

Rekenen-Wiskunde in het VO

# Technisch rapport Peil.Onderwijs Meting 2022-leerjaar 2

Expertisecentrum Nederlands: Joyce Gubbels, Nicole Swart

CITO: Bas Hemker, Jorine Vermeulen

Universiteit Twente: Martina Meelissen, Lyset Rekers-Mombarg

KBA Nijmegen: Tessa Jenniskens, Annemarie van Langen

April 2023

UNIVERSITEIT TWENTE.



EXPERTISECENTRUM  
NEDERLANDS



KBA  Nijmegen



Projectnummer: 2021059

© 2023 KBA Nijmegen

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden vermenigvuldigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op welke andere wijze dan ook, en evenmin in een retrieval systeem worden opgeslagen, zonder de voorafgaande schriftelijke toestemming van KBA Nijmegen.

No part of this book/publication may be reproduced in any form, by print, photoprint, microfilm or any other means without written permission from the publisher.

## Inhoudsopgave

1	Introductie	1
1.1	Inleiding	1
1.2	Achtergrond	1
1.3	Onderzoeksvragen	2
1.4	Invloed Covid-19	2
1.5	Leeswijzer	3
2	Ontwikkeld instrumentarium	4
2.1	Rekenopgaven	4
2.1.1	Theoretische uitgangspunten	4
2.1.2	Definitieve itemverzameling in het onderzoek	6
2.2	Leerling, docent- en schoolvragenlijsten	8
3	Onderzoeksdesign	13
3.1	Inleiding	13
3.2	Steekproefdesign	13
3.2.1	Eerste trap: gewogen steekproef van vestingen	13
3.2.2	Tweede trap: steekproef van één klas per vestiging	14
3.3	Afnamedesign rekenopgaven	14
4	Werving, dataverzameling en respons	19
4.1	Organisatie scholenwerving	19
4.2	Respons van scholen	19
4.3	Selectie van klassen	21
4.3.1	Totaal aantal tweede klassen	21
4.3.2	Selectie van één klas per vestiging	21
4.4	Dataverzameling	22
4.5	Respons	23
5	Schaling rekenvaardigheid	29
5.1	Resultaten van de toetsafname	29
5.2	Klassieke resultaten per itemset	30
5.3	Ontbrekende waarden	32
5.4	Over de toepassing van IRT	36
5.5	Keuze van de te analyseren schalen: toepassing van IRT in dit onderzoek	39
5.5.1	Ter inleiding	39
5.5.2	Schaling op basis van itemkenmerken: losse schalen	40
5.5.3	Samenhang uitkomsten kenmerken van de taakuitvoering	43
5.5.4	Schaling op basis van een enkele rekenschaal	46
5.5.5	Betrouwbaarheid	48
5.5.6	Populatieverdelingen en schattingen van leerlingvaardigheid	49
5.6	Referentieniveaus	52
6	Schaling contextvragenlijsten	56
6.1	Inleiding	56
6.2	Leerlingvragenlijst (versie regulier VO)	56
6.2.1	Ontbrekende waarden	56
6.2.2	Achtergrondkenmerken	56

6.2.3	Domeinspecifieke kenmerken	59
6.3	Leerlingvragenlijst (versie praktijkonderwijs)	63
6.3.1	Ontbrekende waarden	63
6.3.2	Achtergrondkenmerken	63
6.3.3	Domeinspecifieke kenmerken	65
6.4	Docentvragenlijst	67
6.4.1	Ontbrekende waarden	67
6.4.2	Achtergrondkenmerken	67
6.4.3	Domeinspecifieke kenmerken	69
6.5	Schoolvragenlijst	83
6.5.1	Ontbrekende waarden	83
6.5.2	Algemene schoolkenmerken	83
6.5.3	Domeinspecifieke schoolkenmerken	86
7	Leerlingprestaties Rekenen-Wiskunde	91
7.1	Inleiding	91
7.2	Leerlingprestaties Rekentaalvaardigheid (onderzoeksvraag 1)	91
7.3	Leerlingprestaties per domein en per afnameconditie (onderzoeksvraag 2)	95
7.3.1	Ter inleiding	95
7.3.2	Domein B – Getallen en bewerkingen	97
7.3.3	Domein C – Verhoudingen	102
7.3.4	Domein D – Meten en Meetkunde	106
7.3.5	Domein E – Verbanden	110
7.3.6	Domein N1 – Probleemoplossen	114
8	Samenhang tussen leerlingprestaties en kenmerken van leerlingen, docenten en het onderwijsleerproces	117
8.1	Inleiding	117
8.2	Multiple imputatie	117
8.3	Multilevel analyses	118
8.4	Overzicht van de multilevel modellen	118
8.5	Resultaten multilevel analyses	121
8.5.1	Nulmodel	121
8.5.2	Leerlingkenmerken	121
8.5.3	Docentkenmerken	125
8.5.4	Schoolkenmerken	134
8.5.5	Eindmodel	139
9	Samenvatting en beantwoording onderzoeksvragen	144
9.1	Inleiding en context	144
9.2	Beantwoording onderzoeksvragen	144
9.2.1	Het onderwijsleerproces op de scholen (Hoofdstuk 6)	145
9.2.2	De prestaties van de leerlingen (Hoofdstuk 7)	147
9.2.3	Samenhang tussen prestaties en kenmerken van leerlingen, docenten en het onderwijsleerproces (Hoofdstuk 8)	149
	Literatuur	153

Bijlagen	155
Bijlage 1 – Leerlingvragenlijst regulier VO: rechte tellingen	156
Bijlage 2 – Docentvragenlijst: rechte tellingen	160
Bijlage 3 – Schoolvragenlijst: rechte tellingen	166
Bijlage 4 – Syntax hercodering predictoren	170

# 1 Introductie

## 1.1 Inleiding

Peil.onderwijs is het geheel van periodieke peilingsonderzoeken in het primair onderwijs en – sinds de uitbreiding in april 2021 – ook in het voortgezet onderwijs en (voortgezet) speciaal onderwijs. De Inspectie van het Onderwijs voert de regie over de peilingsonderzoeken, in opdracht van het ministerie van Onderwijs, Cultuur en Wetenschap. Peil.Onderwijs geeft input voor een brede dialoog over de inhoud, kwaliteit en het niveau van het onderwijs in verschillende leergebieden. De peilingen brengen de kennis, vaardigheden en houding van leerlingen voor de gepeilde vakken of leergebieden in beeld. Ook geven ze zicht op het aanbod en het onderwijsleerproces op de scholen en de samenhang daarvan met de leerlingprestaties op het betreffende leergebied. De eerste drie peilingen in het voortgezet onderwijs (VO) worden gefinancierd uit NPO-gelden. Het programma is vastgesteld tot 2030 en voor de periode 2025-2030 is het programma gefinancierd door OCW.

Nadat tussen mei en juli 2021 de kwartiermaking van het programma Peil.onderwijs voor het VO had plaatsgevonden duurde de aanbestedingsprocedure voor de eerste VO-peiling, die van Rekenen-Wiskunde<sup>1</sup>, tot eind oktober 2021. De voorbereidingstijd voor dit peilingsonderzoek was hierdoor extreem kort, aangezien de opdrachtgever (OCW) de inspectie verzocht had om de dataverzameling te laten uitvoeren in het voorjaar van 2022. In de vijf maanden tussen aanbesteding en dataverzameling was het niet mogelijk een pretest uit te voeren voor de toets en vragenlijsten. Daarom hebben de uitvoerders van de peiling Rekenen-Wiskunde in het VO, een consortium van onderzoekers van het Expertisecentrum Nederlands, Cito, de Universiteit Twente en KBA Nijmegen, gekozen voor beproefde toetsopgaven en beproefde vragensets uit eerder uitgevoerd peilingsonderzoek.

Voor u ligt het technisch rapport over het peilingsonderzoek Rekenen-Wiskunde in het VO. In dit hoofdstuk lichten we kort de achtergrond van het onderzoek verder toe (1.2), benoemen de onderzoeksvragen die eraan ten grondslag liggen (1.3) en gaan kort in op de rol die Covid-19 in het onderzoek speelde. Daarna volgt een leeswijzer voor de overige hoofdstukken van dit technisch rapport (1.5).

## 1.2 Achtergrond

Goed kunnen rekenen is essentieel voor een succesvolle carrière in het onderwijs en de maatschappij. Rekenen-Wiskunde behoort dan ook tot de kernvakken in zowel het basis- als voortgezet onderwijs. Uit PISA-onderzoek blijkt dat het wiskundeniveau van Nederlandse leerlingen over het algemeen daalt, al bleven de wiskundeprestaties tussen 2015 en 2018 gelijk (Gubbels et al., 2019). Mede op basis van deze resultaten uitte de inspectie in de Staat van het Onderwijs de afgelopen jaren herhaaldelijk haar zorgen en pleitte zij voor meer aandacht voor de basisvaardigheden in het onderwijs (Inspectie van het

---

<sup>1</sup> In het voortgezet onderwijs heet het vakgebied 'Rekenen en wiskunde', maar in navolging van Peil.Rekenen-Wiskunde in het PO (Buisman et al., 2021) wordt in dit rapport 'Rekenen-Wiskunde' gebruikt. Daarnaast wordt in verband met de leesbaarheid de vaardigheid in het vakgebied Rekenen en wiskunde in het VO aangeduid als 'rekenvaardigheid'. Overigens zijn de peilingsonderzoeken Rekenen-Wiskunde in het VO in 2022 en Begrijpend lezen in het VO in 2023 gecombineerd aanbesteed.

Onderwijs, 2018; 2020a). Een pleidooi dat in 2022 door de minister is gehonoreerd met de lancering van een Masterplan Basisvaardigheden (Wiersma, 2022).

Door de coronapandemie en de daarmee gepaard gaande scholensluitingen in de schooljaren 2019/2020, 2020/2021 en 2021/2022 zijn de leerachterstanden van (groepen) leerlingen verder toegenomen (Haelermans et al., 2021; Inspectie van het Onderwijs, 2021, 2022). Om de leer- en onderwijsontwikkeling te herstellen is in 2021 budget beschikbaar gesteld vanuit het Nationaal Programma Onderwijs (NPO; Slob, 2021). Scholen ontvangen dankzij het NPO extra budget om leerlingen en leraren te helpen om leervertragingen en andere problemen ten gevolge van corona aan te pakken.

### 1.3 Onderzoeksvragen

De centrale probleemstelling van dit peilingsonderzoek is: Welke kennis en vaardigheden hebben leerlingen uit het tweede leerjaar van het voortgezet onderwijs op de inhoudsgebieden Rekenen-Wiskunde en in hoeverre hangt dit samen met algemene en domeinspecifieke kenmerken van leerlingen, docenten, scholen en het onderwijsleerproces? Deze probleemstelling is hieronder uitgewerkt in de volgende onderzoeksvragen.

#### **Leerlingprestaties**

1. Welk percentage van de leerlingen einde leerjaar 2 van praktijkonderwijs, vmbo-b/k, vmbo-g/t en havo/vwo beheerst in welke mate de referentieniveaus voor Rekenen-Wiskunde?
2. Hoe kunnen de prestaties van leerlingen op referentieniveaus worden getypeerd naar relevante (sub)domeinen die binnen de referentieniveaus Rekenen-Wiskunde onderscheiden worden?

#### **Samenhang leerlingprestaties met kenmerken van de leerlingen en hun docenten en scholen en met factoren in het onderwijsleerproces:**

3. Welke verschillen in leerlingprestaties Rekenen-Wiskunde zijn er naar algemene en domeinspecifieke leerlingkenmerken en welk deel van de verschillen kunnen door leerlingkenmerken worden verklaard?
4. Welke verschillen in leerlingprestaties Rekenen-Wiskunde zijn er naar kenmerken van het onderwijsleerproces op het gebied van Rekenen-Wiskunde en welk deel van de verschillen in leerlingprestaties Rekenen-Wiskunde wordt verklaard door variantie in scores op kenmerken van het onderwijsleerproces (uitgevoerd curriculum)?
5. Hoe kunnen de deelnemende scholen worden getypeerd naar kenmerken van het onderwijsleerproces op het gebied van Rekenen-Wiskunde?
6. Welke verschillen in leerlingprestaties Rekenen-Wiskunde zijn er naar algemene en domeinspecifieke docentkenmerken en algemene schoolkenmerken en welk deel van de verschillen in leerlingprestaties Rekenen-Wiskunde wordt verklaard door variantie in scores op algemene en domeinspecifieke docentkenmerken, klaskenmerken en algemene schoolkenmerken?
7. Welke kenmerken van leerlingen, docenten, scholen en het onderwijsleerproces op het gebied van Rekenen-Wiskunde zijn in onderlinge samenhang mogelijke verklarende variabelen voor verschillen in leerlingprestaties?

