0.979				1.805	-1.767				0.391
R021K05C	0.969	0.137				1.647	-0.226				
R021K06M	1.489	0.052				2.531	-0.132				0.281
R021K07C	0.682	0.143	0.119	-0.119		1.159		-0.028	-0.332		
R021K08M	0.994	0.354				1.690	-0.598				0.197
R021K09M	1.203	-0.010				2.045	0.020				0.246
R021K10C	0.785	0.778	-0.397	0.397		1.335		-1.568	-2.076		
R021K11M	1.070	0.260				1.819	-0.473				0.240
R021K12C	0.576	-0.110	0.446	-0.084	-0.362	0.979		0.544	0.570	0.323	
R021Y01M	1.097	0.154				1.865	-0.287				0.253
R021Y02M	1.649	-0.204				2.803	0.572				0.288
R021Y03C	0.815	0.564				1.386	-0.781				
R021Y04M	1.273	0.093				2.164	-0.201				0.222
R021Y05M	1.721	0.086				2.926	-0.252				0.226
R021Y06M	1.533	0.042				2.606	-0.109				0.209
R021Y07M	0.792	-1.011				1.346	1.361				0.182
R021Y08M	1.360	-0.271				2.312	0.627				0.261
R021Y09C	0.956	-0.551	0.078	-0.078		1.625		1.022	1.791		
R021Y10C	0.749	0.574				1.273	-0.731				
R021Y11M	1.411	0.035				2.399	-0.084				0.284

Item	PIRLS					MIRT					
	a_i	b_i	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$	a_i	b_i	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$	y_i
R021Y12C	0.706	-0.001	-1.154	1.154		1.200		-1.384	0.002		
R021Y13C	0.760	0.378	0.594	-0.219	-0.375	1.292		0.279	-0.492	-1.465	
R021Y14C	0.576	0.222	-0.549	0.549		0.979		-0.755	-0.435		
R031M01M	1.451	-0.877				2.467	2.163				0.268
R031M02C	1.218	-0.957				2.071	1.982				
R031M03M	1.330	-0.004				2.261	0.009				0.228
R031M04C	0.560	0.357				0.952	-0.340				
R031M05M	1.551	0.112				2.637	-0.295				0.424
R031M06M	1.159	0.368				1.970	-0.725				0.283
R031M07M	1.619	-0.350				2.752	0.963				0.276
R031M08M	1.382	-0.498				2.349	1.170				0.265
R031M09C	0.759	-0.030	1.143	-1.143		1.290		1.514	0.077		
R031M10C	0.623	0.413				1.059	-0.437				
R031M11M	0.854	-0.692				1.452	1.005				0.262
R031M12M	1.196	0.106				2.033	-0.216				0.162
R031M13M	2.100	-0.642				3.570	2.292				0.256
R031M14M	2.283	-0.195				3.881	0.757				0.197
R031M15M	1.298	0.065				2.207	-0.143				0.218
R031M16C	1.207	0.058				2.052	-0.119				
R031M17CZ	0.612	0.003	0.071	0.228	-0.299	1.040		0.071	0.305	-0.009	
R031W01C	0.718	-0.584	0.243	-0.243		1.221		1.009	1.426		
R031W02C	0.800	0.278	-0.107	0.107		1.360		-0.524	-0.756		
R031W03M	1.347	-0.063				2.290	0.144				0.162
R031W04C	0.842	-0.687				1.431	0.983				
R031W05M	1.264	0.497				2.149	-1.068				0.257
R031W06M	0.753	-0.999				1.280	1.279				0.147
R031W07CZ	0.879	0.509	-0.079	0.169	-0.090	1.494		-0.879	-1.387	-2.282	
R031W08M	1.355	-0.093				2.304	0.214				0.307
R031W09M	0.951	0.565				1.617	-0.913				0.178
R031W10M	1.289	0.320				2.191	-0.701				0.164
R031W11C	1.467	0.551				2.494	-1.374				
R031W12M	1.514	0.706				2.574	-1.817				0.219
R031W13C	0.862	0.791				1.465	-1.159				
R041H01M	0.947	-0.667				1.610	1.074				0.582
R041H02M	1.058	-0.743				1.799	1.336				0.196
R041H03C	1.172	0.257				1.992	-0.512				
R041H04C	0.671	1.644				1.141	-1.875				
R041H05M	1.030	0.131				1.751	-0.229				0.286
R041H06C	0.700	-0.264	-0.089	0.089		1.190		0.208	0.628		
R041H07M	0.895	0.873				1.522	-1.328				0.151
R041H08C	0.691	1.492				1.175	-1.753				
R041H09M	0.649	0.272				1.103	-0.300				0.166
R041H10M	1.307	-0.179				2.222	0.398				0.263
R041H11M	1.402	-0.537				2.383	1.280				0.311
R041H12M	1.350	0.002				2.295	-0.005				0.369
R041H13C	0.541	0.753	-0.089	0.270	-0.181	0.920		-0.774	-1.219	-2.078	
R041H14C	0.990	0.307				1.683	-0.517				
R041H15C	1.250	-0.385				2.125	0.818				
R041H16C	1.027	0.577				1.746	-1.007				
R041I01C	0.793	-1.522				1.348	2.052				
R041I02M	1.034	0.623				1.758	-1.095				0.185
R041I03C	0.560	0.201	0.330	-0.330		0.952		0.123	-0.383		
R041I04C	0.685	0.470	0.301	-0.301		1.165		-0.197	-1.095		
R041I05M	1.127	0.164				1.916	-0.314				0.223
R041I06M	1.260	0.487				2.142	-1.043				0.283

Item	PIRLS					MIRT					
	a_i	b_i	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$	a_i	b_i	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$	y_i
R041I07C	0.968	0.502	-0.008	0.008		1.646		-0.839	-1.652		
R041I08M	1.594	-0.579				2.710	1.569				0.264
R041I09C	1.152	-0.035				1.958	0.069				
R041I10M	0.952	0.265				1.618	-0.429				0.244
R041I11C	0.820	0.471	0.256	-0.256		1.394		-0.300	-1.313		
R041I12M	1.061	0.155				1.804	-0.280				0.145
R041I13C	0.804	0.207				1.367	-0.283				
R041I14C	0.801	0.944				1.362	-1.285				
R041I15C	0.755	0.690				1.284	-0.886				
R041O01M	0.942	-0.993				1.601	1.590				0.254
R041O02C	0.859	-0.965				1.460	1.409				
R041O03C	1.020	0.868				1.734	-1.505				
R041O04C	0.595	1.145	-0.160	0.160		1.012		-1.320	-2.316		
R041O05C	0.630	0.228	-0.260	0.260		1.071		-0.523	-0.488		
R041O06M	1.437	-0.212				2.443	0.518				0.315
R041O07C	0.629	-0.953	0.036	-0.036		1.069		1.058	2.038		
R041O08C	0.847	-0.064				1.440	0.092				
R041O09C	1.205	-0.059				2.049	0.121				
R041O10C	0.815	0.236	-0.006	0.006		1.386		-0.335	-0.654		
R041O11M	1.365	0.096				2.321	-0.223				0.295
R041O12M	1.283	-0.341				2.181	0.744				0.169
R041O13C	0.567	0.348	-1.197	0.794	0.403	0.964		-1.489	-1.059	-1.006	
R041T01M	1.085	-0.973				1.845	1.795				0.242
R041T02C	0.690	-0.516	0.269	-0.269		1.173		0.921	1.211		
R041T03C	0.943	0.071	0.034	-0.034		1.603		-0.059	-0.228		
R041T04C	1.239	0.055				2.106	-0.116				
R041T05M	0.756	0.383				1.285	-0.492				0.244
R041T06C	1.381	-0.579				2.348	1.359				
R041T07C	0.780	0.622	-0.202	0.202		1.326		-1.093	-1.650		
R041T08C	1.187	0.058				2.018	-0.117				
R041T09M	1.560	0.500				2.652	-1.326				0.182
R041T10C	1.264	-0.212				2.149	0.456				
R041T11C	0.859	0.518	-0.426	0.319	0.106	1.460		-1.379	-1.669	-2.271	
R041T12M	0.999	-0.050				1.698	0.085				0.346
R041T13M	1.111	0.432				1.889	-0.816				0.112
R041T14C	0.533	0.131				0.906	-0.119				
R041T15M	0.897	0.594				1.525	-0.906				0.235
R041T16M	1.289	-0.046				2.191	0.101				0.286

3.4.2 Schattingsprocedure *plausible values*

Op basis van het geschatte IRT-model werden vervolgens de vaardigheden van de leerlingen geschat. Ook hierbij werden de procedures van de IEA gevolgd. Daarom werden vijf *plausible values* per leerling geschat in de plaats van één puntschatting per leerling. In deze sectie lichten we het schatten van *plausible values* toe.

De methode om *plausible values* te schatten in plaats van één puntschatting is ontwikkeld om te corrigeren voor de onbetrouwbaarheid in vaardigheidsschattingen. De echte vaardigheid van een leerling is immers niet gekend en op basis van een test kan enkel geschat worden wat de meest waarschijnlijke vaardigheid is. Deze schatting is onderworpen aan een meetfout waardoor we de vaardigheid van een individuele leerling oftewel iets te hoog of iets te laag schatten. Het is dan ook correcter om te spreken van een distributie van de meest waarschijnlijke vaardigheid van één leerling. Als we slechts het gemiddelde van deze vaardigheidsdistributie van iedere leerling gebruiken zorgt de meetfout ervoor dat door de spreiding de vaardigheid in de testgroep overschat wordt. De correlaties van deze vaardigheid met andere variabelen zullen dan weer onderschat worden. Verder zorgt de puntschatting voor een vertekende schatting van de steekproeffout (Mislevy, 1991; Wingersky, Kaplan, & Beaton, 1987). De vijf *plausible values* per leerling beschrijven echter de distributie van de meest waarschijnlijke vaardigheid per leerling. Dit is verschillend van een puntschatting die enkel het gemiddelde van deze distributie van de meest waarschijnlijke vaardigheid beschrijft. Door vijf *plausible values* te analyseren kan onderscheiden worden welk deel van de geschatte variatie te wijten is aan de meetfout. Zo kan dan ook de steekproeffout beter geschat worden.

Net zoals de IEA bij PIRLS 2016 gebruikten we conditionering bij het schatten van de *plausible values* om de schattingen betrouwbaarder te maken (Mullis & Martin, 2015). Bij conditionering wordt er bij de schatting van de *plausible values* niet enkel gebruikt gemaakt van de toetsantwoorden, maar ook van andere achtergrondinformatie van de leerlingen. Van de leerlingen die deelnamen aan PIRLS 2016 was immers ook informatie beschikbaar van een leerlingvragenlijst, ouder vragenlijst en administratieve achtergrondkenmerken. PIRLS schatte daarom eerst op basis van de aanwezige informatie 259 principale componenten die meer dan 90% van de totale variantie in de aanwezige informatie beschreef. Deze principale componenten werden vervolgens gebruikt als conditioneringsvariabelen. Ook geslacht werd opgenomen als een conditioneringsvariabele binnen Vlaanderen. Bij de schatting van de *plausible values* bij de herhalingsmeting van PIRLS in 2018 werden bij de leerlingen ook conditioneringsvariabelen opgenomen.

We merken op dat het opnemen van de oorspronkelijke conditioneringsvariabelen, die waren opgenomen in PIRLS 2016, in de herhalingsmeting van PIRLS in 2018 enkel mogelijk was bij de terugkerende leerlingen. Dit zijn de 4079 leerlingen die zowel deelnamen aan PIRLS 2016 als aan de herhalingsmeting van PIRLS in 2018. Enkel voor deze leerlingen konden dezelfde conditioneringsvariabelen gebruikt worden. In totaal waren er echter 4644 leerlingen die deelnamen aan de herhalingsmeting. Er waren dus 565 leerlingen waarvan de oorspronkelijke conditioneringsvariabelen niet bekend zijn. Daarom schatten we twee sets van *plausible values*. De eerste set van *plausible values* is geschat op basis van het IRT-model en de oorspronkelijke conditioneringsvariabelen. Deze set is beschikbaar voor de 4079 terugkerende leerlingen. De tweede set van *plausible values* is geschat op basis van het IRT-model met een andere set conditioneringsvariabelen. Deze conditioneringsvariabelen waren: geslacht, klasgemiddelde voor prestaties in begrijpend lezen, hoeveel boeken er thuis zijn, de zelfzekerheid van leerlingen tegenover lezen, de betrokkenheid van leerlingen bij lezen, de houding van leerlingen tegenover lezen, de nadruk op academisch succes volgens de leerkrachten en de jobtevredenheid van leerkrachten. Deze variabelen werden verzameld tijdens de herhalingsmeting in 2018. Deze set is beschikbaar voor alle 4644 leerlingen. Merk op dat de analyses op beide sets *plausible values* voor de 4079 terugkerende leerlingen tot dezelfde gemiddelde resultaten leiden. Beide sets zijn immers onvertekende schattingen van de onderliggende vaardigheden van leerlingen. De schattingen op basis van de tweede set waarbij de oorspronkelijke conditioneringsvariabelen niet werden opgenomen hebben enkel een iets grotere schattingsfout.

Na het schatten van de vijf *plausible values* per set moesten de *plausible values* nog op de internationale PIRLS-schaal geplaatst worden. Dit werd bereikt door middel van een lineaire transformatie. Als pv_{ij} de geschatte *plausible value* i is van een leerling j , dan geeft de volgende formule weer hoe een *plausible value* op de PIRLS-schaal, pv^*_{ij} , kan bekomen worden:

$$pv^*_{ij} = 516.163 + (pv_{ij} * 97.303)$$

Deze internationale meetschaal van PIRLS voor begrijpend lezen werd bepaald in PIRLS 2001. In de afname van dat jaar werd het gemiddelde van alle landen op 500 gezet en de standaarddeviatie op 100. Elk van de daaropvolgende afnames van PIRLS in 2006, 2011 en 2016 werden geplaatst op dezelfde meetschaal.

3.4.3 Resultaten

Om te onderzoeken of de analyses in deze sectie tot equivalente vaardigheidsschattingen leiden als de IEA hebben we onze analyses ook uitgevoerd op de PIRLS 2016 dataset. Met onze *plausible values* komen we op een Vlaams gemiddelde van 523.40 met een standaarddeviatie van 60.73 (ongewogen over de vijf *plausible values*). Dit is gelijk aan het Vlaams gemiddelde van 523.68 met een standaarddeviatie van 60.62 (ongewogen over de vijf *plausible values*) zoals geschat door de IEA, gegeven een zekere schattingsfout. Ook de *scatterplots* en correlaties met onze *plausible values* van PIRLS 2016 tonen dezelfde resultaten als de oorspronkelijke *plausible values* van de IEA.

Voor de vaardigheidsschatting van de herhalingsmeting van PIRLS in 2018 werd zowel een set *plausible values* geschat voor de 4079 terugkerende leerlingen als een set *plausible values* voor alle 4644 leerlingen. Voor de 4079 terugkerende leerlingen komen we op een Vlaams gemiddelde van 585.64 met een standaarddeviatie van 52.35 (ongewogen over de vijf *plausible values*). Voor alle 4644 leerlingen komen we op een Vlaams gemiddelde van 585.18 met een standaarddeviatie van 53.10 (ongewogen over de vijf *plausible values*).

3.5 Meetschaal Peilingsonderzoek

3.5.1 Parameters van het Peilingsonderzoek

In deze sectie beschrijven we hoe de vaardigheidsschattingen van de leerlingen die deelnamen aan de herhalingsmeting van PIRLS in 2018 op de schaal van het Peilingsonderzoek werd geplaatst. Dit is de meetschaal waarop het Peilingsonderzoek lezen in 2002, 2007, 2013 en het 2018 gesitueerd zijn. In Appendix 2 wordt de syntax weergegeven van de analyses in de gebruikte statistische software, MIRT 1.30 (Chalmers, 2012) in R 3.6.0 (R Core Team, 2018).

Het schatten van de vaardigheidsscores voor de herhalingsmeting startte met het schatten van een IRT-model op basis van de testgegevens van de testgroep. Doorgaans worden in zo'n IRT-model de itemparameters vrij geschat, terwijl het gemiddelde en de standaarddeviatie van de testgroep respectievelijk op 0 en 1 vastgezet worden. Omdat we echter op basis van de meetschaal van het Peilingsonderzoek werkten werden de itemparameters vastgezet volgens de parameterwaarden van het Peilingsonderzoek. Het gemiddelde en de standaarddeviatie van de testgroep werden dus vrij geschat op basis van de itemparameters en de testgegevens. De waarden van de itemparameters van de items die afkomstig zijn van PIRLS 2016 werden wel vrij geschat.

Merk op dat de itemparameters van de meetschaal van het Peilingsonderzoek niet direct toepasbaar waren binnen onze statistische software. Zo gebruikt het peilingsonderzoek een OPLM-model, een variatie van het 2-parameter model, dewelke een andere parameterisatie heeft dan ons 2-parameter model. Zo wordt binnen het Peilingsonderzoek het 2-parameter model op de volgende wijze beschrijven:

$$P(x_i = 1 | \theta_j, a_i, b_i) = \frac{1}{1 + \exp(a_i * (\theta_j - b_i))}$$

Het belangrijkste verschil tussen het OPLM-model zoals het gedefinieerd wordt in het Peilingsonderzoek en het 2-parameter model zoals het gedefinieerd wordt in deze studie is hoe de interceptparameter geschat wordt. Zo wordt er binnen het Peilingsonderzoek gewerkt met “ $-a_i * (\theta_j - b_i)$ ” in de plaats van “ $-a_i * \theta_j + a_i$ ” waardoor de waarde van de interceptparameter afhankelijk wordt van de waarde van discriminatieparameter tijdens de schattingsprocedure. Bijgevolg moeten deze parameter van het Peilingsonderzoek eerst getransformeerd worden voor onze parameterisatie. Om het intercept a_i voor onze parameterisatie te bekomen moeten we “ $a_i * (-a_i)$ ” berekenen volgens de Peilingsonderzoek parameterisatie. Terzijde merken we op dat binnen de literatuur over het OPLM-model eerder gesproken wordt over “discriminatie-index” in de plaats van “discriminatieparameter” (Verhelst & Glas, 1995, p. 217). Deze parameter heeft echter dezelfde waarde in beide modellen, ondanks het verschil in naam.

Tabel 24 toont de waarden van de itemparameters zoals ze beschreven zijn in het Peilingsonderzoek. Vervolgens worden de waarden getoond van de itemparameters na de transformatie voor het gebruik in MIRT. Dit zijn de waardes die vastgezet zijn in het IRT-model dat diende om de vaardigheidsscores op de schaal van het Peilingsonderzoek te schatten.

Tabel 24: Itemparameters zoals gerapporteerd door het Peilingsonderzoek en getransformeerd voor MIRT

Item	Peilingsonderzoek	MIRT	a_i
	b_i	b_i	
Le_01_DSP_03	0.171	-0.684	4
Le_01_SP_02	0.227	-0.908	4
Le_01_SP_05	0.018	-0.090	5
Le_01_SP_06	0.350	-1.050	3
Le_01_SP_07	-0.252	0.756	3
Le_02_DTA_01	-1.111	2.222	2
Le_02_DTA_02	0.178	-0.356	2
Le_02_DTA_03	0.389	-1.167	3
Le_02_DTA_04	-0.046	0.184	4
Le_05_DMS_01	0.254	-1.016	4
Le_05_DMS_02	0.000	0.000	4
Le_05_DMS_03	-0.124	0.496	4
Le_05_DMS_04	-0.077	0.231	3
Le_05_DMS_06	-0.003	0.012	4
Le_05_DMS_07	-0.354	1.416	4
Le_06_DVJ_01	-1.253	2.506	2
Le_06_DVJ_02	-0.580	1.740	3
Le_06_DVJ_03	-0.193	0.386	2
Le_06_DVJ_04	-0.394	1.182	3
Le_06_DVJ_05	-0.111	0.333	3
Le_06_DVJ_06	-0.592	1.776	3
Le_07_DPP_01	-0.064	0.192	3
Le_07_DPP_02	0.339	-1.356	4
Le_07_DPP_03	-0.125	0.500	4
Le01_VL_01	-0.693	1.386	2
Le01_VL_03	0.073	-0.292	4
Le01_VL_04	0.105	-0.420	4
Le01_VL_07	-0.429	1.287	3
Le01_VL_08	-0.036	0.144	4
Le04_GI_02	0.114	-0.342	3
Le04_GI_03	-0.195	0.195	1
Le04_GI_04	0.332	-0.996	3
Le04_GI_05	0.431	-0.862	2
Le06_PZ_01	0.024	-0.096	4
Le06_PZ_03	0.075	-0.375	5
Le06_PZ_06	0.354	-1.062	3
Le06_PZ_07	0.155	-0.620	4
Le06_PZ_09	0.644	-2.576	4
Le07_KM_01	0.422	-0.844	2
Le07_KM_02	0.072	-0.216	3
Le07_KM_03	0.093	-0.186	2
Le07_KM_05	0.115	-0.460	4
Le07_KM_06	0.234	-0.702	3
Le_02_DSC_01	-0.420	1.260	3
Le_02_DSC_02	-0.564	1.692	3
Le_02_DSC_03	-1.468	2.936	2
Le_02_DSC_04	0.121	-0.363	3

3.5.2 Schattingsprocedure WLE-scores

Op basis van het geschatte IRT-model werden vervolgens de vaardigheden van de leerlingen geschat. Ook hierbij werden de procedures van het Peilingsonderzoek gevolgd. Daarom werd *Warm's weighted likelihood estimation* (Warm, 1989) voor het schatten van vaardigheidsscores gebruikt. Hier wordt vaak naar verwezen als WLE-scores. Een WLE-score is een puntschatting van de vaardigheid van een leerling. In deze sectie lichten we het schatten van WLE-scores toe.

Warm's weighted likelihood estimation is een procedure om de onderliggende vaardigheid te schatten en is gebaseerd op de *maximum likelihood estimation* (Warm, 1989). Omdat de echte vaardigheid van een leerling niet gekend is maakt *maximum likelihood estimation* gebruik van een iteratief proces om de meest waarschijnlijke vaardigheid van een leerling te schatten, gegeven het IRT-model. Dit proces verloopt door voor elk patroon van juiste en foute items een waarschijnlijkheidsfunctie op te stellen. Deze functie is daarbij zelf een som van de verschillende waarschijnlijkheidsfuncties voor het al dan niet juist beantwoorden van een item. Wanneer deze waarschijnlijkheidsfuncties berekend zijn kan de meest waarschijnlijke onderliggende vaardigheid van een leerling geschat worden. *Maximum likelihood* als scoringsmechanisme heeft als voordeel dat het volgens een goed gekende statistische theorie verloopt. Deze methode levert dan ook scores op die asymptotisch onvertekend zijn, gegeven dat de toets gericht is op het vaardigheidsniveau van de leerling. Echter, wanneer leerlingen een vaardigheid hebben die lager of hoger is dan het vaardigheidsniveau waarop een toets gericht is, dan worden vaardigheden asymptotisch respectievelijk onderschat en overschat. Het gevolg is dat bij leerlingen die bijna alles juist of bijna alles fout hebben er vaak extreme waarden bekomen worden. Een aansluitend nadeel van deze methode is ook dat de schatting van vaardigheidsscores voor leerlingen die alles juist of alles fout hebben op een toets onmogelijk is. Tevens wordt in deze methode de variantie in de vaardigheden van leerlingen overschat naargelang hoe onbetrouwbaar de items de onderliggende vaardigheid meten.

Warm's weighted likelihood estimation werkt op eenzelfde wijze als *maximum likelihood estimation*, maar gebruikt een correctiefactor in het toekennen van vaardigheidsscores aan leerlingen. Immers, vaardigheidsschattingen met *maximum likelihood* zijn meer vertekend naargelang de vaardigheid van een leerling afwijkt van het vaardigheidsniveau waar een toets zich op richt. Aangezien toetsen doorgaans minder informatie bieden op de extremen van de onderliggende vaardigheid, worden lage vaardigheden en hoge vaardigheden respectievelijk onderschat en overschat. Hierdoor is de variatie in *maximum likelihood* vaardigheidsscores ook een overschatting van de werkelijke variatie. Om deze vertekening naargelang het gebrek van informatie te reduceren wordt in *weighted likelihood estimation* een correctiefactor gebruikt bij het toekennen van de vaardigheidsscore. Deze correctiefactor is daarbij groter naargelang de informatie voor een bepaalde vaardigheid kleiner is en omgekeerd. Op deze manier wordt de vertekening in *maximum likelihood* vaardigheidsschattingen tenietgedaan, wat tevens het belangrijkste voordeel is van deze methode. Een bijkomend voordeel is dat de geschatte variantie van de vaardigheid binnen de steekproef met *weighted likelihood estimation* minder vertekend is.

We merken op dat na het schatten van de WLE-scores er geen transformaties meer nodig zijn. De geschatte WLE-scores op basis van het IRT-model bevinden zich meteen op de meetschaal van het Peilingsonderzoek.

3.5.3 Resultaten

Om te onderzoeken of de analyses in deze sectie tot equivalente vaardigheidsschattingen leiden als het Peilingsonderzoek hebben we onze analyses ook uitgevoerd op de Peilingsonderzoek Nederlands begrijpend lezen in 2018. Met onze WLE-scores (op basis van software MIRT) komen we op een Vlaams gemiddelde van 0.425 met een standaarddeviatie van 0.346. Dit is gelijk aan het Vlaams gemiddelde van 0.425 met een standaarddeviatie van 0.346 zoals geschat door het Peilingsonderzoek. De geschatte WLE-scores van beide studies zijn dus gelijk.

Voor de vaardigheidsschatting in ons onderzoek werden zowel WLE-scores geschat voor de herhalingsmeting van PIRLS in 2018 als voor PIRLS 2016. Voor PIRLS 2016 vinden we een Vlaams gemiddelde van 0.114 met een standaarddeviatie van 0.235. Voor de herhalingsmeting van PIRLS in 2018 vinden we een Vlaams gemiddelde van 0.388 met een standaarddeviatie van 0.228.

3.6 Factorscores attitudes tegenover lezen op PIRLS-meetschaal

3.6.1 Parameters van PIRLS 2016

Voor het schatten van de factorscores voor verschillende attitudes maakte PIRLS 2016 ook gebruik van IRT. Deze concepten werden immers gemeten met verschillende items, waarbij de antwoorden van leerlingen op elk van deze items het onderliggende concept als oorzaak hebben. Een leerling kan dus een hoge of lage waarde hebben voor het onderliggende psychosociale concept, die zijn of haar antwoorden op de items bepaalt. Aangezien in de herhalingsmeting van PIRLS in 2018 vijf van deze concepten opnieuw bevestigd werden, moesten deze ook op dezelfde meetschaal geplaatst worden als in PIRLS 2016. Het gaat om twee concepten die in de leerkrachtvragenlijst van 2016 en van 2018 werden bevestigd en drie concepten uit de leerlingvragenlijst van 2016 en van 2018. Deze attitudes zijn

- “School Emphasis on Academic Success – Teachers’ Reports (ATBGEAS)”,
- “Teacher Job Satisfaction Scale (ATBGTJS)”,
- “Students Confident in Reading Scale (ASBGSCR)”,
- “Students Engaged in Reading Lessons Scale (ASBGERL)”,
- “Students Like Reading Scale (ASBGLR)”.

In deze sectie beschrijven we hoe de niveauschattingen van de leerlingen die deelnamen aan de herhalingsmeting van PIRLS in 2018 op de PIRLS-schaal werden geplaatst. In Appendix 3 wordt de syntax weergegeven van de analyses in de gebruikte statistische software, MIRT 1.30 (Chalmers, 2012) in R 3.6.0 (R Core Team, 2018).

Het schatten van de factorscore voor de herhalingsmeting startte met het schatten van een IRT-model op basis van de antwoorden op de vragenlijsten van de testgroep. De itemparameters werden vastgezet volgens de parameterwaarden van de PIRLS-schaal. Het gemiddelde en de standaarddeviatie van de testgroep werden dus vrij geschat op basis van de itemparameters en de testgegevens.

Merk op dat de itemparameters van de PIRLS-schaal niet direct toepasbaar waren binnen onze statistische software. PIRLS gebruikt *partial credit* modellen voor de vijf psychosociale concepten zoals beschreven in onze sectie over IRT. Dit is een vorm van het *generalized partial credit* model, maar de parameterisatie is verschillend. Binnen het PIRLS-onderzoek wordt het PCM-model voor de concepten op de volgende wijze beschreven:

$$P(x_i = v | \theta_j, b_i, d_{i,1}, \dots, d_{i,m_i-1}, k_{i,0}, \dots, k_{i,m_i-1}) = \frac{\exp(\sum_{k=0}^v (\theta_j - (b_i + d_{i,k})))}{\sum_{g=0}^{m_i-1} \exp(\sum_{k=0}^g (\theta_j - (b_i + d_{i,k})))}$$

Er zijn meerdere verschillen tussen het *partial credit* model zoals het gedefinieerd wordt in PIRLS en het *generalized partial credit* model zoals het gedefinieerd wordt in deze studie. Zo kent het *partial credit* model geen discriminatieparameter, de discriminatiewaarde wordt dus verondersteld gelijk te zijn voor de verschillende items van één concept. Verder wordt er binnen PIRLS gewerkt met de natuurlijke exponent van een som van deelfuncties,

“ $\sum_{k=0}^g (\theta_j - (b_i + d_{i,k}))$,” terwijl er in onze studie wordt gewerkt met de natuurlijk exponent van een deelfunctie, “ $\exp(a_i k_{i,k} (a_i \theta_j + d_{i,k}))$ ”. Bijgevolg moeten de parameters van de PIRLS-schaal voor het PCM-model eerst getransformeerd worden voor onze parameterisatie van het *generalized partial credit* model. De locatieparameter $d_{0,k}$ heeft de waarde 0 over beide parameterisaties. Om de locatieparameter $d_{i,1}$ voor onze parameterisatie te bekomen moeten we “ $-(b_i + d_{i,1})$ ” berekenen volgens de PIRLS parameterisatie. Om de locatieparameter $d_{i,2}$ voor onze parameterisatie te bekomen moeten we “ $-(b_i + d_{i,1}) - (b_i + d_{i,2})$ ” berekenen volgens de PIRLS parameterisatie. Om de locatieparameter $d_{i,3}$ voor onze parameterisatie te bekomen moeten we “ $-(b_i + d_{i,1}) - (b_i + d_{i,2}) - (b_i + d_{i,3})$ ” berekenen volgens de PIRLS parameterisatie. We merken op dat in PIRLS 2016 de items in de databank *reverse coded* (1, 2 en 3 worden respectievelijk 3, 2 en 1) waren voor het toepassen de IRT-modellen. Ook wij werkten daarom met items die *reverse coded* waren.

Tabel 25 toont de waarden van de itemparameters zoals ze beschreven zijn in het internationale PIRLS-onderzoek. Vervolgens worden de waarden getoond van de itemparameters na de transformatie voor het gebruik in MIRT. Dit zijn de waarden die vastgezet zijn in het IRT-model dat diende om de factorscores op de schaal van PIRLS te schatten.

Tabel 25: Itemparameters zoals gerapporteerd door het PIRLS-onderzoek en getransformeerd voor MIRT.

Item	PIRLS				MIRT		
	b_i	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$	$d_{i,1}$	$d_{i,2}$	$d_{i,3}$
School Emphasis on Academic Success – Teachers’ Reports (ATBGEAS)							
ATBG07A	-1.326	-2.862	-0.086	2.948	4.188	5.599	3.977
ATBG07B	-0.933	-3.259	0.009	3.250	4.193	5.117	2.800
ATBG07C	-0.905	-3.150	-0.016	3.167	4.055	4.977	2.716
ATBG07D	-0.957	-3.072	-0.057	3.129	4.029	5.044	2.872
ATBG07E	-0.123	-1.653	-0.249	1.902	1.776	2.149	0.370
ATBG07F	0.992	-2.172	0.070	2.102	1.180	0.117	-2.977
ATBG07G	1.259	-2.514	0.144	2.370	1.256	-0.147	-3.776
ATBG07H	0.108	-2.384	-0.243	2.627	2.276	2.412	-0.323
ATBG07I	1.127	-2.705	0.135	2.570	1.578	0.315	-3.382
ATBG07J	0.205	-2.872	-0.048	2.920	2.668	2.510	-0.614
ATBG07K	0.689	-3.394	0.152	3.242	2.705	1.864	-2.067
ATBG07L	-0.135	-2.388	-0.404	2.792	2.523	3.062	0.405
Teacher Job Satisfaction Scale (ATBGTJS)							
ATBG10A	0.375	-2.264	2.264		1.889	-0.751	
ATBG10B	-0.790	-2.263	2.263		3.053	1.580	
ATBG10C	0.063	-2.280	2.280		2.217	-0.126	
ATBG10D	0.461	-2.325	2.325		1.864	-0.921	
ATBG10E	-0.109	-1.957	1.957		2.066	0.218	
Students Confident in Reading Scale (ASBGSCR)							
ASBR07A	-0.523	-0.324	-0.684	1.008	0.847	2.054	1.569
ASBR07B	-0.601	-0.480	-0.427	0.907	1.081	2.110	1.804
ASBR07C	1.014	-0.938	0.509	0.429	-0.076	-1.599	-3.043
ASBR07D	0.258	-0.389	0.228	0.161	0.131	-0.355	-0.774
ASBR07E	-0.007	-0.117	0.220	-0.103	0.124	-0.089	0.021
ASBR07F	-0.141	-0.060	0.237	-0.177	0.200	0.104	0.423
Students Engaged in Reading Lessons Scale (ASBGERL)							
ASBR01A	0.219	-0.592	-0.744	1.336	0.373	0.899	-0.656
ASBR01B	0.255	-0.683	-0.468	1.150	0.428	0.641	-0.764
ASBR01C	0.064	-0.367	-0.619	0.986	0.302	0.857	-0.193
ASBR01D	-0.130	-0.585	-0.544	1.129	0.715	1.389	0.390
ASBR01E	-0.022	-0.561	-0.560	1.121	0.582	1.164	0.065
ASBR01F	0.358	-0.556	-0.435	0.991	0.198	0.275	-1.073
ASBR01G	0.140	-0.563	-0.425	0.987	0.423	0.708	-0.419
ASBR01H	-0.531	-0.088	-0.476	0.564	0.619	1.625	1.592
ASBR01I	-0.353	-0.162	-0.432	0.594	0.515	1.300	1.059
Students Like Reading Scale (ASBGSLR)							
ASBR06A	0.344	-0.470	-0.358	0.828	0.126	0.139	-1.033
ASBR06B	-0.118	-0.206	-0.369	0.574	0.324	0.811	0.354
ASBR06C	0.071	-0.253	0.139	0.114	0.182	-0.029	-0.214
ASBR06D	0.314	-0.560	-0.118	0.678	0.246	0.050	-0.942
ASBR06E	-0.178	-0.193	-0.352	0.545	0.372	0.902	0.535
ASBR06F	-0.575	-0.509	-0.239	0.747	1.084	1.898	1.726
ASBR06G	-0.115	-0.400	-0.243	0.643	0.515	0.873	0.345
ASBR06H	-0.457	0.030	-0.284	0.254	0.427	1.169	1.372
ASBR05A	0.407	-0.225	-0.470	0.696	-0.182	-0.119	-1.222
ASBR05B	0.307	-0.625	-0.200	0.825	0.318	0.210	-0.922

3.6.2 Schattingsprocedure WLE-scores

Op basis van het geschatte IRT-model werden vervolgens de factorscores van de leerlingen geschat. Ook hierbij werden de procedures van de IEA gevolgd. Daarom werd Warm's *weighted likelihood estimation* (Warm, 1989) voor het schatten van factorscores gebruikt. Hier wordt vaak naar verwezen als WLE-scores. Het schatten van deze WLE-scores werd reeds toegelicht bij de sectie over het schatten van WLE-scores bij het peilingsonderzoek. Na het schatten van de WLE-scores moesten deze nog op de internationale PIRLS-schaal geplaatst worden. Deze internationale meetschaal van PIRLS voor de verschillende attitudes werd bepaald in PIRLS 2016. Het gemiddelde van alle landen werd op 10 gezet en de standaarddeviatie op 2. Dit werd bereikt door middel van een lineaire transformatie. Als WLE_{eij} de geschatte WLE-score i is van een leerling j , dan geeft de volgende formule weer hoe een WLE-score op de PIRLS-schaal, WLE_{ij}^* , kan bekomen worden voor "School Emphasis on Academic Success – Teachers' Reports (ATBGEAS)":

$$WLE_{ij}^* = 9.085861 + (WLE_{eij} * 1.287929)$$

Voor "Teacher Job Satisfaction Scale (ATBGTJS)" werd volgende formule gebruikt:

$$WLE_{ij}^* = 8.192812 + (WLE_{eij} * 0.804273)$$

Voor "Students Confident in Reading Scale (ASBGSCR)" werd volgende formule gebruikt:

$$WLE_{ij}^* = 8.137507 + (WLE_{eij} * 1.753646)$$

Voor "Students Engaged in Reading Lessons Scale (ASBGERL)" werd volgende formule gebruikt:

$$WLE_{ij}^* = 7.347685 + (WLE_{eij} * 1.442440)$$

Voor "Students Like Reading Scale (ASBGLR)" werd volgende formule gebruikt:

$$WLE_{ij}^* = 8.281596 + (WLE_{eij} * 1.704604)$$

3.7 Is er een plafondeffect bij de herhalingsmeting in 2018?

Een bezorgdheid bij de afname van de herhalingsmeting van PIRLS in 2018 was dat de toetsboekjes een 'plafondeffect' zouden hebben. Een plafondeffect wil zeggen dat de toetsboekjes onvoldoende moeilijke toetsvragen hebben die toelaten om de vaardigheidsniveaus van sterkere leerlingen van elkaar te onderscheiden. Bijvoorbeeld, bij een sterk plafondeffect in een toets is het niet mogelijk om 'sterke' leerlingen van 'zeer sterke leerlingen' te onderscheiden. Op de toets met een sterk plafondeffect behalen beide groepen van leerlingen namelijk een maximumscore. Ze bevinden zich tegen 'de grens' van wat een toets kan meten.

De bezorgdheid over een mogelijk plafondeffect bij de herhalingsmeting bestond omdat de toetsboekjes van PIRLS 2016 gericht waren op leerlingen in het vierde leerjaar. De herhalingsmeting was echter een toetsafname bij leerlingen in het zesde leerjaar. Het leek dus behoorlijk plausibel dat we de vaardigheidsniveaus van de sterkere leerlingen niet van elkaar zouden kunnen onderscheiden, en dat veel leerlingen een maximumscore zouden behalen. De redelijk moeilijke tekst 'Chinese horoscoop' (zie [Hoofdstuk 1](#)) werd toegevoegd aan sommige toetsboekjes in de hoop een plafondeffect te vermijden.

Schattingen van parameters kunnen te groot zijn wanneer het plafondeffect te groot is, zelfs al is men niet specifiek geïnteresseerd in de sterkere leerlingen. Literatuur over 'plafondeffecten' bevindt zich voornamelijk binnen de klassieke testtheorie, waarin een somscore gebruikt wordt als vaardigheidsschatting van een individuele leerling. De herhalingsmeting van PIRLS in 2018 maakt echter gebruik van vijf plausible values per leerling die geschat zijn op basis van een IRT-model. Een belangrijk concept binnen IRT is meetfout, deze geeft weer hoeveel 'ruis' er zit op het meten van de vaardigheid. Deze meetfout is verschillend naargelang het vaardigheidsniveau van de leerlingen en welke toetsvragen er onderdeel zijn van een toets. Doorgaans kan je zeggen dat de meetfout problematisch wordt, wanneer ze groter wordt dan de onderliggende spreiding in de vaardigheid. In wat volgt gaan we na hoeveel leerlingen de maximum somscore bereiken en hoeveel toetsafnames van leerlingen een te grote meetfout hebben.

In [Tabel 26](#) wordt weergegeven hoeveel leerlingen hun toetsboekje voor de herhalingsmeting in 2018 volledig correct hebben beantwoord. We vinden dat in de totale steekproef slechts 0.84% van de leerlingen hun toetsboekje volledige correct hebben. In [Tabel 27](#) wordt weergegeven hoeveel leerlingen een geschatte meetfout hebben die de onderliggende spreiding overstijgt. We vinden dat in de totale steekproef slechts 2.05% van de leerlingen een te grote meetfout hebben.

Ter referentie worden in **Tabel 28** de gemiddelde score en spreiding in vaardigheid per toetsboekje van de herhalingsmeting in 2018 beschreven. **Figuur 5** tot en met **Figuur 14** tonen de histogrammen van de somscores van ieder toetsboekje. Hieruit leiden we af dat er meer gemakkelijke items zijn dan moeilijke items in de verschillende toetsboekjes. **Figuur 15** tot en met **Figuur 34** tonen voor elk van de 10 toetsboekje een paar figuren. Elk paar bestaat uit een histogram van de geschatte scores op de PIRLS-meetschaal en de bijhorende meetfout van deze geschatte scores. De meetfout hier is op basis van een IRT-model zonder conditioneringsvariabelen. Er wordt ook steeds een lijn weergegeven die aanduidt wanneer de meetfout groter wordt dan de onderliggende spreiding. Ook hier zien we dat de meeste leerlingen hun vaardigheid voldoende betrouwbaar gemeten wordt. In het algemeen stijgt de meetfout naargelang de leerlingen een hogere vaardigheid hebben. **Figuur 35** vergelijkt hoe de meetfout over de 10 toetsboekjes verschilt. In het algemeen zien we hier geen opvallende verschillen.

Tabel 26: Aantal en percentage leerlingen per toetsboekje die volledig correcte toets hadden

Toetsboekje	<i>n</i> volledig correct	<i>n</i> totaal leerlingen	% volledig correct
1	1	490	0.20%
2	9	468	1.92%
3	5	457	1.09%
4	0	452	0.00%
5	1	459	0.22%
6	3	447	0.67%
7	7	464	1.51%
8	0	451	0.00%
9	5	485	1.03%
10	8	471	1.70%
Totaal	39	4644	0.84%

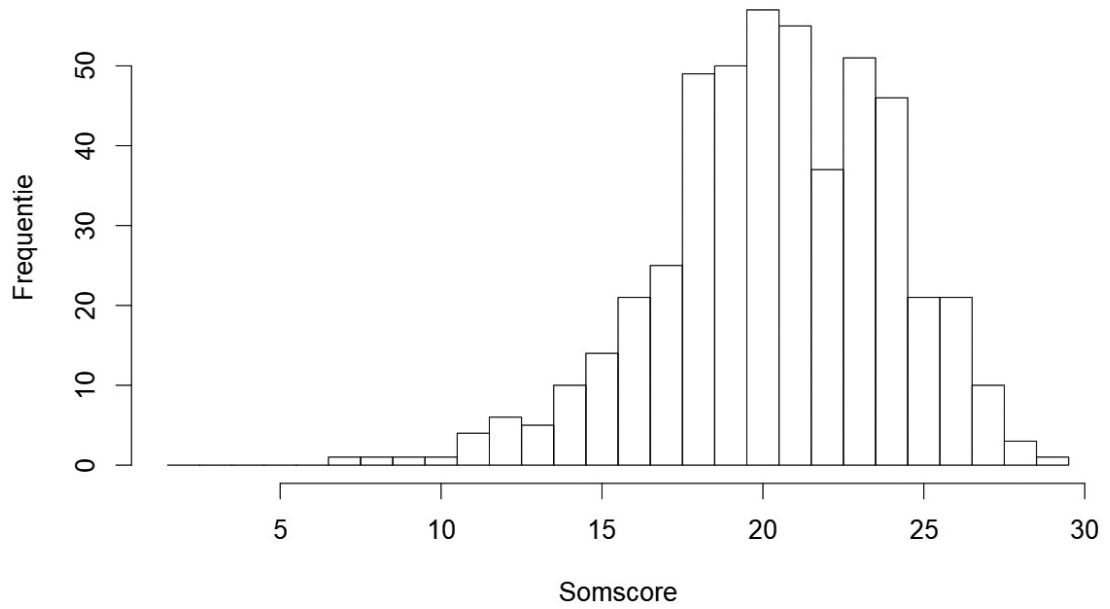
Tabel 27: Aantal en percentage leerlingen per toetsboekje waarbij de meetfout de SD in echte vaardigheid overschrijdt

Toetsboekje	<i>n</i> meetfout > SD vaardigheid	<i>n</i> totaal leerlingen	% meetfout > SD vaardigheid
1	3	490	0.61%
2	14	468	2.99%
3	12	457	2.63%
4	1	452	0.22%
5	2	459	0.44%
6	11	447	2.46%
7	13	464	2.80%
8	3	451	0.67%
9	18	485	3.71%
10	18	471	3.82%
Totaal	95	4644	2.05%

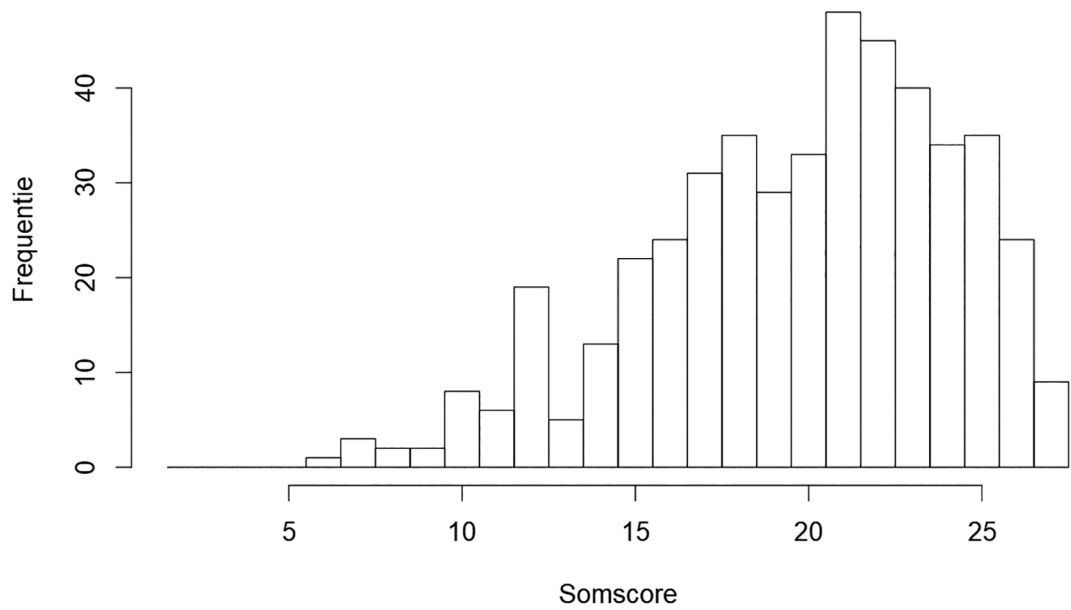
Tabel 28: Gemiddelde en SD in geschatte vaardigheid

Toetsboekje	Gem vaardigheid	SD vaardigheid
1	605.97	42.72
2	588.35	53.18
3	567.51	58.66
4	555.74	47.92
5	580.31	51.48
6	580.05	51.42
7	582.54	52.86
8	575.89	49.15
9	605.39	51.77
10	606.09	45.72
Totaal	585.18	53.10

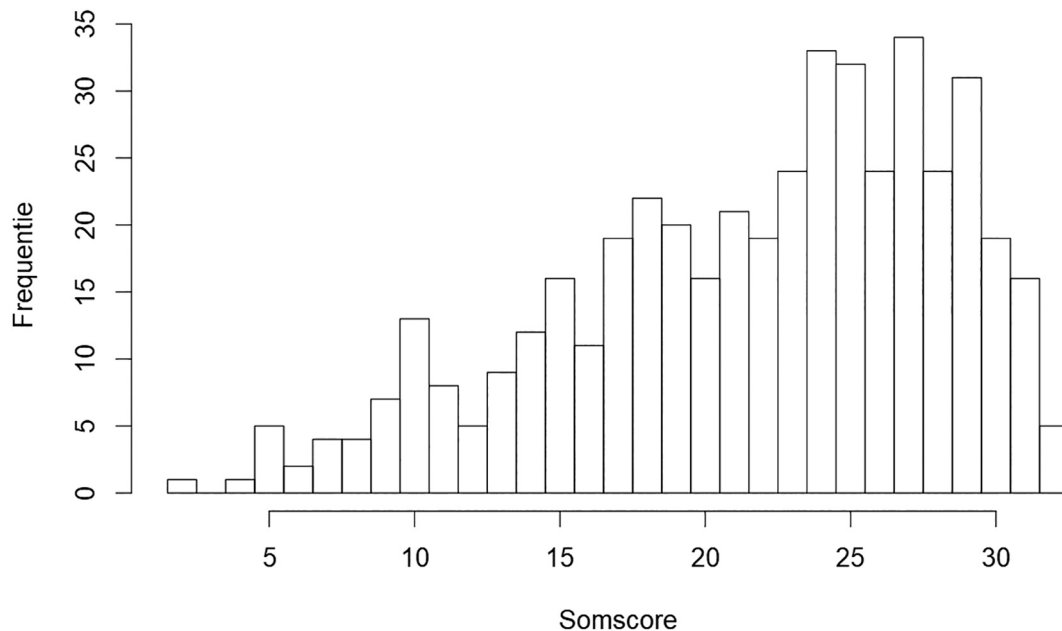
Figuur 5: Somscore toetsboekje 1



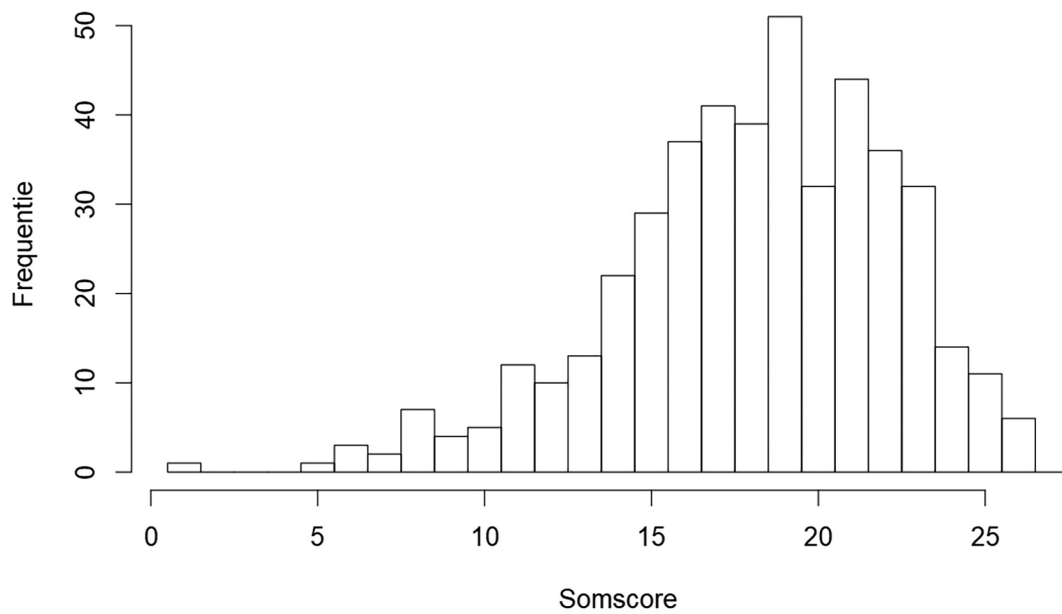
Figuur 6: Somscore toetsboekje 2



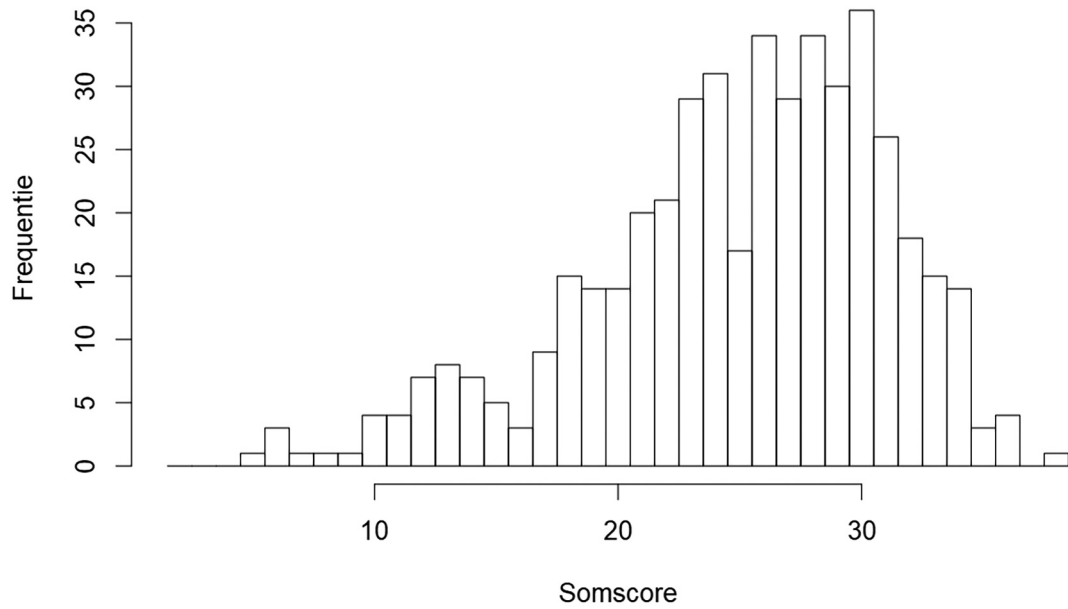
Figuur 7: Somscore toetsboekje 3



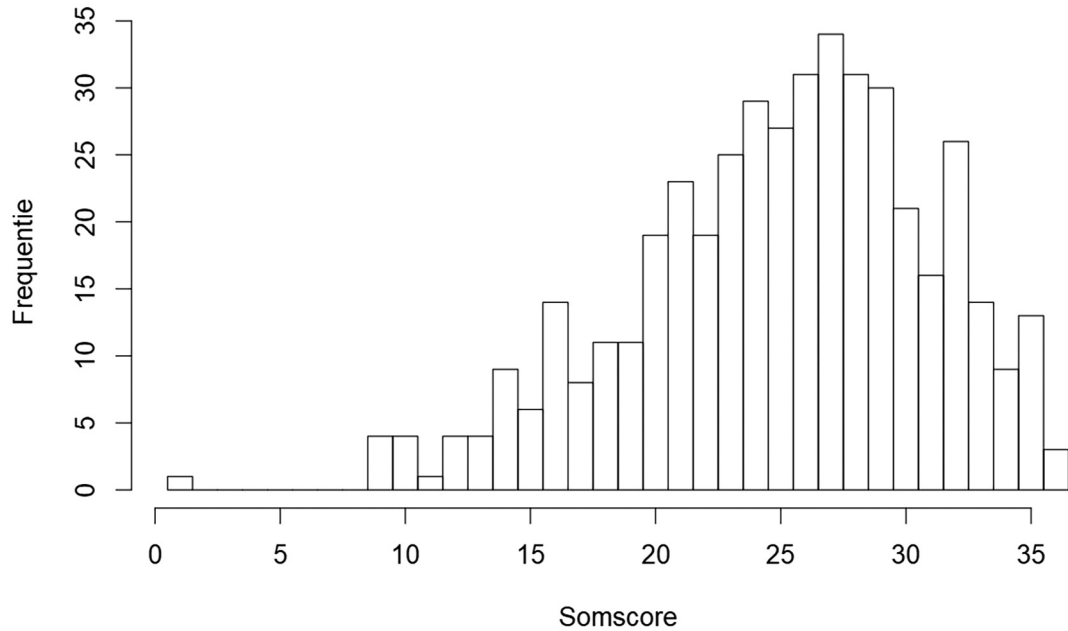
Figuur 8: Somscore toetsboekje 4



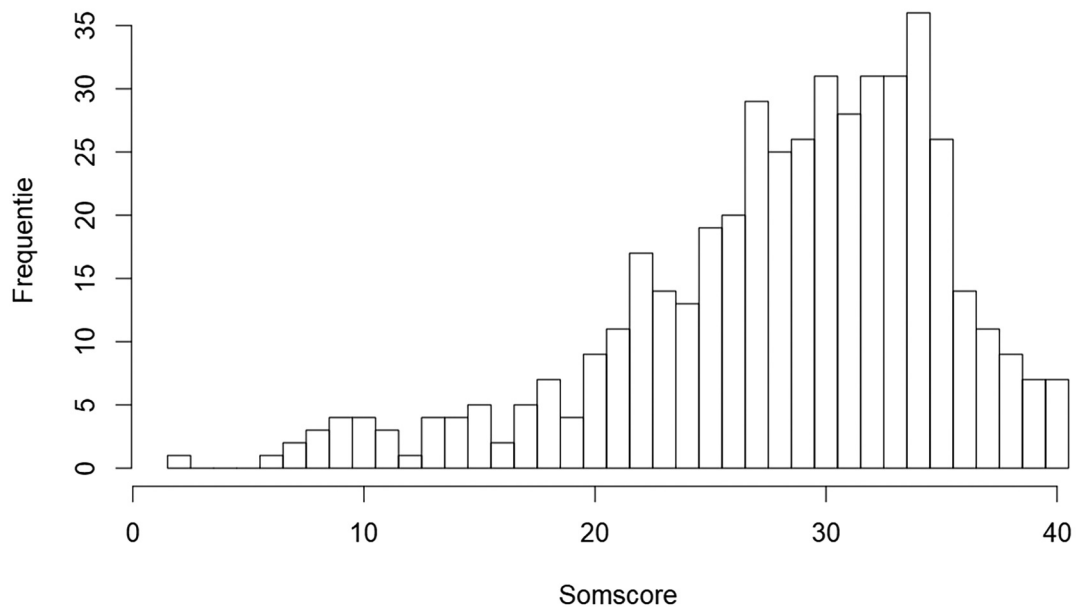
Figuur 9: Somscore toetsboekje 5



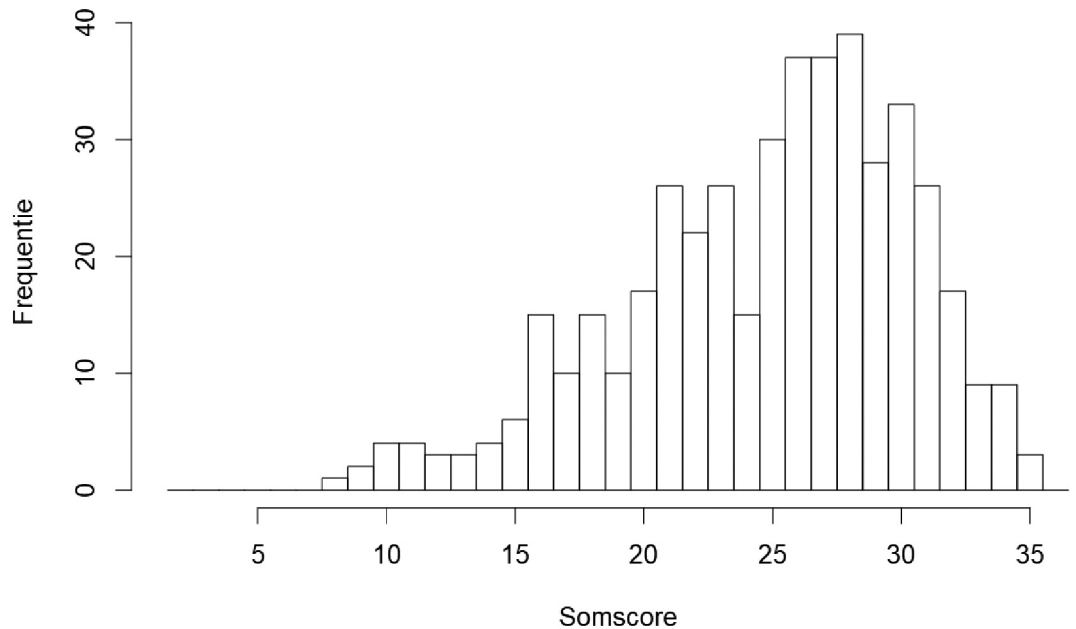
Figuur 10: Somscore toetsboekje 6



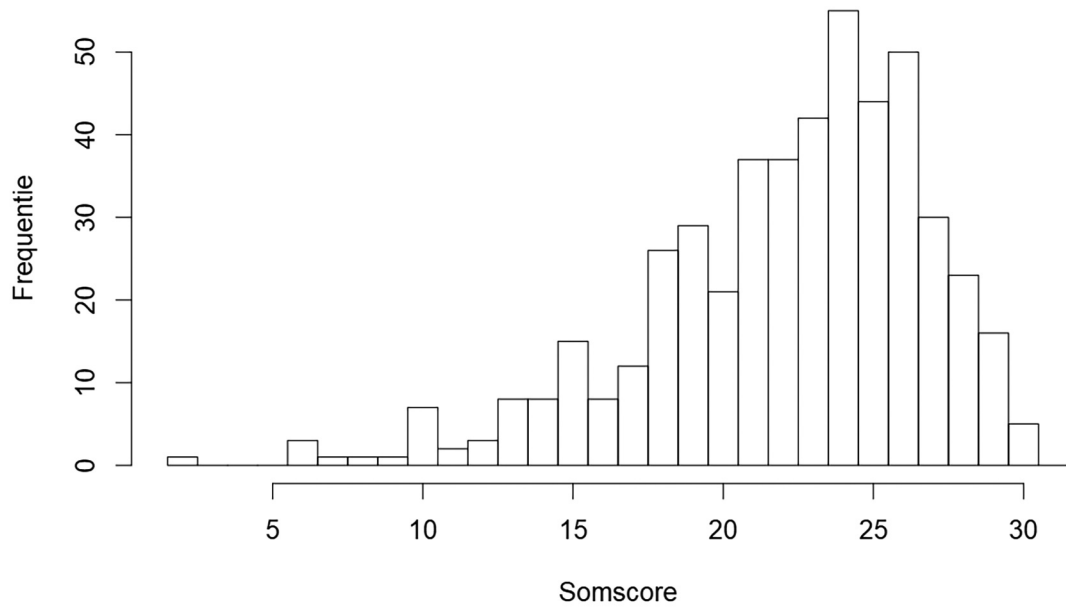
Figuur 11: Somscore toetsboekje 7



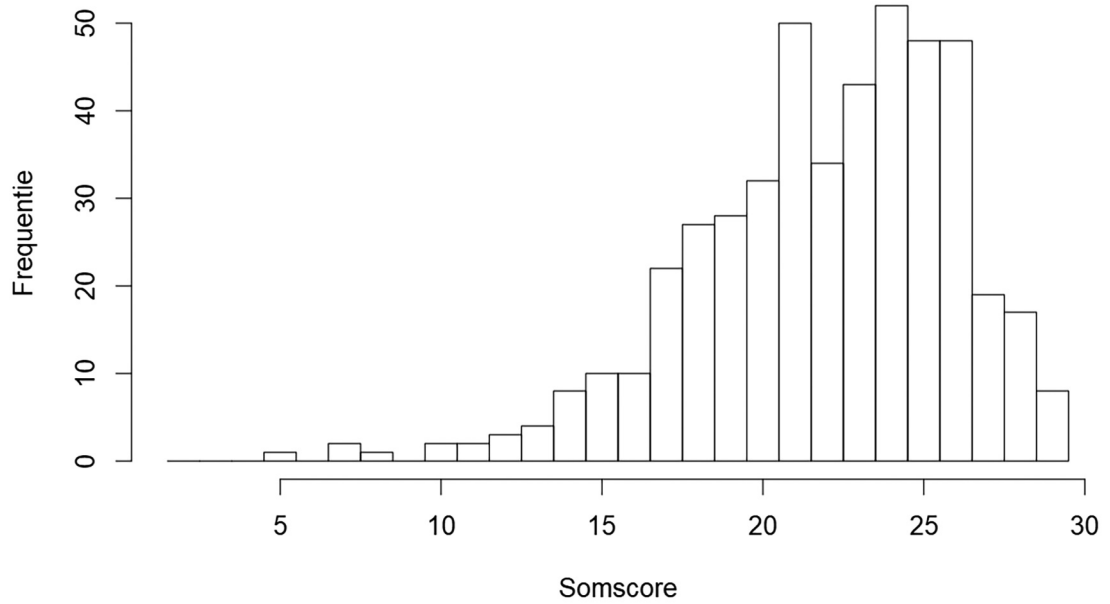
Figuur 12: Somscore toetsboekje 8



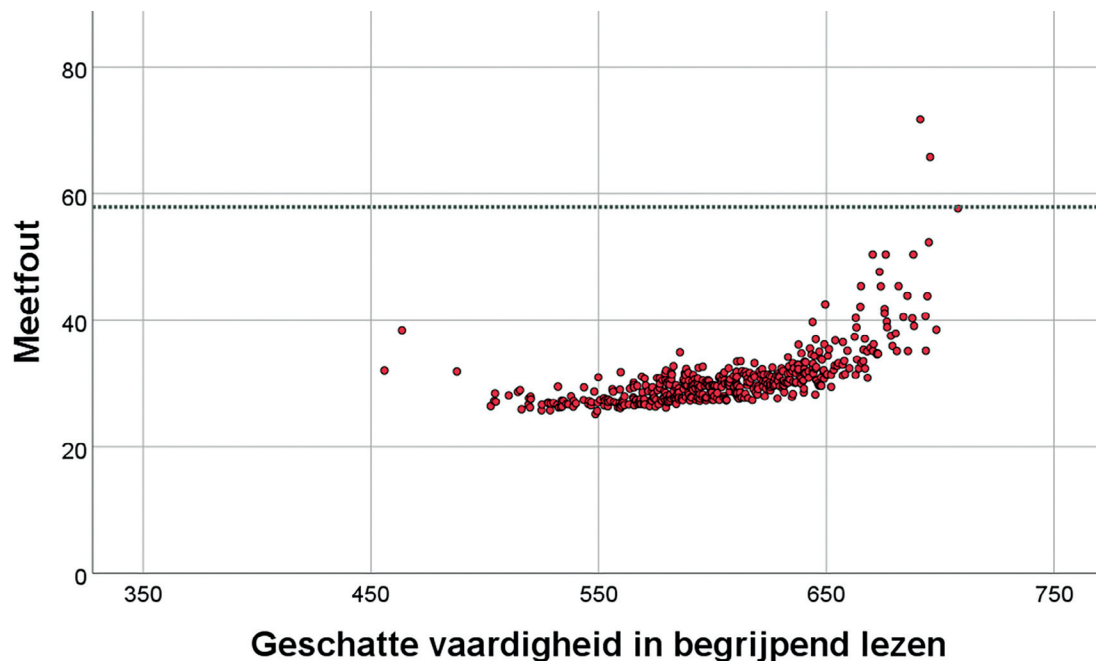
Figuur 13: Somscore toetsboekje 9



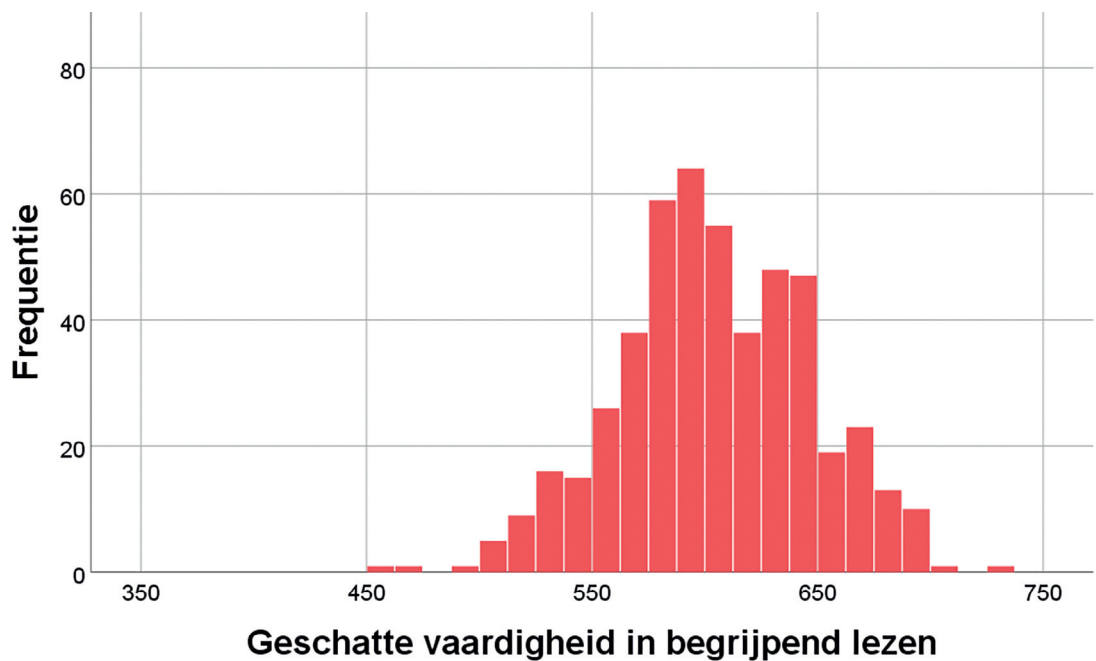
Figuur 14: Somscore toetsboekje 10



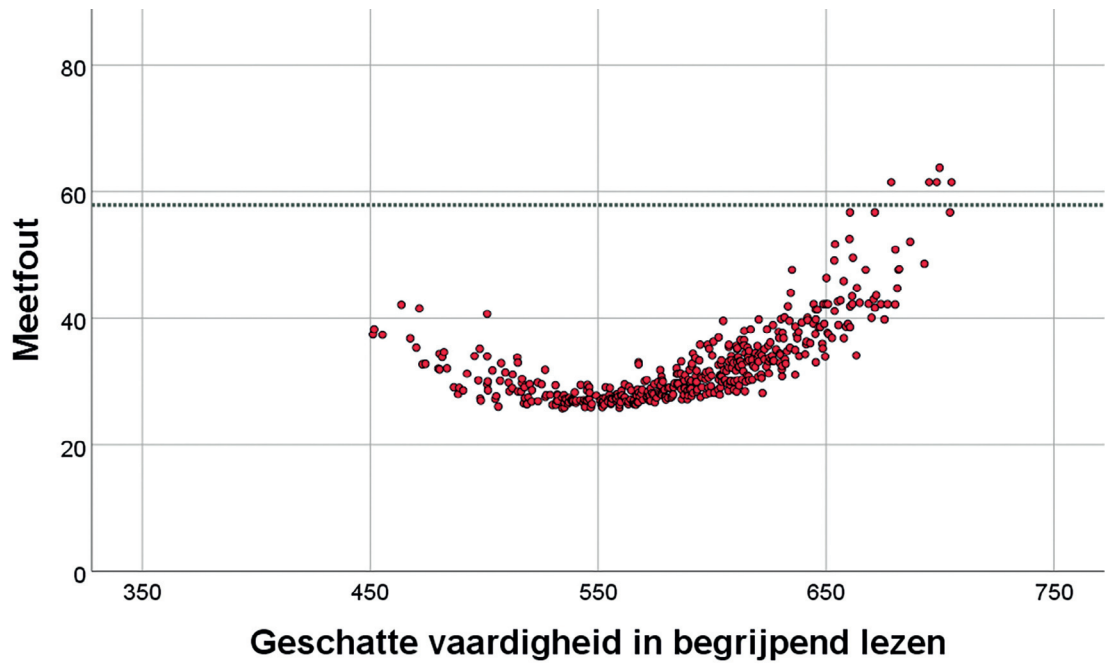
Figuur 15: Scatterplot geschatte vaardigheid en meetfout toetsboekje 1



