



UNIVERSITY
OF TWENTE.

Toetsgedrag van Nederlandse leerlingen in PISA-2018

- Een vergelijking met omliggende landen

Onderdeel van *PISA-2018: De Verdieping*

Remco Feskens
Kimberley Lek
Nathalie Maassen
Martina Meelissen



Alle rechten voorbehouden. Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van Stichting Cito Instituut voor Toetsontwikkeling worden openbaar gemaakt en/of verveelvoudigd door middel van druk, fotokopie, scanning, computersoftware of andere elektronische verveelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook.

Samenvatting

In deze verdiepende studie wordt nader ingegaan op het toetsgedrag van Nederlandse leerlingen in vergelijking met leerlingen uit enkele andere landen en regio's in PISA-2018. Er zijn namelijk indirecte aanwijzingen dat Nederlandse leerlingen minder geneigd zijn om een maximale prestatie te leveren tijdens de afname van zogenaamde *low-stakes* toetsen, in vergelijking met andere landen¹. De landen en regio's waarmee Nederland vergeleken wordt, zijn Denemarken, Duitsland, Groot-Brittannië, Noorwegen, Vlaanderen en Zweden. We gebruiken vier indicatoren voor het in kaart brengen van toetsgedrag: motivatie, responstijden, toetsconcentratie en het aantal overslagen vragen tijdens de toets.

Motivatie

- In vergelijking met leerlingen uit de vergelijkingslanden geven Nederlandse leerlingen zichzelf gemiddeld gezien een redelijk hoog cijfer voor motivatie op de PISA opgaven. Alleen in Groot-Brittannië scoren de leerlingen zichzelf iets hoger op dit vlak. De verschillen tussen landen zijn echter wel gering.
- Net als in andere landen waren leerlingen gemiddeld gezien wel gemotiveerder geweest als de PISA-opgaven tot een beoordeling hadden geleid.

Responstijden

- Gemiddeld gezien besteden Nederlandse leerlingen een half uur aan het maken van de twee wiskunde modules, ongeacht cluster positie. Dit gemiddelde ligt lager dan de gemiddelde responstijden in alle vergelijkingslanden. Deze lage responstijden binnen Nederland worden ook gevonden voor natuurwetenschappen en leesvaardigheid.
- Opvallend hoge responstijden komen in Nederland amper voor, terwijl deze in bijvoorbeeld Zweden veel vaker voorkomen. Ook de variantie in responstijden is relatief beperkt in Nederland.

¹Molen, P. van der, Schouwstra, S., Feskens, R., & Onna, M. van (2019). *Vaardigheidsontwikkelingen volgens PISA en examens*, Arnhem: Cito. https://www.nro.nl/wp-content/uploads/2019/04/Cito_Vaardigheidsontwikkelingen_volgens_PISA_en_examens_maart_2019.pdf.

Toetsconcentratie

- In Nederland daalt net zoals in de meeste andere landen de prestatie op de toets na de pauze. Dit geldt voor zowel wiskunde als natuurwetenschappen.

Overgeslagen vragen

- In Nederland is het percentage niet-gemaakte opgaven gemiddeld gezien laag ten opzichte van de vergelijkingslanden. In Zweden ligt het aantal overgeslagen items gemiddeld gezien bijvoorbeeld ruim twee tot drie keer zo hoog.

Samengevat

- Samengevat kan gezegd worden dat het toetsgedrag van Nederlandse leerlingen wat een aantal aspecten betreft er uit springt ten opzichte van het toetsgedrag van hun leeftijdgenoten in de vergelijkingslanden. Ze besteden relatief weinig tijd aan de opgaven en geven slechts in een gering aantal gevallen geen antwoord op de vragen in de toets.
- Dit gedeeltelijk afwijkende toetsgedrag leidt overigens niet noodzakelijkerwijs tot lage toetsresultaten. Zo is de gemiddelde prestatie voor wiskunde in Nederland de hoogste van de in deze studie opgenomen landen.

Inhoudsopgave

1	Introductie	1
1.1	Inleiding	1
1.2	Toetsgedrag	1
1.3	Leeswijzer	3
2	Data	4
2.1	Afnamedesign	4
2.2	Item type	6
2.3	Aantal leerlingen	7
3	Motivatie	9
3.1	Motivatie-thermometer	9
3.2	Samenhang tussen motivatie bij verschillende toetsen . .	11
4	Responstijden	14
4.1	Gemiddelde responstijd	14
4.2	Responstijd en item type	16
4.3	Responstijd en opgaven	17
4.4	Responstijd en cluster positie	19
4.5	Responstijd en type leerling	21
5	Toetsconcentratie	26
5.1	Wiskunde	26
5.2	Natuurwetenschappen	28
6	Niet-gemaakte opgaven	29
6.1	Percentage niet-gemaakte opgaven	29
6.2	Niet-gemaakte opgaven en item type	31
6.3	Niet-gemaakte opgaven en cluster positie	31
6.4	Niet-gemaakte opgaven en type leerling	32
7	Conclusies en discussie	34

1 Introductie

1.1 Inleiding

Toetsen worden afgenomen om uitspraken te kunnen doen over het vaardigheidsniveau van leerlingen of groepen leerlingen. De prestaties op de toetsen zouden uiteraard bepaald moeten worden door de vaardigheid van de leerlingen, maar deze kunnen ook beïnvloed worden door niet-cognitieve factoren. Een van de andere factoren die naast de vaardigheid van leerlingen toetsresultaten kan beïnvloeden, is het toetsgedrag van leerlingen tijdens de afname van de toets.

In deze verdiepende studie wordt nader ingegaan op het toetsgedrag van Nederlandse leerlingen in vergelijking met leerlingen uit enkele andere landen en regio's in PISA-2018². De landen en regio's waarmee Nederland vergeleken worden, zijn Denemarken, Duitsland, Groot-Brittannië, Noorwegen, Vlaanderen en Zweden. We gebruiken vier indicatoren voor het in kaart brengen van toetsgedrag: motivatie, responstijden, toetsconcentratie en het aantal overslagen vragen tijdens de toets.

1.2 Toetsgedrag

Internationale surveys zoals PISA zijn typische *low-stakes* toetsen. In een low-stakes toetsconditie zijn er geen consequenties verbonden aan de toetsprestatie voor leerlingen zelf. Omgekeerd geldt dat hoe meer *high-stakes* de toetscondities zijn, hoe groter de persoonlijke gevolgen voor leerlingen zijn³. Leerlingen bereiden zich niet voor op de PISA toets, krijgen geen feedback over hun prestatie en er zijn geen persoonlijke gevolgen verbonden aan de toetsprestaties. Meerdere studies hebben laten zien dat leerlingen niet maximaal presteren op het moment dat er geen feedback gegeven wordt of als de prestatie op een vraag niet mee telt bij de bepaling van een cijfer (zie onder an-

²OECD (2019), *PISA 2018 Results (Volume I): What Students Know and Can Do*, PISA, OECD Publishing, Paris, <https://doi.org/10.1787/5f07c754-en>.

³Holland, P.W., & Wightman, L.E. (1982). Section pre-equating: A preliminary investigation. In P.W. Holland & D.R. Rubin (Eds.), *Test Equating* (pp. 271-297). New York: Academic Press.

dere Wise & DeMars, 2005⁴; O'Neill, Sugrue, & Baker, 1996⁵; Kiplinger & Linn, 1996⁶, cf. Keizer-Mittelhaeuser, 2014). Omgekeerd is de verwachting dat leerlingen onder high-stakes toetscondities harder werken en streven naar een maximale prestatie, terwijl low-stakes toetscondities juist een typische prestatie bij leerlingen oproept⁷. Een gevolg daarvan kan zijn dat er verschillen ontstaan in de moeite die leerlingen nemen om de toets voor te bereiden, vragen binnen de toets te beantwoorden en daarmee ook de prestatie die zij uiteindelijk leveren⁸. Wise & DeMars (2005) lieten in een meta-analyse zien dat het verschil in prestaties tussen gemotiveerde en ongemotiveerde leerlingen een gemiddelde effect grootte van 0.59 heeft. Op de PISA-schaal komt dit overeen met ruim 50 scorepunten.

Wise & Kong (2005) suggereren dat een lage motivatie om een toets te maken niet perse constant hoeft te zijn. Het zou kunnen zijn dat leerlingen binnen een toets meer of minder aandacht aan opgaven besteden naarmate de toets vordert. Dit wordt verder ondersteund door een studie van Wise (2007)⁹ waarin wordt beargumenteerd dat leerlingen weliswaar gemotiveerd kunnen beginnen aan een toets, maar op een gegeven moment in de toets steeds minder gemotiveerd kunnen worden.

Of leerlingen een maximale of een typische prestatie leveren op een toets is niet direct zichtbaar. Daarom maken we in deze studie gebruik van verschillende *proxy* variabelen om het toetsgedrag van leerlingen in kaart te brengen. Dat zijn respectievelijk de resultaten van een zelfrapportage van de leerlingen over motivatie (Hoofdstuk 3), de hoeveelheid tijd die leerlingen aan de opgaven hebben besteed (Hoofdstuk 4), toetsconcentratie (Hoofdstuk 5) en de hoeveelheid overgeslagen en niet-bereikte opgaven (Hoofdstuk 6).

⁴Wise, S.L. & DeMars, C.E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10, 1-17.

⁵O'Neill, H.E., Sugrue, B., & Baker, E.L. (1996). Effects of motivational interventions on the National Assessment of Educational Progress mathematics performance. *Educational Assessment*, 3, 135-157.

⁶Kiplinger, V.L., & Linn, R. L. (1996). Raising the stakes of test administration: The impact on student performance on the National Assessment of Educational Progress. *Educational Assessment*, 3, 111-133.

⁷Keizer-Mittelhaeuser, M. (2014). Modeling the Effect of Differential Motivation on Linking Educational Tests. PhD Thesis: Tilburg University.

⁸Wise, S.L. & DeMars, C.E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment*, 10, 1-17.

⁹Wise, S.L. (2007). Examinee effort and test score validity. Paper presented at the annual meeting of the Northeastern Educational Research Association, Rocky Hill, Connecticut.

1.3 Leeswijzer

In PISA worden drie hoofddomeinen in kaart gebracht: Wiskunde, natuurwetenschappen en leesvaardigheid. In PISA-2018 is leesvaardigheid op een adaptieve manier getoetst. Dat betekent dat leerlingen afhankelijk van hun prestatie op de toets later in de toets makkelijkere of juist moeilijker opgaven krijgen voorgelegd. Dat maakt dat het design en daarmee de mogelijke analyses voor leesvaardigheid er anders zien dan voor wiskunde en natuurwetenschappen. Deze twee domeinen zijn namelijk lineair getoetst. Dat wil zeggen dat de opgaven die leerlingen afnemen vastgelegd zijn in een toetsversie. Niet alle analyses die voor wiskunde en natuurwetenschappen zijn uitgevoerd, konden daarom ook voor leesvaardigheid uitgevoerd worden. Ook presenteren we de resultaten niet voor alle drie domeinen op het moment dat deze niet zeer verschillen tussen de domeinen. Deze resultaten zijn wel beschikbaar in de apart gepubliceerde appendix.