### 1.4 Invloed Covid-19

De dataverzameling van het peilingsonderzoek Rekenen-Wiskunde in het VO vond plaats in het voorjaar van 2022. De zorgen over het coronavirus waren op dat moment minder groot dan in het schooljaar



daarvoor, al was er wel nog sprake van een verlengde kerstvakantie in 2021-2022 vanwege Covid-19. Tijdens de dataverzameling hoefde er weinig rekening met eventuele coronamaatregelen te worden gehouden. Voor de interpretatie van de resultaten is het echter wel van belang te beseffen dat de epidemie nog maar pas achter de rug was en dat eventuele achterstanden ten gevolge daarvan mogelijk nog niet waren ingehaald. Dat valt ook op te maken uit de antwoorden die docenten en schoolleiders gaven op enkele vragen over de prestaties, (het gebrek aan) motivatie en het gedrag van de leerlingen. Daarbij speelt ook dat de afgenomen toets een zogenaamde ‘*low stakes*’ toets is; er hangt voor de leerling niets van af in de zin dat een beter resultaat leidt tot een gunstiger rapportcijfer of het behalen van een diploma.

## 1.5 Leeswijzer

In Hoofdstuk 2 beschrijven we het instrumentarium, de afgenomen toets en vragenlijsten. In Hoofdstuk 3 gaan we in op het onderzoeksdesign en in Hoofdstuk 4 op de werving, dataverzameling en respons (inclusief representativiteitsanalyse). In Hoofdstuk 5 wordt de schaling van de afgenomen toets toegelicht. In Hoofdstuk 6 beschrijven we de resultaten van de vragenlijsten voor de leerlingen, docenten en schoolleiders, en lopen daarmee alvast vooruit op het beantwoorden van onderzoeksvraag 5. In Hoofdstuk 7 presenteren we de prestaties van de leerlingen op de afgenomen toets Rekenen-Wiskunde die tevens de basis vormen voor het antwoord op onderzoeksvraag 1 en 2. In Hoofdstuk 8 onderzoeken we de samenhang tussen de rekenprestaties van de leerlingen enerzijds en de algemene en domeinspecifieke leerling-, docent- en schoolkenmerken en kenmerken van het onderwijsleerproces anderzijds, waarmee ook onderzoeksvraag 3, 4 6 en 7 kunnen worden beantwoord. In het slothoofdstuk vatten we het rapport samen en beantwoorden we systematisch alle onderzoeksvragen. Bijlage 1, 2 en 3 bij dit rapport bevatten de frequentieverdelingen van alle oorspronkelijke variabelen in de vragenlijsten voor leerlingen, wiskundedocenten en schoolleiders.

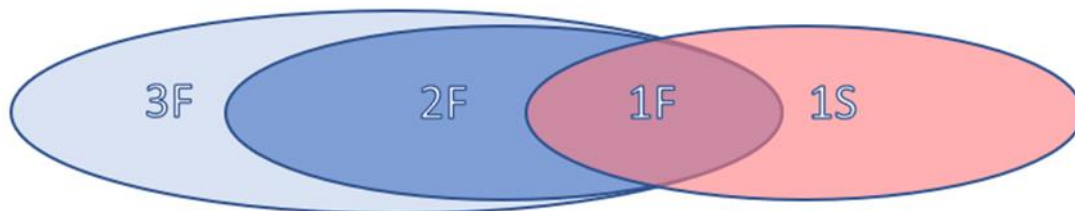
## 2 Ontwikkeld instrumentarium

### 2.1 Rekenopgaven

In deze paragraaf wordt uiteengezet wat de theoretische uitgangspunten waren voor het maken van de opgaven, en tot welke definitieve itemverzameling in het onderzoek dat heeft geleid. Deze opgaven zijn opgenomen in verschillende varianten van toetsen die bij de steekproef zijn afgenomen. De toewijzing van de opgaven aan de toetsen maakt onderdeel uit van het onderzoeksdesign en wordt zodoende beschreven in Hoofdstuk 3.

#### 2.1.1 Theoretische uitgangspunten

Voortbouwend op het peilingsonderzoek Rekenen-Wiskunde in PO is in deze eerste peiling Rekenen-Wiskunde in VO gebruik gemaakt van het Referentiekader rekenen voor 1F, 1S, 2F en 3F (Figuur 1).



Figuur 2.1. Referentiekader rekenen (Expertgroep doorlopende leerlijnen taal en rekenen, 2009).

De referentieniveaus 2S en 3S zijn buiten beschouwing gelaten. Naast de kerndoelen en tussendoelen voor Rekenen en wiskunde in het VO is dit Referentiekader tevens een van de uitgangspunten voor de Domeinbeschrijving Rekenen en wiskunde die door SLO (Sjoers & Schmidt, 2021) is opgesteld ten behoeve van dit peilingsonderzoek. Het Referentiekader beschrijft de doelen die leerlingen aan het einde van een bepaald onderwijsniveau moeten hebben bereikt. Hierbij richt de F-lijn zich op het functioneel rekenen dat noodzakelijk is voor participatie in de samenleving waarbij referentieniveau 2F gezien wordt als het minimale niveau dat mensen moeten beheersen om deel te kunnen nemen aan de maatschappij. De S-lijn richt zich op het streefniveau en kan omschreven worden als de inhoudslijn die leerlingen voorbereidt op de formele en abstracte wiskunde. Zoals in de domeinbeschrijving is genoemd zijn 1F en 1S de uitstroomniveaus van het basisonderwijs, behalve voor leerlingen die uitstromen naar het praktijkonderwijs en het speciaal onderwijs voor zeer moeilijk lerende leerlingen en meervoudig gehandicapte kinderen. Hierbij geldt dat 1F het instroomniveau is voor vmbo-basis/kader en dat 1S het gewenste instroomniveau is voor vmbo-gemengd/theoretisch, havo en vwo.

#### Verwachte prestaties leerjaar 2

Hierboven is beschreven dat de inhouden van de fundamentele- en streefniveaus andere aspecten van rekenvaardigheid meten. Desalniettemin, is op basis van onder andere Peil.Rekenen-Wiskunde in het PO de verwachting dat de opgaven behorende bij de verschillende vaardigheidsniveaus gezamenlijk één rekenvaardigheidsschaal vormen.

Op basis van het gewenste uitstroomniveau uit het PO en het beoogde uitstroomniveau aan het einde van het VO kunnen we de volgende niveaus verwachten voor de **gemiddelde leerling** in leerjaar 2, uitgesplitst naar VO-deelpopulatie:

- Vmbo-basis/kader: tussen 1F en 2F
- Vmbo-gemengd/theoretisch: 1S behaald en functioneren tussen 2F en 3F
- Havo: 1S behaald en functioneren tussen 2F en 3F
- Vwo: 1S behaald en functioneren tussen 2F en 3F

Een van de doelen van dit peilingsonderzoek is in kaart te brengen hoeveel procent van de leerlingen uit de verschillende populaties de verschillende referentieniveaus hebben behaald. In Hoofdstuk 7 zijn de resultaten op onderzoeksvraag 1 en 2 beschreven in relatie tot bovenstaande verwachtingen.

Rekenen en wiskunde in het VO is gericht op negen kerndoelen (19 tot en met 27). De peiling Rekenen-Wiskunde in het VO richt zich op de vier domeinen van het Referentiekader rekenen: Getallen, Verhoudingen, Meten & Meetkunde en Verbanden. Aanvullend wordt er in deze peiling ook naar het Wiskundig probleemoplossend<sup>2</sup> vermogen van leerlingen gekeken. Probleemoplossend vermogen maakt onderdeel uit van het beoogde curriculum zoals omschreven in kerndoel 20: "De leerling leert alleen en in samenwerking met anderen in praktische situaties wiskunde herkennen en te gebruiken om problemen op te lossen" (Ministerie van Onderwijs, Cultuur, & Wetenschap, 2006). Probleemoplossend vermogen maakt geen deel uit van het Referentiekader, maar wordt op basis van internationaal onderzoek wel aanbevolen in de domeinbeschrijving (Sjoers & Schmidt, 2021).

De meeste leerlingen hebben een rekentoets van 44 opgaven gemaakt. Een uitzondering hierop geldt de leerlingen in het praktijkonderwijs (pro), die een toets met 22 opgaven hebben gemaakt. De toetsen bestonden voor 60% uit opgaven met betrekking tot hoofdrekenen (zonder rekenmachine) en voor 40% uit opgaven waarbij leerlingen een rekenmachine mochten gebruiken. Tabel 2.1 geeft de beoogde verdeling van de opgaven over de vijf domeinen weer. Daarnaast wordt er in het Referentiekader onderscheid gemaakt tussen formele en functionele opgaven: formele opgaven bevatten alleen getallen, functionele opgaven hebben een betekenisvolle context. In deze peiling was ongeveer 10% van de opgaven formeel (hierna benoemd als 'Kaal') en 90% functioneel (hierna benoemd als 'Context'). Leerlingen mochten bij alle opgaven een liniaal en een geodriehoek gebruiken. Bij het rekenmachine(RM)-deel was een simpele rekenmachine toegestaan. Anders dan de beschrijving van de handelingsniveaus (Van Groenestijn et al., 2011; over het reken-wiskundig denken en handelen van leerlingen) doet de domeinbeschrijving geen expliciete aanbevelingen over het stimuleren van het gebruik van kladpapier tijdens het peilingsonderzoek. Hoofdrekenen is geoperationaliseerd als rekenen zonder rekenmachine waarbij leerlingen hoofdbewerkingen op papier of uit het hoofd kunnen uitvoeren (p. 65). Kladpapier is bij de toetsafname niet expliciet vernoemd als hulpmiddel, maar leerlingen maakten de toets op papier waardoor zij wel in de gelegenheid waren om berekeningen op te schrijven.

---

<sup>2</sup> In de context van deze peiling wordt met probleemoplossend vermogen 'Wiskundig probleemoplossen' bedoeld. Voor de leesbaarheid van de tekst is dit ingekort.

Tabel 2.1 – Beoogde verdeling van opgaven over de gepeilde domeinen

Domein*	Percentage van de opgaven
Getallen (B)	25 - 30%
Verhoudingen (C)	25 - 30%
Meten en meetkunde (D)	15 - 20%
Verbanden en formules (E)	15 - 20%
Probleemoplossen* (N1)	5 – 10%

\* Probleemoplossen is een vaardigheid en geen officieel domein in het Referentiekader.

De vijf domeinen zijn – in navolging van de domeinbeschrijving - aangegeven als B, C, D, E en N1. Domein B staat voor ‘Getallen’ (met subdomeinen ‘Getallen, getal-systeem en -relaties’ en ‘Rekenen met getallen’, met een grotere nadruk op dit tweede subdomein); C staat voor ‘Verhoudingen’ (met subdomeinen ‘Verhoudingen’ en ‘Procenten’; gelijk verdeeld); D voor ‘Meten en meetkunde’ (met belangrijkste subdomein ‘Rekenen in de meetkunde’ maar ook opgaven over ‘Vormen en figuren’); en E voor ‘Verbanden en formules’ (met belangrijkste subdomein ‘Grafieken, tabellen, verbanden’ maar ook met vragen over ‘Patronen en regelmaat’). Daarnaast zijn er opgaven voor ‘Wiskundig probleemoplossen’ (Kerndoel 20) en ‘Data en statistiek’, waarbij de laatste vanwege de overlap met het domein ‘Verbanden en formules’ hieronder is komen te vallen. De vaardigheid ‘Wiskundig probleemoplossen’ is aangegeven als N1. In eerdere opzetten is gesproken over de subdomeinen ‘Lineaire verbanden’ en ‘Vergelijkingen (en ongelijkheden)’, maar deze zijn in overleg niet opgenomen bij de te meten domeinen.

In aanvulling op het werken vanuit de kerndoelen Rekenen en wiskunde VO en de referentieniveaus werd in de domeinbeschrijving (Bijlage 6, Tabel 1) geadviseerd om het instrumentarium te baseren op de tussendoelen behorend bij deze in Tabel 2.1 genoemde domeinen. In aanvulling hierop is besloten om in het domein Verbanden en formules (E) ook Data en statistiek opgaven onder te brengen in verband met de grote overlap. Daarmee is het domein Data en statistiek komen te vervallen als een apart te peilen domein.

### 2.1.2 Definitieve itemverzameling in het onderzoek

Zoals in Hoofdstuk 1 al is vermeld waren er ter voorbereiding van deze peiling slechts vijf maanden beschikbaar. Hierdoor kon er geen pretest plaatsvinden en is er gebruik gemaakt van eerder ontwikkelde items uit de referentiesets Rekenen-Wiskunde, de toetsen in het leerlingvolgsysteem (LVS) van Cito en eerdere peilingsonderzoeken in het PO (zie Tabel 2.2). Het betreft grotendeels meerkeuze-items en een aantal numerieke invulvragen, waaronder opgaven die in de Peiling Rekenen-Wiskunde in het PO waren afgenomen. De items voor Probleemoplossen (N1) zijn ontwikkeld in de peiling Rekenen-Wiskunde in het PO, maar bij de pretest bleek dat deze te moeilijk waren voor het PO. Voor de items uit het PO geldt dat deze door vijf docenten zijn beoordeeld op geschiktheid voor leerlingen uit het tweede jaar van het VO.