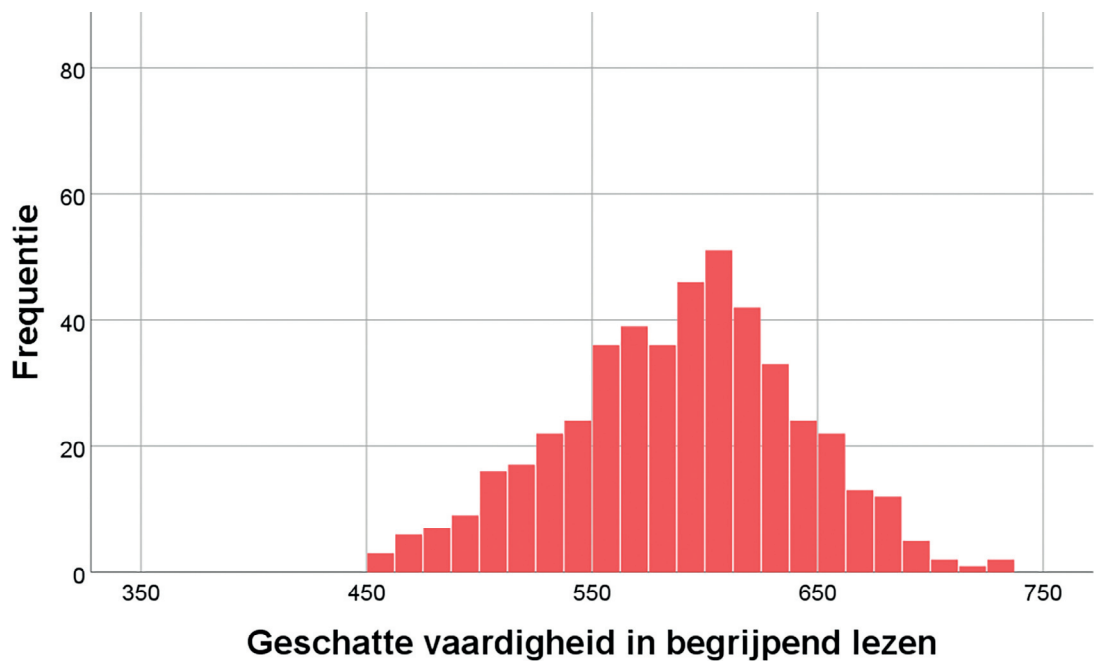
Figuur 16: Histogram geschatte vaardigheid toetsboekje 1



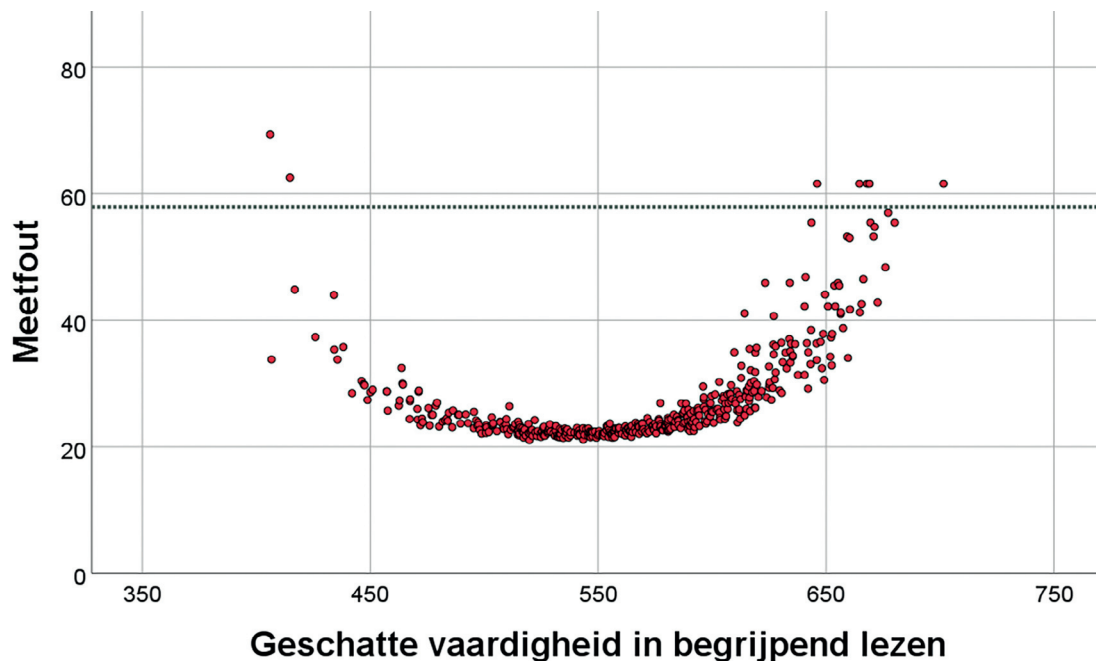
Figuur 17: Scatterplot geschatte vaardigheid en meetfout toetsboekje 2



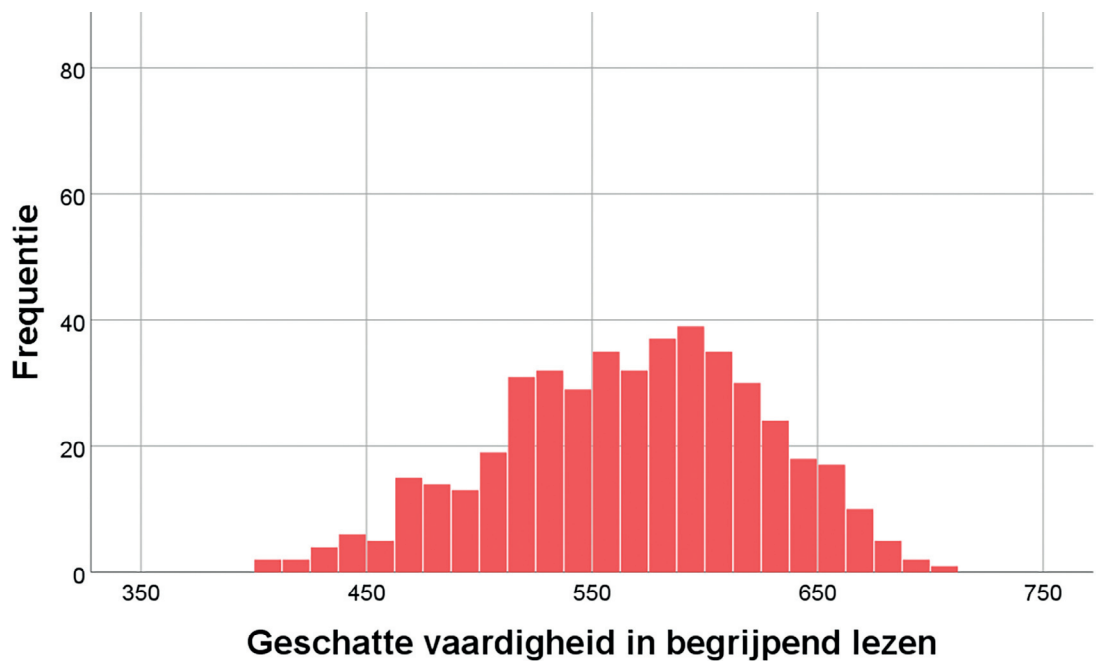
Figuur 18: Histogram geschatte vaardigheid toetsboekje 2



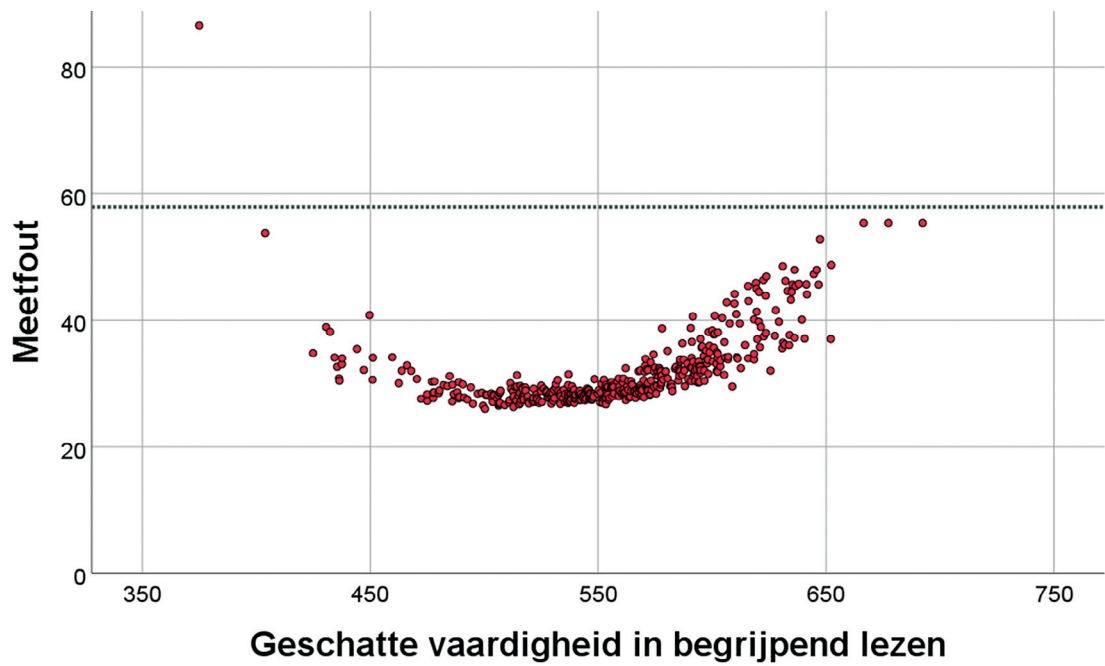
Figuur 19: Scatterplot geschatte vaardigheid en meetfout toetsboekje 3



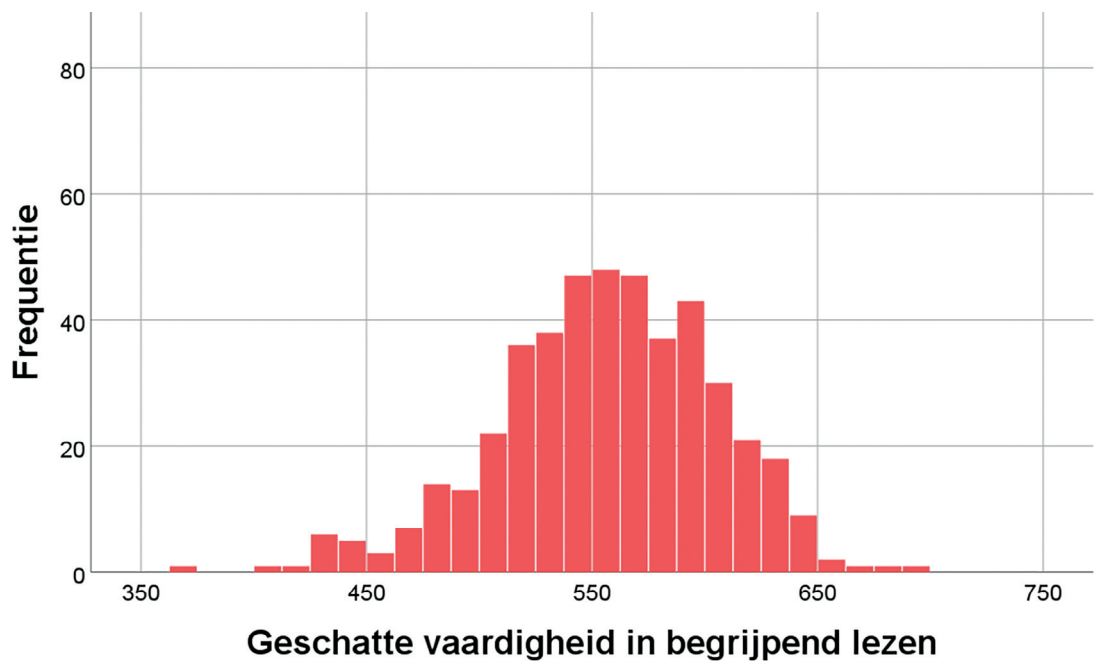
Figuur 20: Histogram geschatte vaardigheid toetsboekje 3



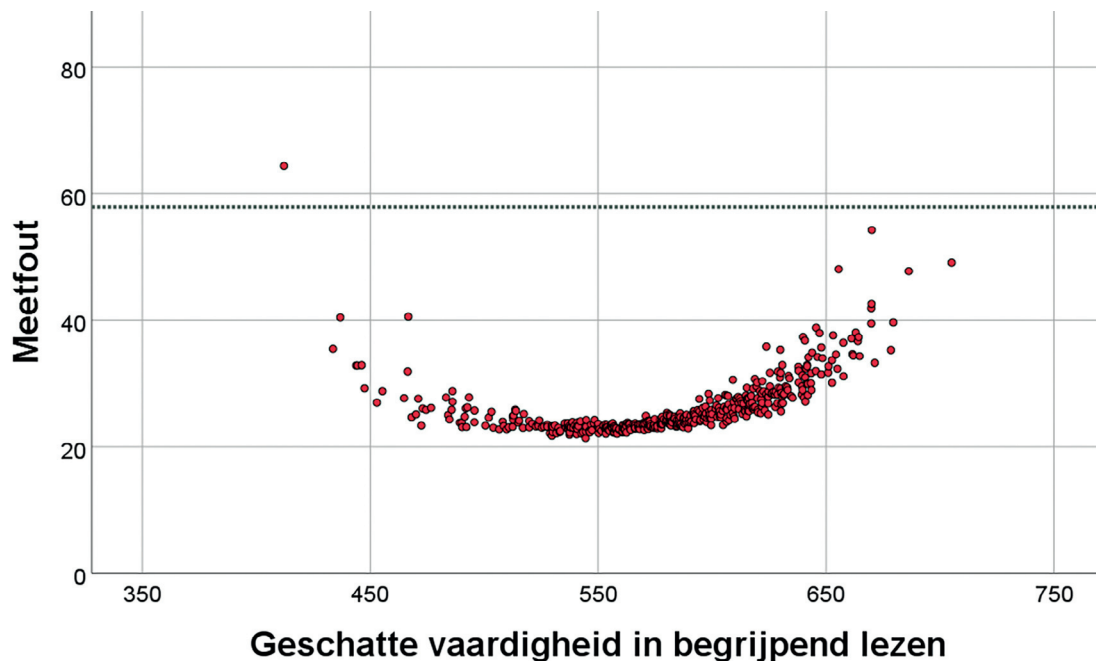
Figuur 21: Scatterplot geschatte vaardigheid en meetfout toetsboekje 4



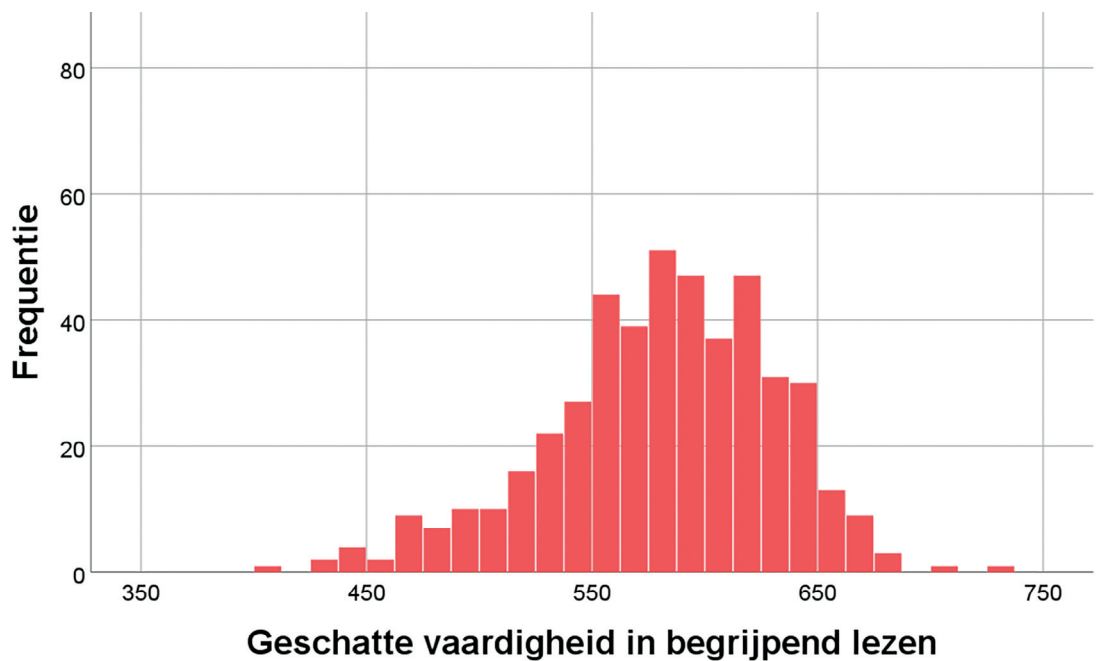
Figuur 22: Histogram geschatte vaardigheid toetsboekje 4



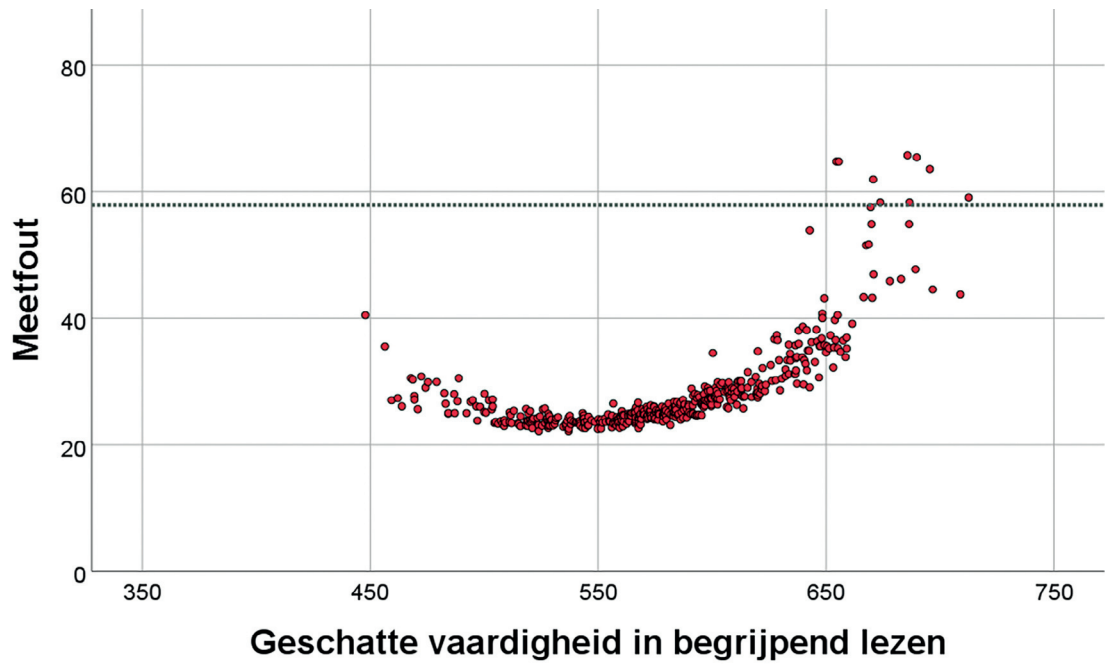
Figuur 23: Scatterplot geschatte vaardigheid en meetfout toetsboekje 5



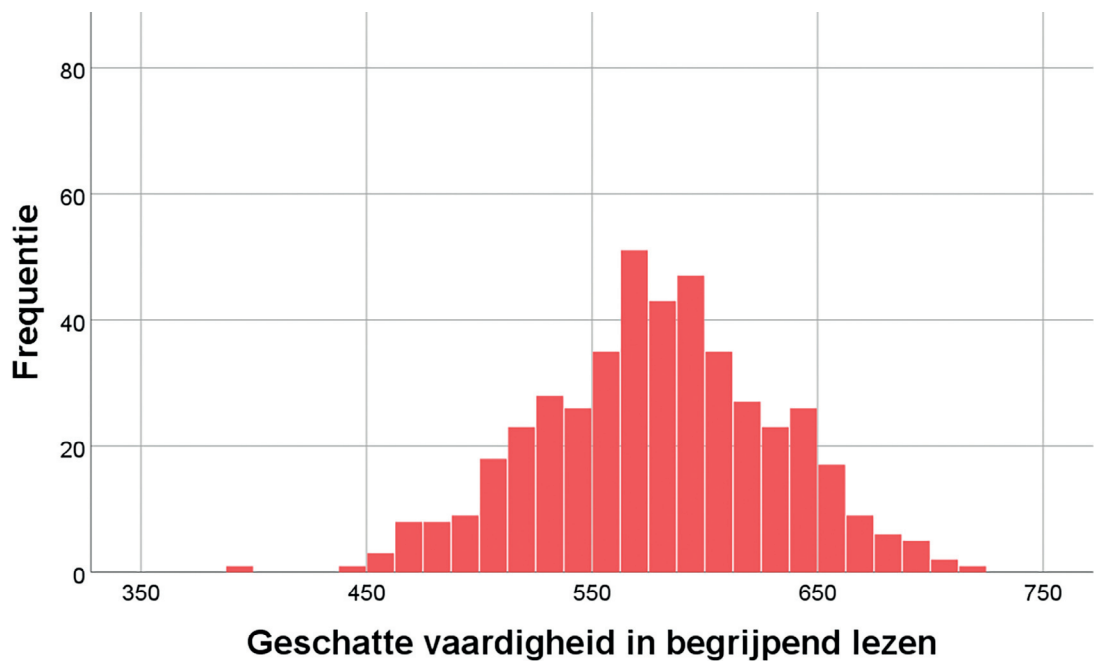
Figuur 24: Histogram geschatte vaardigheid toetsboekje 5



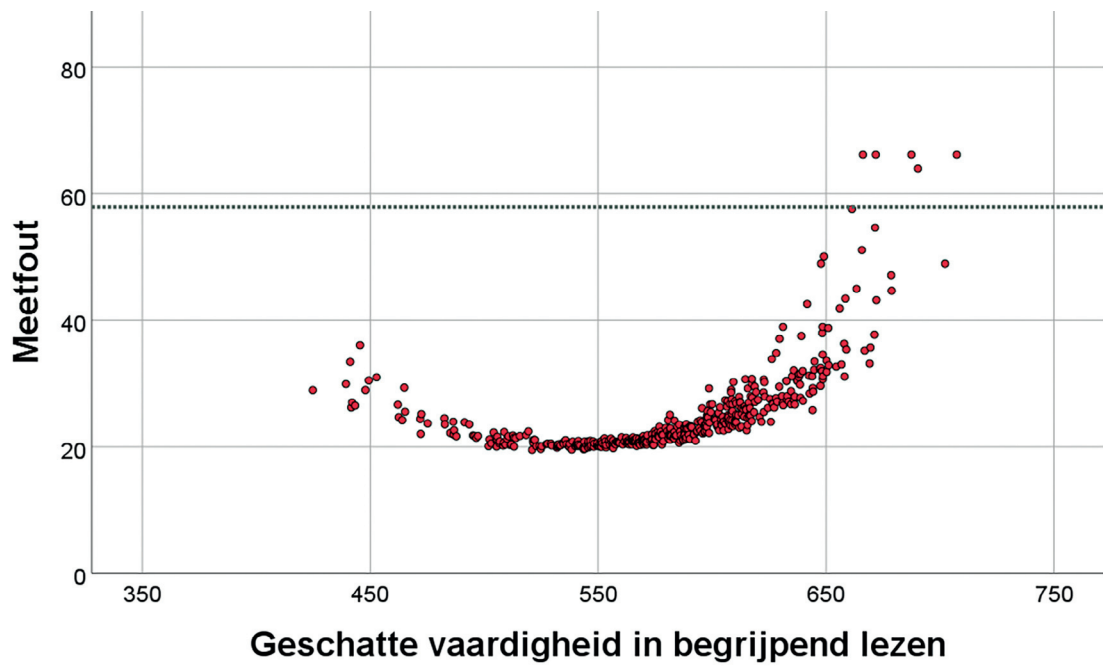
Figuur 25: Scatterplot geschatte vaardigheid en meetfout toetsboekje 6



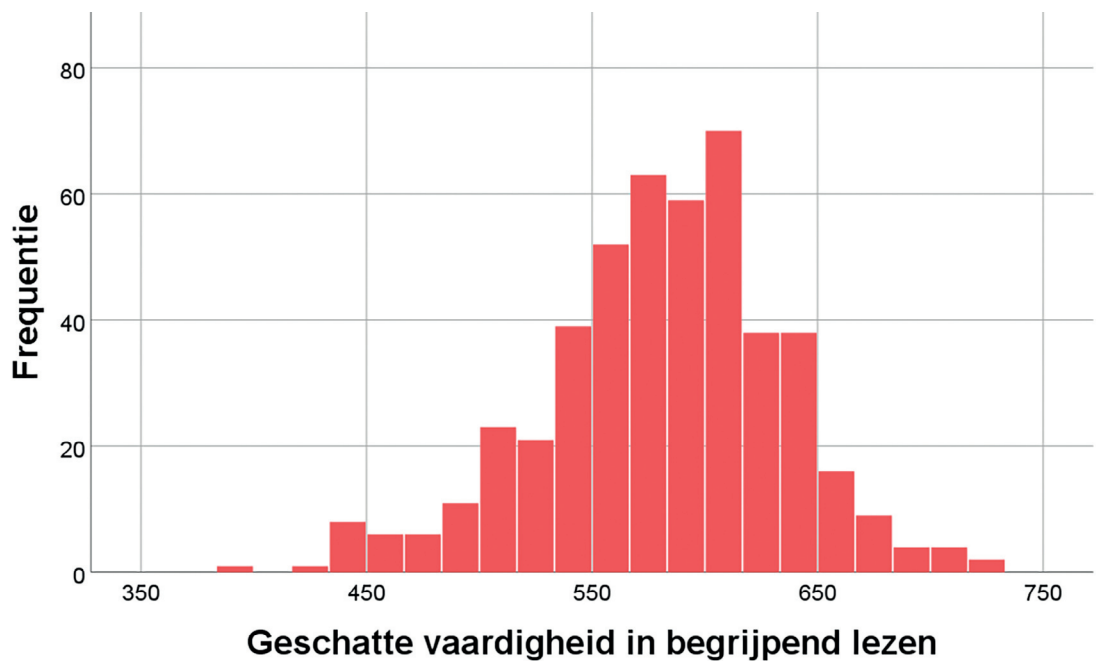
Figuur 26: Histogram geschatte vaardigheid toetsboekje 6



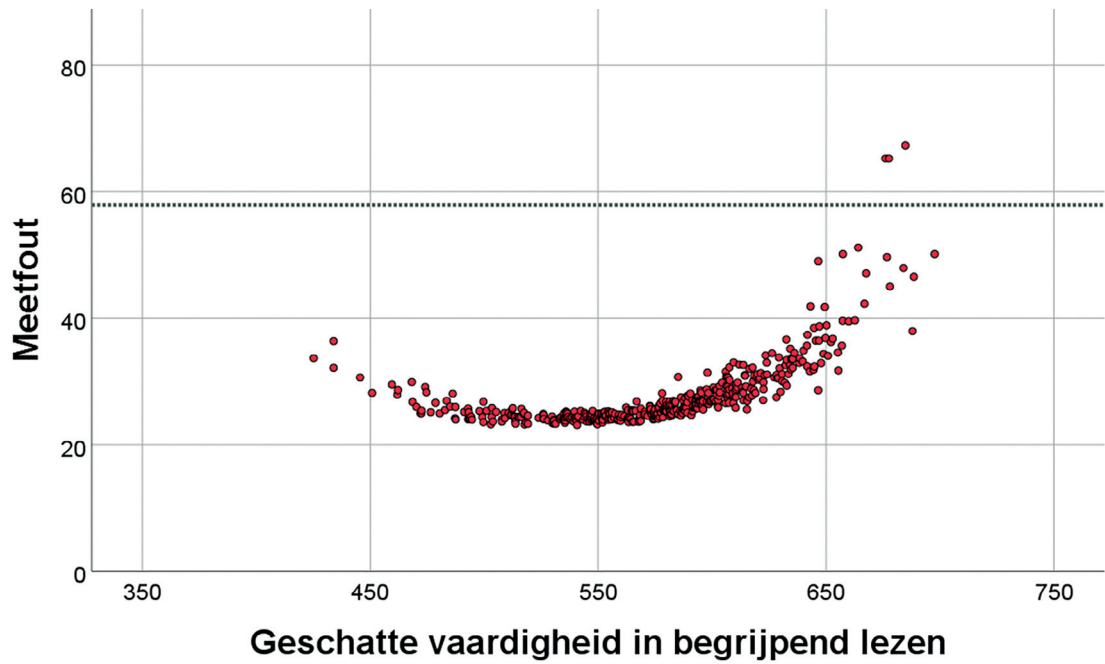
Figuur 27: Scatterplot geschatte vaardigheid en meetfout toetsboekje 7



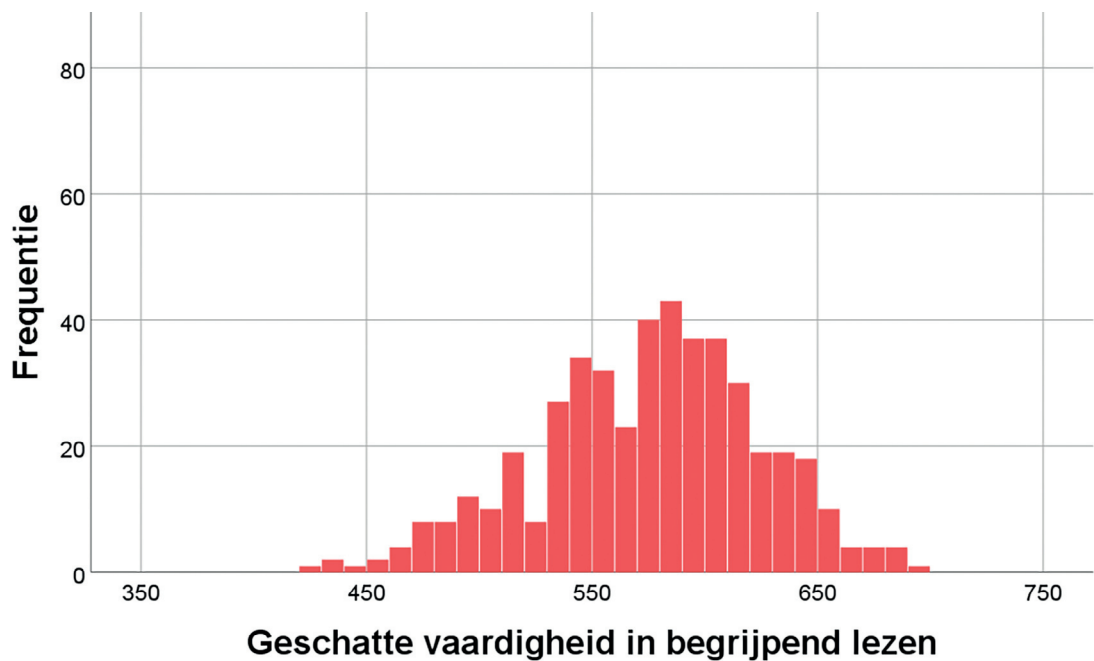
Figuur 28: Histogram geschatte vaardigheid toetsboekje 7



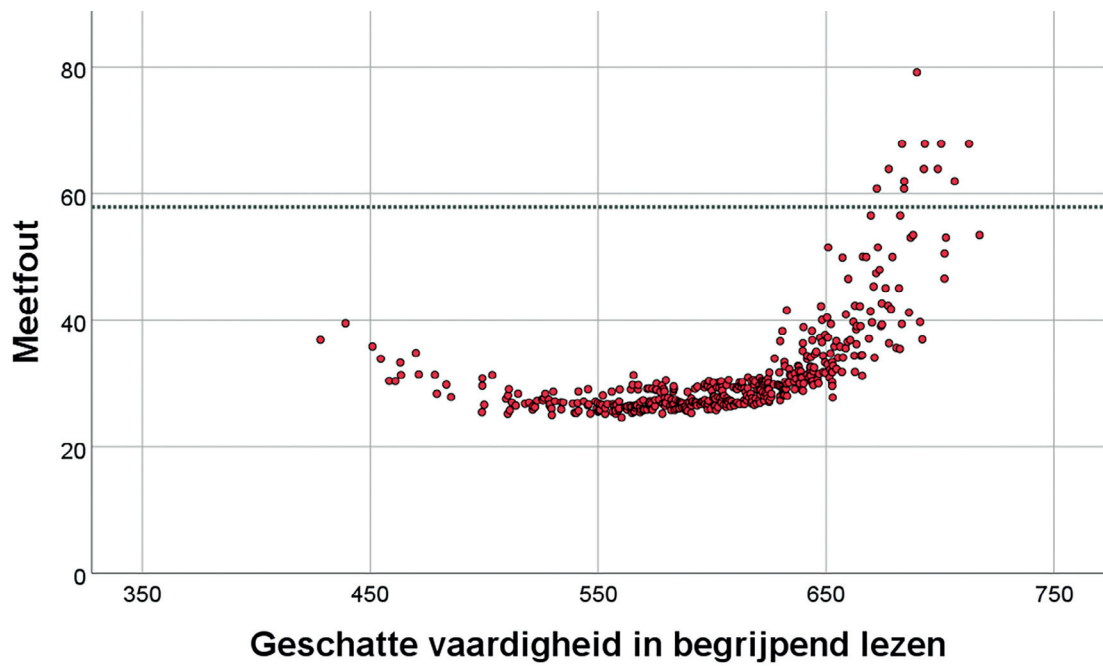
Figuur 29: Scatterplot geschatte vaardigheid en meetfout toetsboekje 8



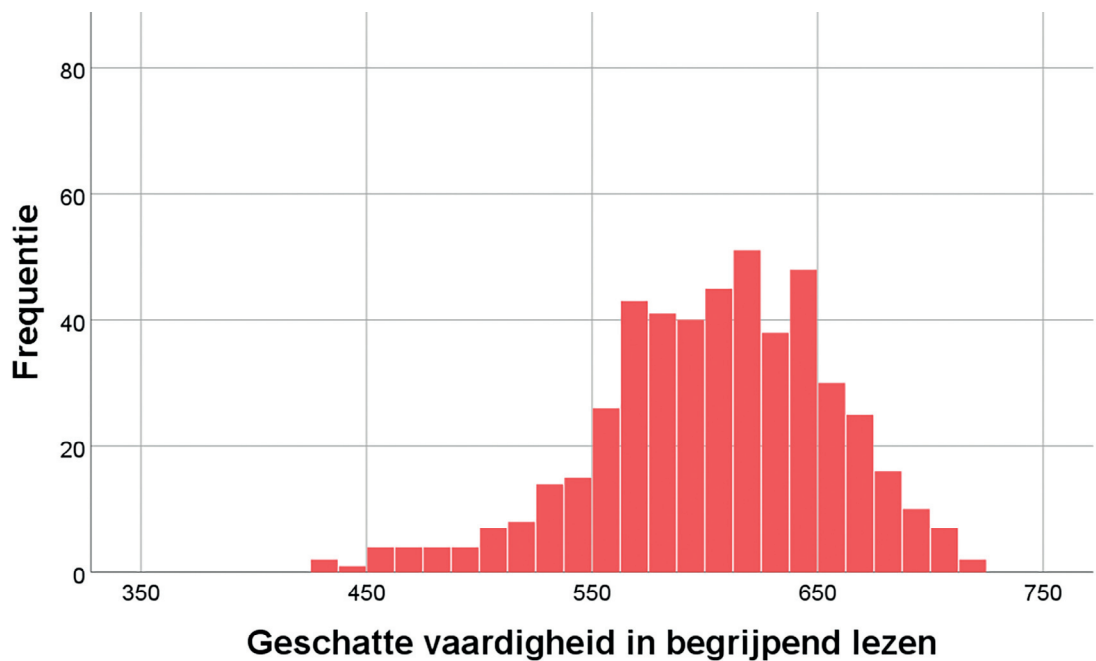
Figuur 30: Histogram geschatte vaardigheid toetsboekje 8



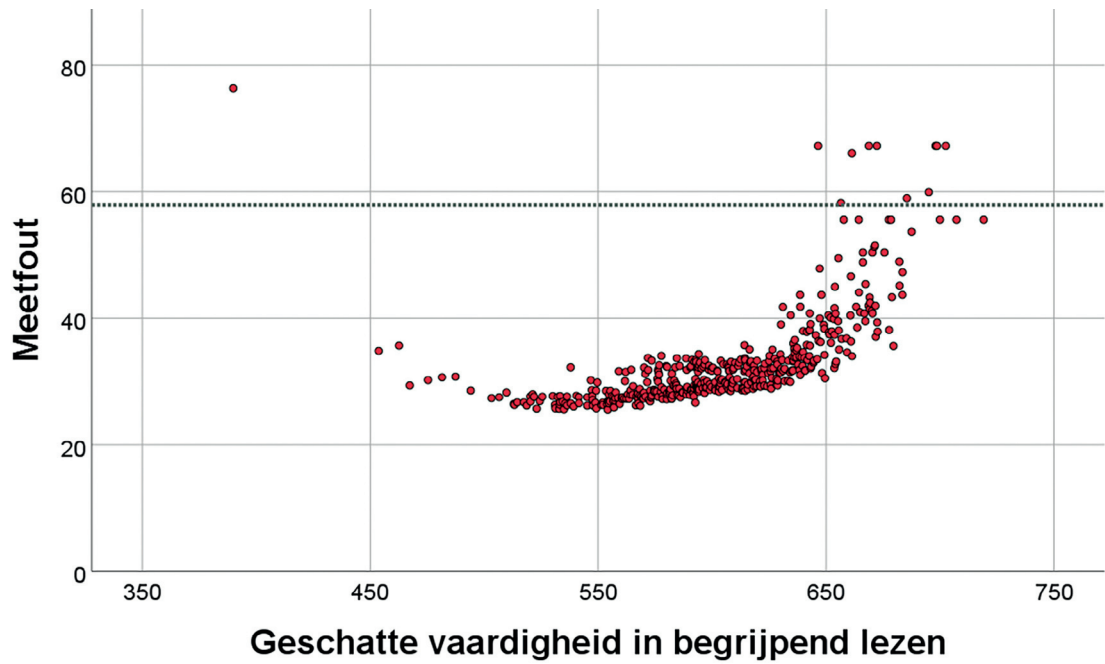
Figuur 31: Scatterplot geschatte vaardigheid en meetfout toetsboekje 9



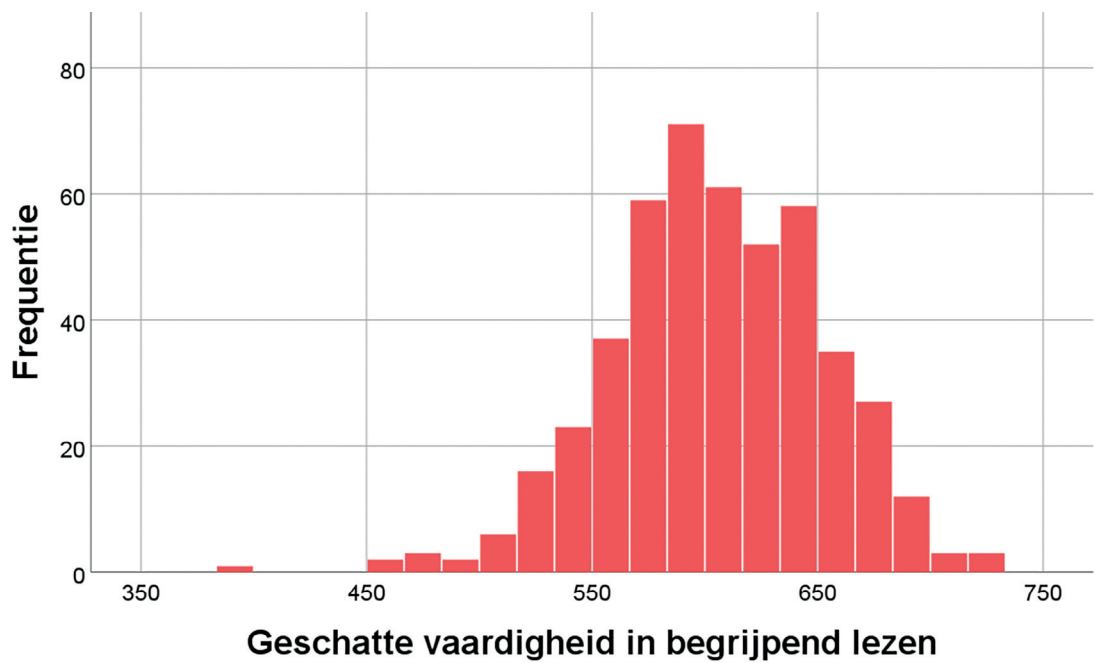
Figuur 32: Histogram geschatte vaardigheid toetsboekje 9



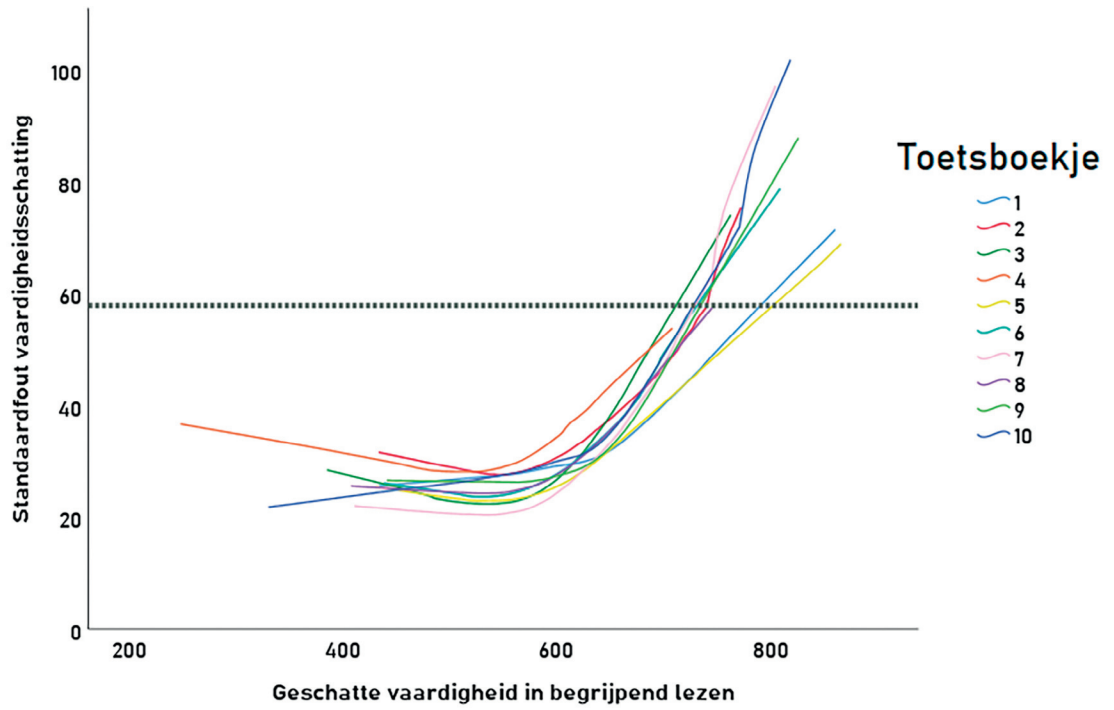
Figuur 33: Scatterplot geschatte vaardigheid en meetfout toetsboekje 10



Figuur 34: Histogram geschatte vaardigheid toetsboekje 10



Figuur 35: Meetfout per toetsboekje naargelang geschatte vaardigheid



4 Resultaten herhalingsmeting 2018

>> In dit hoofdstuk onderzoeken we het niveau van begrijpend lezen van de Vlaamse leerlingen op het einde van het zesde leerjaar in 2018. We starten met analyses naar het algemeen niveau voor begrijpend lezen van leerlingen. Hierbij wordt ook gekeken naar de vier internationale 'standaarden' van prestatieniveaus zoals gedefinieerd door de IEA (*benchmarks*). De resultaten van de herhalingsmeting in 2018 worden vergeleken met de Vlaamse resultaten van 2016, maar ook met de resultaten van andere landen in 2016. Daarna bekijken we hoe het niveau voor begrijpend lezen verschilt tussen verschillende leerlingengroepen. Aansluitend zien we of er verschillen zijn tussen 2016 en 2018 voor de Vlaamse leerlingen in andere uitkomsten zoals attitudes omtrent lezen. We bekijken dan ook of deze als verklarende factor voor de resultaten gebruikt kunnen worden.

In dit hoofdstuk wordt steeds naar de steekproef van 126 scholen gekeken die deel uitmaakten van de herhalingsmeting in 2018. We geven steeds de resultaten van deze scholen weer in 2016 (VLAANDEREN(126)) en 2018 (VLAANDEREN2018). Ook de resultaten van de volledige steekproef van 148 scholen die deelnamen aan PIRLS 2016 worden steeds getoond (VLAANDEREN).

Vlaanderen staat dus driemaal in de tabellen:

VLAANDEREN(126)	Score PIRLS2016, 4e leerjaar	4679 leerlingen in 126 scholen in 2016
VLAANDEREN2018	Score PIRLS2018, 6e leerjaar	4615 leerlingen in 126 scholen in 2018
VLAANDEREN	Score PIRLS2016, 4e leerjaar	5198 leerlingen in 156 scholen in 2016

Verschillen werden steeds getoetst op significantie aan de hand van de standaardfout die geschat werd zoals beschreven in [Hoofdstuk 2](#).

In dit hoofdstuk gebruiken we de term "leerwinst" niet, omdat we die term behouden voor de leerlingen die tweemaal een toets begrijpend lezen invulden. Maar de tabellen in dit hoofdstuk geven wel al aan in hoeverre het Vlaamse gemiddelde voor begrijpend lezen gestegen is tussen het 4e en het 6e leerjaar. De analyses over de leerwinst volgt in [Hoofdstuk 5](#).

4.1 Het gemiddelde prestatieniveau van leerlingen

Voor elk land dat deelnam aan PIRLS 2016 wordt in [Tabel 29](#) het gemiddelde prestatieniveau van de leerlingen, samen met de verdeling van de behaalde scores, weergegeven. Daarbij staan nu ook resultaten in 2016 en 2018 van de 126 scholen die deelnamen aan de herhalingsmeting. De landen zijn gerangschikt van hoog naar laag op basis van hun gemiddelde score.

Uit [Tabel 29](#) blijkt dat Vlaanderen, met een gemiddeld prestatieniveau van 525, het behoorlijk slecht deed voor begrijpend lezen in 2016. Hoewel deze score significant hoger ligt dan het PIRLS 2001-gemiddelde van 500 punten, behaalt Vlaanderen hiermee slechts een 32e plaats in de internationale ranking van de 45 deelnemende landen. Binnen de West-Europese ranking van 14 landen staat Vlaanderen met deze score op een 14e plaats, net voor Frankrijk, Wallonië (FG) en Malta (zie ook Tielmans et al., 2017).

Wanneer we de steekproef van de herhalingsmeting in 2016 bekijken zien we een iets hoger gemiddeld prestatieniveau van 526. Het verschil met de oorspronkelijke steekproef is echter niet significant. De steekproef van de herhalingsmeting staat daarbij ook op dezelfde plaats in de internationale ranking als de oorspronkelijke steekproef. We beschouwen dit als een extra indicatie dat de steekproef van de herhalingsmeting niet merkbaar verschilt van de oorspronkelijke steekproef.

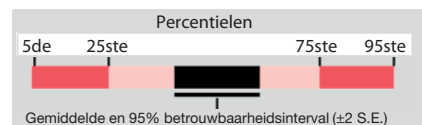
Het gemiddeld prestatieniveau van de herhalingsmeting van de 126 scholen in 2018 is 586. De Vlaamse leerlingen in het zesde leerjaar scoren daarmee net hoger dan de leerlingen in het vierde leerjaar van de hoogst presterende landen. Daarbij moeten we opmerken dat de steekproef van 126 scholen al in 2016 iets sterker was dan de originele steekproef. Verder weten we uit het [Hoofdstuk 2](#) dat in de 126 scholen van de herhalingsmeting veel lager scorende leerlingen voor het zesde leerjaar verdwijnen. Bijgevolg is het waarschijnlijk dat het echte gemiddelde van de oorspronkelijke Vlaamse steekproef in PIRLS 2016 in het zesde leerjaar enigszins lager is dan 586.

De figuur in [Tabel 29](#) toont per land of regio de *verdeling* van de prestaties voor begrijpend lezen. Zo wordt voor elk land of regio getoond binnen welk vaardigheidsinterval 90% van de leerlingen zich bevinden. In het lichtroze wordt aangegeven binnen welk vaardigheidsinterval 50% van de leerlingen zich bevinden. De extremen, de groep van de 5% zwakst en de 5% sterkst presterende leerlingen, worden buiten beschouwing gelaten. In het zwart wordt het 95% betrouwbaarheidsinterval gegeven waarin het gemiddelde van het land of de regio zich bevindt.

Tabel 29: Verdeling van de prestaties voor begrijpend lezen (ASRREA01-05) per land

Landen	Gemiddelde score (S.E.)			Schaalscore begrijpend lezen	
VLAANDEREN2018	586	(1.6)	▲		
Rusland	581	(2.2)	▲		
Singapore	576	(3.2)	▲		
Hong Kong	569	(2.7)	▲		
Ierland	567	(2.5)	▲		
Finland	566	(1.8)	▲		
Polen	565	(2.1)	▲		
Noord-Ierland	565	(2.2)	▲		
Noorwegen (5)	559	(2.3)	▲		
Chinees Taipei	559	(2.0)	▲		
Engeland	559	(1.9)	▲		
Letland	558	(1.7)	▲		
Zweden	555	(2.4)	▲		
Hongarije	554	(2.9)	▲		
Bulgarije	552	(4.2)	▲		
Verenigde Staten	549	(3.1)	▲		
Litouwen	548	(2.6)	▲		
Italië	548	(2.2)	▲		
Denemarken	547	(2.1)	▲		
Macao	546	(1.0)	▲		
Nederland	545	(1.7)	▲		
Australië	544	(2.5)	▲		
Tsjechië	543	(2.1)	▲		
Canada	543	(1.8)	▲		
Slovenië	542	(2.0)	▲		
Oostenrijk	541	(2.4)	▲		
Duitsland	537	(3.2)	▲		
Kazakhstan	536	(2.5)	▲		
Slowakije	535	(3.1)	▲		
Israël	530	(2.5)	▲		
Portugal	528	(2.3)	▲		
Spanje	528	(1.7)	▲		
VLAANDEREN(126)	526	(1.8)	▲		
VLAANDEREN	525	(1.9)	▲		
Nieuw-Zeeland	523	(2.2)	▲		
Frankrijk	511	(2.2)	▲		
PIRLS 2001-gemiddelde	500	(0.0)			
Wallonië (FG)	497	(2.6)			
Chili	494	(2.5)	▼		
Georgië	488	(2.8)	▼		
Trinidad en Tobago	479	(3.3)	▼		
Azerbeidjan	472	(4.2)	▼		
Malta	452	(1.8)	▼		
Ver. Arabische Emiraten	450	(3.2)	▼		
Bahrein	446	(2.3)	▼		
Qatar	442	(1.8)	▼		
Saoedi-Arabië	430	(4.2)	▼		
Oman	418	(3.3)	▼		

▲ Gemiddelde significant hoger dan PIRLS-gemiddelde
 ▼ Gemiddelde significant lager dan PIRLS-gemiddelde



Voor Vlaanderen in PIRLS 2016 blijkt uit de verdeling van de prestaties voor begrijpend lezen dat het verschil tussen zwak en sterk presterende leerlingen eerder klein is. In cijfers wordt deze spreiding van de prestaties rond het gemiddelde uitgedrukt in de standaarddeviatie (SD), die voor Vlaanderen 61 punten bedraagt. De Vlaamse leerlingen vormen voor begrijpend lezen dus een relatief homogene groep. Enkel de Nederlandse leerlingen vormen met een standaarddeviatie van 60 punten een nog meer homogene groep.

De 126 scholen die deelnamen aan de herhalingsmeting behaalden in 2016 een SD van 60 punten, een niet merkbaar kleinere spreiding dan de oorspronkelijke steekproef. Opvallend is dat de spreiding in de vaardigheidsscores van de herhalingsmeting in 2018 met een SD van 57gelijkaardig is aan de spreiding in 2016. De spreiding in vaardigheidsscores verandert dus nauwelijks tussen het vierde en het zesde leerjaar.

Als laatste vinden we dat de vierdejaars van elk land of regio behalve Rusland significant lager scoren dan de zesdejaars in Vlaanderen. We herhalen echter onze opmerking dat de steekproef van 126 scholen al in 2016 iets sterker was dan de originele steekproef en dat veel lager scorende leerlingen tegen het zesde leerjaar verdwijnen.

4.2 Leeftijdverdeling

Tegenover die grote homogeniteit van de Vlaamse prestaties staat echter een relatief grote heterogeniteit van de Vlaamse steekproef van 2016 inzake de leeftijd. De standaarddeviatie bedraagt voor Vlaanderen 0,57 jaar. Enkel de leerlingen uit Trinidad en Tobago en de Verenigde Arabische Emiraten worden met standaarddeviaties van respectievelijk 0,65 jaar en 0,62 jaar gekenmerkt door een nog grotere leeftijdsspreiding. De meest homogene leeftijdsgroepen kunnen gevonden worden in Engeland (0,29 jaar), Noorwegen (5) (0,30 jaar), Zweden (0,30 jaar) en Noord-Ierland (0,30 jaar). Dit zijn landen waar zittenblijven eerder uitzonderlijk wordt toegepast.

Voor de 126 scholen die deelnamen aan de herhalingsmeting worden in 2016 andere resultaten gevonden. De standaarddeviatie is nu slechts 0.46. Dit komt omdat de oudere leerlingen vooral in het BuLO zaten. Aangezien de steekproef van 126 scholen geen BuLO-scholen bevat, wordt dus ook de spreiding in leeftijden gereduceerd. In deze steekproef zien we in 2018 ook een kleinere heterogeniteit voor leeftijd, de standaarddeviatie is nu 0.42. **Tabel 30** toont de geboortejaren voor de verschillende Vlaamse steekproeven.

Tabel 30: Leeftijdverdeling (volgens geboortjaar) in PIRLS 2016, PIRLS 2016 (126) en PIRLS 2018

	2003	2004	2005	2006	2007	2008	TOTAAL
VLAANDEREN2018	1	16	470	4065	63	0	4615
VLAANDEREN(126)	4	54	531	4029	61	1	4679
VLAANDEREN	54	108	638	4332	65	1	5198

De grote leeftijdsspreiding in de Vlaamse steekproef in 2016 kan wellicht verklaard worden door het vele zittenblijven en de opname van BuLO-leerlingen. Uit internationaal onderzoek (o.a. OECD, 2013) blijkt immers dat Vlaanderen een relatief groot aantal zittenblijvers kent in vergelijking met verschillende andere deelnemende landen (bijv. de Scandinavische landen). Wat de opname van BuLO-leerlingen betreft, werd gevraagd leerlingen te selecteren die het curriculum van het vierde leerjaar volgen. Dit betekent echter niet dat deze leerlingen ook daadwerkelijk in het vierde leerjaar zitten. Doorgans zijn de BuLO-leerlingen ouder dan de leerlingen in het vierde leerjaar van het regulier lager onderwijs.

4.3 De internationale prestatieniveaus

In de vorige paragraaf werd reeds de spreiding in de prestaties voor begrijpend lezen besproken. Een andere werkwijze om verschillen tussen leerlingen te verduidelijken, is het berekenen van het percentage leerlingen dat de internationale prestatieniveaus (minstens) bereikt. In PIRLS worden er vier zulke prestatieniveaus, ook wel *benchmarks* of standaarden genoemd, onderscheiden, met name een gevorderd, hoog, middelmatig en laag niveau. Op basis van inhoudelijke en statistische analyses vormen deze vier prestatieniveaus referentiepunten binnen de toetsscores, die vastgelegd zijn in termen van het soort items die leerlingen correct kunnen beantwoorden (zie Tielemans et al., 2019).

In **Tabel 31** wordt een overzicht gegeven van de percentages leerlingen per land die de internationale prestatieniveaus (minstens) bereiken. Daarnaast wordt ook de internationale mediaan weergegeven, die gehanteerd kan worden als een belangrijk referentiepunt.