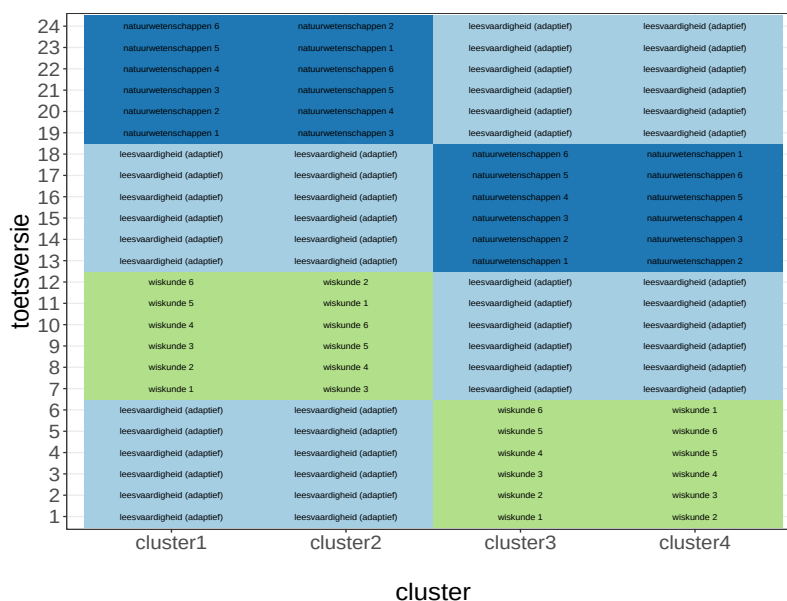
2 Data

In dit hoofdstuk wordt eerst stil gestaan bij afnamedesign van PISA-2018 (Sectie 2.1). Het gebruikte afnamedesign maakt het mogelijk om goed in kaart te brengen hoe het toetsgedrag en prestaties zich in de loop van de toets hebben ontwikkeld. Het gebruikte opgaven-materiaal wordt in Sectie 2.2 besproken. In Sectie 2.3 ten slotte worden het aantal deelnemende leerlingen per land gepresenteerd.

2.1 Afnamedesign

Binnen PISA maken leerlingen verschillende toetsversies die samengesteld zijn uit een bredere set van opgaven¹⁰. De opgaven zijn ingedeeld in clusters en leerlingen maken twee clusters voor en twee clusters met opgaven na de pauze. Zowel voor als na de pauze hebben leerlingen 60 minuten de tijd om de opgaven te maken. In PISA-2018 was leesvaardigheid het hoofddomein. Dat betekent ook dat alle leerlingen twee clusters met leesvaardigheid opgaven hebben gemaakt. Daarnaast hebben leerlingen clusters met òf wiskunde òf natuurwetenschappen opgaven gemaakt. Een overzicht van het afnamedesign van de eerste 24 (primaire) toetsversies binnen PISA-2018 is te vinden in Figuur 1.

¹⁰<https://www.oecd.org/pisa/data/pisa2018technicalreport/>



Figuur 1: Afnamedesign PISA-2018

De analyses voor wiskunde zijn gebaseerd op de toetsversies (of boekjes) 1-12 van PISA-2018. Deze boekjes bestaan voor de helft uit wiskunde opgaven, verdeeld over twee clusters; de andere helft opgaven bestaat uit leesopgaven (zie ook Figuur 1). Ongeveer de helft van de leerlingen start met de twee wiskunde clusters en maken daarna het leesgedeelte. Voor de andere leerlingen is de volgorde omgekeerd; zij eindigen met de twee wiskunde clusters. Daarom is er een onderscheid tussen cluster 1, 2, 3 en 4, waarvan clusters 1-2 en 3-4 bij elkaar horen. Er zijn zes verschillende modules met wiskunde opgaven die op iedere cluster positie worden afgenomen. Iedere module bestaat uit 11 of 12 wiskunde opgaven. Welke module leerlingen maken op welke cluster positie varieert, zodat elke module ongeveer even vaak wordt afgenomen op elke cluster positie. Merk op dat cluster 1 en 2 voor de pauze plaatsvinden, terwijl cluster 3 en 4 na deze pauze worden gemaakt.

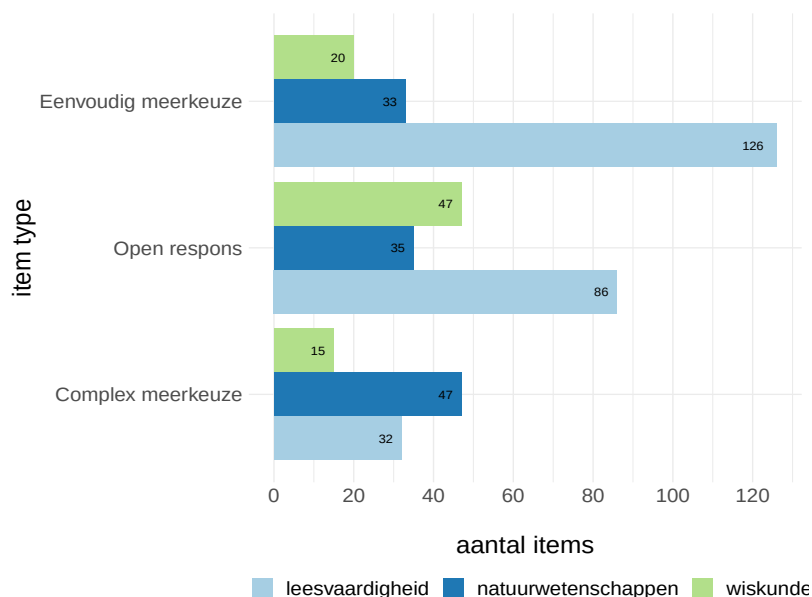
De analyses voor natuurwetenschappen zijn gebaseerd op de boekjes 13-24 van PISA-2018. Deze boekjes bestaan voor de helft uit natuurwetenschappen opgaven, verdeeld over twee clusters; de andere helft uit leesopgaven. Het afnamedesign is vergelijkbaar als bij wiskunde. Ongeveer de helft van de leerlingen start met de twee natuur-

wetenschappen clusters en maken daarna het leesgedeelte. Er zijn zes verschillende modules met natuurwetenschappen opgaven die op iedere cluster positie worden afgenomen. Iedere module bestaat uit 18 tot 20 natuurwetenschappen opgaven. Net zoals bij wiskunde variëren de modules en cluster posities zodat elke module ongeveer even vaak wordt afgenomen op elke cluster positie, waarbij wederom cluster 1 en 2 voor de pauze worden afgenomen, terwijl cluster 3 en 4 na deze pauze worden gemaakt.

In alle PISA-2018-boekjes zit leesvaardigheid verwerkt. Omdat we voor wiskunde en natuurwetenschappen enkel de primaire boekjes – met twee modules wiskunde of twee modules natuurwetenschappen – hebben geselecteerd, hebben we voor leesvaardigheid ook een selectie gemaakt. Dit zijn de 12 eerder geselecteerde boekjes wiskunde en de 12 eerder geselecteerde boekjes natuurwetenschappen, waarvan het andere deel geheel uit leesvaardigheid bestaat.

2.2 Item type

Binnen alle drie de hoofddomeinen worden verschillende item types gebruikt om de vaardigheden te meten. Dit zijn respectievelijk zogenaamde *eenvoudige meerkeuze* -, *complexe meerkeuze* - en *open respons*-opgaven. Het onderscheid naar item type is relevant in de context van toetsgedrag, omdat bijvoorbeeld eenvoudige meerkeuze-opgaven minder inspanning vragen om te beantwoorden dan een open respons-opgave. Leerlingen kunnen in het eerste geval namelijk makkelijker een gokje wagen. De aantallen opgaven per item type die zijn opgenomen in de verschillende domeinen zijn te vinden in Figuur 2.

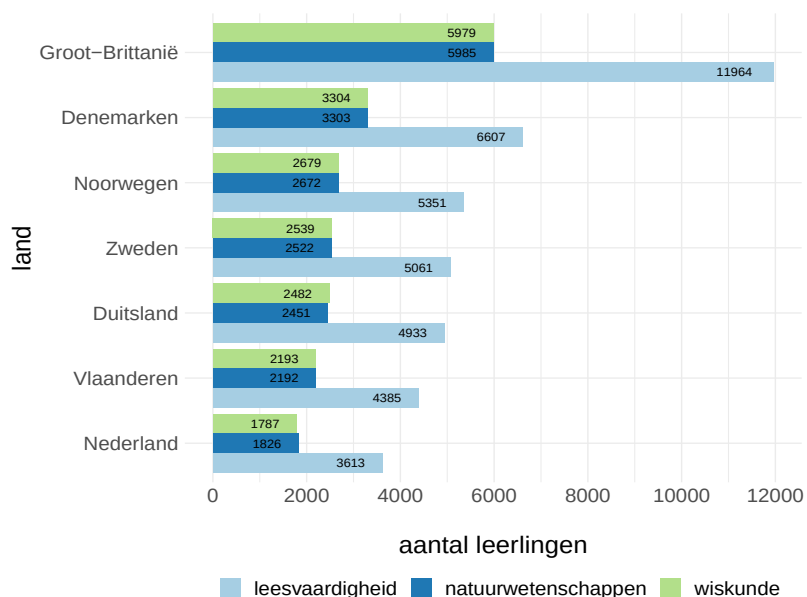


Figuur 2: Aantal type items per domein

Leesvaardigheid bestaat uit veruit de meeste opgaven. Dit komt omdat dit het hoofddomein is en daarom ook het meest in detail in kaart gebracht wordt. Leesvaardigheid is ook gedeeltelijk adaptief afgenomen. Dit vereist automatisch scorebare items en daarom zijn er juist weer relatief minder open respons-opgaven afgenomen voor leesvaardigheid.

2.3 Aantal leerlingen

In deze studie wordt het toetsgedrag van leerlingen uit zes verschillende landen vergeleken. Dit zijn Denemarken, Duitsland, Groot-Brittannië, Nederland, Noorwegen en Zweden. Daarnaast maakt de regio Vlaanderen ook deel uit van deze studie. Leerlingen die één van de 24 eerste toetsversies hebben afgenomen, zijn opgenomen in deze studie. Een overzicht van de aantallen leerlingen per land en domein is te vinden in Figuur 3.



Figuur 3: Aantal deelnemende leerlingen per land en domein

Leesvaardigheid is als hoofddomein in alle toetsversies opgenomen. Vandaar dat het aantal leerlingen dat leesvaardigheid heeft afgenomen in dit geval gelijk is aan de som van het aantal leerlingen dat wiskunde of natuurwetenschappen heeft afgenomen. Opvallend is dat het aantal deelnemende leerlingen binnen Nederland relatief gering is in vergelijking met de andere landen.