Tabel 2.2 – Herkomst van de opgaven uitgesplitst naar referentieniveau

	1F	1S	2F	3F	Totaal
LVS	24	17			41
Peiling in PO	29	31	9	4	73
Referentiesets	12	2	98	38	150
<b>Totaal</b>	<b>65</b>	<b>50</b>	<b>107</b>	<b>42</b>	<b>264</b>

In totaal zijn er 264 opgaven in de peiling Rekenen-Wiskunde in het VO afgenomen. Deze opgaven zijn naar een aantal kenmerken op te delen. De eerste is de opdeling naar gemeten domein, zoals al aangegeven is bij de theoretische uitgangspunten. De betekenis van de domeinen B, C, D, E en N1 is gegeven in Tabel 2.1. De tweede opdeling betreft die naar de referentieniveaus. In de peiling zijn er opgaven op vier niveaus: 1F, 2F, 1S en 3F. De derde opdeling is er een op basis van de afnameconditie. Dit levert de tweedeling 'Rekenmachine' (RM) en 'Hoofdrekenen' (HR) op, die aangeeft of de opgaven met of zonder rekenmachine gemaakt mochten worden. De vierde en laatste opdeling is een tweedeling naar vorm (Kaal – Context). Deze opdeling komt er grof gezegd op neer dat functionele opgaven contextopgaven (zoals redactiesommen) zijn en formele opgaven kale opgaven.

In Tabel 2.3 is aangegeven hoe deze opgaven verdeeld zijn over drie van deze kenmerken, namelijk domein, afnameconditie en vorm. In de omschrijving van de dataverzameling (Hoofdstuk 3) komt het onderzoeksdesign van de hoofdmeting aan bod (Paragraaf 3.1). Daarin is beschreven hoe deze opgaven in de steekproef zijn afgenomen in verschillende toetsboekjes.

Tabel 2.3 – De gerealiseerde verdeling van opgaven over kenmerken

Inhoud			aantallen items					percentage items					
Conditie	Vorm	B	C	D	E	N1	Totaal	B	C	D	E	N1	Totaal
RM	Context	28	28	19	20	2	97	11%	11%	7%	8%	1%	37%
	Kaal	2	1	2	1	0	6	1%	0%	1%	0%	0%	2%
HR	Context	28	36	27	26	20	137	11%	14%	10%	10%	8%	52%
	Kaal	14	7	1	1	0	23	5%	3%	0%	0%	0%	9%
Totaal	RM	30	29	21	21	2	103	11%	11%	8%	8%	1%	39%
	HR	43*	43	28	27	20	161	16%	16%	11%	10%	8%	61%
Totaal	Context	56	64	46	46	22	234	21%	24%	17%	17%	8%	89%
	Kaal	16	8	3	2	0	29	6%	3%	1%	1%	0%	11%
Totaal	Inhoud	73	72	49	48	22	264	28%	27%	19%	18%	8%	100%

\* Dit is 1 item meer dan de som bij de opdeling Kaal-Context, omdat bij 1 opgave niet aangegeven is of dit een contextopgave is

De eigenschappen die in Tabel 2.3 worden onderscheiden, zijn min of meer gelijk over de gehele steekproef afgenomen. Dat is niet het geval bij de opgaven voor de verschillende referentieniveaus, omdat in het VO niet te verwachten is dat deze alle vier voor alle leerlingen even relevant zijn. Zo is uit gegevens in het basisonderwijs bekend welk percentage leerlingen naar verwachting de referentieniveaus 1F en 1S beheerst: op *high stakes* toetsen haalt ongeveer 90% van de leerlingen in groep 8 met een schooladvies vmbo-KB/GT het 1F-niveau, bij leerlingen met schooladvies havo en hoger is dit bijna

100%. Het heeft bij dergelijke populaties daarom weinig zin om opgaven op het referentieniveau 1F op te nemen.

In de Domeinbeschrijving Rekenen en wiskunde in de onderbouw van VO, die verscheen na de *call for proposals* van deze peiling (Sjoers & Schmidt, 2021), wordt ook geadviseerd deze vaardigheden te meten (Figuur 2.2).

Thema	Te peilen vaardigheden
Referentieniveaus	praktijkonderwijs: 1F, 2F vmbo-bb: 1F, 2F vmbo-kb: 1F, 2F vmbo-gt: 1F, 1S, 2F havo: 1S, 2F, 3F vwo: 1S, 2F, 3F

Figuur 2.2 – Deel van Tabel 19 in de Domeinbeschrijving van SLO (Sjoers & Schmidt, 2021).

Gegeven bovenstaande bevindingen is het duidelijk dat bij pro-leerlingen geen 1S- en 3F-items afgenomen hoeven te worden en dat het aantal 2F-items beperkt kan zijn. Bij leerlingen in het vwo kan het aantal aangeboden 2F- en 1S-items beperkt zijn.

In Figuur 2.2 is te zien dat de 2F-opgaven voor alle groepen relevant zijn. In onze itemverzameling is dat met 107 (40,5% van 264) ook de grootste groep. Verder zijn er 65 1F-opgaven (24,6%), 50 opgaven op het 1S-niveau (18,9%) en 42 (15,9%) 3F-opgaven. Hoe deze over de verschillende toetsen en de steekproef verdeeld zijn, is beschreven bij het design van de hoofdmeting in Hoofdstuk 3.

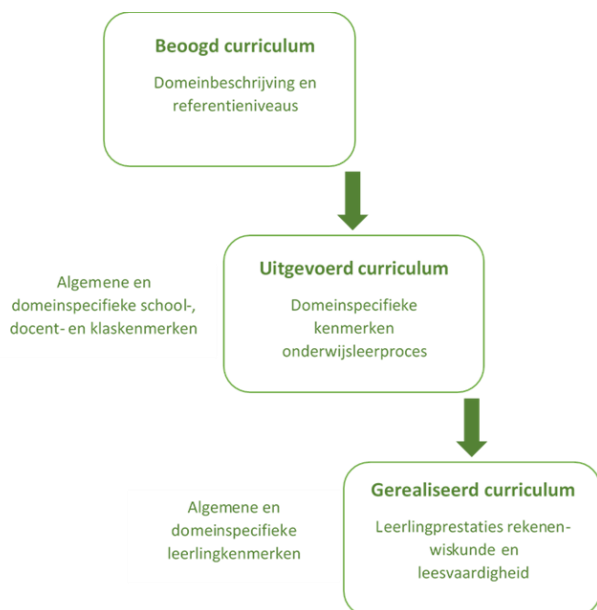
## 2.2 Leerling, docent- en schoolvragenlijsten

De huidige peilingsonderzoeken in het voortgezet onderwijs hebben als doel een betrouwbaar en valide beeld te geven van de vaardigheden van leerlingen in een bepaald domein of vakgebied in het tweede leerjaar van het voortgezet onderwijs. Daarnaast dienen zij inzicht te geven in de kenmerken die mogelijk samenhangen met het vaardigheidsniveau van deze leerlingen. Het gaat dan om algemene en domeinspecifieke kenmerken van leerlingen, docenten/klassen en scholen, en om domeinspecifieke kenmerken van het onderwijsleerproces.

De informatie over de algemene en domeinspecifieke kenmerken is in dit peilingsonderzoek verzameld door middel van vier schriftelijke vragenlijsten met overwegend gesloten vragen:

- een online schoolvragenlijst bestemd voor de schoolleider van de betreffende school/vestiging;
- een online docentvragenlijst bestemd voor de wiskundeleraar van de getoetste klas;
- een op papier afgenomen leerlingvragenlijst bestemd voor de leerlingen in de getoetste klassen die onderwijs volgen op het niveau van vmbo, havo of vwo of een combinatie hiervan;
- een verkorte, op papier afgenomen leerlingvragenlijst bestemd voor de leerlingen in de getoetste klassen die praktijkonderwijs volgen.

Voor de ontwikkeling van de vragenlijsten is uitgegaan van het globale conceptueel model (Figuur 2.3) dat ook als uitgangspunt wordt genomen in internationale peilingsonderzoeken (Mullis & Martin, 2017).



*Figuur 2.3 – Conceptueel model peilingsonderzoek Rekenen-Wiskunde (en Begrijpend lezen, in 2023) in het VO (gebaseerd op Mullis & Martin, 2017).*

Daarnaast zijn ook de onderzoeksvragen (zie Hoofdstuk 1) met de daarin genoemde kenmerken als uitgangspunt genomen. De keuze van met name de domeinspecifieke kenmerken is voornamelijk gebaseerd op de adviezen voortkomend uit de reviewstudies van Hickendorff et al. (2017; 2020) over effectieve rekeninstructie. De items en schalen die deze kenmerken representeren, zijn vooral afkomstig uit de vragenlijsten van Peil.Rekenen-Wiskunde PO in 2019 en de internationale onderwijsstudies TIMSS<sup>3</sup> en PISA<sup>4</sup>. Het gebruik van al bestaande en geteste schalen en vragenlijstitems was voor deze studie van belang omdat het niet mogelijk was om voor deze peiling een pretest uit te voeren. Daarnaast is er voor de school- en docentenvragenlijsten een nieuwe vraag ontwikkeld over de gevolgen van de coronapandemie. Het gaat bij deze vraag om de percepties van de docenten en schoolleiders over de gevolgen van de pandemie voor het leer- en leerlinggedrag van hun leerlingen.

Tabel 2.4, 2.5 en 2.6 geven een overzicht van de kenmerken die in de drie vragenlijsten zijn gemeten. De kenmerken die gemeten zijn aan de hand van een set van items (bijvoorbeeld verschillende stellingen over plezier in wiskunde) zijn cursief weergegeven. In de vragenlijsten voor de docent en de schoolleider gaan de domeinspecifieke vragen over 'wiskunde', nadat in de introductie is toegelicht dat daarmee het gehele vakgebied Rekenen en Wiskunde wordt bedoeld. Ook in de vragenlijst voor leerlingen in regulier VO wordt gesproken over 'wiskunde', in de verkorte vragenlijst voor pro-leerlingen gaan de vragen over 'rekenen'.

<sup>3</sup> IEA's Trends in International Mathematics and Science Study – TIMSS 2019 Copyright © 2020 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College, <https://timssandpirls.bc.edu/timss2019/questionnaires/index.html>  
IEA's Trends in International Mathematics and Science Study – TIMSS 2015 Copyright © 2016 International Association for the Evaluation of Educational Achievement (IEA). Publisher: TIMSS & PIRLS International Study Center, Lynch School of Education and Human Development, Boston College, <https://timssandpirls.bc.edu/timss2015/questionnaires/index.html>

<sup>4</sup> OECD, 2017, <https://www.oecd.org/pisa/data/2018database>

Tabel 2.4 – Vragenlijst leerling, Peiling Rekenen-Wiskunde in het VO (2022)

Categorie	Kenmerk	Gemeten als:
Algemene leerlingkenmerken	Geslacht	Geslacht
	Leeftijd	Geboortejaar en maand
	SES	Aantal boeken thuis Hoogste opleiding moeder Hoogste opleiding vader
	Thuis taal	Meest gesproken taal thuis Frequentie thuis Nederlands gesproken
	Migratieachtergrond	Moeder geboren in Nederland Vader geboren in Nederland Zelf geboren in Nederland
	Onderwijstype	Onderwijstype leerjaar 3*
	Verwacht onderwijsniveau	Verwacht onderwijsniveau met diploma*
Domeinspecifieke leerlingkenmerken	Attitudes	<i>Plezier in wiskunde**</i> <i>Zelfvertrouwen in wiskunde</i> <i>Belang van wiskunde*</i>
	Attributies positieve prestaties	<i>Attributies positieve wiskunde prestaties</i>
	Mindset t.o.v. Rekenen-Wiskunde	<i>Fixed/Growth mindset t.o.v. wiskunde</i>
Domeinspecifieke kenmerken onderwijsleerproces	Ordelijkheid les	<i>Ordelijkheid wiskundeles</i>
	Helderheid instructie	<i>Helderheid instructie wiskundeles*</i>
	Oplossingsstrategieën	<i>Oplossingsstrategieën in de wiskundeles*</i>
	Directe instructie	<i>Directe instructie in de wiskundeles*</i>

\* Niet opgenomen in vragenlijsten voor pro-leerlingen

\*\* De kenmerken die gemeten zijn aan de hand van een set van items zijn cursief weergegeven.