Tabel 31: Percentages leerlingen die de internationale prestatieniveaus voor begrijpend lezen (ASRREA01-05) hebben bereikt

Landen	% leerlingen die internationale standaarden bereiken	Gevorderd (625)	Hoog (550)	Middelmatig (475)	Laag (400)
Singapore		29 (1.6)	66 (1.6)	89 (1.0)	97 (0.5)
Rusland		26 (1.2)	70 (1.3)	94 (0.6)	99 (0.3)
VLAANDEREN2018		24 (1.2)	74 (1.1)	96 (0.3)	100 (0.1)
Noord-Ierland		22 (1.4)	61 (1.3)	87 (0.8)	97 (0.4)
Ierland		21 (1.2)	62 (1.6)	89 (0.9)	98 (0.4)
Polen		20 (1.1)	61 (1.3)	89 (0.7)	98 (0.4)
Engeland		20 (0.9)	57 (1.1)	86 (0.7)	97 (0.4)
Bulgarije		19 (1.3)	55 (2.2)	83 (1.6)	95 (0.9)
Hong Kong		18 (1.3)	65 (1.8)	93 (0.9)	99 (0.3)
Finland		18 (0.8)	62 (1.3)	91 (0.8)	98 (0.3)
Hongarije		17 (1.2)	56 (1.7)	85 (1.0)	97 (0.5)
Verenigde Staten		16 (1.3)	53 (1.6)	83 (1.2)	96 (0.5)
Australië		16 (1.0)	51 (1.4)	81 (1.0)	94 (0.5)
Noorwegen (5)		15 (0.9)	58 (1.7)	90 (0.9)	99 (0.3)
Chinees Taipei		14 (1.1)	59 (1.5)	90 (0.7)	98 (0.2)
Zweden		14 (1.4)	57 (1.6)	88 (0.9)	98 (0.3)
Letland		14 (1.0)	57 (1.3)	90 (0.8)	99 (0.2)
Israël		13 (0.9)	46 (1.3)	75 (1.0)	91 (0.7)
Canada		13 (0.7)	50 (1.0)	83 (0.9)	96 (0.4)
Litouwen		12 (0.9)	52 (1.6)	86 (1.1)	97 (0.5)
Slovenië		11 (0.8)	49 (1.3)	83 (0.9)	96 (0.5)
Denemarken		11 (1.0)	52 (1.3)	86 (1.0)	97 (0.4)
Nieuw-Zeeland		11 (0.6)	41 (1.2)	73 (1.0)	90 (0.7)
Duitsland		11 (0.8)	47 (1.4)	81 (1.4)	95 (1.0)
Italië		11 (0.8)	52 (1.7)	87 (1.0)	98 (0.5)
Slowakije		10 (0.8)	47 (1.4)	81 (1.3)	93 (1.1)
Internationale mediaan		10	47	82	96
Tsjechië		10 (0.7)	49 (1.3)	85 (0.9)	97 (0.5)
Macao		10 (0.6)	50 (0.8)	86 (0.5)	98 (0.3)
Oostentijk		8 (0.8)	47 (1.5)	84 (1.1)	98 (0.4)
Nederland		8 (0.6)	48 (1.3)	88 (0.9)	99 (0.3)
Kazachstan		7 (0.8)	42 (1.8)	84 (1.5)	98 (0.3)
Portugal		7 (0.9)	38 (1.3)	79 (1.3)	97 (0.4)
Spanje		6 (0.4)	39 (0.9)	80 (1.0)	97 (0.6)
Ver. Arabische Emiraten		5 (0.3)	20 (1.0)	43 (1.4)	68 (1.3)
Trinidad en Tobago		4 (0.5)	24 (1.2)	55 (1.7)	80 (1.2)
VLAANDEREN(126)		4 (0.4)	36 (1.1)	81 (1.2)	97 (0.4)
VLAANDEREN		4 (0.4)	35 (1.3)	80 (1.3)	97 (0.4)
Frankrijk		4 (0.6)	30 (1.3)	72 (1.2)	94 (0.5)
Chili		3 (0.4)	25 (1.3)	61 (1.5)	87 (1.1)
Qatar		3 (0.3)	17 (0.6)	42 (1.1)	66 (0.9)
Wallonië (FG)		3 (0.4)	22 (1.2)	65 (1.4)	92 (0.9)
Georgië		2 (0.4)	22 (1.3)	60 (1.6)	86 (1.1)
Bahrein		2 (0.3)	14 (0.6)	41 (1.0)	69 (1.0)
Oman		2 (0.3)	10 (0.8)	32 (1.3)	59 (1.3)
Malta		1 (0.2)	13 (0.7)	45 (1.1)	73 (0.7)

- Gevorderd
- Hoog
- Middelmatig
- Laag

Uit **Tabel 31** blijkt dat 4% van de Vlaamse leerlingen in 2016 voor begrijpend lezen *het gevorderde prestatieniveau* bereikt. Vlaanderen staat daarmee, samen met Frankrijk en Trinidad en Tobago, op de 34e plaats in de internationale ranking. Met 10% van de leerlingen en een gedeelde 25e plaats in de ranking doet de internationale mediaan het opvallend beter. Enkel Wallonië (FG) heeft als enig West-Europees land een kopgroep die nog kleiner is dan de Vlaamse. De leerlingen in 126 scholen die deelnamen aan de herhalingsmeting tonen gelijke resultaten aan de totale steekproef in 2016. Er zijn 4% van de leerlingen die het gevorderde prestatieniveau bereiken. De herhalingsmeting in 2018 toont dat twee jaar later in het zesde leerjaar 24% van de leerlingen het gevorderde prestatieniveau bereiken. Hoewel dit hoog is, is het lager dan Singapore en Rusland in het vierde leerjaar.

Met inbegrip van de vermelde 4% behaalt 35% van de Vlaamse leerlingen minstens *het hoge prestatieniveau*. Vlaanderen staat daarmee op de 33ste plaats. In vergelijking met de internationale mediaan heeft Vlaanderen 12% minder leerlingen die dit hoge prestatieniveau bereiken. Van de West-Europese landen doen enkel Frankrijk en Wallonië (FG) slechter. De leerlingen in de 126 scholen die deelnamen aan de herhalingsmeting tonen gelijkaardige resultaten aan de totale steekproef in 2016. Er zijn 36% die minstens het hoge prestatieniveau bereiken. De herhalingsmeting in 2018 toont dat twee jaar later in het zesde leerjaar 74% van de leerlingen het hoge prestatieniveau bereiken. Er zijn geen landen die in het vierde leerjaar een hoger cijfer bereiken.

Het middelmatige prestatieniveau wordt in Vlaanderen wordt in 2016 door 80% van de leerlingen bereikt. Vlaanderen staat met dit resultaat samen met Spanje op een gedeelde 30e plaats in de internationale ranking. Hoewel de internationale mediaan het nog steeds beter doet dan Vlaanderen, wordt het verschil ditmaal beperkt tot slechts 2%. Binnen West-Europa doet Vlaanderen het net beter dan Portugal, Frankrijk en Wallonië (FG). De leerlingen in de 126 scholen die deelnamen aan de herhalingsmeting tonen gelijkaardige resultaten aan de totale steekproef in 2016. Er zijn 81% leerlingen die minstens het middelmatige prestatieniveau bereiken. De herhalingsmeting in 2018 toont dat twee jaar later in het zesde leerjaar 96% van de leerlingen het middelmatige prestatieniveau bereiken. Er zijn geen landen die in het vierde leerjaar een hoger cijfer bereiken.

Het lage prestatieniveau wordt door 97% van de Vlaamse leerlingen bereikt in 2016. Net zoals Singapore, Noord-Ierland, Hongarije, Litouwen, Denemarken, Portugal en Spanje slaagt Vlaanderen (15e plaats) er daarmee in de ondermaatse prestaties tot een minimum van 3% van de leerlingen te beperken. Vlaanderen heeft hiermee ook voor het eerst de internationale mediaan overschreden. West-Europese landen die het nog beter doen dan Vlaanderen (11e plaats) zijn Ierland, Finland, Noorwegen (5), Zweden, Italië, Oostenrijk en Nederland. De 126 scholen die deelnamen aan de herhalingsmeting tonen gelijkaardige resultaten aan de totale steekproef in 2016. Er zijn 98% die minstens het lage prestatieniveau bereiken. De herhalingsmeting in 2018 toont dat twee jaar later in het zesde leerjaar 100% van de leerlingen het lage prestatieniveau overstijgen. Er zijn geen landen die in het vierde leerjaar een hoger cijfer bereiken.

De kleine kop- en staartgroep in Vlaanderen over de twee afnames wijzen er mogelijk op dat het huidige onderwijsstelsel beter tegemoetkomt aan de noden van de zwakke lezers, maar te weinig inspeelt op die van de sterke lezers.

4.4 Verschillen tussen leerlinggroepen

In wat volgt wordt onderzocht hoe bepaalde eigenschappen van leerlingen samenhangen met hun prestaties. Het gaat om (1) geslacht, (2) socio-economische status (aantal boeken thuis), (3) de houding van leerlingen tegenover lezen, en (4) de zelfzekerheid van leerlingen over hun leesvaardigheid.

4.4.1 Geslacht

In **Tabel 32** wordt een overzicht gegeven van de gemiddelde prestaties voor begrijpend lezen opgedeeld naar het geslacht van de leerlingen. Zowel de originele Vlaamse steekproef van PIRLS 2016 wordt getoond, als de resultaten van de 126 scholen van de steekproef van de herhalingsmeting in 2016 en 2018.

Tabel 32 toont dat in het vierde leerjaar Vlaamse meisjes met een gemiddelde score van 530 significant beter presteren voor begrijpend lezen dan de Vlaamse jongens die een gemiddelde score van 520 behalen. Een dergelijke significante voorsprong voor meisjes wordt overigens vastgesteld in alle West-Europese landen, met uitzondering van Portugal. Dit land slaagt er met andere woorden het best in de ongelijke prestaties tussen jongens en meisjes te beperken.

Ook Vlaanderen houdt de prestatieverschillen tussen beide geslachten redelijk beperkt. Binnen West-Europa staat Vlaanderen immers op een gedeelde vijfde plaats, samen met Nederland en na Portugal, Oostenrijk, Italië en Frankrijk. Vooral de Zuid-Europese landen doen het op dit punt goed. De grootste kloof tussen jongens en meisjes situeert zich voornamelijk in de Scandinavische landen, waarbij de grootste prestatieverschillen worden gevonden in Noorwegen (5) en Finland.

De 126 scholen die deelnamen aan de herhalingsmeting tonen eerder gelijkaardige resultaten aan de totale steekproef in het vierde leerjaar in 2016. De Vlaamse meisjes hebben hier een gemiddelde score van 530 terwijl de jongens een gemiddelde score hebben van 521, wat leidt tot een verschil van 9 punten. Deze scores leiden tot dezelfde plaats in de rangorde voor het verschil tussen jongens en meisjes als bij de originele steekproef.

De herhalingsmeting van de 126 scholen in 2018 toont dat in het zesde leerjaar zowel meisjes als jongens een hogere gemiddelde score hebben dan in het vierde leerjaar. Meisjes hebben een gemiddelde score van 593 en jongens hebben een gemiddelde score van 578, met een gemiddeld verschil van 14. Het verschil tussen jongens en meisjes is van het vierde naar het zesde leerjaar merkbaar groter geworden.

Tabel 32: Prestaties voor begrijpend lezen (ASRREA01-05) opgedeeld naar geslacht (ITSEX)

Land	Meisjes		Jongens		Absoluut verschil	Verschil tussen de geslachten
	%	Gemiddelde score	%	Gemiddelde score		
Portugal	49 (0.8)	529 (2.7)	51 (0.8)	527 (2.5)	1 (2.7)	1
Oostenrijk	48 (0.9)	544 (2.7)	52 (0.9)	538 (2.7)	6 (2.4)	6 ▲
Italië	49 (0.9)	552 (2.7)	51 (0.9)	544 (2.4)	7 (2.6)	7 ▲
Frankrijk	50 (0.9)	515 (2.6)	50 (0.9)	507 (2.5)	8 (2.7)	8 ▲
Spanje	49 (0.6)	532 (1.4)	51 (0.6)	524 (2.7)	8 (2.5)	8 ▲
VLAANDEREN(126)	51 (0.8)	530 (2.0)	49 (0.8)	521 (2.2)	9 (2.1)	9 ▲
VLAANDEREN	51 (0.9)	530 (2.1)	49 (0.9)	520 (2.3)	10 (2.0)	10 ▲
Nederland	51 (1.1)	550 (1.7)	49 (1.1)	540 (2.3)	10 (2.2)	10 ▲
Wallonië (FG)	50 (1.0)	503 (2.5)	50 (1.0)	492 (3.4)	11 (3.0)	11 ▲
Duitsland	49 (0.7)	543 (3.2)	51 (0.7)	532 (3.7)	11 (2.9)	11 ▲
Ierland	50 (1.6)	572 (2.9)	50 (1.6)	561 (3.3)	12 (3.8)	12 ▲
Denemarken	51 (0.9)	554 (2.6)	49 (0.9)	541 (2.7)	13 (3.1)	13 ▲
VLAANDEREN2018	52 (0.8)	593 (1.7)	48 (0.8)	578 (2.0)	14 (1.7)	14 ▲
Engeland	50 (0.9)	566 (2.2)	50 (0.9)	551 (2.4)	15 (2.8)	15 ▲
Zweden	49 (0.9)	563 (2.7)	51 (0.9)	548 (2.6)	15 (2.5)	15 ▲
Noord-Ierland	50 (1.2)	574 (2.8)	50 (1.2)	555 (2.8)	18 (3.5)	18 ▲
Internationaal gem.	49 (0.2)	520 (0.4)	51 (0.2)	501 (0.5)	19 (0.5)	19 ▲
Noorwegen (5)	50 (1.1)	570 (2.6)	50 (1.1)	548 (2.6)	21 (2.3)	21 ▲
Finland	50 (0.9)	577 (1.9)	50 (0.9)	555 (2.3)	22 (2.2)	22 ▲

4.4.2 Socio-economische status (aantal boeken thuis)

De socio-economische status van het gezin wordt hier beschreven aan de hand het aantal boeken thuis op basis van de leerlingenvragenlijst (afgenomen in zowel 2016 als 2018). Voor deze indicator wordt in [Tabel 33](#) een overzicht gegeven van hoeveel leerlingen thuis meer dan 200 boeken hebben, 101 tot 200 boeken hebben, 26 tot 100 boeken hebben, 11 tot 25 boeken hebben, of 0 tot 10 boeken hebben. Daarnaast wordt een overzicht gegeven van de gemiddelde prestaties voor begrijpend lezen naargelang het aantal boeken.

[Tabel 33](#) toont dat er in 2016 minder boeken thuis zijn in Vlaanderen dan in andere landen in West-Europa bij leerlingen in het vierde leerjaar. Daarnaast zien we een duidelijke samenhang tussen het aantal boeken thuis en de gemiddelde prestaties in begrijpend lezen. Zo hebben leerlingen met 101 tot 200 boeken thuis een gemiddelde score van 547, terwijl de leerlingen met 11 tot 25 boeken thuis een gemiddelde score hebben van 510.

Tabel 33: Percentages leerlingen naargelang het aantal boeken thuis (ASBG04), inclusief gemiddelde prestaties voor begrijpend lezen (ASRREA01-05)

Landen	meer dan 200 boeken		101 tot 200 boeken		26 tot 100 boeken		11 tot 25 boeken		0 tot 10 boeken	
	%	Gemiddelde score	%	Gemiddelde score	%	Gemiddelde score	%	Gemiddelde score	%	Gemiddelde score
Duitsland	17 (0.9)	572 (4.2)	18 (0.8)	569 (3.5)	34 (1.1)	553 (2.9)	23 (0.9)	515 (5.3)	8 (0.7)	477 (6.8)
Engeland	16 (0.8)	598 (3.2)	21 (0.8)	588 (2.8)	34 (0.8)	559 (2.5)	20 (0.8)	527 (2.4)	9 (0.6)	496 (4.5)
Wallonië	16 (1.0)	525 (3.8)	19 (0.8)	524 (3.3)	33 (1.0)	501 (2.8)	21 (0.9)	476 (3.7)	11 (0.7)	449 (4.8)
Zweden	16 (1.2)	586 (3.4)	20 (1.0)	575 (3.3)	37 (1.0)	557 (2.8)	19 (0.9)	534 (3.6)	7 (0.6)	492 (4.7)
Noorwegen	16 (1.0)	586 (3.7)	22 (0.8)	571 (3.4)	38 (1.1)	560 (2.9)	18 (0.7)	534 (2.8)	6 (0.5)	517 (6.0)
Denemarken	16 (1.0)	577 (3.9)	21 (0.8)	563 (3.2)	36 (1.1)	549 (2.6)	20 (0.9)	528 (3.1)	8 (0.6)	493 (5.6)
Ierland	15 (0.8)	601 (4.3)	20 (0.9)	598 (3.0)	34 (1.0)	569 (2.6)	21 (0.9)	540 (3.3)	9 (0.6)	498 (6.1)
Oostenrijk	14 (0.9)	576 (3.5)	16 (0.8)	566 (2.6)	36 (0.9)	546 (2.6)	25 (0.9)	517 (3.4)	8 (0.7)	486 (4.0)
Noord-Ierland	14 (0.8)	598 (4.7)	20 (0.8)	595 (3.5)	34 (0.9)	576 (2.4)	21 (0.7)	536 (3.6)	11 (0.7)	494 (4.6)
VLAANDEREN2018	14 (0.7)	610 (3.1)	19 (0.8)	603 (2.5)	38 (0.8)	587 (2.2)	21 (0.7)	567 (2.4)	9 (0.7)	551 (3.2)
Frankrijk	13 (0.8)	541 (5.7)	18 (0.8)	531 (3.3)	36 (0.9)	521 (2.5)	23 (0.9)	492 (2.7)	10 (0.6)	454 (3.3)
Finland	12 (0.7)	590 (3.7)	23 (0.8)	586 (2.5)	44 (1.0)	566 (2.1)	16 (0.7)	538 (3.7)	4 (0.4)	501 (6.2)
Internationaal gemiddelde	12 (0.1)	541 (0.8)	15 (0.1)	545 (0.7)	33 (0.1)	531 (0.5)	25 (0.1)	506 (0.5)	16 (0.1)	473 (0.7)
VLAANDEREN(126)	12 (0.7)	547 (3.4)	17 (0.7)	547 (2.5)	39 (0.8)	529 (1.8)	24 (0.8)	511 (3.0)	8 (0.5)	484 (3.7)
VLAANDEREN	12 (0.8)	549 (3.7)	16 (0.7)	547 (2.6)	38 (0.8)	530 (2.1)	25 (0.9)	510 (2.8)	10 (0.7)	484 (4.9)
Spanje	11 (0.4)	545 (3.7)	16 (0.5)	545 (2.9)	34 (0.7)	538 (2.1)	28 (0.6)	516 (1.8)	11 (0.5)	488 (4.1)
Nederlands	9 (0.7)	564 (4.2)	15 (0.9)	566 (3.2)	39 (1.0)	552 (2.4)	27 (0.9)	529 (2.3)	10 (0.8)	511 (3.6)
Italië	9 (0.6)	565 (4.4)	10 (0.6)	567 (4.7)	30 (0.9)	561 (2.2)	34 (0.9)	541 (2.6)	16 (0.8)	522 (3.9)
Portugal	8 (0.6)	547 (5.2)	13 (0.6)	552 (3.6)	35 (1.0)	540 (2.7)	31 (1.1)	515 (2.6)	13 (0.7)	491 (4.2)

De 126 scholen die deelnamen aan de herhalingsmeting tonen gelijke resultaten aan de originele steekproef in het vierde leerjaar in 2016. Ook hier wordt er aangegeven dat er minder boeken thuis zijn dan in de andere landen in West-Europa. Daarnaast zien we ook hier een duidelijke samenhang tussen het aantal boeken thuis en de gemiddelde prestaties in begrijpend lezen. Ter vergelijking in deze steekproef hebben leerlingen met 101 tot 200 boeken thuis een gemiddelde score van 547, terwijl de leerlingen met 11 tot 25 boeken thuis een gemiddelde score hebben van 511.

De herhalingsmeting van de 126 scholen in het zesde leerjaar in 2018 toont een gelijke verdeling van het aantal boeken thuis als in het vierde leerjaar. Alle categorieën hebben een hogere gemiddelde score in het zesde leerjaar. We vinden wederom een duidelijke samenhang tussen het aantal boeken thuis en de gemiddelde prestaties in begrijpend lezen. Het verschil in gemiddelde prestaties tussen de categorieën lijkt onveranderd te zijn. Nu hebben leerlingen met 101 tot 200 boeken thuis een gemiddelde score van 603, terwijl de leerlingen met 11 tot 25 boeken thuis een gemiddelde score hebben van 567.

4.4.3 De houding van leerlingen tegenover lezen

De houding van leerlingen tegenover lezen werd in PIRLS bevestigd met de schaal *Students Like Reading* (SLR) in de leerlingenvragenlijst. Deze schaal omvat acht stellingen waarbij leerlingen moesten aangeven in welke mate ze ermee akkoord gingen. Daarnaast dienden leerlingen aan te geven hoe vaak ze buiten de schooluren (1) lezen voor het plezier en (2) lezen om iets bij te leren. Op basis van een combinatie van antwoorden werden leerlingen ingedeeld in drie categorieën: zeer positieve, positieve en eerder negatieve houding (zie ook Tielemans et al., 2017). In **Tabel 34** wordt weergegeven hoeveel leerlingen tot deze drie categorieën behoren. Daarnaast wordt een overzicht gegeven van de gemiddelde prestaties voor begrijpend lezen naargelang de houding van de leerlingen tegenover lezen.

Tabel 34 toont dat Vlaanderen in 2016 één van de hoogste percentages leerlingen met een eerder negatieve houding ten aanzien van lezen heeft. De West-Europese tendens, hoe positiever de houding ten aanzien van lezen, hoe beter de prestaties, geldt ook voor Vlaanderen. Zo behalen Vlaamse leerlingen met een heel positieve houding een gemiddelde score van 536 terwijl leerlingen met eerder negatieve houding een gemiddelde score van 513 hebben.

De 126 scholen die deelnamen aan de herhalingsmeting tonen gelijke resultaten aan de originele steekproef van 148 scholen in het vierde leerjaar in 2016. Ook hier vinden we dat er in Vlaanderen meer leerlingen zijn met een eerder negatieve houding dan in de meeste andere West-Europese landen. Ook hier zien we een duidelijke samenhang tussen de houding tegenover lezen en de gemiddelde prestaties in begrijpend lezen. Zo behalen Vlaamse leerlingen met een heel positieve houding een gemiddelde score van 535 terwijl leerlingen met eerder negatieve houding een gemiddelde score van 512 hebben.

De herhalingsmeting van de 126 scholen in het zesde leerjaar in 2018 toont een sterk veranderde verdeling voor de houding van leerlingen tegenover lezen. Het aantal leerlingen dat een zeer positieve houding heeft tegenover lezen is gehalveerd, terwijl het aantal leerlingen dat een eerder negatieve houding heeft is toegenomen. Het aantal leerlingen met een positieve houding tegenover lezen is dus gedaald tussen het vierde en het zesde leerjaar. In vergelijking met het vierde leerjaar van de West-Europese landen en regio's heeft Vlaanderen in het zesde leerjaar de minste leerlingen met een heel positieve houding. Alle categorieën hebben wel een hogere gemiddelde score in het zesde leerjaar tegenover het vierde leerjaar. We vinden wederom een duidelijke samenhang tussen de houding tegenover lezen en de gemiddelde prestaties in begrijpend lezen. Het verschil in gemiddelde prestaties tussen de categorieën lijkt wel toegenomen. Nu hebben Vlaamse leerlingen met een zeer positieve houding een gemiddelde score van 613 terwijl leerlingen met eerder negatieve houding een gemiddelde score van 574 hebben.

Tabel 34: Percentages leerlingen met een zeer positieve, een positieve en een eerder negatieve houding tegenover lezen (ASDGSLR), inclusief gemiddelde prestaties voor begrijpend lezen (ASRREA01-05)

Landen	Zeer positieve houding		Positieve houding		Eerder negatieve houding	
	Gemiddelde score		Gemiddelde score		Gemiddelde score	
	%		%		%	
Portugal	72 (1.1)	530 (2.4)	23 (1.0)	526 (3.2)	5 (0.4)	513 (5.3)
Spanje	56 (0.8)	534 (1.5)	34 (0.7)	523 (3.5)	10 (0.5)	512 (3.7)
Ierland	46 (1.1)	580 (3.0)	40 (1.0)	565 (2.9)	15 (0.8)	534 (4.6)
Internationaal gemiddelde	43 (0.2)	523 (0.5)	41 (0.1)	507 (0.5)	16 (0.1)	486 (1.0)
Italië	41 (1.1)	554 (2.5)	44 (1.0)	546 (2.7)	16 (0.8)	539 (3.7)
Frankrijk	40 (1.0)	519 (3.0)	46 (1.0)	510 (3.0)	14 (0.9)	497 (3.4)
Noord-Ierland	39 (1.3)	580 (2.9)	42 (1.0)	567 (2.7)	19 (0.9)	531 (3.8)
Oostenrijk	37 (1.1)	550 (3.1)	45 (0.8)	541 (2.6)	18 (0.8)	524 (3.4)
England	35 (1.0)	575 (2.5)	45 (0.9)	559 (2.2)	20 (0.9)	530 (3.3)
Wallonië (FG)	33 (1.1)	504 (3.4)	44 (0.9)	500 (2.8)	23 (1.2)	484 (3.3)
Duitsland	32 (1.3)	563 (2.9)	43 (1.0)	548 (2.7)	25 (1.2)	510 (5.0)
Finland	28 (0.9)	584 (2.3)	49 (0.8)	568 (2.3)	23 (0.7)	540 (2.5)
VLAANDEREN(126)	25 (1.0)	535 (2.7)	46 (0.8)	530 (1.7)	29 (1.2)	512 (2.4)
VLAANDEREN	24 (1.0)	536 (3.0)	45 (0.8)	529 (2.0)	31 (1.1)	513 (2.6)
Nederland	24 (0.9)	560 (2.4)	46 (0.9)	550 (2.0)	31 (1.1)	527 (2.5)
Noorwegen (5)	22 (0.9)	575 (3.2)	51 (1.0)	565 (2.4)	27 (1.2)	536 (2.8)
Denemarken	20 (0.8)	569 (2.9)	53 (1.0)	551 (2.4)	27 (1.2)	528 (3.2)
Zweden	18 (1.0)	572 (4.3)	50 (1.0)	563 (2.5)	31 (1.1)	535 (3.2)
VLAANDEREN2018	12 (0.6)	613 (3.4)	41 (1.1)	592 (1.9)	46 (1.3)	574 (2.0)

4.4.4 De zelfzekerheid van leerlingen over hun leesvaardigheid

De zelfzekerheid van leerlingen over hun vaardigheid voor lezen werd in PIRLS bevraagd met de schaal *Students Confident in Reading* (SCR). Leerlingen moesten daarbij voor zes stellingen aangeven in welke mate ze ermee akkoord gingen. De combinatie van antwoorden bepaalde in welke van volgende drie categorieën de leerling terecht kwam: zeer zelfzeker, enigszins zelfzeker en niet zelfzeker voor wat betreft hun leesvaardigheid. In **Tabel 35** wordt weergegeven hoeveel leerlingen tot deze drie categorieën behoren. Daarnaast wordt een overzicht gegeven van de gemiddelde prestaties voor begrijpend lezen naargelang de zelfzekerheid van leerlingen over hun leesvaardigheid.

Tabel 35 toont dat in 2016 bijna de helft van de Vlaamse leerlingen in het vierde leerjaar zeer zelfzeker is over zijn of haar leesvaardigheid. Daarnaast is één derde van de leerlingen enigszins zelfzeker en een niet te verwaarlozen percentage leerlingen, ongeveer één vijfde, geeft aan niet zelfzeker te zijn over zijn leesvaardigheid. De West-Europese tendens, hoe meer zelfzeker ten aanzien van leesvaardigheid, hoe beter de prestaties, geldt ook voor Vlaanderen. Zo behalen Vlaamse leerlingen die zeer zelfzeker zijn een gemiddelde score van 544 terwijl leerlingen die niet zelfzeker zijn een gemiddelde score van 493 hebben.

De 126 scholen die deelnamen aan de herhalingsmeting tonen gelijke resultaten aan de originele steekproef in het vierde leerjaar in 2016. Ook hier vinden we dat in Vlaanderen bijna de helft van de leerlingen zeer zelfzeker is hun leesvaardigheid. Ook met deze steekproef geeft ongeveer één vijfde van de leerlingen aan niet zelfzeker te zijn over hun leesvaardigheid. Daarnaast zien we ook hier een duidelijke samenhang tussen zelfzekerheid in leesvaardigheid en de gemiddelde prestaties in begrijpend lezen. Zo behalen Vlaamse leerlingen die zeer zelfzeker zijn een gemiddelde score van 546 terwijl leerlingen die niet zelfzeker zijn een gemiddelde score van 493 hebben.

De herhalingsmeting van de 126 scholen in het zesde leerjaar in 2018 toont een eerder gelijke verdeling in zelfzekerheid als in het vierde leerjaar. Het percentage leerlingen dat niet zelfzeker is in lezen is nu wel enigszins gedaald. Alle categorieën hebben een hogere gemiddelde score in het zesde leerjaar tegenover het vierde leerjaar. We vinden wederom een duidelijke samenhang tussen zelfzekerheid in lezen en de gemiddelde prestaties in begrijpend lezen. Het verschil in gemiddelde prestaties tussen de categorieën is ongeveer gelijk gebleven. Zo behalen Vlaamse leerlingen die zeer zelfzeker zijn een gemiddelde score van 603 terwijl leerlingen die niet zelfzeker zijn een gemiddelde score van 554.

Tabel 35: Percentages leerlingen die zeer, enigszins en niet zelfzeker zijn over hun leesvaardigheid (ASDGSCR), inclusief gemiddelde prestaties voor begrijpend lezen (ASRREA01-05)

Landen	Zeer zelfzeker		Enigszins zelfzeker		Niet zelfzeker	
	%	Gemiddelde score	%	Gemiddelde score	%	Gemiddelde score
Zweden	65 (0.8)	575 (2.3)	28 (0.8)	532 (3.2)	8 (0.5)	488 (5.2)
Finland	60 (0.9)	586 (1.9)	31 (0.9)	549 (2.2)	10 (0.5)	503 (4.5)
Oostenrijk	56 (1.0)	562 (2.1)	29 (0.9)	526 (3.1)	14 (0.7)	493 (3.5)
Ierland	55 (1.0)	593 (2.6)	31 (0.8)	550 (2.7)	14 (0.8)	505 (4.2)
Duitsland	55 (1.0)	569 (2.4)	30 (0.9)	529 (3.6)	15 (0.7)	487 (4.7)
Engeland	53 (0.9)	591 (1.9)	31 (0.8)	541 (2.6)	16 (0.5)	488 (3.1)
Noorwegen (5)	53 (0.9)	584 (2.2)	33 (0.8)	545 (2.7)	14 (0.8)	505 (3.6)
Denemarken	51 (0.8)	577 (2.3)	36 (0.8)	532 (2.7)	14 (0.6)	486 (4.4)
Noord-Ierland	50 (1.1)	598 (2.2)	33 (1.0)	553 (3.0)	17 (0.7)	493 (4.1)
Nederland	49 (1.1)	565 (1.9)	30 (0.8)	536 (2.2)	21 (0.9)	511 (2.8)
VLAANDEREN2018	48 (1.2)	603 (1.8)	38 (1.0)	577 (2.3)	14 (0.6)	554 (3.0)
VLAANDEREN (126)	46 (0.9)	546 (2.2)	34 (0.9)	519 (2.2)	19 (0.7)	493 (2.3)
VLAANDEREN	46 (1.0)	544 (2.3)	34 (0.8)	519 (2.5)	19 (0.7)	493 (2.6)
Italië	46 (1.0)	567 (2.2)	38 (1.0)	545 (2.7)	16 (0.5)	505 (4.3)
Internationaal gemiddelde	45 (0.2)	545 (0.4)	35 (0.1)	503 (0.5)	21 (0.1)	455 (0.6)
Frankrijk	44 (1.1)	539 (2.6)	39 (1.1)	506 (2.6)	17 (0.7)	455 (3.7)
Wallonië (FG)	41 (0.9)	528 (2.8)	37 (0.8)	493 (2.8)	22 (0.7)	450 (3.4)
Spanje	40 (0.7)	554 (1.8)	40 (0.6)	525 (2.1)	20 (0.6)	483 (2.7)
Portugal	38 (1.3)	555 (2.7)	40 (1.0)	526 (2.7)	22 (0.9)	483 (3.0)

4.5 Eigenschappen van de school en leerkracht

In wat volgt, worden vijf eigenschappen van de school en leerkracht onderzocht. Het gaat om (1) de gerichtheid op academisch succes volgens leerkrachten, (2) de tevredenheid van leerkrachten, (3) onderwijservaring van leerkrachten, (4) nascholing begrijpend lezen van leerkrachten, en (5) gebruikte werkvormen.

4.5.1 Gerichtheid op academisch succes

De gerichtheid op academisch succes werd in PIRLS gemeten met de schaal *School Emphasis on Academic Success* (SEAS). De leerkrachten kregen een aantal stellingen omtrent het streven van leerlingen, ouders, leerkrachten en directie naar goede prestaties. De leerkrachten (zowel van het 4e als 6e leerjaar in respectievelijk 2016 en 2018) duiden aan in welke mate de stellingen typerend zijn voor de eigen school, gaande van zeer hoog tot zeer laag. Leerlingen werden uiteindelijk, op basis van een combinatie van antwoorden van hun leerkracht, ingedeeld in een van volgende categorieën van gerichtheid: zeer grote nadruk, grote nadruk en matige nadruk op academisch succes. In [Tabel 36](#) wordt weergegeven hoeveel leerlingen tot deze drie categorieën behoren.

[Tabel 36](#) toont dat in 2016 meer dan de helft van de leerlingen een leerkracht heeft die aangeeft dat er een hoge nadruk is op academisch succes. Slechts één percent van de leerlingen heeft een leerkracht die aangeeft dat er een zeer hoge nadruk is op academisch succes. Dit is opvallend laag binnen West-Europa. Uit de internationale en West-Europese vergelijking blijkt dat Vlaanderen relatief minder nadruk legt op het belang van goede prestaties. Landen met een zeer grote tot grote nadruk op academisch succes zijn onder andere Engeland, Noord-Ierland en Ierland. De West-Europese tendens, hoe meer gerichtheid op academisch succes, hoe beter de prestaties, geldt ook voor Vlaanderen. Zo behalen Vlaamse leerlingen met een leerkracht die aangeeft dat er een hoge nadruk is op academisch succes een gemiddelde score van 532. Leerlingen met een leerkracht die aangeeft dat er een matige nadruk is op academisch succes behalen een lagere gemiddelde score van 517.

De 126 scholen die deelnamen aan de herhalingsmeting tonen gelijke resultaten aan de originele steekproef in het vierde leerjaar in 2016. Hier vinden we dat in Vlaanderen slechts twee percent van de leerlingen een leerkracht heeft die aangeeft dat er een zeer hoge nadruk is op academisch succes. Ook in deze steekproef heeft het merendeel van de leerlingen een leerkracht die aangeeft dat er een hoge nadruk is op academisch succes.

Daarnaast zien we ook hier dezelfde duidelijke samenhang tussen gerichtheid op academisch succes en de gemiddelde prestaties in begrijpend lezen. Zo behalen Vlaamse leerlingen met een leerkracht die aangeeft dat er een hoge nadruk is op academisch succes een gemiddelde score van 532. Leerlingen met een leerkracht die aangeeft dat er een matige nadruk is op academisch succes behalen een lagere gemiddelde score van 518.

De herhalingsmeting van de 126 scholen in het zesde leerjaar in 2018 toont een sterk veranderde verdeling in van wat de leerkrachten aangeven over de gerichtheid op academisch succes. In het zesde leerjaar hebben twee derde van de leerlingen een leerkracht die aangeeft dat er een matige nadruk is op academisch succes. Eén derde van de leerlingen hebben een leerkracht die aangeeft dat er een hoge nadruk is op academisch succes. Ook in het zesde jaar heeft slechts één percent van de leerlingen een leerkracht heeft die aanduidt dat er een zeer hoge nadruk is op academisch succes. Alle categorieën hebben wel een hogere gemiddelde score in het zesde leerjaar tegenover het vierde leerjaar. We vinden wederom een duidelijke samenhang tussen gerichtheid op academisch succes en de gemiddelde prestaties in begrijpend lezen. Zo behalen Vlaamse leerlingen met een leerkracht die aangeeft dat er een hoge nadruk is op academisch succes een gemiddelde score van 595. Leerlingen met een leerkracht die aangeeft dat er een matige nadruk is op academisch succes behalen een lagere gemiddelde score van 581.

Tabel 36: Percentages leerlingen van wie de leerkracht een zeer grote, grote en matige nadruk op academisch succes rapporteert (ATDGEAS), inclusief gemiddelde prestaties voor begrijpend lezen (ASRREA01-05)

Landen	Zeer hoge nadruk		Hoge nadruk		Matige nadruk	
	%	Gemiddelde score	%	Gemiddelde score	%	Gemiddelde score
Engeland	19 (2.6)	568 (4.4)	62 (3.7)	559 (2.6)	18 (3.1)	548 (4.7)
Noord-Ierland	19 (3.1)	579 (5.6)	68 (3.9)	561 (3.0)	14 (3.0)	556 (8.2)
Ierland	16 (2.7)	579 (6.7)	67 (3.5)	571 (2.9)	17 (2.6)	539 (6.1)
Spanje	10 (1.8)	544 (4.1)	62 (3.6)	531 (2.3)	28 (3.3)	516 (3.5)
Oostenrijk	9 (2.1)	563 (4.9)	63 (3.5)	546 (2.4)	28 (3.1)	523 (4.8)
Internationaal gemiddelde	8 (0.3)	522 (1.9)	55 (0.5)	518 (0.6)	37 (0.4)	497 (0.9)
Portugal	5 (1.5)	547 (5.4)	53 (3.7)	534 (3.4)	41 (3.4)	518 (3.0)
Frankrijk	4 (1.6)	537 (10.3)	51 (3.4)	518 (2.9)	44 (3.2)	502 (4.0)
Denemarken	4 (1.9)	564 (11.4)	54 (3.6)	553 (3.1)	42 (3.5)	540 (3.0)
Italië	3 (1.4)	556 (12.6)	55 (3.7)	551 (3.3)	42 (3.4)	546 (3.1)
Finland	2 (0.8)	~ ~	66 (2.8)	570 (2.0)	32 (2.9)	558 (3.7)
Duitsland	2 (1.0)	~ ~	54 (3.5)	555 (2.7)	44 (3.3)	512 (6.6)
Nederland	2 (1.2)	~ ~	50 (3.5)	550 (2.4)	48 (3.6)	539 (2.6)
VLAANDEREN(126)	2 (1.0)	~ ~	54 (3.7)	532 (2.2)	44 (3.8)	518 (2.9)
VLAANDEREN2018	1 (0.7)	~ ~	35 (3.6)	595 (2.6)	64 (3.6)	581 (2.1)
VLAANDEREN	1 (0.6)	~ ~	53 (3.9)	532 (2.3)	46 (3.9)	517 (3.1)
Wallonië (FG)	1 (0.6)	~ ~	49 (3.8)	511 (2.9)	51 (3.7)	485 (3.8)
Norwegen (5)	1 (0.6)	~ ~	58 (3.6)	563 (2.7)	41 (3.7)	551 (3.4)

4.5.2 Tevredenheid van leerkrachten

De tevredenheid van leerkrachten werd in PIRLS 2016 en de herhalingsmeting in 2018 onderzocht met de *Teacher Job Satisfaction Scale* (TJS). De leerkrachten kregen een aantal stellingen omtrent de tevredenheid van leerkracht over zijn of haar job. Deze werd in de leerkrachtvragenlijst bevestigd aan de hand van stellingen zoals 'Ik vind mijn werk belangrijk en zinvol'. De leerkrachten duiden aan in welke mate de stellingen typerend zijn voor zichzelf, gaande van zeer hoog tot zeer laag. Leerlingen werden uiteindelijk, op basis van een combinatie van antwoorden van hun leerkracht, ingedeeld in een van volgende categorieën van tevredenheid van de leerkracht: zeer tevreden, redelijk tevreden en niet tevreden. In [Tabel 37](#) wordt weergegeven hoeveel leerlingen tot deze drie categorieën behoren. Daarnaast wordt een overzicht gegeven van de gemiddelde prestaties voor begrijpend lezen naargelang de tevredenheid van leerkrachten over hun job.

Tabel 37 toont dat meer dan 90% van de Vlaamse leerkrachten aangeeft hier *heel* tot *redelijk tevreden* over te zijn, terwijl slechts 3% *niet tevreden* is. In vergelijking met andere West-Europese landen neemt Vlaanderen qua leerkrachttevredenheid een eerder gunstige positie in. Het verband tussen de jobtevredenheid van de leerkracht en de prestaties voor begrijpend lezen van diens leerlingen, vertoont eerder atypische resultaten (zowel in Vlaanderen als internationaal). In beide gevallen geldt immers dat leerlingen die onderwezen worden door een leerkracht die *redelijk tevreden* is, de laagste gemiddelde score behalen, terwijl leerlingen met *niet tevreden* leerkrachten de hoogste gemiddelde score hebben. Voor Vlaanderen is het evenwel zo dat de prestaties van leerlingen van heel, redelijk en niet tevreden leerkrachten telkens niet significant verschillen van het algemeen Vlaams gemiddelde. Anders gezegd: in Vlaanderen is er nauwelijks een verband tussen de jobtevredenheid van de leerkracht van het vierde leerjaar en de prestaties voor begrijpend lezen van de leerlingen. Ter informatie, Vlaamse leerlingen met een leerkracht die aangeeft zeer tevreden te zijn hebben een gemiddelde score van 526. Leerlingen met een leerkracht die aangeeft redelijk tevreden te zijn behalen een gelijkaardige gemiddelde score van 524.

De 126 scholen die deelnamen aan de herhalingsmeting tonen ongeveer gelijke resultaten aan de originele steekproef in het vierde leerjaar in 2016. Ook hier vinden we dat in Vlaanderen meer dan 90% van de leerlingen een leerkracht heeft die aanduidt dat zeer of redelijk tevreden te zijn over zijn of haar job, terwijl slechts 2% niet tevreden is. Het gebrek aan verband tussen de tevredenheid van leerkrachten en de gemiddelde prestaties van leerlingen vinden we ook hier terug. Ter informatie, Vlaamse leerlingen met een leerkracht die aangeeft zeer tevreden te zijn hebben een gemiddelde score van 527. Leerlingen met een leerkracht die aangeeft redelijk tevreden te zijn behalen een gelijkaardige gemiddelde score van 524.

De herhalingsmeting van de 126 scholen in het zesde leerjaar in 2018 toont een eerder gelijke verdeling in tevredenheid van leerkrachten als in het vierde leerjaar. Er is een beperkte daling in het aantal leerkrachten dat zeer tevreden is en een kleine stijging in het aantal leerkrachten dat redelijk tevreden is. Alle categorieën hebben een hogere gemiddelde score in het zesde leerjaar tegenover het vierde leerjaar. We vinden wederom geen verband tussen de tevredenheid van leerkrachten en de gemiddelde prestaties van leerlingen. Ter informatie, Vlaamse leerlingen met een leerkracht 6e leerjaar die aangeeft zeer tevreden te zijn hebben een gemiddelde score van 589. Leerlingen met een leerkracht die aangeeft redelijk tevreden te zijn behalen een gelijkaardige gemiddelde score van 583.

Tabel 37: Percentages leerlingen met een leerkracht die zeer, redelijk en niet tevreden is over zijn/haar job (ATDGTJS), inclusief gemiddelde prestaties voor begrijpend lezen (ASRREA01-05)

Landen	Zeer tevreden		Redelijk tevreden		Niet tevreden	
	%	Gemiddelde score	%	Gemiddelde score	%	Gemiddelde score
Spanje	73 (3.0)	529 (2.4)	25 (3.1)	524 (3.4)	2 (0.9)	~ ~
Noord-Ierland	62 (4.7)	564 (3.4)	31 (4.2)	567 (4.0)	7 (2.5)	548 (9.0)
Ierland	60 (3.6)	570 (3.8)	36 (3.5)	561 (3.9)	4 (1.3)	561 (8.9)
Oostenrijk	59 (3.9)	542 (2.8)	40 (3.8)	540 (3.7)	1 (0.6)	~ ~
Italië	58 (3.4)	549 (3.1)	38 (3.4)	547 (3.8)	3 (1.6)	559 (11.2)
Internationaal gemiddelde	57 (0.5)	513 (0.6)	37 (0.5)	508 (0.9)	6 (0.2)	525 (2.3)
VLAANDEREN(126)	55 (3.1)	527 (2.3)	43 (3.1)	524 (2.7)	2 (1.0)	527 (13.2)
VLAANDEREN	53 (3.5)	526 (2.9)	44 (3.3)	524 (2.8)	3 (1.2)	532 (13.9)
Engeland	51 (3.7)	558 (3.4)	42 (3.8)	559 (2.8)	7 (2.0)	563 (7.1)
Wallonië (FG)	51 (3.3)	503 (3.5)	40 (3.4)	494 (4.3)	9 (2.3)	484 (10.2)
Portugal	49 (3.7)	531 (2.8)	41 (3.8)	526 (4.9)	10 (2.1)	523 (4.2)
VLAANDEREN2018	46 (3.5)	589 (2.5)	50 (3.7)	583 (2.5)	4 (1.5)	585 (9.0)
Nederland	44 (4.0)	542 (3.3)	53 (4.0)	549 (2.2)	3 (1.1)	528 (11.9)
Denemarken	43 (3.6)	548 (3.5)	45 (3.9)	546 (3.3)	11 (2.5)	551 (5.1)
Noorwegen (5)	42 (4.1)	560 (3.3)	53 (3.9)	560 (2.6)	4 (1.2)	544 (19.4)
Zweden	41 (4.3)	554 (4.0)	52 (4.6)	557 (3.3)	6 (2.1)	549 (9.4)
Finland	41 (3.5)	565 (2.6)	49 (3.2)	567 (2.6)	11 (1.8)	568 (4.4)
Duitsland	38 (3.6)	544 (4.8)	53 (4.0)	533 (5.4)	10 (2.3)	524 (10.0)
Frankrijk	26 (2.8)	513 (4.4)	62 (3.3)	511 (2.8)	12 (2.2)	506 (6.6)

4.5.3 Onderwijservaring van leerkrachten

De leerkrachten werd ook gevraagd in de leerkrachtvragenlijst hoeveel jaren onderwijservaring ze hadden. Er worden vier categorieën van ervaring onderscheiden: 20 jaar of meer, 10 tot 20 jaar, 5 tot 10 jaar, of 5 jaar of minder. In **Tabel 38** wordt per categorie weergegeven hoeveel leerlingen een leerkracht hebben met deze onderwijservaring. Daarnaast wordt een overzicht gegeven van de gemiddelde prestaties voor begrijpend lezen volgens elk van de categorieën van onderwijservaring. Er wordt ook weergegeven wat het gemiddelde jaren onderwijservaring is van leerkrachten.

Tabel 38 toont dat bij de leerlingen van het vierde leerjaar maar liefst 37% van de Vlaamse leerkrachten 10 tot 20 jaar ervaring en heeft nog eens 37% 20 jaar of meer ervaring. Vlaanderen telt daarnaast 17% beginnende leerkrachten (0 tot 5 jaar ervaring) en 10% leerkrachten met 5 tot 10 jaar ervaring. Binnen West-Europa zijn er heel wat landen (zoals Portugal, Italië, en Wallonië (FG)), waar leerkrachten met een beperkte onderwijservaring eerder uitzonderlijk zijn. Vanuit internationaal perspectief blijkt er een positief verband te zijn tussen het aantal jaren ervaring en de gemiddelde prestaties voor begrijpend lezen. Dit is ook het geval in Vlaanderen.

De 126 scholen die deelnamen aan de herhalingsmeting tonen gelijke resultaten aan de originele steekproef in het vierde leerjaar in 2016. Hier vinden we dat in Vlaanderen 38% van de Vlaamse leerkrachten 10 tot 20 jaar ervaring en heeft nog eens 37% 20 jaar of meer ervaring. Er zijn in deze steekproef 16% beginnende leerkracht en 9% leerkrachten met 5 tot 10 jaar onderwijservaring. Ook hier is er een verband tussen aantal jaren onderwijservaring en de gemiddelde prestaties van leerlingen in begrijpend lezen.

De herhalingsmeting van de 126 scholen in het zesde leerjaar in 2018 toont een eerder gelijke verdeling in de onderwijservaring van de leerkrachten. Ook hier vinden we dat in Vlaanderen het merendeel van de leerkrachten zesde leerjaar ofwel 20 jaar of meer onderwijservaring hebben, of tussen 10 en 20 jaar onderwijservaring hebben. Leerkrachten met 5 tot 10 jaar of minder dan 5 jaar onderwijservaring zijn hier in de minderheid. Binnen alle categorieën is er vooruitgang in de prestaties voor begrijpend lezen. Het is echter opvallend dat er tussen de verschillende categorieën van onderwijservaring nu nauwelijks verschillen zijn in gemiddelde prestaties van leerlingen voor begrijpend lezen.

4.5.4 Nascholing begrijpend lezen van leerkrachten

De leerkrachten werd ook gevraagd in de leerkrachtvragenlijst (zowel in 2016 als in 2018) hoeveel uren nascholing m.b.t. begrijpend lezen werden gevolgd. Er worden vier categorieën van uren nascholing onderscheiden: 16 uur of meer, 6 tot 15 uur, minder dan 6 uur en geen nascholing m.b.t. begrijpend lezen. In **Tabel 39** wordt per categorie weergegeven hoeveel leerlingen een leerkracht hebben die een bepaalde hoeveelheid aan nascholing m.b.t. begrijpend lezen hebben gevolgd.

Tabel 39 toont dat in Vlaanderen leerkrachten in het vierde leerjaar weinig deelnamen aan professionele vormingsactiviteiten. Zo geeft 35% van de Vlaamse leerkrachten aan de voorbije twee jaar geen bijscholing gevolgd te hebben inzake begrijpend lezen. Van de Vlaamse leerkrachten 4e leerjaar die wel nascholing inzake lezen volgden, geeft 36% aan minder dan 6 uur hieraan besteed te hebben. Ongeveer 20% investeerde 6 tot 15 uur van zijn tijd in zo'n nascholingen en slechts 6% volgde 16 uur of meer. West-Europese landen die daarentegen meer inzetten op nascholingen voor begrijpend lezen zijn Spanje en Zweden. In Vlaanderen kan bovendien een negatief verband gevonden worden tussen het aantal uren nascholing die de leerkracht volgde en de prestaties voor begrijpend lezen. Met andere woorden: leerlingen die in een klas zitten bij een leerkracht die meer nascholingen volgde, presteren slechter voor begrijpend lezen dan leerlingen van een leerkracht die minder of geen nascholing volgde. Een mogelijke verklaring kan zijn dat leerkrachten die ervoor kiezen een nascholing voor begrijpend lezen te volgen, zich wellicht meer onzeker voelen over het onderwijzen van begrijpend lezen of dat leerkrachten van minder goed presterende leerlingen gestimuleerd worden of zelfs verplicht worden nascholing te volgen.

De 126 scholen die deelnamen aan de herhalingsmeting tonen gelijke resultaten aan de originele steekproef in het vierde leerjaar in 2016. Ook hier geeft 37% van de Vlaamse leerkrachten aan de voorbije twee jaar geen bijscholing gevolgd te hebben inzake begrijpend lezen. Van de Vlaamse leerkrachten die wel nascholing inzake lezen volgden, geeft 41% aan minder dan 6 uur hieraan besteed te hebben. 18% investeerde 6 tot 15 uur van zijn tijd in zo'n nascholingen en slechts 5% volgde 16 uur of meer. Ook hier zien we dat hoe meer nascholing, hoe lager de gemiddelde prestaties van leerlingen in begrijpend lezen.

De herhalingsmeting van de 126 scholen in het zesde leerjaar in 2018 toont een eerder gelijke verdeling in het aantal uren professionele ontwikkeling m.b.t. begrijpend lezen bij leerkrachten. Nu geeft 33% van de Vlaamse leerkrachten aan de voorbije twee jaar geen bijscholing gevolgd te hebben inzake begrijpend lezen. Dit is iets minder dan in het vierde leerjaar.

Tabel 38: Percentages leerlingen volgens het aantal jaren onderwijservaring van de leerkracht (ATDG01 & ATBG01), inclusief gemiddelde prestaties voor begrijpend lezen (ASRREA01-05)

Landen	20 jaar of meer		10-20 jaar		5-10 jaar		5 jaar of minder		Gemiddeld aantal jaar ervaring
	%	Gemiddelde score	%	Gemiddelde score	%	Gemiddelde score	%	Gemiddelde score	
Italië	71 (3.0)	552 (2.7)	22 (2.9)	543 (5.4)	3 (0.8)	530 (12.9)	4 (1.7)	524 (13.8)	25 (0.8)
Oostenrijk	59 (3.7)	541 (3.1)	19 (2.9)	551 (5.3)	12 (2.2)	528 (6.8)	10 (2.0)	542 (6.8)	22 (0.9)
Wallonië (FG)	47 (3.8)	503 (3.5)	31 (3.5)	494 (5.2)	15 (2.7)	494 (8.0)	6 (1.6)	479 (8.8)	19 (0.8)
Spanje	47 (3.2)	527 (2.3)	26 (2.5)	531 (3.2)	18 (2.5)	526 (7.5)	8 (1.6)	525 (4.4)	19 (0.8)
Portugal	46 (2.6)	529 (2.5)	48 (3.3)	527 (4.2)	4 (2.2)	538 (8.4)	2 (0.9)	~	21 (0.4)
Noord-Ierland	45 (4.2)	558 (3.3)	29 (4.2)	570 (5.2)	15 (3.6)	577 (6.0)	11 (2.6)	554 (8.6)	18 (0.9)
Internationaal gemiddelde	42 (0.5)	513 (0.9)	30 (0.5)	511 (0.9)	15 (0.4)	510 (1.3)	13 (0.3)	505 (1.6)	17 (0.1)
Nederland	39 (3.9)	548 (2.7)	29 (3.2)	545 (3.6)	15 (3.0)	545 (4.7)	17 (2.9)	540 (7.3)	17 (0.9)
VLAANDEREN2018	39 (3.6)	587 (2.6)	35 (3.4)	586 (2.7)	13 (2.1)	583 (5.1)	13 (2.6)	586 (4.9)	18 (0.8)
Duitsland	38 (3.8)	536 (6.7)	32 (3.7)	544 (4.0)	17 (2.8)	527 (8.7)	13 (2.5)	544 (8.0)	18 (0.9)
Finland	38 (3.1)	567 (3.2)	25 (2.5)	565 (4.9)	18 (2.5)	563 (3.6)	18 (2.8)	571 (3.4)	16 (0.8)
VLAANDEREN(126)	37 (3.2)	530 (2.6)	38 (3.4)	527 (3.0)	9 (1.9)	522 (6.4)	16 (2.4)	516 (4.7)	17 (0.7)
VLAANDEREN	37 (3.5)	529 (3.2)	37 (3.6)	529 (2.8)	10 (1.9)	516 (7.8)	17 (2.7)	514 (5.0)	17 (0.7)
Denemarken	31 (3.8)	554 (3.6)	38 (4.3)	543 (3.7)	21 (3.1)	547 (4.3)	10 (2.2)	546 (6.6)	16 (0.8)
Frankrijk	30 (3.6)	521 (3.6)	41 (3.7)	510 (3.3)	18 (2.6)	509 (8.1)	12 (2.2)	495 (7.0)	16 (0.7)
Noorwegen (5)	28 (3.5)	562 (4.0)	37 (4.1)	559 (2.8)	19 (3.3)	559 (4.5)	16 (2.9)	554 (5.2)	15 (0.8)
Ierland	20 (3.3)	567 (5.0)	29 (3.4)	569 (4.0)	34 (4.1)	567 (4.8)	17 (2.6)	563 (5.5)	13 (0.9)
Engeland	20 (3.0)	563 (4.7)	26 (3.3)	554 (4.6)	18 (2.8)	563 (4.3)	35 (3.3)	557 (4.1)	11 (0.7)
Zweden	19 (3.2)	554 (4.9)	43 (4.7)	559 (3.1)	17 (3.8)	563 (6.4)	21 (3.5)	544 (4.3)	13 (0.7)

Tabel 39: Percentages leerlingen met een leerkracht die geen, minder dan 6 uur, tussen 6 en 15 uur en meer dan 16 uur nascholingen m.b.t. begrijpend lezen volgde (ATBG06), inclusief gemiddelde prestaties voor begrijpend lezen (ASRREA01-05)

Landen	16 uur of meer		6-15 uur		Minder dan 6 uur		Geen	
	%	Gemiddelde score	%	Gemiddelde score	%	Gemiddelde score	%	Gemiddelde score
Spanje	48 (2.9)	527 (2.1)	21 (2.0)	526 (5.8)	15 (2.4)	528 (4.4)	17 (2.3)	535 (3.3)
Zweden	37 (3.9)	555 (3.8)	25 (3.6)	556 (5.3)	24 (3.8)	557 (4.6)	15 (2.9)	554 (5.7)
Internationaal gemiddelde	36 (0.5)	510 (0.9)	27 (0.5)	512 (1.0)	22 (0.4)	513 (1.1)	16 (0.4)	514 (1.5)
Ierland	34 (3.3)	564 (5.9)	22 (3.3)	567 (4.8)	35 (3.6)	567 (4.2)	10 (2.2)	572 (6.2)
Portugal	33 (2.9)	531 (4.5)	20 (2.4)	529 (4.8)	20 (2.8)	529 (5.6)	28 (2.8)	523 (3.3)
Noorwegen (5)	32 (3.5)	561 (3.6)	32 (3.5)	559 (4.0)	18 (3.1)	560 (4.7)	18 (2.7)	553 (4.0)
Nederland	28 (3.9)	539 (4.1)	30 (4.0)	546 (3.8)	33 (4.4)	546 (4.0)	10 (2.0)	560 (4.1)
Oostenrijk	28 (3.2)	542 (4.3)	49 (3.4)	539 (3.5)	16 (2.4)	547 (5.3)	7 (1.9)	544 (5.2)
Noord-Ierland	25 (4.4)	566 (5.6)	31 (4.8)	559 (5.1)	34 (4.9)	565 (4.2)	10 (2.7)	568 (8.5)
Italië	25 (3.7)	549 (4.1)	26 (3.2)	549 (4.7)	23 (3.3)	551 (4.9)	26 (3.4)	546 (5.4)
Wallonië (FG)	8 (1.8)	496 (7.3)	24 (3.4)	505 (6.0)	30 (3.3)	491 (4.8)	38 (3.7)	499 (4.2)
VLAANDEREN	6 (1.5)	516 (9.7)	20 (3.2)	527 (4.8)	39 (3.5)	522 (3.5)	35 (3.6)	530 (3.4)
VLAANDEREN(126)	5 (1.4)	514 (10.8)	18 (3.0)	526 (5.1)	41 (3.5)	524 (3.1)	37 (3.2)	529 (3.1)
Frankrijk	6 (1.4)	502 (12.0)	18 (3.1)	515 (4.2)	38 (3.8)	509 (4.4)	38 (4.3)	514 (3.1)
Duitsland	5 (1.8)	563 (9.4)	26 (3.3)	530 (9.4)	41 (3.8)	539 (4.3)	27 (3.4)	539 (5.4)
Finland	5 (1.1)	549 (5.5)	12 (2.2)	563 (5.3)	24 (3.0)	562 (4.3)	59 (3.1)	569 (2.1)
VLAANDEREN2018	4 (1.6)	590 (5.3)	16 (2.6)	585 (5.3)	47 (3.7)	587 (2.5)	33 (3.3)	585 (2.9)

Van de Vlaamse leerkrachten die wel nascholing inzake lezen volgden, geeft 47% aan minder dan 6 uur hieraan besteed te hebben. Dit is gelijkaardig aan 2016. 16% investeerde 6 tot 15 uur van zijn tijd in zo'n nascholingen en slechts 46% volgde 16 uur of meer. Ook deze cijfers zijn gelijkaardig aan het vierde leerjaar. Opvallend is dat er in het zesde leerjaar geen duidelijke relatie meer is tussen het aantal uren bijscholing en de gemiddelde prestaties van leerlingen voor begrijpend lezen.

4.5.5 Gebruikte werkvormen

Tabel 40 omvat een overzicht van de mate waarin de werkvormen (1) frontaal lesgeven, (2) homogene groepen, (3) heterogene groepen en (4) zelfstandig werk ingezet worden tijdens de leeslessen in het vierde en zesde leerjaar. Omdat de percentages voor werkvormen vaak zeer klein zijn, kozen we om hier geen gemiddelde prestaties weer te geven. Uit de tabel blijkt dat de voorkeur van Vlaamse leerkrachten in 2016 uitgaat naar zelfstandig werk, gevolgd door frontaal lesgeven, heterogene groepen en pas in de laatste plaats homogene groepen. Vlaamse leerkrachten lijken ook eerder voorzichtig te antwoorden, aangezien de extremen (altijd of bijna altijd en nooit) nauwelijks werden aangeduid. Andere West-Europese landen hebben een duidelijkere voorkeur. Zo zal ongeveer de helft van de Zuid-Europese leerkrachten altijd of bijna altijd kiezen voor frontaal lesgeven tijdens de leeslessen. Grote voorstanders van homogene groepen (niveaugroepen) zijn onder andere Noord-Ierland, Engeland en Nederland. Heterogene groepen zijn eerder populair in Italië. Italië en Oostenrijk kiezen het meest frequent van alle West-Europese landen voor zelfstandig werk.

De 126 scholen die deelnamen aan de herhalingsmeting tonen gelijke resultaten aan de originele steekproef in het vierde leerjaar in 2016. Ook hier blijkt dat de voorkeur van Vlaamse leerkrachten uitgaat naar zelfstandig werk, gevolgd door frontaal lesgeven, heterogene groepen en pas in de laatste plaats homogene groepen.

De herhalingsmeting van de 126 scholen in het zesde leerjaar in 2018 toont ongeveer gelijke antwoorden over het gebruik van de verschillende werkvormen. Vooral de dominantie van de antwoordcategorie "soms of vaak" blijft opvallend. We vinden wel dat de categorie "altijd of bijna altijd" voor heterogene groepen en zelfstandig werk stijgt tot respectievelijke 8% en 9%. De categorie "altijd of bijna altijd" homogene groepen daalt in het zesde leerjaar dan weer tot 1%.

4.6 Conclusie

In dit hoofdstuk onderzochten we het niveau van begrijpend lezen van de Vlaamse leerlingen op het einde van het zesde leerjaar in 2018. We bekeken eerst het algemene niveau voor begrijpend lezen en vergeleken deze met de resultaten in het vierde leerjaar. Daarbij gebruikten we ook de internationale 'standaarden' van de IEA. Ook de resultaten van verschillende leerlingengroepen werden bekeken.

We stellen vast dat de Vlaamse leerlingen in het zesde leerjaar in deze steekproef een hoger niveau bereikten dan de leerlingen in het vierde leerjaar van alle andere landen en regio's. We moeten echter nuanceren dat de steekproef van scholen die gebruikt werd voor het zesde leerjaar in 2018 al gemiddeld iets hoger scoorde voor begrijpend lezen in het vierde leerjaar in 2016. Daarbij komt nog dat tussen het vierde en zesde leerjaar zwakkere leerlingen het lager onderwijs verlaten. Het is dus eerder waarschijnlijk dat de leerlingen in het vierde leerjaar in Vlaanderen na twee jaar extra onderwijs niet hoger scoren dan de leerlingen in het vierde leerjaar van de topperpresterende landen. Het lijkt ons correcter te besluiten dat leerlingen in Vlaanderen pas in het zesde leerjaar hetzelfde niveau bereiken als de leerlingen in het vierde leerjaar in topperpresterende landen zoals Rusland en Singapore.

Aansluitend stellen we ook vast dat alle leerlingen in het zesde leerjaar het lage prestatieniveau (*low benchmark*) van de IEA bereikt hebben. Opvallend is echter dat slechts 24% het gevorderd niveau (*advanced benchmark*) bereikt hebben in het zesde leerjaar, dit is minder dan Singapore en Rusland in het vierde leerjaar.

Wat betreft verschillen tussen leerlingengroepen vinden we dat het verschil tussen jongens en meisjes iets groter wordt. Zowel in het vierde als het zesde leerjaar scoren meisjes hoger dan jongens voor begrijpend lezen. Het verschil tussen de leerlingengroepen met verschillende SES is echter onveranderd tussen 2016 en 2018. Ook het verschil in prestaties tussen leerlingen volgens zelfzekerheid over hun leesvaardigheid is afgenomen. Het verschil in prestaties tussen leerlingen volgens hun houding tegenover lezen is echter ongeveer gelijk gebleven.

Tabel 40: Percentages leerlingen waarvoor de werkvormen (1) frontaal lesgeven, (2) homogene groepen, (3) heterogene groepen en (4) zelfstandig werk altijd of bijna altijd, vaak of soms en nooit tijdens de leeslessen worden ingezet (ATBR08A, ATBR08B, ATBR08C & ATBR08E)

Landen	Frontaal lesgeven			Homogene groepen			Heterogene groepen			Zelfstandig werk		
	Altijd of bijna altijd	Vaak of soms	Nooit	Altijd of bijna altijd	Vaak of soms	Nooit	Altijd of bijna altijd	Vaak of soms	Nooit	Altijd of bijna altijd	Vaak of soms	Nooit
Denemarken	15 (2.7)	85 (2.7)	0 (0.2)	5 (1.5)	91 (2.3)	5 (1.7)	3 (1.1)	92 (2.1)	5 (1.8)	13 (2.8)	78 (3.7)	9 (2.2)
Duitsland	26 (2.9)	72 (3.0)	2 (1.0)	2 (1.0)	86 (2.5)	12 (2.4)	3 (0.9)	85 (2.8)	12 (2.6)	9 (1.9)	86 (2.4)	5 (1.7)
Engeland	12 (2.7)	85 (2.7)	3 (1.3)	26 (3.3)	72 (3.3)	3 (1.2)	7 (1.9)	90 (2.3)	3 (1.2)	9 (1.9)	87 (2.4)	4 (1.5)
Finland	26 (2.8)	70 (2.8)	3 (1.1)	1 (0.7)	83 (2.4)	15 (2.4)	4 (1.7)	86 (2.3)	10 (2.0)	6 (1.4)	90 (1.9)	5 (1.4)
Frankrijk	32 (3.4)	65 (3.5)	3 (1.4)	6 (1.8)	78 (3.2)	16 (2.7)	3 (1.5)	67 (3.8)	29 (3.7)	5 (1.7)	68 (3.7)	28 (3.3)
Ierland	25 (3.3)	74 (3.3)	1 (0.4)	9 (2.0)	79 (3.0)	13 (2.5)	5 (1.6)	86 (2.6)	10 (2.1)	5 (1.7)	88 (2.5)	7 (1.8)
Internationaal gemiddelde	32 (0.4)	65 (0.5)	3 (0.2)	11 (0.3)	74 (0.4)	15 (0.4)	13 (0.3)	79 (0.4)	8 (0.3)	14 (0.4)	81 (0.4)	5 (0.2)
Italië	56 (4.0)	44 (4.0)	0 (0.0)	2 (1.2)	70 (3.5)	29 (3.6)	10 (2.7)	79 (3.6)	11 (2.4)	23 (3.4)	73 (3.4)	4 (1.4)
Nederland	7 (2.2)	84 (2.8)	9 (2.0)	17 (3.1)	71 (3.5)	11 (2.6)	7 (2.1)	80 (3.3)	14 (2.9)	11 (2.7)	82 (3.2)	7 (2.1)
Noord-Ierland	6 (2.3)	85 (3.7)	9 (3.0)	55 (4.4)	44 (4.3)	1 (0.8)	1 (0.8)	87 (3.1)	12 (2.9)	10 (2.5)	84 (3.0)	6 (1.9)
Noorwegen (5)	24 (3.0)	76 (3.0)	0 (0.2)	3 (1.4)	78 (3.1)	19 (2.9)	4 (1.6)	87 (2.5)	9 (2.2)	6 (1.9)	86 (2.6)	8 (2.2)
Oostenrijk	20 (2.7)	75 (2.9)	5 (1.4)	4 (1.3)	80 (3.1)	15 (2.8)	4 (1.5)	85 (2.5)	11 (2.1)	23 (2.9)	73 (3.1)	4 (1.2)
Portugal	41 (3.7)	59 (3.7)	0 (0.3)	1 (0.4)	78 (2.8)	21 (2.8)	8 (2.1)	83 (2.8)	9 (2.0)	13 (2.7)	85 (2.8)	2 (1.0)
Spanje	55 (3.1)	45 (3.0)	0 (0.2)	2 (0.7)	70 (2.1)	29 (2.0)	8 (1.4)	71 (2.5)	22 (2.7)	8 (1.9)	77 (2.4)	15 (2.3)
VLAANDEREN	3 (1.6)	92 (2.1)	5 (1.4)	5 (1.9)	83 (2.8)	12 (2.4)	4 (1.6)	88 (2.6)	7 (2.0)	4 (1.4)	95 (1.4)	1 (0.4)
VLAANDEREN(126)	2 (1.1)	93 (1.8)	5 (1.4)	3 (1.1)	85 (2.8)	13 (2.7)	4 (1.5)	89 (2.4)	7 (1.9)	5 (1.5)	94 (1.6)	1 (0.6)
VLAANDEREN2018	5 (1.6)	93 (1.9)	2 (0.9)	1 (0.8)	80 (2.9)	18 (2.9)	8 (2.2)	90 (2.5)	3 (1.1)	9 (1.8)	91 (1.8)	0 (n.v.t.)
Wallonië (FG)	35 (3.7)	64 (3.8)	1 (0.7)	1 (0.5)	68 (4.1)	32 (4.0)	5 (1.9)	73 (3.8)	22 (3.7)	7 (2.1)	84 (3.0)	9 (2.4)
Zweden	23 (3.5)	73 (3.6)	4 (1.7)	2 (1.1)	82 (2.9)	16 (2.8)	6 (2.0)	80 (3.0)	13 (2.7)	6 (2.1)	85 (3.3)	8 (2.7)

Inzake het schoolklimaat is het opvallend dat in het zesde leerjaar er minder leerkrachten aangeven dat er een hoge nadruk is academisch succes dan in het vierde leerjaar. De tevredenheid van leerkrachten blijft daarentegen wel eerder stabiel tussen het vierde leerjaar en het zesde leerjaar. De relatie tussen deze indicatoren en prestaties in begrijpend lezen van leerlingen is relatief beperkt in het vierde leerjaar en ook relatief beperkt in het zesde leerjaar. De opleiding van leerkrachten in het zesde leerjaar is niet verschillend van leerkrachten in het vierde leerjaar. Opvallend is dat de positieve relatie tussen de ervaring van leerkrachten en de leerlingenprestaties in begrijpend lezen verdwijnt in het zesde leerjaar. Het aantal nascholingen blijft ook bij de leerkrachten in het zesde leerjaar beperkt. Ook opvallend is dat de negatieve relatie tussen het volgen van nascholingsactiviteiten en de leerlingenprestaties in begrijpend lezen verdwijnt in het zesde leerjaar. Welke werkvormen gebruikt worden blijft eerder stabiel tussen het vierde en zesde leerjaar.

5 Leerwinst tussen vierde en zesde leerjaar

5.1 Inleiding

>> In dit hoofdstuk onderzoeken we de leerwinst in begrijpend lezen van de Vlaamse leerlingen tussen het vierde leerjaar in 2016 en het zesde leerjaar in 2018. In dit hoofdstuk kijken we ook naar verschillen in leerwinst tussen verschillende leerlingengroepen.

In dit hoofdstuk wordt steeds naar de steekproef van 126 scholen gekeken die deel uitmaakten van de herhalingsmeting in 2018. In dit hoofdstuk zoomen we in op de leerlingen die zowel deelnamen aan de toetsen van PIRLS 2016 als de herhalingsmeting van 2018. Dit zijn dus 4093 leerlingen die tussen het vierde en het zesde leerjaar op dezelfde school blijven en niet zijn blijven zitten. Daardoor handelt dit hoofdstuk dus over de evolutie van de leerlingen tijdens de laatste twee jaar van het lager onderwijs. Verschillen werden steeds getoetst op significantie aan de hand van de standaardfout die geschat werd zoals beschreven in [Hoofdstuk 2](#).

5.1.1 Gemiddelde leerwinst

In [Tabel 41](#) wordt het gemiddelde niveau en de standaarddeviatie in prestaties getoond van dezelfde leerlingen bij PIRLS 2016 en de herhalingsmeting van PIRLS in 2018. Het verschil tussen het prestatieniveau in 2016 en het prestatieniveau in 2018 is 57 punten (S.E.=1.8), een significant verschil.

Tabel 41: Prestaties voor begrijpend lezen (ASRREA01-05)

Jaar	Steekproef (4093 lln in 126 scholen)			
	Gemiddelde (S.E.)		Standaarddeviatie (S.E.)	
2018	586	(1.7)	56	(1.0)
2016	529	(1.6)	59	(0.9)

Is deze leerwinst voor begrijpend lezen tussen het vierde leerjaar en het zesde leerjaar veel of weinig? Welke vooruitgang kunnen we verwachten na 24 maanden onderwijs en ontwikkeling³? Om te duiden of 57 punten op de PIRLS-meetschaal 'veel' of 'weinig' leerwinst is, zijn er geen vaststaande criteria. Daarom presenteren we hier vijf mogelijke *maatstaven* om de Vlaamse leerwinst in de derde graad te kunnen inschatten: (1) landen die met twee opeenvolgende leerjaren deelnamen aan PIRLS, (2) onderzoek van Rindermann, (3) onderzoek van Luyten, (4) onderzoek van Retelsdorf en (5) het Vlaamse SiBO-onderzoek. Het gebruik van meerdere maatstaven wordt beschouwd als een goede praktijk voor het inschatten van de grootte van een effect (Hill, Bloom, Black, & Lipsey, 2008). Naast deze vijf, zijn er ook andere mogelijke maatstaven maar die waren ons niet bekend bij het uitschrijven van dit rapport.

Maatstaf 1: landen die met twee leerjaren deelnamen aan PIRLS

Als eerste maatstaf bekijken we gegevens van andere landen die in het verleden ook herhalings-metingen van PIRLS hebben uitgevoerd. Deze vier herhalingsmetingen staan in [Tabel 42](#) weergegeven.

Tabel 42: Landen die met opeenvolgende leerjaren deelnemen aan PIRLS

Land	Deelname met 2 opeenvolgende leerjaren	Scoreverschil tussen beide leerjaren
Zweden	Grade 3 en Grade 4 in PIRLS 2001	41.71 punten
IJsland	Grade 4 en Grade 5 in PIRLS 2006	38.65 punten
Noorwegen	Grade 4 en Grade 5 in PIRLS 2006	42.28 punten
Noorwegen	Grade 4 en Grade 5 in PIRLS 2016	42.02 punten

3 Doorgaans is ongeveer 1/3e van de leerwinst toe te schrijven aan rijping (*age*) en 2/3e aan onderwijs (*grade*) (Van Damme et al., 2010). Voor de inschattingen van de leerwinst tussen mei 2016 (L4) en mei 2018 (L6) hoeven we geen onderscheid te maken tussen leeftijds effecten en leerjareffecten.

We vinden dat de verschillende landen op één jaar tijd gemiddeld 41.26 punten leerwinst maken op de PIRLS-schaal. Vlaanderen maakt op twee jaar tijd 56.09 punten leerwinst op de PIRLS-schaal. Als we veronderstellen dat de leerwinst per leerjaar even groot is, dan moeten we besluiten dat er in Vlaanderen relatief weinig leerwinst is in begrijpend lezen tussen het vierde en zesde leerjaar.

Maatstaf 2: Rindermann

In 2007 schreef Reiner Rindermann, een Duitse psycholoog, een artikel over de samenhang tussen intelligentiemetingen en de prestatiemetingen van PISA, TIMSS en PIRLS. Omdat de leeftijd van de leerlingen die deelnemen aan PIRLS soms schommelt van land tot land, doet hij een voorstel voor een leeftijdscorrectie. Voor landen waar de leerlingen 1 jaar ouder zijn dan het gemiddelde, stelt hij voor om 42 punten af te trekken om te komen tot een faire score.

Die 42 punten zijn gebaseerd op de gemiddelde groei in één schooljaar volgens verschillende internationale studies (in wiskunde en wetenschappen). Rindermann gebruikt die 42 punten om de PIRLS-scores uit 2001 te corrigeren.

Voor 1 jaar schoollopen is het effect: "... in mathematics from grades 3 to 4, about 60 points (Mullis, Martin, Beaton, Gonzalez, Kelly, & Smith, 1997, p. 31); in mathematics from grades 4 to 8, about 30 points each year (Mullis et al., 1997, p. 43); in science from grades 3 to 4, about 55 points (Martin, Mullis, Beaton, Gonzalez, Smith, & Kelly, 1997, p. 29); in science from grades 4 to 8, about 40 points each year (Martin et al., 1997, p. 41), in mathematics from grades 7 to 8 about 30 points (Beaton, Mullis, Martin, Gonzalez, Kelly, & Smith, 1996, p. 29); in science from grades 7 to 8, about 35 points (Beaton, Martin, Mullis, Gonzalez, Smith, & Kelly, 1996, p. 29). The overall mean of these gains is 42 points." (Rindermann, 2007, p. 675).

De Vlaamse leerlingen gaan op 2 jaar tijd 56 punten vooruit in PIRLS, wat lager ligt dan kan verwacht worden op basis van de schatting van Rindermann dat één extra leerjaar ongeveer 42 punten oplevert.

Maatstaf 3: Luyten

Luyten, Merell en Tymms (2017) analyseerden de leesprestaties⁴ van leerlingen uit Engeland. Het gaat in deze studie om twee verschillende cohorten leerlingen.

Tabel 43: Prestaties 'reading' uit Luyten, Merell en Tymms (2017)

	N	gemiddelde	SD
Cohorte 4 (= 4e leerjaar)	621	9,88	1,60
Cohorte 6 (=6e leerjaar)	600	11,53	1,46

In het vierde leerjaar behalen de leerlingen een gemiddelde van 9,88 punten en in het zesde leerjaar een gemiddelde van 11,53 punten. Hun leerwinst bedraagt 1,65 punten. Om de leerwinst uit verschillende studies te kunnen vergelijken, delen we de leerwinst door de standaarddeviatie van de score in het vierde leerjaar. Als we 1,65 punten delen door de SD van 1,60 bekomen we een leerwinst van 1,03 standaarddeviaties.

De leerwinst tussen het vierde leerjaar en het zesde leerjaar in PIRLS Vlaanderen bedraagt 57 punten. Dat is 0.97 keer de standaarddeviatie van het vierde leerjaar (SD L4 = 59).

De leerwinst in PIRLS in de derde graad in Vlaanderen is dus een beetje lager dan in Engeland.

⁴ In de studie van Luyten et al. (2017) verwijst 'reading' niet enkel naar begrijpend lezen. Het artikel vermeldt: Reading includes word recognition (the pupil hears a word and is asked to choose the correct written version out of five options), word decoding (the pupils must decide which of five written options matches a nonsense word), comprehension (the pupils must fill in the missing word in a sentence), and spelling.

Maatstaf 4: Retelsdorf

Jan Retelsdorf en collega's (2012; 2011) onderzochten de evolutie van begrijpend lezen bij leerlingen in Schleswig-Holstein. Op drie momenten beantwoordden deze leerlingen PIRLS-toetsen en Duitse toetsen begrijpend lezen.

Tabel 44: Prestaties 'reading comprehension' uit Retelsdorf (2012, 2011 en persoonlijke communicatie)

Toets	Gemiddelde	SD	Min.	Max	N
Beginning Grade 5	50.00	10.00	7.73	78.51	1194
End Grade 6	54.73	9.71	17.97	91.63	1194

De Duitse leerlingen gaan op 18 maanden tijd van een gemiddelde score van 50 punten naar een gemiddelde van 54,66 punten. Dat komt overeen met 0,446 standaarddeviaties op 18 maanden tijd. Als we dit doortrekken naar 24 maanden, bedraagt de leerwinst in begrijpend lezen in Duitsland 0,595 standaarddeviaties. De leerwinst in PIRLS in de derde graad in Vlaanderen (0,97 keer de standaarddeviatie) is dus beduidend groter dan de leerwinst in begrijpend lezen in Duitsland.

Maatstaf 5: het Vlaamse SiBO-onderzoek

We kunnen echter ook een vergelijking maken met oudere gegevens van Vlaanderen omtrent de leerwinst tussen het vierde en zesde leerjaar. Zo heeft het SiBO-project de leerwinst tussen het vierde en zesde leerjaar gemeten. De toetsen lagen echter niet op de PIRLS-schaal. Daarom delen we de leerwinst door de standaarddeviatie van begrijpend lezen in het vierde leerjaar. De resultaten van SiBO staan in [Tabel 45](#) weergegeven.

Tabel 45: Descriptieve statistieken SiBO-toetsen begrijpend lezen L4 en L6 voor leerlingen met 2 meetmomenten (N=4022)

Toets	Gemiddelde	SD	Min.	Max	N
SiBOL4 (0607)	49.65	7.47	30,29	68.81	4022
SiBOL6 (0809)	59.09	7.88	37,50	79.59	4022

De leerwinst tussen het vierde leerjaar en het zesde leerjaar bedraagt 9.44 punten in het SiBO-project. Dat is 1.26 keer de standaarddeviatie van het vierde leerjaar. Het cijfer 1,26 kan beschouwd worden als een effectgrootte. Na twee jaar onderwijs én rijping zijn de SiBO-leerlingen 1,26 standaarddeviaties gegroeid in begrijpend lezen.

De leerwinst tussen het vierde leerjaar en het zesde leerjaar in PIRLS Vlaanderen bedraagt 0,97 keer de standaarddeviatie van het vierde leerjaar.

De leerwinst in SiBO bedraagt 1,26 standaarddeviaties en de leerwinst in PIRLS bedraagt 0,97 standaarddeviaties. Als we de veronderstelling maken dat beide studies dezelfde vorm van begrijpend lezen meten en het een gelijkaardige steekproeftrekking betreft, dan wordt er minder leerwinst gemaakt tussen 2016 en 2018.

Vijf maatstaven

Vier maatstaven geven aan dat de Vlaamse leerwinst van 57 punten kleiner is dan verwacht. Slechts één maatstaf (onderzoek van Retelsdorf) geeft aan dat de Vlaamse leerwinst groter is dan verwacht.

We concluderen daarom dat er is dus sprake is van leerwinst in de derde graad, maar er is geen duidelijke inhaalbeweging. Vlaanderen scoort ondermaats voor begrijpend lezen, zowel in het vierde als in het zesde leerjaar.

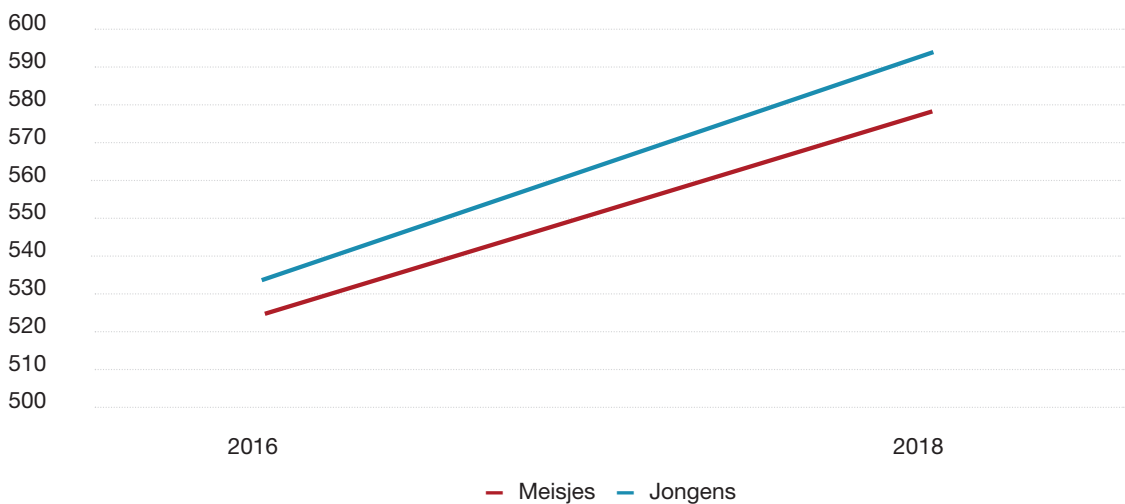
5.1.2 Leerwinst volgens geslacht

Tabel 46 toont de gemiddelde prestatieniveaus voor begrijpend lezen in 2016 en 2018, opgedeeld volgens geslacht. **Figuur 36** toont de gemiddelde leerwinst tussen beide afnamemomenten volgens geslacht. De steekproef bestond uit 51.4% meisjes en 48.6% jongens. Uit de gegevens blijkt dat in 2016 Vlaamse meisjes in met een gemiddelde score van 534 significant beter presteren voor begrijpend lezen dan de Vlaamse jongens die een gemiddelde score van 525 behalen. De gegevens tonen dat ook in 2018 de Vlaamse meisjes met een gemiddelde score van 593 significant beter presteren voor begrijpend lezen dan de Vlaamse jongens die een gemiddelde score van 578 behalen. Het verschil tussen jongens en meisjes is 9 punten in 2016 en 15 punten in 2018. De kloof tussen jongens en meisjes is dus enigszins toegenomen tussen 2016 en 2018.

Tabel 46: Prestaties voor begrijpend lezen (ASRREA01-05) opgedeeld naar geslacht (ITSEX)

Jaar	Meisjes		Jongens	
	Gemiddelde (S.E.)		Standaarddeviatie (S.E.)	
2018	593	(1.7)	578	(2.2)
2016	534	(1.9)	525	(2.2)

Figuur 36: Leerwinst voor begrijpend lezen (ASRREA01-05) opgedeeld naar geslacht (ITSEX)



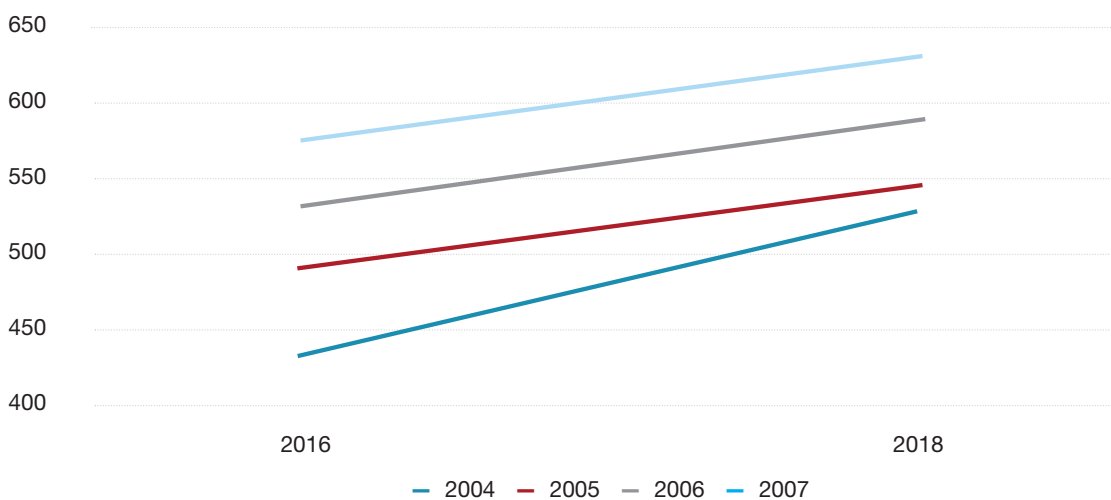
5.1.3 Leerwinst volgens geboortjaar

Tabel 47 toont de gemiddelde prestatieniveaus voor begrijpend lezen in 2016 en 2018, opgedeeld volgens geboortjaar. **Figuur 37** toont de gemiddelde leerwinst tussen beide afnamemomenten volgens geboortjaar. De steekproef bestond uit 0.3% leerlingen geboren in 2004, 8.8% leerlingen geboren in 2005, 89.7% leerlingen geboren in 2006 en 1.2% leerlingen geboren in 2007. Omdat het aantal leerlingen geboren in 2004 en 2007 beperkt is, vergelijken we enkel de normaalvorderende leerlingen geboren in 2006 en de leerlingen met één jaar achterstand geboren in 2005. In 2016 presteren de leerlingen geboren in 2006 significant beter dan de leerlingen geboren in 2005. Ook in 2018 presteren de leerlingen geboren in 2006 significant beter dan de leerlingen geboren in 2005. Het verschil tussen beide groepen leerlingen is 39 punten in 2016 en 43 punten in 2018. Het verschil tussen de leerlingen geboren in 2006 en de leerlingen geboren in 2005 is daarmee stabiel tussen 2016 en 2018.

Tabel 47: Prestaties voor begrijpend lezen (ASRREA01-05) opgedeeld naar geboortjaar (ITBIRTHY)

Jaar	2004		2005		2006		2007	
	Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)	
2018	531	(12.9)	547	(3.2)	590	(1.7)	631	(8.6)
2016	437	(14.3)	493	(3.7)	533	(1.5)	577	(6.7)

Figuur 37: Leerwinst voor begrijpend lezen (ASRREA01-05) opgedeeld naar geboortjaar (ITBIRTHY)



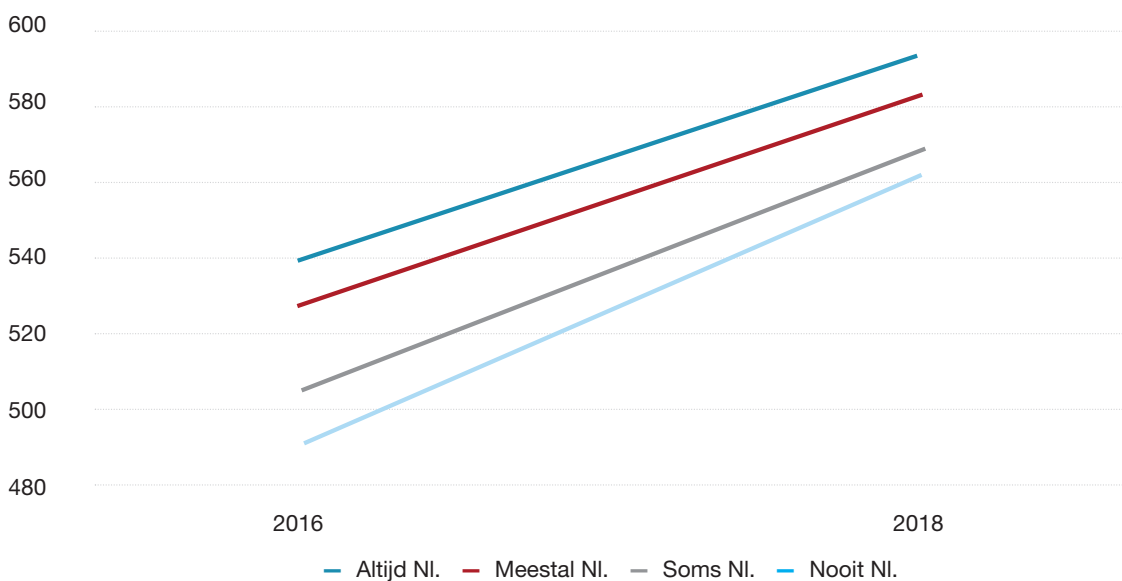
5.1.4 Leerwinst volgens thuistaal

Tabel 48 toont de gemiddelde prestatieniveaus voor begrijpend lezen in 2016 en 2018, opgedeeld naargelang hoe vaak thuis Nederlands gesproken wordt. **Figuur 38** toont de gemiddelde leerwinst tussen beide afnamemomenten volgens hoe vaak thuis Nederlands gesproken wordt. De steekproef bestond uit 66.8% leerlingen die thuis altijd Nederlands spreken, 10.0% leerlingen die thuis meestal Nederlands spreken, 19.2% leerlingen die thuis soms Nederlands spreken en 4.1% leerlingen die thuis nooit Nederlands spreken. We vergelijken hier de leerlingen die thuis altijd Nederlands spreken en de leerlingen die thuis nooit Nederlands spreken. In 2016 presteren de leerlingen die thuis altijd Nederlands spreken met een gemiddelde van 539 significant beter dan de leerlingen die thuis nooit Nederlands spreken met een gemiddelde van 490. Ook in 2018 presteren de leerlingen die thuis altijd Nederlands spreken met een gemiddelde van 593 significant beter dan de leerlingen die thuis nooit Nederlands spreken met een gemiddelde van 562. Het verschil tussen beide groepen leerlingen is 49 in 2016 en 32 in 2018. Het verschil tussen de leerlingen die thuis altijd Nederlands spreken en de leerlingen die thuis nooit Nederlands spreken is dus afgenomen tussen 2016 en 2018. De leerwinst bij de anderstalige leerlingen is dus iets groter dan de leerwinst bij de Nederlandstalige leerlingen.

Tabel 48: Prestaties voor begrijpend lezen (ASRREA01-05) opgedeeld naar thuistaal (ASBG03)

Jaar	Altijd NI.		Meestal NI.		Soms NI.		Nooit NI.	
	Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)	
2018	593	(1.9)	583	(3.2)	569	(3.1)	562	(6.7)
2016	539	(1.6)	528	(3.6)	505	(3.5)	490	(6.9)

Figuur 38: Leerwinst voor begrijpend lezen (ASRREA01-05) opgedeeld naar thuistaal (ASBG03)



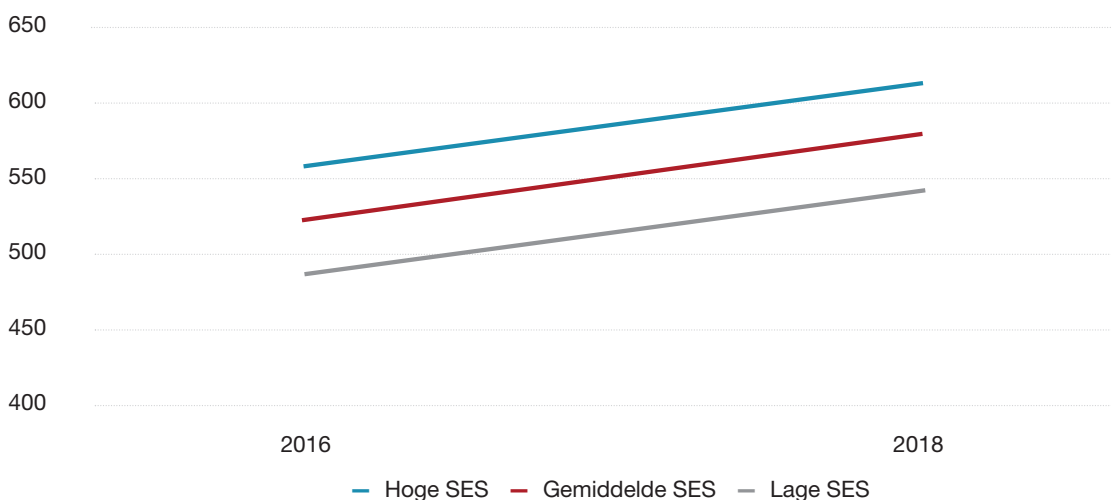
5.1.5 Leerwinst volgens socio-economische status

Tabel 49 toont de gemiddelde prestaties voor begrijpend lezen in 2016 en 2018, opgedeeld volgens de klasse van socio-economische status (*Home Resources for Learning* (HRL))⁵ waartoe de leerlingen behoren. **Figuur 39** toont de gemiddelde leerwinst tussen beide afnamemomenten volgens de drie SES-groepen. De steekproef bestond uit 29.9% leerlingen met een hoge SES, 68.9% leerlingen met een gemiddelde SES en 1.3% leerlingen met een lage SES. We vergelijken de leerlingen met een hoge SES met de leerlingen met een gemiddelde SES. In 2016 presteren de leerlingen met een hoge SES met een gemiddelde van 558 significant beter dan de leerlingen met een gemiddelde SES met een gemiddelde van 523. Ook in 2018 presteren de leerlingen met een hoge SES met een gemiddelde van 613 significant beter dan de leerlingen met een gemiddelde SES met een gemiddelde van 578. Het verschil tussen beide groepen leerlingen is 36 in 2016 en 35 in 2018. Het verschil tussen de leerlingen met een hoge SES en de leerlingen met gemiddelde SES is dus stabiel gebleven.

Tabel 49: Prestaties voor begrijpend lezen (ASRREA01-05) opgedeeld naar SES of home resources (ASDGHRL)

Jaar	Hoge SES	Gemiddelde SES	Lage SES
	Gemiddelde (S.E.)	Gemiddelde (S.E.)	Gemiddelde (S.E.)
2018	613 (2.3)	578 (1.6)	543 (9.6)
2016	558 (2.1)	523 (1.7)	487 (9.5)

Figuur 39: Leerwinst voor begrijpend lezen (ASRREA01-05) opgedeeld naar home resources (ASDGHRL)



⁵ SES is gemeten in PIRLS met de *Home Resources for Learning* schaal. Deze schaal combineert gegevens van leerlingen en hun ouders. Leerlingen gaven in de leerlingvragenlijst aan hoeveel boeken er thuis zijn en welke middelen ze hebben om thuis te studeren (de beschikbaarheid van een eigen bureau en/of een computer). Ouders gaven aan hoeveel kinderboeken er thuis zijn, wat hun opleidingsniveau is, en wat hun beroepsniveau is. Op basis van deze informatie werd een schaalscore berekend.

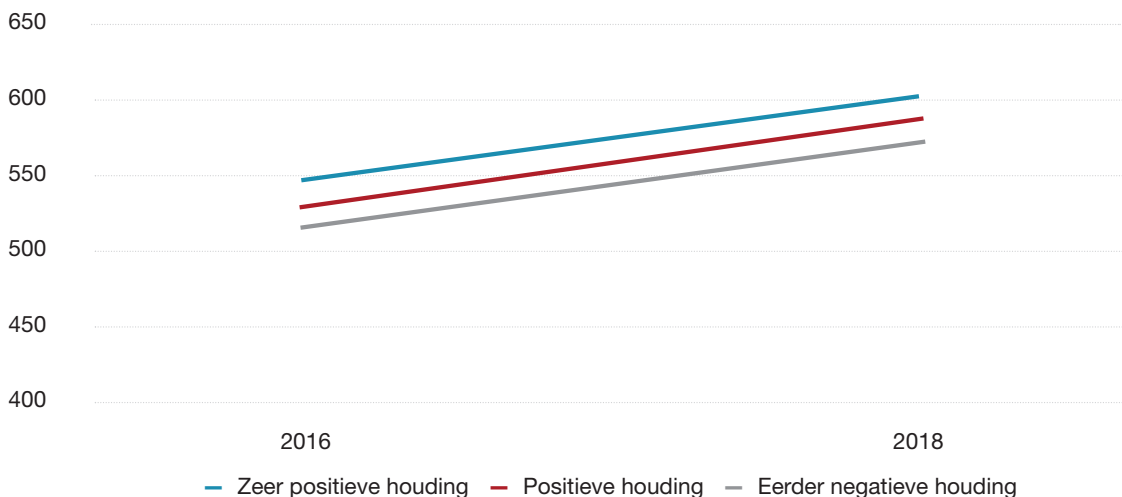
5.1.6 Leerwinst naarmate de houding van ouders tegenover lezen

Tabel 50 toont de gemiddelde prestatieniveaus voor begrijpend lezen in 2016 en 2018, opgedeeld naargelang de houding van ouders tegenover lezen. **Figuur 40** toont de gemiddelde leerwinst tussen beide afnamemomenten volgens de schaal 'Parents like reading' die gebaseerd is op 8 stellingen uit de oudervragenlijst. De steekproef bestond uit 30.1% leerlingen waarvan de ouders een zeer positieve houding hebben tegenover lezen, 47.1% leerlingen waarvan de ouders een positieve houding hebben tegenover lezen en 22.9% leerlingen waarvan de ouders een eerder negatieve houding hebben tegenover lezen. We vergelijken de leerlingen waarvan de ouders een zeer positieve houding hebben tegenover lezen met leerlingen waarvan de ouders een eerder negatieve houding hebben tegenover lezen. In 2016 presteren de leerlingen waarvan de ouders een zeer positieve houding hebben tegenover lezen met een gemiddelde van 548 significant beter dan de leerlingen waarvan de ouders een eerder negatieve houding hebben tegenover lezen met een gemiddelde van 516. Ook in 2018 presteren de leerlingen waarvan de ouders een zeer positieve houding hebben tegenover lezen met een gemiddelde van 601 significant beter dan de leerlingen waarvan de ouders een eerder negatieve houding hebben tegenover lezen met een gemiddelde van 572. Het verschil tussen beide groepen leerlingen is 32 in 2016 en 29 in 2018. Het verschil tussen de leerlingen waarvan de ouders een zeer positieve houding hebben tegenover lezen en de leerlingen waarvan de ouders een eerder negatieve houding hebben tegenover lezen is dus stabiel tussen 2016 en 2018.

Tabel 50: Prestaties voor begrijpend lezen (ASRREA01-05) naarmate houding van ouders tegenover lezen (ASDHPLR)

Jaar	Zeer positieve houding		Positieve houding		Eerder negatieve houding	
	Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)	
2018	601	(2.0)	586	(1.8)	572	(2.5)
2016	548	(2.1)	529	(1.9)	516	(2.8)

Figuur 40: Leerwinst voor begrijpend lezen (ASRREA01-05) naarmate houding van ouders tegenover lezen (ASDHPLR)



5.1.7 Leerwinst naarmate leerlingen betrokken zijn bij lezen

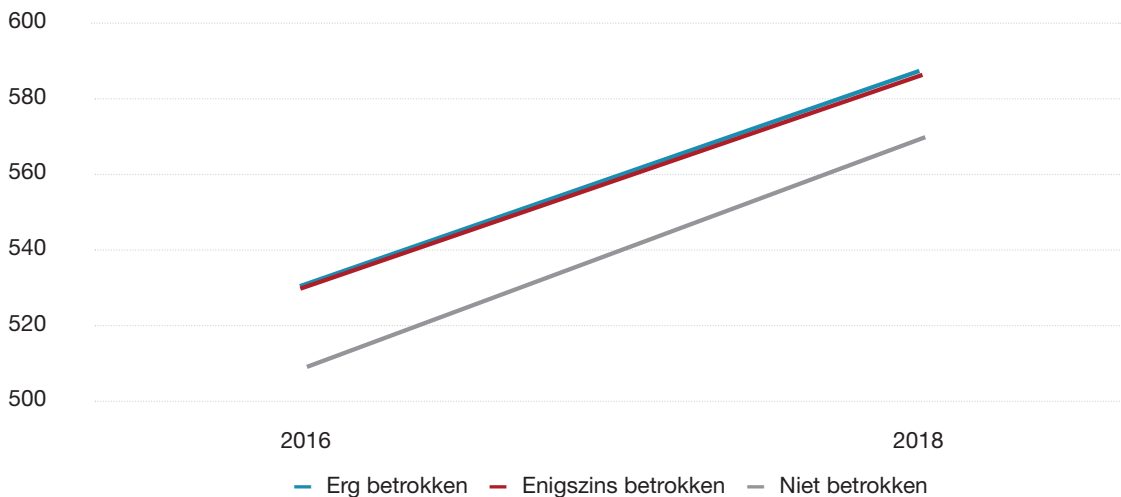
Tabel 51 toont de gemiddelde prestatieniveaus voor begrijpend lezen in 2016 en 2018, opgedeeld naargelang de mate waarin leerlingen betrokken waren bij lezen in 2016. **Figuur 41** toont de gemiddelde leerwinst tussen beide afnamemomenten volgens de mate waarin de leerlingen betrokken zijn bij lezen. De leerlingen worden opgedeeld in drie groepen volgens hun score op de schaal 'Students engaged in reading lessons'. Deze schaal is gebaseerd op 9 vragen uit de leerlingvragenlijst.

De steekproef bestond uit 49.5% leerlingen die erg betrokken zijn bij lezen, 46.8% leerlingen die enigszins betrokken zijn bij lezen en 3.7% leerlingen die niet betrokken zijn bij lezen. We vergelijken de leerlingen die erg betrokken zijn bij lezen met de leerlingen die enigszins betrokken zijn bij lezen. In 2016 presteren de leerlingen die erg betrokken zijn bij lezen met een gemiddelde van 531 gelijk aan de leerlingen die enigszins betrokken zijn bij lezen met een gemiddelde van 530. Ook in 2018 presteren de leerlingen die erg betrokken zijn bij lezen met een gemiddelde van 587 gelijk aan de leerlingen die enigszins betrokken zijn bij lezen met een gemiddelde van 586. Het verschil tussen beide groepen leerlingen is 1 in 2016 en 1 in 2018. Het verschil tussen de leerlingen die erg betrokken zijn bij lezen en de leerlingen die enigszins betrokken zijn bij lezen is dus stabiel gebleven tussen 2016 en 2018.

Tabel 51: Prestaties voor begrijpend lezen (ASRREA01-05) naarmate leerlingen betrokken zijn bij lezen (ASDGERL)

Jaar	Erg betrokken		Enigszins betrokken		Niet betrokken	
	Gemiddelde	(S.E.)	Gemiddelde	(S.E.)	Gemiddelde	(S.E.)
2018	587	(1.8)	586	(2.2)	570	(6.1)
2016	531	(2.1)	530	(2.0)	509	(6.2)

Figuur 41: Leerwinst voor begrijpend lezen (ASRREA01-05) naarmate leerlingen betrokken zijn bij lezen (ASDGERL)



5.1.8 Leerwinst naarmate de houding van leerlingen tegenover lezen

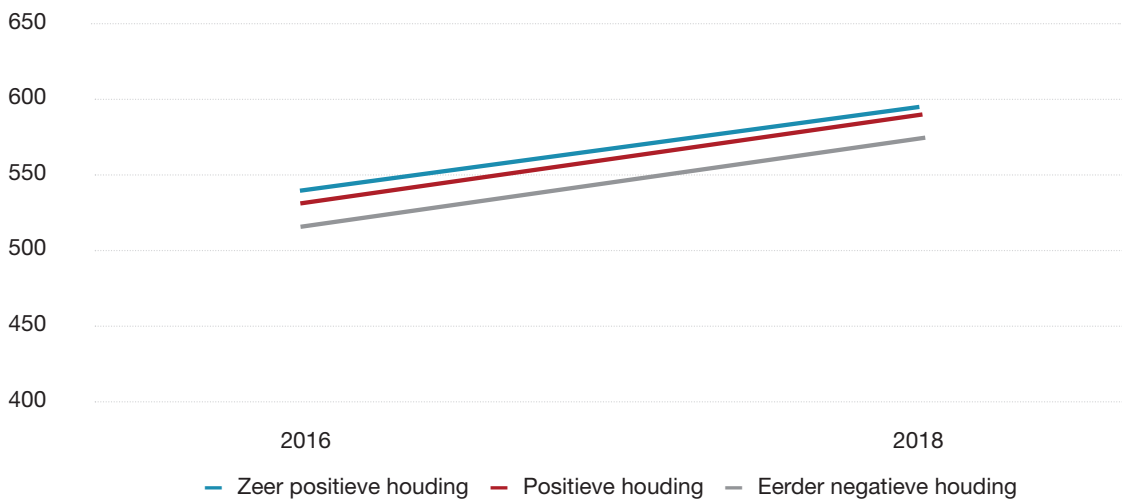
Tabel 52 toont de gemiddelde prestatieniveaus voor begrijpend lezen in 2016 en 2018, opgedeeld naargelang de houding van leerlingen tegenover lezen in 2016. **Figuur 42** toont de gemiddelde leerwinst tussen beide afname-momenten voor de drie groepen leerlingen. De leerlingen worden opgedeeld in drie groepen volgens de schaal 'Students like reading' (SLR), die gebaseerd is op 8 stellingen die door elke leerling in 2016 werden beoordeeld (Tielemans et al., 2017).

De steekproef bestond uit 24.9% leerlingen met een zeer positieve houding, 46.5% leerlingen met een positieve houding en 28.6% leerlingen met een eerder negatieve houding. We vergelijken de leerlingen met een zeer positieve houding met de leerlingen met een eerdere negatieve houding. In 2016 presteren de leerlingen met een zeer positieve houding (gemiddelde: 539) significant beter dan de leerlingen die een eerder negatieve houding hebben (gemiddelde: 516). Ook in 2018 presteren de leerlingen met een zeer positieve houding met een gemiddelde van 594 significant beter dan de leerlingen met een eerder negatieve houding met een gemiddelde van 575. Het verschil tussen beide groepen leerlingen is 23 in 2016 en 19 in 2018. Het verschil tussen de leerlingen met een zeer positieve houding en de leerlingen met een eerder negatieve houding is dus stabiel gebleven tussen 2016 en 2018.

Tabel 52: Prestaties voor begrijpend lezen (ASRREA01-05) naarmate leerlingen van lezen houden (ASDGSLR)

Jaar	Zeer positieve houding		Positieve houding		Eerder negatieve houding	
	Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)	
2018	594	(2.5)	589	(1.8)	575	(2.3)
2016	539	(2.5)	533	(1.7)	516	(2.3)

Figuur 42: Leerwinst voor begrijpend lezen (ASRREA01-05) naarmate leerlingen van lezen houden (ASDGSLR)



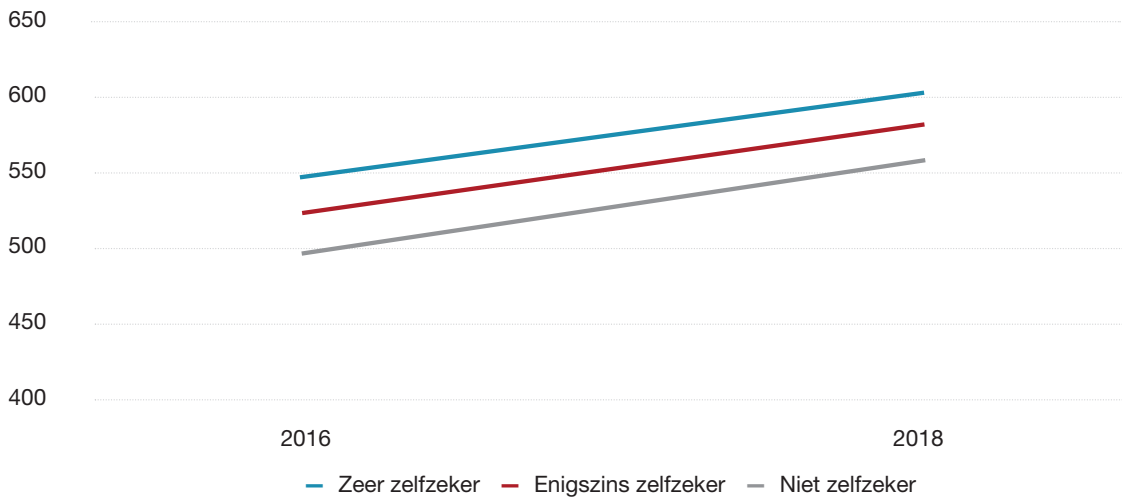
5.1.9 Leerwinst naarmate leerlingen zelfzeker zijn in hun leesvaardigheid

Tabel 53 toont de gemiddelde prestatieniveaus voor begrijpend lezen in 2016 en 2018, opgedeeld naargelang leerlingen hun zelfzekerheid in begrijpend lezen in 2016. **Figuur 43** toont de gemiddelde leerwinst tussen beide afnamemomenten volgens de schaal 'Students confident in reading' (SCR). De steekproef bestond uit 46.9% leerlingen die zeer zelfzeker zijn in hun leesvaardigheid, 34.2% leerlingen die enigszins zelfzeker zijn in hun leesvaardigheid en 18.9% leerlingen die niet zelfzeker zijn in hun leesvaardigheid. We vergelijken de leerlingen die zeer zelfzeker zijn in hun leesvaardigheid met de leerlingen die niet zelfzeker zijn in hun leesvaardigheid. In 2016 presteren de leerlingen die zeer zelfzeker zijn in hun leesvaardigheid met een gemiddelde van 548 significant beter dan de leerlingen die niet zelfzeker zijn in hun leesvaardigheid met een gemiddelde van 497. Ook in 2018 presteren de leerlingen die zeer zelfzeker zijn in hun leesvaardigheid met een gemiddelde van 601 significant beter dan de leerlingen die niet zelfzeker zijn in hun leesvaardigheid met een gemiddelde van 558. Het verschil tussen beide groepen leerlingen is 51 in 2016 en 43 in 2018. Het verschil tussen de leerlingen die zeer zelfzeker zijn in hun leesvaardigheid en de leerlingen die niet zelfzeker zijn in hun leesvaardigheid is dus enigszins afgenomen.

Tabel 53: Prestaties voor begrijpend lezen (ASRREA01-05) naarmate leerlingen zelfzeker zijn in hun leesvaardigheid (ASDGSCR)

Jaar	Zeer zelfzeker		Enigszins zelfzeker		Niet zelfzeker	
	Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)	
2018	601	(1.9)	581	(2.3)	558	(2.4)
2016	548	(1.9)	522	(2.4)	497	(2.4)

Figuur 43: Leerwinst voor begrijpend lezen (ASRREA01-05) naarmate leerlingen zelfzeker zijn in hun leesvaardigheid (ASDGSCR)



5.2 Conclusie

In dit hoofdstuk onderzochten we de leerwinst voor begrijpend lezen voor de Vlaamse leerlingen tussen het einde van het vierde leerjaar in 2016 en het einde van het zesde leerjaar in 2018. We bekeken eerst de gemiddelde leerwinst en vergeleken deze met vijf maatstaven over de leerwinst voor begrijpend lezen. In dit hoofdstuk keken we ook naar verschillen in leerwinst tussen verschillende leerlingengroepen.

De resultaten voor de gemiddelde leerwinst duiden er niet op dat Vlaanderen een inhaalbeweging maakt voor begrijpend lezen tussen het vierde en zesde leerjaar (zie Vlor, 2018). We moeten zelfs vaststellen dat de leerwinst op jaarbasis waarschijnlijk kleiner is dan verwacht kan worden. We concluderen dat het Vlaamse lager onderwijs zowel in het vierde leerjaar als in het zesde leerjaar gekenmerkt wordt door ondermaatse prestaties in begrijpen lezen.

Wanneer we verschillende leerlingengroepen onderzoeken vinden we ook enkele verschillen in leerwinst voor begrijpend lezen. Zo maken meisjes meer leerwinst tussen het vierde en zesde jaar dan de jongens. Leerlingen die thuis altijd Nederlands spreken hebben dan weer minder leerwinst dan leerlingen die thuis soms Nederlands spreken. Gemiddeld genomen zien we ook dat leerlingen die laag scoorden in 2016 meer leerwinst maken dan de leerlingen die hoog scoorden in 2016. Daarentegen vinden we geen verschil in leerwinst naargelang de SES-status van leerlingen, hoeveel ze van lezen houden, hoe sterk betrokken ze zijn bij lezen en hoeveel de ouders van lezen houden. Leerlingen die niet zelfzeker waren in hun leesvaardigheid in 2016 maken dan weer meer leerwinst.

6 Evolutie in attitudes tussen vierde en zesde leerjaar

6.1 Inleiding

>> In dit hoofdstuk onderzoeken we de evolutie van attitudes m.b.t. begrijpend lezen van de Vlaamse leerlingen tussen het vierde leerjaar in 2016 en het zesde leerjaar in 2018. We richten ons daarbij op verschillen in trends tussen verschillende leerlingengroepen. We onderzoeken drie attitudes:

1. zelfzekerheid van leerlingen over hun leesvaardigheid ('Students Confident in Reading Scale', 6 items)⁶
2. betrokkenheid van leerlingen bij lezen ('Students Engaged in Reading Lessons Scale', 9 items)
3. de houding van leerlingen tegenover lezen ('Students Like Reading Scale', 8 items + 2 items)

De leerlingen beoordeelden dezelfde stellingen in de leerlingvragenlijst van 2016 en in de leerlingvragenlijst van 2018. De wijze waarop IRT werd toegepast om tot scores per leerling te komen, is toegelicht in **Hoofdstuk 3**. Voor elke leerling werden dus twee IRT-scores berekend: eentje voor 2016 (berekend door IEA) en eentje voor 2018 (berekend door PIRLS-team Vlaanderen). In sommige analyses worden de leerlingen in drie groepen opgedeeld volgens hun IRT-score.

In dit hoofdstuk wordt steeds naar de steekproef van 126 scholen gekeken die deel uitmaakten van de herhalingsmeting in 2018. In dit hoofdstuk analyseren we de gegevens van de 4093 leerlingen die tweemaal deelnamen aan PIRLS: in 2016 en 2018. Het gaat om leerlingen die niet van school veranderden en die niet zijn blijven zitten tussen het vierde en het zesde leerjaar. Daardoor handelt dit hoofdstuk dus over de evolutie in attitudes van de leerlingen tijdens de laatste twee jaar van het lager onderwijs. Verschillen werden steeds getoetst op significantie aan de hand van de standaardfout die geschat werd zoals beschreven in **Hoofdstuk 2**.

6.2 Evolutie van zelfzekerheid van leerlingen over hun leesvaardigheid tussen 4^{de} en 6^{de} leerjaar

De houding tegenover lezen werd zowel in de leerlingvragenlijst van 2016 als in de leerlingvragenlijst van 2018 bevraagd met de volgende 6 items:

Hoe goed ben je in lezen? Geef telkens aan in hoeverre je akkoord gaat met de uitspraak.

1. Meestal ben ik goed in lezen
2. Lezen is gemakkelijk voor mij
3. Ik vind het moeilijk om verhalen te lezen met moeilijke woorden erin (-)
4. Lezen is voor mij moeilijker dan voor veel van mijn klasgenoten (-)
5. Lezen is voor mij moeilijker dan alle andere vakken op school (-)
6. Ik ben gewoon niet goed in lezen (-)

De leerlingen gaven aan op een 4-puntenschaal in hoeverre ze akkoord gingen: helemaal akkoord / eerder akkoord / eerder niet akkoord / helemaal niet akkoord

Opmerking: (-) duidt op inverse codering

⁶ We merken op dat in het rapport van Tielemans et al. (2017) zelfzekerheid werd omschreven met de term 'zelfconcept'.

6.2.1 Gemiddelde evolutie

In **Tabel 54** wordt het gemiddelde niveau en de standaarddeviatie in zelfzekerheid in leesvaardigheid getoond van dezelfde leerlingen bij PIRLS 2016 en de herhalingsmeting van PIRLS in 2018. Het verschil tussen het niveau in 2016 en het niveau in 2018 is 0.11 (S.E.=0.04), een significant verschil, maar verwaarloosbaar in grootte. De leerlingen zijn in de periode van 2 jaar iets zelfzekerder geworden in leesvaardigheid.

Tabel 54: Zelfzekerheid in leesvaardigheid (ASBGSCR)

Jaar	Steekproef (N = 4093)			
	Gemiddelde (S.E.)		Standaarddeviatie (S.E.)	
2018	10.14	(0.04)	1.93	(0.03)
2016	10.02	(0.04)	2.02	(0.03)

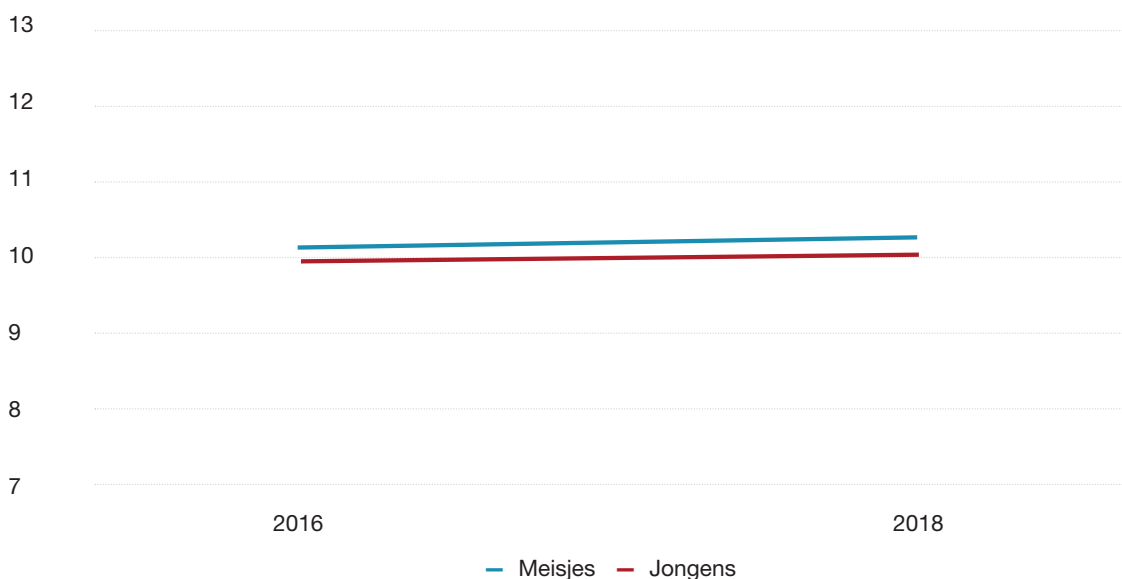
6.2.2 Evolutie volgens geslacht

Tabel 55 toont de gemiddelde niveaus in zelfzekerheid in 2016 en 2018, opgedeeld volgens geslacht. **Figuur 44** toont de gemiddelde evolutie tussen beide afnamemomenten volgens geslacht. De steekproef bestond uit 51.8% meisjes en 48.2% jongens. Uit de resultaten blijkt dat in 2016 Vlaamse meisjes in met een gemiddelde van 10.09 significant meer zelfzeker zijn in hun leesvaardigheid dan Vlaamse jongens die een gemiddelde van 9.94 behalen. De gegevens tonen dat ook in 2018 de Vlaamse meisjes met een gemiddelde van 10.23 significant meer zelfzeker zijn in hun leesvaardigheid dan de Vlaamse jongens die een gemiddelde score van 10.03 behalen. Het verschil tussen jongens en meisjes is 0.15 in 2016 en 0.20 in 2018. De verschillen zijn echter verwaarloosbaar in grootte. Het verschil tussen jongens en meisjes is niet significant gewijzigd tussen 2016 en 2018.

Tabel 55: Zelfzekerheid in leesvaardigheid (ASBGSCR) opgedeeld naar geslacht (ITSEX)

Jaar	Meisjes		Jongens	
	Gemiddelde (S.E.)		Gemiddelde (S.E.)	
2018	10.23	(0.05)	10.03	(0.05)
2016	10.10	(0.05)	9.94	(0.06)

Figuur 44: Evolutie in zelfzekerheid in leesvaardigheid (ASBGSCR) opgedeeld naar geslacht (ITSEX)



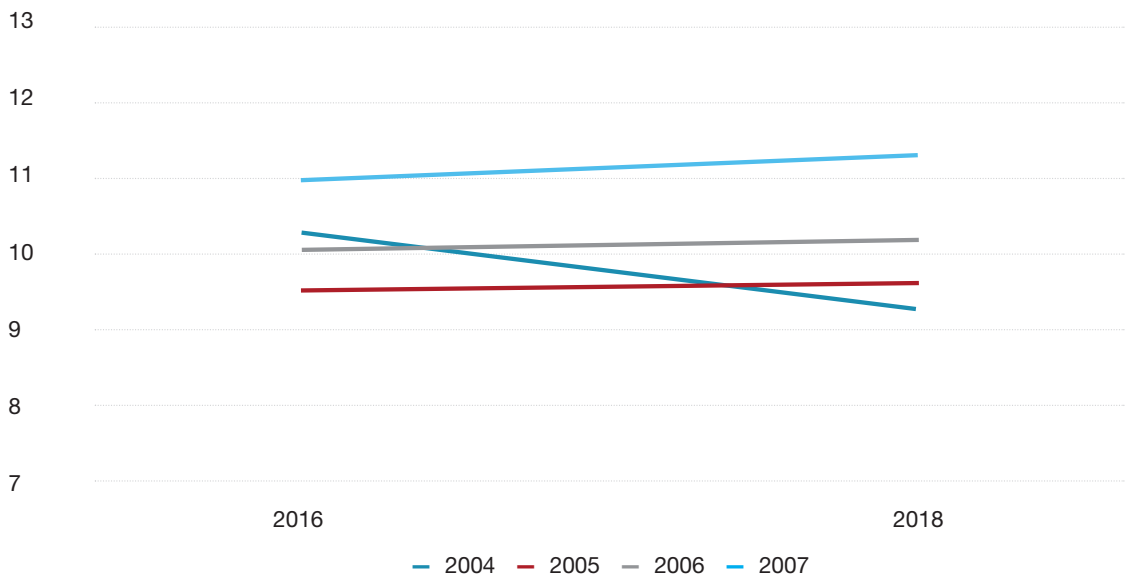
6.2.3 Evolutie volgens geboortjaar

Tabel 56 toont de gemiddelde niveaus in zelfzekerheid in 2016 en 2018, opgedeeld volgens geboortjaar. **Figuur 45** toont de gemiddelde evolutie tussen beide afnamemomenten volgens geboortjaar. Omdat het aantal leerlingen geboren in 2004 en 2007 beperkt is, vergelijken we enkel de normaalvorderende leerlingen geboren in 2006 en de leerlingen met één jaar achterstand geboren in 2005. In 2016 hebben de leerlingen geboren in 2006 een significant hoger gemiddelde van 10.06 dan de leerlingen geboren in 2005 met een gemiddelde van 9.51. Ook in 2018 hebben de leerlingen geboren in 2006 een significant hoger gemiddelde van 10.17 dan de leerlingen geboren in 2005 met een gemiddelde van 9.61. Het verschil tussen beide groepen leerlingen is 0.54 in 2016 en 0.56 in 2018. Het verschil tussen de leerlingen geboren in 2006 en de leerlingen geboren in 2005 is dus stabiel van het vierde naar het zesde leerjaar.

Tabel 56: Zelfzekerheid in leesvaardigheid (ASBGSCR) opgedeeld naar geboortjaar (ITBIRTHY)

Jaar	2004		2005		2006		2007	
	Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)	
2018	9.28	(0.45)	9.61	(0.12)	10.17	(0.04)	11.30	(0.26)
2016	10.28	(0.48)	9.51	(0.11)	10.06	(0.04)	10.97	(0.22)

Figuur 45: Evolutie in zelfzekerheid (ASBGSCR) opgedeeld naar geboortjaar (ITBIRTHY)



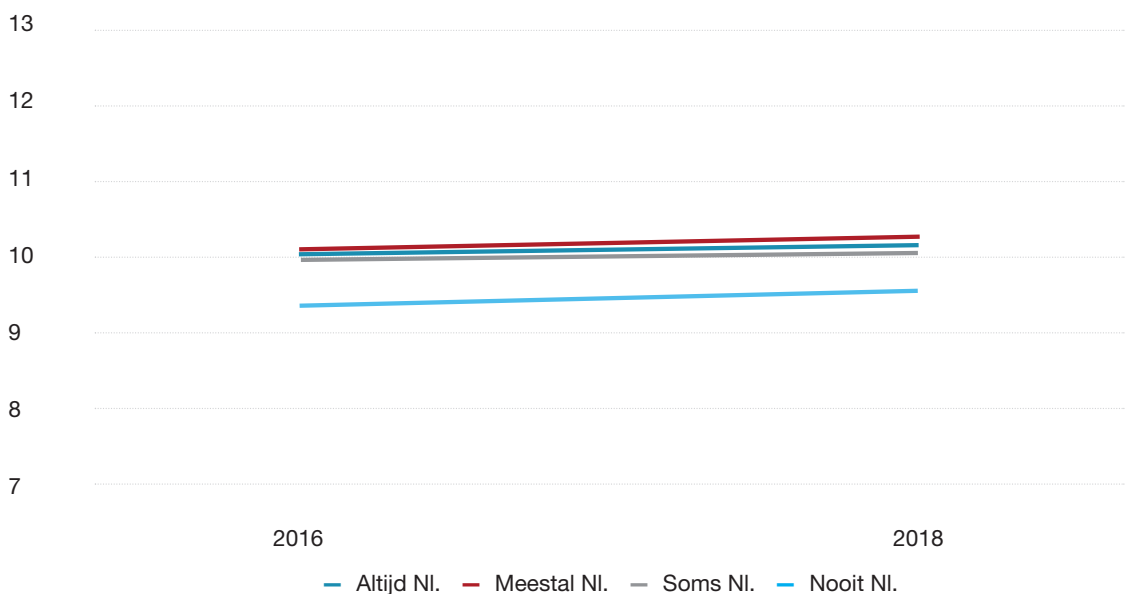
6.2.4 Evolutie volgens thuistaal

Tabel 57 toont de gemiddelde niveaus in zelfzekerheid in 2016 en 2018, opgedeeld volgens hoe vaak thuis Nederlands gesproken wordt. **Figuur 46** toont de gemiddelde evolutie tussen beide afnamemomenten volgens hoe vaak thuis Nederlands gesproken wordt. We vergelijken hier de leerlingen die thuis altijd Nederlands spreken en de leerlingen die thuis nooit Nederlands spreken. In 2016 hebben de leerlingen die thuis altijd Nederlands spreken een significant hoger gemiddelde van 10.06 dan de leerlingen die thuis nooit Nederlands spreken met een gemiddelde van 9.35. Ook in 2018 hebben de leerlingen die thuis altijd Nederlands spreken een significant hoger gemiddelde van 10.17 dan de leerlingen die thuis nooit Nederlands spreken met een gemiddelde van 9.56. Het verschil tussen beide groepen leerlingen is 0.71 in 2016 en 0.62 in 2018. Het verschil tussen de leerlingen die thuis altijd Nederlands spreken en de leerlingen die thuis nooit Nederlands spreken is dus stabiel van het vierde naar het zesde leerjaar.

Tabel 57: zelfzekerheid in leesvaardigheid (ASBGSCR) opgedeeld naar thuistaal (ASBG03)

Jaar	Altijd NI.	Meestal NI.	Soms NI.	Nooit NI.
	Gemiddelde (S.E.)	Gemiddelde (S.E.)	Gemiddelde (S.E.)	Gemiddelde (S.E.)
2018	10.17 (0.05)	10.32 (0.10)	10.05 (0.08)	9.56 (0.11)
2016	10.06 (0.05)	10.14 (0.12)	9.98 (0.07)	9.35 (0.15)

Figuur 46: Zelfzekerheid in leesvaardigheid (ASBGSCR) opgedeeld naar thuistaal (ASBG03)



6.2.5 Evolutie volgens benchmarks 2016

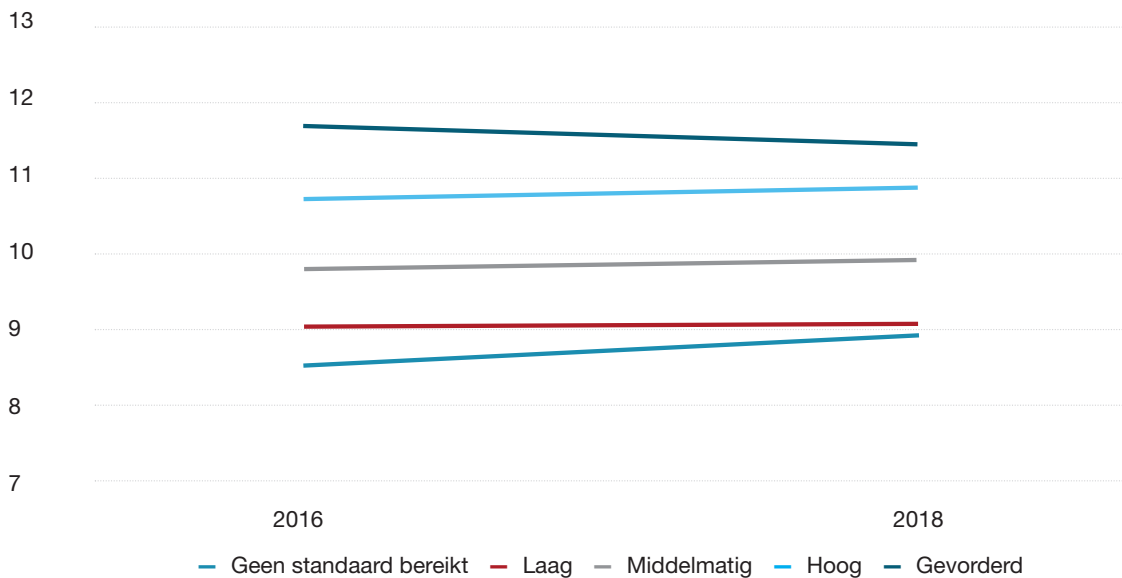
Tabel 58 toont de gemiddelde niveaus in zelfzekerheid in leesvaardigheid in 2016 en 2018, opgedeeld volgens de *benchmark* of standaard voor begrijpend lezen die de leerlingen behaalden in 2016. **Figuur 47** toont de gemiddelde evolutie tussen beide afnamemomenten volgens de standaard voor begrijpend lezen waartoe de leerlingen behoorden in 2016.

We vergelijken de leerlingen die tot de lage benchmark behoorden in 2016 met leerlingen die tot de hoge standaard behoorden in 2016. In 2016 hebben de leerlingen die tot de hoge standaard behoorden een significant hoger gemiddelde van 10.72 dan de leerlingen die tot de lage standaard behoorden met een gemiddelde van 9.06. In 2018 hebben de leerlingen die tot de hoge standaard behoorden een significant hoger gemiddelde van 10.87 dan de leerlingen die tot de lage standaard behoorden met een gemiddelde van 9.13. Het verschil tussen beide groepen leerlingen is 1.65 in 2016 en 1.75 in 2018. Het verschil tussen de leerlingen die de lage standaard behaalden in 2016 en de leerlingen die de hoge standaard behaalden in 2018 is dus eerder stabiel.

Tabel 58: Zelfzekerheid in leesvaardigheid (ASBGSCR) opgedeeld naar benchmark 2016

Jaar	Laag		Middelmatig		Hoog		Gevorderd	
	<400	>400 en <475	>475 en <550	>550 en <625	>550 en <625	>625	>625	>625
	Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)	
2018	8.93 (0.22)	9.13 (0.09)	9.86 (0.05)	10.87 (0.06)	10.87 (0.06)	11.45 (0.17)	11.45 (0.17)	11.45 (0.17)
2016	8.50 (0.21)	9.06 (0.09)	9.75 (0.06)	10.72 (0.07)	10.72 (0.07)	11.70 (0.14)	11.70 (0.14)	11.70 (0.14)

Figuur 47: Evolutie in zelfzekerheid in leesvaardigheid (ASBGSCR) opgedeeld naar benchmark 2016



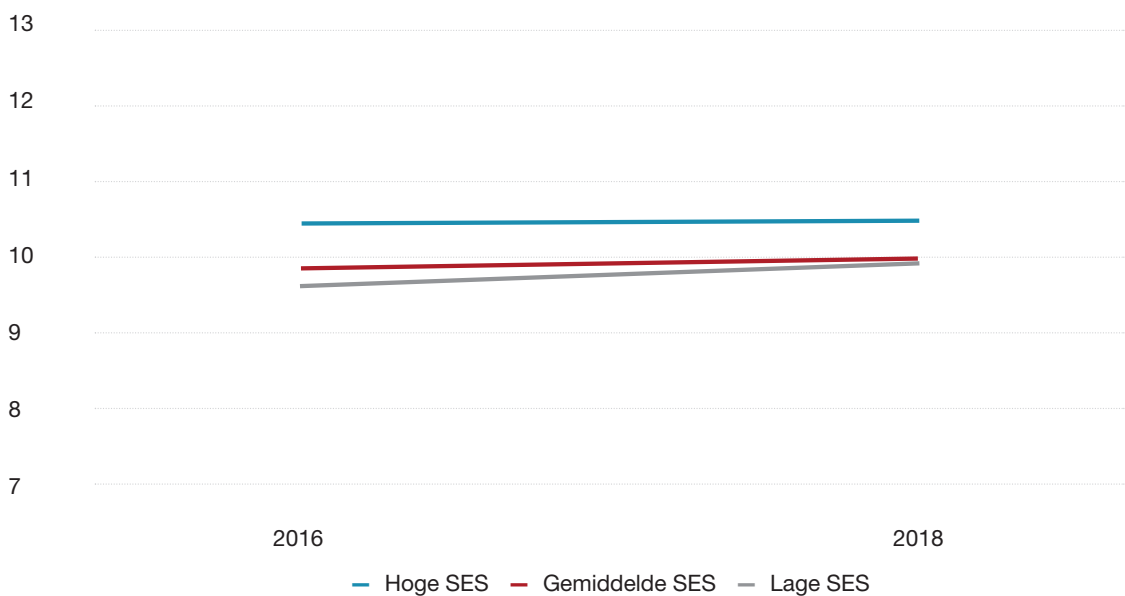
6.2.6 Evolutie volgens socio-economische status

Tabel 59 toont de gemiddelde niveaus in zelfzekerheid in 2016 en 2018, opgedeeld volgens tot welke klasse van socio-economische status (*Home Resources for Learning* (HRL)) de leerlingen behoren. **Figuur 48** toont de gemiddelde evolutie tussen beide afnamemomenten volgens de klasse van socio-economische status waartoe de leerlingen behoren. We vergelijken de leerlingen met een hoge SES met de leerlingen met een gemiddelde SES. In 2016 hebben de leerlingen met een hoge SES een significant hoger gemiddelde van 10.45 dan de leerlingen met een gemiddelde SES met een gemiddelde van 9.86. Ook in 2018 hebben de leerlingen met een hoge SES een significant hoger gemiddelde van 10.51 dan de leerlingen met een gemiddelde SES met een gemiddelde van 9.98. Het verschil tussen beide groepen leerlingen is 0.59 in 2016 en 0.55 in 2018. Het verschil tussen de leerlingen met een hoge SES en de leerlingen met gemiddelde SES is dus stabiel gebleven.

Tabel 59: zelfzekerheid in leesvaardigheid (ASBGSCR) opgedeeld naar SES of home resources (ASDGHRL)

Jaar	Hoge SES	Gemiddelde SES	Lage SES
	Gemiddelde (S.E.)	Gemiddelde (S.E.)	Gemiddelde (S.E.)
2018	10.51 (0.07)	9.98 (0.05)	9.94 (0.21)
2016	10.45 (0.07)	9.86 (0.04)	9.63 (0.29)

Figuur 48: Evolutie in zelfzekerheid in leesvaardigheid (ASBGSCR) opgedeeld naar home resources (ASDGHRL)



6.3 Evolutie van betrokkenheid bij lezen tussen 4^{de} en 6^{de} leerjaar

De betrokkenheid bij lezen werd zowel in de leerlingvragenlijst van 2016 als in de leerlingvragenlijst van 2018 bevestigd met de volgende 9 items:

Denk aan wat je moet lezen voor school. In hoeverre ga je akkoord met volgende uitspraken over de leeslessen?

1. Ik vind de dingen die ik lees op school leuk.
2. De meester of juf geeft me interessante dingen om te lezen
3. Ik weet wat ik van mijn meester of juf moet doen.
4. Ik begrijp mijn juf of meester goed.
5. Ik ben geïnteresseerd in wat mijn juf of meester vertelt.
6. Mijn juf of meester moedigt me aan om te vertellen wat ik heb gelezen.
7. Mijn juf of meester geeft me de kans te laten zien wat ik geleerd heb.
8. Mijn juf of meester doet verschillende dingen om ons te helpen met leren.
9. Mijn juf of meester vertelt me hoe ik het beter kan doen als ik een fout maak.

De leerlingen gaven aan op een 4-puntenschaal in hoeverre ze akkoord gingen: helemaal akkoord / eerder akkoord / eerder niet akkoord / helemaal niet akkoord

6.3.1 Gemiddelde evolutie

In **Tabel 60** wordt het gemiddelde niveau en de standaarddeviatie in de mate van betrokkenheid bij lezen (Students Engaged in Reading Lessons Scale) getoond van dezelfde leerlingen bij PIRLS 2016 en de herhalingsmeting van PIRLS in 2018. Het verschil tussen het niveau in 2016 en het niveau in 2018 is 0.58 (S.E.=0.05), een significante daling.

Tabel 60: Betrokkenheid bij lezen (ASBGERL)

Jaar	Steekproef			
	Gemiddelde (S.E.)		Standaarddeviatie (S.E.)	
2018	8.92	(0.04)	1.18	(0.03)
2016	9.50	(0.05)	1.59	(0.03)

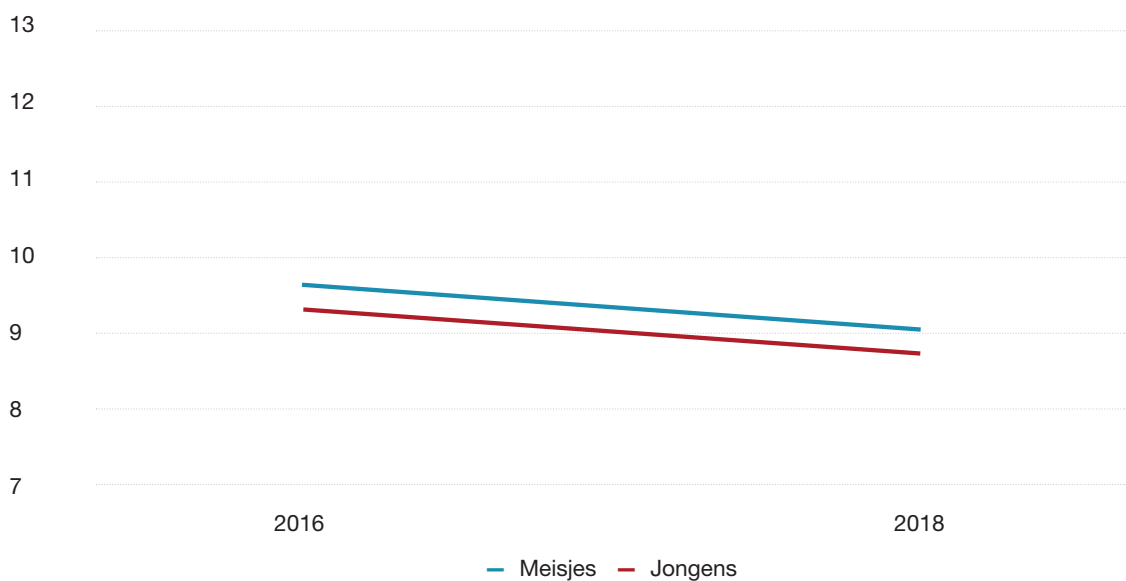
6.3.2 Evolutie volgens geslacht

Tabel 61 toont de gemiddelde niveaus in betrokkenheid in 2016 en 2018, opgedeeld volgens het geslacht van leerlingen. **Figuur 49** toont de gemiddelde evolutie tussen beide afnamemomenten volgens geslacht. Uit de gegevens blijkt dat in 2016 Vlaamse meisjes een significant hoger gemiddelde hebben van 9.65 dan de Vlaamse jongens die een gemiddelde hebben van 9.07. De gegevens tonen dat ook in 2018 de Vlaamse meisjes een significant hoger gemiddelde hebben van 9.35 dan de Vlaamse jongens die een gemiddelde hebben van 8.76. Het verschil tussen jongens en meisjes is 0.30 in 2016 en 0.32 in 2018. Het verschil tussen jongens en meisjes is dus stabiel gebleven tussen 2016 en 2018.

Tabel 61: Betrokkenheid bij lezen (ASBGERL) opgedeeld naar geslacht (ITSEX)

Jaar	Meisjes		Jongens	
	Gemiddelde (S.E.)		Gemiddelde (S.E.)	
2018	9.07	(0.04)	8.76	(0.05)
2016	9.65	(0.05)	9.35	(0.06)

Figuur 49: Evolutie in betrokkenheid (ASBGERL) opgedeeld naar geslacht (ITSEX)



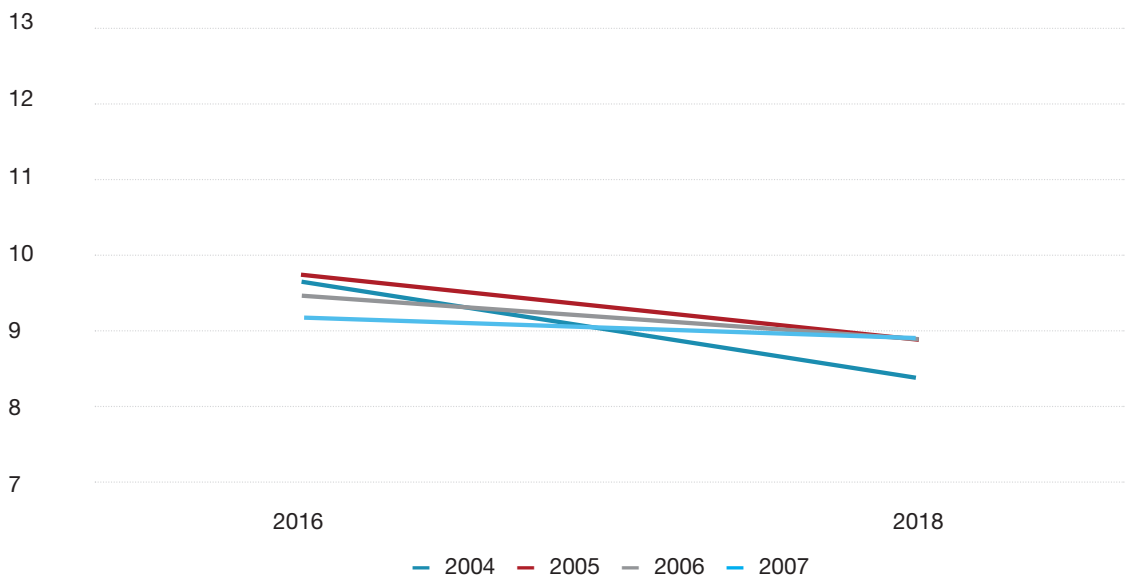
6.3.3 Evolutie volgens geboortjaar

Tabel 62 toont de gemiddelde niveaus in betrokkenheid in 2016 en 2018, opgedeeld volgens geboortjaar. **Figuur 50** toont de gemiddelde evolutie tussen beide afnamemomenten volgens geboortjaar. Omdat het aantal leerlingen geboren in 2004 en 2007 beperkt is, vergelijken we enkel de normaalvorderende leerlingen geboren in 2006 en de leerlingen met één jaar achterstand geboren in 2005. In 2016 hebben de leerlingen geboren in 2006 een significant lager gemiddelde van 9.48 dan de leerlingen geboren in 2005 met een gemiddelde van 9.76. In 2018 hebben de leerlingen geboren in 2006 een gemiddelde van 8.92, wat niet significant verschilt van de leerlingen geboren in 2005 met een gemiddelde van 8.93. Het verschil tussen beide groepen leerlingen is 0.28 in 2016 en 0.10 in 2018. Het verschil tussen de leerlingen geboren in 2006 en de leerlingen geboren in 2005 verdwijnt dus tussen het vierde en zesde leerjaar.

Tabel 62: Betrokkenheid bij lezen (ASBGERL) opgedeeld naar geboortjaar (ITBIRTHY)

Jaar	2004		2005		2006		2007	
	Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)	
2018	8.40	(0.18)	8.93	(0.07)	8.92	(0.04)	8.93	(0.14)
2016	9.68	(0.35)	9.76	(0.09)	9.48	(0.05)	9.18	(0.23)

Figuur 50: Evolutie in betrokkenheid (ASBGERL) opgedeeld naar geboortjaar (ITBIRTHY)



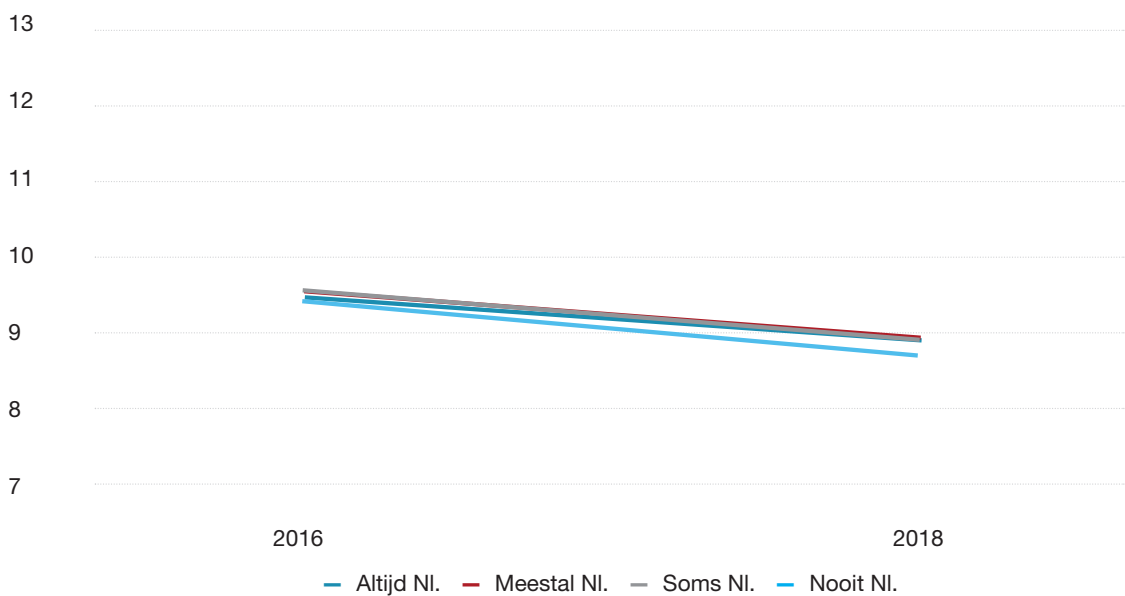
6.3.4 Evolutie volgens thuistaal

Tabel 63 toont de gemiddelde niveaus in betrokkenheid in 2016 en 2018, opgedeeld volgens hoe vaak thuis Nederlands gesproken wordt. **Figuur 51** toont de gemiddelde evolutie tussen beide afnamemomenten volgens hoe vaak thuis Nederlands gesproken wordt. We vergelijken hier de leerlingen die thuis altijd Nederlands spreken en de leerlingen die thuis nooit Nederlands spreken. In 2016 hebben de leerlingen die thuis altijd Nederlands spreken een gemiddelde van 9.49, wat niet significant verschilt van de leerlingen die thuis nooit Nederlands spreken met een gemiddelde van 9.43. In 2018 hebben de leerlingen die thuis altijd Nederlands spreken een gemiddelde van 8.93, wat niet significant verschilt van de leerlingen die thuis nooit Nederlands spreken met een gemiddelde van 8.70. Het verschil tussen beide groepen leerlingen is 0.06 in 2016 en 0.23 in 2018. Er is dus zowel in 2016 als 2018 geen verschil tussen de leerlingen die thuis altijd Nederlands spreken en de leerlingen die thuis nooit Nederlands spreken.

Tabel 63: Betrokkenheid bij lezen (ASBGERL) opgedeeld naar thuistaal (ASBG03)

Jaar	Altijd NI.	Meestal NI.	Soms NI.	Nooit NI.
	Gemiddelde (S.E.)	Gemiddelde (S.E.)	Gemiddelde (S.E.)	Gemiddelde (S.E.)
2018	8.93 (0.05)	8.95 (0.07)	8.93 (0.05)	8.70 (0.15)
2016	9.49 (0.06)	9.55 (0.08)	9.55 (0.07)	9.43 (0.15)

Figuur 51: Evolutie in betrokkenheid (ASBGERL) opgedeeld naar thuistaal (ASBG03)



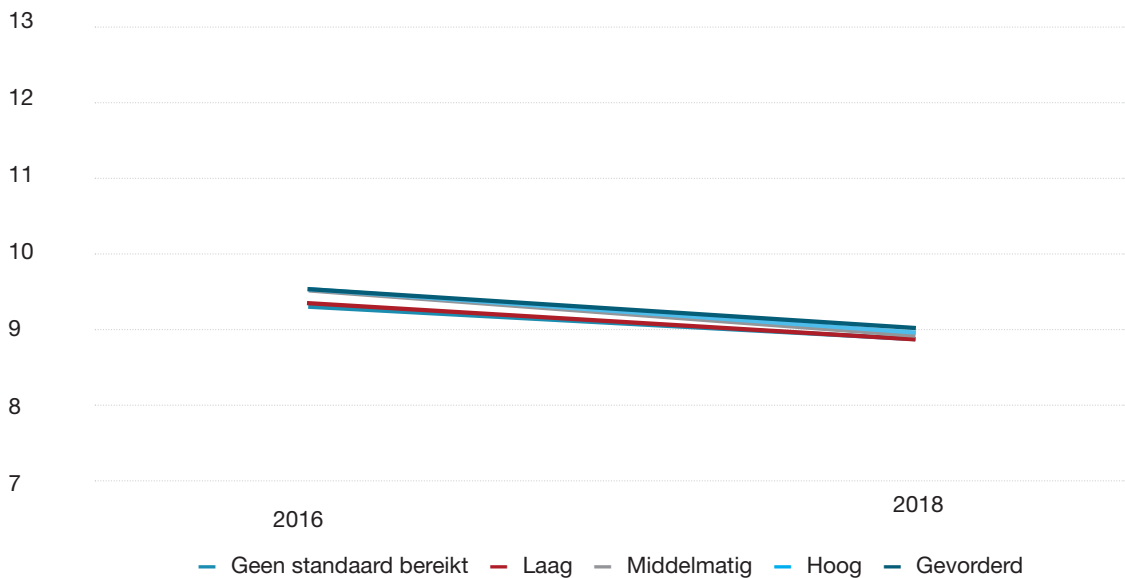
6.3.5 Evolutie volgens benchmarks 2016

Tabel 64 toont de gemiddelde niveaus in betrokkenheid in 2016 en 2018, opgedeeld volgens de *benchmark* of standaard voor begrijpend lezen die de leerlingen behaalden in 2016. **Figuur 52** toont de gemiddelde evolutie tussen beide afnamemomenten, opgedeeld volgens de *benchmark* of standaard voor begrijpend lezen die de leerlingen behaalden in 2016. We vergelijken de leerlingen die tot de lage benchmark behoorden in 2016 met leerlingen die tot de hoge standaard behoorden in 2016. In 2016 hebben de leerlingen die tot de hoge standaard behoorden een significant hoger gemiddelde van 9.55 dan de leerlingen die tot de lage standaard behoorden met een gemiddelde van 9.35. Echter, in 2018 verschillen de leerlingen die tot de hoge standaard behoorden met een gemiddelde van 8.95 niet significant van de leerlingen die tot de lage standaard behoorden met een gemiddelde van 8.85. Het verschil tussen beide groepen leerlingen is 0.20 in 2016 en 0.10 in 2018. Het verschil tussen de leerlingen die de lage standaard behaalden in 2016 en de leerlingen die de hoge standaard behaalden in 2018 is dus stabiel gebleven tussen 2016 en 2018.

Tabel 64: Betrokkenheid bij lezen (ASBGERL) opgedeeld naar benchmark 2016

Jaar	Laag		Middelmatig		Hoog		Gevorderd	
	<400	>400 en <475	>475 en <550	>550 en <625	>625	Gemiddelde (S.E.)	Gemiddelde (S.E.)	
2018	8.89 (0.16)	8.85 (0.06)	8.92 (0.04)	8.95 (0.05)	9.04 (0.12)			
2016	9.29 (0.33)	9.35 (0.06)	9.52 (0.05)	9.55 (0.06)	9.55 (0.14)			

Figuur 52: Evolutie in betrokkenheid (ASBGERL) opgedeeld naar benchmark 2016



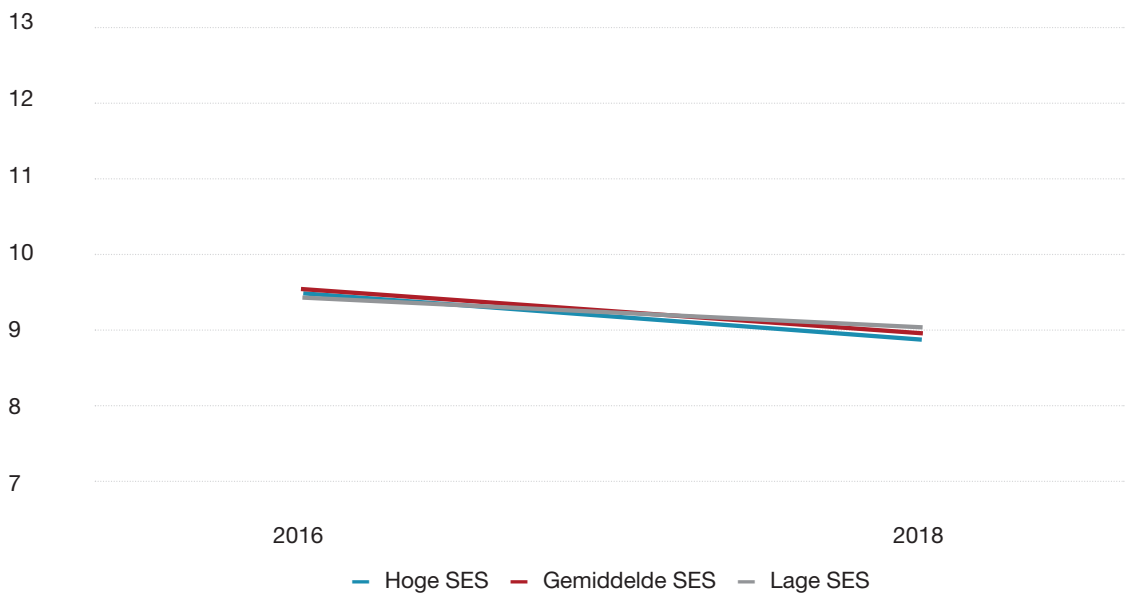
6.3.6 Evolutie volgens socio-economische status

Tabel 65 toont de gemiddelde niveaus in betrokkenheid in 2016 en 2018, opgedeeld volgens de klasse van socio-economische status (*Home Resources for Learning* (HRL)) waartoe de leerlingen behoren. **Figuur 53** toont de gemiddelde evolutie tussen beide afnamemomenten volgens tot welke klasse van socio-economische status de leerlingen behoren. In 2016 was de gemiddelde betrokkenheid niet verschillend voor de drie SES-groepen. Ook in 2018 waren de verschillen in leesbetrokkenheid tussen de drie SES-groepen beperkt. Er is dus zowel in 2016 als in 2018 geen verband tussen SES en betrokkenheid bij lezen.

Tabel 65: Betrokkenheid bij lezen (ASBGERL) opgedeeld naar SES of home resources (ASDGHRL)

Jaar	Hoge SES	Gemiddelde SES	Lage SES
	Gemiddelde (S.E.)	Gemiddelde (S.E.)	Gemiddelde (S.E.)
2018	8.88 (0.05)	8.94 (0.04)	9.03 (0.18)
2016	9.49 (0.07)	9.52 (0.05)	9.39 (0.21)

Figuur 53: Evolutie in betrokkenheid (ASBGERL) opgedeeld naar home resources (ASDGHRL)



6.4 Evolutie van houding van leerlingen tegenover lezen tussen 4^{de} en 6^{de} leerjaar

De houding van leerlingen tegenover lezen werd zowel in de leerlingvragenlijst van 2016 als in de leerlingvragenlijst van 2018 bevestigd met de volgende 8 items en met twee leesactiviteiten buiten school:

Wat vind je van lezen? Geef telkens aan in hoeverre je akkoord gaat met de uitspraak.

1. Ik vind het leuk om met andere mensen te praten over wat ik heb gelezen.
2. Ik vind het leuk als ik een boek als cadeau krijg.
3. Ik vind lezen saai. (-)
4. Ik wou dat ik meer tijd had om te lezen.
5. Ik vind lezen leuk.
6. Ik leer veel door te lezen.
7. Ik vind het leuk om dingen te lezen die me doen nadenken.
8. Ik vind het leuk wanneer een boek me helpt fantaseren over andere werelden.

De leerlingen gaven aan op een 4-puntenschaal in hoeverre ze akkoord gingen: helemaal akkoord / eerder akkoord / eerder niet akkoord / helemaal niet akkoord

Hoe vaak doe je de volgende dingen buiten de school ?

1. Ik lees voor mijn plezier.
2. Ik lees om dingen uit te zoeken waarover ik iets wil leren.

De leerlingen kozen uit vier antwoordmogelijkheden: elke dag of bijna elke dag / één of twee keer per week / één of twee keer per maand / nooit of bijna nooit

Opmerking: (-) duidt op inverse codering

6.4.1 Gemiddelde evolutie

In **Tabel 66** wordt het gemiddelde niveau en de standaarddeviatie in de houding tegenover lezen getoond van dezelfde leerlingen bij PIRLS 2016 en de herhalingsmeting van PIRLS in 2018. Het verschil tussen het niveau in 2016 en het niveau in 2018 is 0.84 (S.E.=0.4), een significant verschil. Het leesplezier is erop achteruit gegaan tussen het vierde en het zesde leerjaar.

Tabel 66: Houding tegenover lezen (ASBGSLR)

Jaar	Steekproef			
	Gemiddelde (S.E.)		Standaarddeviatie (S.E.)	
2018	8.36	(0.05)	1.79	(0.03)
2016	9.21	(0.04)	1.82	(0.03)

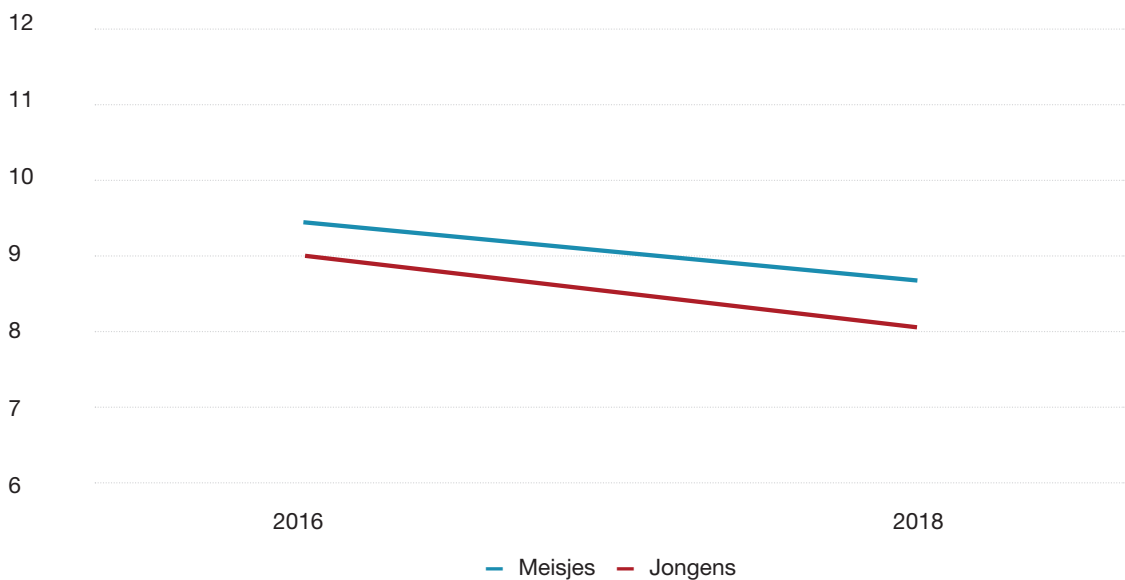
6.4.2 Evolutie volgens geslacht

Tabel 67 toont de gemiddelde niveaus in de houding tegenover lezen in 2016 en 2018, opgedeeld volgens geslacht. **Figuur 54** toont de gemiddelde evolutie tussen beide afnamemomenten voor jongens en meisjes. Uit de gegevens blijkt dat in 2016 Vlaamse meisjes een significant hoger gemiddelde hebben van 9.42 dan de Vlaamse jongens die een gemiddelde hebben van 8.97. De gegevens tonen dat ook in 2018 de Vlaamse meisjes een significant hoger gemiddelde hebben van 8.67 dan de Vlaamse jongens die een gemiddelde hebben van 8.03 behalen. Het verschil tussen jongens en meisjes is 0.45 in 2016 en 0.63 in 2018. Het verschil tussen jongens en meisjes is dus enigszins toegenomen tussen 2016 en 2018. Anders gezegd: de mate van positieve houding tegenover lezen is bij de jongens iets sterker gedaald dan bij de meisjes tussen het 4e en het 6e leerjaar.

Tabel 67: Houding tegenover lezen (ASBGSLR) opgedeeld naar geslacht (ITSEX)

Jaar	Meisjes		Jongens	
	Gemiddelde (S.E.)		Gemiddelde (S.E.)	
2018	8.67	(0.05)	8.03	(0.06)
2016	9.42	(0.05)	8.97	(0.06)

Figuur 54: Houding tegenover lezen (ASBGSLR) opgedeeld naar geslacht (ITSEX)



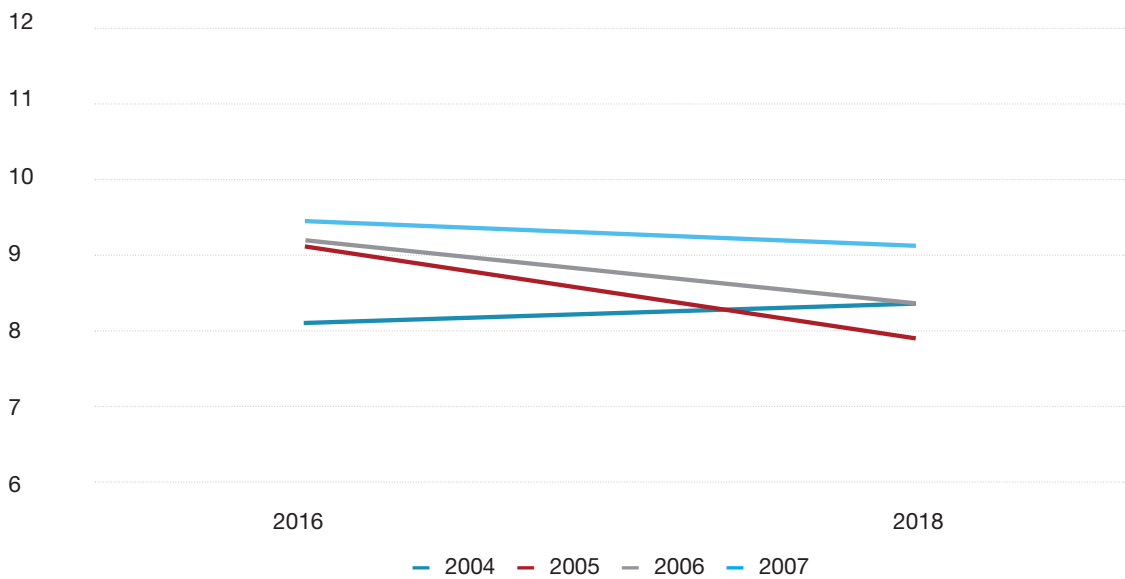
6.4.3 Evolutie volgens geboortejaar

Tabel 68 toont de gemiddelde niveaus in de houding tegenover lezen in 2016 en 2018, opgedeeld volgens geboortejaar. **Figuur 55** toont de gemiddelde evolutie tussen beide afnamemomenten volgens geboortejaar. Omdat het aantal leerlingen geboren in 2004 en 2007 beperkt is, vergelijken we enkel de normaalvorderende leerlingen geboren in 2006 en de leerlingen met één jaar achterstand geboren in 2005. In 2016 verschillen de leerlingen geboren in 2006 met een gemiddelde van 9.21 niet significant van de leerlingen geboren in 2005 met een gemiddelde van 9.14. Echter, in 2018 hebben de leerlingen geboren in 2006 een significant hoger gemiddelde van 8.39 dan de leerlingen geboren in 2005 met een gemiddelde van 7.92. Het verschil tussen beide groepen leerlingen is 0.08 in 2016 en 0.48 in 2018. Het verschil tussen de leerlingen geboren in 2006 en de leerlingen geboren in 2005 is dus toegenomen tussen 2016 en 2018.

Tabel 68: Houding tegenover lezen (ASBGSLR) opgedeeld naar geboortejaar (ITBIRTHY)

Jaar	2004		2005		2006		2007	
	Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)	
2018	8.36	(0.32)	7.92	(0.12)	8.39	(0.05)	9.14	(0.24)
2016	8.12	(0.72)	9.14	(0.10)	9.21	(0.04)	9.48	(0.33)

Figuur 55: Houding tegenover lezen (ASBGSLR) opgedeeld naar geboortejaar (ITBIRTHY)



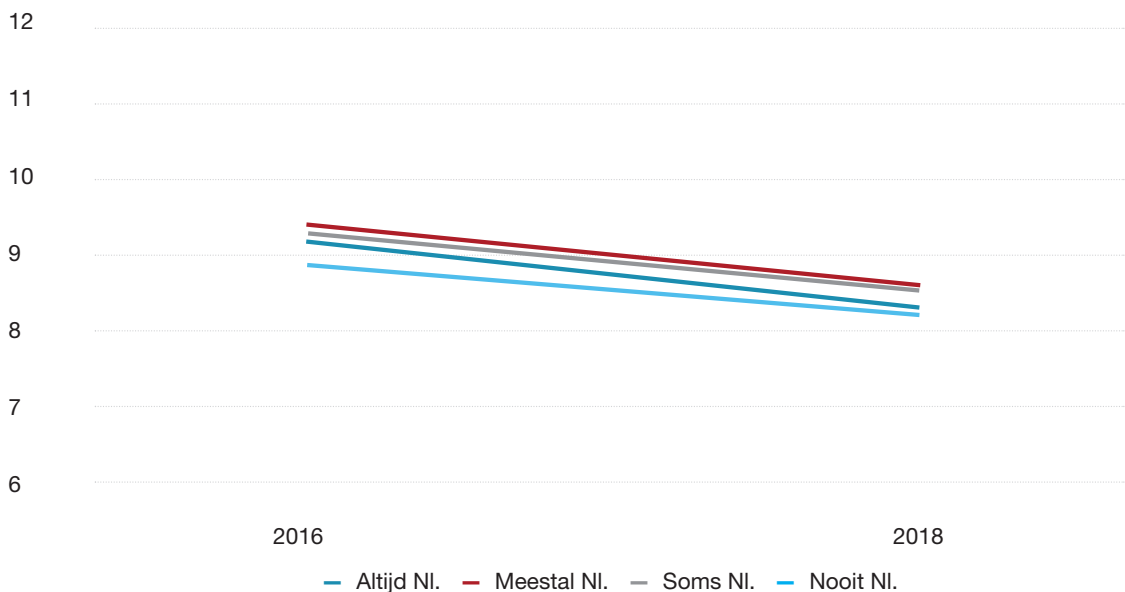
6.4.4 Evolutie volgens thuistaal

Tabel 69 toont de gemiddelde niveaus in de houding tegenover lezen in 2016 en 2018, opgedeeld volgens hoe vaak thuis Nederlands gesproken wordt. **Figuur 56** toont de gemiddelde evolutie tussen beide afnamemomenten volgens hoe vaak thuis Nederlands gesproken wordt. We vergelijken hier de leerlingen die thuis altijd Nederlands spreken en de leerlingen die thuis nooit Nederlands spreken. In 2016 hebben de leerlingen die thuis altijd Nederlands spreken een significant hoger gemiddelde van 9.18 dan de leerlingen die thuis nooit Nederlands spreken met een gemiddelde van 8.87. Echter in 2018 verschillen de leerlingen die thuis altijd Nederlands spreken met een gemiddelde van 8.30 niet significant van de leerlingen die thuis nooit Nederlands spreken met een gemiddelde van 8.50. Het verschil tussen beide groepen leerlingen is 0.30 in 2016 en 0.09 in 2018. Het verschil tussen de leerlingen die thuis altijd Nederlands spreken en de leerlingen die thuis nooit Nederlands spreken is dus afgenomen tussen 2016 en 2018.

Tabel 69: Houding tegenover lezen (ASBGSLR) opgedeeld naar thuistaal (ASBG03)

Jaar	Altijd NI.	Meestal NI.	Soms NI.	Nooit NI.
	Gemiddelde (S.E.)	Gemiddelde (S.E.)	Gemiddelde (S.E.)	Gemiddelde (S.E.)
2018	8.30 (0.05)	8.57 (0.10)	8.50 (0.09)	8.21 (0.18)
2016	9.18 (0.06)	9.41 (0.09)	9.29 (0.07)	8.87 (0.15)

Figuur 56: Houding tegenover lezen (ASBGSLR) opgedeeld naar thuistaal (ASBG03)



6.4.5 Evolutie volgens benchmarks 2016

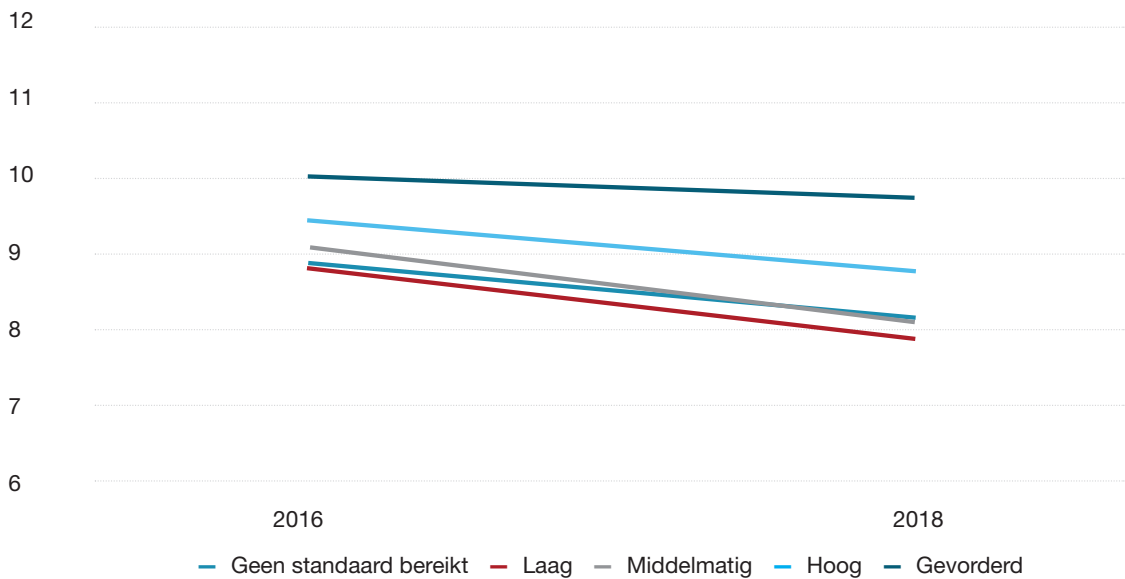
Tabel 70 toont de gemiddelde niveaus in de houding tegenover lezen in 2016 en 2018, opgedeeld volgens de *benchmark* of standaard voor begrijpend lezen die de leerlingen behaalden in 2016. **Figuur 57** toont de gemiddelde evolutie tussen beide afnamemomenten, opgedeeld volgens de *benchmark* of standaard voor begrijpend lezen die de leerlingen behaalden in 2016.

We vergelijken de leerlingen die tot de lage benchmark behoorden in 2016 met leerlingen die tot de hoge standaard behoorden in 2016. In 2016 hebben de leerlingen die tot de hoge standaard behoorden een significant hoger gemiddelde van 9.46 dan de leerlingen die tot de lage standaard behoorden met een gemiddelde van 8.82. Ook in 2018 hebben de leerlingen die tot de hoge standaard behoorden een significant hoger gemiddelde van 8.78 dan de leerlingen die tot de lage standaard behoorden met een gemiddelde van 7.90. Het verschil tussen beide groepen leerlingen is 0.64 in 2016 en 0.88 in 2018. Het verschil tussen de leerlingen die de lage standaard behaalden in 2016 en de leerlingen die de hoge standaard behaalden in 2018 is dus toegenomen tussen 2016 en 2018.

Tabel 70: Houding tegenover lezen (ASBGSLR) opgedeeld naar benchmark 2016

Jaar	Laag		Middelmatig		Hoog		Gevorderd	
	<400	>400 en <475	>475 en <550	>550 en <625	>550 en <625	>625	>625	>625
	Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)		Gemiddelde (S.E.)	
2018	8.19 (0.22)	7.90 (0.08)	8.11 (0.05)	8.78 (0.06)	8.78 (0.06)	9.74 (0.16)	9.74 (0.16)	9.74 (0.16)
2016	8.88 (0.21)	8.82 (0.08)	9.10 (0.06)	9.46 (0.05)	9.46 (0.05)	10.03 (0.15)	10.03 (0.15)	10.03 (0.15)

Figuur 57: Houding tegenover lezen (ASBGSLR) opgedeeld naar benchmark 2016



6.4.6 Evolutie volgens socio-economische status

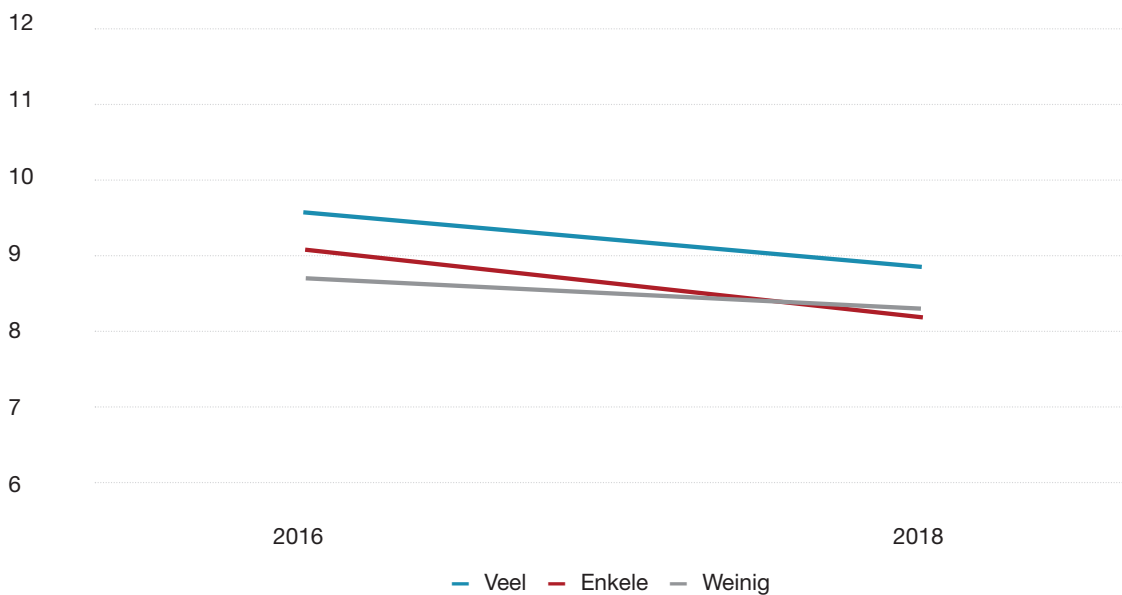
Tabel 71 toont de gemiddelde niveaus in de houding tegenover lezen in 2016 en 2018, opgedeeld volgens de klasse van socio-economische status (*Home Resources for Learning* (HRL)) waartoe de leerlingen behoren. **Figuur 58** toont de gemiddelde evolutie tussen beide afnamemomenten volgens tot welke klasse van socio-economische status de leerlingen behoren.

We vergelijken de leerlingen met een hoge SES met de leerlingen met een gemiddelde SES. In 2016 hebben de leerlingen met een hoge SES een significant hoger gemiddelde van 9.56 dan de leerlingen met een gemiddelde SES met een gemiddelde van 9.08. Ook in 2018 hebben de leerlingen met een hoge SES een significant hoger gemiddelde van 8.81 dan de leerlingen met een gemiddelde SES met een gemiddelde van 8.19. Het verschil tussen beide groepen leerlingen is 0.48 in 2016 en 0.63 in 2018. Het verschil tussen de leerlingen met een hoge SES en de leerlingen met gemiddelde SES is dus enigszins afgenomen.

Tabel 71: Houding tegenover lezen (ASBGSLR) opgedeeld naar SES of home resources (ASDGHRL)

Jaar	Hoge SES	Gemiddelde SES	Lage SES
	Gemiddelde (S.E.)	Gemiddelde (S.E.)	Gemiddelde (S.E.)
2018	8.81 (0.06)	8.19 (0.05)	8.29 (0.23)
2016	9.56 (0.06)	9.08 (0.05)	8.69 (0.21)

Figuur 58: Houding tegenover lezen (ASBGSLR) opgedeeld naar home resources (ASDGHRL)



6.5 Conclusie

In dit hoofdstuk onderzochten we de evolutie van attitudes m.b.t. begrijpend lezen van de Vlaamse leerlingen tussen het vierde leerjaar in 2016 en het zesde leerjaar in 2018. We bekeken evoluties voor (1) zelfzekerheid in leesvaardigheid, (2) betrokkenheid bij lezen en (3) de houding tegenover lezen

De resultaten voor de evoluties duiden erop dat de Vlaamse leerlingen gemiddeld relatief stabiel blijven in hun zelfzekerheid in leesvaardigheid. De leerlingen worden echter wel minder betrokken bij lezen en hebben een minder positieve houding tegenover lezen na twee jaar.

De kleine verschillen tussen jongens en meisjes voor zelfzekerheid en betrokkenheid blijven relatief stabiel tussen het vierde en zesde leerjaar. Voor een positieve houding tegenover lezen zien we dat jongens een grotere daling kennen dan meisjes, terwijl de jongens initieel ook al een minder positieve houding hadden in lezen. Het verschil in de houding tegenover lezen tussen jongens en meisjes wordt dus groter.

Voor geboortjaar zien we dat zowel oudere leerlingen als normaalvorderende leerlingen niet veranderen voor hun niveau van zelfzekerheid. De oudere leerlingen behouden daarmee hun relatief lagere zelfzekerheid. We vinden echter ook dat oudere leerlingen minder dalen in hun betrokkenheid dan normaalvorderende leerlingen. Daarmee verdwijnt hun initieel lagere betrokkenheid. Voor de houding tegenover lezen wordt het verschil tussen beide groepen echter groter, oudere leerlingen hebben een minder positieve houding tegenover lezen.

Voor thuistaal vinden we dat leerlingen die thuis altijd Nederlands spreken en leerlingen die thuis nooit Nederlands spreken dezelfde evolutie kennen voor zelfzekerheid en betrokkenheid. Voor zelfzekerheid was er in het vierde leerjaar ook geen verschil. Zowel in het vierde leerjaar als in het zesde leerjaar zijn leerlingen die thuis altijd Nederlands spreken enigszins meer betrokken, dit verschil blijft stabiel. Leerlingen die thuis altijd Nederlands spreken hadden in het vierde leerjaar een positievere houding tegenover lezen dan leerlingen thuis nooit Nederlands spreken. Dit verschil is echter verdwenen in het zesde leerjaar.

Leerlingen met hogere prestaties voor begrijpend lezen in 2016 waren meer zelfzeker over hun leesvaardigheid, waren meer betrokken bij lezen en hadden een positievere houding tegenover lezen in 2016. Voor zelfzekerheid blijft dit verschil stabiel, terwijl het verschil in betrokkenheid afneemt tegen 2018. Het verschil in de houding tegenover lezen neemt dan weer toe tussen het vierde leerjaar en het zesde leerjaar.

We vinden dat leerlingen met een hogere SES in 2016 meer zelfzeker waren in lezen, meer betrokken waren lezen en een positievere houding hadden tegenover lezen. Voor zelfzekerheid en de houding tegenover lezen is het verschil tussen de verschillende groepen leerlingen stabiel gebleven tussen het vierde en zesde leerjaar. Voor betrokkenheid is het verschil tussen de verschillende groepen leerlingen naargelang SES dan weer afgenomen.

7 Unidimensionaliteit toetsen herhalingsmeting 2018

7.1 Inleiding

>> In dit hoofdstuk behandelen we de onderzoeksvraag of de toetsvragen van PIRLS 2016 en de toetsvragen van het Peilingsonderzoek lezen 2018 dezelfde onderliggende vaardigheid voor begrijpend lezen meten. Dit hoofdstuk start met een duiding van de begrippen unidimensionaliteit en multidimensionaliteit. Vervolgens voeren we verschillende tests uit op basis van IRT om na te gaan of PIRLS 2016 en het Peilingsonderzoek lezen 2018 een of twee aparte onderliggende vaardigheden meten. Als laatste gaan we na hoe de meetschaal van PIRLS 2016 zich verhoudt tot de meetschaal van het Peilingsonderzoek lezen 2018.

7.2 Unidimensionaliteit of multidimensionaliteit?

Bij het opstellen van een toets is het doel doorgaans om tot een eindscore te komen die iets zegt over de vaardigheid van de leerling. De geobserveerde toetsscore is dus een weergave van hoe vaardig een leerling is. De reden dat we een toets moeten afnemen om uitspraken over de vaardigheid van een leerling te kunnen doen, is omdat 'vaardigheid' niet direct observeerbaar is. Je kan de vaardigheid enkel afleiden uit hoe goed of hoe slecht een leerling de vragen van een toets beantwoordt.

Dat de vaardigheid van een leerling kan afgeleid worden op basis van zijn of haar antwoorden op toetsvragen veronderstelt het idee van een zogenaamd meetmodel. Als er beoogd wordt dat een toets één bepaalde vaardigheid meet, dan is elke toetsvraag van die toets een meting van die onderliggende vaardigheid. In dit geval is de toets unidimensioneel, wat wil zeggen dat de antwoorden op elke toetsvraag enkel het gevolg zijn van die onderliggende vaardigheid plus wat onsystematische meetfout (Harrison, 1986; Hattie, 1985; Reise, Moore, & Haviland, 2010). In het unidimensionele geval wordt in het meetmodel dus verondersteld dat er één onderliggende vaardigheid is.

Een toets hoeft echter niet unidimensioneel te zijn als beoogd wordt dat een toets meer dan één vaardigheid meet. Hiervoor zijn er meerdere mogelijkheden. Zo kunnen er verschillende groepen items zijn die verschillende vaardigheden meten, feitelijk is dan elke groep items unidimensioneel op zich, ze maken gewoon toevallig deel uit van één toets. In dit geval zijn de antwoorden op elke toetsvraag enkel het gevolg van de onderliggende vaardigheid die dat groepje toetsvragen beoogt te meten plus wat onsystematische meetfout. Het is echter ook mogelijk dat dezelfde items tegelijkertijd verschillende vaardigheden meten. Dit wil zeggen dat de antwoorden op elke toetsvraag het gevolg mogen zijn van meerdere onderliggende vaardigheden plus wat onsystematische meetfout. In het multidimensionele geval wordt in het meetmodel dus verondersteld dat er meerdere onderliggende vaardigheden zijn.

De afweging of er beoogd wordt dat een toets één onderliggende vaardigheid meet of meerdere onderliggende vaardigheden meet hangt af van het te meten concept. Sommige concepten zijn namelijk inherent multidimensioneel en een reductie naar een unidimensionele toets veronderstelt het verwijderen van bepaalde aspecten van dit concept. Een alternatief is een multidimensioneel concept toch meten alsof het unidimensioneel is. Dit leidt tot het risico dat onderzoeksresultaten verschillend zijn voor verschillende dimensies, maar dat door de verschillende dimensies te combineren dit verschil niet opgemerkt wordt. Een ander risico is dat op een onbewuste wijze aan één dimensie van het multidimensionele concept te veel gewicht gegeven wordt. In andere gevallen is het concept inherent unidimensioneel en de antwoorden van een toets rond dit concept zullen dan ook maar tot één onderliggende vaardigheid te herleiden zijn.

We merkten eerder al op dat de vraag of een concept unidimensioneel of multidimensioneel meestal niet éénduidig te beantwoorden is. Zo kan een vaardigheid vaak op verschillende manieren gedefinieerd worden en kan het resultaat ook verschillen per steekproef. Vaak wordt ook geobserveerd dat toetsvragen van één bepaalde afnamevorm sterker onderling steker samenhangen dan toetsvragen van een andere afnamevorm (vb. meerkeuzevragen versus open vragen), ook al meten ze dezelfde onderliggende vaardigheid. Dit is een vorm van methode-effect. In empirisch onderzoek is de vraag dan ook vaker of de toets voldoende unidimensioneel is.

7.3 Unidimensionaliteit PIRLS 2016 en het Peilingsonderzoek lezen 2018

Voor deze studie werden toetsen opgesteld die zowel uit teksten met bijhorende toetsvragen van PIRLS 2016 als van het Peilingsonderzoek lezen 2018 bestaat. Het toetsmateriaal van het Peilingsonderzoek lezen 2018 werd immers gebruikt als aanvulling op het toetsmateriaal van PIRLS 2016. Er waren immers onvoldoende ‘moeilijke’ toetsvragen in PIRLS 2016 om te kunnen differentiëren tussen sterkere leerlingen in het zesde leerjaar. Doordat we teksten met bijhorende toetsvragen opnamen van twee verschillende studies wordt de veronderstelling gemaakt dat beide studies⁷ hetzelfde meten.

PIRLS 2016 beschrijft begrijpend lezen als de vaardigheid om geschreven taal te begrijpen en te gebruiken, zoals vereist door de maatschappij en/of het individu. Lezers kunnen daarbij betekenis construeren uit teksten van verschillende vormen. Ze lezen om te leren, om deel te nemen aan leesgroepen op school en in het alledaagse leven en voor het plezier (Mullis & Martin, 2015; p. 12). Bij de constructie van toetsen voor PIRLS 2016 is internationaal samengewerkt met experts in begrijpend lezen om de vaardigheid voor begrijpend lezen van een internationaal leerlingenpubliek te meten (Martin, Mullis, & Hooper, 2017; p. 3). De consensus is dat de teksten en bijhorende toetsvragen van PIRLS een vorm van begrijpend lezen meten die relevant is voor alle deelnemende landen en regio's. Bij het vooronderzoek van PIRLS 2016 werd dit ook empirisch getoetst en werd er inderdaad gevonden dat de antwoorden op de toetsvragen tot één onderliggende vaardigheid te herleiden zijn.

Het Peilingsonderzoek lezen 2018 beschrijft begrijpend lezen volgens de Vlaamse eindtermen voor lezen. De eindterm lezen bestaat uit drie verwerkingsniveaus: “begrijpen”, “structureren” en “beoordelen. Deze eindtermen worden weergegeven in [Tabel 72](#).

Tabel 72: eindtermen lezen Vlaams lager onderwijs

De leerlingen kunnen (verwerkingsniveau=beschrijven) de informatie achterhalen in:

- 3.1 voor hen bestemde instructies voor handelingen van gevarieerde aard;
- 3.2 de gegevens in schema's en tabellen ten dienste van het publiek;
- 3.3 voor hen bestemde teksten in tijdschriften.

De leerlingen kunnen (verwerkingsniveau = structureren) de informatie ordenen die voorkomt in:

- 3.4 voor hen bestemde school- en studieteksten en instructies bij schoolopdrachten;
- 3.5 voor hen bestemde verhalen, kinderromans, dialogen, gedichten, kindertijdschriften en jeugencyclopedieën.

De leerlingen kunnen (verwerkingsniveau = beoordelen) op basis van, hetzij de eigen mening, hetzij informatie uit andere bronnen, informatie beoordelen die voorkomt in:

- 3.6 verschillende voor hen bestemde brieven of uitnodigingen;
- 3.7 reclameteksten die rechtstreeks verband houden met hun leefwereld.

Bij de constructie van toetsen van het Peilingsonderzoek lezen 2018 is samengewerkt met Vlaamse experts in begrijpend lezen om de vaardigheid voor begrijpend lezen van het Vlaamse leerlingenpubliek te meten (Steunpunt Toetsontwikkeling en Peilingen, 2014). De consensus is dat de teksten en bijhorende toetsvragen van het Peilingsonderzoek lezen 2018 een vorm van begrijpend lezen meten die het meest relevant is voor de Vlaamse leerlingen. Bij het vooronderzoek van het Peilingsonderzoek lezen 2018 werd dit ook empirisch getoetst en werd er inderdaad gevonden dat de antwoorden op de toetsvragen tot één onderliggende vaardigheid te herleiden zijn.

Hoewel er dus voldoende evidentie is uit voorgaand onderzoek dat beide studies elk één onderliggende vaardigheid voor begrijpend lezen meten, meten ze daarom niet dezelfde vorm van onderliggende vaardigheid. Immers, PIRLS richt zich op een internationaal leerlingenpubliek in het vierde leerjaar terwijl het Peilingsonderzoek lezen 2018 zich richt op het Vlaamse leerlingenpubliek in het zesde leerjaar. Beide studies hebben volledig los van elkaar toetsen begrijpend lezen ontwikkeld. Er is tot nu toe nooit onderzocht of beide studies ook dezelfde vorm van begrijpend lezen meten.

⁷ We merken op dat de tekst ‘Chinese Horoscoop’ van een derde bron afkomstig is, namelijk het Centrum voor Schoolfeedback. Aangezien deze tekst slechts vijf items bevat was het niet mogelijk om deze als aparte dimensie te onderzoeken. De vergelijking tussen de items van PIRLS 2016 en de items van het Peilingsonderzoek lezen 2018 is echter zowel uitgevoerd met de tekst ‘Chinese Horoscoop’ als zonder de tekst ‘Chinese Horoscoop’. Dit leidde echter niet tot merkbaar andere resultaten. In dit hoofdstuk wordt ‘Chinese Horoscoop’ dan ook beschouwd alsof het één van de teksten van het Peilingsonderzoek is.

7.4 Gelijkenissen en verschillen tussen PIRLS en peiling volgens Vlaamse experts

In de periode februari 2019 - april 2019 interviewde studente Heleen Hoebrechts (Master Educatieve Studies, KU Leuven) 10 Vlaamse experts over de gelijkenissen en verschillen tussen de PIRLS-toetsen en de peilingstoetsen lezen. Deze 10 sleutelinformanten waren pedagogisch begeleider, onderwijsinspecteur, onderzoeker of lerarenopleider. In een semigestructureerd interview werd gepolst naar hun mening over de PIRLS-toetsen en peilingstoetsen. De toetsboekjes van PIRLS, de eindtermen lezen en voorbeelditems van de peiling lezen werden ter plaatse getoond zodat de respondenten het toetsmateriaal konden inkijken.

Daarnaast bezorgde het PIRLS-team de 16 toetsboekjes van PIRLS2016 aan één pedagogisch begeleider die een uitgebreide analyse maakte van de waarde van PIRLS voor de Vlaamse onderwijscontext.

7.4.1 Lengte en moeilijkheidsgraad van de teksten

De experts gaven allemaal aan dat de PIRLS-teksten behoorlijk lang zijn. De teksten in de peilingstoets zijn beperkter in omvang. Er wordt aangegeven dat die lange PIRLS-teksten voor de Vlaamse leerlingen behoorlijk uitdagend zijn, en dat ze concentratie vereisen. Dit soort teksten en vragen zijn de Vlaamse leerlingen niet gewoon. De PIRLS-teksten lijken niet op het aanbod in de huidige leesmethodes. Het lezen van lange teksten wordt weinig geoefend. Sommige respondenten noemen de PIRLS-teksten 'demotiverend' of ze vrezen dat leerlingen ontmoedigd worden door de teksten. De peilingsteksten daarentegen sluiten sterk aan bij wat de Vlaamse leerlingen gewoon zijn.

De meerderheid van de respondenten ziet de lengte van de PIRLS-teksten als één van de grootste verschillen met de peilingstoetsen. De peilingstoetsen werden omschreven als korter met minder items in vergelijking met PIRLS.

Ook het niveau van de teksten en vragen in PIRLS wordt als 'pittig' en 'uitdagend' omschreven voor de leerlingen van het vierde leerjaar. De woordenschat is niet altijd voor de hand liggend.

Eén expert geeft aan dat de PIRLS-teksten opvallend korte zinnen bevatten en vraagt zich af of de vertaling vanuit het Engels daarvoor verantwoordelijk zou kunnen zijn. Dit kan effecten hebben omdat korte zinnen taaldenkwaties wegwerken en de tekst daardoor mogelijk moeilijker wordt voor de leerling.

Eén respondent zegt: *"Als je me de twee teksten (van PIRLS) had gegeven en me had gevraagd voor welke leeftijdscategorie zou je deze hanteren, ik vermoed dat mijn intuïtie mij had geleid naar het zesde leerjaar."*

7.4.2 Tekstsoorten

Wat opvalt is dat er in PIRLS maar twee tekstsoorten aan bod komen (verhalende tekst en informatieve tekst), terwijl er in de peiling een grotere diversiteit aan tekstsoorten is aangezien die tekstsoorten ook vermeld worden in de eindtermen. Het gaat in de peiling om schema's, tabellen, teksten in tijdschriften, school- en studieteksten, instructies bij schoolopdrachten, verhalen, brieven, uitnodigingen, en reclameteksten.

Een aantal respondenten storen zich aan de 'oubollige' teksten in PIRLS en soms ook ouderwetse tekeningen. Meermaals wordt vermeld dat de peilingsteksten meer hedendaags en aantrekkelijker zijn. De peilingstoetsen sluiten volgens een aantal geïnterviewden sterker aan bij de leefwereld van de leerlingen. Het volgende citaat illustreert dit: *"Eigenlijk zijn deze teksten (PIRLS) quasi tijdloos, maar dat weet ik nu niet of dat een compliment is."*

7.4.3 Vraagsoorten

Qua soorten vragen is er in PIRLS meer diversiteit dan in de peilingstoetsen. In PIRLS zijn er zowel meerkeuzevragen, korte open vragen en lange open vragen. In PIRLS worden redelijk veel waarom-vragen gesteld. De peilingitems zijn haast allemaal meerkeuzevragen. Er zijn ook enkele vragen in de peiling waarbij de leerling vier tekeningen krijgt die verbonden moeten worden met vier antwoorden. Er zijn ook een paar vragen waarbij leerlingen vier elementen in de juiste volgorde moeten zetten. Er worden geen open vragen gesteld in de peilingstoets, met uitzondering van een vraag naar één woord of naar een tijdstip dat kan gevonden worden in een reclametekst.

De meeste experten geven aan dat de vraagstelling in de peiling en in PIRLS redelijk gelijkaardig is. Maar één expert vermeldt dat de peilingsitems beperkt blijven tot het beschrijvende/begrijpende niveau. Er missen doordenkvragen, interpretatievragen en beoordelingsvragen. Deze komen wel aan bod binnen PIRLS. Een aantal experten vermoeden dus dat de PIRLS-vragen vaker een beroep doen op leesvaardigheden van een hogere orde.

Eén expert zegt: 'PIRLS gaat niet aan de slag met 'kenmerken van de teksten' en hanteert ze enkel als 'bron van informatie'. Daarmee wordt bedoeld dat er in PIRLS zelden vragen worden gesteld over de bedoeling van de schrijver, tekstsoorten, tekstkenmerken, de relatie met de eigen voorkennis. Volgens deze respondent konden er nog bijkomende vragen worden gesteld, bijvoorbeeld: 'Welke soort tekst is dit?' of 'Hoe weet je dat dit geen reclametekst is maar een informatieve tekst?' In het Vlaamse onderwijs worden teksten vaak gekaderd in een communicatieschema waarbij de auteur een bepaald doel heeft en de lezer eveneens een bepaald doel heeft. In de PIRLS-toetsen wordt er vooral gevraagd naar de inhoudelijke betekenis van de tekst en minder naar het doel van de tekst. Maar ook de peilingsvragen polsen niet naar tekstsoorten of tekstkenmerken.

7.4.4 Samenhang tussen PIRLS en peiling

In het interview werd de vraag gesteld: 'Denkt u dat een leerling die hoog scoort op de PIRLS-toets automatisch hoog zal scoren op de peilingstoets en omgekeerd?' . Dit sluit aan bij de vraag naar unidimensionaliteit. Drie respondenten menen dat er een hoge correlatie zal zijn. Vijf respondenten geven aan dat ze vermoeden dat een leerling die hoog scoort voor PIRLS, ook hoog zou scoren voor de peiling maar dat het omgekeerde niet vanzelfsprekend is. Een hoge score voor de peiling betekent volgens hen niet automatisch een hoge score voor PIRLS. Eén respondent zegt: "een zwakke lezer op PIRLS zou misschien weleens beter kunnen uitkomen bij de peilingonderzoeken. Omdat de peiling wat diverser meet en ook die zwakke lezer eigenlijk wat meer kansen geeft om toch te tonen wat die kan, minder ontmoedigd te worden door de tekst, minder ontmoedigd worden om ook door het groot aantal vragen bij die tekst, steeds weer terug te moeten gaan naar die tekst en op een bepaald moment te zeggen ik geef het op." Volgens twee respondenten zou er helemaal geen correlatie zijn tussen PIRLS en de peiling. De toetsen zijn te verschillend en hebben een andere focus waardoor het niet noodzakelijk betekent dat het goed of slecht scoren op het ene toets automatisch resulteert in het goed of slecht scoren op de andere toets.

7.5 Onderzoeksvragen en analyseplan unidimensionaliteit

PIRLS 2016 en het Peilingsonderzoek lezen 2018 meten niet noodzakelijk dezelfde vorm van begrijpend lezen. In de herhalingsmeting van PIRLS in 2018 werd echter verondersteld dat de teksten met bijhorende toetsvragen zowel van PIRLS 2016 en het Peilingsonderzoek lezen 2018 dezelfde vorm van begrijpend lezen meten. Immers, in **Hoofdstuk 3** werden de IRT-analyses uitgevoerd in de veronderstelling dat de antwoorden van de leerlingen op toetsvragen te wijten is aan één onderliggende vaardigheid en wat onsystematische meetfout. Specifiek werd er gebruik gemaakt van de itemparameters van het PIRLS 2016, in de veronderstelling dat de toetsvragen van het Peilingsonderzoek lezen 2018 hier ook op geplaatst kunnen worden. Als aan deze veronderstelling niet voldaan is, dan zijn de geschatte vaardigheden van de herhalingsmeting vertekend.

De onderzoeksvraag van dit hoofdstuk is of de toetsen van PIRLS 2016 en de toetsen van het Peilingsonderzoek lezen 2018 dezelfde onderliggende vaardigheid voor begrijpend lezen meten. Aangezien de onderzoeksvraag betrekking heeft op 'wat er gemeten wordt' beschouwen we dit als een vraag naar de validiteit van onze toetsen. We onderscheiden daarbij in de eerste plaats de constructvaliditeit, tonen de antwoorden van leerlingen een patroon dat te herleiden is naar één onderliggende vaardigheid. In de tweede plaats kijken we naar de predictieve validiteit, voorspellen de scores op de toetsvragen van PIRLS enerzijds hetzelfde als de scores op de toetsvragen van het Peilingsonderzoek lezen 2018 anderzijds.

Om dit te onderzoeken gebruiken we de toetsgegevens van leerlingen die deelnamen aan de herhalingsmeting van PIRLS in 2018. Zoals vermeld in **Hoofdstuk 1** bestonden zes van de tien toetsboekjes uit teksten met bijhorende toetsvragen van zowel PIRLS 2016 als het Peilingsonderzoek lezen 2018. 2850 leerlingen vulden deze zes toetsboekjes in. Het zijn dan ook de antwoorden van deze leerlingen die we gebruiken om te testen of er sprake is van één onderliggende vaardigheid in de herhalingsmeting.

Er is géén eenduidigheid in de literatuur hoe onderzocht moet worden of een toets één (unidimensionaliteit) of meer onderliggende vaardigheden (multidimensionaliteit) meet. Bijgevolg benaderen we deze onderzoeksvraag met twee methoden: exploratieve factoranalyse en IRT-modellen.

7.5.1 Exploratieve factoranalyses

Exploratieve factoranalyse is een methode die dient om van een set variabelen de onderliggende structuur te vinden. Op basis van de set variabelen worden er in een exploratieve factoranalyse namelijk factoren geschat die de samenhang tussen de variabelen verklaren. Hoe goed een factor de samenhang tussen variabelen beschrijft wordt uitgedrukt in hoeveel variantie in variabelen door de factor verklaard wordt. De eerste geëxtraheerde factor verklaart steeds de meeste variantie in de variabelen, de tweede geëxtraheerde verklaart steeds het tweede meeste de variantie in de variabelen, en dit gaat door tot het volledige aantal factoren geschat zijn. Er worden evenveel factoren geschat als er variabelen zijn. De samenhang tussen de variabelen wordt hierdoor meestal al goed beschreven door de eerste factoren. Hoeveel factoren er nodig zijn om een set van variabelen te beschrijven is echter afhankelijk van wat de exploratieve factoranalyse oplevert en welke inhoudelijke overwegingen gemaakt worden.

Exploratieve factoranalyse om unidimensionaliteit te onderzoeken gebeurt doorgaans door na te gaan hoeveel variantie in de toetsvragen verklaard wordt de eerste geëxtraheerde factor versus de tweede geëxtraheerde factor. De consensus is dat de eerste geëxtraheerde factor minstens driemaal of viermaal de variantie in de toetsvragen moet verklaren versus de tweede geëxtraheerde factor. Slechts dan kan een toets als unidimensioneel beschouwd worden. De drempelwaarden zijn echter eerder arbitrair bepaald, er is geen onderzoek dat deze waarden ondersteunt. Het gaat om een vuistregel. Logischerwijs is het echter wel zo dat hoe meer de eerste factor kan verklaren, hoe beter. We gebruiken daarom de exploratieve factoranalyses om te onderzoeken in welke mate één onderliggende factor de antwoorden op de toetsvragen verklaart. Dit doen we zowel voor de zes toetsboekjes die zowel toetsvragen van PIRLS als het Peilingsonderzoek lezen 2018 bevatten, als de vier toetsboekjes die enkel toetsvragen van PIRLS bevatten. We vergelijken de resultaten echter ook met de oorspronkelijke 16 toetsboekjes van PIRLS 2016 en de vier toetsboekjes van het Peilingsonderzoek lezen 2018. We krijgen zo een indruk van de verschillende mate waarin de toetsen van PIRLS 2016, het Peilingsonderzoek lezen 2018 en deze herhalingsmeting aan de voorwaarde van unidimensionaliteit voldoen.

Exploratieve factoranalyses zijn echter beperkt in het beantwoorden van de vraag of een toets een undimensionele vaardigheid meet. Zoals reeds vermeld is er geen duidelijke drempelwaarde aan de hand waarvan unidimensionaliteit vastgesteld kan worden. Een ander nadeel van deze methode is dat de literatuur over exploratieve factoranalyse hoofdzakelijk gericht is op het ontleden van de onderliggende structuur van continue indicatoren. Onze indicatoren zijn echter categorisch. Er is dan ook weinig geweten over hoe goed exploratieve factoranalyses zijn in het vaststellen van de onderliggende structuur van categorische indicatoren. Een ander nadeel is dat exploratieve analyses 'blind' werken, ze weten niet welke specifieke meetmodellen we expliciet willen vergelijken. Deze nadelen worden enigszins gereduceerd door het vergelijken van IRT-modellen.

We gebruikten Mplus 8.0 voor het uitvoeren van de exploratieve factoranalyses met orthogonale *geomin* rotatie. In Appendix 4 wordt het inputbestand van toetsboekje 1 van de herhalingsmeting PIRLS in 2018 weergegeven. De exploratieve factoranalyses zijn uitgevoerd voor elk van de tien toetsboekjes. Er is namelijk geen consensus in de literatuur over hoe de resultaten van exploratieve factoranalyses te interpreteren wanneer er ontbrekende gegevens zijn. Ook de 16 toetsboekjes van PIRLS 2016 en de vier toetsboekjes van het Peilingsonderzoek lezen 2018 zijn steeds apart onderzocht.

7.5.2 Vergelijking 1-factor, 2-factor en bifactor IRT-modellen

Door verschillende IRT-modellen te vergelijken kan onderzocht worden welke van deze IRT-modellen het best de onderliggende structuur van een set variabelen beschrijft. Binnen een IRT-model worden namelijk de antwoorden van leerlingen verklaard aan de hand van een op voorhand bepaalde structuur van onderliggende factoren.

We vergelijken drie IRT-modellen. In het eerste model wordt verondersteld dat één onderliggende vaardigheid de antwoorden van leerlingen op de toetsvragen verklaart. De antwoorden op de toetsvragen van PIRLS 2016 en de antwoorden op toetsvragen van het Peilingsonderzoek lezen 2018 worden dus door dezelfde factor verklaard. Dit is het 1-factor IRT-model. In het tweede model wordt verondersteld dat twee onderliggende vaardigheden de antwoorden van leerlingen op de toetsvragen verklaren. De antwoorden op de toetsvragen van PIRLS 2016 hebben hun eigen factor en de antwoorden op de toetsvragen van het Peilingsonderzoek lezen 2018 hebben hun eigen factor. De mate waarin de twee factoren samenhangen in het IRT-model wordt vrij geschat. Dit is het 2-factor IRT-model. Het derde model is hetzelfde als het 1-factor IRT-model, waarbij één factor alle antwoorden verklaart. Nu zijn er echter twee 'storingsfactoren' in het model, één storingsfactor voor de toetsvragen van PIRLS 2016 en één storingsfactor voor de toetsvragen van het Peilingsonderzoek lezen 2018. Een storingsfactor beschrijft, bovenop de ene factor voor alle toetsantwoorden, hoe sterk de toetsantwoorden verklaard worden door tot eenzelfde tekst te behoren.

Deze drie modellen worden vergeleken aan de hand van hoe goed ze de antwoorden verklaren. Dit kunnen we afleiden aan de hand van de factorladingen, deze geven weer goed een toetsvraag beschreven wordt door een onderliggende factor. We bekijken ook de verklaarde variantie van elke toetsvraag, deze geeft de mate weer waarin de antwoorden op een toetsvraag verklaard wordt door het gekozen model.

Hoe de verschillende IRT-modellen de vergelijking doorstaan wordt bepaald aan de hand van zogenaamde fit-indices, indicatoren van hoe goed een model bij de toetsgegevens past. Fitindices zijn indicatoren van hoe goed het gekozen IRT-model bij de vastgestelde gegevens past. Dit wordt geëvalueerd aan de hand van hoe goed het model de samenhang tussen de variabelen voorspelt. Afhankelijk van de fitindex wordt de complexiteit van het model (uitgedrukt in het aantal parameters) strenger of minder streng bestraft. Een fitindex kan zowel absoluut als relatief zijn. De waarde van een absolute fitindex kan op zich geïnterpreteerd worden, de waarde geeft dus op zich weer hoe goed het model bij de gegevens past. De waarde van een relatieve fitindex moet daarentegen altijd vergeleken worden tussen verschillende modellen, de waarde geeft dus slechts aan of een model beter of slechter past bij de gegevens dan een ander model. We bekijken de volgende fitindices:

- **RMSEA:** De *Root Mean Square Error of Approximation* (RMSEA) is gebaseerd op de χ^2 -statistiek⁸ en gaat na hoe goed het model past in de populatie. Deze index is ongevoelig voor steekproefgrootte en corrigeert voor modelcomplexiteit. Optimaal is dat deze een waarde van 0 heeft, maar dit is in de praktijk niet haalbaar. Hu & Bentler (1999) stellen dat een passend model waarde van 0.06 of lager moet hebben.
- **SRMSR:** De *Standardized Root Mean Square Residual* (SRMSR) kan geïnterpreteerd worden als het gemiddelde verschil tussen de correlaties tussen de verschillende items van een test en de correlaties tussen de items zoals het toegepaste model voorspelt. Optimaal is dat deze een waarde van 0 heeft, de correlaties worden dan volledig correct voorspeld. Een waarde van 0 is echter in de praktijk niet haalbaar. Hu & Bentler (1999) stellen dat een passend model waarde van 0.08 of lager moet hebben.
- **CFI:** De *Comparative Fit Index* (CFI) wordt berekend aan de hand van de fit van het toegepaste model en de fit van het nulmodel volgens de χ^2 -statistiek. Daarbij wordt nagegaan in welke mate het toegepaste model de fit verbetert, gegeven de maximale verbetering die mogelijk is. De waarde van deze index ligt tussen 0 en 1. Hoe hoger de waarde van de CFI, hoe beter de fit (Brown, 2006). Hu & Bentler (1999) stellen dat een passend model een waarde van 0.95 of hoger moet hebben. We merken echter op dat deze drempelwaarde bepaald is voor continue indicatoren, terwijl onze indicatoren categorisch zijn.
- **TLI:** De *Tucker–Lewis index* (TLI) kan beschouwd worden als een extensie van de CFI. Met deze index wordt modelcomplexiteit echter alsnog minder wenselijk beschouwd dan bij CFI. Algemeen genomen zal deze index nog simpelere modellen verkiezen dan CFI. De waarde van deze index ligt tussen 0 en 1, waarbij een hogere waarde een betere fit aangeeft (Brown, 2006). Ook voor deze index stellen Hu & Bentler (1999) een waarde van 0.95 voorop om het model als passend te beschouwen. We merken echter op dat deze drempelwaarde bepaald is voor continue indicatoren, terwijl onze indicatoren categorisch zijn.
- **AIC:** Het *Akaike Information Criterion* (AIC) is een index die gebruikt wordt bij het vergelijken van verschillende modellen, toegepast op eenzelfde set gegevens. De AIC maakt gebruik van de log-likelihood waarde van de verschillende modellen, waarbij een correctie wordt uitgevoerd voor de steekproefgrootte. Op deze manier worden complexere modellen met een grote steekproef niet te snel aanvaard. Een complexer model dient een lagere waarde in AIC te hebben dan een minder complex model. Als vuistregel geldt dat als het complexe model twee eenheden of meer AIC lager is dan het minder complexe model, het complexe model verkozen wordt (Kang & Cohen, 2007). Het verschil met BIC is dat de AIC geen correctie heeft voor modelcomplexiteit. De AIC is hierdoor meer gericht op het vinden van een model dat de oorspronkelijke variabelen goed reproduceert.
- **BIC:** Het *Bayesian Information Criterion* (BIC) is een index die gebruikt wordt bij het vergelijken van verschillende modellen, toegepast op eenzelfde set gegevens. De BIC maakt gebruik van de log-likelihood waarde van de verschillende modellen, waarbij een correctie wordt uitgevoerd voor de steekproefgrootte en de modelcomplexiteit. Op deze manier worden complexere modellen niet te snel aanvaard. Een complexer model dient een lagere waarde in BIC te hebben dan een minder complex model. Als vuistregel geldt dat als het complexe model twee eenheden of meer BIC lager is dan het minder complexe model, het complexe model verkozen wordt (Kang & Cohen, 2007). Het verschil met AIC is dat de BIC een correctie heeft voor modelcomplexiteit. De BIC is hierdoor meer gericht op het vinden van de meest eenvoudige structuur.

8 χ^2 is een klassieke statistiek die de mate van afwijking beschrijft van de geobserveerde waarden tegenover de voorspelde waarden van een model dat geschat wordt met maximum-likelihood. Deze statistiek wordt echter als te streng beschouwd wegens een sterke gevoeligheid voor steekproefgrootte. We gebruiken deze statistiek dus niet om model-fit te evalueren.

De fitindices zullen echter waarschijnlijk geen sluitend antwoord bieden op de vraag of de toetsvragen van PIRLS en het Peilingsonderzoek te kunnen verklaren zijn voor één onderliggende vaardigheid. Verder geven ze niet weer hoe sterk de toetsvragen van PIRLS met de toetsvragen van het Peilingsonderzoek samenhangen. Verder beschrijven de fitindices enkel de mate waarin de constructvaliditeit van de verschillende modellen te behouden of te verwerpen zijn. De fitindices zeggen niets over de predictieve validiteit. Daarom geven we ook de volgende zaken weer.

- **Verklaarde variantie toetsvragen:** We geven weer hoeveel variantie in de toetsvragen verklaard wordt door de verschillende modellen. Hoe meer variantie in de toetsvragen verklaard wordt door een model, hoe beter dat model de antwoorden op de toetsvragen voorspelt. Een nadeel van deze methode is dat ze steeds het meest complexe model zal verkiezen. We krijgen echter wel een inschatting van de hoeveelheid variantie we niet verklaren door bijvoorbeeld het 1-factor IRT-model te gebruiken in de plaats van het 2-factor IRT-model.
- **Correlatie tussen factoren in 2-factor IRT-model.** We geven de latente correlatie weer tussen de factor van de toetsvragen van PIRLS en de factor van de toetsvragen van het Peilingsonderzoek. Bij een perfecte samenhang is de correlatie 1 en bij geen samenhang is de waarde 0. Om de samenhang tussen beide factoren te interpreteren, wordt deze ook vergeleken met de toetsboekjes waarin enkel toetsvragen van PIRLS zijn opgenomen. Hierin wordt dan steeds een aparte factor geschat voor de toetsvragen van de eerste PIRLS-tekst en een aparte factor voor de toetsvragen van de tweede PIRLS-tekst. Deze samenhang wordt ook onderzocht in de 16 toetsboekjes van PIRLS 2016.
- **Correlatie factoren in 2-factor IRT-model met andere variabelen.** We onderzoeken of de vaardigheid in begrijpend lezen volgens PIRLS en de vaardigheid in begrijpend lezen volgens het Peilingsonderzoek hetzelfde voorspellen. Hiervoor worden in het 2-factor IRT-model de factoren vrij gecorreleerd met een derde variabele. We bekijken zowel de samenhang met de prestaties van leerlingen zoals ze geschat zijn in 2016 en de SES van leerlingen (ASBGRHL). Vervolgens bekijken we ook of het gemiddelde van jongens en meisjes (ITSEX) hetzelfde is voor beide factoren. We doen dit ook door een vergelijking te maken tussen leerlingen die thuis altijd Nederlands spreken en leerlingen die thuis soms Nederlands spreken (ASBG03).

We gebruikten MIRT 1.30 (Chalmers, 2012) in R 3.6.0 (R Core Team, 2018) en Mplus 8.0 (Muthén & Muthén, 2015). In Appendix 5 wordt de syntax weergegeven van de analyses in MIRT van toetsboekje 1. In Appendix 6 wordt de syntax weergegeven van de analyses in Mplus van toetsboekje 1. We merken op dat MIRT voor de meeste analyses gebruikt werd. Mplus werd enkel gebruikt om de samenhang tussen de twee factoren in het 2-factor IRT-model met andere variabelen te beschrijven. MIRT en Mplus kwamen tot exact dezelfde schatting voor het 2-factor IRT-model.

7.6 Resultaten

7.6.1 Exploratieve factoranalyses

In **Tabel 73** staan de resultaten van de exploratieve factoranalyse van de tien toetsboekjes van de herhalingsmeting van PIRLS in 2018. Deze tabel toont de eerste vier eigenwaardes van de exploratieve factoranalyses. De eigenwaarde van een factor is in deze context gelijk aan de totale variantie die de factor verklaart in de antwoorden op de toetsvragen. De eigenwaardes in **Tabel 73** beschrijven dus hoeveel variantie verklaard wordt door respectievelijk de eerste factor, tweede factor, derde factor en vierde factor. Er zijn steeds evenveel factoren of eigenwaardes als er items per toetsboekje zijn. In de laatste kolom wordt de verhouding tussen de eerste en tweede eigenwaarde weergegeven. In **Tabel 74** worden de scree plots getoond van de tien toetsboekjes van de herhalingsmeting van PIRLS in 2018. Een scree plot geeft de eigenwaarde van iedere factor weer, waarbij de factoren ook gerangschikt zijn volgens de grootte van hun eigenwaarde. Hoe groter de eigenwaarde, hoe hoger de factor staat op de grafiek.

Voor de zes toetsboekjes die toetsvragen van zowel PIRLS als het Peilingsonderzoek bevatten is de gemiddelde verhouding tussen de eerste eigenwaarde en de tweede eigenwaarde 3.19. De kleinste verhouding is 2.05 en de grootste verhouding is 3.93. Voor de vier toetsboekjes die enkel twee PIRLS-teksten bevatten is de gemiddelde verhouding tussen de eerste eigenwaarde en de tweede eigenwaarde 4.46. De kleinste verhouding is 3.96 en de grootste verhouding is 4.98.

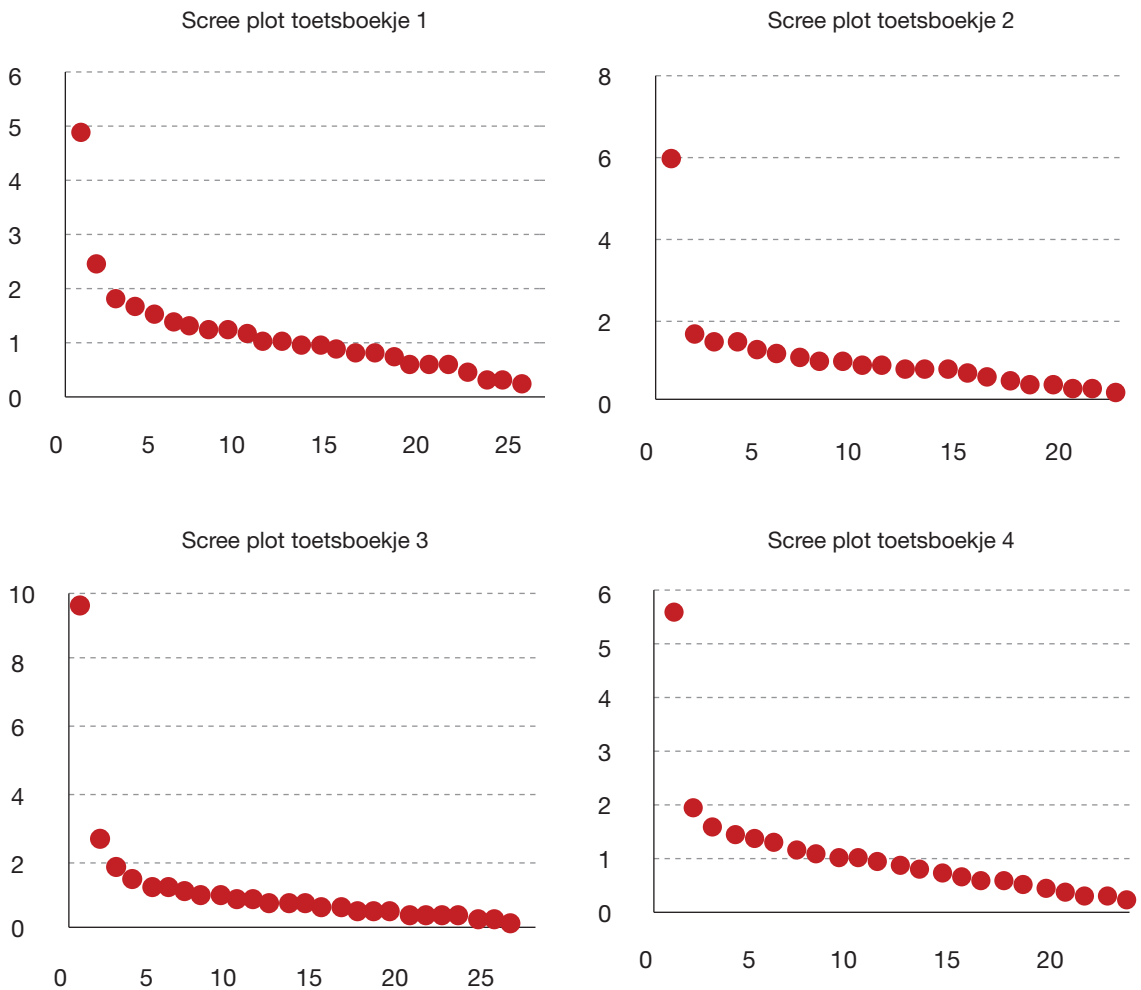
Ter vergelijking vermelden we ook de gemiddelden van de oorspronkelijke gegevens van PIRLS 2016 en het Peilingsonderzoek lezen 2018. Voor de 16 toetsboekjes van PIRLS 2016 is de gemiddelde verhouding tussen de eerste eigenwaarde op de tweede eigenwaarde 4.02, met een minimum van 2.07 en een maximum van 5.57. Voor de vier toetsboekjes van het Peilingsonderzoek lezen 2018 is de gemiddelde verhouding tussen de eerste eigenwaarde op de tweede eigenwaarde 3.89, met een minimum van 3.00 en een maximum van 4.57.

Tabel 73: Vier eerste eigenwaardes van exploratieve factoranalyses toetsboekjes met verhouding eerste en tweede eigenwaarde

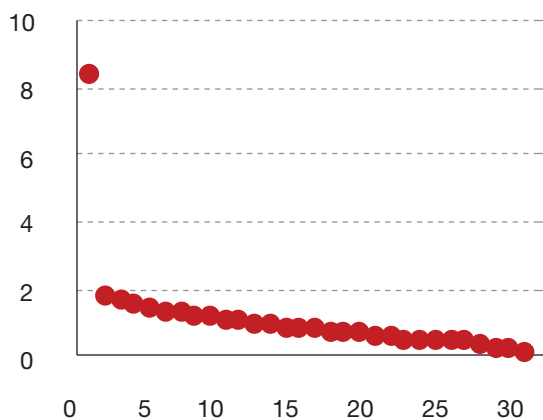
Toets-boekje	Inhoud	Eigenwaarde 1	Eigenwaarde 2	Eigenwaarde 3	Eigenwaarde 4	EW1/EW2
1	PIRLS+peiling	4.786	2.330	1.680	1.533	2.054
2	PIRLS+peiling	5.863	1.592	1.418	1.373	3.683
3	PIRLS+peiling	9.499	2.513	1.668	1.366	3.780
4	PIRLS+peiling	5.544	1.878	1.487	1.380	2.952
5	2 PIRLS-teksten	8.377	1.682	1.647	1.535	4.980
6	2 PIRLS-teksten	6.536	1.641	1.431	1.389	3.983
7	2 PIRLS-teksten	9.049	1.834	1.500	1.312	4.934
8	2 PIRLS-teksten	7.212	1.822	1.553	1.519	3.958
9	PIRLS+peiling	7.852	1.996	1.543	1.356	3.934
10	PIRLS+peiling	5.190	1.879	1.686	1.393	2.762

Voor de zes toetsboekjes met items van PIRLS en het Peilingsonderzoek vinden we dat de eerste geëxtraheerde factor van de toetsboekjes gemiddeld iets meer dan driemaal zoveel variantie verklaard als de tweede geëxtraheerde factor. Dit toont dat er één dominante factor is die de antwoorden op de toetsvragen verklaart. Echter, de vier toetsboekjes van de herhalingsmeting die enkel toetsvragen bevatten van PIRLS, de 16 toetsboekjes van PIRLS 2016 en de vier toetsboekjes van het Peilingsonderzoek tonen een gemiddelde verhouding van vier. Deze vergelijking toont dat de toetsboekjes met een combinatie van toetsvragen van beide studies minder unidimensioneel is.

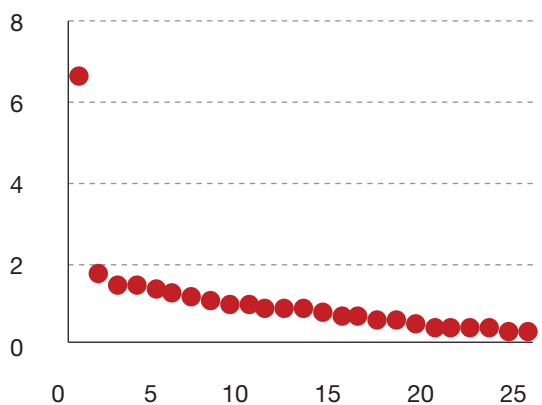
Tabel 74: Scree plots toetsboekjes op basis van exploratieve factoranalyses



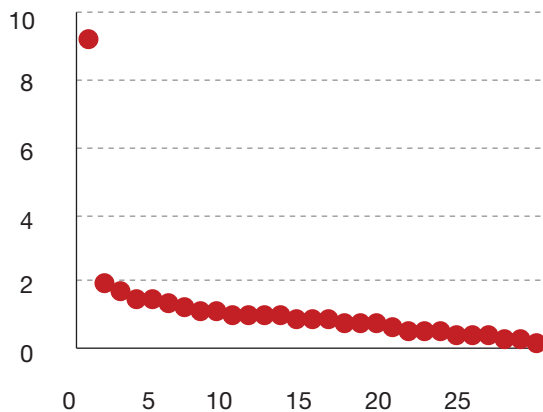
Scree plot toetsboekje 5



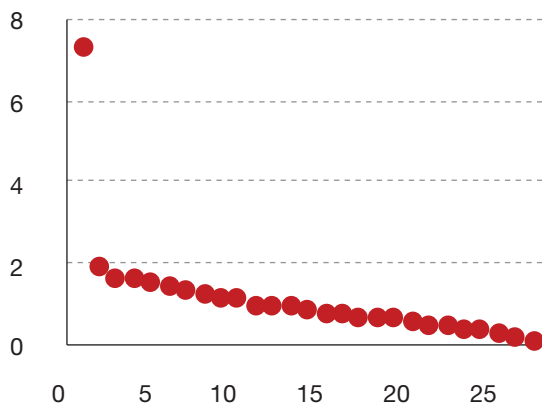
Scree plot toetsboekje 6



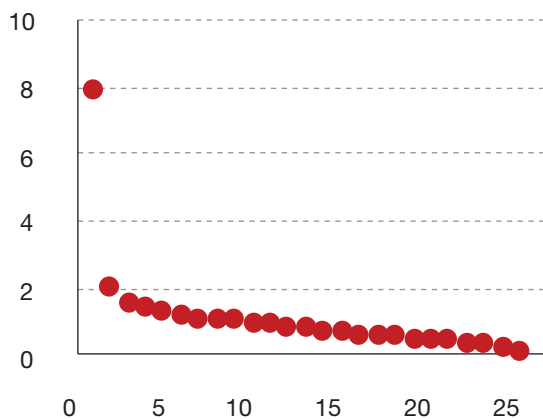
Scree plot toetsboekje 7



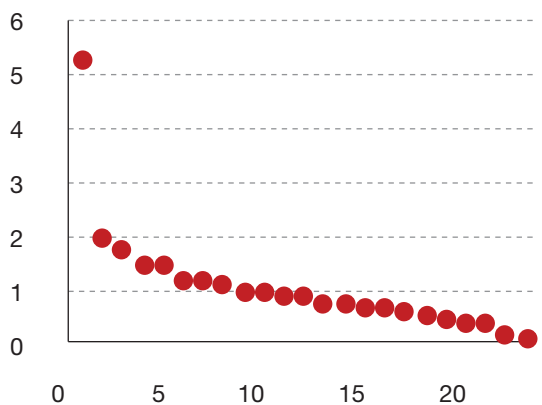
Scree plot toetsboekje 8



Scree plot toetsboekje 9



Scree plot toetsboekje 10



7.6.2 Vergelijking 1-factor, 2-factor en bifactor IRT-modellen

In [Tabel 75](#) tot [Tabel 85](#) staan de resultaten van de 1-factor, 2-factor en bifactor IRT-modellen van de tien toetsboekjes van de herhalingsmeting van PIRLS in 2018. Iedere tabel toont de geschatte discriminatieparameters en de geschatte verklaarde varianties van de drie verschillende modellen. Per model worden ook de fitindices RMSEA, SRMSR, CFI, TLI, AIC en BIC getoond. Daarbij tonen we ook de verklaarde variantie over alle toetsvragen en de correlatie tussen de factoren van het 2-factor IRT-model. In de volgende paragrafen interpreteren we deze resultaten per toetsboekje vooraleer we tot een algemene conclusie komen.

Voor toetsboekje 1 vinden we dat alle IRT-modellen aan de drempelwaarden voor de RMSEA en de SRMSR voldoen. Voor de drempelwaarden van de CFI en TLI voldoet enkel het bifactor IRT-model, het 2-factor IRT-model voldoet net niet. De AIC duidt erop dat het bifactor IRT-model het best de antwoorden verklaart, terwijl de BIC erop duidt dat het 2-factor IRT-model het best de antwoorden verklaart. Relatief gezien verklaart het 1-factor IRT-model 86.15% van de variantie die het 2-factor IRT-model verklaart en 65.03% van de variantie die het bifactor IRT-model verklaart. Er is een correlatie van 0.748 tussen de factoren van het 2-factor IRT-model.

Voor toetsboekje 2 vinden we dat alle IRT-modellen aan de drempelwaarden voor de RMSEA en de SRMSR voldoen. Voor de drempelwaarden van de CFI voldoen alle modellen. Voor de drempelwaarde van de TLI voldoen zowel het 2-factor IRT-model als het bifactor IRT-model. Het 1-factor IRT-model voldoet net niet. Zowel de AIC als BIC duiden erop dat het 2-factor IRT-model het best de antwoorden verklaart. Relatief gezien verklaart het 1-factor IRT-model 87.58% van de variantie die het 2-factor IRT-model verklaart en 77.90% van de variantie die het bifactor IRT-model verklaart. Er is een correlatie van 0.760 tussen de factoren van het 2-factor IRT-model.

Voor toetsboekje 3 vinden we dat alle IRT-modellen aan de drempelwaarden voor de RMSEA en de SRMSR voldoen. Voor de drempelwaarden van de TLI en CFI voldoen enkel het 2-factor IRT-model en het bifactor IRT-model. De AIC duidt erop dat het bifactor IRT-model het best de antwoorden verklaart, terwijl de BIC erop duidt dat het 2-factor IRT-model het best de antwoorden verklaart. Relatief gezien verklaart het 1-factor IRT-model 85.64% van de variantie die het 2-factor IRT-model verklaart en 75.22% van de variantie die het bifactor IRT-model verklaart. Er is een correlatie van 0.720 tussen de factoren van het 2-factor IRT-model.

Voor toetsboekje 4 vinden we dat alle IRT-modellen aan de drempelwaarden voor de RMSEA en de SRMSR voldoen. Voor de drempelwaarden van de TLI en CFI voldoen enkel het 2-factor IRT-model en het bifactor IRT-model. Zowel de AIC als BIC duiden erop dat het 2-factor IRT-model het best de antwoorden verklaart. Relatief gezien verklaart het 1-factor IRT-model 83.88% van de variantie die het 2-factor IRT-model verklaart en 72.01% van de variantie die het bifactor IRT-model verklaart. Er is een correlatie van 0.647 tussen de factoren van het 2-factor IRT-model.

Voor toetsboekje 5 (met 2 PIRLS-teksten) vinden we dat alle IRT-modellen aan de drempelwaarden voor de RMSEA en de SRMSR voldoen. Voor de drempelwaarden van de TLI en CFI voldoen ook alle modellen. Zowel de AIC als BIC duiden erop dat het 2-factor IRT-model het best de antwoorden verklaart. Relatief gezien verklaart het 1-factor IRT-model 93.33% van de variantie die het 2-factor IRT-model verklaart en 81.03% van de variantie die het bifactor IRT-model verklaart. Er is een correlatie van 0.869 tussen de factoren van het 2-factor IRT-model.

Voor toetsboekje 6 (met 2 PIRLS-teksten) vinden we dat alle IRT-modellen aan de drempelwaarden voor de RMSEA en de SRMSR voldoen. Voor de drempelwaarden van de TLI en CFI voldoen ook alle modellen. De AIC duidt erop dat het bifactor IRT-model het best de antwoorden verklaart, terwijl de BIC erop duidt dat het 1-factor IRT-model het best de antwoorden verklaart. Relatief gezien verklaart het 1-factor IRT-model 97.52% van de variantie die het 2-factor IRT-model verklaart en 80.00% van de variantie die het bifactor IRT-model verklaart. Er is een correlatie van 0.953 tussen de factoren van het 2-factor IRT-model.

Voor toetsboekje 7 (met 2 PIRLS-teksten) vinden we dat alle IRT-modellen aan de drempelwaarden voor de RMSEA en de SRMSR voldoen. Voor de drempelwaarden van de TLI en CFI voldoen ook alle modellen. De AIC duidt erop dat het bifactor IRT-model het best de antwoorden verklaart, terwijl de BIC erop duidt dat het 2-factor IRT-model het best de antwoorden verklaart. Relatief gezien verklaart het 1-factor IRT-model 93.88% van de variantie die het 2-factor IRT-model verklaart en 77.09% van de variantie die het bifactor IRT-model verklaart. Er is een correlatie van 0.869 tussen de factoren van het 2-factor IRT-model.

Voor toetsboekje 8 (met 2 PIRLS-teksten) vinden we dat alle IRT-modellen aan de drempelwaarden voor de RMSEA en de SRMSR voldoen. Voor de drempelwaarden van de TLI en CFI voldoen ook alle modellen. Zowel de AIC als BIC duiden erop dat het 2-factor IRT-model het best de antwoorden verklaart. Relatief gezien verklaart het 1-factor IRT-model 91.80% van de variantie die het 2-factor IRT-model verklaart en 74.13% van de variantie die het bifactor IRT-model verklaart. Er is een correlatie van 0.858 tussen de factoren van het 2-factor IRT-model.

Voor toetsboekje 9 vinden we dat alle IRT-modellen aan de drempelwaarden voor de RMSEA en de SRMSR voldoen. Voor de drempelwaarden van de TLI en CFI voldoen ook alle modellen. Zowel de AIC als BIC duiden erop dat het 2-factor IRT-model het best de antwoorden verklaart. Relatief gezien verklaart het 1-factor IRT-model 93.02% van de variantie die het 2-factor IRT-model verklaart en 79.19% van de variantie die het bifactor IRT-model verklaart. Er is een correlatie van 0.868 tussen de factoren van het 2-factor IRT-model.

Voor toetsboekje 10 vinden we dat alle IRT-modellen aan de drempelwaarden voor de RMSEA en de SRMSR voldoen. Voor de drempelwaarden van de CFI voldoen ook alle modellen. Voor de drempelwaarden van de TLI voldoen enkel het 2-factor IRT-model en het bifactor IRT-model. Het 1-factor IRT-model voldoet net niet. Zowel de AIC als BIC duiden erop dat het 2-factor IRT-model het best de antwoorden verklaart. Relatief gezien verklaart het 1-factor IRT-model 89.80% van de variantie die het 2-factor IRT-model verklaart en 73.89% van de variantie die het bifactor IRT-model verklaart. Er is een correlatie van 0.776 tussen de factoren van het 2-factor IRT-model.

Voor het model op basis van alle gegevens (**Tabel 85**) vinden we de AIC erop duidt dat het bifactor IRT-model het best de antwoorden verklaart. De BIC duidt erop dat het 2-factor IRT-model het best de antwoorden verklaart. Relatief gezien verklaart het 1-factor IRT-model 92.49% van de variantie die het 2-factor IRT-model verklaart en 83.13% van de variantie die het bifactor IRT-model verklaart. Er is een correlatie van 0.765 tussen de factoren van het 2-factor IRT-model.

Tabel 75: Factorladingen (a, a₁ en a₂), verklaarde varianties (h²) en fitindices van 1-factor model, 2-factor model en bifactor model toetsboekje 1 (Marie en de rode kip / Chinese horoscoop / Postzegel)

Item	1-factor model		2-factor model			Bifactor model			
	a	h ²	a ₁	a ₂	h ²	a	n ₁	n ₂	h ²
R41H01M	0.900	0.219	0.897		0.217	0.897	0.140		0.222
R41H02M	0.485	0.075	0.567		0.100	0.457	0.387		0.110
R41H05M	1.100	0.295	1.164		0.319	1.097	0.273		0.306
R41H07M	0.627	0.119	0.739		0.159	0.534	0.606		0.184
R41H09M	1.014	0.262	1.129		0.306	0.835	0.705		0.292
R41H10M	0.852	0.200	1.026		0.266	0.628	1.116		0.362
R41H11M	1.100	0.295	1.756		0.516	0.873	5.215		0.906
R41H12M	1.538	0.450	1.811		0.531	1.515	1.122		0.551
R41H03C	0.794	0.179	0.943		0.235	0.622	0.777		0.255
R41H04C	0.388	0.050	0.501		0.080	0.251	0.831		0.206
R41H06C	0.426	0.059	0.439		0.063	0.387	0.198		0.061
R41H08C	0.284	0.027	0.360		0.043	0.164	0.577		0.111
R41H13C	0.368	0.045	0.384		0.049	0.366	0.087		0.047
R41H14C	0.378	0.047	0.416		0.057	0.335	0.208		0.051
R41H15C	0.839	0.196	0.863		0.205	0.764	0.395		0.203
R41H16C	0.311	0.032	0.326		0.036	0.246	0.181		0.031
i2091	0.726	0.154		0.742	0.160	0.869		-0.241	0.219
i2092	1.946	0.567		2.052	0.593	1.915		0.134	0.560
i2094	0.351	0.041		0.389	0.050	0.387		0.034	0.050
i2095	0.433	0.061		0.482	0.074	0.450		0.142	0.071
i2096	0.841	0.196		0.700	0.145	1.273		-0.703	0.422
Le6PZ1	1.374	0.395		1.568	0.459	1.501		0.136	0.440
Le6PZ3	1.400	0.404		1.491	0.434	1.468		0.177	0.430
Le6PZ6	0.883	0.212		0.982	0.250	1.202		1.844	0.626
Le6PZ9	1.085	0.289		1.205	0.334	1.999		3.522	0.850
Le6PZ7	1.131	0.307		1.247	0.349	1.351		0.115	0.388
Gem. h ²	0.199		0.231			0.306			
df	296	295			270				
RMSEA	0.023		0.018			0.010			
SRMSR	0.048		0.046			0.041			
TLI	0.912	0.944			0.984				
CFI	0.919	0.949			0.987				
AIC	13027.7	13012.990			12997.6				
BIC	13258.4	13247.870			13337.4				
f _{factor1} *f _{factor2}			0.748						

Opmerking: a = factorlading, a₁ = factorlading op factor 1, a₂ = factorlading op factor 2, n₁ = lading op storingsfactor 1, n₂ = lading op storingsfactor 2, h² = verklaarde variantie item

Tabel 76: Factorladingen (a, a₁ en a₂), verklaarde varianties (h²) en fitindices van 1-factor model, 2-factor model en bifactor model toetsboekje 2 (Haaien / Vliegtuig / Postzegel)

Item	1-factor model		2-factor model			Bifactor model			
	a	h ²	a ₁	a ₂	h ²	a	n ₁	n ₂	h ²
R21K03M	0.747	0.162	0.735		0.157	0.703	0.171		0.153
R21K04M	1.402	0.404	1.655		0.486	1.321	1.382		0.558
R21K06M	1.278	0.361	1.393		0.401	1.120	0.726		0.381
R21K08M	0.792	0.178	0.867		0.206	0.682	0.508		0.200
R21K09M	1.149	0.313	1.352		0.387	0.970	1.238		0.460
R21K11M	0.748	0.162	0.809		0.184	0.597	0.583		0.194
R21K01C	0.430	0.060	0.447		0.065	0.375	0.253		0.066
R21K02C	0.786	0.176	0.864		0.205	0.616	0.733		0.240
R21K05C	0.957	0.240	1.008		0.260	0.859	0.420		0.240
R21K07C	0.703	0.146	0.805		0.183	0.601	0.482		0.170
R21K10C	0.750	0.163	0.890		0.215	0.691	0.597		0.223
R21K12C	0.566	0.100	0.637		0.123	0.466	0.468		0.131
Le6PZ1	1.501	0.438		1.589	0.466	1.586		0.202	0.469
Le6PZ3	1.885	0.551		2.227	0.631	2.247		0.732	0.658
Le6PZ6	0.769	0.170		0.844	0.197	0.949		1.139	0.431
Le6PZ9	2.290	0.644		3.200	0.780	6.573		5.412	0.962
Le6PZ7	1.304	0.370		1.453	0.422	1.427		0.391	0.430
Le1VL1	0.581	0.104		0.528	0.088	0.611		-0.469	0.170
Le1VL4	1.083	0.288		1.223	0.341	1.334		-0.316	0.394
Le1VL3	1.557	0.456		1.626	0.477	1.729		-0.224	0.512
Le1VL8	1.041	0.272		1.120	0.302	1.274		-0.509	0.394
Le1VL7	1.119	0.302		1.168	0.320	1.173		-0.208	0.329
Gem. h ²	0.275		0.314			0.353			
df	204	203			182				
RMSEA	0.031		0.022			0.015			
SRMSR	0.049		0.045			0.042			
TLI 0.948		0.975			0.987				
CFI 0.953		0.977			0.990				
AIC 11652.9		11615.2			11624.8				
BIC 11856.2		11822.6			11919.4				
f ² factor1*factor2			0.760						

Opmerking: a = factorlading, a₁ = factorlading op factor 1, a₂ = factorlading op factor 2, n₁ = lading op storingsfactor 1, n₂ = lading op storingsfactor 2, h² = verklaarde variantie item

Tabel 77: Factorladingen, verklaarde varianties en fitindices van 1-factor model, 2-factor model en bifactor model toetsboekje 3 (Magische sleutel / Moeilijk gaat ook / De reis van de groene zeeschildpad)

Item	1-factor model		2-factor model			Bifactor model			
	a	h ²	a ₁	a ₂	h ²	a	n ₁	n ₂	h ²
R041T01M	2.353	0.657	2.327		0.652	2.358	0.141		0.658
R41T05M	0.757	0.165	0.769		0.169	1.246	-1.116		0.491
R41T09M	1.627	0.477	1.703		0.500	1.670	0.287		0.498
R41T12M	1.905	0.556	2.011		0.583	2.136	1.352		0.688
R41T13M	1.336	0.381	1.403		0.405	1.343	0.700		0.442
R41T15M	1.150	0.313	1.232		0.344	1.597	2.007		0.694
R41T16M	1.579	0.463	1.641		0.482	2.189	2.354		0.781
R41T02C	1.067	0.282	1.090		0.291	1.200	-0.168		0.337
R41T03C	1.350	0.386	1.299		0.368	1.582	-0.263		0.470
R41T04C	1.719	0.505	1.744		0.512	2.052	-0.412		0.602
R41T06C	2.000	0.580	2.063		0.595	2.005	0.386		0.590
R41T07C	0.651	0.128	0.656		0.129	0.671	0.017		0.135
R41T08C	1.885	0.551	1.990		0.578	1.923	0.380		0.570
R41T10C	1.814	0.532	1.894		0.553	1.855	0.352		0.552
R41T11C	1.166	0.319	1.274		0.359	1.229	0.327		0.358
R41T14C	0.484	0.075	0.509		0.082	0.443	0.604		0.162
Le5DMS1	0.938	0.233		1.355	0.388	0.931		1.005	0.393
Le5DMS4	1.018	0.264		1.154	0.315	0.945		0.536	0.290
Le5DMS2	1.047	0.275		1.410	0.407	1.025		0.967	0.407
Le5DMS6	1.144	0.311		1.462	0.425	1.108		1.009	0.437
Le5DMS3	1.232	0.344		1.997	0.579	1.472		1.698	0.636
Le5DMS7	1.127	0.305		1.457	0.423	1.166		0.868	0.422
Le1SP6	0.915	0.224		1.160	0.317	0.789		1.012	0.362
Le1SP2	1.040	0.272		1.486	0.432	1.036		1.236	0.473
Le1SP3	1.034	0.270		1.146	0.312	0.992		0.469	0.294
Le1SP5	1.002	0.257		1.383	0.398	1.036		0.977	0.412
Le1SP7	0.931	0.230		1.122	0.303	0.853		0.641	0.282
Gem. h ²	0.346		0.404			0.460			
df	319	318			292				
RMSEA	0.054		0.039			0.030			
SRMSR	0.062		0.054			0.047			
TLI	0.923	0.961			0.977				
CFI	0.929	0.964			0.980				
AIC	13874.0	13765.7			13713.0				
BIC	14117.3	14013.2			14067.7				
I ² _{factor1*factor2}			0.720						

Opmerking: a = factorlading, a₁ = factorlading op factor 1, a₂ = factorlading op factor 2, n₁ = lading op storingsfactor 1, n₂ = lading op storingsfactor 2, h² = verklaarde variantie item

Tabel 78: Factorladingen, verklaarde varianties en fitindices van 1-factor model, 2-factor model en bifactor model toetsboekje 4 (Miegtuig / Postzegel / Bloemen op het dak)

Item	1-factor model		2-factor model			Bifactor model			
	a	h ²	a ₁	a ₂	h ²	a	n ₁	n ₂	h ²
R11F01M	1.612	0.473	1.727		0.507	0.975	1.554		0.538
R11F02M	0.303	0.031	0.293		0.029	0.271	0.125		0.030
R11F03M	1.061	0.280	1.033		0.269	0.861	0.594		0.274
R11F04M	1.660	0.487	1.746		0.513	1.190	1.239		0.505
R11F05M	1.214	0.337	1.265		0.356	0.972	0.781		0.349
R11F11M	0.619	0.117	0.619		0.117	0.455	0.423		0.117
R11F13M	0.942	0.235	0.993		0.254	0.491	0.981		0.294
R11F06C	0.819	0.188	0.866		0.206	0.552	0.671		0.207
R11F07C	0.533	0.089	0.546		0.093	0.386	0.375		0.091
R11F08C	1.402	0.404	1.512		0.441	0.923	1.288		0.464
R11F09C	1.374	0.394	1.455		0.422	0.983	1.114		0.433
R11F10C	0.795	0.179	0.848		0.199	0.366	0.884		0.240
R11F12C	0.628	0.120	0.690		0.141	0.468	0.455		0.128
Le6PZ1	0.934	0.232		1.053	0.277	1.102		-0.169	0.300
Le6PZ3	1.429	0.413		2.305	0.647	2.533		-0.129	0.690
Le6PZ6	0.452	0.066		0.524	0.087	0.685		-0.530	0.206
Le6PZ9	1.724	0.507		1.915	0.559	3.052		-1.327	0.793
Le6PZ7	0.667	0.133		0.815	0.187	0.826		0.040	0.191
Le1VL1	0.190	0.012		0.401	0.053	0.303		0.301	0.059
Le1VL4	0.527	0.087		0.738	0.158	0.691		0.728	0.258
Le1VL3	0.981	0.249		1.247	0.349	1.296		1.054	0.491
Le1VL8	0.656	0.129		0.963	0.242	0.911		0.662	0.305
Le1VL7	0.589	0.107		0.775	0.172	0.685		1.024	0.344
Gem. h ²	0.229		0.273			0.318			
df	226.0	225.0			203.0				
RMSEA	0.030		0.017			0.014			
SRMSR	0.050		0.045			0.042			
TLI 0.930		0.978			0.985				
CFI 0.937		0.980			0.988				
AIC 11481.5		11436.3			11448.6				
BIC 11687.2		11646.1			11748.9				
r ² factor1*factor2			0.647						

Opmerking: a = factorlading, a₁ = factorlading op factor 1, a₂ = factorlading op factor 2, n₁ = lading op storingsfactor 1, n₂ = lading op storingsfactor 2, h² = verklaarde variantie item

Tabel 79: Factorladingen, verklaarde varianties en fitindices van 1-factor model, 2-factor model en bifactor model toetsboekje 5 (Ijslandse paarden / Marie en de rode kip)

Item	1-factor model		2-factor model			Bifactor model			
	a	h ²	a ₁	a ₂	h ²	a	n ₁	n ₂	h ²
R41H01M	1.532	0.448	1.596		0.468	1.507	0.271		0.447
R41H02M	0.710	0.148	0.723		0.153	0.675	0.253		0.152
R41H05M	0.918	0.225	0.925		0.228	1.018	-0.260		0.276
R41H07M	1.070	0.283	1.086		0.289	1.021	0.251		0.276
R41H09M	1.167	0.320	1.224		0.341	1.158	0.154		0.320
R41H10M	1.235	0.345	1.287		0.364	1.143	0.732		0.389
R41H11M	1.699	0.499	1.864		0.545	1.560	1.052		0.550
R41H12M	1.541	0.451	1.697		0.499	2.032	2.818		0.807
R41I02M	0.956	0.240		0.995	0.255	1.011		-0.085	0.262
R41I05M	0.875	0.209		0.947	0.237	0.896		0.279	0.233
R41I06M	0.970	0.245		0.972	0.246	0.971		0.116	0.248
R41I08M	1.866	0.546		2.060	0.594	2.400		2.641	0.815
R41I10M	1.071	0.284		1.145	0.311	1.129		0.996	0.439
R41I12M	1.047	0.275		1.113	0.299	1.060		0.379	0.305
R41H03C	1.525	0.446	1.594		0.467	1.504	0.417		0.457
R41H04C	0.460	0.068	0.492		0.077	0.429	0.199		0.072
R41H06C	0.901	0.219	0.942		0.235	0.887	0.170		0.220
R41H08C	0.574	0.102	0.568		0.100	0.583	-0.005		0.105
R41H13C	0.517	0.084	0.533		0.089	0.530	0.032		0.089
R41H14C	0.787	0.176	0.872		0.208	0.819	1.616		0.531
R41H15C	1.292	0.366	1.394		0.401	1.217	0.977		0.457
R41H16C	0.360	0.043	0.394		0.051	0.309	0.301		0.061
R41I01C	0.616	0.116		0.563	0.099	0.556		0.256	0.115
R41I03C	0.354	0.041		0.369	0.045	0.391		-0.165	0.059
R41I04C	0.512	0.083		0.532	0.089	0.591		-0.239	0.123
R41I07C	0.729	0.155		0.779	0.173	0.755		0.234	0.177
R41I09C	1.275	0.360		1.371	0.394	1.352		0.872	0.472
R41I11C	0.892	0.215		0.955	0.240	0.921		0.136	0.230
R41I13C	1.069	0.283		1.145	0.312	1.106		0.010	0.297
R41I14C	1.085	0.289		1.130	0.306	1.130		0.241	0.315
R41I15C	0.984	0.250		0.991	0.253	1.121		-0.459	0.336
Gem. h ²	0.252		0.270			0.311			
df	427	426			396				
RMSEA	0.012		0.000			0.000			
SRMSR	0.045		0.043			0.041			
TLI	0.993	1.000			1.007				
CFI	0.993	1.000			1.000				
AIC	16759.4	16741.6			16742.1				
BIC	17044.3	17030.6			17155.0				
f _{factor1} *f _{factor2}			0.869						

Opmerking: a = factorlading, a₁ = factorlading op factor 1, a₂ = factorlading op factor 2, n₁ = lading op storingsfactor 1, n₂ = lading op storingsfactor 2, h² = verklaarde variantie item

Tabel 80: Factorladingen, verklaarde varianties en fitindices van 1-factor model, 2-factor model en bifactor model toetsboekje 6 (Olivier en de griffioen / Haaien)

Item	1-factor model		2-factor model			Bifactor model			
	a	h ²	a ₁	a ₂	h ²	a	n ₁	n ₂	h ²
R21K03M	0.845	0.198	0.844		0.197	0.837	0.477		0.198
R21K04M	1.308	0.371	1.341		0.383	1.310	-0.074		0.488
R21K06M	1.822	0.534	1.852		0.542	2.141	-0.356		0.612
R21K08M	1.126	0.305	1.151		0.314	1.278	-0.383		0.796
R21K09M	1.334	0.380	1.331		0.380	1.419	-0.368		0.375
R21K11M	1.350	0.386	1.356		0.388	1.692	-0.841		0.409
R41O01M	1.154	0.315		1.204	0.334	1.116		1.871	0.508
R41O06M	1.273	0.359		1.289	0.365	1.273		0.822	0.378
R41O11M	1.270	0.358		1.275	0.360	1.266		0.244	0.357
R41O12M	1.163	0.318		1.171	0.321	1.150		0.248	0.320
R21K01C	0.781	0.174	0.783		0.175	1.015	1.027		0.230
R21K02C	0.521	0.086	0.531		0.089	0.547	0.260		0.103
R21K05C	0.900	0.219	0.913		0.223	0.961	0.341		0.258
R21K07C	0.917	0.225	0.944		0.235	0.980	0.345		0.237
R21K10C	0.850	0.200	0.869		0.207	0.891	-0.157		0.216
R21K12C	0.337	0.038	0.343		0.039	0.367	0.213		0.074
R41O02C	1.201	0.332		1.228	0.343	1.162		0.621	0.326
R41O03C	1.038	0.271		1.053	0.277	1.034		0.235	0.273
R41O04C	0.431	0.060		0.444	0.064	0.402		0.676	0.315
R41O05C	0.448	0.065		0.465	0.070	0.412		0.469	0.088
R41O07C	0.864	0.205		0.873	0.209	0.845		0.269	0.214
R41O08C	0.466	0.070		0.471	0.071	0.413		0.505	0.101
R41O09C	1.004	0.258		1.043	0.273	0.922		0.827	0.274
R41O10C	0.616	0.116		0.626	0.119	0.558		0.373	0.132
R41O13C	0.440	0.063		0.440	0.063	0.478		-0.042	0.096
Gem. h ²	0.236		0.242			0.295			
df	264	263			239				
RMSEA	0.027		0.027			0.020			
SRMSR	0.047		0.047			0.042			
TLI	0.962	0.962			0.979				
CFI	0.965	0.965			0.983				
AIC	14791.1	14792.2			14773.7				
BIC	15041.4	15046.5			15126.5				
f ² _{factor1*factor2}			0.953						

Opmerking: a = factorlading, a₁ = factorlading op factor 1, a₂ = factorlading op factor 2, n₁ = lading op storingsfactor 1, n₂ = lading op storingsfactor 2, h² = verklaarde variantie item

Tabel 81: Factorladingen, verklaarde varianties en fitindices van 1-factor model, 2-factor model en bifactor model toetsboekje 7 (De reis van de groene zeeschildpad / Olivier en de griffioen)

Item	1-factor model		2-factor model			Bifactor model			
	a	h ²	a ₁	a ₂	h ²	a	n ₁	n ₂	h ²
R41T01M	1.004	0.258	1.015		0.262	1.115	-0.427		0.330
R41T05M	0.981	0.250	0.958		0.241	1.245	-0.521		0.386
R41T09M	1.299	0.368	1.378		0.396	1.287	0.314		0.377
R41T12M	1.568	0.459	1.645		0.483	1.594	1.302		0.594
R41T13M	1.733	0.509	1.822		0.534	1.783	0.609		0.551
R41T15M	1.113	0.300	1.149		0.313	1.179	1.495		0.556
R41T16M	1.486	0.433	1.590		0.466	4.635	7.473		0.964
R41O01M	0.791	0.178		0.836	0.194	0.727		0.336	0.181
R41O06M	1.625	0.477		1.728	0.508	1.627		0.270	0.484
R41O11M	1.372	0.394		1.457	0.423	1.349		0.387	0.405
R41O12M	0.872	0.208		0.974	0.247	0.821		0.836	0.322
R41T02C	0.955	0.239	0.941		0.234	1.077	-0.308		0.302
R41T03C	0.791	0.178	0.799		0.181	0.810	0.043		0.185
R41T04C	1.163	0.318	1.204		0.334	1.220	-0.016		0.340
R41T06C	1.654	0.486	1.701		0.500	1.652	0.145		0.487
R41T07C	0.369	0.045	0.381		0.048	0.412	-0.106		0.059
R41T08C	1.476	0.429	1.576		0.462	1.549	0.215		0.458
R41T10C	1.913	0.558	1.958		0.570	2.096	-0.169		0.604
R41T11C	0.864	0.205	0.948		0.237	0.893	0.339		0.239
R41T14C	0.425	0.059	0.451		0.066	0.423	0.143		0.064
R41O02C	1.122	0.303		1.176	0.323	1.238		-0.187	0.351
R41O03C	1.129	0.305		1.237	0.346	1.110		0.193	0.305
R41O04C	0.559	0.097		0.557	0.097	0.606		-0.062	0.113
R41O05C	0.531	0.089		0.558	0.097	0.531		0.120	0.093
R41O07C	0.819	0.188		0.845	0.198	0.795		0.112	0.182
R41O08C	0.854	0.201		0.952	0.238	0.843		0.221	0.208
R41O09C	1.053	0.277		1.143	0.311	1.033		0.593	0.329
R41O10C	0.664	0.132		0.715	0.150	0.625		0.244	0.135
R41O13C	0.418	0.057		0.477	0.073	1.479		2.792	0.775
Gem. h ²	0.276		0.294			0.358			
df 366		365			337				
RMSEA	0.023		0.019			0.000			
SRMSR	0.047		0.046			0.041			
TLI 0.981		0.987			1.002				
CFI 0.982		0.988			1.000				
AIC 16039.8		16017.6			15982.1				
BIC 16325.4		16307.4			16387.8				
f ² factor1*factor2			0.869						

Opmerking: a = factorlading, a₁ = factorlading op factor 1, a₂ = factorlading op factor 2, n₁ = lading op storingsfactor 1, n₂ = lading op storingsfactor 2, h² = verklaarde variantie item

Tabel 82: Factorladingen, verklaarde varianties en fitindices van 1-factor model, 2-factor model en bifactor model toetsboekje 8 (Bloemen op het dak / Ijslandse paarden)

Item	1-factor model		2-factor model			Bifactor model			
	a	h ²	a ₁	a ₂	h ²	a	n ₁	n ₂	h ²
R11F01M	0.769	0.170	0.836		0.194	1.489	3.210		0.812
R11F02M	0.443	0.063	0.484		0.075	0.479	0.065		0.075
R11F03M	0.942	0.235	1.011		0.261	0.997	0.105		0.258
R11F04M	0.963	0.243	1.061		0.280	1.102	0.447		0.328
R11F05M	1.413	0.408	1.601		0.470	1.549	0.121		0.455
R11F11M	0.904	0.220	1.023		0.265	0.946	-0.009		0.236
R11F13M	1.052	0.276	1.061		0.280	3.044	-4.101		0.900
R41I02M	0.933	0.231		0.930	0.230	1.023		-0.037	0.266
R41I05M	1.403	0.405		1.431	0.414	1.236		0.717	0.413
R41I06M	1.125	0.304		1.158	0.316	1.009		0.573	0.317
R41I08M	1.695	0.498		1.911	0.558	1.418		2.283	0.714
R41I10M	0.677	0.137		0.717	0.151	0.479		0.987	0.294
R41I12M	1.776	0.521		1.796	0.527	1.648		0.716	0.527
R11F06C	0.538	0.091	0.565		0.099	0.584	-0.114		0.109
R11F07C	0.608	0.113	0.689		0.141	0.666	0.041		0.133
R11F08C	1.199	0.332	1.346		0.385	1.287	0.224		0.371
R11F09C	1.311	0.372	1.489		0.434	1.424	0.173		0.416
R11F10C	0.698	0.144	0.763		0.167	0.750	0.200		0.172
R11F12C	0.462	0.069	0.498		0.079	0.490	0.158		0.084
R41I01C	0.832	0.193		0.856	0.202	0.767		0.260	0.185
R41I03C	0.498	0.079		0.524	0.087	0.459		0.195	0.079
R41I04C	0.657	0.130		0.677	0.137	0.605		0.287	0.134
R41I07C	0.957	0.240		0.980	0.249	0.897		0.320	0.239
R41I09C	0.948	0.237		0.985	0.251	0.821		0.623	0.268
R41I11C	0.749	0.162		0.768	0.169	0.699		0.311	0.168
R41I13C	1.026	0.267		1.084	0.289	0.889		1.070	0.401
R41I14C	0.865	0.205		0.863	0.204	0.948		-0.048	0.237
R41I15C	0.971	0.246		1.004	0.258	0.854		0.652	0.285
Gem. h ²	0.235		0.256			0.317			
df	342	341			314				
RMSEA	0.012		0.006			0.000			
SRMSR	0.044		0.043			0.040			
TLI	0.991	0.998			1.002				
CFI	0.991	0.998			1.000				
AIC	14215.4	14202.5			14217.7				
BIC	14478.6	14469.7			14596.0				
f _{factor1*factor2}			0.858						

Opmerking: a = factorlading, a₁ = factorlading op factor 1, a₂ = factorlading op factor 2, n₁ = lading op storingsfactor 1, n₂ = lading op storingsfactor 2, h² = verklaarde variantie item

Tabel 83: Factorladingen, verklaarde varianties en fitindices van 1-factor model, 2-factor model en bifactor model toetsboekje 9 (Magische sleutel / Chinese horoscoop / IJslandse paarden)

Item	1-factor model		2-factor model			Bifactor model			
	a	h ²	a ₁	a ₂	h ²	a	n ₁	n ₂	h ²
R41I02M	1.094	0.292	1.077		0.286	1.218	-0.019		0.339
R41I05M	1.498	0.436	1.513		0.442	1.218	1.220		0.506
R41I06M	1.141	0.310	1.153		0.315	1.127	0.200		0.312
R41I08M	4.638	0.881	5.068		0.899	5.524	5.099		0.951
R41I10M	1.139	0.310	1.166		0.320	1.003	0.627		0.326
R41I12M	1.139	0.309	1.142		0.310	1.093	0.324		0.310
R41I01C	0.950	0.238	0.994		0.254	0.621	1.099		0.355
R41I03C	0.479	0.073	0.491		0.077	0.415	0.327		0.088
R41I04C	0.633	0.122	0.630		0.121	0.595	0.203		0.120
R41I07C	1.083	0.288	1.081		0.288	1.133	0.141		0.311
R41I09C	1.175	0.323	1.195		0.330	0.990	0.815		0.362
R41I11C	1.014	0.262	1.072		0.284	0.947	0.698		0.323
R41I13C	1.143	0.311	1.154		0.315	1.027	0.560		0.321
R41I14C	1.245	0.349	1.269		0.358	1.196	0.398		0.354
R41I15C	0.659	0.131	0.688		0.140	0.580	0.534		0.177
i2091	0.796	0.179		0.819	0.188	0.803		-0.003	0.182
i2092	2.084	0.600		2.355	0.657	2.418		0.294	0.672
i2094	0.235	0.019		0.233	0.018	0.261		-0.120	0.028
i2095	0.632	0.121		0.686	0.140	0.683		0.102	0.141
i2096	1.162	0.318		1.239	0.347	1.793		2.693	0.783
Le5DMS1	1.213	0.337		1.293	0.366	1.287		0.286	0.375
Le5DMS4	0.930	0.230		1.156	0.316	2.440		4.663	0.905
Le5DMS2	1.390	0.400		1.618	0.475	1.521		0.195	0.448
Le5DMS6	0.739	0.159		0.864	0.205	0.819		0.330	0.212
Le5DMS3	0.998	0.256		1.090	0.291	1.054		-0.079	0.278
Le5DMS7	1.276	0.360		1.583	0.464	1.498		0.325	0.448
Gem. h ²	0.293		0.315			0.370			
df	294	293			268				
RMSEA	0.017		0.008			0.000			
SRMSR	0.049		0.047			0.044			
TLI	0.985	0.997			1.010				
CFI	0.986	0.997			1.000				
AIC	12963.9	12955.1			12964.3				
BIC	13202.4	13197.8			13311.6				
f ^{factor1} *f ^{factor2}			0.868						

Opmerking: a = factorlading, a₁ = factorlading op factor 1, a₂ = factorlading op factor 2, n₁ = lading op storingsfactor 1, n₂ = lading op storingsfactor 2, h² = verklaarde variantie item

Tabel 84: Factorladingen, verklaarde varianties en fitindices van 1-factor model, 2-factor model en bifactor model toetsboekje 10 (Moeilijk gaat ook / Chinese horoscoop / Olivier en de griffioen)

Item	1-factor model		2-factor model			Bifactor model			
	a	h ²	a ₁	a ₂	h ²	a	n ₁	n ₂	h ²
R41O01M	1.480	0.431	1.463		0.425	1.667	-0.015		0.490
R41O06M	1.370	0.393	1.542		0.451	1.235	0.741		0.417
R41O11M	1.176	0.323	1.104		0.296	1.656	-0.450		0.504
R41O12M	0.591	0.108	0.666		0.133	0.421	1.209		0.361
R41O02C	1.115	0.300	1.182		0.325	0.884	0.802		0.330
R41O03C	1.127	0.305	1.194		0.330	1.044	0.470		0.312
R41O04C	0.327	0.036	0.362		0.043	0.279	0.230		0.043
R41O05C	0.411	0.055	0.431		0.060	0.383	0.153		0.055
R41O07C	0.651	0.128	0.704		0.146	0.600	0.336		0.140
R41O08C	0.776	0.172	0.785		0.176	0.816	0.011		0.187
R41O09C	0.602	0.111	0.694		0.143	0.418	0.627		0.164
R41O10C	0.545	0.093	0.592		0.108	0.461	0.333		0.100
R41O13C	0.421	0.058	0.464		0.069	0.408	0.518		0.131
i2091	1.187	0.327		1.377	0.396	1.343		0.079	0.384
i2092	1.905	0.556		2.188	0.623	2.065		0.410	0.605
i2094	0.378	0.047		0.428	0.060	0.302		0.981	0.267
i2095	1.028	0.267		1.222	0.340	2.360		4.357	0.895
i2096	0.880	0.211		0.976	0.248	0.914		0.313	0.244
Le1SP6	0.634	0.122		0.680	0.138	0.611		0.254	0.131
Le1SP2	1.081	0.287		1.172	0.322	1.129		0.309	0.321
Le1SP3	0.843	0.197		0.883	0.212	0.905		-0.040	0.221
Le1SP5	1.403	0.405		1.508	0.440	1.357		0.407	0.409
Le1SP7	0.627	0.120		0.693	0.142	0.678		-0.014	0.137
Gem. h ²	0.220		0.245			0.298			
df	224	223			201				
RMSEA	0.022		0.015			0.012			
SRMSR	0.048		0.045			0.042			
TLI 0.945		0.974			0.985				
CFI 0.950		0.977			0.988				
AIC 11547.0		11531.8			11534.8				
BIC 11763.1		11752.0			11846.4				
r _{factor1*factor2}			0.776						

Opmerking: a = factorlading, a₁ = factorlading op factor 1, a₂ = factorlading op factor 2, n₁ = lading op storingsfactor 1, n₂ = lading op storingsfactor 2, h² = verklaarde variantie item

Tabel 85: Factorladingen, verklaarde varianties en fitindices van 1-factor model, 2-factor model en bifactor model met alle 111 items

Item	1-factor model		2-factor model			Bifactor model			
	a	h ²	a ₁	a ₂	h ²	a	n ₁	n ₂	h ²
R41H01M	1.346	0.385	1.354		0.388	1.305	0.359		0.387
R41H02M	0.725	0.154	0.749		0.162	0.665	0.311		0.157
R41H05M	0.993	0.254	0.988		0.252	1.139	-0.106		0.311
R41H07M	0.914	0.224	0.961		0.242	0.812	0.460		0.231
R41H09M	1.190	0.328	1.215		0.338	1.036	0.530		0.319
R41H10M	1.325	0.378	1.374		0.395	1.069	1.050		0.436
R41H11M	1.756	0.516	1.933		0.563	1.335	2.019		0.669
R41H12M	1.726	0.507	1.781		0.523	1.556	1.987		0.687
R21K03M	0.770	0.170	0.774		0.172	0.779	0.122		0.177
R21K04M	1.318	0.375	1.395		0.402	1.298	0.384		0.387
R21K06M	1.529	0.447	1.622		0.476	1.314	1.307		0.542
R21K08M	0.950	0.238	1.014		0.262	0.782	0.864		0.319
R21K09M	1.238	0.346	1.335		0.381	1.056	0.903		0.400
R21K11M	1.008	0.260	1.062		0.280	0.815	0.787		0.307
R41T01M	1.726	0.507	1.747		0.513	1.759	0.247		0.521
R41T05M	0.854	0.201	0.865		0.205	1.249	-0.621		0.402
R41T09M	1.421	0.411	1.466		0.426	1.402	0.398		0.423
R41T12M	1.816	0.532	1.874		0.548	1.767	1.560		0.657
R41T13M	1.463	0.425	1.515		0.442	1.387	0.750		0.462
R41T15M	1.160	0.317	1.204		0.334	1.237	1.882		0.637
R41T16M	1.578	0.462	1.618		0.475	2.227	3.509		0.856
R11F01M	1.251	0.351	1.289		0.365	0.865	1.083		0.399
R11F02M	0.411	0.055	0.397		0.052	0.315	0.282		0.058
R11F03M	1.111	0.299	1.098		0.294	0.866	0.685		0.296
R11F04M	1.372	0.394	1.396		0.402	0.979	1.125		0.434
R11F05M	1.392	0.401	1.428		0.413	1.099	0.994		0.431
R11F11M	0.797	0.180	0.795		0.179	0.581	0.572		0.187
R11F13M	1.065	0.281	1.093		0.292	0.801	0.691		0.279
R41I02M	1.051	0.276	1.042		0.273	0.988	0.315		0.271
R41I05M	1.233	0.344	1.236		0.345	1.226	0.139		0.344
R41I06M	1.113	0.300	1.110		0.299	1.129	0.064		0.306
R41I08M	2.427	0.670	2.452		0.675	2.819	-0.290		0.735
R41I10M	0.984	0.251	0.994		0.255	1.035	-0.002		0.270
R41I12M	1.325	0.377	1.333		0.380	1.267	0.366		0.375
R41O01M	1.130	0.306	1.135		0.308	1.250	-0.084		0.351
R41O06M	1.545	0.452	1.593		0.467	1.606	0.188		0.474
R41O11M	1.377	0.396	1.371		0.394	1.318	0.398		0.396
R41O12M	0.889	0.214	0.925		0.228	0.828	0.440		0.233
R41H03C	1.454	0.422	1.522		0.445	1.263	0.790		0.434
R41H04C	0.502	0.080	0.541		0.092	0.380	0.510		0.122
R41H06C	0.749	0.162	0.767		0.169	0.676	0.307		0.160
R41H08C	0.462	0.069	0.499		0.079	0.410	0.277		0.078
R41H13C	0.485	0.075	0.495		0.078	0.478	0.108		0.077
R41H14C	0.736	0.158	0.761		0.167	0.529	0.942		0.287
R41H15C	1.319	0.375	1.340		0.383	1.088	0.896		0.407
R41H16C	0.486	0.075	0.497		0.079	0.373	0.401		0.094
R21K01C	0.597	0.110	0.624		0.119	0.596	0.145		0.115
R21K02C	0.656	0.129	0.691		0.141	0.551	0.446		0.148
R21K05C	0.946	0.236	0.988		0.252	0.810	0.600		0.260
R21K07C	0.813	0.186	0.884		0.213	0.716	0.479		0.204
R21K10C	0.801	0.181	0.864		0.205	0.722	0.439		0.198
R21K12C	0.434	0.061	0.460		0.068	0.370	0.267		0.067
R41T02C	1.042	0.273	1.059		0.279	1.149	-0.021		0.313
R41T03C	1.059	0.279	1.043		0.273	1.113	0.058		0.300

Item	1-factor model		2-factor model			Bifactor model			
	a	h ²	a ₁	a ₂	h ²	a	n ₁	n ₂	h ²
R41T04C	1.452	0.421	1.470		0.427	1.574	0.012		0.461
R41T06C	1.884	0.551	1.923		0.561	1.831	0.557		0.558
R41T07C	0.504	0.081	0.508		0.082	0.540	-0.004		0.092
R41T08C	1.657	0.487	1.712		0.503	1.662	0.395		0.502
R41T10C	1.856	0.543	1.912		0.558	1.905	0.308		0.562
R41T11C	1.021	0.265	1.068		0.283	1.010	0.409		0.291
R41T14C	0.436	0.062	0.451		0.066	0.373	0.365		0.086
R11F06C	0.692	0.142	0.709		0.148	0.495	0.531		0.154
R11F07C	0.603	0.112	0.618		0.116	0.443	0.488		0.130
R11F08C	1.370	0.393	1.425		0.412	1.028	1.125		0.445
R11F09C	1.402	0.404	1.471		0.428	1.080	1.161		0.465
R11F10C	0.815	0.186	0.853		0.201	0.392	0.940		0.264
R11F12C	0.545	0.093	0.562		0.098	0.443	0.384		0.106
R41I01C	0.779	0.173	0.811		0.185	0.800	0.049		0.182
R41I03C	0.456	0.067	0.460		0.068	0.479	0.001		0.073
R41I04C	0.622	0.118	0.621		0.118	0.642	0.011		0.125
R41I07C	0.939	0.233	0.937		0.233	0.933	0.154		0.236
R41I09C	1.113	0.299	1.125		0.304	1.178	-0.012		0.324
R41I11C	0.875	0.209	0.897		0.217	0.903	0.091		0.221
R41I13C	1.121	0.303	1.120		0.302	1.193	0.000		0.330
R41I14C	1.108	0.298	1.117		0.301	1.081	0.196		0.294
R41I15C	0.918	0.226	0.939		0.234	0.898	0.250		0.231
R41O02C	1.165	0.319	1.195		0.330	1.140	0.348		0.329
R41O03C	1.147	0.312	1.172		0.322	1.110	0.359		0.320
R41O04C	0.452	0.066	0.469		0.071	0.571	-0.166		0.109
R41O05C	0.514	0.084	0.526		0.087	0.559	-0.002		0.097
R41O07C	0.824	0.190	0.849		0.199	0.805	0.241		0.196
R41O08C	0.716	0.150	0.722		0.153	0.835	-0.167		0.200
R41O09C	0.945	0.236	0.987		0.252	1.038	-0.021		0.271
R41O10C	0.660	0.131	0.681		0.138	0.654	0.188		0.138
R41O13C	0.463	0.069	0.476		0.073	0.427	0.250		0.078
i2091	0.912	0.223		0.967	0.244	0.919		0.019	0.226
i2092	2.090	0.601		2.320	0.650	2.012		0.307	0.589
i2094	0.309	0.032		0.341	0.039	0.345		-0.031	0.040
i2095	0.683	0.139		0.765	0.168	0.685		0.127	0.144
i2096	1.030	0.268		1.027	0.267	1.069		-0.087	0.284
Le6PZ1	1.312	0.373		1.484	0.432	1.439		0.100	0.418
Le6PZ3	1.657	0.487		2.031	0.588	1.960		0.278	0.575
Le6PZ6	0.729	0.155		0.819	0.188	1.276		2.418	0.721
Le6PZ9	1.571	0.460		1.801	0.528	3.600		4.532	0.920
Le6PZ7	0.998	0.256		1.169	0.321	1.167		0.096	0.321
Le1VL1	0.381	0.048		0.447	0.065	0.506		-0.245	0.098
Le1VL4	0.791	0.178		0.967	0.244	1.022		-0.191	0.272
Le1VL3	1.182	0.325		1.370	0.393	1.469		-0.173	0.430
Le1VL8	0.862	0.204		1.038	0.271	1.098		-0.117	0.296
Le1VL7	0.832	0.193		0.961	0.242	1.033		-0.365	0.293
Le5DMS1	1.041	0.272		1.293	0.366	1.056		0.776	0.372
Le5DMS4	0.977	0.248		1.198	0.331	0.939		0.609	0.302
Le5DMS2	1.215	0.338		1.545	0.452	1.239		0.955	0.458
Le5DMS6	0.949	0.237		1.166	0.319	0.945		0.951	0.383
Le5DMS3	1.192	0.329		1.630	0.478	1.358		1.324	0.554
Le5DMS7	1.143	0.311		1.479	0.430	1.232		0.941	0.454
Le1SP6	0.865	0.205		0.998	0.256	0.834		0.921	0.348
Le1SP2	1.066	0.282		1.323	0.377	1.127		1.202	0.484

Item	1-factor model		2-factor model			Bifactor model			
	a	h ²	a ₁	a ₂	h ²	a	n ₁	n ₂	h ²
Le1SP3	1.028	0.267		1.134	0.307	1.014		0.300	0.279
Le1SP5	1.158	0.316		1.444	0.419	1.181		0.764	0.406
Le1SP7	1.035	0.270		1.199	0.332	0.964		0.601	0.308
Gem. h ²	0.271		0.293			0.326			
AIC	136555.1	136312.7			136149.3				
BIC	138159.5	137923.6			138468.9				
r _{factor1*factor2}			0.765						

Opmerking: a = factorlading, a₁ = factorlading op factor 1, a₂ = factorlading op factor 2, n₁ = lading op storingsfactor 1, n₂ = lading op storingsfactor 2, h² = verklaarde variantie item

7.6.2.1 Samenvatting en conclusie

In deze sectie onderzochten we welk IRT-model het best de antwoorden op toetsvragen verklaart in de zes toetsboekjes die zowel toetsvragen van PIRLS als het Peilingsonderzoek bevatten. De resultaten zijn wisselend van aard.

Volgens de RMSEA en SRMSR verklaren de drie modellen steeds in voldoende mate de antwoorden. Voor drie van de zes toetsboekjes voldoet het 1-factor IRT-model aan de drempelwaarde van de CFI, en voor slechts een van de zes toetsboekjes voldoet het 1-factor IRT-model aan de drempelwaarde van de TLI. De gemiddelde waarden voor de CFI en het TLI van het 1-factor IRT-model zijn respectievelijk 0.946 en 0.940, wat dan weer relatief dicht is bij de drempelwaarden. Ter vergelijking, in de vier toetsboekjes met enkel toetsvragen van PIRLS voldoet het 1-factor IRT-model wel steeds aan de drempelwaarden voor de CFI en TLI. De gemiddeldes zijn respectievelijk 0.983 en 0.981. In vijf van de zes toetsboekjes voldoet het 2-factor IRT-model aan de drempelwaarden van de CFI en TLI, terwijl het bifactor IRT-model altijd voldoet. De BIC verkiest steeds het 2-factor IRT-model, terwijl de AIC wisselt tussen het 2-factor IRT-model en het bifactor IRT-model. Gemiddeld genomen verklaart het 1-factor IRT-model 87.68% van de variantie van het 2-factor IRT-model in de zes toetsboekjes met toetsvragen van PIRLS en het Peilingsonderzoek. Ter vergelijking, in de vier toetsboekjes met enkel toetsvragen van PIRLS verklaart het 1-factor IRT-model 94.13% van de variantie in de antwoorden tegenover het 2-factor IRT-model. Dit is geen merkbaar verschil. De gemiddelde correlatie tussen de twee factoren in het 2-factor IRT-model is 0.753 in de zes toetsboekjes met toetsvragen van PIRLS en het Peilingsonderzoek. Ter vergelijking, in de vier toetsboekjes met enkel toetsvragen van PIRLS is de gemiddelde correlatie tussen de twee factoren in het 2-factor IRT-model 0.887. Ook dit is geen merkbaar verschil.

De resultaten van deze sectie tonen dat het twee-factor IRT-model en het bifactor IRT-model vaak het best de variantie in de zes toetsboekjes met toetsvragen van PIRLS en het Peilingsonderzoek verklaren. We stellen echter ook vast dat het 1-factor IRT-model de drempelwaarden van de absolute fit indices eveneens overschrijdt of ten minste dicht benadert. Ook zien we dat het overgrote deel van de variantie die verklaard wordt door het 2-factor IRT-model ook verklaard wordt door het 1-factor IRT-model. Wanneer we vergelijken met de vier toetsboekjes die enkel toetsvragen van PIRLS bevatten vinden we gelijkaardige resultaten. We concluderen dan ook dat verondersteld mag worden dat de zes toetsboekjes met toetsvragen van PIRLS en het Peilingsonderzoek hoofdzakelijk één onderliggende vaardigheid meten.

7.6.3 Predictieve validiteit volgens 2-factor IRT-model

De predictieve validiteit van de toetsvragen van PIRLS en het Peilingsonderzoek werd onderzocht door in het 2-factor IRT-model beide factoren met exogene variabelen te laten correleren. We bespreken hier vier exogene variabelen: leesprestaties in 2016, SES, geslacht en thuistaal. In de zes toetsboekjes met toetsvragen van zowel PIRLS als het Peilingsonderzoek is er namelijk één factor voor de toetsvragen van PIRLS en één factor voor de toetsvragen van het Peilingsonderzoek. Als de correlaties tussen de exogene factoren en beide factoren hetzelfde zijn, dan is dit een indicatie dat beide vormen van begrijpend lezen hetzelfde voorspellen. Als illustratie bekijken we ook de correlaties van exogene variabelen met de factoren in het 2-factor IRT-model van de vier toetsboekjes met enkel toetsvragen van PIRLS. In dit geval heeft elk van twee teksten van PIRLS een aparte factor. Ten slotte worden deze correlaties ook getoetst in een 2-factor IRT-model op basis van de gegevens van alle toetsboekjes. Hierin wordt een factor voor PIRLS en een factor voor het Peilingsonderzoek onderscheiden.

In **Tabel 86** worden de correlaties weergegeven van de twee factoren in het 2-factor IRT-model met de prestaties van leerlingen in PIRLS 2016. In **Tabel 87** worden de correlaties weergegeven van de twee factoren in het 2-factor IRT-model met de SES van leerlingen. In **Tabel 88** wordt het verschil tussen jongens en meisjes weergegeven volgens elk van de twee factoren in het 2-factor IRT-model. Een waarde hoger dan 0 geeft aan dat meisjes gemiddeld hoger scoren dan jongens. In **Tabel 89** wordt het verschil tussen leerlingen die thuis altijd Nederlands spreken en leerlingen die thuis soms Nederlands spreken weergegeven volgens elk van de twee factoren in het 2-factor IRT-model. Een waarde hoger dan 0 geeft aan dat leerlingen die thuis altijd Nederlands spreken gemiddeld hoger scores. In de tabellen wordt ook het verschil weergegeven in de correlaties of de verschillen tussen groepen naargelang welk van de twee factoren gebruikt is. De resultaten worden eerst getoond voor het model op basis van de gegevens van ieder toetsboekje apart. In de laatste rij van iedere tabel worden de resultaten getoond van het model op basis van de gegevens van alle toetsboekjes samen

Voor de correlatie met de prestaties voor begrijpend lezen in PIRLS 2016 vinden we voor de zes toetsboekjes met toetsvragen van PIRLS en het Peilingsonderzoek drie significante verschillen (boekje 3, boekje 9 en boekje 10). In elk van deze drie gevallen heeft factor 2 een hogere correlatie, dit zijn de toetsvragen van het Peilingsonderzoek. Dit betekent dat de score begrijpend lezen die de leerling behaalde in het vierde leerjaar (PIRLS2016) sterker samenhangt met zijn score op de peilingsitems (in boekje 3, 9 of 10) dan met zijn score op de PIRLS-items (in boekje 3, 9 of 10). Dit is een onverwacht resultaat dat eigenlijk extra aangeeft dat PIRLS en de peiling lezen een gelijkaardige vaardigheid meten. Bij de vier toetsboekjes die enkel toetsvragen van PIRLS bevatten is er eenmaal een significant verschil (boekje 6). Gemiddeld over de zes toetsboekjes met toetsvragen van PIRLS en het Peilingsonderzoek vinden we echter geen significant verschil tussen de factoren. Ook wanneer alle gegevens in één model opgenomen worden vinden we geen significant verschil.

De correlatie tussen de leesprestaties in het vierde leerjaar (2016) en de twee factoren van leesprestaties in het zesde leerjaar bedraagt 0.690 en 0.670. Dit zijn relatief hoge correlaties die aangeven dat leesvaardigheid redelijk stabiel blijft over een periode van 2 jaren.

Tabel 86: Correlaties factor 1 en factor 2 met prestaties begrijpend lezen (ASRREA) in PIRLS 2016

Toetsboekje	r_{f1}	r_{f1} se	r_{f2}	r_{f2} se	Verskil	Verskil se
1	0.596	0.052	0.509	0.052	0.087	0.068
2	0.680	0.041	0.609	0.041	0.071	0.044
3	0.664	0.037	0.754	0.029	-0.090	0.045
4	0.534	0.049	0.506	0.055	0.027	0.059
5	0.678	0.041	0.687	0.036	-0.008	0.040
6	0.678	0.040	0.808	0.031	-0.131	0.043
7	0.668	0.037	0.656	0.039	0.012	0.039
8	0.651	0.040	0.706	0.033	-0.055	0.039
9	0.605	0.040	0.746	0.039	-0.141	0.046
10	0.471	0.058	0.663	0.047	-0.192	0.062
Alle toetsboekjes	0.690	0.012	0.670	0.016	0.020	0.018

Opmerking: r_{f1} = correlatie factor 1 met ASRREA, r_{f1} se = standaardfout correlatie factor 1 met ASRREA, r_{f2} = correlatie factor 2 met ASRREA, r_{f2} se = standaardfout correlatie factor 2 met ASRREA

Voor de correlatie met SES (*home resources for learning* (ASBGHRL)) vinden we voor de zes toetsboekjes met toetsvragen van PIRLS en het Peilingsonderzoek twee significante verschillen (boekje 3 en boekje 10). In beide gevallen heeft factor 2 een hogere correlatie, dit zijn de toetsvragen van het Peilingsonderzoek. Bij de vier toetsboekjes die enkele toetsvragen van PIRLS bevatten is er nooit een significant verschil. Gemiddeld over de zes toetsboekjes met toetsvragen van PIRLS en het Peilingsonderzoek vinden we dat de factor van de toetsvragen van het Peilingsonderzoek hoger correleren met SES. Ook wanneer alle gegevens in één model opgenomen worden vinden we dat SES hoger correleert met de factor van de toetsvragen van het Peilingsonderzoek.

Tabel 87: Correlaties factor 1 en factor 2 met SES (ASBGRHL)

	r_{f1}	$r_{f1\ se}$	r_{f2}	$r_{f2\ se}$	Verskil	Verskil se
Toetsboekje 1	0.274	0.073	0.246	0.060	0.028	0.072
Toetsboekje 2	0.446	0.053	0.433	0.060	0.013	0.059
Toetsboekje 3	0.314	0.054	0.491	0.049	-0.177	0.054
Toetsboekje 4	0.206	0.076	0.269	0.076	-0.064	0.076
Toetsboekje 5	0.242	0.063	0.317	0.052	-0.075	0.051
Toetsboekje 6	0.339	0.056	0.275	0.062	0.063	0.048
Toetsboekje 7	0.368	0.053	0.324	0.056	0.044	0.044
Toetsboekje 8	0.411	0.054	0.411	0.053	0.000	0.053
Toetsboekje 9	0.332	0.054	0.361	0.065	-0.029	0.050
Toetsboekje 10	0.210	0.066	0.439	0.057	-0.229	0.065
Alle toetsboekjes	0.349	0.017	0.415	0.022	-0.066	0.023

Opmerking: r_{f1} = correlatie factor 1 met ASBGRHL, $r_{f1\ se}$ = standaardfout correlatie factor 2 met ASBGRHL, r_{f2} = correlatie factor 2 met ASBGRHL, $r_{f2\ se}$ = standaardfout correlatie factor 2 met ASBGRHL

Voor het verschil tussen jongens en meisjes vinden we voor de zes toetsboekjes met toetsvragen van PIRLS en het Peilingsonderzoek dat de twee factoren hetzelfde weergeven. Een waarde hoger dan 0 geeft aan dat meisjes gemiddeld hoger scoren dan de jongens. We zien dat voor alle toetsboekjes en beide factoren dat meisjes hoger scoren dan jongens. Het verschil tussen jongens en meisjes bij de toetsboekjes die enkel toetsvragen van PIRLS bevatten is stabiel over beide teksten. Ook bij de toetsboekjes die zowel toetsvragen van PIRLS als de Peilingen bevatten is het verschil tussen jongens en meisjes stabiel over de teksten. Gemiddeld over de zes toetsboekjes met toetsvragen van PIRLS en het Peilingsonderzoek vinden we dat de twee factoren hetzelfde weergeven. Ook wanneer alle gegevens in één model opgenomen worden vinden dat beide factoren op een gelijk verschil tussen jongens en meisjes uitkomen.

Tabel 88: Verschil in factor 1 en factor 2 naargelang geslacht (ITSEX)

	$Effect_{f1}$	$Effect_{f1\ se}$	$Effect_{f2}$	$Effect_{f2\ se}$	Verskil	Verskil se
Toetsboekje 1	0.349	0.133	0.312	0.117	0.037	0.139
Toetsboekje 2	0.338	0.113	0.403	0.111	-0.065	0.112
Toetsboekje 3	0.056	0.102	0.158	0.117	-0.102	0.105
Toetsboekje 4	0.313	0.110	0.302	0.125	0.011	0.129
Toetsboekje 5	0.264	0.116	0.306	0.108	-0.042	0.100
Toetsboekje 6	0.083	0.113	0.202	0.112	-0.118	0.097
Toetsboekje 7	0.265	0.106	0.443	0.101	-0.178	0.093
Toetsboekje 8	0.285	0.123	0.094	0.099	0.192	0.107
Toetsboekje 9	0.208	0.102	0.221	0.120	-0.013	0.103
Toetsboekje 10	0.260	0.123	0.266	0.119	-0.006	0.130
Alle toetsboekjes	0.226	0.033	0.270	0.045	-0.044	0.044

Opmerking: $effect_{f1}$ = effect op factor 1 van ITSEX, $effect_{f1\ se}$ = standaardfout effect op factor 1 van ITSEX, $effect_{f2}$ = effect op factor 2 van ITSEX, $effect_{f2\ se}$ = standaardfout effect op factor 2 van ITSEX

Voor het verschil tussen leerlingen die thuis altijd Nederlands spreken en leerlingen die thuis soms Nederlands spreken vinden we voor de zes toetsboekjes met toetsvragen van PIRLS en het Peilingsonderzoek dat de twee factoren verschillen. In twee toetsboekjes (boekje 2 en boekje 9) is het effect van thuistaal significant groter voor factor 2, de factoren voor de toetsvragen van het Peilingsonderzoek. Bij de vier toetsboekjes die enkele toetsvragen van PIRLS bevatten vinden we dat de twee factoren hetzelfde uitkomen. Gemiddeld over de zes toetsboekjes met toetsvragen van PIRLS en het Peilingsonderzoek vinden we dat de factor van de toetsvragen van het Peilingsonderzoek een groter verschil tussen leerlingen weergeeft. Ook wanneer alle gegevens in één model opgenomen worden vinden we dat het verschil tussen beide groepen leerlingen groter is voor de factor van het Peilingsonderzoek. Dat betekent dat thuistaal (zoals gemeten in de leerlingvragenlijst in 2016) sterker samenhangt met de score op de peilingsitems in 2018 dan met de PIRLS-items in 2018.

Tabel 89: Verschil in factor 1 en factor 2 naargelang thuistaal (ASB03) tussen leerlingen die thuis altijd Nederlands spreken en leerlingen die thuis soms Nederlands spreken.

	Effect Factor 1	Effect Factor 1 se	Effect Factor 2	Effect Factor 2 se	Verschil	Verschil se
Toetsboekje 1	0.547	0.157	0.293	0.151	0.255	0.163
Toetsboekje 2	0.229	0.143	0.510	0.159	-0.281	0.141
Toetsboekje 3	0.341	0.145	0.492	0.146	-0.151	0.143
Toetsboekje 4	0.141	0.145	0.313	0.154	-0.172	0.167
Toetsboekje 5	0.227	0.163	0.399	0.144	-0.177	0.134
Toetsboekje 6	0.162	0.177	0.330	0.175	-0.168	0.142
Toetsboekje 7	0.290	0.145	0.320	0.128	-0.030	0.112
Toetsboekje 8	0.602	0.152	0.669	0.137	-0.068	0.151
Toetsboekje 9	-0.013	0.159	0.336	0.169	-0.348	0.141
Toetsboekje 10	0.502	0.215	0.650	0.182	-0.148	0.217
Alle toetsboekjes	0.384	0.047	0.530	0.061	-0.147	0.060

7.6.3.1 Samenvatting en conclusie

In deze sectie onderzochten we of de factor op basis van de toetsvragen van PIRLS en de factor op basis van de toetsvragen van het peilingsonderzoek dezelfde relaties hebben met vier exogene variabelen. De relaties met prestaties voor begrijpend lezen in 2016, SES, geslacht en thuistaal werden onderzocht. Het gebruikte model was het 2-factor IRT-model, waarbij beide factoren vrij konden correleren met de exogene variabele. We beschouwen dit als een test van predictieve validiteit.

De resultaten tonen dat begrijpend lezen zoals gemeten door de toetsvragen van PIRLS en begrijpend lezen zoals gemeten door de toetsvragen van het Peilingsonderzoek dezelfde samenhang hebben met begrijpend lezen in 2016 en geslacht. SES en thuistaal hebben echter een enigszins grotere samenhang met de factor voor begrijpend lezen zoals gemeten door de toetsvragen van het Peilingsonderzoek. Het verschil is wel relatief beperkt. Daarom concluderen we dat als er bepaalde relaties zijn tussen begrijpend lezen en andere variabelen in het PIRLS-onderzoek, deze waarschijnlijk ook teruggevonden zullen worden in het Peilingsonderzoek. Omgekeerd geldt dit natuurlijk ook. Er is dan ook evidentie dat voorspellingen op basis van PIRLS en het Peilingsonderzoek gelijklopend zijn.

7.7 Algemene conclusie

In dit hoofdstuk onderzochten we of de toetsvragen van PIRLS 2016 en de toetsvragen van het Peilingsonderzoek Nederlands 2018 dezelfde onderliggende vaardigheid voor begrijpend lezen meten. Daarom werd getest of de toetsboekjes van de herhalingsmeting van PIRLS in 2018 één onderliggende vaardigheid of twee onderliggende vaardigheden voor begrijpend lezen meten. Daarom voerden we exploratieve factoranalyses uit en vergeleken we verschillende IRT-modellen. We bekeken de constructvaliditeit aan de hand van de vraag hoe de toetsvragen onderling samenhangen. Daarna bekeken we de predictieve validiteit door na te gaan of de toetsvragen van PIRLS en de toetsvragen van het Peilingsonderzoek iets anders voorspellen.

De resultaten duiden erop dat de samenhang tussen de verschillende toetsvragen voornamelijk naar één onderliggende vaardigheid terug te brengen is. Een model met twee onderliggende vaardigheden of een model waarbij gecorrigeerd wordt voor de extra samenhang tussen toetsvragen van PIRLS onderling en de extra samenhang tussen toetsvragen van het Peilingsonderzoek onderling passen wel iets beter. Toch zien we dat het model met één onderliggende vaardigheid het merendeel van de variantie in deze modellen al verklaart. We vinden ook dat de vaardigheid in begrijpend lezen zoals geoperationaliseerd door PIRLS en zoals geoperationaliseerd door het Peilingsonderzoek een gelijkaardige samenhang hebben met vier leerlingkenmerken. We besluiten dus dat het aanvaardbaar is om bij de gegevens van de herhalingsmeting, die bestaat uit toetsvragen van zowel PIRLS als het Peilingsonderzoek, één onderliggende vaardigheid te veronderstellen.

Deze conclusie voor Vlaanderen sluit aan bij onderzoek uit Nederland (Van Diepen, 2007) waar 1228 leerlingen uit 62 scholen enerzijds 4 PIRLS-teksten lezen (uit PIRLS2001) en anderzijds de CITO-toets begrijpend lezen maakten. Er werd vastgesteld dat de vijf metingen van begrijpend lezen (4 PIRLS-teksten + CITO-test) zodanig sterk samenhangen dat ze samengevat kunnen worden in één latente variabele. Dus ook in Nederland werd vastgesteld dat de meting van begrijpend lezen aan de hand van een internationale toets sterk samenhangt met de meting van begrijpend lezen met een landspecifieke toets.

7.8 PIRLS 2016 op de meetschaal van het Peilingsonderzoek.

Dat de herhalingsmeting van PIRLS in 2018 (de 111 items) als unidimensioneel beschouwd kan worden toont ook dat PIRLS 2016 op de schaal van het Peilingsonderzoek begrijpend lezen geplaatst kan worden. In **Hoofdstuk 3** verwezen we er kort naar hoe we PIRLS 2016 op de schaal van het Peilingsonderzoek hadden geplaatst. Een onderzoeksvraag die we in deze studie immers willen beantwoorden is hoeveel leerlingen van het vierde jaar reeds de eindtermen voor begrijpend lezen bereikt hebben.

Het is echter niet vanzelfsprekend om de resultaten voor begrijpend lezen in PIRLS 2016 te plaatsen op de schaal van het Peilingsonderzoek begrijpend lezen. PIRLS 2016 en het Peilingsonderzoek begrijpend lezen gebruiken immers verschillende IRT-modellen en een andere methode voor vaardigheidsschatting (plausible values en WLE). Daarbij is er natuurlijk ook nog een schattingsfout op de itemparameters. Zelfs al zijn de antwoorden op de toetsvragen van beide onderzoeken dus unidimensioneel te verklaren, het gebruik van verschillende methoden en de schattingsfout verhindert een eenvoudige transformatie van beide schalen. Echter, omdat de leerlingen in PIRLS 2016 zowel een vaardigheidsschatting hebben op basis van de PIRLS-meetschaal als de Peilingmeetschaal, kunnen we goed benaderen hoe beide meetschalen zich ten opzichte van elkaar verhouden. In de volgende paragrafen beschrijven we twee benaderingen om de resultaten van PIRLS 2016 op de meetschaal van het Peilingsonderzoek begrijpend lezen te plaatsen.

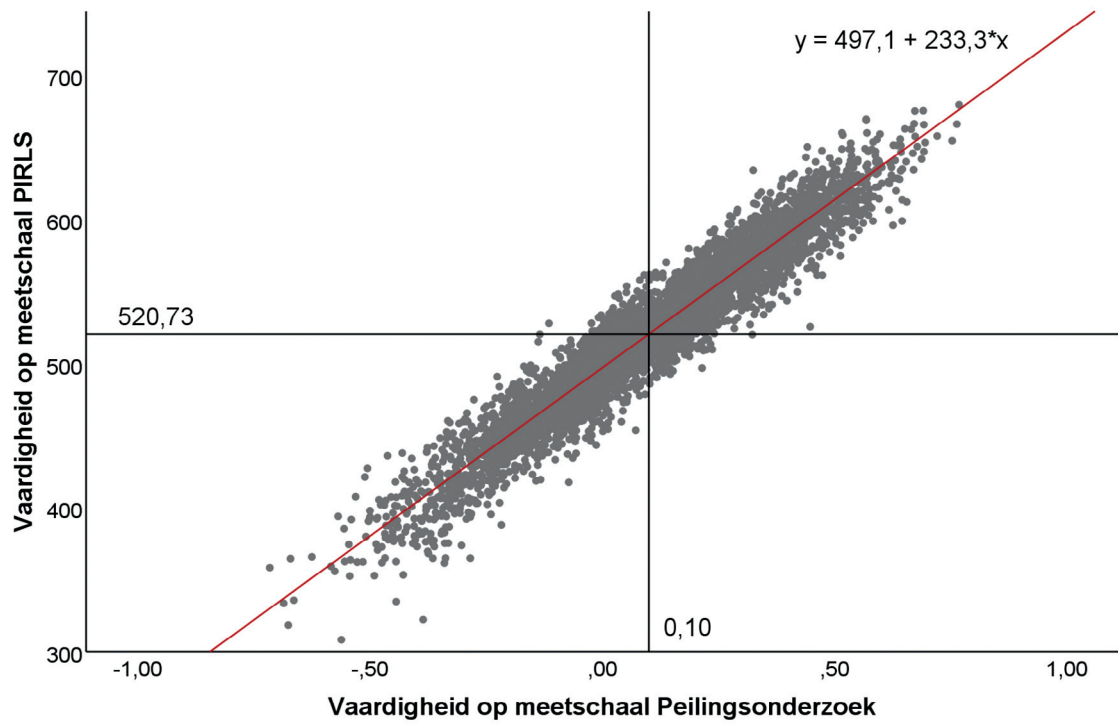
Een eerste benadering voor het plaatsen van PIRLS 2016 op de schaal van het Peilingsonderzoek is via de latent geschatte distributies van de IRT-modellen. Hierbij berekenen we wat de Z-waarde is van de cesuur op de meetschaal van het Peilingsonderzoek, om vervolgens deze Z-waarde om te zetten naar de meetschaal van PIRLS. Met deze berekeningsmethode vinden we dat de cesuur met waarde 0.10 op de meetschaal van het Peilingsonderzoek overeenkomt met een waarde van 520.73 op de meetschaal van PIRLS. Een nadeel van deze benadering is dat dit berekend is op basis van latent geschatte distributies, terwijl de onderzoeksresultaten van PIRLS en het Peilingsonderzoek steeds op basis van geschatte vaardigheden plaatsvinden.

Een tweede benadering is door middel van een lineaire regressie. Hierbij worden de geschatte vaardigheden van de leerlingen op de meetschaal van PIRLS voorspeld door de geschatte vaardigheden van deze leerlingen op de meetschaal van het Peilingsonderzoek. Voor de geschatte vaardigheden op de meetschaal van PIRLS wordt het gemiddelde per leerling over de vijf plausible values gebruikt. Vervolgens voorspellen we wat voor de cesuur van 0.10 op de meetschaal van het Peilingsonderzoek de voorspelde waarde is op de meetschaal van PIRLS. Met deze methode vinden we dat de cesuur van 0.10 op de meetschaal van het Peilingsonderzoek overeenkomt met een waarde van 520.43 op de meetschaal van PIRLS. De scatterplot van de gegevens worden getoond in **Figuur 59**. Hierbij staan ook de cesuur volgens het Peilingsonderzoek begrijpend lezen weergegeven (verticale lijn), de voorspelde cesuur voor PIRLS 2016 en de voorspelde regressielijn. Een nadeel van deze benadering is dat door de attenuatie van de regressiecoëfficiënt de relatie tussen de vaardigheden volgens beide meetschalen onderschat zal zijn.

Beide benaderingen om te bepalen waar de cesuur van het Peilingsonderzoek ligt op de meetschaal van PIRLS komen dus tot bijna dezelfde uitkomst. Als we de benadering met lineaire regressie gebruiken dan vinden we dat 53.25% van de leerlingen op het einde van het vierde leerjaar de eindtermen bereikt hebben. Dit zijn alle leerlingen die zich rechts bevinden van de verticale lijn (de cesuur) in **Figuur 59**. Wanneer we gewichten van PIRLS 2016 gebruiken, dan vinden we dat 53.99% van de leerlingen op het einde van het vierde leerjaar de eindtermen bereikt hebben. Deze uitspraken zijn op basis van de steekproef van 5198 leerlingen van PIRLS 2016. De cesuur met een waarde van 520.43 op de meetschaal van PIRLS ligt tussen de middelmatige standaard (*intermediate benchmark*, 475) en hoge standaard (*high benchmark*, 550) van PIRLS voor het vierde leerjaar.

We merken op dat de resultaten hier natuurlijk schattingen zijn waarbij verondersteld wordt dat PIRLS en het Peilingsonderzoek dezelfde vaardigheid voor begrijpend lezen meten. Zoals beschreven in dit hoofdstuk vonden we dat PIRLS en het Peilingsonderzoek grotendeels dezelfde vaardigheid meten. PIRLS en het Peilingsonderzoek meten echter niet exact dezelfde vorm van begrijpen lezen. er kan dus een kleine vertekening zijn op dit resultaat.

Figuur 59: regressievergelijking geschatte vaardigheden volgens Peilingsonderzoek en PIRLS 2016



8 Algemeen besluit

8.1 Zesde leerjaar in Vlaanderen op niveau vierde leerjaar in toppersterende landen

>> De resultaten van de herhalingsmeting van PIRLS in 2018 tonen dat de Vlaamse leerlingen in het zesde leerjaar een hoger niveau bereiken dan de leerlingen van andere landen in het vierde leerjaar. De Vlaamse leerlingen zijn daarmee in het zesde leerjaar op ongeveer hetzelfde niveau als de leerlingen van toppersterende landen (Rusland, Singapore, Hong Kong, Ierland) in het vierde leerjaar. Opvallend daarbij is de kleine spreiding van prestatieniveaus tussen Vlaamse leerlingen. Alle Vlaamse leerlingen in het zesde leerjaar behalen zo minstens het lage niveau voor begrijpend lezen van de IEA (*low benchmark*). Er zijn echter relatief minder leerlingen die het gevorderd niveau (*advanced benchmark*) bereiken in het zesde leerjaar dan de leerlingen in het vierde leerjaar van de toppersterende landen.

Verschillen die bestaan tussen verschillende leerlingengroepen in het vierde leerjaar zijn doorgaans ook aanwezig in het zesde leerjaar. Daarbij blijft de grootte van het verschil tussen jongens en meisjes hetzelfde. Verschillen in prestaties voor begrijpend lezen naargelang SES van leerlingen, of lezen leuk vinden of hoe zelfzeker ze zijn tegenover lezen zijn dan weer afgenomen.

8.2 Vlaanderen maakt geen inhaalbeweging tussen het vierde en zesde leerjaar, integendeel

De tegenvallende resultaten voor Vlaanderen in PIRLS 2016 toonden dat Vlaanderen de grootste achteruitgang had in begrijpend lezen tussen 2006 en 2016. Er werd toen gehoopt dat de Vlaamse leerlingen tussen het vierde en zesde leerjaar een grote inhaalbeweging maken, omdat in de derde graad de focus dan van technisch naar begrijpend lezen verplaatst wordt. Daarom onderzochten we ook hoe groot de vooruitgang is van de Vlaamse leerlingen voor begrijpend lezen tussen het vierde en zesde leerjaar.

Deze studie vindt echter niet dat Vlaanderen een inhaalbeweging maakt voor begrijpend lezen tussen het vierde en zesde leerjaar. De leerwinst op jaarbasis is daarentegen zelfs kleiner dan in enkele andere landen. Ook wanneer we een historische vergelijking maken tussen de resultaten van deze studie en resultaten voor Vlaanderen tien jaar geleden (het SIBO-onderzoek), maken we nu minder leerwinst tussen het vierde en zesde leerjaar. Onze conclusie is dan ook dat de hele duur van Vlaamse lager onderwijs gekenmerkt is door weinig leerwinst in begrijpend lezen.

Er zijn ook verschillen tussen groepen van leerlingen in leerwinst voor begrijpend lezen. Zo maken meisjes meer leerwinst tussen het vierde en zesde jaar dan jongens. Leerlingen die thuis altijd Nederlands spreken hebben dan weer minder leerwinst dan leerlingen die thuis soms Nederlands spreken. Gemiddeld genomen zien we ook dat leerlingen die laag scoorden in 2016 meer leerwinst maken dan leerlingen die hoog scoorden in 2016. We vinden echter geen verschil in leerwinst naargelang de SES-status van leerlingen, hoeveel ze van lezen houden, hoe sterk betrokken ze zijn bij lezen en hoeveel de ouders van lezen houden. Leerlingen die niet zelfzeker waren in hun leesvaardigheid in 2016 maken dan weer meer leerwinst.

8.3 Attitudes van leerlingen ten opzichte van lezen worden doorgaans negatiever

PIRLS 2016 toonde dat Vlaamse leerlingen in het vierde leerjaar relatief weinig zelfzeker waren in lezen en een relatief negatieve houding hadden tegenover lezen. De vraag werd dan ook gesteld hoe de attitudes van Vlaamse leerlingen ten opzichte van lezen veranderen tussen het vierde en zesde leerjaar.

De resultaten tonen dat Vlaamse leerlingen zich in het zesde leerjaar nog minder betrokken voelen bij lezen en een nog negatievere houding hebben ten aanzien van lezen. Enkel hun zelfzekerheid in hun leesvaardigheid blijft stabiel. De evolutie in de attitudes is dus eerder negatief van aard.

Aansluitend zijn er ook verschillen gevonden in deze evoluties in attitudes tussen diverse leerlingengroepen. Deze verschillen hebben echter een weinig consistent patroon.

8.4 PIRLS meet begrijpend lezen op een manier die relevant is voor Vlaanderen

Een vraag die bij de resultaten van PIRLS 2016 werd gesteld is of begrijpend lezen werd gemeten op een manier die relevant is voor de Vlaamse context. Eén van de doelen van de herhalingsmeting van PIRLS in 2018 was dan ook om te onderzoeken of begrijpend lezen zoals gemeten door PIRLS relevant is voor Vlaanderen. Daarom werden in zes toetsboekjes zowel vragen opgenomen van PIRLS als vragen van de peiling lezen. Het Peilingsonderzoek richt zich namelijk specifiek op de vaardigheid in begrijpend lezen volgens de Vlaamse eindtermen. Door toetsvragen van beide studies op te nemen kunnen we onderzoeken hoe sterk beide studies dezelfde vorm van begrijpend lezen meten.

De resultaten wijzen erop dat de antwoorden op de toetsvragen van PIRLS en de toetsvragen van het Peilingsonderzoek begrijpend lezen voornamelijk naar één onderliggende vaardigheid te herleiden zijn. Hoewel de vragen van PIRLS enerzijds en de vragen van het Peilingsonderzoek anderzijds elk een unieke samenhang kennen is deze unieke samenhang relatief beperkt. Ook vinden we dat wat voorspeld wordt op basis de vragen van PIRLS en wat voorspeld wordt op basis de vragen van het peilingsonderzoek begrijpend lezen zeer gelijkend is. PIRLS en het Peilingsonderzoek meten dus hoofdzakelijk dezelfde vaardigheid in begrijpend lezen. Hoe PIRLS begrijpend lezen meet is dus relevant voor de Vlaamse context.

Dankwoord

Op 16 februari 2018 kreeg ons onderzoekscentrum de vraag om in mei 2018 toetsen begrijpend lezen af te nemen van de Vlaamse PIRLS-leerlingen.

Tijdens de weken die daarop volgden, toonden diverse mensen hun organisatorische ‘superkrachten’. Jan Van Damme belde alle 141 PIRLS-scholen persoonlijk op. Nathalie Vandenberghe organiseerde in ijtempo de dataverzameling, bijgestaan door Sabine Rouffaert, Ineke Stienaers en Kelly Tielemans. Zij verwerkten alle klaslijsten, en zorgden voor een heldere databank van leerlingen, leerkrachten en scholen. Deze dataverzameling was een logistiek huzarenstukje dat nacht- en weekendwerk vroeg, maar die lukte dankzij de inzet van al deze mensen. We willen ook drukkerij Acco bedanken voor het drukken van de stapels toetsboekjes en vragenlijsten. We willen de jobstudenten bedanken voor hun grote verantwoordelijkheidsgevoel om in de zomer van 2018 de gegevens in te voeren en de antwoorden te scoren. We willen hier ook onze teleurstelling uitdrukken over de gebrekkige werking van Bpost. Zij slaagden er in om vijf pakketten met toetsen en vragenlijsten kwijt te raken, waardoor de betrokken scholen een onvolledig feedbackrapport kregen.

Bij het opstellen van de toetsen begrijpend lezen kregen we statistische ondersteuning van Jonas Dockx. Ook de collega's van het Peilingsonderzoek (Rianne Janssen, Mieke Heyvaert, Koen Aesaert, Jo Denis) willen we bedanken om hun toetsen en vragenlijsten met ons te delen.

We wensen ook de IEA (International Association for the Evaluation of Educational Achievement) te bedanken voor hun toelating om zes teksten en vragen uit PIRLS2016 te hergebruiken in dit vervolgonderzoek, zonder enige financiële compensatie. Bedankt aan Dirk Hastedt en Andrea Netten van IEA.

Bij het opstellen van de vragenlijst voor de leerkrachten konden we een beroep op de expertise van diverse taalspecialisten: Hilde Van Keer, Thoni Houtveen, Kris Van den Branden, Marieke Vanbuel, en Simon Bequoye.

In mei-juni 2018, legden bijna vijfduizend leerlingen uit 127 lagere scholen de toetsen af. We willen alle leerlingen, leerkrachten en directeurs héél erg bedanken om in dat drukke laatste trimester van het zesde leerjaar nog tijd te maken om aan dit onderzoek deel te nemen. Het was helemaal niet evident om last-minute nog zo'n grootschalig onderzoek in te plannen, maar jullie beseften dat jullie daarmee een bijdrage konden leveren aan het begrijpen van het Vlaamse (lees)onderwijs. Heel erg bedankt!

De complexe analyses van dit atypische project lagen in de deskundige handen van Jonas Dockx en Georges Van Landeghem. Hun grondige aanpak leidde tot dit uitgebreide rapport en tot feedbackrapporten voor elk van de 127 deelnemende lagere scholen.

In de loop van dit project hebben we met zeer veel mensen gepraat over taalonderwijs, leesonderwijs, het toetsen van begrijpend lezen, de waarde van internationaal vergelijkend onderzoek, ... Dat overleg vond onder meer plaats in de IVOO-stuurgroep. Studente Heleen Hoebrechts interviewde 10 Vlaamse experts. Al die gesprekken en discussies hebben bijgedragen aan dit onderzoeksrapport.

Kortom: onderzoek doe je nooit alleen. We willen onderwijsonderzoekers zijn die een vinger aan de pols houden en daarmee bijdragen aan de kwaliteit van het Vlaams onderwijs. We willen iedereen bedanken die inzichten en expertise met ons heeft gedeeld, die met ons in discussie is gegaan en vragen heeft gesteld. We hopen met dit rapport ons steentje bij te dragen aan een onderbouwd onderwijsdebat.

Leuven, juni 2019
Bieke De Fraine, promotor
Koen Asaert, copromotor
Jan Van Damme, copromotor

Referenties

- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York City, NY: Guilford Press.
- Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48(6), 1–29.
- Dorans, N. J., Moses, T. P., & Eignor, D. R. (2010). Principles and practices of test score equating. *ETS Research Report Series*, 2010(2), i--41.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Harrison, D. A. (1986). Robustness of IRT parameter estimation to violations of the unidimensionality assumption. *Journal of Educational Statistics*, 11(2), 91–115.
- Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9(2), 139–164.
- Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, 2(3), 172–177.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55.
- IEA. (2019). About us. Retrieved from <http://www.iea.nl/about-us>
- Kang, T., & Cohen, A. S. (2007). IRT model selection methods for dichotomous items. *Applied Psychological Measurement*, 31(4), 331–358.
- Kline, P. (1986). *A handbook of test construction : introduction to psychometric design*. London, United Kingdom: Methuen.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). New York City, NY: Springer Science & Business Media.
- Li, D., Jiang, Y., & von Davier, A. A. (2012). The accuracy and consistency of a series of IRT true score equatings. *Journal of Educational Measurement*, 49(2), 167–189.
- Luyten, H., Merrell, C., & Tymms, P. (2017). The contribution of schooling to learning gains of pupils in years 1 to 6. *School Effectiveness and School Improvement*, 28(3), 374–405.
- Martin, M. O., Mullis, I. V. S., & Hooper, M. (2017). *Methods and Procedures in PIRLS 2016*. Boston, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Mislevy, R. J. (1991). Randomization-based inference about latent variables from complex samples. *Psychometrika*, 56(2), 177–196.
- Mullis, I. V. S., & Martin, M. O. (2015). *PIRLS 2016 Assessment Framework*. Boston, MA: TIMSS & PIRLS International Study Center, Lynch School of Education & IEA.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2017). *PIRLS 2016 International Results in Reading*. Boston, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Mullis, I. V. S., Martin, M. O., Goh, S., & Prendergast, C. (2017). *PIRLS 2016 Encyclopedia: Education Policy and Curriculum in Reading*. Boston, MA: TIMSS & PIRLS International Study Center, Lynch School of Education & IEA.
- Muthén, B., & Muthén, L. (2015). *Mplus Statistical Analysis With Latent Variables User's Guide*. Los Angeles, CA: Muthén & Muthén.
- Oliveri, M. E., Rutkowski, D., & Rutkowski, L. (2018). *Bridging validity and evaluation to match international large-scale assessment claims and country aims (research report RR-18-27)*. Princeton, NJ: ETS.
- R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, 92(6), 544–559.
- Retelsdorf, J., Becker, M., Köller, O., & Möller, J. (2012). Reading development in a tracked school system: A longitudinal study over 3 years using propensity score matching. *British Journal of Educational Psychology*, 82(4), 647–671.
- Retelsdorf, J., Köller, O., & Möller, J. (2011). On the effects of motivation on reading performance growth in secondary school. *Learning and Instruction*, 21(4), 550–559.
- Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality: Published for the European Association of Personality Psychology*, 21(5), 667–706.
- Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16(1), 32–71.
- Steunpunt Toetsontwikkeling en Peilingen. (2014). *Brochure Peiling Nederlands: lezen en luisteren*. Brussel, België.
- Steunpunt Toetsontwikkeling en Peilingen. (2019). *Brochure Peiling Nederlands: lezen, luisteren en schrijven in het basisonderwijs*. Brussel, België.

- Tielemans, K., Vandenbroeck, M., Bellens, K., Van Damme, J., & De Fraine, B. (2017). *Het Vlaams lager onderwijs in PIRLS 2016. Begrijpend lezen in internationaal perspectief en in vergelijking met 2006*. Leuven: CO&E. Retrieved from <https://onderwijs.vlaanderen.be/nl/progress-in-international-reading-literacy-study-pirls>
- Tielemans, K., Vanlaar, G., Van Damme, J., & De Fraine, B. (2019). *Lessen door en voor het Vlaams begrijpend leesonderwijs*. Leuven: CO&E. Retrieved from <https://data-onderwijs.vlaanderen.be/onderwijsonderzoek/?nr=178>
- Van Diepen, M. (2007). *Variation in reading literacy: A cross-national approach (Doctoral thesis)*. Nijmegen, the Netherlands: Universiteit Nijmegen.
- Verhelst, N. D., & Glas, C. A. W. (1995). The one parameter logistic model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, Recent Developments, and Applications* (pp. 215–237). New York City, NY: Springer.
- Vlor. (2018). *Advies over begrijpend lezen (PIRLS 2016) (RBO-ADV-1718-002)*. Brussel: VLOR. Retrieved from <https://www.vlor.be/adviezen/advies-over-begrijpend-lezen-pirls-2016>
- Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54(3), 427–450.
- Wingersky, M., Kaplan, B. A., & Beaton, A. E. (1987). Joint estimation procedures. In *Implementing the new design: The NAEP 1983–84 technical report (pp.285– 92) (No. 15-TR-20)*. Princeton, NJ: Educational Testing Service, National Assessment of Educational Progress.

Appendices

Appendix 1: Syntax model IRT met parameters van PIRLS in MIRT

```
## IRT-analyses PIRLS 2018 terugkerende studenten###

#Nodige pakketten
install.packages("mirt") #Installatie MIRT 1.30 software.
install.packages("psych") #installatie psych 1.8.12 software#
install.packages("Cairo") #installatie Cairo 1.5-10 software#

library(mirt) #Activatie MIRT 1.30 software.
library(psych) #activatie psych 1.8.12 software#
library(Cairo) #activatie Cairo 1.5-10 software#

#Inlezen bestanden via read.csv()
PIRLS18=read.csv("PIRLS2018.csv", sep=";")

#Her codering toetscodes 8 0.
PIRLS18[PIRLS18==8]<-NA

describe(PIRLS18[,c(2:112)])

total<-rowSums(PIRLS18[,c(2:112)], na.rm=TRUE) #Somscore 0/1 scoring
berekenen en in "Total" opslagen ter controle#
hist(total,breaks=30) #Histogram somscore ter controle#

#Inlezen classmeans PIRLS
classmeans=read.csv("PIRLS18_Classmeans.csv", sep=";")
PIRLS18<-merge(PIRLS18,classmeans, by="LeerlingID")

#Inlezen Gender PIRLS
Gender=read.csv("Gender.csv", sep=";")
PIRLS18<-merge(PIRLS18,Gender, by="LeerlingID")

#Inlezen componenten PIRLS
components=read.csv("Components.csv", sep=";")
names(components)[1]<- "LeerlingID"
PIRLS18<-merge(PIRLS18, components, by="LeerlingID")

#Recode R041I15C as described in appendix 10D
PIRLS18$R041I15C[PIRLS18$R041I15C==2]<-1

covdataset=PIRLS18[,c(114:374)]

Svalues<-mirt(PIRLS18[,c(2:112)], model=1,
itemtype=c('2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL', '2PL', 'gpcm', '2PL', 'gpcm', '2PL',
'2PL', '2PL', 'gpcm', '2PL', '2PL', 'gpcm', 'gpcm', 'gpcm', 'gpcm',
'gpcm', '2PL', '2PL', 'gpcm', '2PL', '2PL', '2PL', '2PL', '2PL',
'gpcm', 'gpcm', 'gpcm', '2PL', '2PL', '2PL', '2PL', '2PL',
'gpcm', '2PL', '2PL', '2PL', 'gpcm', 'gpcm', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL'),
SE=TRUE,
covdata=covdataset,
```

```

formula = ~
EAP+ITSEX+Prin1+Prin2+Prin3+Prin4+Prin5+Prin6+Prin7+Prin8+Prin9+Prin10+Prin
11+Prin12+Prin13+Prin14+Prin15+Prin16+Prin17+Prin18+Prin19+Prin20+Prin21+Pr
in22+Prin23+Prin24+Prin25+Prin26+Prin27+Prin28+Prin29+Prin30+Prin31+Prin32+
Prin33+Prin34+Prin35+Prin36+Prin37+Prin38+Prin39+Prin40+Prin41+Prin42+Prin4
3+Prin44+Prin45+Prin46+Prin47+Prin48+Prin49+Prin50+Prin51+Prin52+Prin53+Pri
n54+Prin55+Prin56+Prin57+Prin58+Prin59+Prin60+Prin61+Prin62+Prin63+Prin64+P
rin65+Prin66+Prin67+Prin68+Prin69+Prin70+Prin71+Prin72+Prin73+Prin74+Prin75
+Prin76+Prin77+Prin78+Prin79+Prin80+Prin81+Prin82+Prin83+Prin84+Prin85+Prin
86+Prin87+Prin88+Prin89+Prin90+Prin91+Prin92+Prin93+Prin94+Prin95+Prin96+Pr
in97+Prin98+Prin99+Prin100+Prin101+Prin102+Prin103+Prin104+Prin105+Prin106+
Prin107+Prin108+Prin109+Prin110+Prin111+Prin112+Prin113+Prin114+Prin115+Pri
n116+Prin117+Prin118+Prin119+Prin120+Prin121+Prin122+Prin123+Prin124+Prin12
5+Prin126+Prin127+Prin128+Prin129+Prin130+Prin131+Prin132+Prin133+Prin134+P
rin135+Prin136+Prin137+Prin138+Prin139+Prin140+Prin141+Prin142+Prin143+Prin
144+Prin145+Prin146+Prin147+Prin148+Prin149+Prin150+Prin151+Prin152+Prin153
+Prin154+Prin155+Prin156+Prin157+Prin158+Prin159+Prin160+Prin161+Prin162+Pr
in163+Prin164+Prin165+Prin166+Prin167+Prin168+Prin169+Prin170+Prin171+Prin1
72+Prin173+Prin174+Prin175+Prin176+Prin177+Prin178+Prin179+Prin180+Prin181+
Prin182+Prin183+Prin184+Prin185+Prin186+Prin187+Prin188+Prin189+Prin190+Pri
n191+Prin192+Prin193+Prin194+Prin195+Prin196+Prin197+Prin198+Prin199+Prin20
0+Prin201+Prin202+Prin203+Prin204+Prin205+Prin206+Prin207+Prin208+Prin209+P
rin210+Prin211+Prin212+Prin213+Prin214+Prin215+Prin216+Prin217+Prin218+Prin
219+Prin220+Prin221+Prin222+Prin223+Prin224+Prin225+Prin226+Prin227+Prin228
+Prin229+Prin230+Prin231+Prin232+Prin233+Prin234+Prin235+Prin236+Prin237+Pr
in238+Prin239+Prin240+Prin241+Prin242+Prin243+Prin244+Prin245+Prin246+Prin2
47+Prin248+Prin249+Prin250+Prin251+Prin252+Prin253+Prin254+Prin255+Prin256+
Prin257+Prin258+Prin259,
pars ="values",
technical=list(NCYCLES=2000)

###Set estimation to FALSE
Svalues$est<-FALSE

### Parameters 2PL items PIRLS

Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B01C']<-1.150441
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B04C']<-1.200897
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B07C']<-1.267163
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B08C']<-1.427796
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B09C']<-1.454231
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B15C']<-0.71281
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E01C']<-1.3532
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E02C']<-1.311176
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E10C']<-1.459909
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E12C']<-1.897812
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E13C']<-0.897702
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E15C']<-1.351262
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E16C']<-1.200064
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011F06C']<-1.319914
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011F08C']<-1.953096
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011F10C']<-1.437911
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011L03C']<-1.046877
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011L06C']<-1.114418
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021K02C']<-1.371679
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021K05C']<-1.647402
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021Y03C']<-1.385738
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021Y10C']<-1.272858
Svalues$value[Svalues$name=='a1' & Svalues$item=='R031M02C']<-2.07111
Svalues$value[Svalues$name=='a1' & Svalues$item=='R031M04C']<-0.952187
Svalues$value[Svalues$name=='a1' & Svalues$item=='R031M10C']<-1.059202

```

Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031M16C']<-2.051509
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031W04C']<-1.431434
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031W11C']<-2.494308
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031W13C']<-1.466131
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041H03C']<-1.993046
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041H04C']<-1.140955
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041H08C']<-1.174547
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041H14C']<-1.683289
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041H15C']<-2.125221
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041H16C']<-1.745849
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041I01C']<-1.34742
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041I09C']<-1.958196
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041I13C']<-1.366698
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041I14C']<-1.362057
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041I15C']<-1.283857
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041O02C']<-1.459875
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041O03C']<-1.733796
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041O08C']<-1.439934
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041O09C']<-2.049027
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T04C']<-2.106198
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T06C']<-2.348516
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T08C']<-2.017594
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T10C']<-2.148851
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T14C']<-0.906015

Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021B01C']<-2.911075906
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021B04C']<-1.513190265
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021B07C']<-1.470466632
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021B08C']<-1.940974438
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021B09C']<-1.658434117
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021B15C']<--0.226395584
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E01C']<-4.234852932
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E02C']<-2.774736875
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E10C']<-3.964806263
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E12C']<-4.053384826
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E13C']<-0.246903958
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E15C']<-2.998220663
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E16C']<-1.150501357
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011F06C']<-0.200956907
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011F08C']<-0.640224869
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011F10C']<-2.040036231
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011L03C']<-0.495842822
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011L06C']<--0.212887271
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021K02C']<-0.766192456
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021K05C']<--0.225908236
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021Y03C']<--0.782068955
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021Y10C']<--0.730060434
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M02C']<-1.98186587
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M04C']<--0.340216415
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M10C']<--0.437662266
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M16C']<--0.118105373
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031W04C']<-0.98379596
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031W11C']<--1.373266212
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031W13C']<--1.159431056
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041H03C']<--0.512053378
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041H04C']<--1.876209221
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041H08C']<--1.75295267
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041H14C']<--0.516146906
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041H15C']<-0.818380103
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041H16C']<--1.007634209
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041I01C']<-2.050274695

```

Svalues$value[Svalues$name=='d' & Svalues$item=='R041I09C']<-0.068145221
Svalues$value[Svalues$name=='d' & Svalues$item=='R041I13C']<--0.282523811
Svalues$value[Svalues$name=='d' & Svalues$item=='R041I14C']<--1.286285769
Svalues$value[Svalues$name=='d' & Svalues$item=='R041I15C']<--0.885476173
Svalues$value[Svalues$name=='d' & Svalues$item=='R041O02C']<-1.408852369
Svalues$value[Svalues$name=='d' & Svalues$item=='R041O03C']<--1.504414789
Svalues$value[Svalues$name=='d' & Svalues$item=='R041O08C']<-0.091939786
Svalues$value[Svalues$name=='d' & Svalues$item=='R041O09C']<-0.121753184
Svalues$value[Svalues$name=='d' & Svalues$item=='R041T04C']<--0.116788679
Svalues$value[Svalues$name=='d' & Svalues$item=='R041T06C']<-1.360542289
Svalues$value[Svalues$name=='d' & Svalues$item=='R041T08C']<--0.11687922
Svalues$value[Svalues$name=='d' & Svalues$item=='R041T10C']<-0.454610918
Svalues$value[Svalues$name=='d' & Svalues$item=='R041T14C']<--0.118280258

```

Parameters 3PL items PIRLS

```

Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B02M']<-1.62809
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B03M']<-1.225139
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B05M']<-1.501899
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B06M']<-0.748289
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B10M']<-1.122765
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B11M']<-1.664402
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B12M']<-1.254838
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B14M']<-1.828231
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E03M']<-2.283865
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E04M']<-1.87935
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E05M']<-1.900073
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E06M']<-1.591455
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E08M']<-1.738862
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E09M']<-1.053167
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E11M']<-2.048449
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E17M']<-1.780529
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011F01M']<-2.267256
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011F02M']<-1.132523
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011F03M']<-1.564017
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011F04M']<-2.22241
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011F05M']<-1.59715
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011F11M']<-1.25562
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011F13M']<-1.911259
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011L01M']<-0.903754
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011L02M']<-1.273215
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011L05M']<-2.01688
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011L07M']<-1.312689
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011L09M']<-1.637933
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011L11M']<-1.549975
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021K03M']<-1.707225
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021K04M']<-1.804737
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021K06M']<-2.532133
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021K08M']<-1.690293
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021K09M']<-2.045406
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021K11M']<-1.818422
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021Y01M']<-1.86405
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021Y02M']<-2.803215
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021Y04M']<-2.163692
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021Y05M']<-2.924986
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021Y06M']<-2.605726
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021Y07M']<-1.345669
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021Y08M']<-2.312238
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021Y11M']<-2.397952
Svalues$value[Svalues$name=='a1' & Svalues$item=='R031M01M']<-2.467091
Svalues$value[Svalues$name=='a1' & Svalues$item=='R031M03M']<-2.261425
Svalues$value[Svalues$name=='a1' & Svalues$item=='R031M05M']<-2.63704

```


Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031M06M']<-1.969705
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031M07M']<-2.752385
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031M08M']<-2.348601
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031M11M']<-1.451018
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031M12M']<-2.032469
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031M13M']<-3.569405
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031M14M']<-3.881134
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031M15M']<-2.206702
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031W03M']<-2.289373
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031W05M']<-2.148035
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031W06M']<-1.28044
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031W08M']<-2.303823
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031W09M']<-1.616479
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031W10M']<-2.190909
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R031W12M']<-2.573919
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041H01M']<-1.609118
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041H02M']<-1.799229
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041H05M']<-1.75151
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041H07M']<-1.521245
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041H09M']<-1.10364
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041H10M']<-2.221118
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041H11M']<-2.38408
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041H12M']<-2.294167
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041I02M']<-1.756967
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041I05M']<-1.915186
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041I06M']<-2.141235
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041I08M']<-2.70912
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041I10M']<-1.618213
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041I12M']<-1.803411
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041O01M']<-1.601094
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041O06M']<-2.443291
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041O11M']<-2.320415
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041O12M']<-2.180352
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T01M']<-1.843786
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T05M']<-1.285013
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T09M']<-2.652238
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T12M']<-1.698453
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T13M']<-1.889448
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T15M']<-1.525699
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T16M']<-2.191623

Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021B02M']<-3.830098006
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021B03M']<-2.947586423
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021B05M']<-1.974486539
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021B06M']<-1.381887745
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021B10M']<-0.379707895
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021B11M']<-2.037577572
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021B12M']<-2.762726631
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021B14M']<-1.0918744
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E03M']<-1.190510309
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E04M']<-0.709548593
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E05M']<-2.97997949
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E06M']<-3.92295249
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E08M']<-1.303068406
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E09M']<-0.488795868
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E11M']<-1.672947814
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='L021E17M']<-0.831916565
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011F01M']<-1.421093388
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011F02M']<-0.960119024
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011F03M']<-1.042276569

Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011F04M']<-1.847622778
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011F05M']<-0.406889934
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011F11M']<--0.260566262
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011F13M']<-0.38003474
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011L01M']<-2.055724036
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011L02M']<--0.87241965
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011L05M']<--0.709679566
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011L07M']<--0.622752788
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011L09M']<-1.324825728
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R011L11M']<--0.548861647
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021K03M']<--0.137858419
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021K04M']<--1.767415039
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021K06M']<--0.132405235
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021K08M']<--0.597619993
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021K09M']<-0.021026774
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021K11M']<--0.472898825
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021Y01M']<--0.287883882
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021Y02M']<-0.573061242
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021Y04M']<--0.200574248
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021Y05M']<--0.252923539
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021Y06M']<--0.110352496
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021Y07M']<-1.360686666
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021Y08M']<-0.625807215
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R021Y11M']<--0.083736484
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M01M']<-2.164576302
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M03M']<-0.008073287
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M05M']<--0.29600774
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M06M']<--0.725422654
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M07M']<-0.962701701
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M08M']<-1.168875232
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M11M']<-1.00361111
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M12M']<--0.215462039
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M13M']<-2.29270022
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M14M']<-0.758024282
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031M15M']<--0.144075574
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031W03M']<-0.144047349
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031W05M']<--1.067315631
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031W06M']<-1.279210778
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031W08M']<-0.214739342
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031W09M']<--0.912971174
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031W10M']<--0.700499335
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R031W12M']<--1.816337421
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041H01M']<-1.07320125
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041H02M']<-1.336251394
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041H05M']<--0.229920718
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041H07M']<--1.32807731
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041H09M']<--0.300322517
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041H10M']<-0.397224743
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041H11M']<-1.280441686
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041H12M']<--0.003670667
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041I02M']<--1.093940363
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041I05M']<--0.314454389
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041I06M']<--1.042845682
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041I08M']<-1.569880858
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041I10M']<--0.429279545
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041I12M']<--0.279131955
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041O01M']<-1.58985432
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041O06M']<-0.517391302
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041O11M']<--0.221970899
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041O12M']<-0.744066924
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041T01M']<-1.794870357

Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041T05M']<--0.491928677
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041T09M']<--1.324792881
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041T12M']<-0.085669969
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041T13M']<--0.816033697
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041T15M']<--0.905624412
Svalues\$value[Svalues\$name=='d' & Svalues\$item=='R041T16M']<-0.101187234

Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021B02M']<-0.10136
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021B03M']<-0.25
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021B05M']<-0.22267
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021B06M']<-0.25
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021B10M']<-0.19174
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021B11M']<-0.18603
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021B12M']<-0.12727
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021B14M']<-0.1494
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021E03M']<-0.26973
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021E04M']<-0.26216
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021E05M']<-0.24221
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021E06M']<-0.30402
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021E08M']<-0.20112
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021E09M']<-0.33952
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021E11M']<-0.22596
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021E17M']<-0.17116
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011F01M']<-0.14794
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011F02M']<-0.24302
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011F03M']<-0.15674
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011F04M']<-0.22812
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011F05M']<-0.21673
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011F11M']<-0.19218
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011F13M']<-0.2702
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011L01M']<-0.14643
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011L02M']<-0.24066
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011L05M']<-0.20576
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011L07M']<-0.1544
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011L09M']<-0.2261
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011L11M']<-0.18939
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021K03M']<-0.18378
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021K04M']<-0.39112
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021K06M']<-0.28064
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021K08M']<-0.1967
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021K09M']<-0.24618
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021K11M']<-0.23994
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y01M']<-0.25259
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y02M']<-0.28768
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y04M']<-0.22186
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y05M']<-0.22563
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y06M']<-0.20883
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y07M']<-0.18192
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y08M']<-0.26053
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y11M']<-0.28385
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M01M']<-0.26826
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M03M']<-0.22822
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M05M']<-0.42435
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M06M']<-0.28256
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M07M']<-0.27562
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M08M']<-0.26526
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M11M']<-0.26229
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M12M']<-0.16209
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M13M']<-0.25588
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M14M']<-0.19745

```

Svalues$value[Svalues$name=='g' & Svalues$item=='R031M15M']<-0.21785
Svalues$value[Svalues$name=='g' & Svalues$item=='R031W03M']<-0.16194
Svalues$value[Svalues$name=='g' & Svalues$item=='R031W05M']<-0.25664
Svalues$value[Svalues$name=='g' & Svalues$item=='R031W06M']<-0.147
Svalues$value[Svalues$name=='g' & Svalues$item=='R031W08M']<-0.30706
Svalues$value[Svalues$name=='g' & Svalues$item=='R031W09M']<-0.17846
Svalues$value[Svalues$name=='g' & Svalues$item=='R031W10M']<-0.16449
Svalues$value[Svalues$name=='g' & Svalues$item=='R031W12M']<-0.21876
Svalues$value[Svalues$name=='g' & Svalues$item=='R041H01M']<-0.58233
Svalues$value[Svalues$name=='g' & Svalues$item=='R041H02M']<-0.19599
Svalues$value[Svalues$name=='g' & Svalues$item=='R041H05M']<-0.28585
Svalues$value[Svalues$name=='g' & Svalues$item=='R041H07M']<-0.1511
Svalues$value[Svalues$name=='g' & Svalues$item=='R041H09M']<-0.16601
Svalues$value[Svalues$name=='g' & Svalues$item=='R041H10M']<-0.26344
Svalues$value[Svalues$name=='g' & Svalues$item=='R041H11M']<-0.31077
Svalues$value[Svalues$name=='g' & Svalues$item=='R041H12M']<-0.36935
Svalues$value[Svalues$name=='g' & Svalues$item=='R041I02M']<-0.18475
Svalues$value[Svalues$name=='g' & Svalues$item=='R041I05M']<-0.22272
Svalues$value[Svalues$name=='g' & Svalues$item=='R041I06M']<-0.2827
Svalues$value[Svalues$name=='g' & Svalues$item=='R041I08M']<-0.26413
Svalues$value[Svalues$name=='g' & Svalues$item=='R041I10M']<-0.24422
Svalues$value[Svalues$name=='g' & Svalues$item=='R041I12M']<-0.14524
Svalues$value[Svalues$name=='g' & Svalues$item=='R041O01M']<-0.25359
Svalues$value[Svalues$name=='g' & Svalues$item=='R041O06M']<-0.31459
Svalues$value[Svalues$name=='g' & Svalues$item=='R041O11M']<-0.29507
Svalues$value[Svalues$name=='g' & Svalues$item=='R041O12M']<-0.16907
Svalues$value[Svalues$name=='g' & Svalues$item=='R041T01M']<-0.2421
Svalues$value[Svalues$name=='g' & Svalues$item=='R041T05M']<-0.24397
Svalues$value[Svalues$name=='g' & Svalues$item=='R041T09M']<-0.18213
Svalues$value[Svalues$name=='g' & Svalues$item=='R041T12M']<-0.34556
Svalues$value[Svalues$name=='g' & Svalues$item=='R041T13M']<-0.11168
Svalues$value[Svalues$name=='g' & Svalues$item=='R041T15M']<-0.23507
Svalues$value[Svalues$name=='g' & Svalues$item=='R041T16M']<-0.28628

```

Parameters GPCM met 3 categorieë items PIRLS

```

Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B13C']<-0.920635
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B16C']<-0.740027
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021B17C']<-1.19918
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E07C']<-0.786811
Svalues$value[Svalues$name=='a1' & Svalues$item=='L021E14C']<-0.837488
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011F07C']<-0.855882
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011F09C']<-1.718309
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011F12C']<-1.050158
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011L08C']<-1.361224
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011L10C']<-1.243771
Svalues$value[Svalues$name=='a1' & Svalues$item=='R011L12C']<-1.250112
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021K01C']<-0.717043
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021K07C']<-1.159978
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021K10C']<-1.334126
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021Y09C']<-1.625047
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021Y12C']<-1.199605
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021Y14C']<-0.979965
Svalues$value[Svalues$name=='a1' & Svalues$item=='R031M09C']<-1.290946
Svalues$value[Svalues$name=='a1' & Svalues$item=='R031W01C']<-1.220328
Svalues$value[Svalues$name=='a1' & Svalues$item=='R031W02C']<-1.359779
Svalues$value[Svalues$name=='a1' & Svalues$item=='R041H06C']<-1.189286
Svalues$value[Svalues$name=='a1' & Svalues$item=='R041I03C']<-0.951388
Svalues$value[Svalues$name=='a1' & Svalues$item=='R041I04C']<-1.164959
Svalues$value[Svalues$name=='a1' & Svalues$item=='R041I07C']<-1.646382
Svalues$value[Svalues$name=='a1' & Svalues$item=='R041I11C']<-1.394459
Svalues$value[Svalues$name=='a1' & Svalues$item=='R041O04C']<-1.011704

```

Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041O05C']<-1.071646
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041O07C']<-1.068586
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041O10C']<-1.386061
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T02C']<-1.17215
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T03C']<-1.603423
Svalues\$value[Svalues\$name=='a1' & Svalues\$item=='R041T07C']<-1.326272

Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='L021B13C']<-1.528318544
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='L021B16C']<-0.20527609
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='L021B17C']<-1.315932165
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='L021E07C']<-0.622776643
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='L021E14C']<-0.557373389
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R011F07C']<--1.087783228
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R011F09C']<-1.205118834
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R011F12C']<--1.169718488
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R011L08C']<-0.123939445
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R011L10C']<--0.559659637
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R011L12C']<-0.376308714
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R021K01C']<-0.772083221
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R021K07C']<--0.027769873
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R021K10C']<--1.566810916
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R021Y09C']<-1.022333318
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R021Y12C']<--1.383264526
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R021Y14C']<--0.756327187
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R031M09C']<-1.514589485
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R031W01C']<-1.009199053
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R031W02C']<--0.522943808
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R041H06C']<-0.208886193
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R041I03C']<-0.122377038
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R041I04C']<--0.197437251
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R041I07C']<--0.839029195
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R041I11C']<--0.300631416
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R041O04C']<--1.319707166
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R041O05C']<--0.523102562
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R041O07C']<-1.056692638
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R041O10C']<--0.334511962
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R041T02C']<-0.919844713
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R041T03C']<--0.059727507
Svalues\$value[Svalues\$name=='d1' & Svalues\$item=='R041T07C']<--1.09207889

Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='L021B13C']<-1.780765868
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='L021B16C']<-0.273869192
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='L021B17C']<-0.992081614
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='L021E07C']<-1.274114525
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='L021E14C']<--0.311763283
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='R011F07C']<--0.641860147
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='R011F09C']<-2.155378077
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='R011F12C']<--1.349369017
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='R011L08C']<--1.665239768
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='R011L10C']<--1.694961368
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='R011L12C']<--1.27221398
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='R021K01C']<-1.277784967
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='R021K07C']<--0.331266517
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='R021K10C']<--2.075633231
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='R021Y09C']<-1.791516815
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='R021Y12C']<-0.001559487
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='R021Y14C']<--0.435810035
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='R031M09C']<-0.076940382
Svalues\$value[Svalues\$name=='d2' & Svalues\$item=='R031W01C']<-1.425806829

```

Svalues$value[Svalues$name=='d2' & Svalues$item=='R031W02C']<--0.756200297
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041H06C']<-0.628299794
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041I03C']<--0.38293367
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041I04C']<--1.095084759
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041I07C']<--1.652967528
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041I11C']<--1.313915048
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041O04C']<--2.31635701
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041O05C']<--0.489056369
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041O07C']<-2.036233366
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041O10C']<--0.652917895
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041T02C']<-1.209283712
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041T03C']<--0.228359504
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041T07C']<--1.648741774

```

Parameters GPCM met 4 categorie ^뽁items PIRLS

```

Svalues$value[Svalues$name=='a1' & Svalues$item=='R011L04C']<-1.134699
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021K12C']<-0.97852
Svalues$value[Svalues$name=='a1' & Svalues$item=='R021Y13C']<-1.292102
Svalues$value[Svalues$name=='a1' & Svalues$item=='R031M17C']<-1.039924
Svalues$value[Svalues$name=='a1' & Svalues$item=='R031W07C']<-1.49396
Svalues$value[Svalues$name=='a1' & Svalues$item=='R041H13C']<-0.919887
Svalues$value[Svalues$name=='a1' & Svalues$item=='R041O13C']<-0.963509
Svalues$value[Svalues$name=='a1' & Svalues$item=='R041T11C']<-1.46064

```

```

Svalues$value[Svalues$name=='d1' & Svalues$item=='R011L04C']<-1.389484313
Svalues$value[Svalues$name=='d1' & Svalues$item=='R021K12C']<-0.544106046
Svalues$value[Svalues$name=='d1' & Svalues$item=='R021Y13C']<-0.279249084
Svalues$value[Svalues$name=='d1' & Svalues$item=='R031M17C']<-0.07074603
Svalues$value[Svalues$name=='d1' & Svalues$item=='R031W07C']<--0.878687514
Svalues$value[Svalues$name=='d1' & Svalues$item=='R041H13C']<--0.774278087
Svalues$value[Svalues$name=='d1' & Svalues$item=='R041O13C']<--1.488659945
Svalues$value[Svalues$name=='d1' & Svalues$item=='R041T11C']<--1.378902586

```

```

Svalues$value[Svalues$name=='d2' & Svalues$item=='R011L04C']<--0.277672192
Svalues$value[Svalues$name=='d2' & Svalues$item=='R021K12C']<-0.569107232
Svalues$value[Svalues$name=='d2' & Svalues$item=='R021Y13C']<--0.491476838
Svalues$value[Svalues$name=='d2' & Svalues$item=='R031M17C']<-0.305280089
Svalues$value[Svalues$name=='d2' & Svalues$item=='R031W07C']<--1.386140907
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041H13C']<--1.218519116
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041O13C']<--1.05828938
Svalues$value[Svalues$name=='d2' & Svalues$item=='R041T11C']<--1.66971601

```

```

Svalues$value[Svalues$name=='d3' & Svalues$item=='R011L04C']<--1.423355079
Svalues$value[Svalues$name=='d3' & Svalues$item=='R021K12C']<-0.322725681
Svalues$value[Svalues$name=='d3' & Svalues$item=='R021Y13C']<--1.464390881
Svalues$value[Svalues$name=='d3' & Svalues$item=='R031M17C']<--0.00889135
Svalues$value[Svalues$name=='d3' & Svalues$item=='R031W07C']<--2.281575712
Svalues$value[Svalues$name=='d3' & Svalues$item=='R041H13C']<--2.077868352
Svalues$value[Svalues$name=='d3' & Svalues$item=='R041O13C']<--1.005325291
Svalues$value[Svalues$name=='d3' & Svalues$item=='R041T11C']<--2.27149969

```

###Peilingen & SFB met gokparameters

```

Svalues$value[Svalues$name=='g' & Svalues$item=='i2091']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='i2092']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='i2094']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='i2095']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='i2096']<-0.15

```

```

Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_SP_02']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_SP_06']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_SP_07']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_VL_01']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_VL_03']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_VL_07']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='Le05_DMS_02']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='Le05_DMS_03']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='Le05_DMS_04']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='Le05_DMS_06']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='Le05_DMS_07']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='Le06_PZ_01']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='Le06_PZ_03']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='Le06_PZ_09']<-0.15

```

```

Svalues$est[Svalues$name=='a1' & Svalues$item=='i2091']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='i2092']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='i2094']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='i2095']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='i2096']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le01_SP_02']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le01_SP_06']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le01_SP_07']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le01_VL_01']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le01_VL_03']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le01_VL_07']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le05_DMS_02']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le05_DMS_03']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le05_DMS_04']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le05_DMS_06']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le05_DMS_07']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le06_PZ_01']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le06_PZ_03']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le06_PZ_09']<-TRUE

```

```

Svalues$est[Svalues$name=='d' & Svalues$item=='i2091']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='i2092']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='i2094']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='i2095']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='i2096']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le01_SP_02']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le01_SP_06']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le01_SP_07']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le01_VL_01']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le01_VL_03']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le01_VL_07']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le05_DMS_02']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le05_DMS_03']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le05_DMS_04']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le05_DMS_06']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le05_DMS_07']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le06_PZ_01']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le06_PZ_03']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le06_PZ_09']<-TRUE

```

```
###Peilingen & SFB zonder gokparameter
```

```
#Alles wordt vrij geschat, geen constraints te plaatsen
```

```

Svalues$est[Svalues$name=='a1' & Svalues$item=='Le01_SP_03']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le01_SP_05']<-TRUE

```

```
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le01_VL_04']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le01_VL_08']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le05_DMS_01']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le06_PZ_06']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='Le06_PZ_07']<-TRUE
```

```
Svalues$est[Svalues$name=='d' & Svalues$item=='Le01_SP_03']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le01_SP_05']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le01_VL_04']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le01_VL_08']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le05_DMS_01']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le06_PZ_06']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='Le06_PZ_07']<-TRUE
```

```
### Vrij zetten groepsparameters
```

```
Svalues$est[Svalues$name=='MEAN_1']<-TRUE
Svalues$est[Svalues$name=='COV_11']<-TRUE
Svalues$est[Svalues$class=='lrPars']<-TRUE
Svalues$est[Svalues$name=='Fl_(Intercept)']<-TRUE
```

```
#Unidimensionele model
```

```
unidim<-mirt(PIRLS18[,c(2:112)],model=1,
itemtype=c('2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL', '2PL', 'gpcm', '2PL', 'gpcm', '2PL',
'2PL', '2PL', 'gpcm', '2PL', '2PL', 'gpcm', 'gpcm', 'gpcm', 'gpcm',
'gpcm', '2PL', '2PL', 'gpcm', '2PL', '2PL', 'gpcm', '2PL', '2PL',
'gpcm', '2PL', 'gpcm', '2PL', 'gpcm', '2PL', 'gpcm', 'gpcm', 'gpcm',
'2PL', 'gpcm', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', 'gpcm', 'gpcm',
'gpcm', '2PL', '2PL', 'gpcm', 'gpcm', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL'),
covdata=covdataset,
formula = ~
EAP+ITSEX+Prin1+Prin2+Prin3+Prin4+Prin5+Prin6+Prin7+Prin8+Prin9+Prin10+Prin
11+Prin12+Prin13+Prin14+Prin15+Prin16+Prin17+Prin18+Prin19+Prin20+Prin21+Pr
in22+Prin23+Prin24+Prin25+Prin26+Prin27+Prin28+Prin29+Prin30+Prin31+Prin32+
Prin33+Prin34+Prin35+Prin36+Prin37+Prin38+Prin39+Prin40+Prin41+Prin42+Prin4
3+Prin44+Prin45+Prin46+Prin47+Prin48+Prin49+Prin50+Prin51+Prin52+Prin53+Pri
n54+Prin55+Prin56+Prin57+Prin58+Prin59+Prin60+Prin61+Prin62+Prin63+Prin64+P
rin65+Prin66+Prin67+Prin68+Prin69+Prin70+Prin71+Prin72+Prin73+Prin74+Prin75
+Prin76+Prin77+Prin78+Prin79+Prin80+Prin81+Prin82+Prin83+Prin84+Prin85+Prin
86+Prin87+Prin88+Prin89+Prin90+Prin91+Prin92+Prin93+Prin94+Prin95+Prin96+Pr
in97+Prin98+Prin99+Prin100+Prin101+Prin102+Prin103+Prin104+Prin105+Prin106+
Prin107+Prin108+Prin109+Prin110+Prin111+Prin112+Prin113+Prin114+Prin115+Pri
n116+Prin117+Prin118+Prin119+Prin120+Prin121+Prin122+Prin123+Prin124+Prin12
5+Prin126+Prin127+Prin128+Prin129+Prin130+Prin131+Prin132+Prin133+Prin134+P
rin135+Prin136+Prin137+Prin138+Prin139+Prin140+Prin141+Prin142+Prin143+Prin
144+Prin145+Prin146+Prin147+Prin148+Prin149+Prin150+Prin151+Prin152+Prin153
+Prin154+Prin155+Prin156+Prin157+Prin158+Prin159+Prin160+Prin161+Prin162+Pr
in163+Prin164+Prin165+Prin166+Prin167+Prin168+Prin169+Prin170+Prin171+Prin1
72+Prin173+Prin174+Prin175+Prin176+Prin177+Prin178+Prin179+Prin180+Prin181+
Prin182+Prin183+Prin184+Prin185+Prin186+Prin187+Prin188+Prin189+Prin190+Pri
n191+Prin192+Prin193+Prin194+Prin195+Prin196+Prin197+Prin198+Prin199+Prin20
0+Prin201+Prin202+Prin203+Prin204+Prin205+Prin206+Prin207+Prin208+Prin209+P
rin210+Prin211+Prin212+Prin213+Prin214+Prin215+Prin216+Prin217+Prin218+Prin
219+Prin220+Prin221+Prin222+Prin223+Prin224+Prin225+Prin226+Prin227+Prin228
```



```

+Prin229+Prin230+Prin231+Prin232+Prin233+Prin234+Prin235+Prin236+Prin237+Pr
in238+Prin239+Prin240+Prin241+Prin242+Prin243+Prin244+Prin245+Prin246+Prin2
47+Prin248+Prin249+Prin250+Prin251+Prin252+Prin253+Prin254+Prin255+Prin256+
Prin257+Prin258+Prin259,
pars=Svalues,
technical=list(NCYCLES=2000))

summary(unidim) #Parameterwaarden voor 3-parameter model volgens
factormodel. Ter controle.
coef(unidim, simplify=TRUE, IRTpars=TRUE) #Extractie parameterwaarden voor
3-parameter model volgens IRT-model. Ter controle.

#schatting plausible values
set.seed(12345)
pv<-fcores(unidim,plausible.draws=5)
pv<-as.data.frame(pv)
names(pv)<-c("pv1","pv2","pv3","pv4","pv5")

cor(pv$pv1,pv$pv2)
cor(pv$pv1,pv$pv3)
cor(pv$pv1,pv$pv4)
cor(pv$pv1,pv$pv5)

cor(pv$pv2,pv$pv3)
cor(pv$pv2,pv$pv4)
cor(pv$pv2,pv$pv5)

cor(pv$pv3,pv$pv4)
cor(pv$pv3,pv$pv5)

cor(pv$pv4,pv$pv5)

AverageMeanpv=(mean(pv$pv1)+mean(pv$pv2)+mean(pv$pv3)+mean(pv$pv4)+mean(pv$
pv5))/5
AverageSDpv=sqrt((var(pv$pv1)+var(pv$pv2)+var(pv$pv3)+var(pv$pv4)+var(pv$pv
5))/5)

#Lineare transformatie naar PIRLS schaal
pv$pv1<-(pv$pv1*97.30265)+516.163
pv$pv2<-(pv$pv2*97.30265)+516.163
pv$pv3<-(pv$pv3*97.30265)+516.163
pv$pv4<-(pv$pv4*97.30265)+516.163
pv$pv5<-(pv$pv5*97.30265)+516.163

AverageMeanpv=(mean(pv$pv1)+mean(pv$pv2)+mean(pv$pv3)+mean(pv$pv4)+mean(pv$
pv5))/5
AverageSDpv=sqrt((var(pv$pv1)+var(pv$pv2)+var(pv$pv3)+var(pv$pv4)+var(pv$pv
5))/5)

pv=cbind(PIRLS18$LeerlingID,pv)

#Wegschrijven plausible values

write.csv(pv,file="2019-05-01 Plausiblevalues terugkerende studenten
2018.csv")

#exporteren geschatte itemparameters
Parameters_IRT<-
as.data.frame(coef(unidim,simplify=TRUE,IRTpars=TRUE)$items)
Parameters_MIRT<-as.data.frame(coef(unidim,simplify=TRUE)$items)

```

```
options(max.print=10000)
sink('parameters2018 terugkerende studenten.txt')
cbind(Parameters_IRT,Parameters_MIRT$d,Parameters_MIRT$ak0,Parameters_MIRT$
ak1,Parameters_MIRT$ak2,Parameters_MIRT$ak3,Parameters_MIRT$d0,Parameters_M
IRT$d1,Parameters_MIRT$d2,Parameters_MIRT$d3)
sink()
```



```

R031M12M', 'R031M13M', 'R031M14M', 'R031M15M', 'R031W03M', 'R031W05M', 'R031W06M',
'R031W08M', 'R031W09M', 'R031W10M', 'R031W12M', 'R041H01M', 'R041H02M', 'R041H05
M', 'R041H07M', 'R041H09M', 'R041H10M', 'R041H11M', 'R041H12M', 'R041I02M', 'R041I
05M', 'R041I06M', 'R041I08M', 'R041I10M', 'R041I12M', 'R041O01M', 'R041O06M', 'R04
1O11M', 'R041O12M', 'R041T01M', 'R041T05M', 'R041T09M', 'R041T12M', 'R041T13M', 'R
041T15M', 'R041T16M', 'L021B13C', 'L021B16C', 'L021B17C', 'L021E07C', 'L021E14C',
'R011F07C', 'R011F09C', 'R011F12C', 'R011L08C', 'R011L10C', 'R011L12C', 'R021K01C
', 'R021K07C', 'R021K10C', 'R021Y09C', 'R021Y12C', 'R021Y14C', 'R031M09C', 'R031W0
1C', 'R031W02C', 'R041H06C', 'R041I03C', 'R041I04C', 'R041I07C', 'R041I11C', 'R041
O04C', 'R041O05C', 'R041O07C', 'R041O10C', 'R041T02C', 'R041T03C', 'R041T07C', 'R0
11L04C', 'R021K12C', 'R021Y13C', 'R031M17C', 'R031W07C', 'R041H13C', 'R041O13C', '
R041T11C', 'i2091', 'i2092', 'i2094', 'i2095', 'i2096'))), 'free_means', 'free_var
'))

```

```

###Set estimation to FALSE
Svalues$est<-FALSE

```

```

### Parameters 2PL items PIRLS

```

```

Svalues$est[Svalues$name=='al' & Svalues$item=='L021B01C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='L021B04C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='L021B07C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='L021B08C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='L021B09C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='L021B15C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='L021E01C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='L021E02C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='L021E10C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='L021E12C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='L021E13C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='L021E15C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='L021E16C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R011F06C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R011F08C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R011F10C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R011L03C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R011L06C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R021K02C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R021K05C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R021Y03C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R021Y10C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R031M02C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R031M04C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R031M10C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R031M16C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R031W04C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R031W11C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R031W13C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R041H03C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R041H04C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R041H08C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R041H14C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R041H15C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R041H16C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R041I01C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R041I09C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R041I13C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R041I14C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R041I15C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R041O02C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R041O03C']<-TRUE
Svalues$est[Svalues$name=='al' & Svalues$item=='R041O08C']<-TRUE

```

```
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041O09C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041T04C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041T06C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041T08C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041T10C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041T14C']<-TRUE
```

```
Svalues$est[Svalues$name=='d' & Svalues$item=='L021B01C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='L021B04C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='L021B07C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='L021B08C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='L021B09C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='L021B15C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='L021E01C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='L021E02C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='L021E10C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='L021E12C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='L021E13C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='L021E15C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='L021E16C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R011F06C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R011F08C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R011F10C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R011L03C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R011L06C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R021K02C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R021K05C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R021Y03C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R021Y10C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R031M02C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R031M04C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R031M10C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R031M16C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R031W04C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R031W11C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R031W13C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041H03C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041H04C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041H08C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041H14C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041H15C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041H16C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041I01C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041I09C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041I13C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041I14C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041I15C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041O02C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041O03C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041O08C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041O09C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041T04C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041T06C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041T08C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041T10C']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='R041T14C']<-TRUE
```

```
### Parameters 3PL items PIRLS
```

```
Svalues$est[Svalues$name=='a1' & Svalues$item=='L021B02M']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='L021B03M']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='L021B05M']<-TRUE
```


Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021E08M']<-0.20112
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021E09M']<-0.33952
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021E11M']<-0.22596
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='L021E17M']<-0.17116
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011F01M']<-0.14794
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011F02M']<-0.24302
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011F03M']<-0.15674
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011F04M']<-0.22812
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011F05M']<-0.21673
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011F11M']<-0.19218
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011F13M']<-0.2702
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011L01M']<-0.14643
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011L02M']<-0.24066
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011L05M']<-0.20576
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011L07M']<-0.1544
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011L09M']<-0.2261
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R011L11M']<-0.18939
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021K03M']<-0.18378
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021K04M']<-0.39112
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021K06M']<-0.28064
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021K08M']<-0.1967
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021K09M']<-0.24618
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021K11M']<-0.23994
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y01M']<-0.25259
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y02M']<-0.28768
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y04M']<-0.22186
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y05M']<-0.22563
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y06M']<-0.20883
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y07M']<-0.18192
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y08M']<-0.26053
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R021Y11M']<-0.28385
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M01M']<-0.26826
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M03M']<-0.22822
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M05M']<-0.42435
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M06M']<-0.28256
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M07M']<-0.27562
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M08M']<-0.26526
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M11M']<-0.26229
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M12M']<-0.16209
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M13M']<-0.25588
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M14M']<-0.19745
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031M15M']<-0.21785
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031W03M']<-0.16194
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031W05M']<-0.25664
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031W06M']<-0.147
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031W08M']<-0.30706
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031W09M']<-0.17846
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031W10M']<-0.16449
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R031W12M']<-0.21876
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R041H01M']<-0.58233
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R041H02M']<-0.19599
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R041H05M']<-0.28585
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R041H07M']<-0.1511
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R041H09M']<-0.16601
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R041H10M']<-0.26344
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R041H11M']<-0.31077
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R041H12M']<-0.36935
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R041I02M']<-0.18475
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R041I05M']<-0.22272
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R041I06M']<-0.2827
Svalues\$value[Svalues\$name=='g' & Svalues\$item=='R041I08M']<-0.26413

```

Svalues$value[Svalues$name=='g' & Svalues$item=='R041I10M']<-0.24422
Svalues$value[Svalues$name=='g' & Svalues$item=='R041I12M']<-0.14524
Svalues$value[Svalues$name=='g' & Svalues$item=='R041O01M']<-0.25359
Svalues$value[Svalues$name=='g' & Svalues$item=='R041O06M']<-0.31459
Svalues$value[Svalues$name=='g' & Svalues$item=='R041O11M']<-0.29507
Svalues$value[Svalues$name=='g' & Svalues$item=='R041O12M']<-0.16907
Svalues$value[Svalues$name=='g' & Svalues$item=='R041T01M']<-0.2421
Svalues$value[Svalues$name=='g' & Svalues$item=='R041T05M']<-0.24397
Svalues$value[Svalues$name=='g' & Svalues$item=='R041T09M']<-0.18213
Svalues$value[Svalues$name=='g' & Svalues$item=='R041T12M']<-0.34556
Svalues$value[Svalues$name=='g' & Svalues$item=='R041T13M']<-0.11168
Svalues$value[Svalues$name=='g' & Svalues$item=='R041T15M']<-0.23507
Svalues$value[Svalues$name=='g' & Svalues$item=='R041T16M']<-0.28628

```

```
### Parameters GPCM met 3 categorie items PIRLS
```

```

Svalues$est[Svalues$name=='a1' & Svalues$item=='L021B13C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='L021B16C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='L021B17C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='L021E07C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='L021E14C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R011F07C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R011F09C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R011F12C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R011L08C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R011L10C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R011L12C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R021K01C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R021K07C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R021K10C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R021Y09C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R021Y12C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R021Y14C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R031M09C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R031W01C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R031W02C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041H06C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041I03C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041I04C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041I07C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041I11C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041O04C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041O05C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041O07C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041O10C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041T02C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041T03C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041T07C']<-TRUE

```

```

Svalues$est[Svalues$name=='d1' & Svalues$item=='L021B13C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='L021B16C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='L021B17C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='L021E07C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='L021E14C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R011F07C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R011F09C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R011F12C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R011L08C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R011L10C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R011L12C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R021K01C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R021K07C']<-TRUE

```

```

Svalues$est[Svalues$name=='d1' & Svalues$item=='R021K10C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R021Y09C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R021Y12C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R021Y14C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R031M09C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R031W01C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R031W02C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041H06C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041I03C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041I04C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041I07C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041I11C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041O04C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041O05C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041O07C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041O10C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041T02C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041T03C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041T07C']<-TRUE

```

```

Svalues$est[Svalues$name=='d2' & Svalues$item=='L021B13C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='L021B16C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='L021B17C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='L021E07C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='L021E14C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R011F07C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R011F09C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R011F12C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R011L08C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R011L10C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R011L12C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R021K01C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R021K07C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R021K10C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R021Y09C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R021Y12C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R021Y14C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R031M09C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R031W01C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R031W02C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041H06C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041I03C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041I04C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041I07C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041I11C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041O04C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041O05C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041O07C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041O10C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041T02C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041T03C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041T07C']<-TRUE

```

```

### Parameters GPCM met 4 categorie items PIRLS

```

```

Svalues$est[Svalues$name=='a1' & Svalues$item=='R011L04C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R021K12C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R021Y13C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R031M17C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R031W07C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041H13C']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041O13C']<-TRUE

```

```
Svalues$est[Svalues$name=='a1' & Svalues$item=='R041T11C']<-TRUE
```

```
Svalues$est[Svalues$name=='d1' & Svalues$item=='R011L04C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R021K12C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R021Y13C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R031M17C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R031W07C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041H13C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041O13C']<-TRUE
Svalues$est[Svalues$name=='d1' & Svalues$item=='R041T11C']<-TRUE
```

```
Svalues$est[Svalues$name=='d2' & Svalues$item=='R011L04C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R021K12C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R021Y13C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R031M17C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R031W07C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041H13C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041O13C']<-TRUE
Svalues$est[Svalues$name=='d2' & Svalues$item=='R041T11C']<-TRUE
```

```
Svalues$est[Svalues$name=='d3' & Svalues$item=='R011L04C']<-TRUE
Svalues$est[Svalues$name=='d3' & Svalues$item=='R021K12C']<-TRUE
Svalues$est[Svalues$name=='d3' & Svalues$item=='R021Y13C']<-TRUE
Svalues$est[Svalues$name=='d3' & Svalues$item=='R031M17C']<-TRUE
Svalues$est[Svalues$name=='d3' & Svalues$item=='R031W07C']<-TRUE
Svalues$est[Svalues$name=='d3' & Svalues$item=='R041H13C']<-TRUE
Svalues$est[Svalues$name=='d3' & Svalues$item=='R041O13C']<-TRUE
Svalues$est[Svalues$name=='d3' & Svalues$item=='R041T11C']<-TRUE
```

```
### SFB met gokparameters
```

```
Svalues$value[Svalues$name=='g' & Svalues$item=='i2091']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='i2092']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='i2094']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='i2095']<-0.15
Svalues$value[Svalues$name=='g' & Svalues$item=='i2096']<-0.15
```

```
Svalues$est[Svalues$name=='a1' & Svalues$item=='i2091']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='i2092']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='i2094']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='i2095']<-TRUE
Svalues$est[Svalues$name=='a1' & Svalues$item=='i2096']<-TRUE
```

```
Svalues$est[Svalues$name=='d' & Svalues$item=='i2091']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='i2092']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='i2094']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='i2095']<-TRUE
Svalues$est[Svalues$name=='d' & Svalues$item=='i2096']<-TRUE
```

```
###Peilingen
```

```
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_SP_02']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_SP_03']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_SP_05']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_SP_06']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_SP_07']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_VL_01']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_VL_03']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_VL_04']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_VL_07']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le01_VL_08']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le05_DMS_01']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le05_DMS_02']<-0
```

```

Svalues$value[Svalues$name=='g' & Svalues$item=='Le05_DMS_03']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le05_DMS_04']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le05_DMS_06']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le05_DMS_07']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le06_PZ_01']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le06_PZ_03']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le06_PZ_06']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le06_PZ_07']<-0
Svalues$value[Svalues$name=='g' & Svalues$item=='Le06_PZ_09']<-0

Svalues$value[Svalues$name=='a1' & Svalues$item=='Le01_SP_02']<-4
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le01_SP_03']<-4
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le01_SP_05']<-5
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le01_SP_06']<-3
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le01_SP_07']<-3
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le01_VL_01']<-2
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le01_VL_03']<-4
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le01_VL_04']<-4
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le01_VL_07']<-3
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le01_VL_08']<-4
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le05_DMS_01']<-4
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le05_DMS_02']<-4
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le05_DMS_03']<-4
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le05_DMS_04']<-3
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le05_DMS_06']<-4
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le05_DMS_07']<-4
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le06_PZ_01']<-4
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le06_PZ_03']<-5
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le06_PZ_06']<-3
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le06_PZ_07']<-4
Svalues$value[Svalues$name=='a1' & Svalues$item=='Le06_PZ_09']<-4

Svalues$value[Svalues$name=='d' & Svalues$item=='Le01_SP_02']<--0.908
Svalues$value[Svalues$name=='d' & Svalues$item=='Le01_SP_03']<--0.684
Svalues$value[Svalues$name=='d' & Svalues$item=='Le01_SP_05']<--0.09
Svalues$value[Svalues$name=='d' & Svalues$item=='Le01_SP_06']<--1.05
Svalues$value[Svalues$name=='d' & Svalues$item=='Le01_SP_07']<-0.756
Svalues$value[Svalues$name=='d' & Svalues$item=='Le01_VL_01']<-1.386
Svalues$value[Svalues$name=='d' & Svalues$item=='Le01_VL_03']<--0.292
Svalues$value[Svalues$name=='d' & Svalues$item=='Le01_VL_04']<--0.42
Svalues$value[Svalues$name=='d' & Svalues$item=='Le01_VL_07']<-1.287
Svalues$value[Svalues$name=='d' & Svalues$item=='Le01_VL_08']<-0.144
Svalues$value[Svalues$name=='d' & Svalues$item=='Le05_DMS_01']<--1.016
Svalues$value[Svalues$name=='d' & Svalues$item=='Le05_DMS_02']<-0
Svalues$value[Svalues$name=='d' & Svalues$item=='Le05_DMS_03']<-0.496
Svalues$value[Svalues$name=='d' & Svalues$item=='Le05_DMS_04']<-0.231
Svalues$value[Svalues$name=='d' & Svalues$item=='Le05_DMS_06']<-0.012
Svalues$value[Svalues$name=='d' & Svalues$item=='Le05_DMS_07']<-1.416
Svalues$value[Svalues$name=='d' & Svalues$item=='Le06_PZ_01']<--0.096
Svalues$value[Svalues$name=='d' & Svalues$item=='Le06_PZ_03']<--0.375
Svalues$value[Svalues$name=='d' & Svalues$item=='Le06_PZ_06']<--1.062
Svalues$value[Svalues$name=='d' & Svalues$item=='Le06_PZ_07']<--0.62
Svalues$value[Svalues$name=='d' & Svalues$item=='Le06_PZ_09']<--2.576

### Vrij zetten groepsparameters
Svalues$est[Svalues$name=='MEAN_1']<-TRUE
Svalues$est[Svalues$name=='COV_11']<-TRUE

#Unidimensionele model

```



```
WLE<-fscores (unidim,type='WLE',full.scores.SE=TRUE)
WLE<-as.data.frame (WLE)

WLE=cbind(PIRLS1618$_LeerlingID,PIRLS1618$Groep,WLE)

mean (WLE$F1[WLE$`PIRLS1618$Groep`=="PIRLS2016"])
sqrt (var (WLE$F1[WLE$`PIRLS1618$Groep`=="PIRLS2016"]))

mean (WLE$F1[WLE$`PIRLS1618$Groep`=="PIRLS2018"])
sqrt (var (WLE$F1[WLE$`PIRLS1618$Groep`=="PIRLS2018"]))

#Wegschrijven plausible values

write.csv(WLE,file="2019-05-01_WLE20162018_iedereen_schaalpeilingen.csv")

sum(WLE$F1[PIRLS1618$Groep=='PIRLS2016']>0.10)
sum(WLE$F1[PIRLS1618$Groep=='PIRLS2018']>0.10)

#exporteren geschatte itemparameters
Parameters_IRT<-
as.data.frame(coef (unidim,simplify=TRUE,IRTpars=TRUE) $PIRLS2018$items)
Parameters_MIRT<-as.data.frame(coef (unidim,simplify=TRUE) $PIRLS2018$items)
options(max.print=100000)
sink('parameters20162018_schaalpeilingen.txt')
cbind(Parameters_IRT,Parameters_MIRT$d,Parameters_MIRT$ak0,Parameters_MIRT$
ak1,Parameters_MIRT$ak2,Parameters_MIRT$ak3,Parameters_MIRT$d0,Parameters_M
IRT$d1,Parameters_MIRT$d2,Parameters_MIRT$d3)
sink()
```

Appendix 3: Syntax model IRT met parameters van PIRLS in MIRT voor attitudes

```
## IRT-analyses PIRLS 2018 schaal leerlingvragenlijst###

#Nodige pakketten
install.packages("mirt") #Installatie MIRT 1.30 software.
install.packages("psych") #installatie psych 1.8.12 software#
install.packages("Cairo") #installatie Cairo 1.5-9 software#

library(mirt) #Activatie MIRT 1.30 software.
library(psych) #activatie psych 1.8.12 software#
library(Cairo) #activatie Cairo 1.5-9 software#

#Inlezen bestanden via read.csv()
PIRLS18=read.csv("2019-05-14 Items schalen PIRLS2018.csv", sep=";")

#Hercodering toetscodes missing naar NA.
PIRLS18[PIRLS18==999999]<-NA
PIRLS18[1:44][PIRLS18[1:44]==9]<-NA

#Behoud enkel de cases die meer dan ☒ geldig antwoord hebben op de items
van de schaal
PIRLS18<-PIRLS18[rowSums(!is.na(PIRLS18[20:28]))>1,]

#Hercodering van 4,3,2,1 naar 1,2,3,4
PIRLS18[,c(20:28)]<-5-PIRLS18[,c(20:28)]

#Beschrijvende gegevens
describe(PIRLS18[,c(20:28)])
total<-rowSums(PIRLS18[,c(20:28)],na.rm=TRUE) #Somscore 0/1 scoring
berekenen en in "Total" opslagen ter controle#
hist(total,breaks=11) #Histogram somscore ter controle#

#Configuratie IRT-model
Svalues<-mirt(PIRLS18[,c(20:28)],model=1,
itemtype=c('Rasch','Rasch','Rasch','Rasch','Rasch','Rasch','Rasch','Rasch',
'Rasch'),
SE=TRUE,
pars ="values",
technical=list(NCYCLES=2000,removeEmptyRows=TRUE))

###Set estimation to FALSE
Svalues$est<-FALSE

### Parameters PCM met 4 categorie씩items PIRLS
Svalues$value[Svalues$name=='d1' & Svalues$item=='ASBR01A']<-0.373
Svalues$value[Svalues$name=='d1' & Svalues$item=='ASBR01B']<-0.428
Svalues$value[Svalues$name=='d1' & Svalues$item=='ASBR01C']<-0.302
Svalues$value[Svalues$name=='d1' & Svalues$item=='ASBR01D']<-0.715
Svalues$value[Svalues$name=='d1' & Svalues$item=='ASBR01E']<-0.582
Svalues$value[Svalues$name=='d1' & Svalues$item=='ASBR01F']<-0.198
Svalues$value[Svalues$name=='d1' & Svalues$item=='ASBR01G']<-0.423
Svalues$value[Svalues$name=='d1' & Svalues$item=='ASBR01H']<-0.619
Svalues$value[Svalues$name=='d1' & Svalues$item=='ASBR01I']<-0.515

Svalues$value[Svalues$name=='d2' & Svalues$item=='ASBR01A']<-0.899
Svalues$value[Svalues$name=='d2' & Svalues$item=='ASBR01B']<-0.641
Svalues$value[Svalues$name=='d2' & Svalues$item=='ASBR01C']<-0.857
Svalues$value[Svalues$name=='d2' & Svalues$item=='ASBR01D']<-1.389
```



```

Svalues$value[Svalues$name=='d2' & Svalues$item=='ASBR01E']<-1.164
Svalues$value[Svalues$name=='d2' & Svalues$item=='ASBR01F']<-0.275
Svalues$value[Svalues$name=='d2' & Svalues$item=='ASBR01G']<-0.708
Svalues$value[Svalues$name=='d2' & Svalues$item=='ASBR01H']<-1.625
Svalues$value[Svalues$name=='d2' & Svalues$item=='ASBR01I']<-1.300

Svalues$value[Svalues$name=='d3' & Svalues$item=='ASBR01A']<--0.656
Svalues$value[Svalues$name=='d3' & Svalues$item=='ASBR01B']<--0.764
Svalues$value[Svalues$name=='d3' & Svalues$item=='ASBR01C']<--0.193
Svalues$value[Svalues$name=='d3' & Svalues$item=='ASBR01D']<-0.390
Svalues$value[Svalues$name=='d3' & Svalues$item=='ASBR01E']<-0.065
Svalues$value[Svalues$name=='d3' & Svalues$item=='ASBR01F']<--1.073
Svalues$value[Svalues$name=='d3' & Svalues$item=='ASBR01G']<--0.419
Svalues$value[Svalues$name=='d3' & Svalues$item=='ASBR01H']<-1.592
Svalues$value[Svalues$name=='d3' & Svalues$item=='ASBR01I']<-1.059

### Vrij zetten groepsparameters
Svalues$est[Svalues$name=='MEAN_1']<-TRUE
Svalues$est[Svalues$name=='COV_11']<-TRUE

unidim<-mirt(PIRLS18[,c(20:28)],model=1,
itemtype=c('Rasch','Rasch','Rasch','Rasch','Rasch','Rasch','Rasch','Rasch',
'Rasch'),
SE=TRUE,
pars=Svalues,
technical=list(NCYCLES=2000,removeEmptyRows=TRUE))

### Controle geschatte model
summary(unidim) #Parameterwaarden voor 3-parameter model volgens
factormodel. Ter controle.
coef(unidim,simplify=TRUE,printSE=TRUE) #Extractie parameterwaarden voor 3-
parameter model volgens (M)IRT-model. Ter controle.
coef(unidim, simplify=TRUE,IRTpars=TRUE) #Extractie parameterwaarden voor
3-parameter model volgens klassiek IRT-model. Ter controle.

### Wegeschrijven parameterwaarden volgens verschillende parameterisaties
sink('2019-05-14 Factorparameters_SERLS.txt')
summary(unidim)
sink()

sink('2019-05-14 IRTparameters_SERLS.txt')
coef(unidim,simplify=TRUE,IRTpars=TRUE)
sink()

sink('2019-05-14 MIRTparameters_SERLS.txt')
coef(unidim,simplify=TRUE,printSE=TRUE)
sink()

### Schatten van WLE-scores
FSCORESWLE<-fscores(unidim,method="WLE",full.scores.SE=TRUE)

colnames(FSCORESWLE)<-c("WLE","WLE_SE")

FSCORESWLE[,1]<-(FSCORESWLE[,1]*1.442440)+7.347685
FSCORESWLE[,2]<-(FSCORESWLE[,2]*1.442440)

summary(FSCORESWLE[,1])

Abilityestimates<-cbind(PIRLS18[,c("IDSTUD","IDCLASS")],FSCORESWLE)

```

```
#Categoriseren van geschatte scores
Abilityestimates$IDX<-cut(Abilityestimates$WLE,c(-100,7.1,9.5,100))
Abilityestimates$IDX<-(4-as.numeric(Abilityestimates$IDX))

summary(as.factor(Abilityestimates$IDX))

#Wegschrijven WLE-scores
write.csv(Abilityestimates,file="2019-05-14 Abilityestimates_SERLS.csv")
```

Appendix 4: Syntax voor exploratieve factoranalyse in Mplus voor toetsboekje 1

```

TITLE:    Onderzoek multidimensionaliteit;
DATA:     FILE IS Boek1.csv;
          Format is free;
VARIABLE:
Names are
STUDID
R41H01M R41H02M R41H05M R41H07M R41H09M R41H10M R41H11M R41H12M
R41H03C R41H04C R41H06C R41H08C R41H13C R41H14C R41H15C R41H16C
i2091 i2092 i2094 i2095 i2096
Le6PZ1 Le6PZ3 Le6PZ6 Le6PZ9 Le6PZ7
IDBOOK BIRTHY ITSEX
ASBG04 ASBGSCR ASDGSCR ATBGEAS
ATDGEAS ASBGERL ASDGERL ASBGLR
ASDGLR ATBGTJS ATDGTJS ATBG04
ATBG01 ATDG01 ATBG06R ASBG03 BENCHM ASBGHRL ASDGHRL ASRREAM
;

IDVARIABLE=STUDID;

MISSING ARE ALL (9999999);

USEVARIABLES=
R41H01M R41H02M R41H05M R41H07M R41H09M R41H10M R41H11M R41H12M
R41H03C R41H04C R41H06C R41H08C R41H13C R41H14C R41H15C R41H16C
i2091 i2092 i2094 i2095 i2096
Le6PZ1 Le6PZ3 Le6PZ6 Le6PZ9 Le6PZ7;

CATEGORICAL ARE
R41H01M R41H02M R41H05M R41H07M R41H09M R41H10M R41H11M R41H12M
R41H03C R41H04C R41H06C R41H08C R41H13C R41H14C R41H15C R41H16C
i2091 i2092 i2094 i2095 i2096
Le6PZ1 Le6PZ3 Le6PZ6 Le6PZ9 Le6PZ7 ;

Analysis:
ESTIMATOR=MLR;
TYPE = EFA 1 2;
ROTATION=GEOMIN(ORTHOGONAL);

OUTPUT:
MODINDICES;

```

Appendix 5: Syntax modellen IRT voor onderzoek unidimensionaliteit toetsboekje 1 in MIRT

```

### Test multidimensionaliteit ###

## Opzetten van omgeving om individuele toets te analyseren met IRT##

install.packages("mirt") #installatie mirt 1.29 software#
install.packages("psych") #installatie psych 1.8.10 software#
install.packages("Cairo") #installatie Cairo 1.5-9 software#

library(mirt) #activatie mirt 1.29 software#
library(psych) #activatie psych 1.8.10 software#
library(Cairo) #activatie Cairo 1.5-9 software#

Dataset1=read.csv("Boek1.csv", sep=";", encoding="UTF-8-BOM") #opladen
toetsgegevens#

head(Dataset1) #Titels kolommen nalezen ter controle#

describe(Dataset1[,c(2:27)])
alpha(Dataset1[,c(2:27)])

#Opmerking: vaak zullen we gebruik maken van "Dataset1[,c(2:27)]" in plaats
van simpelweg "Dataset1" te gebruiken. De reden hiervoor is dat de eerste
kolom variabelen zijn die we niet willen opnemen in de analyses#

total<-rowSums(Dataset1[,c(2:27)]) #Somscore 0/1 scoring berekenen en in
"Total" opslagen ter controle#
hist(total,breaks=25) #Histogram somscore ter controle#

## Vergelijking unidimensionele en multidimensionele model ##

#Unidimensioneel model

unidim<- mirt(Dataset1[,c(2:27)],model=1,
itemtype=c('2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', 'gpcm', '2PL', 'gpcm', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL'),
guess=c(0.582, 0.196, 0.286, 0.151, 0.166, 0.263, 0.311, 0.369, 0,
0, 0, 0, 0, 0, 0, 0, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15,
0.15, 0, 0.15, 0),
SE=TRUE,
technical=list(NCYCLES=2000))

summary(unidim) #Parameterwaarden voor model volgens factormodel. Ter
controle.
coef(unidim, simplify=TRUE) #Extractie parameterwaarden voor model volgens
IRT-model. Ter controle en extractie waarde gokparameter.
M2(unidim) #extractie fit indices (df,RMSEA,SRMSR,TLI,CFI) model
c(extract.mirt(unidim,"AIC"),extract.mirt(unidim,"BIC"))

#Multidimensioneel model
s<-
'F1=1-16
F2=17-26
COV = F1*F2'
Model <- mirt.model(s)

multidim<- mirt(Dataset1[,c(2:27)],model=Model,

```

```

itemtype=c('2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', 'gpcm', '2PL', 'gpcm', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL'),
guess=c(0.582, 0.196, 0.286, 0.151, 0.166, 0.263, 0.311, 0.369, 0,
0, 0, 0, 0, 0, 0, 0, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15,
0.15, 0, 0.15, 0),
SE=TRUE,
technical=list(NCYCLES=2000))

summary(multidim) #Parameterwaarden voor model volgens factormodel. Ter
controle.
coef(multidim, simplify=TRUE) #Extractie parameterwaarden voor model
volgens IRT-model. Ter controle en extractie waarde gokparameter.
M2(multidim) #extractie fit indices (df,RMSEA,SRMSR,TLI,CFI) model
c(extract.mirt(multidim,"AIC"),extract.mirt(multidim,"BIC"))

#Bifactor model
Model<-c(rep(1,16),rep(2,10))
bfmod<-bfactor(Dataset1[,c(2:27)],model=Model,
itemtype=c('2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', 'gpcm', '2PL', 'gpcm', '2PL', '2PL', '2PL', '2PL',
'2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL', '2PL'),
guess=c(0.582, 0.196, 0.286, 0.151, 0.166, 0.263, 0.311, 0.369, 0,
0, 0, 0, 0, 0, 0, 0, 0.15, 0.15, 0.15, 0.15, 0.15, 0.15,
0.15, 0, 0.15, 0))

summary(bfmod) #Parameterwaarden voor model volgens factormodel. Ter
controle.
coef(bfmod, simplify=TRUE) #Extractie parameterwaarden voor model volgens
IRT-model. Ter controle en extractie waarde gokparameter.
M2(bfmod) #extractie fit indices (df,RMSEA,SRMSR,TLI,CFI) model
c(extract.mirt(bfmod,"AIC"),extract.mirt(bfmod,"BIC"))

```

Appendix 6: Syntax modellen IRT voor onderzoek unidimensionaliteit toetsboekje 1 in Mplus

```

TITLE:    Onderzoek multidimensionaliteit;
DATA:     FILE IS Boek1.csv;
          Format is free;
VARIABLE:
Names are
STUDID
R41H01M R41H02M R41H05M R41H07M R41H09M R41H10M R41H11M R41H12M
R41H03C R41H04C R41H06C R41H08C R41H13C R41H14C R41H15C R41H16C
i2091 i2092 i2094 i2095 i2096
Le6PZ1 Le6PZ3 Le6PZ6 Le6PZ9 Le6PZ7
IDBOOK BIRTHY ITSEX
ASBG04 ASBGSCR ASDGSCR ATBGEAS
ATDGEAS ASBGERL ASDGERL ASBGLR
ASDGLR ATBGTJS ATDGTJS ATBG04
ATBG01 ATDG01 ATBG06R ASBG03 BENCHM ASBGHRL ASDGHRL ASRREAM
;

IDVARIABLE=STUDID;

MISSING ARE ALL (9999999);

USEVARIABLES=
R41H01M R41H02M R41H05M R41H07M R41H09M R41H10M R41H11M R41H12M
R41H03C R41H04C R41H06C R41H08C R41H13C R41H14C R41H15C R41H16C
i2091 i2092 i2094 i2095 i2096
Le6PZ1 Le6PZ3 Le6PZ6 Le6PZ9 Le6PZ7;

CATEGORICAL ARE
R41H01M(3pl) R41H02M(3pl) R41H05M(3pl) R41H07M(3pl) R41H09M(3pl) R41H10M(3pl)
R41H11M(3pl)
R41H12M(3pl)
i2091(3pl) i2092(3pl) i2094(3pl) i2095(3pl) i2096(3pl)
Le6PZ1(3pl) Le6PZ3(3pl) Le6PZ9(3pl);

CATEGORICAL ARE
R41H06C(gpcm) R41H13C(gpcm);

CATEGORICAL ARE
R41H03C R41H04C R41H08C R41H14C R41H15C R41H16C
Le6PZ6 Le6PZ7;

Analysis:
ESTIMATOR=MLR;

Model:
ABIL1 by
R41H01M* R41H02M* R41H05M* R41H07M* R41H09M* R41H10M* R41H11M* R41H12M*

```

R41H03C* R41H04C* R41H06C* R41H08C* R41H13C* R41H14C* R41H15C* R41H16C* ;

[ABIL1@0];

ABIL1@1;

ABIL2 by

i2091* i2092* i2094* i2095* i2096*

Le6PZ1* Le6PZ3* Le6PZ6* Le6PZ9* Le6PZ7*;

[ABIL2@0];

ABIL2@1;

[R41H01M\$2@-0.331];

[R41H02M\$2@1.411];

[R41H05M\$2@0.915];

[R41H07M\$2@1.727];

[R41H09M\$2@1.614];

[R41H10M\$2@1.030];

[R41H11M\$2@0.795];

[R41H12M\$2@0.537];

[i2091\$2@1.735];

[i2092\$2@1.735];

[i2094\$2@1.735];

[i2095\$2@1.735];

[i2096\$2@1.735];

[Le6PZ1\$2@1.735];

[Le6PZ3\$2@1.735];

[Le6PZ9\$2@1.735];

PLOT:

TYPE=PLOT3;

OUTPUT:

STANDARDIZED;

SVALUES;

SAVEDATA:

FILE IS TwoFactorScores.DAT;

SAVE = FSCORES;