Het digitale platform dat gebruikt is bij de toetsafname slaat ook de hoeveelheid tijd op die leerlingen besteden aan elke vraag in de toets¹¹. Deze responstijden zijn in milliseconden geregistreerd. Leerlingen die meer dan 30 minuten aan één opgave hebben besteed, zijn uitgesloten van deze analyses. Zo'n ruime responstijd duidt waarschijnlijk op een registratie-fout in het platform. Dit betreft 0.2% van het totaal aantal leerlingen in deze studie. We presenteren in deze studie altijd de ongewogen resultaten.

¹¹ OECD (2017). *PISA 2015 technical report. Technical report*, OECD, Paris: OECD Publications. <https://www.oecd.org/pisa/sitedocument/PISA-2015-technical-report-final.pdf>.

3 Motivatie

De eerste indicator die beschikbaar is om toetsgedrag in kaart te brengen is een zelf-rapportage voor motivatie. De resultaten van deze meting in de verschillende landen worden besproken in Sectie 3.1. In Sectie 3.2 wordt nader ingegaan op de samenhang tussen de motivatie op een low-stakes toets zoals PISA en de motivatie op een (hypothetische) toets waar wel een individuele beoordeling aan gekoppeld is.

3.1 Motivatie-thermometer

Een manier om te achterhalen hoeveel moeite leerlingen in het maken van een toets hebben gestopt, is om na de toetsafname hen zelf hiernaar te vragen¹². PISA-2018 heeft ook een aantal vragen in de achtergrondvragenlijst opgenomen die motivatie meten. Daarnaast is ook gevraagd naar de hoeveelheid moeite die de leerlingen in een toetsafname zouden hebben gestopt op het moment dat deze wel mee zou tellen in een beoordeling. Een illustratie van het meetinstrument zoals dat in PISA-2018 is opgenomen, is te vinden in Figuur 4.

In Figuur 5 zijn de uitkomsten te vinden van de Nederlandse leerlingen en die uit de vergelijkingslanden op deze motivatie-thermometer.

¹²De validiteit van zelf-rapportages wordt ook wel ter discussie gesteld. Enerzijds kunnen zelf-rapportages van potentieel sensitief gedrag tot over-rapportage leiden op het moment dat respondenten zichzelf zo in een gunstig perspectief kunnen stellen (zie onder andere Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., & Tourangeau, R. (2004): *Survey Methodology*, Hoboken, NJ.). Anderzijds, kunnen leerlingen voor zichzelf ook een lagere motivatie invullen om zo mogelijk te verhullen dat een slechte prestatie op de toets niet zozeer het gevolg was van een gebrek aan vaardigheid, maar juist van een gebrek aan motivatie (Baumert, J. & A. Demmrich (2001), Test motivation in the assessment of student skills: The effects of incentives on motivation and performance, *European Journal of Psychology of Education*, Vol. 16(3), pp. 441-462. Zie ook OECD (2019), *PISA 2018 Results (Volume I): What Students Know and Can Do*, PISA, OECD Publishing, Paris.)

¹³OECD (2019), *PISA 2018 Results (Volume I): What Students Know and Can Do*, PISA, OECD Publishing, Paris, <https://doi.org/10.1787/5f07c754-en>.

PISA 2018

Effort Thermometer

How much effort did you invest?

Please try to imagine an actual situation (at school or in some other context) that is highly important to you personally, so that you would try your very best and put in as much effort as you could to do well.

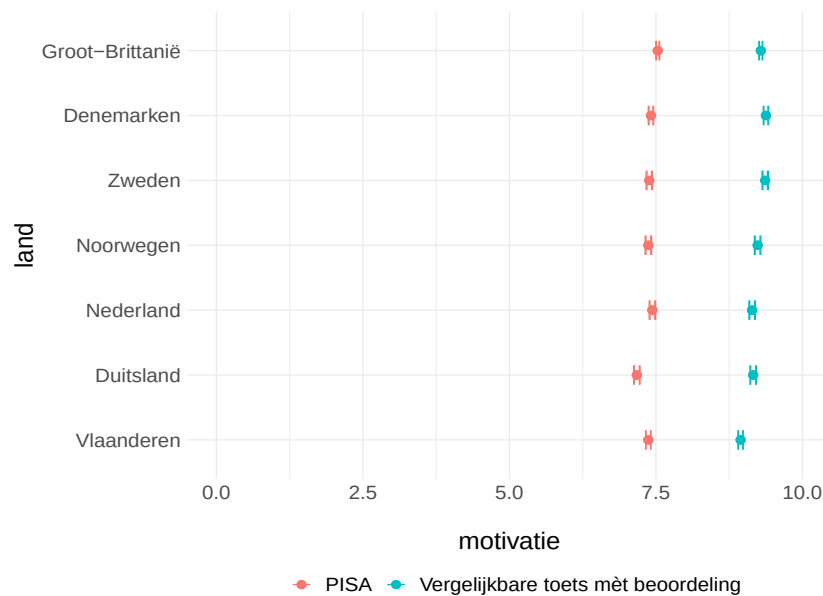
In this situation you would mark the highest value on the "effort thermometer", as shown below:

Compared to the situation you have just imagined, how much effort did you put into doing this test?

How much effort would you have invested if your marks from the test were going to be counted in your school marks?

Click on the NEXT arrow to continue.

Figuur 4: De *effort* thermometer in PISA-2018¹³

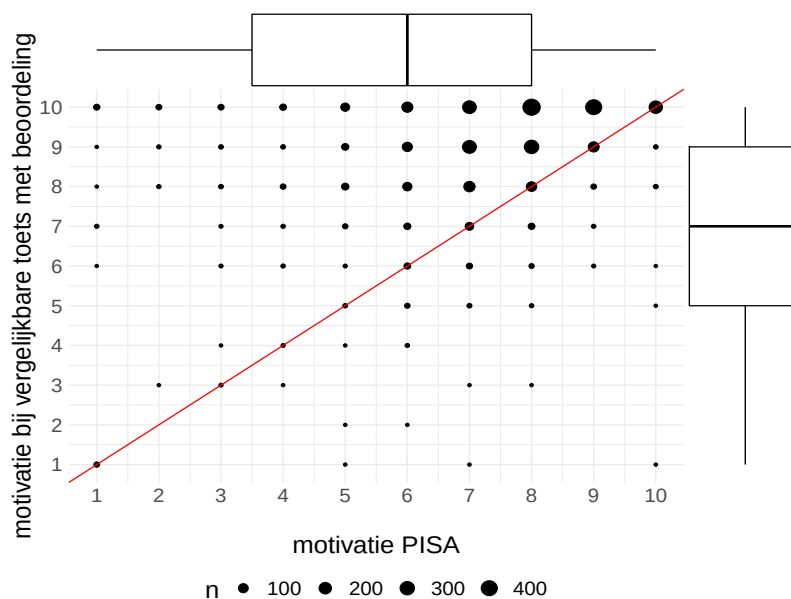


Figuur 5: Gemiddelde motivatie-scores op PISA en toetsen met beoordeling

In vergelijking met leerlingen uit de vergelijkingslanden geven Nederlandse leerlingen zichzelf gemiddeld gezien een redelijk hoog cijfer voor motivatie op de PISA-opgaven, zie Figuur 5. De verschillen tussen landen zijn echter klein. In Duitsland scoren de leerlingen zichzelf iets lager op dit vlak. Net als in andere landen waren leerlingen gemid-

deld gezien wel gemotiveerder geweest als de PISA-opgaven tot een beoordeling hadden geleid. Voor een vergelijkbare toets met beoordeling zouden leerlingen in Nederland zichzelf voor motivatie wel relatief gezien een wat lager cijfer geven dan in de andere vergelijkingslanden, op Vlaanderen na. Door de relatief hoge inschatting qua motivatie voor PISA en de relatief lage inschatting qua motivatie voor een vergelijkbare toets met beoordeling is het verschil tussen beide scenario's relatief klein in Nederland. Hieronder zoomen we in op de relatie tussen beide motivaties (voor de PISA-opgaven en voor een vergelijkbare toets met beoordeling) voor de verschillende landen apart.

3.2 Samenhang tussen motivatie bij verschillende toetsen



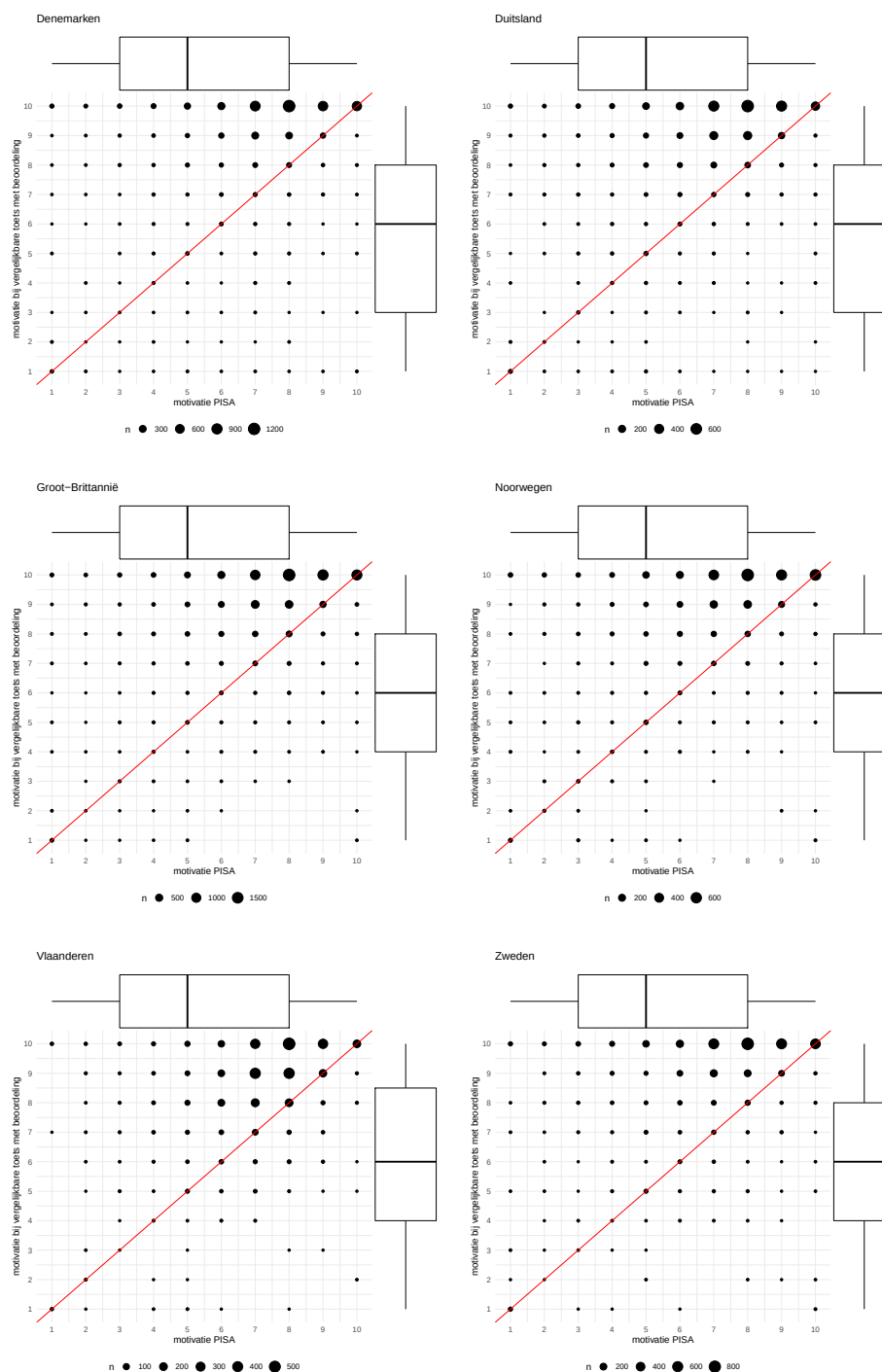
Figuur 6: Samenhang motivatie scores op PISA en toetsen met beoordeling in Nederland

In Figuur 6 is te zien dat in Nederland de cijfers 9 en 10 relatief vaak gegeven worden voor motivatie bij een vergelijkbare toets met beoordeling (y-as). Deze cijfers zijn vaak gecombineerd met iets lagere cijfers voor motivatie bij de PISA-opgaven (x-as). De combinatie waarbij leerlingen hun motivatie bij PISA-opgaven hoger beoordelen dan bij een vergelijkbare toets met beoordeling komt ook af en toe voor (alle

punten onder de rode lijn). Enkele leerlingen hebben voor beide motivaties een '1' gegeven. De boxplotjes in de marges van de figuur laten zien dat de variatie in cijfers voor de motivatie bij de PISA-opgaven wat hoger ligt dan de variatie in cijfers voor de motivatie bij een vergelijkbare toets met beoordeling.