Tabel 2.5 – Vragenlijst schoolleider, Peiling Rekenen-Wiskunde in het VO (2022)

Categorie	Kenmerk	Gemeten als:
Algemene schoolkenmerken	Schoolgrootte	Vestigingsgrootte
	Onderwijsaanbod	Opleidingsaanbod vestiging
	Denominatie	DUO, Onderwijsconcept (wel/niet specifiek concept)
	Kenmerken leerlingpopulatie	Komt uit een economisch achterstandsgezin Komt uit een economisch welgesteld gezin Aandeel leerlingen Nederlands eerste taal
	Urbanisatiegraad	Directe omgeving school (stedelijkheid)
	Prestatiegerichtheid school	<i>Prestatiegericht schoolklimaat_docenten*</i> <i>Prestatiegericht schoolklimaat_ouders</i> <i>Prestatiegericht schoolklimaat_leerlingen</i>
	Leerlingvolgsysteem	Gebruik leerlingvolgsysteem <i>Belang leerlingvolgsysteem</i>
Domeinspecifieke schoolkenmerken	Buitencurriculaire activiteiten	Aanbod extra wiskundelessen buiten schooltijd (wel/niet en voor wie) Gevolg pandemie voor aanbod Extra activiteiten voor leerlingen in tweede leerjaar <i>Bevordering interesse in wiskunde</i>
	Stimuleren nascholing	Aanbod professionele ontwikkeling wiskundedocenten (het aantal gebieden)
Coronapandemie	Gepercipieerde veranderingen door de pandemie	<i>Verandering in leerprestaties</i> <i>Verandering in leergedrag</i> <i>Verandering in leerlinggedrag</i>

\* De kenmerken die gemeten zijn aan de hand van een set van items zijn cursief weergegeven.



Tabel 2.6 – Vragenlijst wiskundedocent, Peiling Rekenen-Wiskunde in het VO (2022)

Categorie	Kenmerk	Gemeten als:
Domeinspecifieke kenmerken onderwijsleerproces	Tijdsbesteding domeinen Rekenen-Wiskunde	Aantal minuten wiskunde per week Tijdsbesteding wiskundedomeinen
	Leermethode en aanvullend materiaal (incl. ICT)	Hoofdzakelijk gebruikte reken- en wiskundemethode Gebruik digitale apparaten <i>Intensiteit ICT-gebruik*</i> Gebruik rekenmachine tijdens de wiskundeles
	Geschiktheid methode	<i>Geschiktheid wiskundemethode</i>
	Werkvormen	<i>Leerkrachtgestuurde klassikale lesactiviteiten</i> <i>Probleemoplossen</i> Overige lesactiviteiten
	Differentiatie	<i>Differentiatie</i>
	Omvang en bespreking huiswerk	Frequentie huiswerk voor wiskunde Hoeveelheid huiswerk per keer Feedback huiswerk
	Vormen en doelen van evaluatie prestaties en toetsing	<i>Intensiteit evaluatie prestaties en leerproces</i> Belang van beoordelingsstrategieën Gebruik typen toetsen
	Directe instructie	Wel/niet Directe instructiemodel Wel/niet Ontdekkend of onderzoekend leren
	Oplossingsstrategieën	Wel/niet alle leerlingen meerdere oplossingsstrategieën aanbieden
Algemene docentkenmerken	Geslacht	Geslacht
	Leeftijd	Leeftijd
	Opleiding en bevoegdheid	Hoogste opleiding met diploma afgerond Lesbevoegdheid
	Leservaring	In totaal lesgegeven Lesgegeven in wiskunde
Algemene klaskenmerken	Klassamenstelling naar onderwijstype en klassengroote	Onderwijstype Aantal leerlingen getoetste klas
Domeinspecifieke docentkenmerken	Professionaliseringsactiviteiten	Gevolgde en gewenste (bij)scholing Tijdsbesteding (bij)scholing
	Zelfvertrouwen in vakinhoudelijke en vakdidactische vaardigheden	<i>Zelfvertrouwen wiskundeinstructie</i>
	<i>Mindset t.o.v. Rekenen-Wiskunde</i>	<i>Fixed/Growth mindset t.o.v. wiskunde</i>
Coronapandemie	Gepercipieerde veranderingen door de pandemie	<i>Verandering in leerprestaties</i> <i>Verandering in leergedrag</i> <i>Verandering in leerlinggedrag</i>

\* De kenmerken die gemeten zijn aan de hand van een set van items zijn cursief weergegeven.



## 3 Onderzoeksdesign

### 3.1 Inleiding

Zoals eerder aangegeven was het niet mogelijk een pretest uit te voeren voor het peilingsonderzoek Rekenen-Wiskunde in het VO. Het design van de steekproef en toetsafname dat we in dit hoofdstuk beschrijven heeft dus uitsluitend betrekking op de hoofdmeting in 2022.

### 3.2 Steekproefdesign

Het peilingsonderzoek Rekenen-Wiskunde in het VO werd uitgevoerd bij klassen in het tweede leerjaar van een landelijk representatieve steekproef van vestigingen van middelbare scholen. In overleg met de Inspectie is er gekozen voor een getrapte steekproef waarbij eerst een steekproef van schoolvestigingen werd getrokken door de Inspectie, waarna in tweede instantie door het consortium één tweede klas werd geselecteerd die wiskunde volgde binnen de getrokken en geworven vestiging.

#### 3.2.1 Eerste trap: gewogen steekproef van vestingen

De doelpopulatie bestond uit leerlingen in leerjaar twee in het bekostigd voortgezet onderwijs per 1 oktober 2020. Uitgesloten waren leerlingen in het niet-bekostigde onderwijs en leerlingen die een internationaal curriculum (IB-opleiding of Engelse Stroom) volgden. Ook uitgesloten waren buitenlandse scholen, vestigingen die waren opgeheven of genomineerd waren om opgeheven te worden in schooljaar 2021/22, vestigingen met minder dan vijf leerlingen en vestigingen met alleen internationale schakelklassen/eerste opvang anderstaligen.

In de eerste trap van het steekproefkader heeft de Inspectie een steekproef getrokken uit alle vestigingen (combinatie van brin- en vestigingsnummer) met leerlingen uit deze doelpopulatie. In schooljaar 2020/21 waren dit in totaal 1288 vestigingen. Impliciet criterium bij de trekking was het aantal leerlingen in leerjaar 2 op de vestiging: scholen hadden een grotere kans in de steekproef terecht te komen naarmate het aantal leerlingen in leerjaar 2 groter was (gewogen steekproeftrekking). Expliciet stratificatiecriterium was het aanbod van onderwijssoorten in de brugklas (waar leerjaar 2 onder valt) op de vestigingen. Daarbij werden negen strata onderscheiden: 'Pro' (alleen praktijkonderwijs), 'Pro plus' (praktijkonderwijs en andere afdeling(en)), 'Vmbo' (vmbo-b, -k en/of -gt), 'Vmbo-gt', 'Vmbo-gt en havo', 'Vmbo-gt en havo en/of vwo', 'Breed' (vmbo-b en/of -k en -gt en havo en/of vwo, 'Havo-vwo' en 'Vwo'. Per stratum is een aantal vestigingen getrokken, proportioneel naar het aantal leerlingen dat een stratum omvatte ten opzichte van het totaal aantal leerlingen (zie Tabel 3.1). Er zijn drie tranches van ieder 175 vestigingen getrokken. Elke vestiging in de eerste tranche (de hoofdsteekproef) had dus een qua stratificatiecriteria (onderwijsaanbod en aantal leerlingen in leerjaar 2) gelijkwaardige vervanger in de tweede en in de derde tranche. Bij het trekken van deze steekproef is zoveel mogelijk rekening gehouden met de steekproeven van de internationale PISA- en ICCS-onderzoeken die ongeveer gelijktijdig plaatsvonden in het Nederlandse voortgezet onderwijs.

Tabel 3.1 – Stratificatietabel en omvang steekproef van schoolvestigingen

Vestigingsstrata	% leerlingen in populatie	n steekproef per tranche	n steekproef totaal
Pro	2,9	5	15
Pro plus	2,9	5	15
Vmbo	19,3	34	102
Vmbo-gt	2,9	5	15
Vmbo-gt en havo	2,7	5	15
Vmbo-gt en havo en/of vwo	27,4	48	144
Breed	22,0	38	114
Havo/vwo	16,4	29	87
Vwo	3,5	6	18
Totaal leerlingen (populatie)	186.746		
Totaal vestigingen	1.288	175	525

### 3.2.2 Tweede trap: steekproef van één klas per vestiging

Een getrapte procedure was noodzakelijk omdat de landelijke registerdata onvoldoende betrouwbaar zijn wat betreft de aangeboden onderwijstypen in de brugklassen (waartoe leerjaar 2 behoort) en het aantal leerlingen daarbinnen. De tweede trap in de steekproeftrekking kon daarom ook pas plaatsvinden na de fase van de scholenwerving (zie Hoofdstuk 4). Elke schoolvestiging die in de wervingsfase deelname aan het peilingsonderzoek toezegde, kreeg het verzoek om op te geven hoeveel tweede klassen er op de vestiging waren, welke schoolsoort/schoolsoorten er in deze klassen werd/werden aangeboden en hoeveel leerlingen er in de klassen zaten. Uit het totale aantal tweede klassen per vestiging selecteerde het consortium vervolgens random één klas, met als doel een representatieve groep tweede klassen te verkrijgen en daar het onderzoek uit te voeren.

### 3.3 Afnamedesign rekenopgaven

Zoals genoemd in Paragraaf 2.1.2 zijn er 264 items gebruikt om de leerlingprestaties Rekenen-Wiskunde te meten. Het uitgangspunt voor het design was dat leerlingen in praktijkonderwijs maximaal ongeveer 22 items kunnen maken en leerlingen in de overige onderwijssoorten maximaal ongeveer 44 opgaven. Met deze aantallen hebben de leerlingen voldoende tijd om de toetsen te maken (en daarnaast de leerlingvragenlijst in te vullen) en wordt het aantal ontbrekende waarden door tijdsgebrek geminimaliseerd. Op basis van deze uitgangspunten zijn met de 264 opgaven verschillende toetsen samengesteld.

Deze toetsen zijn toegewezen aan leerlingen op basis van hun verwachte rekenniveau zoals beschreven in Paragraaf 2.1.2, rekening houdend met de moeilijkheid van opgaven. Het design is dus zo opgezet dat de (veronderstelde) moeilijkheid van de opgaven varieert voor de verschillen groepen leerlingen. Door de items op te delen in itemsets (is01, is02, etc.) die aan verschillende groepen leerlingen worden voorgelegd (zie Tabel 3.2a) kunnen leerlingen door middel van analyses waarin gebruik gemaakt van Item Respons Theorie (IRT) alsnog worden vergeleken.

Bij het opdelen van de opgaven in itemsets moesten we ook rekening houden met een belangrijke afnameconditie die de opgaven typeert, namelijk het al dan niet mogen gebruiken van een rekenmachine. Deze twee typen opgaven mogen niet door elkaar heen aangeboden worden, zodat wordt voorkomen dat de leerling toegang heeft tot een rekenmachine bij opgaven waar dat niet mag. Er waren dus

twee parallele designs nodig, waarbij de 103 opgaven met rekenmachine (RM-items) en de 161 opgaven zonder rekenmachine (HR-items; HR staat voor hoofdrekenen) apart in 12 itemsets opgedeeld zijn. Elke toets (T01, T02, etc.) werd vervolgens samengesteld uit twee itemsets per afnameconditie.

In Tabel 3.2a worden de 12 RM-itemsets (RMis01 – RMis12), en de 12 HR-itemsets (HRis01 – HRis12) beschreven. De RM-itemsets bestaan uit 8 of 9 items, de HR-itemsets uit 13 of 14 items. Op deze manier zijn RM-toets 01 tot en met 12 (RM01-RM12) samengesteld. Hetzelfde principe is toegepast bij HR, hetgeen HR-toets 01 tot en met 12 (HR01-HR12) oplevert. Deze twee designs kunnen in elkaar geschoven worden zodat voor de analyses gewerkt kan worden met een twaalftal toetsen (T01-T12; Tabel 3.2b).

Tabel 3.2a – Afnamedesign rekentoetsen

Test	N/test	RM												HR												N items	
		is01	is02	is03	is04	is05	is06	is07	is08	is09	is10	is11	is12	is01	is02	is03	is04	is05	is06	is07	is08	is09	is10	is11	is12	RM	HR
T01	80																								9	13	
T02	168																									17	27
T03	149																									17	27
T04	155																									18	26
T05	164																									17	27
T06	174																									17	27
T07	189																									17	27
T08	186																									17	27
T09	262																									17	27
T10	437																									17	27
T11	293																									17	27
T12	467																									17	27
Totaal	2724	248	317	304	319	338	363	375	448	699	730	760	467	248	317	304	319	338	363	375	448	699	730	760	467	2724	2724
N items		9	8	9	9	8	9	8	9	8	9	8	9	13	14	13	13	14	13	14	13	14	13	14	13	103	161

Tabel 3.2b – Analysedesign – uitgangspunt (met itemsets oplopend in veronderstelde moeilijkheid)

Test	N/test	is01		is02		is03		is04		is05		is06		is07		is08		is09		is10		is11		is12		N items
		RM	HR	RM	HR	RM	HR	RM	HR	RM	HR	RM	HR	RM	HR	RM	HR	RM	HR	RM	HR	RM	HR	RM	HR	
T01	80																									22
T02	168																									44
T03	149																									44
T04	155																									44
T05	164																									44
T06	174																									44
T07	189																									44
T08	186																									44
T09	262																									44
T10	437																									44
T11	293																									44
T12	467																									44
Totaal	2724	248	317	304	319	338	363	375	448	699	730	760	467	248	317	304	319	338	363	375	448	699	730	760	467	2724
N items		22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	22	264

Als leerlingen een in Tabel 3.2b gegeven combinatie van itemsets maken, hebben ze de toetsen ‘volgens design’ gemaakt. Het aantal verschillende toetsen waar volgens design vanuit gegaan wordt is dus 12 (T01-T12). De leerlingen die T01 maakten (leerlingen in praktijkonderwijs) kregen twee itemsets voorgelegd (een RM-itemset en een HR-itemset, met in totaal 22 items), de leerlingen die een van de andere toetsen maakten (leerlingen in alle overige onderwijssoorten) kregen ieder 4 itemsets voorgelegd (twee RM-sets en twee HR-sets, met in totaal 44 opgaven). Bij T02-T12 zou dit ook gedefinieerd

kunnen worden als een toets die bestaat uit twee itemsets van 22 items. We benoemen dit echter hier als 4 itemsets, omdat we bij de evaluatie van de ontbrekende waarden dit ook specifiek voor ieder van de vier itemsets afzonderlijk doen. De in Tabel 3.2 vermelde aantallen leerlingen per toets in het design houden geen rekening met deze niet afgenomen delen. Bij de analyses van de ontbrekende waarden in de aangeleverde data (zie Hoofdstuk 5) wordt hier verder op ingegaan, want afhankelijk van wanneer een itemset als 'niet gemaakt' beschouwd wordt, verandert deze verdeling.

Zoals aangegeven, was het onze doelstelling om toetsen te maken die voor de verschillende vaardigheidsniveaus geschikt zijn. De opgaven wilden we zo ook ordenen naar moeilijkheid. Deze opdeling van de opgaven naar moeilijkheid heeft ook een relatie met de inschatting van de opgaven op basis van de referentieniveaus, zoals beschreven is in Hoofdstuk 2 (itemverzameling). De verdeling van de opgaven op basis van referentieniveau is zodoende niet gelijkmatig verdeeld over de itemsets. Opgaven van niveau 1F komen alleen voor in itemsets is01 tot en met is07, waarbij in de eerste drie itemsets minstens de helft van de opgaven van niveau 1F is. Opgaven van niveau 1S komen alleen voor in itemsets is06 tot en met is12, waarbij in is08-is09 minstens de helft 1S-items betreft. De 3F-items komen alleen voor in drie itemsets, namelijk is10, is11 en is12, waarbij minstens de helft van de opgaven niveau 3F betreft. De 2F-opgaven komen in alle itemsets voor, waarbij in is03 tot en met is08 minstens de helft van de opgaven 2F betreft. De exacte verdeling is gegeven in Tabel 3.3.

*Tabel 3.3 – Aantal items (en percentage) per itemset, per referentieniveau*

Itemset	Beoogd niveau	1F (% is)	2F (% is)	1S (% is)	3F (% is)	totaal
is 01	pro	16 (73%)	6 (27%)	0	0	22
is 02	BB-KB	13 (59%)	9 (41%)	0	0	22
is 03	BB-KB	11 (50%)	11 (50%)	0	0	22
is 04	BB-KB/GT	9 (41%)	13 (59%)	0	0	22
is 05	BB-KB/GT	6 (27%)	16 (73%)	0	0	22
is 06	GT	6 (27%)	11 (50%)	5 (23%)	0	22
is 07	GT	4 (18%)	11 (50%)	7 (32%)	0	22
is 08	GT/HV	0	11 (50%)	11 (50%)	0	22
is 09	GT/HV	0	7 (32%)	15 (68%)	0	22
is 10	HV	0	5 (23%)	5 (23%)	12 (55%)	22
is 11	HV	0	4 (18%)	4 (18%)	14 (64%)	22
is 12	HV	0	3 (14%)	3 (14%)	16 (73%)	22
<b>Totaal aantal</b>		<b>65</b>	<b>107</b>	<b>50</b>	<b>42</b>	<b>264</b>

De verdeling naar de andere drie kenmerken (domein, afnameconditie, vorm) is zo gelijk mogelijk gedaan per itemset, en volgt zo veel mogelijk de percentages van Tabel 2.2. Zo kon per toets een zo evenwichtig mogelijke verdeling naar inhoud gegeven worden. De verdeling van de vier typen opgaven over de toetsversies is gegeven in Tabel 3.4. In de tabel verwijzen deze letters naar de vijf domeinen:

- B. Getallen
- C. Verhoudingen
- D. Meten en meetkunde
- E. Verbanden en formules
- N1. Wiskundig probleemoplossen

Tabel 3.4 – Verdeling van de opgaven naar type, per toetsversie (toetsboekje)

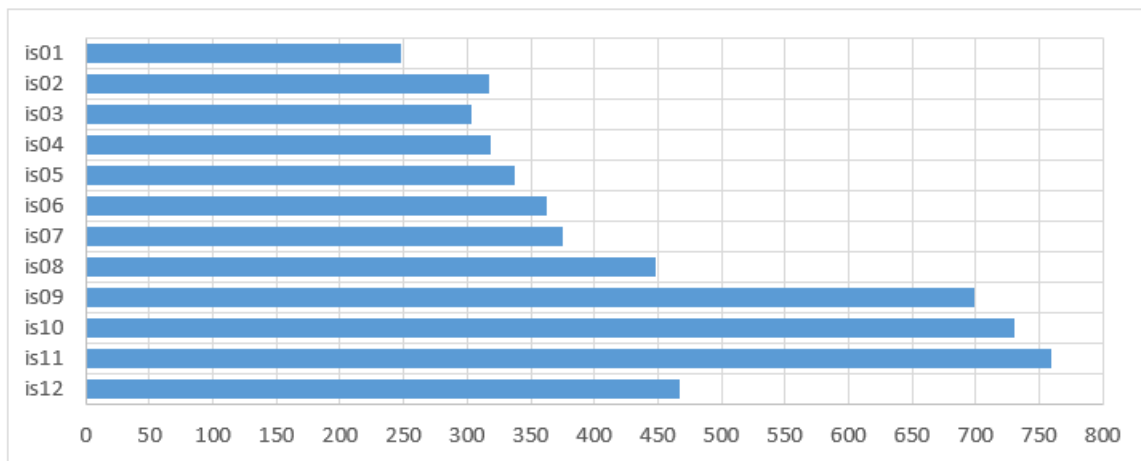
Aantal items per toets		T01	T02	T03	T04	T05	T06	T07	T08	T09	T10	T11	T12
<b>Totaal</b>		<b>22</b>	<b>44</b>	<b>44</b>	<b>44</b>	<b>44</b>	<b>44</b>	<b>44</b>	<b>44</b>	<b>44</b>	<b>44</b>	<b>44</b>	<b>44</b>
Afname- conditie	RM	9	17	17	18	17	17	17	17	17	17	17	17
	HR	13	27	27	26	27	27	27	27	27	27	27	27
Vaardigheid	B	7	13	12	13	13	11	11	12	12	12	12	12
	C	6	11	11	13	13	12	12	12	12	12	12	12
	D	4	9	9	7	7	8	10	10	8	8	9	7
	E	4	8	8	7	7	8	8	8	8	8	8	9
	N1	1	3	4	4	4	5	3	2	4	4	3	4
Referentie- niveau	1F	16	29	24	20	15	12	10	4				
	2F	6	15	20	24	29	27	22	22	18	12	9	7
	1S						5	12	18	26	20	9	7
	3F										12	26	30
Vorm	Functioneel	20	39	40	40	39	39	39	41	41	40	38	35
	Formeel	2	5	4	4	5	5	5	2	2	4	6	9
	—								1	1			

Slechts in een enkel geval had een toets een opgave meer of minder kunnen hebben om de percentages van Tabel 3.3 beter te benaderen. De grootste afwijking die gevonden is, betreft toetsen T08 en T09 die iets meer functionele opgaven (93%) hebben en toets T12 die juist iets minder functionele opgaven heeft (80%). De afwijking is zeer beperkt. Deze verdeling zorgt ervoor dat een leerling bij elk kenmerk van de taakuitvoering een relevant aantal opgaven maakt. Bij de inhoudskenmerken Probleemoplossen (N1) en Formeel is dit door het lage aantal opgaven dat totaal is ingezet (22 N1, 29 Formeel) soms lager dan vier, maar bij de meeste kenmerken zijn het er meer; ook bij toets T01 die uit slechts 22 opgaven bestaat. Hierdoor is het mogelijk om op stelselniveau een uitspraak te doen over de beheersing van de verschillende kenmerken van de taakuitvoering door de leerlingen als dat nodig is. De afnameconditie met of zonder rekenmachine kan opgevat worden als een itemkenmerk aangezien de psychometrische eigenschappen van het item, en de te meten vaardigheid, hierdoor in de meeste gevallen (bij de meeste leerlingen, en bij de meeste opgaven) beïnvloed worden. De opdeling van opgaven naar vorm (formeel/functioneel) is eerder besproken.

Zoals aangegeven zijn de opgaven ingedeeld naar vermeende oplopende moeilijkheid, waarbij een heel strikte ordening niet noodzakelijk is. Het doel was vooral dat de leerlingen opgaven voorgelegd krijgen op een min of meer passend niveau. Doordat in dit design iedere leerling opgaven aangeboden krijgt op het best passende niveau, was het goed mogelijk de hele range van vaardigheden – van pro tot en met havo/vwo – te meten. Het verschil tussen die twee groepen is aan het einde van groep 8 al ongeveer vier standaarddeviaties op de IRT-vaardigheidsschaal, dus er valt een behoorlijke spreiding in vaardigheid te verwachten tussen de leerlingen die de verschillende boekjes maken. Door deze werkwijze was het mogelijk een grote hoeveelheid opgaven met een behoorlijke spreiding in moeilijkheden in te zetten om zo bij al die groepen goed te kunnen meten. Deze werkwijze bevordert de betrouwbaarheid van de meting.

Een mogelijke fout in de ordening van opgaven, zeker op relatief detailniveau, is overigens niet onoverkomelijk. Het kan en mag dus voorkomen dat een relatief makkelijke opgave in een wat hoger genummerde toets terecht komt, en een iets moeilijkere opgave in een lager genummerde toets. Dat is geen probleem zolang dat niet stelselmatig gebeurt en deze mismatch er maar niet toe leidt dat (vrijwel) alle opgaven goed of fout beantwoord zijn door de gehele set leerlingen die de test maakt. In deze peiling was dit niet aan de orde.

De toewijzing van de aantallen leerlingen (N) per test is belangrijk voor het realiseren van voldoende waarnemingen per opgave zodat de meting zo betrouwbaar mogelijk is. In dit design is het uitgangspunt dat iedere opgave door minimaal 300 leerlingen gemaakt wordt zodat de itemkenmerken voldoende goed in beeld gebracht kunnen worden. In Tabel 3.2 is te zien dat dit is gelukt, op de items in itemset is01 na (de gemakkelijkste opgaven bij RM en bij HR; maximaal 248 observaties per opgave). Bij itemsets is02 tot en met is07 (ook weer zonder onderscheid te maken tussen RM en HR) ligt het aantal observaties tussen de 300 en 400. Bij de overige itemsets (is08 tot en met is12) is het aantal waarnemingen boven de 400, waarbij is09 tot en met is11 door (ruim) 700 leerlingen is gemaakt. Gemiddeld is ieder item aan 447 leerlingen voorgelegd, met een mediaan van 369 observaties per item. In Figuur 3.1 is het overzicht gegeven van het maximale aantal observaties per itemset en betreft zodoende de maximumaantallen van alle items binnen de itemset. Bij de selectie van kandidaten wordt een verder onderscheid gemaakt tussen de RM-itemsets en de HR-itemsets. Bij de beschrijving van de analyses (Hoofdstuk 5) wordt dieper ingegaan op deze aantallen.



Figuur 3.1 – Maximum aantal te gebruiken observaties (leerlingen) per itemset



## 4 Werving, dataverzameling en respons

### 4.1 Organisatie scholenwerving

De werving van scholen is uitgevoerd door onderzoeksassistenten, begeleid door onderzoekers van het consortium. Deze assistenten ontvingen voorafgaand aan hun werkzaamheden een training, waarin het doel en belang van het onderzoek werden toegelicht en praktische zaken (zoals het belscript) gezamenlijk werden doorgenomen. Voor de werving was relatief weinig tijd beschikbaar, doordat er slechts enkele maanden zaten tussen de start van de opdracht en de voorgeschreven dataverzamelingsperiode (zie ook Hoofdstuk 1). Eind 2021 werd het onderzoek door middel van een informatiebrief aangekondigd bij alle schoolbesturen met scholen in het voorgezet onderwijs. In deze brief aan het bestuur werd de opzet, het belang en de opbrengsten van het onderzoek beschreven. Schoolleiders van vestigingen uit de hoofdstreekproef (eerste tranche) ontvingen in diezelfde periode een uitnodigingsbrief uit naam van de Inspecteur-generaal van het Onderwijs met daarin het verzoek om deel te nemen aan het onderzoek. In deze brief werd naast informatie over het doel en het belang van het onderzoek ook ingegaan op de inzet die van de schoolleider, wiskundedocent en leerlingen gevraagd werd en de manier waarop anonimiteit gegarandeerd werd. Om deelname aan het onderzoek zo aantrekkelijk mogelijk te maken, werd uitgelegd dat de onderzoekers en onderzoeksassistenten de voorbereidingen en de toetsafname zoveel mogelijk voor hun rekening zouden nemen en dat er voorzien was in een terugkoppeling van resultaten (het schoolrapport) en in een klein presentje voor de deelnemende respondenten. Enkele dagen na het versturen van de brieven nam een onderzoeksassistent telefonisch contact op met de schoolleider. In dit telefoongesprek werd aan de hand van een belscript meer informatie gegeven over het onderzoek en werd de school gevraagd om deel te nemen. Schoolleiders konden in het telefoongesprek alle vragen stellen die ze hadden met betrekking tot het onderzoek.

Scholen die toezegden deel te nemen aan het onderzoek kregen het verzoek hun aanmelding te bevestigen via een digitaal aanmeldformulier, waarin ze ook een schoolcontactpersoon aanwezen en administratieve gegevens verstrekten zoals het aantal tweede klassen op hun vestiging, en per klas het aangeboden onderwijstype en aantal leerlingen. Nadat op basis van deze informatie de tweede klas was geselecteerd die aan het onderzoek zou deelnemen (zie ook Paragraaf 3.2.2), informeerde een onderzoeksassistent de schoolcontactpersoon hierover en voorzag deze tevens van informatiebrieven voor ouders en leerlingen. In de informatiebrief voor ouders werd het onderzoek toegelicht, waarbij onder andere aandacht was voor de (anonieme) manier waarop gegevens verwerkt worden. Ook werd uitgelegd dat ouders die geen toestemming gaven voor deelname van hun kind aan het peilingsonderzoek, dat aan de school door konden geven. Deze leerlingen werden uitgesloten van deelname aan het onderzoek. In de informatiebrief voor leerlingen werd de opzet van het onderzoek toegelicht en werd de leerlingen gevraagd een liniaal en rekenmachine mee te nemen op de dag van het onderzoek.

### 4.2 Respons van scholen

De hiervoor beschreven wervingsprocedure leverde een respons op van 138 schoolvestigingen (78% van 175), waarvan 98 uit de hoofdstreekproef (56%). Tabel 4.1 laat hun verdeling over de negen strata zien. In de tabel is ook het totale aantal leerlingen in leerjaar 2 op deze scholen opgenomen, volgens hun eigen opgave.

Tabel 4.1 – Geworven schoolvestigingen en hun totale aantal leerlingen in leerjaar 2, naar stratum

Strata	Scholen		Leerlingen in leerjaar 2	
	n	%	n	%
Pro	6	4,3	246	0,9
Pro plus	2	1,4	668	2,4
Vmbo	28	20,3	4446	16,3
Vmbo-gt	3	2,2	573	2,1
Vmbo-gt en havo	3	2,2	545	2,0
Vmbo-gt en havo en/of vwo	32	23,2	8116	29,7
Breed	35	25,4	6933	25,4
Havo/vwo	24	17,4	5012	18,3
Vwo	5	3,6	778	2,8
<b>Totaal</b>	<b>138</b>	<b>100%</b>	<b>27.317</b>	<b>100%</b>

### Redenen non-respons

Wanneer een school afzag van deelname aan het onderzoek werd de school uit de eerste vervangende steekproef (tweede tranche) benaderd via dezelfde procedure. Indien ook deze weigerde werd de school uit de tweede vervangende steekproef (derde tranche) benaderd. Er waren echter ook scholen die de beslissing over deelname heel lang in het midden lieten en/of telefonisch slecht bereikbaar waren. In verband met de tijdsdruk zijn in deze gevallen – dus zonder dat er uitsluitel was over de deelname van de eerste school – vaak alvast de vervangende scholen benaderd, zeker tegen het einde van de wervingsperiode. In totaal hebben zodoende ongeveer 225 scholen (waaronder alle 175 scholen uit de hoofdstekproef) de uitnodigingsbrief ontvangen. De meeste van de 87 scholen die niet zijn ingegaan op de uitnodiging om deel te nemen zijn tenminste één keer – maar soms veel vaker – telefonisch benaderd door een onderzoeksassistent, maar van slechts 62 van hen hebben we een definitieve afmelding ontvangen. Tabel 4.2 toont de opgegeven reden(en) daarvoor, gesorteerd naar frequentie.

Tabel 4.2 – Redenen voor non-respons van scholen

Reden(en) afmelding	Aantal scholen
Te druk/kost te veel tijd/moeite	13
Corona-gerelateerd	10
Willen zelf klas kunnen kiezen	5
Prioriteiten elders	4
Deelname andere onderzoeken	4
Te belastend voor leerlingen	4
Hoge lesuitval	4
Personeelsgebrek/ziekteverzuim	3
Willen focus op regulier 'normaal' onderwijs	3
Grote veranderingen (zoals verhuizing)	3
Afmelding zonder reden	3
Onrust in school/al genoeg problemen	2
Zien het belang niet in	2
Doen nooit mee aan onderzoek	2
<b>Totaal</b>	<b>62</b>

## 4.3 Selectie van klassen

### 4.3.1 Totaal aantal tweede klassen

Bij de 138 geworven schoolvestigingen is informatie opgevraagd over alle tweede klassen die ze in het voorjaar van 2022 in huis hadden. Per klas ging het om de klasnaam, het aantal leerlingen en het onderwijsaanbod. In totaal gaven de scholen 1170 tweede klassen op die in veertien categorieën zijn in te delen wat betreft aangeboden (combinaties van) onderwijstypen<sup>5</sup>. Deze zijn opgenomen in Tabel 4.3, inclusief het totale aantal klassen en leerlingen.

Tabel 4.3 – Alle tweede klassen (naar onderwijsaanbod) en leerlingen daarin, op 138 geworven scholen

Aangeboden onderwijstype(n) in tweede klas	Aantal klassen	Aantal leerlingen
Pro	13	157
Pro/vmbo*	2	25
Vmbo-b	82	1.177
Vmbo-b/vmbo-k	109	1.944
Vmbo-b/vmbo-k/vmbo-(g)t	9	171
Vmbo-k	95	1.970
Vmbo-k/vmbo-(g)t	55	1.202
Vmbo-(g)t	172	4.323
Vmbo-(g)t/havo	71	1.749
Vmbo-(g)t/havo/vwo	17	471
Havo	189	4.905
Havo/vwo	112	2.876
Vwo	234	6.059
Breed	10	249
<b>Totaal</b>	<b>1.170</b>	<b>27.278**</b>

\* De klassen met aanbod pro/vmbo komen voor op pro-scholen. \*\* Zie voetnoot voor verklaring van het verschil met Tabel 4.1.

Gemiddeld hebben de 138 schoolvestigingen 8,5 tweede klassen, met een minimum van één en een maximum van twintig tweede klassen. Het gemiddeld aantal leerlingen per klas varieert van twaalf in pro- en pro/vmbo-klassen tot 28 in vmbo-t/havo/vwo-klassen.

### 4.3.2 Selectie van één klas per vestiging

Per aangemelde schoolvestiging hebben de onderzoekers random één tweede klas geselecteerd voor deelname aan het peilingonderzoek. Vanwege de korte voorbereidingstijd voor het peilingsonderzoek is deze klasselectie in twee batches uitgevoerd. De eerste klasselectie vond plaats op 1 maart 2022 en betrof de 98 scholen die zich op dat moment via het aanmeldformulier hadden aangemeld. Op 25 mei 2022 is de klasselectie uitgevoerd voor de veertig scholen die zich aanvullend hadden aangemeld voor het onderzoek.

<sup>5</sup> Feitelijk werden er nog twee klassen opgegeven, maar daarin 39 leerlingen. De door de school opgegeven niveau-aanduiding voor deze klassen is echter onduidelijk ('E' en 'Masterclass tto').

Om ook met de lage leerlingaantallen per pro-klas voldoende pro-leerlingen te laten deelnemen is er na de originele selectie van klassen een aanvullende selectie van pro-klassen uitgevoerd. Aan scholen die alleen praktijkonderwijs aanboden (stratum 1) is gevraagd of we random nog een tweede pro-klas mochten selecteren voor deelname aan het onderzoek. Dit resulteerde uiteindelijk in drie extra pro-klassen. Om voldoende vmbo-leerlingen te laten deelnemen aan het onderzoek zijn na de klasseselectie voor batch 1 een aantal havo-, vwo- of havo-/vwo-klassen vervangen door vmbo-klassen. Dit betrof in totaal elf klassen. Omdat voorafgaand aan de klasseselectie voor batch 2 bleek dat deze vervanging niet nodig was (geweest), is na de klasseselectie voor batch 2 de omgekeerde procedure gevolgd: geselecteerde vmbo-klassen werden vervangen door havo-, vwo- of havo-/vwo-klassen op scholen waar dit mogelijk was. Dit betrof vijf scholen en dus klassen.

De scholen werden door de onderzoeksassistenten op de hoogte gesteld van de geselecteerde klas die aan het onderzoek zou deelnemen. Op één school heeft dat geleid tot een wijziging, omdat de betreffende klas toevallig ook was geselecteerd voor deelname aan het internationaal vergelijkende ICCS-onderzoek. Uiteindelijk resulteerde de procedure van de klasseselectie in de verdeling die is weergegeven in Tabel 4.4. De tabel laat ook het aantal leerlingen zien dat in totaal in de geselecteerde klassen zou zitten volgens de opgave van de scholen. Over alle niveaus heen zijn dat er 3226.

*Tabel 4.4 – Geselecteerde klassen naar onderwijsaanbod, aantal leerlingen volgens opgave school*

Aangeboden onderwijstype(n) in tweede klas	N klassen	N leerlingen	% leerlingen
Pro	7	85	2,6
Pro/vmbo	1	9	0,3
Vmbo-b	7	91	2,8
Vmbo-b/vmbo-k	15	297	9,2
Vmbo-b/vmbo-k/vmbo-(g)t	1	26	0,8
Vmbo-k	12	247	7,7
Vmbo-k/vmbo-(g)t	6	146	4,5
Vmbo-(g)t	23	579	17,9
Vmbo-(g)t/havo	7	175	5,4
Vmbo-(g)t/havo/vwo	4	103	3,2
Havo	18	454	14,1
Havo/vwo	9	217	6,7
Vwo	28	720	22,3
Breed	3	77	2,4
<b>Totaal</b>	<b>141</b>	<b>3.226</b>	<b>100</b>

#### 4.4 Dataverzameling

Op de afgesproken toetsdatum bezocht een onderzoeksassistent de school om de toets en de leerlingvragenlijst bij de geselecteerde klas af te nemen. De onderzoeksassistenten waren speciaal getraind om ervoor te zorgen dat de dataverzameling op elke school zoveel mogelijk hetzelfde zou verlopen. Zij brachten ook de papieren boekjes mee waarin de toetsen en vragenlijsten waren opgenomen. De leerlingen waren vooraf geïnstrueerd (via hun school) om hun rekenmachine en liniaal (of geodriehoek) mee te brengen, de toetsleiders hadden enkele reserve-exemplaren bij zich voor het geval deze

onvoldoende beschikbaar zouden zijn. De scholen verzorgden in principe enkel de ruimte. In overleg met de school kon wel worden afgesproken dat een docent of andere schoolmedewerker bij de afname aanwezig zou zijn, om onrust in de klas te voorkomen. De onderzoeksassistenten vulden bij elke afname een formulier in waarmee geregistreerd werd hoe de afname was verlopen (bijv. aantal afwezige en uitgesloten leerlingen, overige bijzonderheden).

Omdat de toets uit een deel met en een deel zonder gebruik van een rekenmachine bestond en om te voorkomen dat de leerlingen onnodig afgeleid zouden worden door het afgaan van een schoolbel, was de afname van de toets opgesplitst in twee delen. Tabel 4.5 geeft de timing van de afname weer, voor leerlingen in pro-classes (die een verkorte toets en vragenlijst kregen) en in alle overige vo-classes.

*Tabel 4.5 – Toets- en vragenlijstafnameduur voor leerlingen in minuten, per niveau en activiteit*

Afnameschema	Pro	Overig VO
Uitleg onderzoek en instructie toets	10	10
Toetsafname I	15	40
Pauze	5	5
Toetsafname II	15	25
Instructie leerlingvragenlijst	5	5
Invullen leerlingvragenlijst	10	15
<b>Totale afnameduur (minuten)</b>	<b>60</b>	<b>100</b>

De toets en de vragenlijst waren opgenomen in twee (papieren) boekjes, die correspondeerden met deel I en II van het afnameschema. Bij het uitdelen zorgde de onderzoeksassistent ervoor dat op de twee boekjes waarin een leerling werkte hetzelfde respondentnummer vermeld stond. Zodoende konden achteraf alle antwoorden van die leerling aan elkaar gekoppeld worden. De anonimiteit van de leerlingen werd gewaarborgd, omdat van hen alleen dat nummer opgeslagen werd in de databestanden. Uit het nummer was af te leiden welke school het betrof, maar niet welke leerling.

Zoals is beschreven in Hoofdstuk 3 bestond het toetsinstrumentarium uit twaalf toetsen met een oplopende moeilijkheidsgraad. Het onderwijsaanbod in de geselecteerde klas (opgegeven door de school bij de aanmelding) bepaalde welke van deze toetsen de onderzoeksassistent meebracht naar de school. De leerlingen van één klas kregen echter niet allen dezelfde toets; gemiddeld werden er per klas 3 tot 4 toetsvarianten random uitgedeeld. In pro-classes maakten wel alle leerlingen dezelfde (verkorte) toets.

De docent en schoolleider kregen voorafgaand aan het schoolbezoek de link naar de digitale docent- of schoolvragenlijst toegestuurd. Uit de gebruikte link konden we afleiden welke school het betrof, maar niet de naam van de respondent. Hiermee werd ook de anonimiteit van docenten en schoolleiders gewaarborgd.

## 4.5 Respons

De dataverzamelingsperiode liep van begin april tot in juli 2022. Tegen het einde van deze periode trokken 8 van de 138 geworven scholen hun deelname aan het onderzoek helaas weer in. Het gaat om scholen die de afspraak voor de toetsafname pas op het laatste moment annuleerden, waarna het niet

meer lukte een nieuwe afnamedatum te vinden vóór de zomervakantie. Redenen voor de late annulering waren wijzigingen in de jaarplanning en onverwachte problemen of drukte op school. Op de scholen die wel deelnamen kwam het volgens de terugkoppeling van de toetsleiders geregeld voor dat één of enkele leerlingen afwezig waren tijdens de afname. Tabel 4.6 toont het aantal leerlingen dat uiteindelijk aan het onderzoek heeft deelgenomen, en laat ook zien over hoeveel scholen en klassen (naar onderwijsaanbod) zij verdeeld waren.

Tabel 4.6 – Definitieve aantallen deelnemers, naar onderwijsaanbod in de tweede klas

Aangeboden onderwijstype(n) in tweede klas	N scholen	N klassen	N leerlingen
Pro	4	7	71
Pro/vmbo	1	1	8
Vmbo-b	7	7	80
Vmbo-b/vmbo-k	14	14	236
Vmbo-b/vmbo-k/vmbo-(g)t	1	1	26
Vmbo-k	11	11	201
Vmbo-k/vmbo-(g)t	4	4	83
Vmbo-(g)t	21	21	469
Vmbo-t/havo	7	7	159
Vmbo-t/havo/vwo	4	4	85
Havo	17	17	397
Havo/vwo	9	9	201
Vwo	27	27	633
Breed	3	3	75
<b>Totaal</b>	<b>130</b>	<b>133</b>	<b>2.724</b>

Het streven was dat minimaal 148 scholen (85% van de bruto steekproef) deel zouden nemen aan het onderzoek, waarvan de helft uit de hoofdstekproef zou komen. Het laatste is gelukt, ook na verdiscontering van de 8 late afvallers namen er nog 93 scholen uit de hoofdstekproef deel (plus 30 scholen uit de tweede en 7 uit de derde tranche). Het totale nagestreefde responspercentage is echter niet behaald; 130 scholen komt neer op 74%. Hierna onderzoeken we in hoeverre de (non-)respons varieert naar kenmerken van scholen en klassen.

#### SCHOLEN

In Tabel 4.7 zien we in hoeverre de respons van scholen, berekend als percentage van de totale steekproef, varieert tussen de negen strata. De respons blijkt het laagst onder de scholen in stratum 'pro plus' (20%), daarnaast is de deelname van scholen in de strata 'vmbo-gt', 'vmbo-gt en havo' en 'vmbo-gt en havo en/of vwo' met circa 60% relatief laag.

Tabel 4.7 – Steekproef en deelnemende scholen, respons naar stratum

Vestigingsstrata	Steekproef	Deelnemers	Respons (%)
Pro	5	5	100
Pro plus	5	1	20
Vmbo	34	28	82
Vmbo-gt	5	3	60
Vmbo-gt en havo	5	3	60
Vmbo-gt en havo en/of vwo	48	30	63
Breed	38	31	82
Havo/vwo	29	24	83
Vwo	6	5	83
Totaal	175	130	74

Tabel 4.8 toont hoe de scholen in de steekproef zijn verdeeld over drie categorieën van schoolgrootte (gebaseerd op het totale aantal leerlingen in leerjaar 2 per 1 oktober 2020 volgens DUO). Daarnaast zien we de verdeling voor de deelnemende scholen. De laatste kolom laat zien dat de respons relatief het laagst was voor de middelgrote scholen (66%).

Tabel 4.8 – Steekproef en deelnemende scholen, respons naar schoolgrootte (aantal leerlingen in leerjaar 2 in oktober 2020)

Schoolgrootte (leerjaar 2)	Steekproef	Deelnemers	% respons
Tot 150 leerlingen	59	45	76%
151-250 leerlingen	65	43	66%
250-545 leerlingen	51	42	82%
Totaal	175	130	74%

### Klassen en leerlingen

Al voorafgaand aan het peilingsonderzoek werd in de notitie van de Inspectie over het steekproefkader opgemerkt dat, door het ontbreken van betrouwbare informatie over het aantal leerlingen in de verschillende brugklasstypen op landelijk niveau (de registerdata van DUO), niet nagerekend zou kunnen worden hoe representatief de leerlingen in de geselecteerde klassen zijn voor de populatie.

Een oplossing voor dit probleem ligt in de variabele ‘Verwacht onderwijsniveau in leerjaar 3’ die bij de deelnemende tweedeklassers verzameld is met behulp van een vraag in de leerlingvragenlijst<sup>6</sup> en die kan worden vergeleken met de landelijke verdeling van leerlingen in leerjaar 3. De antwoorden van de leerlingen op de betreffende vraag zijn eerst, voor zover mogelijk, gevalideerd met behulp van de informatie uit de docentvragenlijst over het onderwijsaanbod in de tweede klas. Daardoor konden zeer onwaarschijnlijke antwoorden uitgefilterd worden; bijvoorbeeld van leerlingen in een 2vwo-klas die aangeven in leerjaar 3 naar verwachting in vmbo-basis te zitten, of omgekeerd. Enkele leerlingen hebben wel de toets gemaakt maar niet de leerlingvragenlijst ingevuld of de betreffende vraag overgeslagen, sommige andere leerlingen gaven een zodanig antwoord dat – ondanks de informatie over het

<sup>6</sup> In de vragenlijst voor pro-leerlingen kwam deze vraag niet voor; de aanname was/is dat zij in leerjaar 3 ook in pro zitten.

onderwijsaanbod in leerjaar 2 – het verwachte niveau in leerjaar 3 niet met zekerheid kon worden bepaald. Zij vallen in de categorie ‘Onbekend’. Van het overgrote deel van de deelnemende leerlingen (92%) is het verwachte niveau in leerjaar 3 echter bekend. In Tabel 4.9 zijn zij in vier niveaugroepen ingedeeld, die dus verwijst naar het verwachte onderwijsniveau in leerjaar 3: pro, vmbo-b/k, vmbo-g/t en havo/vwo. In het rechterdeel van de tabel is te zien hoe de totale leerlingenpopulatie in leerjaar 3 in 2022 verdeeld was over deze vier niveaugroepen (bron: DUO).

Tabel 4.9 – (Verwacht) onderwijsniveau leerjaar 3: deelnemende leerlingen en landelijke populatie 2022

(Verwacht) onderwijsniveau in leerjaar 3	Deelnemers onderzoek		Landelijke populatie 2022	
	N	%	n	%
Pro	79	3	5.760	3
Vmbo-b/k	491	19	45.767	23
Vmbo-g/t	619	24	52.200	27
Havo/vwo	1.342	53	91.272	47
<b>Totaal</b>	<b>2.531</b>	<b>100</b>	<b>194.999</b>	<b>100</b>
Onbekend	193			

Onder de onderzoeksdeelnemers is het aandeel havo/vwo leerlingen iets groter dan in de landelijke populatie, het aandeel vmbo-leerlingen is juist wat kleiner. De verschillen zijn volgens de uitgevoerde Chi-kwadraattoets echter niet significant ( $p=1$ ).

De representativiteit op leerlingniveau kan daarnaast ook worden vastgesteld binnen de geworven schoolvestigingen, door na te gaan of de voor het onderzoek geselecteerde tweede klassen representatief zijn voor alle tweede klassen op de geworven scholen (Tabel 4.10 linkerdeel). Bovendien kunnen we nagaan of de leerlingen in de geselecteerde klassen vergelijkbaar verdeeld zijn over de aangeboden onderwijstypen als de leerlingen in alle tweede klassen van de deelnemende scholen (Tabel 4.10 rechterdeel). Met behulp van Chi-kwadraattoetsen is nagegaan of er significante verschillen zijn. Wat betreft de verdeling van de klassen blijkt dat niet het geval ( $p=0,999$ ) en hetzelfde geldt voor de verdeling van de leerlingen ( $p=1$ ).

Gezien het voorafgaande zien we geen noodzaak tot het toepassen van weging op de verzamelde data over de rekenvaardigheid van de onderzochte leerlingen.



Tabel 4.10 – Verdeling klassen/leerlingen naar onderwijsaanbod in de tweede klas; alle klassen/leerlingen en geselecteerde/deelnemende klassen/leerlingen (130 scholen, 133 klassen)

Aangeboden onderwijstype(n) in tweede klas	Alle klassen	Geselecteerde klassen	Alle leerlingen	Deelnemende leerlingen
Pro	1,1%	5,3%	0,6%	2,6%
Pro/vmbo	0,2%	0,8%	0,1%	0,3%
Vmbo-b	7,2%	5,3%	4,4%	2,9%
Vmbo-b/vmbo-k	8,6%	10,5%	6,6%	8,7%
Vmbo-b/vmbo-k/vmbo-(g)t	0,8%	0,8%	0,7%	1,0%
Vmbo-k	8,0%	8,3%	7,1%	7,4%
Vmbo-k/vmbo-(g)t	4,7%	3,0%	4,4%	3,0%
Vmbo-(g)t	14,7%	15,8%	15,8%	17,2%
Vmbo-(g)t/havo	6,2%	5,3%	6,5%	5,8%
Vmbo-(g)t/havo/vwo	1,6%	3,0%	1,9%	3,1%
Havo	15,9%	12,8%	17,7%	14,6%
Havo/vwo	9,9%	6,8%	10,9%	7,4%
Vwo	20,2%	20,3%	22,5%	23,2%
Breed	0,9%	2,3%	1,0%	2,8%
<b>Totaal</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>

#### Schoolleiders en docenten

Van de 130 deelnemende scholen hebben in totaal 101 schoolleiders de schoolvragenlijst ingevuld, een respons van 78%. De docentvragenlijst is ingevuld door 100 docenten van de 133 geselecteerde tweede klassen (75%). Meer informatie over het onderwijsaanbod op de scholen en in de tweede klassen van deze respondenten is te vinden in Hoofdstuk 6.



## 5 Schaling rekenvaardigheid

Opmerking bij dit rapport (april 2023). In dit hoofdstuk wordt in voetnoten (bijv. 9) verwezen naar (tabbladen van) Excelbestanden. Deze zijn mogelijk handig voor de Inspectie om analysebestanden bij Cito op te vragen, maar moeten worden verwijderd als dit TR voor algemeen publiek wordt uitgebracht.

### 5.1 Resultaten van de toetsafname

De toetsen met (RM) en zonder (HR) gebruik van een rekenmachine zijn als aparte toetsen afgenomen, en ook als twee verschillende datasets aangeleverd. Deze konden aan elkaar gerelateerd worden doordat de leerlingen unieke identificatie-labels hadden. Het design was zo opgezet dat iedere leerling hetzelfde nummer van de RM- en de HR-toets maakte. Vervolgens is per itemset (zie Paragraaf 3.3) gekeken hoeveel observaties er beschikbaar waren. Het komt namelijk voor dat een leerling een of meer itemsets geheel, of grotendeel heeft overgeslagen, of dat deze niet voorgelegd zijn. Een leerling kan dus een toets gemaakt hebben, maar (bijvoorbeeld) zonder dat het HR-deel is voorgelegd of zonder dat het tweede deel (de tweede itemset) van het RM-deel is voorgelegd. In de analyses zijn daarom de itemsets als aparte toetsboekjes in de analyses meegenomen. In Tabel 5.1 is aangegeven hoeveel leerlingen de toetsen 'volgens design' gemaakt hebben en in hoeverre dat afwijkt van het voorgestelde toetsdesign in Tabel 3.2b. Deze tabel gaat uitsluitend over geheel niet voorgelegde itemsets, niet over ontbrekende waarden op itemniveau.

Tabel 5.1 – Aantal leerlingen dat toetsen 'volgens design' gemaakt heeft, in vergelijking tot uitgangspunt

	T01	T02	T03	T04	T05	T06	T07	T08	T09	T10	T11	T12
Uitgangspunt	80	168	149	155	164	174	189	186	262	437	293	467
Design in data	78	163	146	153	159	163	178	180	259	433	289	456
Uitval (n)	2	5	3	2	5	11	11	6	3	4	4	11
Uitval (%)	3%	3%	2%	1%	3%	6%	6%	3%	1%	1%	1%	2%

In dit hoofdstuk gaan we in op de antwoorden die op de opgaven gegeven zijn. We rapporteren eerst over de klassieke resultaten per itemset. We maken hierbij onderscheid tussen de itemset in de afnameconditie met rekenmachine (RMis01-RMis12) en zonder rekenmachine (HRis01-HRis12). We nemen in deze fase alle leerlingen mee, ongeacht het aantal ontbrekende waarden. Op basis van die resultaten gaan we in Paragraaf 5.3 dieper in op de wijze waarop met ontbrekende waarden is omgegaan, en wat de impact is op de verdere analyses.

In de verdere analyses komen we kort terug op de klassieke resultaten, nu wel rekening houdend met het ontbreken van antwoorden op opgaven, en gaan daarna dieper in op de IRT-analyses. Hierbij kijken we naar de resultaten voor afzonderlijke rekenschalen zoals die te vormen zijn op basis van de vier verschillende kenmerken. Dat betekent dat we onderzoek doen naar de twee schalen die gevormd kunnen worden op basis van de afnameconditie (de RM- en HR-schaal), naar de vijf schalen op basis van de indeling naar inhoud (B-, C-, D-, E-, en N1-schaal), naar de vier schalen op basis van referentieniveau (1F-, 2F-, 1S- en 3F-schaal) en naar de twee schalen op basis van vorm (Context- en Kaal-schaal). Onder andere door alvast een specifieke onderzoeksvraag te beantwoorden – namelijk die naar de

samenhang tussen verschillende schalen – kunnen we toelichten waarom het mogelijk en ook aan te raden is, om gebruik te maken van een enkele rekenschaal.

## 5.2 Klassieke resultaten per itemset

Voor alle opgaven is per itemset een klassieke toets- en itemanalyse uitgevoerd. Het eerste doel was er achter te komen of er mogelijk sleutelfouten waren. Bij de 52 meerkeuzeopgaven<sup>7</sup> zou een fout antwoordalternatief als correct gezien kunnen zijn, en bij de 212 open opgaven zou mogelijk kunnen blijken dat bepaalde alternatieven ook goed gerekend kunnen worden, terwijl die in eerste instantie niet automatisch als correct geregistreerd zijn. Als voorbeeld noemen we dat naast het antwoord '4' ook '4 uur', '4u' of '4,0' goed gerekend zouden kunnen worden.

De klassieke analyses in deze fase worden bij voorkeur gedaan per itemset. Een item zit in een specifieke itemset, en zo worden in de analyses alle leerlingen die een specifiek item gemaakt hebben bij elkaar genomen. Uit deze analyses bleek inderdaad dat er sleutels waren die (enigszins) aangepast moesten worden. Dat was voor 14 open opgaven het geval en voor 6 daarvan betrof het een significante aanpassing. De rest van de analyses kon na deze aanpassing met de juiste scores uitgevoerd worden. Na het aanpassen van de sleutels waren er geen opgaven meer met negatieve  $r_{ir}$ -waarden<sup>8</sup>. Bij 98% van de opgaven was de  $r_{ir}$ -waarde groter dan 0,10, en bij 86% groter dan 0,20. De gemiddelde  $r_{ir}$ -waarde is 0,30 en meer dan de helft van de items (53%) had een  $r_{ir}$ -waarde daarboven. Als we kijken naar de p-waarden zien we dat de gemiddelde p-waarde gelijk is aan 51. Gegeven de kans op gokken (die beperkt is bij open opgaven) en het geringe aantal meerkeuzeopgaven is deze p-waarde zeer acceptabel, aangezien deze dicht tegen het punt ligt waarmee de opgave het beste onderscheid maakt tussen de leerlingen op de vaardigheidsschaal<sup>9</sup>. Driekwart van de opgaven heeft een p-waarde tussen 20 en 80, met ongeveer een gelijke verdeling van (zeer) moeilijke (4% met een p-waarde lager dan 10) en (zeer) makkelijke opgaven (5% met een p-waarde groter dan 90). Deze klassieke resultaten laten zien dat het zowel op basis van de  $r_{ir}$ -waarden als van de p-waarden voldoende goed gelukt is om, bij de selectie van items in itemsets en het toewijzen van de toetsen aan de leerlingen, de juiste leerlingen de juiste opgaven voor te leggen.

In Tabel 5.2 wordt het overzicht gegeven van alle itemsets, onderverdeeld in RM- en HR-itemsets. Op het niveau van de itemsets is de betrouwbaarheid al voldoende als we de COTAN-richtlijnen (Evers et al., 2010) hanteren met betrekking tot het vergelijken van leerlingen op systeemniveau, zoals bij deze peiling het geval is. In deze richtlijnen wordt een betrouwbaarheid vanaf 0,60 als voldoende beschouwd en een betrouwbaarheid vanaf 0,70 als goed. Hiermee is te zien dat op het niveau van de itemsets de betrouwbaarheid van ongeveer de helft al voldoende is en een kwart zelf goed. Als we realiseren dat leerlingen over het algemeen geëvalueerd worden op basis van 4 itemsets, en gezien de positieve relatie tussen het aantal items en de betrouwbaarheid, is dit veelbelovend voor de betrouwbaarheid op toetsniveau<sup>10</sup>.

<sup>7</sup> De meeste (46) meerkeuzeopgaven hebben vier antwoordalternatieven. Daarnaast zijn er 5 meerkeuzeopgaven met vijf alternatieven en is er 1 opgave met drie antwoordalternatieven.

<sup>8</sup> De  $r_{ir}$ -waarde geeft de correlatie tussen de opgaven en de rest van de opgaven in de itemset weer. Een alternatief voor deze waarde is de  $r_{it}$ -waarde die de correlatie van de opgave met de score op de gehele itemset weergeeft, waarbij het item zelf ook bijdraagt aan deze totaalscore. Aangezien de itemsets uit slecht 8 tot en met 14 opgaven bestaan, is de impact van de autocorrelatie bij het rapporteren van de  $r_{it}$ -waarde relatief groot, en is hier gekozen voor de evaluatie via de  $r_{ir}$ -waarde.

<sup>9</sup> Bij deze opgaven, rekening houdend met de kans op gokken, zou de ideale p-waarde 52 zijn (zie tabblad 'alle items per bk (hardcopy)' in '220725\_PRV.TIA\_overzicht\_NEW\_JV.xlsx').

<sup>10</sup> Bij de voetnoot bij Tabel 5.5 wordt verder ingegaan op afwezigheid van de impact van de clustering van de steekproef en de afwezigheid, dan wel beperkte invloed van de ICC op de hoogte van alpha.

Tabel 5.2 – Klassieke toets- en itemkenmerken op het niveau van de itemsets

Label	Aantal		Scores			P-waarde			Rir-waarde		
	Personen	Items	Gem	SD	Alpha (se)	Min	Gem	Max	Min	Gem	Max
RMis01	243	9	3,2	2,1	,71 (.03)	5	36	72	23	38	56
RMis02	310	8	2,0	1,6	,53 (.04)	4	25	51	10	25	35
RMis03	300	9	2,4	1,8	,61 (.03)	10	26	68	4	31	48
RMis04	315	9	3,0	2,0	,61 (.03)	11	34	57	17	30	42
RMis05	332	8	3,5	2,0	,60 (.03)	36	44	61	17	30	40
RMis06	357	9	3,9	2,3	,68 (.03)	26	44	59	26	35	44
RMis07	369	8	3,4	1,9	,57 (.03)	24	42	58	14	28	37
RMis08	445	9	5,2	2,4	,72 (.02)	38	57	77	32	40	47
RMis09	694	8	4,7	1,9	,63 (.02)	32	59	80	23	32	41
RMis10	726	9	4,0	1,7	,45 (.03)	8	45	78	12	19	32
RMis11	755	8	3,7	1,8	,55 (.02)	11	47	70	18	27	39
RMis12	460	9	3,6	1,8	,56 (.03)	5	40	79	4	27	43
HRis01	248	13	8,4	2,9	,74 (.02)	33	65	92	25	37	62
HRis02	317	14	7,7	2,5	,65 (.03)	9	55	97	2	27	40
HRis03	304	13	6,5	2,1	,57 (.04)	2	50	99	6	22	39
HRis04	316	13	7,5	2,4	,69 (.03)	3	58	97	11	31	61
HRis05	328	14	7,3	2,5	,68 (.03)	7	52	95	13	30	41
HRis06	349	13	6,0	2,7	,69 (.02)	10	46	88	19	32	41
HRis07	364	14	9,8	2,6	,64 (.03)	49	70	92	4	27	46
HRis08	444	13	8,0	2,8	,72 (.02)	25	61	94	19	34	48
HRis09	697	14	9,8	2,5	,68 (.02)	21	70	95	16	30	40
HRis10	725	13	6,6	2,6	,69 (.02)	12	51	87	17	31	45
HRis11	755	14	7,5	2,6	,63 (.02)	18	54	88	13	26	41
HRis12	461	13	6,0	2,8	,70 (.02)	10	46	82	17	33	47

De moeilijkheid in termen van de p-waarde die hier gegeven is, betreft alleen de relatieve moeilijkheid ten opzichte van de groep aan wie de opgave is voorgelegd. Aangezien de toetsen specifiek aan leerlingen zijn toegewezen op basis van hun (verwachte) rekenniveau kunnen op basis van deze resultaten geen definitieve conclusies getrokken worden over de absolute moeilijkheid van de opgaven. Een item met een lagere p-waarde in een moeilijke taak kan makkelijker zijn dan een item met een hogere p-waarde in een makkelijke taak. Om opgaven van een bepaald domein over boekjes heen te kunnen vergelijken kunnen we geen klassieke analyses gebruiken, maar zullen met behulp van item respons theorie (IRT) de items per domein op één en dezelfde meetschaal geplaatst worden.

Tot slot kijken we naar de percentages ontbrekende waarden bij de opgaven. Het blijkt dat bij slecht 9 opgaven het percentage ontbrekende waarden (afgerond) 0 is. Deze 9 items zitten bij ongeveer de helft van de opgaven waar minder dan 5% van de te verwachten antwoorden ontbreekt. Bij een kleine 70% van de opgaven ontbreekt 10% of minder. Het maximum percentage ontbrekende antwoorden was 41%. Dat is een van de 6 items in Tabel 5.3 met meer dan 30% ontbrekende waarden.

Tabel 5.3 – Overzicht van ontbrekende antwoorden

	Percentage ontbrekende antwoorden per item				
	<5%	5-10%	11-20%	21-30%	>30%
Aantal items dat dit betreft	134	49	60	15	6
Percentage items dat dit betreft	50,7%	18,6%	22,7%	5,7%	2,3%

Zoals vaker in dergelijke situaties wordt ook hier gevonden dat de open opgaven vaker overgeslagen worden dan de gesloten opgaven. Bij de meerkeuzevragen ontbreekt gemiddeld 2,3% van de antwoorden, terwijl dat bij de open opgaven 9,1% is.

Bij deze klassieke analyses zijn alle leerlingen meegenomen, ook als deze het grootste deel van een itemset niet gemaakt hebben (of soms zelfs in het geheel niet), maar wel geadmistreerd zijn als leerlingen aan wie de itemset is voorgelegd. Het is duidelijk dat bij de verdere analyses met deze ontbrekende waarnemingen rekening gehouden moet worden. Dit wordt in de volgende paragraaf beschreven, waarna opnieuw naar de klassieke resultaten gekeken wordt.

### 5.3 Ontbrekende waarden

Bij de analyses van de ontbrekende waarden is bij alle itemsets gekeken naar leerlingen die veel ontbrekende waarden hebben, en naar leerlingen die specifieke patronen van ontbrekende waarden hebben. Wat betreft het eerste moet onderscheid gemaakt worden tussen leerlingen die de toetsen serieus gemaakt hebben – en bij wie er dus werkelijk iets gemeten wordt – en leerlingen die dat evident niet gedaan hebben. In deze paragraaf is beschreven hoe met verschillende patronen van ontbrekende waarden is omgegaan in de verschillende analyses.

Bij het evalueren van de opgaven is ook gekeken naar opgaven met veel ontbrekende waarden. