

This item is the archived peer-reviewed author-version of:

Twee jaar na TIMSS-2019 : TIMSS-repeat en de leerwinst in Vlaanderen

Reference:

Verhelst Dries, Gijsen Marijn, Appels Lies, De Maeyer Sven, Van Petegem Peter.- Twee jaar na TIMSS-2019 : TIMSS-repeat en de leerwinst in Vlaanderen
Tijdschrift voor onderwijsrecht en onderwijsbeleid - ISSN 0778-0443 - 1-2(2023-2024), p. 87-102
To cite this reference: <https://hdl.handle.net/10067/2024140151162165141>

Twee jaar na TIMSS-2019: TIMSS-repeat en de Leerwinst in Vlaanderen

Dries Verhelst¹, Marijn Gijsen², Lies Appels², Sven De Maeyer³ en Peter Van Petegem³

I. TIMSS in Vlaanderen

Vlaanderen tekent al verschillende cycli in op de TIMSS-onderzoeken. TIMSS staat voor *Trends in International Mathematics and Science Study* en is een internationaal vergelijkend onderzoek dat zich richt op wiskunde en wetenschapsprestaties van leerlingen in het vierde leerjaar (Mullis et al., 2020). Doel is om leerprestaties voor wiskunde en wetenschappen in kaart te brengen, te meten en zowel internationaal als temporeel te vergelijken. De internationale onderzoeksleiding ligt bij de IEA oftewel *the International Association for the Evaluation of Educational Achievement* en Boston College.

Vlaanderen neemt sinds 1995 deel aan het TIMSS onderzoek. De resultaten van het TIMSS-onderzoek bieden steeds een scala aan beleidsrelevante inzichten. De cyclus in 2003 participeert Vlaanderen I ook met de doelpopulatie van het vierde leerjaar basisonderwijs (Tabel 1). Vanaf 2011 wordt er enkel nog met het vierde leerjaar deelgenomen. Door om de vier jaar de TIMSS-toets af te nemen, wordt het mogelijk om trends in prestaties op te volgen.

Tabel 1: deelname van Vlaanderen aan TIMSS

Doelpopulatie	CYCLI						
	1995	1999	2003	2007	2011	2015	2019
4 ^{de} leerjaar Basionderwijs			X		X	X	X
2 ^{de} leerjaar Secundair onderwijs	X	X	X				

Bron: Rapport TIMSS 2019
(Faddar et al., 2020)

Daarnaast biedt TIMSS ook mogelijkheden om het Vlaamse prestatieniveau af te toetsen aan internationale maatstaven en te vergelijken met andere landen. De TIMSS-onderzoeken situeren de Vlaamse prestatie immers in een internationale ranking. Bovendien vullen de leerlingen naast de vaardigheidstesten voor wiskunde en wetenschappen ook contextvragenlijsten in. Ook bij de ouders, leerkrachten en scholen wordt heel wat informatie verzameld wat het mogelijk maakt om diverse contextfactoren mee op te nemen in de beschouwing van de Vlaamse prestaties (Faddar et al., 2020; Mullis et al., 2020).

II. De dip van 2019: wiskunde en wetenschappen in het slop

Bij de publicatie van het TIMSS-2019 onderzoek bleek dat de Vlaamse leerlingen uit het vierde leerjaar basisonderwijs een duidelijke terugval kenden wat betreft hun wiskunde en wetenschapsprestaties (Faddar et al., 2020). Wiskunde kende tussen 2015 en 2019 een significante terugval van 14 punten, wetenschappen eveneens met significante daling van 11 punten in vergelijking met 2015. Deze daling zette zich door overheen zo goed als alle cognitieve en inhoudelijke domeinen. Voor wiskunde en wetenschappen worden volgende cognitieve domeinen

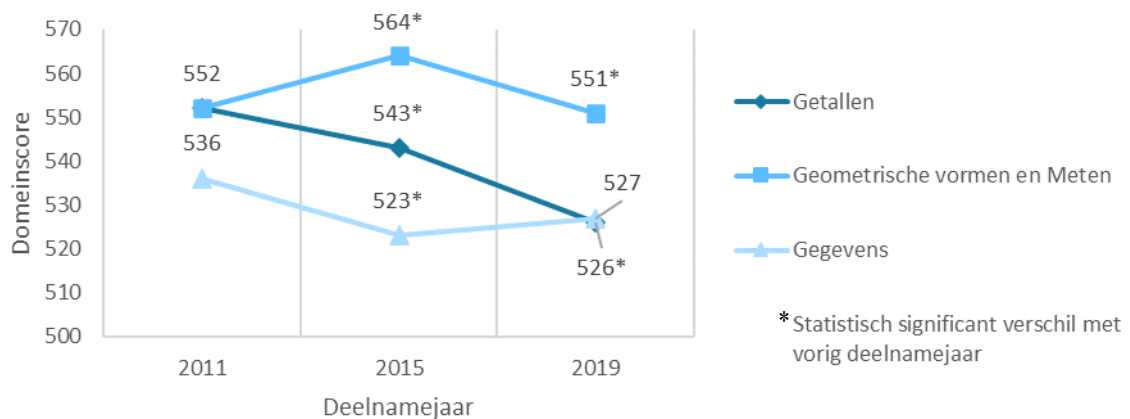
¹ Onderzoekskoördinator TIMSS 2023 & TIMSS-repeat, Universiteit Antwerpen

² Projectmedewerker TIMSS-repeat, Universiteit Antwerpen

³ Promotor TIMSS 2023 en TIMSS-repeat, Universiteit Antwerpen

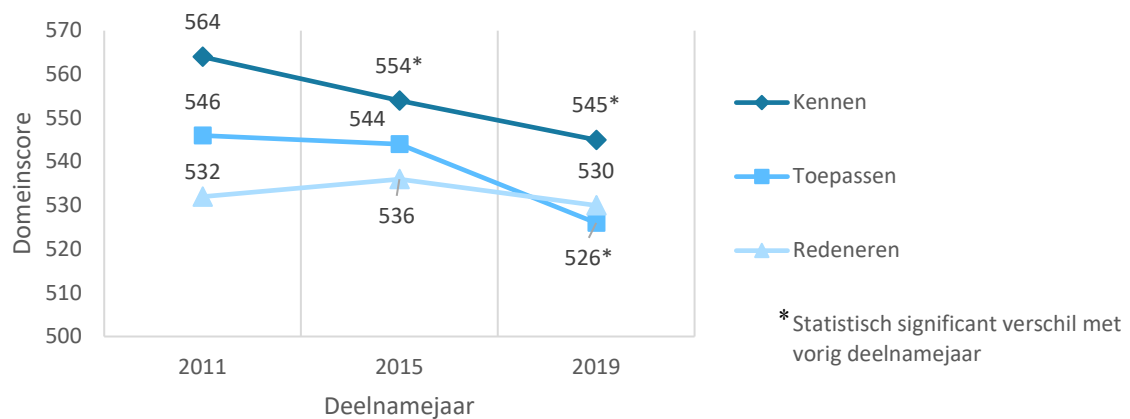
onderscheiden: *Kennen, Toepassen en Redeneren*. De inhoudelijke domeinen verschillen voor wiskunde en wetenschappen. Bij wiskunde gaat het om *Getallen, Gegevens en Geometrische vormen en Meten*. Bij wetenschappen om *Biologie, Natuurkunde en Aardrijkskunde*.

Voor wiskunde is er een duidelijke daling ten opzichte van 2015 merkbaar voor de inhoudelijke domeinen *Getallen* en *Geometrische vormen & Meten* (Figuur 1). Opmerkelijk is dat het domein *Getallen* sinds 2011 een duidelijke negatieve trend kent terwijl *Geometrische vormen & Meten* tussen 2011 nog een stijging kende, om dan in 2019 terug te vallen. Het domein *Gegevens* kent in 2019 geen merkbare toe- of afname ten opzichte van 2015 (Faddar et al., 2020).



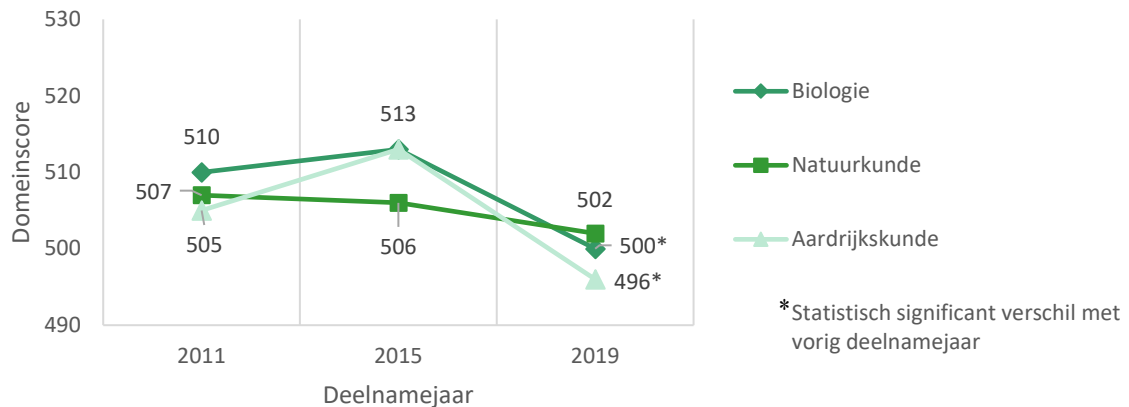
Figuur 1: Vlaamse wiskundescore in inhoudelijke domeinen doorheen de tijd (overgenomen uit Rapport TIMSS 2019 (Faddar et al., 2020))

Ook de cognitieve domeinen *Kennen* en *Toepassen* kennen een significantie daling tussen 2015 en 2019 (Figuur 2). Voor *Redeneren* blijft het prestatieniveau vrij gelijklopend (Faddar et al., 2020).



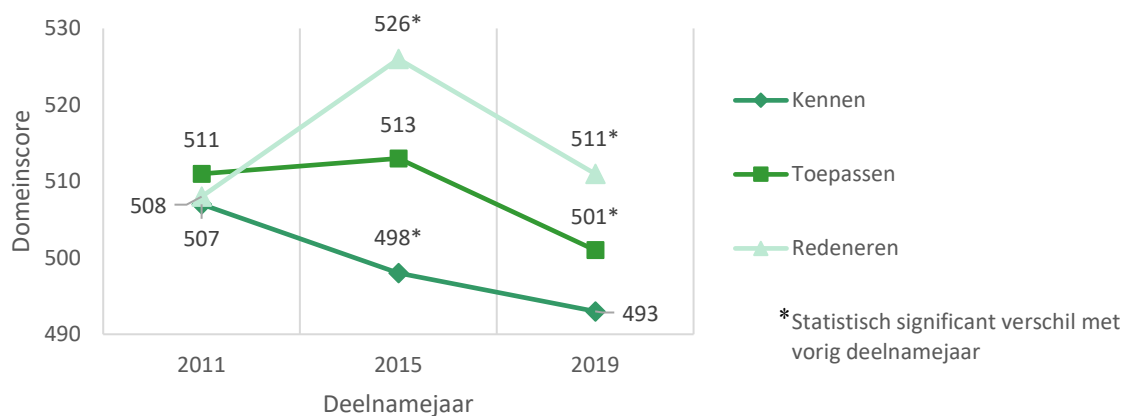
Figuur 2: Vlaamse wiskundescore in cognitieve domeinen doorheen de tijd (overgenomen uit Rapport TIMSS 2019 (Faddar et al., 2020))

Ook voor wetenschappen is de terugval overheen de verschillende deeldomeinen treffend. Ten opzichte van 2015 scoorden de Vlaamse leerlingen in 2019 17 punten lager voor *Aardrijkskunde* en 13 punten lager voor *Biologie*. *Natuurkunde* kent een niet-significante daling van 4 punten. Doorheen de tijd blijkt *Natuurkunde* minder fluctuaties kent in vergelijking met de andere domeinen, al kant het wel een constante dalende trend.



Figuur 3: Vlaamse wetenschapsscore in inhoudelijke domeinen doorheen de tijd (overgenomen uit Rapport TIMSS 2019 (Faddar et al., 2020))

De drie cognitieve domeinen bij wetenschappen kennen ook een terugval tussen 2015 en 2019. Het domein *Kennen* zit in een dalende trend sinds 2011. *Toepassen* bleef relatief stabiel tussen 2011 en 2015 om dan in 2019 een sterke daling te kennen. Opvallend is de bijzonder sterke stijging voor *Redeneren* tussen 2011 en 2015, gevolgd door de opmerkelijke daling in 2019 (Faddar et al., 2020).



Figuur 4: Vlaamse wetenschapsscore in cognitieve domeinen doorheen de tijd (overgenomen uit Rapport TIMSS 2019 (Faddar et al., 2020))

III. TIMSS-repeat: wat gebeurt er tijdens de twee laatste jaren van het Vlaamse basisonderwijs?

De sterke achteruitgang maakte de nood aan vervolgonderzoek duidelijk, waarop in navolging van TIMSS 2019 het TIMSS-repeat onderzoek opgezet werd.

TIMSS-repeat is een herhalingsmeting op TIMSS 2019. Hierbij werd nagegaan in welke mate de Vlaamse leerlingen leerwinst boeken doorheen de twee laatste jaren van het Vlaamse basisonderwijs, specifiek dan voor wiskunde en wetenschappen. Hiervoor werden de leerlingen die in 2019 deelnamen in het 4^{de} leerjaar, twee jaar later opnieuw bevroegd toen ze in het 6^{de} leerjaar zaten (of het 5^{de} leerjaar voor de zittenblijvers).

In vergelijking met de traditionele TIMSS-onderzoeken, kent TIMSS-repeat enkele belangrijke verschillen. Ten eerste focust TIMSS-repeat enkel op het Vlaams onderwijs en wordt er geen internationale vergelijking gemaakt. De focus van TIMSS-repeat gaat dus uit naar de evolutie van de

Vlaamse leerlingen binnen de Vlaamse onderwijscontext. Ten tweede zijn de leerlingen die aan TIMSS-repeat deelnamen, dezelfde als zij die in 2019 reeds deelnamen aan TIMSS. Dit terwijl TIMSS traditioneel steeds nieuwe leerlingen bevaart (telkens in het 4^{de} leerjaar). Ten derde zijn er ook belangrijke verschillen in de gehanteerde meetinstrumenten en vragenlijsten. In de toets voor de leerlingen werden verschillende items vervangen. Dit om een plafondeffect te vermijden. Bij de contextvragenlijsten is het belangrijk om te vermelden dat TIMSS-repeat geen ouders bevaart heeft; enkel de leerlingen, leerkrachten en de schooldirecties namen deel.

Het huidige artikel gaat in op de volgende onderzoeksvraag die centraal stond in het TIMSS-repeat onderzoek (Verhelst et al., 2022): “Wat is de geboekte leerwinst van de leerlingen tussen het vierde en het zesde leerjaar op de TIMSS-toetsen voor wiskunde en wetenschappen?”.

IV. Methode en analyses

In onderstaande secties gaan we dieper in op het praktisch opzet van de herhalingsmeting TIMSS-repeat en de gehanteerde methodes en analyses.

Steekproef

De steekproef van TIMSS-repeat bestaat uit scholen die in 2019 ook deelnamen aan TIMSS 2019. Net zoals de deelname aan TIMSS 2019, gebeurde de deelname aan TIMSS-repeat geheel op vrijwillige basis. 133 scholen stemden in om deel te nemen aan TIMSS-repeat, wat neerkomt op 91,9% van de oorspronkelijke steekproef uit 2019. De TIMSS-steekproef is een gestratificeerde, tweestapssteekproef die steeds representatief is op leerlingniveau (Fishbein, 2021). Om na te gaan of de representativiteit behouden bleef in TIMSS-repeat werd er zowel op schoolniveau als op leerlingniveau gekeken of beide steekproeven voldoende gelijkend zijn.

Steekproef op schoolniveau

Op schoolniveau blijkt dat de geografische spreiding van de deelnemende scholen overheen de provincies en het Brussel-Hoofdstedelijk Gewest, vrijwel gelijk loopt overheen beide steekproeven (Tabel 2).

Tabel 2: verdeling van de scholen overheen de provincies in 2019 en 2021

	TIMSS 2019	TIMSS-repeat
Antwerpen	31%	33%
Brussels-Hoofdstedelijk Gewest	6%	6%
Limburg	16%	16%
Oost-Vlaanderen	16%	16%
Vlaams Brabant	16%	13%
West-Vlaanderen	16%	15%

Bron: Rapport TIMSS-repeat
(Verhelst et al., 2022)

Kijkend naar de verhoudingen op vlak van de onderwijsnetten, zien we ook dat de steekproef van TIMSS-repeat gelijkaardig is aan de steekproef van TIMSS 2019 (Tabel 3).

Tabel 3. Verdeling van de scholen overheen de onderwijsnetten in 2019 en 2021

	TIMSS 2019	TIMSS-repeat
Gemeenschapsonderwijs	18%	19%
Officieel gesubsidieerd onderwijs	27%	28%
Vrij gesubsidieerd onderwijs	55%	53%

Bron: Rapport
TIMSS-repeat
(Verhelst et al.,
2022)

Een ander belangrijk aspect, is de deelname aan TIMSS-repeat op basis van de (school)resultaten bij een eerdere deelname. Immers, onderzoek uitgevoerd door de IEA toont aan dat non-respons frequenter voorkomt bij scholen die eerder minder goed presteerden. Daarom controleerden we ook of scholen die in 2021 deelnamen significant hoger (of lager) scoorden in vergelijking met scholen uit 2019 die niet deelnamen aan TIMSS-repeat. Hieruit blijkt dat scholen in de steekproef van TIMSS-repeat niet significant hoger of lager scoren voor de TIMSS-toets in vergelijking met scholen die enkel in 2019 deelnamen.

Deze vergelijking op schoolniveau toont dus aan dat er geen sprake is van een selectie-effect wat betreft de provincie, onderwijsnet of eerdere prestaties.

Steekproef op leerlingniveau

Naast de vergelijking op schoolniveau, is het ook van belang dat beide steekproeven vergelijkbaar zijn op leerlingniveau. Daarom gaan we na in welke mate de steekproef van TIMSS-repeat vergelijkbaar is met de steekproef van TIMSS 2019 op vlak van volgende leerlingkenmerken: geslacht, thuistaal en indicatoren voor sociaaleconomische status (SES). Zoals eerder aangegeven, werden er geen ouders bevroegd binnen TIMSS-repeat. Dit maakt dat we op vlak van SES de “*Home Resources for Learning-schaal*” niet kunnen hanteren daar deze gebruik maakt van zowel leerling- als ouderdata (Martin et al., 2020; Mullis et al., 2020). Om de SES van de leerlingen in kaart te brengen kijken we daarom naar verschillende indicatoren die bij de leerlingen bevroegd werden zoals het aantal boeken dat een leerling thuis heeft en het al dan niet beschikken over een eigen kamer.

De verdeling naar geslacht blijkt identiek tussen TIMSS 2019 en TIMSS-repeat. Er deden 2178 meisjes en 2123 jongens mee aan TIMSS-repeat, tabel 4 geeft de verhouding tussen jongens en meisjes weer voor de twee steekproeven.

Tabel 4. Geslacht in 2019 en 2021

	2019	TIMSS-repeat
Meisjes	51%	51%
Jongens	49%	49%

Bron: Rapport TIMSS-repeat (Verhelst et al., 2022)

Ook op vlak van thuistaal zijn beide steekproeven zeer goed vergelijkbaar (Tabel 5).

Tabel 5. Thuistaal in 2019 en 2021

	2019	TIMSS-repeat
Altijd Nederlands	62%	66%
Meestal Nederlands	11%	12%
Soms Nederlands, soms een andere taal	22%	19%
Nooit Nederlands	4%	4 %

Bron: Rapport TIMSS-repeat (Verhelst et al., 2022)

Het aantal boeken dat een leerling thuis heeft werd binnen TIMSS-repeat als een proxy gehanteerd voor SES. Hierbij worden vijf categorieën onderscheiden:

- Geen of erg weinig boeken (0-10 boeken)
- Genoeg boeken om één rek te vullen (11-25 boeken)
- Genoeg boeken om één boekenkast te vullen (26-100 boeken)
- Genoeg boeken om twee boekenkasten te vullen (101-200 boeken)
- Genoeg boeken om drie of meer boekenkasten te vullen (Meer dan 200 boeken)

Uit de vergelijking in Tabel 6 blijkt dat ook wat betreft het aantal boeken thuis beide steekproeven goed te vergelijken zijn.

Tabel 6. Aantal boeken thuis in 2019 en 2021

	2019	TIMSS-repeat
Meer dan 200 boeken	9%	9%
101-200 boeken	15%	17%
26-100 boeken	36%	36%
11-25 boeken	27%	23%
0-10 boeken	13%	15%

Bron: Rapport TIMSS-repeat
(Verhelst et al., 2022)

Deze vergelijking op leerlingniveau leert ons dat de beide steekproeven vergelijkbaar samengesteld zijn. Hieruit kunnen we dus concluderen dat de steekproef van TIMSS-repeat ook representatief is op leerlingniveau.

Ontwikkeling toetsmaterialen

De gebruikte toetsmaterialen van TIMSS-repeat werden grondig tegen het licht gehouden om bias en ongewenste effecten te vermijden. Hieronder gaan we in op de itemselectie van TIMSS-repeat en op de verdeling van de items overheen de verschillende toetsboekjes en leerlingen.

Itemselectie

In de eerste fase van de itemselectie werd er gekeken naar de items uit de toets van TIMSS 2019. Om een plafondeffect te vermijden (dat wil zeggen, wanneer de items te makkelijk zijn waardoor het prestatieniveau plafonneert), werden de makkelijkste items uit TIMSS 2019 niet weerhouden. Die items werden vervangen door moeilijkere items uit het peilingsonderzoek wiskunde en wereldoriëntatie (STEP, 2015, 2021). Naast de moeilijkheid van de items, werd er ook rekening gehouden met de verdeling overheen de inhoudelijke en cognitieve domeinen zoals deze beschreven zijn binnen TIMSS (Mullis et al., 2020; Mullis and Martin, 2017). Tenslotte werd bij de selectie en samenstelling van de testitems ook rekening gehouden met de verhouding tussen de verschillende formats: *antwoorditems, meerkeuzevragen, open vragen en doevragen*.

Door een minutieuze selectie van de items op basis van de hierboven omschreven criteria, was het mogelijk om een toets op te stellen die de vaardigheden van de leerlingen in 2021 op een betrouwbare en valide manier kon vastleggen. Tabel 8 t.e.m. 12 bieden een overzicht van de itemverdeling op basis van de criteria voor zowel 2019 en 2021.

Tabel 7. Soorten vragen binnen wiskunde en wetenschappen

	Wiskunde		Wetenschappen	
	TIMSS 2019	TIMSS-repeat	TIMSS 2019	TIMSS-repeat
Antwoord	76	85	27	20
Multiple choice	76	65	104	116
Open vraag	7	11	44	38
Doevraag	14	12	0	1

Bron: Rapport TIMSS-repeat
(Verhelst et al., 2022)

Tabel 8. Verdeling domeinen wiskunde

	TIMSS 2019	TIMSS-repeat
Inhoudelijke domeinen		

Bron: Rapport TIMSS-repeat
(Verhelst et al., 2022)

Getallen	84	49%	84	49%
Geometrische vormen en Meten	51	29%	52	30%
Gegevens	38	22%	37	21%
Cognitieve domeinen				
Kennen	62	36%	59	34%
Toepassen	73	42%	73	42%
Redeneren	38	22%	41	24%

Tabel 9. Verdeling domeinen wetenschappen

	TIMSS 2019		TIMSS-repeat	
Inhoudelijke domeinen				
Biologie	78	45%	78	45%
Natuurkunde	62	36%	62	36%
Aardrijkskunde	35	20%	35	20%
Cognitieve domeinen				
Kennen	62	36%	59	34%
Toepassen	73	42%	73	42%
Redeneren	38	22%	41	24%

Bron: Rapport TIMSS-repeat (Verhelst et al., 2022)

Samenstelling toetsboekjes

Na de selectie van de testitems werden deze gecombineerd in 14 toetsboekjes, evenveel als bij TIMSS 2019. Ieder toetsboekje bestond uit een deel wiskunde en een deel wetenschappen waarbij de even boekjes starten met wetenschappen en de oneven met wiskunde. Om een hertestingseffect te vermijden werd met twee zaken rekening gehouden bij de verdeling van de toetsboekjes overheen de leerlingen (Scharfen et al., 2018). Ten eerste werd ervoor gezorgd dat leerlingen niet dezelfde items als in TIMSS 2019 voorgelegd kregen. Ten tweede werd er ook rekening gehouden met de toetsafname van het peilingsonderzoek wiskunde 2021 die tijdens dezelfde periode als TIMSS-repeat gepland was (STEP, 2021). Hierbij werd nagegaan dat leerlingen die zowel deelnamen aan TIMSS-repeat als aan de peilingstoets geen overlap in items hadden op beide testen, aangezien een aantal items uit de peilingstoets ook opgenomen waren in de TIMSS-repeattoets.

Afname toetsen en vragenlijsten

De afname van TIMSS-repeat vond plaats in mei 2021. Hierbij werd geopteerd om deze afname zo te organiseren dat de procedures gelijk waren aan die van de afname van TIMSS 2019. Zo werden de leerlingtoetsen en vragenlijsten ook op papier ingevuld (de leerkrachten en directies vulden hun vragenlijsten online in). De scholen konden deze afname zelf organiseren tussen 3 mei en 28 mei. Hiervoor ontvingen de scholen een duidelijk draaiboek dat strikt gevolgd diende te worden.

Dataverwerking en analyses

Na de dataverzameling werd de data gecleand om er vervolgens de nodige analyses op te kunnen doen. De TIMSS-data heeft een zekere eigenheid waardoor specifieke stappen genomen dienen te worden alvorens de analyses kunnen gebeuren. Zo werkt TIMSS-data niet met een enkele puntschatting om de vaardigheid van de leerlingen te berekenen. Per leerling worden vijf *plausible values* berekend voor de algemene prestaties en de diverse deeldomeinen. Aan de hand van deze *plausible values* wordt de vaardigheid van iedere leerling vastgesteld op de TIMSS-schaal, de welke een bereik heeft van 0 tot 1000 punten (Fishbein, 2021; Martin et al., 2020). Internationaal wordt het gemiddelde vastgelegd op 500 met een standaardafwijking van 100. Omdat TIMSS-repeat focust op de Vlaamse context, zonder een internationale vergelijking, werden de itemparameters opnieuw

geschat op basis van een dataset die de steekproeven uit 2019 en 2021 combineert. Vervolgens werd een lineaire transformatie toegepast om de resultaten van TIMSS-repeat toch eenduidig te kunnen interpreteren. Daarnaast wordt er ook gebruik gemaakt van gewichten om te compenseren voor de clustering eigen aan de steekproeftrekking. Voor een diepgaande toelichting over de dataverwerking binnen TIMSS en TIMSS repeat verwijzen we naar de technische handleiding van TIMSS (Martin et al., 2020) en het volledige rapport van TIMSS-repeat (Verhelst et al., 2022).

Alle analyses gebeurden in R met het pakket EdSurvey (Bailey, 2020). Dit pakket is specifiek gericht internationaal vergelijkend onderwijsonderzoek zoals TIMSS en laat een vlot gebruik van de plausible values en weights toe (Bailey et al., 2021).

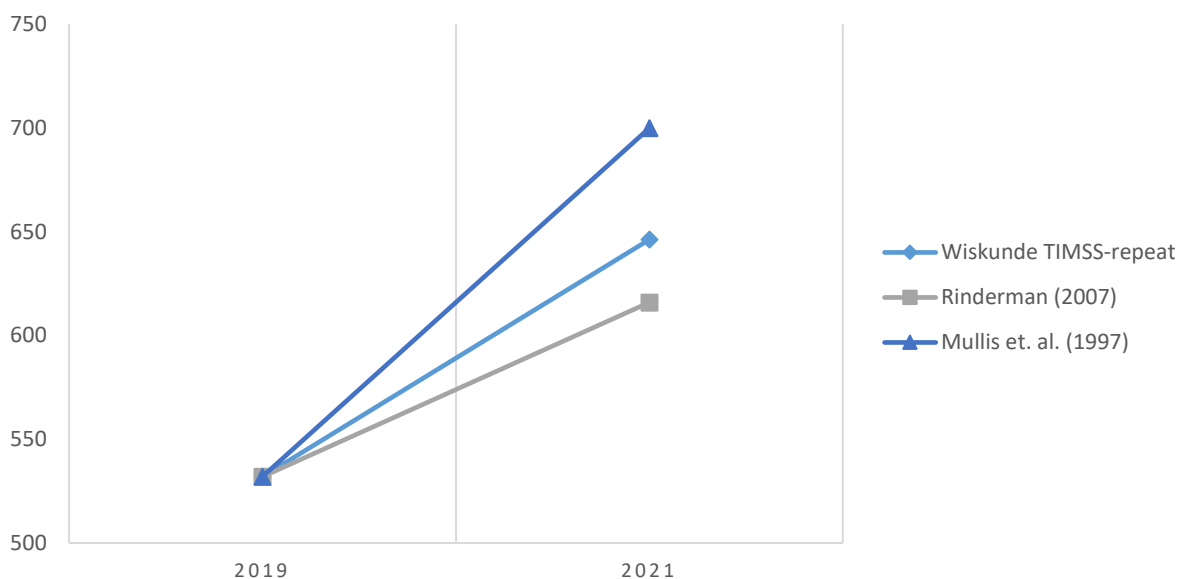
V. Leerwinst bij wiskunde en wetenschappen van het vierde naar het zesde leerjaar

In de volgende sectie gaan we in op de leerwinst die de Vlaamse leerlingen maakten tussen 2019 en 2021. Hierbij kijken we naar de leerwinst voor wiskunde en wetenschappen op populatieniveau maar gaan we ook in op de samenhang tussen leerlingkenmerken zoals geslacht, thuistaal en SES met de geboekte leerwinst.

Om de leerwinst uit te drukken rapporteren we zowel de leerwinst uitgedrukt in punten als in een effectgrootte. Om de leerwinst te interpreteren geven we ook vergelijkingspunten mee. Zo kijken we onder andere naar de leerwinst die gerapporteerd werd in een longitudinaal TIMSS-onderzoek uit de jaren negentig in Nederland (Martin et al., 1997; Mullis et al., 1997) en een theoretische schatting voor de leerwinst die verwacht mag worden (Rindermann, 2007). Daarnaast vergelijken we de effectgrootte uit TIMSS-repeat ook met de effectgrootte voor leerwinst gerapporteerd door Bloom et al. (2008) in de Verenigde Staten. We zetten de gevonden leerwinst voor wiskunde en wetenschappen ook af ten opzichte van de resultaten van PIRLS-repeat, een herhalingsmeting rond begrijpend lezen met een gelijkaardig opzet aan TIMSS-repeat (Dockx et al., 2019). Hierbij merken we voorafgaandelijk op dat alle gevonden vergelijkingspunten vrij ver verwijderd staan van de Vlaamse context, ofwel liggen ze ver in het verleden, ofwel zijn ze te situeren in een heel andere context. De huidige literatuur biedt echter geen vergelijkingsbasis die geschikter is.

Wiskunde

De gemiddelde leerling in Vlaanderen behaalt 117 punten leerwinst tijdens de twee laatste jaren van het basisonderwijs (Figuur 5). In vergelijking met het theoretisch construct van Rinderman (die aangeeft dat er per leerjaar 42 punten leerwinst te verwachten zijn) scoort de Vlaamse leerling dus beter. Echter, in vergelijking met de resultaten van Nederlandse leerlingen doen de Vlaamse leerlingen het minder goed. De Nederlandse leerlingen behaalden immers 84 punten leerwinst op één jaar onderwijs (Mullis et al., 1997).



Figuur 5: leerwinst wiskunde

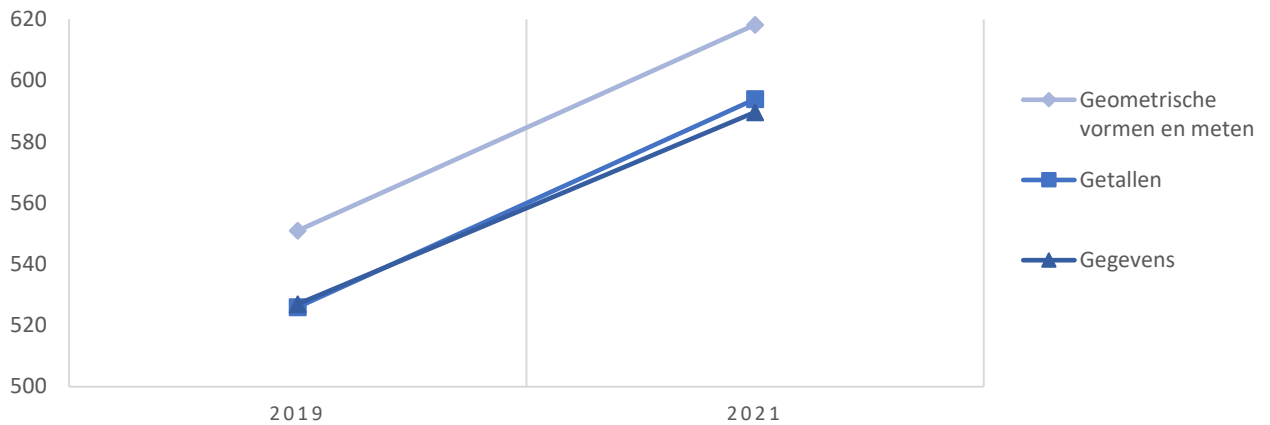
We kunnen de leerwinst ook uitdrukken in een effectgrootte. Dit is een statistische maatstaf die de grootte van het verschil tussen twee groepen aangeeft. In deze context houdt de effectgrootte dus het effect van twee jaar onderwijs op de leerprestatie in. Dit houdt in dat wanneer we kijken naar de effectgrootte van twee jaar onderwijs, van leerjaar vier naar leerjaar zes, voor wiskunde (Tabel 12), we zien dat de Vlaamse leerlingen 0,99 standaarddeviaties toenemen. Dit is iets meer dan wat gerapporteerd werd door Bloom et al. (2008) basis van gegevens uit de VS in de jaren '90. De Nederlandse leerlingen behaalden wel opvallend meer leerwinst (2,58 standaarddeviaties op twee jaar). We merken hierbij op dat de leerwinst vaak sterker is in de lagere leerjaren in vergelijking met hogere leerjaren en dat de Nederlandse cijfers gebaseerd zijn op het traject derde naar vierde leerjaar (Bloom et al., 2008; Mullis et al., 1997).

Het bekijken van de leerwinst als een effectgrootte stelt ons ook in staat om na te gaan hoe de leerwinst wiskunde verschilt met de vastgestelde leerwinst bij Vlaamse leerlingen voor begrijpend lezen op basis van het PIRLS-onderzoek (Dockx et al., 2019). Daarbij stellen we vast dat de leerwinst wiskunde vrij gelijk loopt tussen Vlaanderen en de bevindingen van Bloom et al. (2008). Voor begrijpend lezen is er echter wel een discrepantie te zien (Tabel 10).

Tabel 10: effectgrootte van twee jaar onderwijs op wiskunde

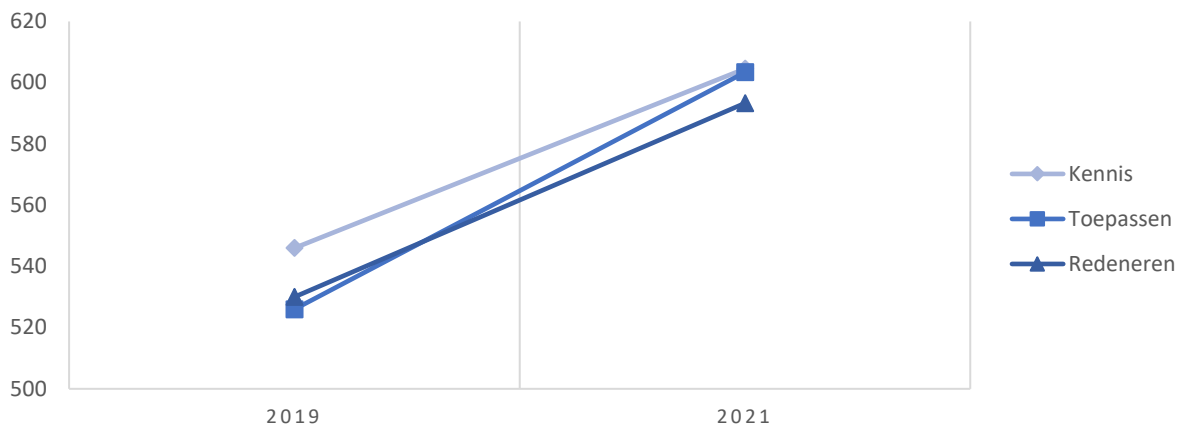
	TIMSS-repeat	Bloom et. al., 2008	Mullis et al., 1997
Wiskunde	0,99 SD	0,97 SD	2,58 SD
Begrijpend Lezen (Dockx et al., 2019)	0,97 SD	0,72 SD	n.v.t.

Wanneer we kijken naar de deeldomeinen zien we overheen de inhoudelijke domeinen wiskunde een vrij gelijke trend (Figuur 6). De respectievelijke toename van 67 punten leerwinst voor *Geometrische vormen en meten*, 68 punten voor *Getallen*, en 63 punten voor *Gegevens* maakt dus dat de verhoudingen tussen inhoudelijke deeldomeinen in 2021 gelijk is aan de verhoudingen in 2019. De Vlaamse leerlingen presteren verhoudingsgewijs nog steeds beter voor *Geometrische vormen en meten* in vergelijking met de andere domeinen.



Figuur 6: leerwinst wiskunde naar inhoudelijke deeldomeinen

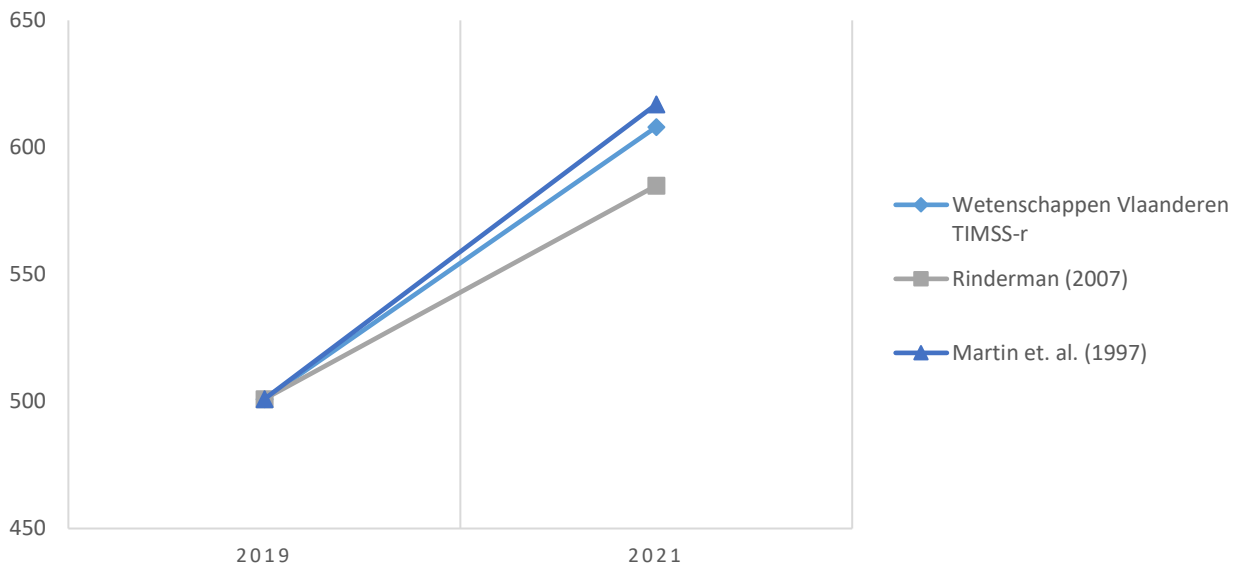
Bij de cognitieve domeinen is wel een opmerkelijke verschuiving zichtbaar: binnen het domein *Toepassen* hebben de Vlaamse leerlingen met 77 punten merkbaar meer vooruitgang geboekt. *Kennen* en *Redeneren* boeken respectievelijk 59 en 63 punten leerwinst (Figuur 7).



Figuur 7: leerwinst wiskunde naar cognitieve domeinen

Wetenschappen

Voor wetenschappen boekten de Vlaamse leerlingen 107 punten leerwinst overheen twee jaar (Figuur 8). Dit is meer dan wat verwacht kan worden op basis van Rindermann (2007) die 84 punten vooropstelt voor twee jaar onderwijs. In vergelijking met de leerwinst van de Nederlandse leerlingen in 1997, met 58 punten per jaar, zien we dat de Vlaamse scholieren tussen 2019 en 2021 net iets minder leerwinst boeken (Martin et al., 1997). Opvallend is wel dat het verschil voor wiskunde, met 51 punten groter blijkt te zijn dan voor wetenschappen waar het verschil slechts 9 punten bedraagt.



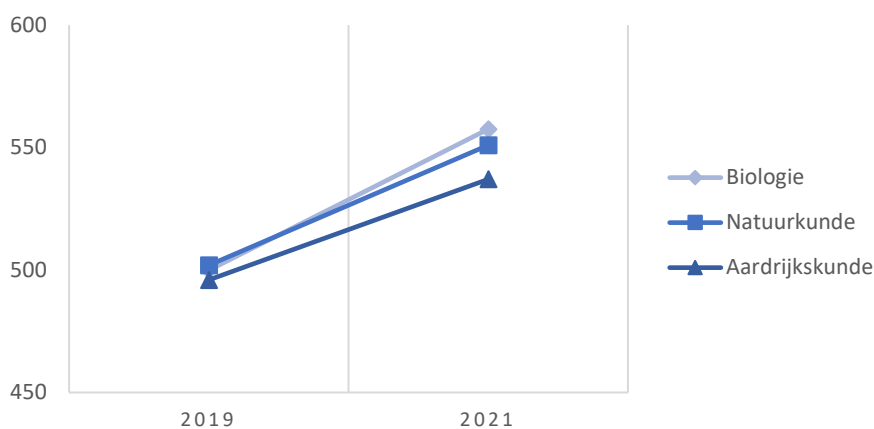
Figuur 8: leerwinst wetenschappen

Wanneer we kijken naar de effectgrootte van twee jaar onderwijs op wetenschappen, zien we ook hier dat de kloof met de Nederlandse leerlingen uit 1997 beduidend kleiner is dat bij wiskunde. Binnen Vlaanderen zelf is de effectgrootte tussen wiskunde en wetenschappen echter identiek, mat(?) 0,99 standaarddeviaties (Tabel 11).

Tabel 11: effectgrootte van twee jaar onderwijs op wetenschappen

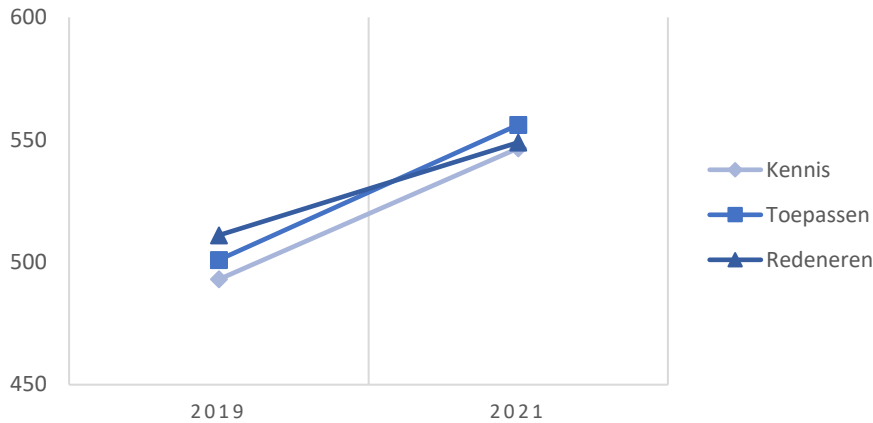
	TIMSS-repeat	Bloom et. al., 2008	Martin et al., 1997
Wetenschappen	0,92 SD	0,67 SD	1,84
Begrijpend Lezen (Dockx et al., 2019)	0,97 SD	0,72 SD	n.v.t.

Bij de inhoudelijke domeinen voor wetenschappen, zien we de sterkste stijging in het domein *Biologie*, met 58 punten (Figuur 9). Hierdoor is dit het domein in 2021 waar de Vlaamse leerlingen het sterkst op scoren terwijl dit in 2019 nog voor *Natuurkunde* was. Voor *Natuurkunde* bedraagt de leerwinst 49 punten. De leerwinst voor het domein *Aardrijkskunde* is iets beperkter met 41 punten. In 2019 was dit ook al het inhoudelijke domein waar de Vlaamse leerlingen het minst goed voor presteerden. De kloof tussen *Aardrijkskunde* en de andere domeinen neemt dus verder toe overheen de twee laatste jaren van het basisonderwijs.



Figuur 9: inhoudelijke domeinen wetenschappen

Bij cognitieve domeinen is het opmerkelijk *Redeneren* een beperkte leerwinst kent van 38 punten, terwijl dit domein in 2019 net bovenaan stond. *Kennis* en *Toepassen* hebben daarentegen een gelijkaardig patroon met 54 en 55 punten leerwinst. De voorsprong in het hogere cognitieve domein *Redeneren* uit 2019 blijkt in 2021 weg te vallen (Figuur 10).



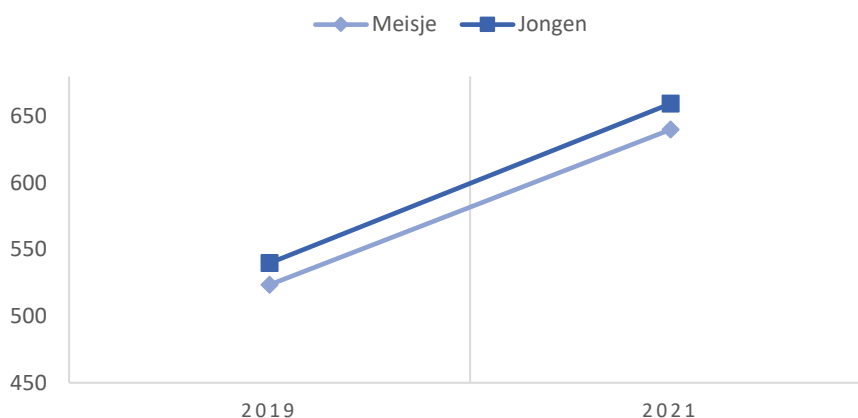
Figuur 10: cognitieve domeinen wetenschappen

Leerwinst en leerlingkenmerken

Na de leerwinst in detail te bekijken op vlak van de cognitieve en inhoudelijke domeinen, gaan we in wat volgt verder in op de relatie tussen bepaalde leerlingkenmerken en de leerwinst voor wiskunde en wetenschappen.

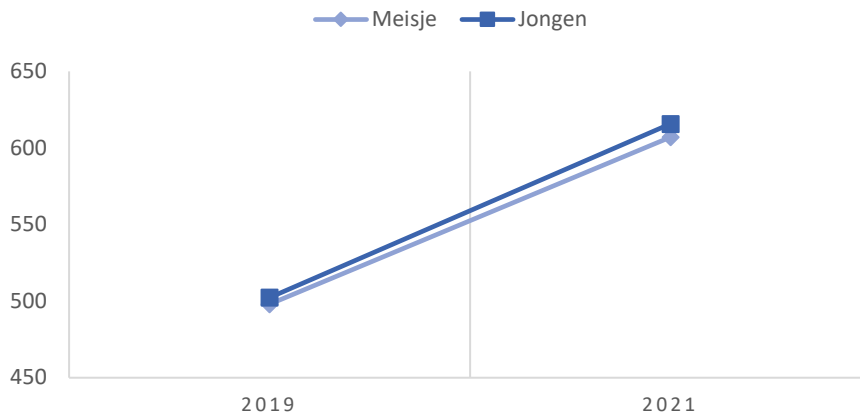
Geslacht

Op vlak van wiskunde behaalden jongens 4 punten meer leerwinst dan meisjes, met respectievelijk 120 en 116 punten leerwinst (Figuur 11). Jongens blijven dus iets beter presteren dan meisjes in 2021, maar het lijkt er niet op dat ze veel verder uitlopen. Langs de andere kant betekent dit ook dat meisjes geen inhaalbeweging maken ten opzichte van jongens voor wiskunde.



Figuur 11: leerwinst wiskunde naar geslacht

Ook voor wetenschappen zien we een gelijkaardige trend. Jongens behalen ook hier 4 punten meer leerwinst dan meisjes: 113 punten ten opzichte van 109. Hierbij stellen we ook vast dat het verschil in prestaties tussen jongens en meisjes in 2021 statistisch significant is. Jongens lijken voor wetenschappen dus beperkt uit te lopen op meisjes.

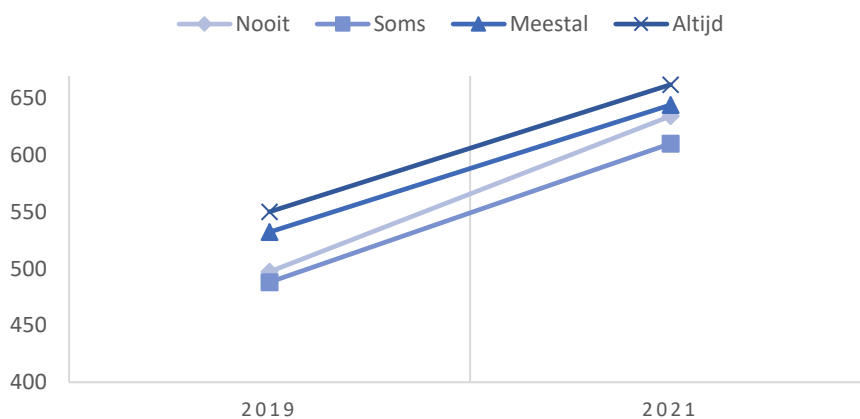


Figuur 12: leerwinst wetenschappen naar geslacht

Thuis taal

Op vlak van thuistaal zien we opvallende verschillen naargelang de leerwinst voor wiskunde (Figuur 13). De beste presteerders uit 2019 waren de leerlingen die thuis meestal of altijd Nederlands spreken. De leerwinst voor deze twee categorieën bedraagt 112 punten en ze zijn ook in 2021 de beste presteerders voor wiskunde.

Toch situeert de opvallendste bevinding zich bij de leerlingen die thuis nooit Nederlands spreken. Zij boeken met 137 punten de grootste leerwinst in de twee laatste jaren basisonderwijs. Door deze inhaalbeweging halen ze hun klasgenoten die thuis meestal Nederlands spreken zo goed als in. Waar het verschil tussen deze groepen in 2019 nog 35 punten was, bedraagt dit in 2021 nog slechts 9 punten. Leerlingen die thuis soms Nederlands spreken maken ook een inhaalbeweging ten opzichte van leerlingen die wel altijd Nederlands spreken thuis, met 122 punten leerwinst. Desondanks blijft deze groep wat achterop hangen en vergroot de kloof met de leerlingen nooit Nederlands spreken van 9 punten naar 25 punten.

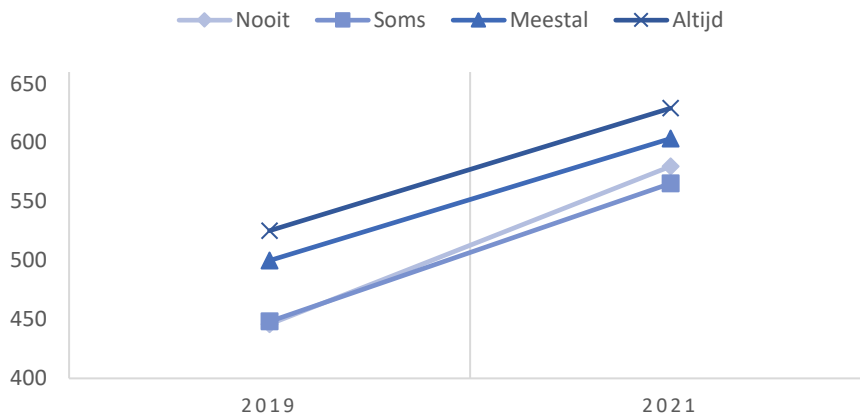


Figuur 13: leerwinst wiskunde naar thuistaal

Net zoals bij wiskunde vinden we de sterkste stijging voor wetenschappen bij de leerlingen die thuis nooit Nederlands spreken, met 134 punten (Figuur 14). Anders dan bij wiskunde is hier het verschil leerlingen die thuis soms Nederlands spreekt wel beperkter, deze groep behaalt 117 punten leerwinst voor wetenschappen.

De inhaalbeweging die we bij wiskunde zagen voor de leerlingen die dus nooit of soms Nederlands spreken thuis zien we ook bij wetenschappen terugkomen al is deze hier minder sterk uitgesproken

in vergelijking met wiskunde. Leerlingen die thuis meestal of altijd Nederlands spreken boeken ook voor wetenschappen evenveel leerwinst (104 punten) en blijven zo de sterkste presteerders.



Figuur 14: leerwinst wetenschappen naar thuistaal

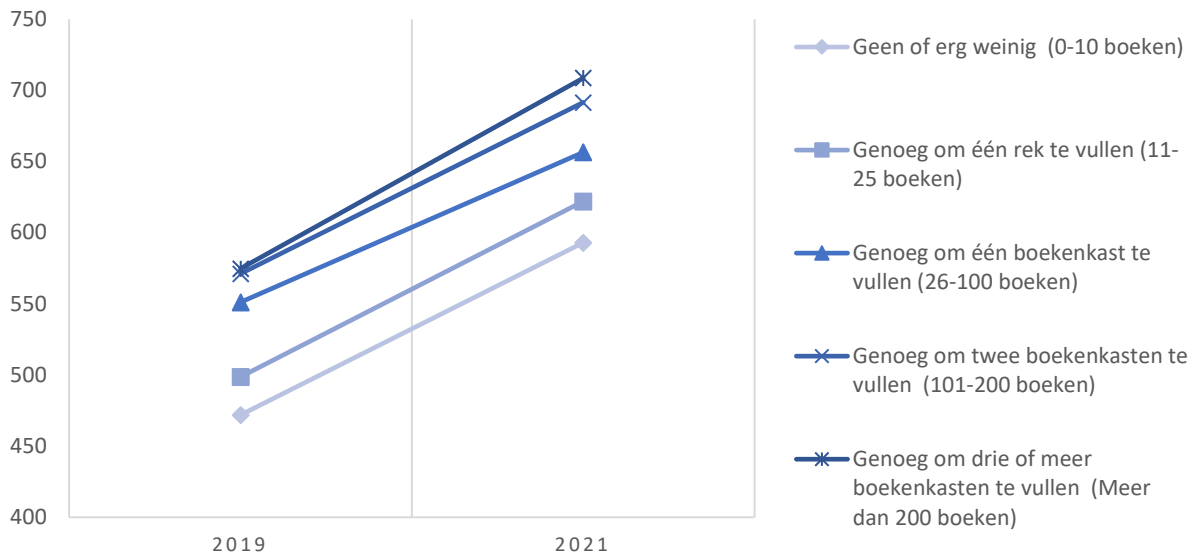
Aantal boeken thuis (sociaal economische status)

Zoals aangegeven in de voorgaande secties, wordt het aantal boeken dat een leerling thuis heeft gebruikt als proxy voor de sociaal economische status (SES) van de leerling. Er zijn 5 categorieën naargelang het aantal boeken dat een leerling thuis heeft:

- Geen of erg weinig boeken (0-10 boeken)
- Genoeg boeken om één rek te vullen (11-25 boeken)
- Genoeg boeken om één boekenkast te vullen (26-100 boeken)
- Genoeg boeken om twee boekenkasten te vullen (101-200 boeken)
- Genoeg boeken om drie of meer boekenkasten te vullen (Meer dan 200 boeken)

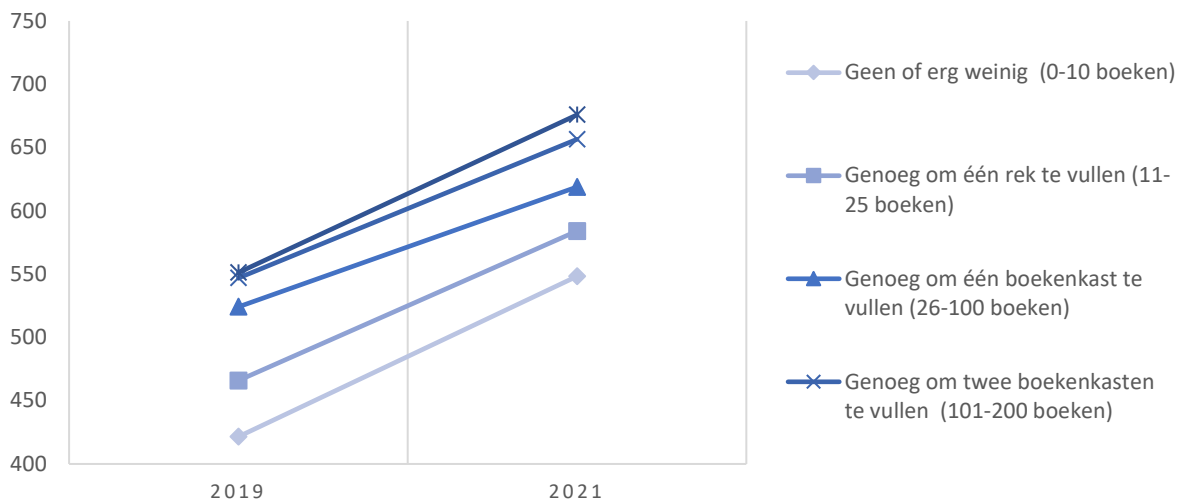
Uit de resultaten voor wiskunde blijkt dat naarmate een leerling thuis meer boeken heeft (wat gezien kan worden als meer *home resources*), er ook meer leerwinst geboekt wordt (Figuur 15). De sterkste presteerders uit 2019, leerlingen die thuis genoeg boeken hebben om drie of meer boekenkasten mee te vullen, sterk uitlopen op de andere leerlingen met 134 punten leerwinst. Leerlingen met thuis genoeg boeken voor twee boekenkasten hebben, behalen 120 punten leerwinst.

De resultaten voor de leerlingen in de categorieën met de minste boeken zijn opvallend. Zo behaalt de middelste categorie, de leerlingen met thuis voldoende boeken voor één boekenkast, de minste leerwinst met 105 punten. Leerlingen met geen of erg weinig boeken of genoeg boeken voor één rek kennen een beperkte inhaalbeweging, met respectievelijk 121 en 123 punten leerwinst. Qua leerwinst doet deze groep het dus beter leerlingen die thuis twee boekenkasten vol hebben (120 punten) en leerlingen die één boekenkast vol hebben (105 punten). Deze inhaalbeweging blijkt echter niet voldoende om de kloof helemaal dicht te rijden.



Figuur 15: leerwinst wiskunde naar aantal boeken thuis (SES)

Waar voor wiskunde de leerlingen met thuis de meeste boeken de sterkste stijgers zijn (de categorieën “genoeg boeken om twee boekenkasten te vullen” of “genoeg boeken om drie of meer boekenkasten te vullen”), is dat niet het geval voor wetenschappen (Figuur 16). Leerlingen met genoeg boeken voor drie of meer boekenkasten behalen 125 punten leerwinst voor wetenschappen. Leerlingen met twee volle boekenkasten, behalen 110 punten leerwinst. De sterkste stijgers zijn opvallend genoeg leerlingen met geen of weinig boeken. Ondanks dat deze groep 127 punten leerwinst behaalt en dus een inhaalbeweging kent, blijven ze achterop in vergelijking met de sterkste presteerders. Hetzelfde zien we terugkomen bij leerlingen met genoeg boeken voor één rek. Zij behalen een gemiddelde leerwinst van 118 punten en kunnen ondanks een beperkte inhaalbeweging, de kloof niet dichtrijden. Tenslotte valt op dat leerlingen met één boekenkast merkbaar minder leerwinst boeken in vergelijking met de andere groepen. Zij behalen slechts 95 punten leerwinst waardoor de kloof tussen de hen en de kopgroep 30 punten groter is dan in 2019.



Figuur 16: leerwinst wetenschappen naar aantal boeken thuis (SES)

VI. Inhaalbeweging of niet? Bedenkingen en conclusies

Leerwinst! Maar in welke mate?

Overheen de twee laatste jaren van het lager onderwijs gaan de Vlaamse leerlingen er duidelijk op vooruit. Met 117 punten voor wiskunde en 107 punten voor wetenschappen wordt er duidelijk leerwinst geboekt. De vraag is (en blijft) echter of dit nu een goed of een slecht resultaat is.

Een van de belangrijkste oorzaken waardoor het moeilijk is om in te schatten of de vastgestelde leerwinst goed of slecht is, is het gebrek aan ijkingspunten. In het huidig artikel en in het uitgebreide rapport van TIMSS-repeat (litref) werden verschillende ijkingspunten naar voren geschoven. Echter, de keuze voor deze ijkingspunten kwam eerder voort uit het beperkt aanbod dan wel uit de geschiktheid van de gehanteerde bronnen.

Een eerste ijkingspunt, is de theoretische berekening van Rindermann (2007). Rindermann legde de te verwachte leerwinst vast op 42 punten per jaar. Dit cijfer werd bekomen door het gemiddelde te nemen van verschillende (internationaal vergelijkende) studies voor wiskunde en wetenschappen. Hoewel hij zich dus baseerde op empirische studies, is de maatstaf zelf niet te definiëren als een empirisch gevalideerde waarde. Bovendien is de maatstaf sterk geaggregeerd en houdt deze geen rekening met temporele evoluties of met de specifieke Vlaamse context.

Een ander vergelijkingspunt dat in deze studie en in het rapport van TIMSS-repeat gehanteerd werd, waren de bevindingen van Martin et al. (1997) en Mullis et al. (1997) op basis van de derde TIMSS-studie. In deze cyclus werd er uitzonderlijk ook longitudinale data verzameld: leerlingen werden gevolgd van het 3^{de} naar het 4^{de} leerjaar. Hoewel dit op zich interessante informatie biedt, zijn ook hier enkele aandachtspunten bij op te merken. Allereerst was er geen Vlaamse deelname aan deze longitudinale TIMSS-studie. Een één-op-één vergelijking met andere Vlaamse leerlingen uit het verleden is dus niet mogelijk. Daarom werd geopteerd om telkens de Nederlandse leerwinst te rapporteren. Ten tweede werden de leerlingen in deze studie gevolgd in de lagere leerjaren van het basisonderwijs. Dit maakt de vergelijking met TIMSS-repeat moeilijker aangezien wetenschappelijke literatuur aangeeft dat leerwinst veelal hoger ligt in de lagere onderwijsjaren in vergelijking met de hogere onderwijsjaren (Bloom et al., 2008). Ten derde kan men zich ook afvragen in welke mate een vergelijking met een studie van meer dan 25 jaar geleden nog actueel te noemen is, zeker wanneer ook de socio-culturele context verschillend is.

Het derde vergelijkingspunt betrof de studie van Bloom et al. (2008). Om een vergelijking met deze studie te kunnen maken, werden de resultaten omgezet in een effectgrootte. Dit maakt dat vergelijkingen overheen verschillende schalen mogelijk worden. In de studie van Bloom werd duidelijk gerapporteerd over de effectgrootte van één jaar onderwijs voor wiskunde en wetenschappen per leerjaar van het basisonderwijs, wat dus op het eerste zicht een bijzonder goede vergelijkingsbasis biedt. Echter, de data waarop de gerapporteerde effectgroottes gebaseerd zijn, werd reeds in de jaren '90 verzameld. De kritische lezer zal zich dus ook kunnen afvragen in welke mate leerlingen uit de Verenigde Staten in de jaren '90 een goed ijkingspunt zijn voor de Vlaamse leerlingen in 2021.

Tenslotte werd er ook gekeken naar de leerwinst zoals die vastgesteld werd binnen het PIRLS-repeat onderzoek (Dockx et al., 2019). Hoewel PIRLS ook uitgedrukt wordt op de TIMSS-schaal, dezelfde doelgroep bevraagd werd en hetzelfde tijdspad voor de longitudinale aspecten van de studie gehanteerd werden (van het 4^{de} naar het 6^{de} leerjaar basisonderwijs) kan men zich wel de vraag stellen in welke mate het relevant is om wiskunde en wetenschappen af te zetten tegenover begrijpend lezen.

Vergelijkingen binnen de Vlaamse leerlingpopulatie

De kritische bedenkingen rond de ijkingspunten terzijde, kunnen we echter wel stellen dat de Vlaamse leerlingen leerwinst boeken. Hierbij kan dan gekeken worden naar de deeldomeinen voor wiskunde en wetenschappen of specifieke leerlingkenmerken.

Een van de hypothesen waarmee het rapport van TIMSS 2019 mee afsloot, ging ervanuit dat de lagere prestaties voor de hogere cognitieve domeinen (zoals *Toepassen* en *Redeneren*) toe te schrijven zijn aan de focus die het Vlaams curriculum legt in de verschillende leerjaren van het basisonderwijs (Faddar et al., 2020). Zo werd er gesteld dat het Vlaams onderwijs in de lagere leerjaren sterk focust op kennis en dat er pas in de latere jaren een verschuiving is naar het toepassen van die kennis en redeneren. De hogere leerwinst voor wiskunde binnen het domein *Toepassen*, lijkt deze hypothese te onderbouwen. Echter, bij wetenschappen is de leerwinst minder uitgesproken voor de hogere cognitieve deeldomeinen. Wel stellen we vast dat de leerwinst hoger is voor de domeinen van een hoger cognitief beheersingsniveau, voornamelijk voor wiskunde. Bij de cognitieve domeinen zien we wel dat de Vlaamse leerlingen een sterke sprong maken op het vlak van *Toepassen*. De hypothese dat het Vlaams onderwijs de eerste jaren voornamelijk inzet op kennis en pas in de latere jaren op toepassen en redeneren, lijkt hier dus bevestigd te worden (Faddar et al., 2020). Voor wetenschappen komt de stijging voor cognitief hogere domeinen zoals *Toepassen* en *Redeneren* komt minder uitgesproken terug. Daar stellen we zelfs vast dat de sterkere prestatie in het domein *Redeneren* uit 2019 in 2021 niet meer merkbaar is.

De dataset van TIMSS-repeat bleek uitermate geschikt om verschillen in leerwinst naargelang leerlingkenmerken te beschrijven. Dit biedt een bijzonder interessant perspectief op hoe verschillende groepen leerlingen evolueren overheen de laatste twee jaren van het basisonderwijs. Een van de opvallendste vaststellingen hierbij is dat leerlingen die thuis nooit Nederlands spreken voor zowel wiskunde als wetenschappen net de grootste leerwinst boeken. Hoewel de kloof tussen de sterkste presteerders (leerlingen die meestal of altijd Nederlands spreken) niet helemaal gedicht wordt, stellen we wel vast dat leerlingen die nooit Nederlands spreken een duidelijke inhaalbeweging maken. Ook bij het aantal boeken dat een leerling in huis heeft, als proxy voor SES, zien we opvallende verbanden met de geboekte leerwinst. Hier stellen we vast dat leerlingen met veel boeken (100 of meer) hun voorsprong op hun klasgenoten vergroot zien. Dit lijkt erop te wijzen dat het Vlaamse onderwijs er niet in slaagt om de kloof tussen leerlingen met veel of weinig middelen thuis dicht te rijden. Voor geslacht kwamen er geen opvallende verschillen in leerwinst terug. Jongens boeken dan wel iets meer leerwinst dan meisjes, maar dit verschil is niet significant voor wiskunde. Bij wetenschappen was dit verschil wel significant al bleef het beperkt tot 4 punten.

Beleidsrelevantie en nood aan verder onderzoek

Het TIMSS-repeat onderzoek bracht de leerwinst in de laatste jaren van het Vlaamse basisonderwijs in kaart om zo na te gaan in welke mate er sprake is van een inhaalbeweging voor wiskunde en wetenschappen. Dat er leerwinst geboekt werd, is duidelijk. Toch ontbreekt het ons aan voldoende vergelijkingspunten om een claim rond een eventuele inhaalbeweging van de Vlaamse leerlingpopulatie hard te maken wegens het gebrek aan valide ijkingspunten. Toch bieden de resultaten van TIMSS-repeat, en met name de resultaten over de evoluties binnen de Vlaamse leerlingpopulatie heel wat stof tot nadenken. Gezien de focus van menig onderwijsbeleidsmaker op STEM-onderwijs is het bijzonder relevant om na te gaan op welke manier de vastgestelde verschillen binnen de Vlaamse leerlingpopulatie te verklaren zijn.

Verder onderzoek, Vlaams en internationaal, kan bijdragen aan een verder begrip van de leerwinst van Vlaamse leerlingen in wiskunde en wetenschappen. Aangezien Vlaanderen niet deelneemt aan

de vervolgstudie van TIMSS in 2023, TIMSS Longitudinal, betekent dat de resultaten van TIMSS-repeat niet direct vergeleken kunnen worden met eventuele bevindingen uit TIMSS Longitudinal. De beperkingen van het huidige onderzoek illustreren immers ook mogelijke valkuilen voor initiatieven zoals de Vlaamse centrale toetsen. Hoewel het onmiskenbaar is dat een gestandaardiseerde toetsing binnen het Vlaamse onderwijs bijzonder relevant is voor het onderwijsbeleid op alle niveaus, tonen de beperkingen TIMSS-repeat het belang van internationaal vergelijkend onderzoek aan om de kwaliteit van het Vlaamse onderwijs kritisch te kunnen beschouwen.

VII. Bibliografie

- Bailey, P., Lee, M., Nguyen, T., & Zhang, T. (2020). *Using EdSurvey to Analyze TIMSS Data*. In Bloom, H. S., Hill, C. J., Black, A. R., and Lipsey, M. W. (2008). Performance Trajectories and Performance Gaps as Achievement Effect-Size Benchmarks for Educational Interventions. *Journal of Research on Educational Effectiveness*, 1(4), 289-328. <https://doi.org/10.1080/19345740802400072>
- Dockx, J., Van Landeghem, G., Aesaert, K., Van Damme, J., and De Fraine, B. (2019). *Begrijpend lezen van het vierde naar het zesde leerjaar: Herhalingsmeting van PIRLS in 2018 vergeleken met PIRLS 2016*.
- Faddar, J., Appels, L., Merckx, B., Boeve-de Pauw, J., Delrue, K., De Maeyer, S., and Van Petegem, P. (2020). *Vlaanderen in TIMSS 2019. Wiskunde- en wetenschapsprestaties van het vierde leerjaar in internationaal perspectief en doorheen de tijd*.
- Fishbein, B., Foy, P., & Yin, L. (2021). *TIMSS 2019 User Guide for the International Database*.
- Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzalez, E. J., Smith, T. A., and Kelly, D. L. (1997). *Science achievement in the primary school years: IEA's Third International Mathematics and Science Study (TIMSS)*.
- Martin, M. O., von Davier, M., and Mullis, I. V. S. (2020). *Methods and Procedures: TIMSS 2019 Technical Report*. T. P. I. S. C. Boston College. <https://timssandpirls.bc.edu/timss2019/methods>
- Mullis, I., Martin, M., Foy, P., Kelly, D., and Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. <https://timssandpirls.bc.edu/timss2019/international-results/>
- Mullis, I. V. S., and Martin, M. O. (2017). *TIMSS 2019 Assessment Frameworks*.
- Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., and Smith, T. A. (1997). *Mathematics achievement in the primary school years: IEA's Third international Mathematics and Science Study (TIMSS)*.
- Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: the homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations [\[https://doi.org/10.1002/per.634\]](https://doi.org/10.1002/per.634). *European Journal of Personality*, 21(5), 667-706. <https://doi.org/https://doi.org/10.1002/per.634>
- Scharfen, J., Peters, J. M., and Holling, H. (2018). Retest effects in cognitive ability tests: A meta-analysis. *Intelligence*, 67, 44-66. <https://doi.org/https://doi.org/10.1016/j.intell.2018.01.003>
- STEP. (2015). Peiling wereldoriëntatie: natuur en techniek.
- STEP. (2021). Peiling Wiskunde in het Basisonderwijs.
- Verhelst, D., Meulemans, C., Gijssen, M., Appels, L., De Maeyer, S., and Van Petegem, P. (2022). *TIMSS-REPEAT: herhalingsmeting op TIMSS 2019 in 2021*. <https://www.uantwerpen.be/nl/projecten/timss-vlaanderen/studies/2019-repeat/>