Vergelijken we hetzelfde figuur voor andere landen met die van Nederland, dan valt op dat bijvoorbeeld de Duitse leerlingen meer variatie laten zien in het cijfer wat ze geven voor motivatie (zie Figuur 7). Dit geldt zowel voor de PISA-opgaven als vergelijkbare toetsen met beoordeling. Deze variatie is onder andere zichtbaar in de breedte van de boxplotjes. Ook in Duitsland komt de combinatie van een hoog cijfer op de y-as en een ietswat lager cijfer op de x-as relatief het meest voor. Doordat wat lagere cijfers relatief vaker voorkomen dan in Nederland, drukt dit op de mediaan voor zowel de motivatie van PISA als de motivatie voor een vergelijkbare toets met beoordeling.

Ook ten opzichte van de andere vergelijkingslanden zijn de Nederlandse leerlingen consistent in de cijfergeving rondom motivatie. De variatie in motivatie-cijfers ligt in Nederland relatief laag en de mediaan - mede door het relatief weinig aanwezig zijn van heel lage motivatie-cijfers - relatief hoog.



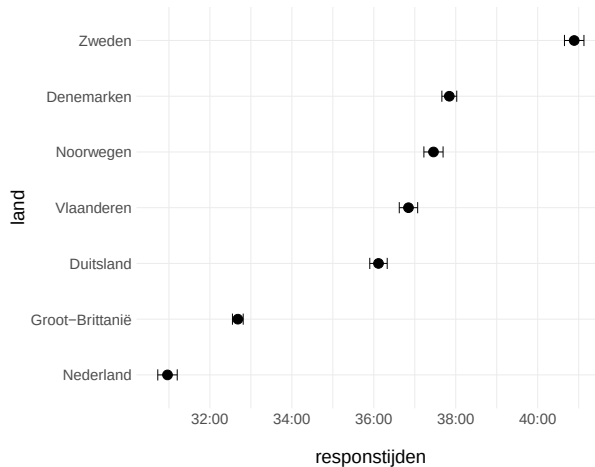
Figuur 7: Samenhang motivatie scores op PISA en toetsen met beoordeling in andere landen

4 Responstijden

De tweede belangrijke indicator voor toetsgedrag is de tijd die leerlingen aan het maken van opgaven besteden. We vergelijken de gemiddelde responstijden die leerlingen aan de toets besteden in de verschillende landen in Sectie 4.1. Dit splitsen we verder uit naar item type (Sectie 4.2), opgaven (Sectie 4.3) en cluster positie (4.4). Ook proberen we de leerlingen te classificeren aan de hand van de verandering van de tijd die zij aan de opgaven besteden, in het verloop van de toets (Sectie 4.5).

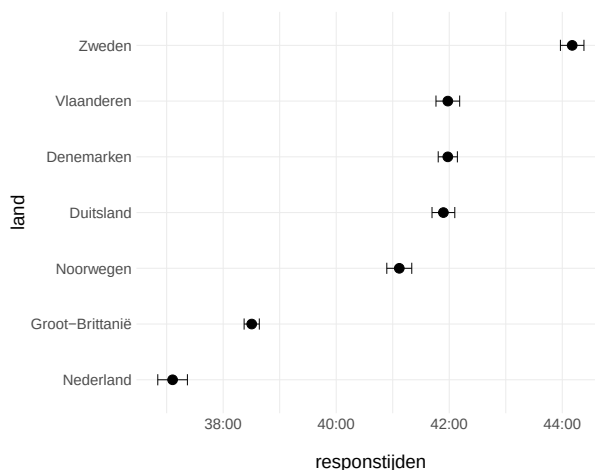
4.1 Gemiddelde responstijd

Gemiddeld gezien besteden Nederlandse leerlingen een half uur aan het maken van de twee wiskunde modules (zie Figuur 8). Dit gemiddelde ligt lager dan de gemiddelde responstijden in de vergelijkingslanden Denemarken, Duitsland, Groot-Brittannië, Noorwegen, Vlaanderen en Zweden. In Zweden besteden leerlingen zelfs - gemiddeld gezien - bijna 10 minuten extra aan de opgaven wiskunde.



Figuur 8: Gemiddelde responstijd wiskunde

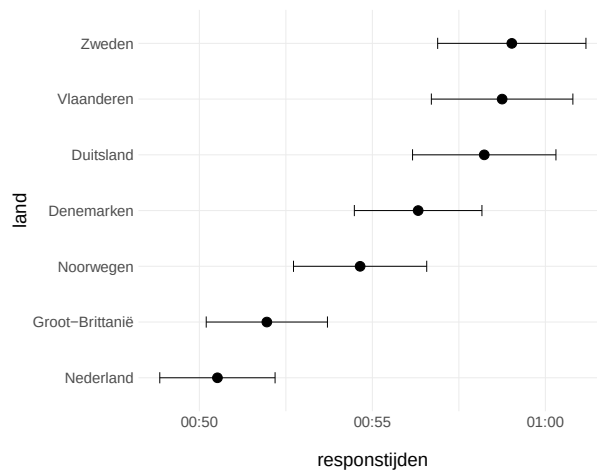
Ook voor natuurwetenschappen geldt dat Nederlandse leerlingen gezien het minste tijd aan de opgaven besteden in vergelijking met de leerlingen in andere landen (zie Figuur 9).



Figuur 9: Gemiddelde responstijd natuurwetenschappen

Gemiddeld gezien besteden Nederlandse leerlingen ongeveer 37 minuten aan het maken van de twee natuurwetenschappen modules. Dit gemiddelde ligt lager dan de gemiddelde responstijden in de vergelijkingslanden. In Zweden besteden leerlingen zelfs - gemiddeld gezien – ruim 7 minuten extra. Het patroon is hetzelfde als bij wiskunde: Nederlandse leerlingen besteden gemiddeld gezien de minste tijd aan de opgaven, daarna leerlingen uit Groot-Brittannië, daarna alle andere vergelijkingslanden met als duidelijke hekkensluiter Zweden. Het verschil tussen wiskunde en natuurwetenschappen is wel dat alle leerlingen – ongeacht (vergelijkings)land – gemiddeld gezien meer tijd besteden aan de natuurwetenschappen-opgaven dan aan de wiskunde-opgaven. Dit kan wellicht verklaard worden uit het feit dat de modules voor wiskunde slechts 11-12 opgaven bevatten, terwijl de modules voor natuurwetenschappen 18-20 opgaven tellen.

Tenslotte zijn de gemiddelde responstijden voor leesvaardigheid te vinden in Figuur 10. Deze gemiddelde responstijden zijn echter niet zoals voor wiskunde en natuurwetenschappen per leerling weergegeven, maar per opgave.

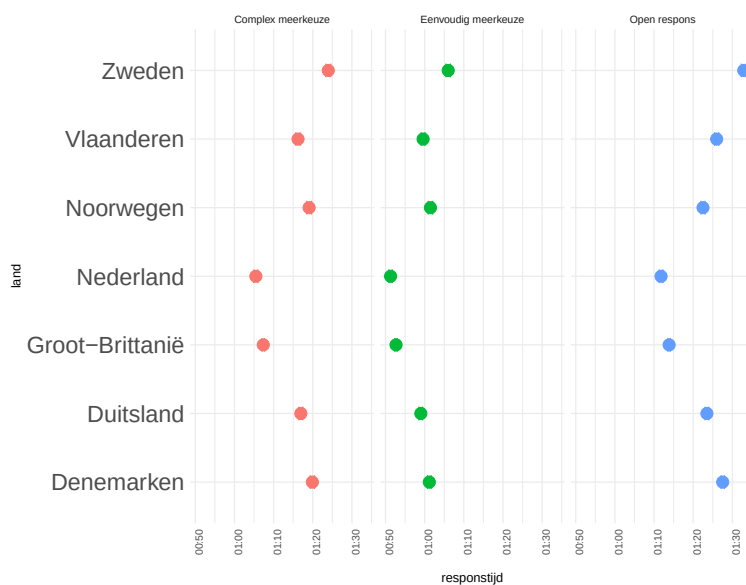


Figuur 10: Gemiddelde responstijd leesvaardigheid

Net als bij wiskunde en natuurwetenschappen besteden Nederlandse leerlingen – gemiddeld gezien – minder tijd in de PISA opgaven dan leerlingen van vergelijkbare landen. In Nederland besteden de leerlingen 50 seconden aan het maken van een opgave. In bijvoorbeeld Vlaanderen - een regio waar de opgaven in dezelfde taal zijn gepresenteerd als in Nederland - is de gemiddelde responstijd een stuk hoger dan in Nederland. Hier besteden leerlingen gemiddeld bijna een minuut aan het oplossen van een opgave. In relatief opzicht is dit een groot verschil.

4.2 Responstijd en item type

Zoals in Sectie 2.2 aangegeven, wordt in PISA gebruik gemaakt van verschillende type opgaven. Eenvoudige meerkeuze-, complexe meerkeuze- en open respons-vragen zijn opgenomen in de toets. Het is voorstelbaar dat leerlingen verschillend reageren op deze verschillende item types. In Figuur 11 zijn de gemiddelde responstijden voor wiskunde uitgesplitst naar item type gepresenteerd.



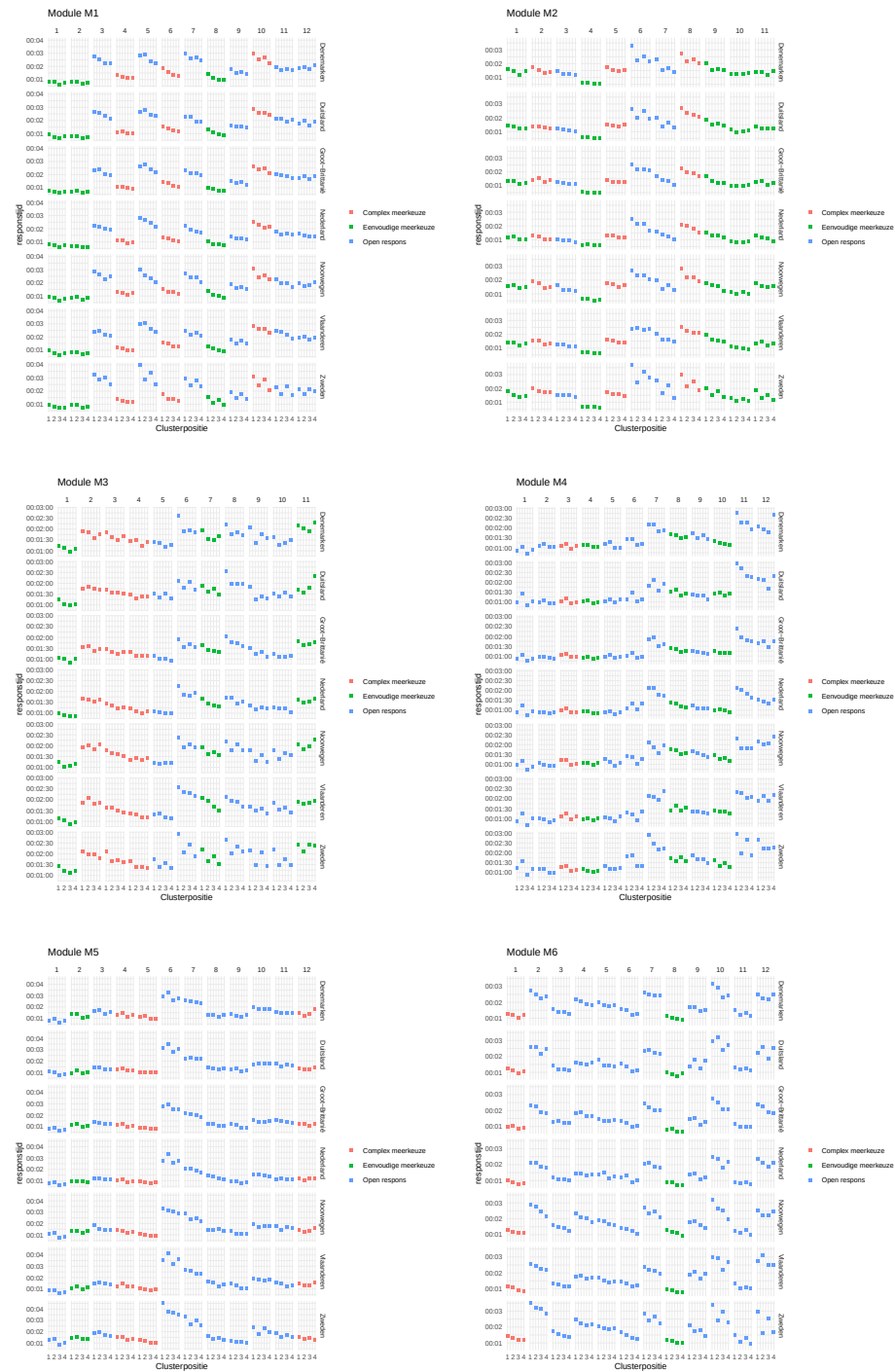
Figuur 11: Gemiddelde responstijden wiskunde per item type

In Figuur 11 is te zien dat Nederlandse leerlingen op alle item types relatief weinig tijd besteden. Leerlingen besteden - niet verrassend - de meeste tijd aan de open respons vragen. Dit patroon is consistent voor alle landen. Dus ondanks dat de gemiddelde responstijd varieert per item type, investeren Nederlandse leerlingen structureel relatief minder tijd. We presenteren deze figuren alleen voor wiskunde, omdat de andere resultaten zeer vergelijkbaar zijn.

4.3 Responstijd en opgaven

In de bovenstaande figuren is steeds het gemiddelde genomen over alle opgaven. Dit maskeert dat er grote verschillen zijn in responstijd voor de verschillende opgaven, die slechts gedeeltelijk afhankelijk zijn van het item type. Figuur 12 laat een meer fijnmazig beeld zien van de responstijden voor wiskunde, afhankelijk van het item type, de opgave, de module en de cluster positie waarin de module (en dus de opgave) zijn afgenomen. De nummers 1-11 of 1-12 staan voor de respectievelijk 11 of 12 opgaven binnen de module.

Toetsgedrag van Nederlandse leerlingen in PISA-2018

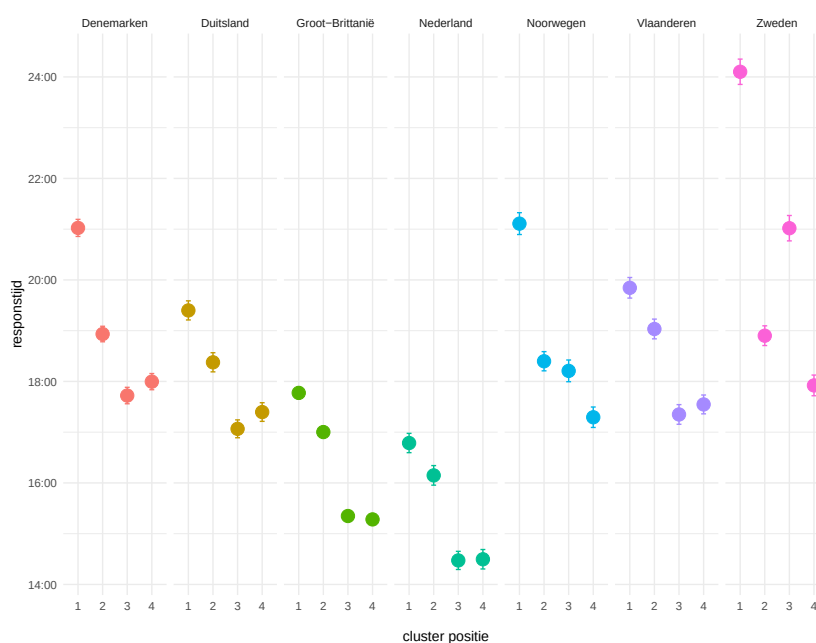


Figuur 12: Responstijden per opgave

Figuur 12 laat zien dat de responstijden van afzonderlijke opgaven behoorlijk beïnvloed worden door de positie binnen de toets. Dit geldt met name voor open respons-opgaven en in mindere mate ook voor de complexe meerkeuze-opgaven.

4.4 Responstijd en cluster positie

De tijd en aandacht die leerlingen aan opgaven besteden, kan dus afnemen in het verloop van de toets zoals in Sectie 4.3 naar voren is gekomen. In Figuur 13 zoomen we in op responstijden voor de verschillende cluster posities voor wiskunde.

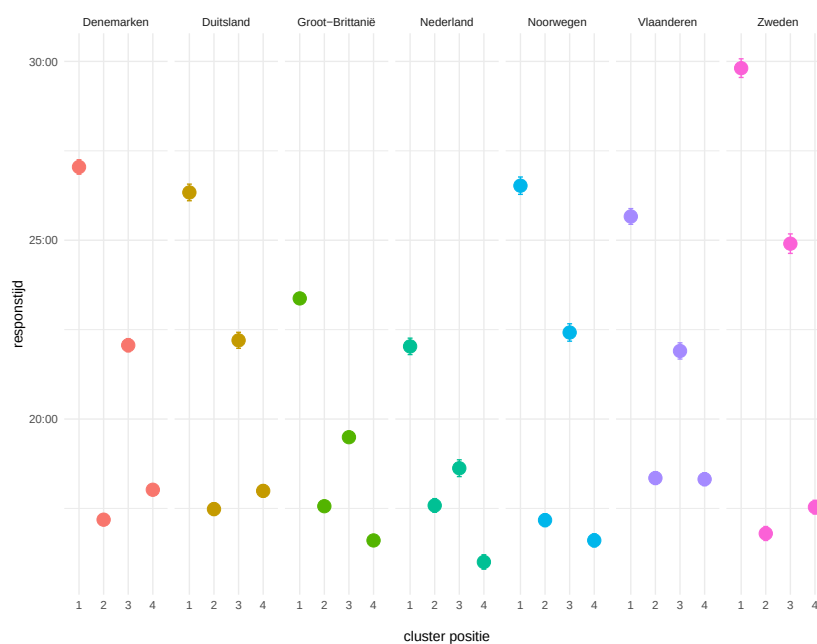


Figuur 13: Responstijden per cluster wiskunde

Allereerst is in Figuur 13 te zien dat responstijd gedeeltelijk afhangt van het cluster waarin de wiskunde-modules zijn gemaakt. Zo is het in ieder land zo dat aan cluster 1 de meeste tijd wordt besteed. In sommige landen is het daarnaast zo dat er een groot verschil is tussen de twee clusters voor de pauze (1-2) en de twee clusters na de pauze (3-4). Na de pauze nemen leerlingen gemiddeld minder tijd om de opgaven te voltooien. Zweden vormt hierop een opmerkelijke uitzondering: hier is een 'opleving' in termen van responstijd te zien na de pauze

in het eerste cluster. Vergelijken we Nederland met de vergelijkingslanden, dan valt op dat Nederlandse leerlingen bij elke cluster positie - gemiddeld gezien - minder tijd besteden aan het beantwoorden van de opgaven. De gemiddelde responstijden voor de clusters voor (1-2) en na de pauze (3-4) liggen daarbij relatief dichtbij elkaar (ongeveer 17 minuten voor de pauze en ruim 14 minuten na de pauze).

Anders dan voor wiskunde is het bij natuurwetenschappen niet zo dat er meer tijd wordt besteed aan de twee clusters voor de pauze (1-2) dan de twee clusters na de pauze (3-4), zoals in Figuur 14 te zien is. Het is juist zo dat leerlingen de meeste tijd besteden aan de eerste cluster voor of na de pauze (cluster 1 en 3) en relatief minder tijd aan het daaropvolgende cluster (cluster 2 of cluster 4). Het is daarbij wel zo dat cluster 1 relatief meer tijd kost – gemiddeld gezien – dan cluster 3, ondanks dat deze beide de eerste van de twee natuurwetenschappen clusters zijn. De verschillen tussen de eerste twee clusters (clusters 1 en 3) en de tweede set clusters (cluster 2 en 4) zijn relatief het grootst in Zweden en relatief het kleinst in Nederland.

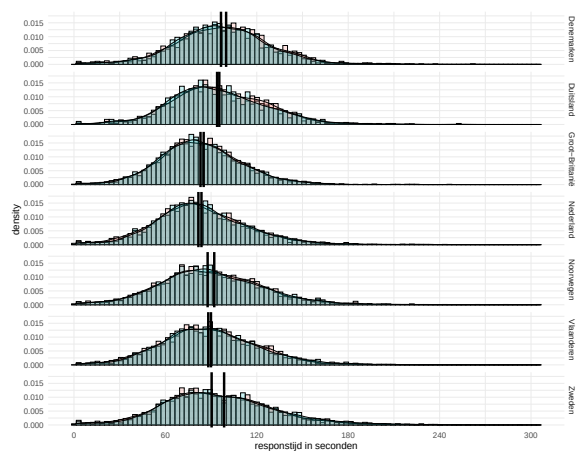


Figuur 14: Responstijden per cluster natuurwetenschappen

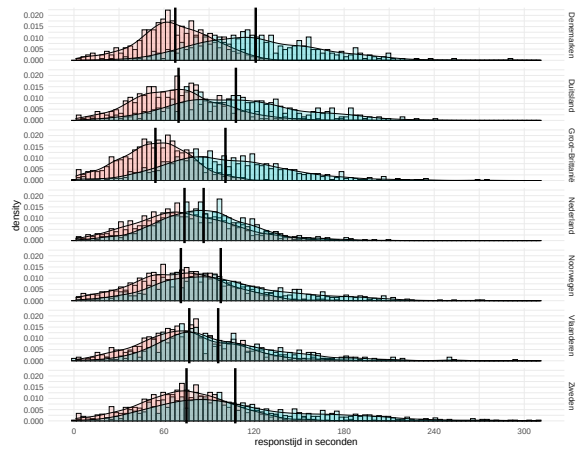
4.5 Responstijd en type leerling

Hierboven is het beeld geschetst dat Nederlandse leerlingen relatief minder tijd investeren in de PISA-opgaven dan hun leeftijdsgenoten uit de vergelijkingslanden en dat deze tijdsinvestering afneemt naarmate de toets vordert. Deze conclusies zijn gebaseerd op gemiddelden over alle Nederlandse deelnemers. Het hoeft dus niet zo te zijn dat alle Nederlandse leerlingen steeds minder tijd investeren. Er kunnen verschillende type leerlingen bestaan: (1) leerlingen die ongeveer evenveel tijd investeren in de eerste en tweede module, (2) leerlingen die meer tijd gaan investeren van de eerste naar de tweede module en (3) leerlingen die minder tijd gaan investeren van de eerste naar de tweede module¹⁴. De frequentie van deze verschillende type leerlingen en het verschil in hun tijdsinvestering van de eerste naar tweede cluster positie (ofwel 1 - 2 ofwel 3 - 4) kan per land verschillen. Deze verschillen zijn voor wiskunde weergegeven in Figuur 15. De responstijden voor de eerste module zijn weergegeven in blauw, de responstijden voor de twee module zijn weergegeven in roze.

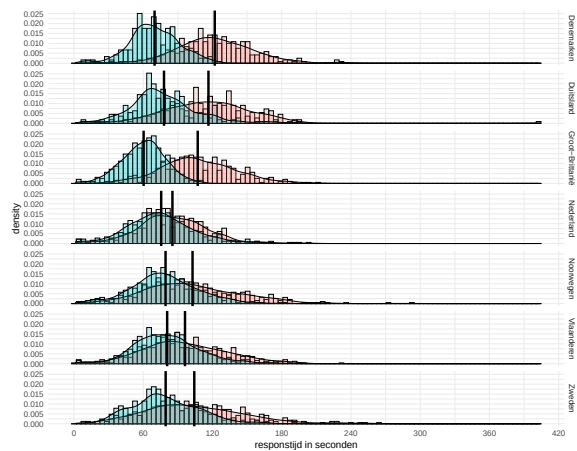
¹⁴Om te bepalen of een leerling valt onder respectievelijk type 1, 2 of 3 is op basis van de responstijden van de leerlingen een Bayes factor bepaald voor de hypothesen “tijdsinvestering module 1 = module 2”, “tijdsinvestering module 1 > module 2” en “tijdsinvestering module 1 < module 2”. De indeling in type 1, 2 en 3 heeft plaats gevonden op basis van de hoogste Bayes factor.



(a) Type I leerlingen



(b) Type II leerlingen



(c) Type III leerlingen

Figuur 15: Type leerlingen

Type 1

Voor de “type 1” leerlingen geldt dat ze ongeveer evenveel tijd hebben besteed aan de eerste als de tweede module. Daarom overlappen in alle (vergelijkings)landen de verdelingen voor de responstijden voor de eerste en tweede module grotendeels (zie Figuur 15a). Wel verschillen de gemiddelde responstijden over de landen. Nederland heeft een relatief lage gemiddelde responstijd t.o.v. de vergelijkingslanden.

Type 2

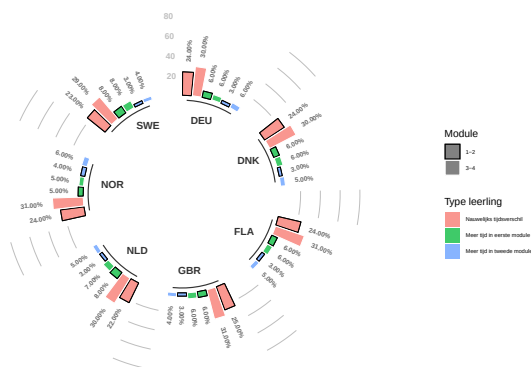
Ondanks dat het aantal “type 2” leerlingen - waarbij de responstijd in de eerste module (in blauw weergegeven) hoger ligt dan in de tweede module (roze) - relatief gezien ongeveer even hoog ligt in elk van de vergelijkingslanden, zijn er duidelijke verschillen zichtbaar tussen de landen. In Nederland ligt de verdeling van de responstijden voor de eerste en tweede module het dichtst bij elkaar voor de type 2 leerlingen. Er is een lichte verschuiving te zien van de roze naar de blauwe verdeling, maar er is een grote mate van overlap. De variantie in de responstijden is ongeveer gelijk voor de eerste en tweede module. Dit is niet het geval voor alle vergelijkingslanden; in deze landen is er meer variatie in responstijden zichtbaar in de eerste module terwijl de responstijden voor de tweede module dichterbij elkaar liggen. In sommige landen - zoals Duitsland, Denemarken en Groot-Brittannië - is er een veel duidelijker verschil te zien tussen de gemiddelde responstijd in module 1 en 2. Als leerlingen dus relatief meer tijd besteden in de eerste module dan de tweede, dan is dit verschil in vergelijkingslanden vaker groter en lopen de responstijden op de twee modules vaker meer uiteen. Bovendien is te zien dat in de vergelijkingslanden er veel meer ‘uitschieters’ zijn; leerlingen met een relatief hoge responstijd in de eerste module. Deze uitschieters zijn er niet in Nederland. Merk op dat in landen als Duitsland, Denemarken en Groot-Brittannië de gemiddelde responstijd in de tweede module lager ligt dan die in Nederland. Dit ‘compenseren’ deze leerlingen echter met een hogere gemiddelde responstijd in de eerste module.

Type 3

Als leerlingen behoren tot “type 3” (dus: minder tijd besteden aan de eerste dan aan de tweede module), dan is het verschil in tijdsinvestering in vergelijkingslanden (met name Duitsland, Denemarken en Groot-Brittannië) vaker groter dan in Nederland. Bovendien is het we-

derom zo dat de variantie in responstijden niet hetzelfde is in de eerste en tweede module in deze vergelijkingslanden. Uitschieters - leerlingen met een opvallend hoge responstijd in de tweede module - komen ook minder voor in Nederland dan in vergelijkingslanden.

In Figuur 16 is te zien hoeveel leerlingen - procentueel - voor wiskunde vallen onder type 1 (nauwelijks tijdverschil), type 2 (meer tijd besteed aan de opgaven in de eerste module) en type 3 (meer tijd in tweede module) binnen de verschillende landen.



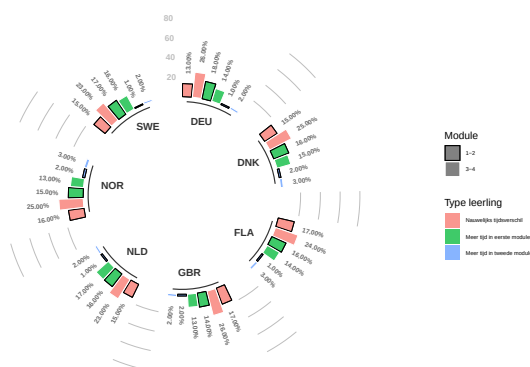
Figuur 16: Responstijden en type leerlingen wiskunde

Interessant is dat er geen grote verschillen tussen de landen lijken te bestaan. In elk van de landen is de tijdsinvestering van de meeste leerlingen (23-31 procent) in de eerste en tweede module ongeveer gelijk. Op Noorwegen na, geldt voor alle landen dat de een na grootste groep bestaat uit leerlingen die wat meer tijd in de eerste module dan in de tweede module stoppen.

Samengevat: in Nederland komen - in vergelijking met Duitsland, Denemarken, Vlaanderen, Groot-Brittannië, Noorwegen en Zweden - ongeveer evenveel leerlingen voor die ongeveer evenveel (type 1), minder (type 2) of meer (type 3) tijd investeren naarmate het wiskunde onderdeel vordert. Zoomen we in op deze drie typen leerlingen, dan valt op dat Nederlandse leerlingen in vergelijking consistent zijn in hun tijdsinvestering. Verschillen in responstijd tussen de eerste en tweede module vallen relatief gezien mee, ook als leerlingen tot de type 2 en type 3 worden gerekend. Opvallend hoge respons-

tijden komen in Nederland amper voor, terwijl deze in vergelijkingslanden Zweden (het vergelijkingsland met de hoogste gemiddelde responstijd, zie Figuur 8) veel vaker voorkomen en ook de variantie in responstijden is relatief beperkter in Nederland. Het feit dat Nederlandse leerlingen gemiddeld gezien minder tijd stoppen in het beantwoorden van de vragen betekent niet dat dit altijd het geval is; in vergelijkingslanden waar type 2- en 3-leerlingen een groot verschil laten zien in responstijd (zie bijvoorbeeld Duitsland), ligt de gemiddelde responstijd voor de snelst afgeronde module lager dan in Nederland.

Anders dan bij wiskunde is het zo dat het percentage type 1, type 2 en type 3 leerlingen wel verschilt bij natuurwetenschappen afhankelijk van of de leerling zijn modules in cluster 1-2 of 3-4 heeft gemaakt (zie Figuur 17). Meer specifiek: bij cluster 1-2 heeft een leerling relatief vaker meer tijd in de eerste module gestoken dan in de tweede module, terwijl bij cluster 3-4 het zo is dat er vaker nauwelijks tijdsverschil is. Het patroon over landen heen is redelijk vergelijkbaar.



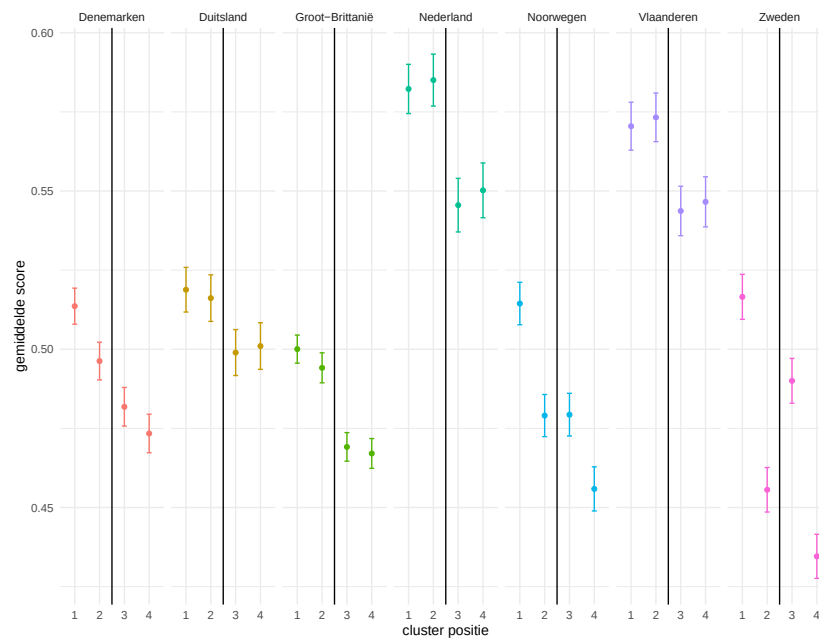
Figuur 17: Responstijden en type leerlingen natuurwetenschappen

5 Toetsconcentratie

Tussen het tweede en het derde cluster is er een korte pauze in de PISA-afname. In dit hoofdstuk wordt de vraag geadresseerd of leerlingen de opgaven voor deze pauze even goed maken als na de pauze. Door de combinatie van toevallige toewijzing van leerlingen aan toetsversies en het afnamedesign, die ervoor zorgen dat voor wiskunde en natuurwetenschappen alle opgaven voor en na de pauze afgenomen worden, kan deze vergelijking goed gemaakt worden. Voor leesvaardigheid is deze vergelijking minder betekenisvol door het adaptieve karakter van de toetsing van dit domein. De opgaven worden gedurende de toets namelijk makkelijker of moeilijker afhankelijk van de prestatie op de eerdere opgaven. Een vergelijking van het aandeel leerlingen dat de opgaven goed maakt naarmate de toetsafname vordert is dan niet meer informatief om toetsgedrag in kaart te brengen. De resultaten voor wiskunde en natuurwetenschappen worden gepresenteerd in respectievelijk Sectie 5.1 en Sectie 5.2.

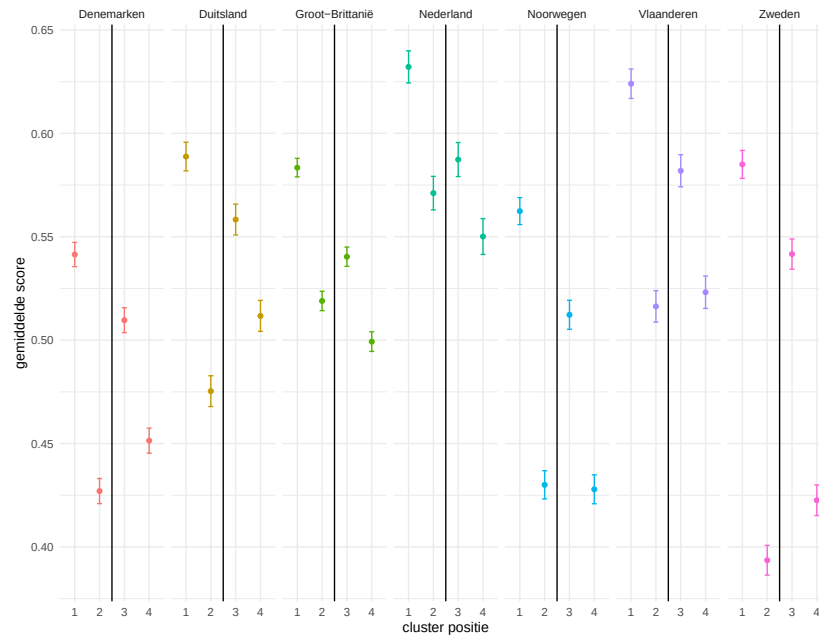
5.1 Wiskunde

In Figuur 18 is te zien dat in elk land geldt dat leerlingen met wiskunde clusters vóór de pauze gemiddeld gezien een hogere score halen dan vergelijkbare leerlingen die deze clusters na de pauze maken. In Nederland is het verschil tussen de laatste cluster voor de pauze (cluster 2) en de eerste cluster na de pauze (cluster 3) het grootst.



Figuur 18: Proportie behaalde scores en cluster positie wiskunde

5.2 Natuurwetenschappen



Figuur 19: Proportie behaalde scores en cluster positie natuurwetenschappen

In Figuur 19 is te zien dat in elk land geldt dat leerlingen met natuurwetenschappen clusters vóór de pauze gemiddeld gezien een hogere score halen dan vergelijkbare leerlingen die deze clusters na de pauze maken. Het patroon van Nederlandse leerlingen bij natuurwetenschappen is wel anders dan bij wiskunde. Waar bij de wiskunde clusters het grootste verschil zichtbaar was tussen het laatste cluster voor de pauze (cluster 2) en het eerste cluster na de pauze (cluster 3), is bij natuurwetenschappen het verschil tussen het eerste cluster voor de pauze (cluster 1) en het laatste cluster na de pauze (cluster 4) het grootst.

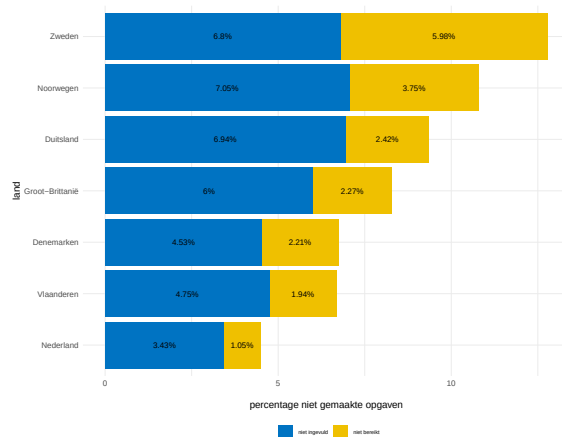
6 Niet-gemaakte opgaven

De vierde en laatste indicator die we gebruiken om toetsgedrag in kaart te brengen, is het aantal niet gemaakte opgaven van de landen te vergelijken. Leerlingen kunnen een opgave niet maken doordat zij deze overslaan en verder werken aan de andere opgaven. Ze kunnen echter ook een opgave niet bereiken omdat zij zelf of de toetsleider de toets afsluit. Eerst laten we in Sectie 6.1 het percentage niet-gemaakte opgaven per land zien. In Secties 6.2, 6.3 en 6.4 wordt dit verder uitgediept voor respectievelijk item type, cluster positie en type leerling.

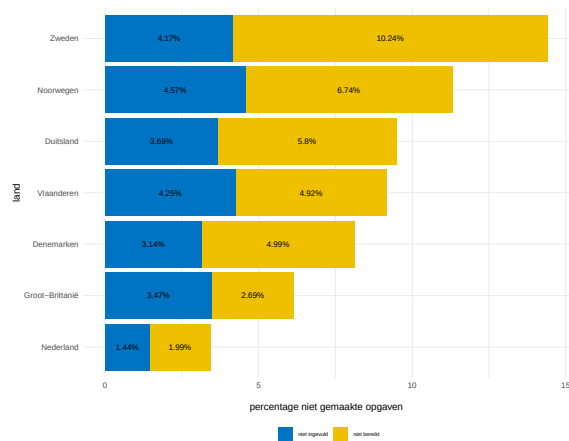
6.1 Percentage niet-gemaakte opgaven

Leerlingen maken vaak niet alle opgaven van hun PISA-toetsversie. In de Figuur 20 is te zien dat in Nederland het aandeel vragen dat niet door de leerlingen gemaakt is - gemiddeld gezien - laag is ten opzichte van de vergelijkingslanden. In Zweden ligt het aantal overgeslagen opgaven gemiddeld gezien bijvoorbeeld ruim twee keer zo hoog voor wiskunde en ruim drie keer zo hoog voor natuurwetenschappen.

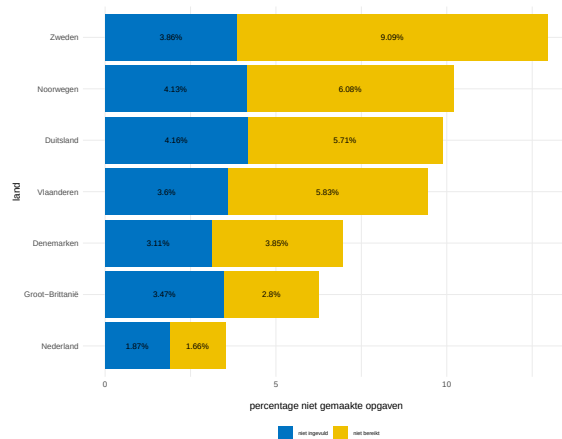
Toetsgedrag van Nederlandse leerlingen in PISA-2018



(a) Percentage niet-gemaakte opgaven wiskunde



(b) Percentage niet-gemaakte opgaven natuurwetenschappen

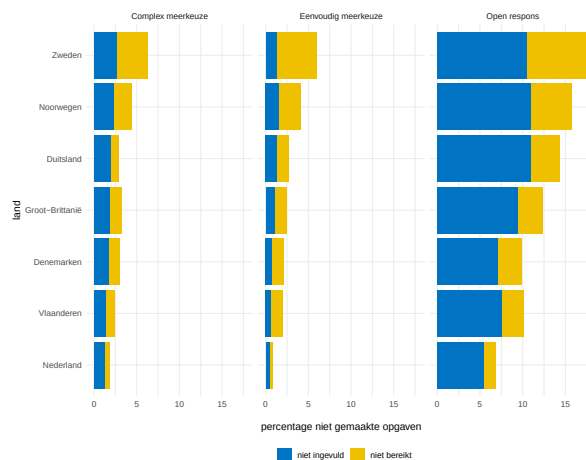


(c) Percentage niet-gemaakte opgaven leesvaardigheid

Figuur 20: Percentage niet-gemaakte opgaven

6.2 Niet-gemaakte opgaven en item type

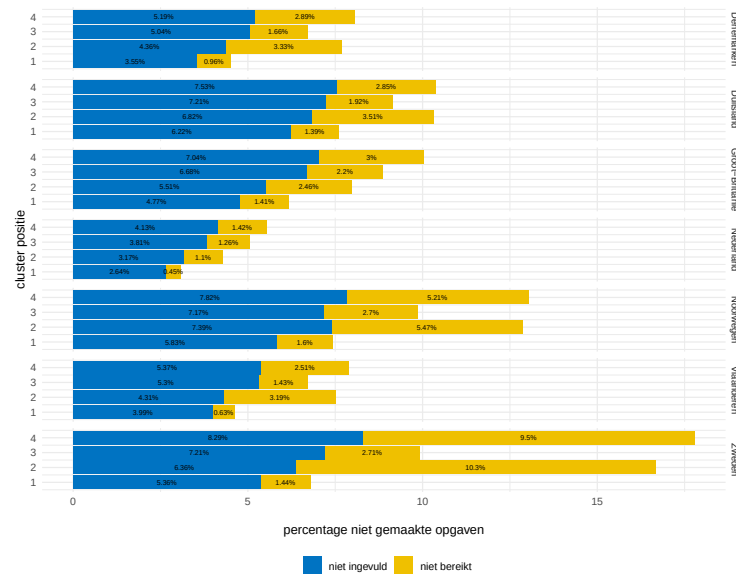
Als we het aantal niet gemaakt vragen opsplitsen naar soort item, dan valt allereerst op dat bij open respons-opgaven het hoogste aantal niet-gemaakte opgaven voorkomen in elk van de landen. De simpele meerkeuzevragen wordt in elk van de landen het minst overgeslagen. De open respons opgaven vragen een grotere cognitieve inspanning om een antwoord te geven dan de meerkeuze opgaven. Het is dus niet verwonderlijk dat in deze toets, juist deze vragen tot een hoger aandeel niet ingevulde vragen leidt. Voor elk van deze type opgaven geldt dat Nederland gemiddeld gezien een relatief laag percentage overgeslagen opgaven kent, zie Figuur 21 voor de resultaten van wiskunde.



Figuur 21: Percentage niet-gemaakte opgaven per type item wiskunde

6.3 Niet-gemaakte opgaven en cluster positie

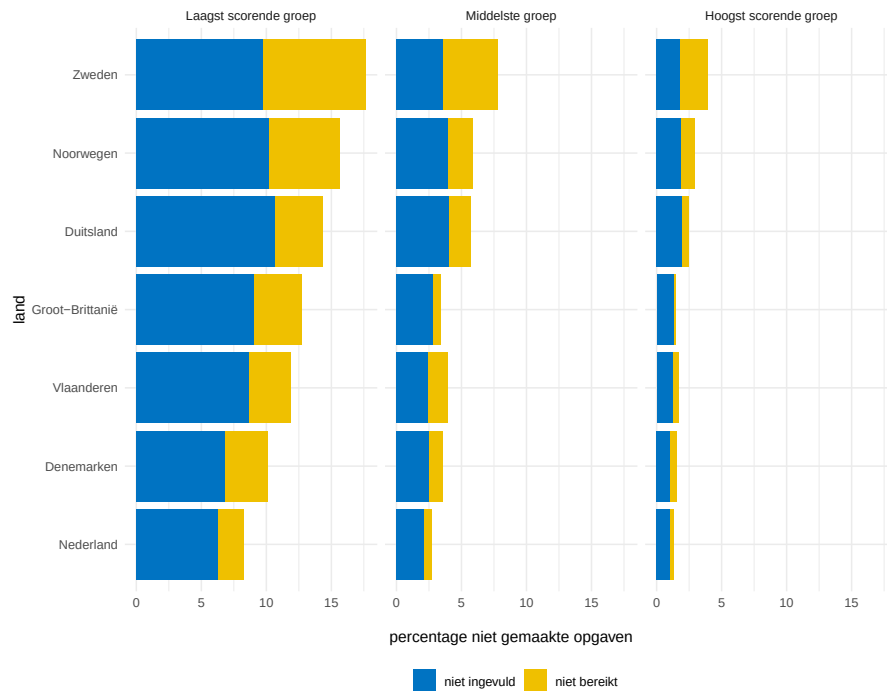
In de Figuur 22 is zien hoe de het percentage overgeslagen opgaven zich verhoudt naar de verschillende cluster posities. In Nederland (net als in Groot-Brittannië) is te zien dat de percentage niet gemaakte opgaven iets toeneemt van cluster positie 1 naar cluster positie 4. In Zweden en Noorwegen worden het hoogst aandeel opgaven niet-gemaakt op cluster positie 2 (voor de pauze) en 4 (bij het einde van de toets). Met name in Zweden is het aandeel opgaven dat niet bereikt wordt binnen deze twee cluster posities hoog.



Figuur 22: Percentage niet-gemaakte opgaven per cluster wiskunde

6.4 Niet-gemaakte opgaven en type leerling

Om te beoordelen of er verschillen bestaan tussen goed en slecht presterende leerlingen en de mate waarin zij opgaven niet invullen hebben we drie groepen leerlingen gemaakt. Een groep leerlingen die boven-gemiddeld goed presteert, een gemiddeld presterende groep en een onder-gemiddelde presterende groep leerlingen. Het aantal niet-ingevulde opgaven ligt iets hoger bij de laagst-scorende groep. In elk van deze groepen scoort Nederland laag ten opzichte van vergelijkbare landen. De resultaten zijn te vinden in Figuur 23.



Figuur 23: Percentage niet-gemaakte opgaven per type leerling wiskunde

7 Conclusies en discussie

In deze verdiepende studie hebben we het toetsgedrag van Nederlandse leerlingen vergeleken met het toetsgedrag van leerlingen in de omliggende landen en regio's (Groot-Brittannië, Denemarken, Noorwegen, Zweden, Duitsland en Vlaanderen). Toetsgedrag is relevant in de context van (internationale) peilingsonderzoeken zoals PISA, omdat hier vaak voor leerlingen geen externe motivatie bestaat om een maximale prestaties te leveren op de toets door het ontbreken van persoonlijke gevolgen voor de prestaties op de toets. In hoeverre leerlingen wel gemotiveerd zijn en moeite stoppen in het leveren van een goede prestatie is onderzocht in deze studie. Daarbij is ook expliciet gekeken naar de vergelijking van het toetsgedrag van Nederlandse leerlingen met leerlingen uit de zes bovengenoemde andere landen. Leerlingen zijn vergeleken op de door hen zelf ingevulde motivatie, responstijden, toetsconcentratie en het aantal overslagen vragen tijdens de toets.

De door leerlingen zelf ingevulde motivatie om de PISA te maken en de aangegeven motivatie voor een hypothetische toets waarbij wel een beoordeling wordt gegeven verschillen niet veel tussen de landen. De resultaten van de Nederlandse leerlingen vertonen wel minder spreiding dan de resultaten van de leerlingen uit andere landen.

Nederlandse leerlingen maken gemiddeld gezien de opgaven die opgenomen zijn in de PISA sneller dan leerlingen in de andere landen. Dit kan overigens (gedeeltelijk) ook veroorzaakt worden door verschillen in taal. Internationaal gezien geldt Engels bijvoorbeeld als een heel compacte taal¹⁵. Niet alle verschillen in responstijd zijn dus toe te schrijven aan verschillen in toetsgedrag. In die zin zijn de verschillen met de responstijd in Vlaanderen wellicht informatiever. Ook voor de vergelijking met Vlaanderen geldt dat Nederlandse leerlingen relatief snel de toets maken. Voor wiskunde en natuurwetenschappen - de domeinen die wat responstijd betreft het best met elkaar te vergelijken dat Nederlandse leerlingen gemiddeld rond respectievelijk 31 en 37 minuten aan deze opgaven besteden. Vlaamse leerlingen besteden gemiddeld 37 minuten aan de wiskunde- en 42 minuten aan de natuurwetenschappen opgaven.

¹⁵<https://www.w3.org/International/articles/article-text-size.en>

Voor Nederland geldt net zoals voor Denemarken, Duitsland, Groot-Brittannië en Vlaanderen dat de toetsconcentratie - uitgedrukt in het aandeel goed gemaakte opgaven voor wiskunde afneemt na de pauze. Voor Noorwegen geldt dit in iets minder mate. Zweden heeft een opvallend patroon: De prestaties neemt tijdens de toets voor de pauze af, om na de pauze weer toe te nemen en vervolgens weer af te nemen. Voor natuurwetenschappen is een vergelijkbaar beeld voor alle landen te zien zoals voor Zweden bij wiskunde. De prestatie neemt af, stijgt weer naar de pauze en neemt uiteindelijk weer af.

Opvallend is dat het aandeel niet-gemaakte opgaven in Nederland heel erg laag is in vergelijking met de andere landen. Dit geldt zowel voor de opgaven die overgeslagen zijn als de opgaven die niet bereikt zijn. In Zweden en Noorwegen worden telkens de hoogste percentages opgaven gevonden die niet door leerlingen zijn gemaakt. Dat geldt voor alle drie hoofddomeinen. Voor Zweden geldt zelfs dat tussen de 10 en 15% van de opgaven niet gemaakt worden. Het zou kunnen dat Nederlandse leerlingen meer dan hun leeftijdgenoten in het buitenland over een soort “toetswijsheid” beschikken, waardoor ze sneller een antwoord geven op een opgave, ongeacht of ze het antwoord weten. Dat kan zeker bij meerkeuze opgaven geen onverstandige strategie zijn.

Samengevat kan gezegd worden dat het toetsgedrag van Nederlandse leerlingen wat een aantal aspecten betreft er uit springt ten opzichte van het toetsgedrag van hun leeftijdgenoten in de vergelijkingslanden. Ze besteden relatief weinig tijd aan de opgaven en geven slechts in een gering aantal gevallen geen antwoord op de vragen in de toets. Tenslotte, het is nog van belang om af te sluiten met de opmerking dat het gedeeltelijk afwijkende toetsgedrag in Nederland niet noodzakelijkerwijs leidt tot lage toetsresultaten. Zo is de gemiddelde prestatie voor wiskunde in Nederland de hoogste van de in deze studie opgenomen landen¹⁶.

¹⁶OECD (2019), *PISA 2018 Results (Volume I): What Students Know and Can Do*, PISA, OECD Publishing, Paris, <https://doi.org/10.1787/5f07c754-en>.

Lijst van figuren

1	Afnamedesign PISA-2018	5
2	Aantal type items per domein	7
3	Aantal deelnemende leerlingen per land en domein	8
4	De <i>effort</i> thermometer in PISA-2018	10
5	Gemiddelde motivatie-scores op PISA en toetsen met beoordeling	10
6	Samenhang motivatie scores op PISA en toetsen met beoordeling in Nederland .	11
7	Samenhang motivatie scores op PISA en toetsen met beoordeling in andere lan- den	13
8	Gemiddelde responstijd wiskunde	14
9	Gemiddelde responstijd natuurwetenschappen	15
10	Gemiddelde responstijd leesvaardigheid	16
11	Gemiddelde responstijden wiskunde per item type	17
12	Responstijden per opgave	18
13	Responstijden per cluster wiskunde	19
14	Responstijden per cluster natuurwetenschappen	20
15	Type leerlingen	22
16	Responstijden en type leerlingen wiskunde	24
17	Responstijden en type leerlingen natuurwetenschappen	25
18	Proportie behaalde scores en cluster positie wiskunde	27
19	Proportie behaalde scores en cluster positie natuurwetenschappen	28
20	Percentage niet-gemaakte opgaven	30
21	Percentage niet-gemaakte opgaven per type item wiskunde	31
22	Percentage niet-gemaakte opgaven per cluster wiskunde	32
23	Percentage niet-gemaakte opgaven per type leerling wiskunde	33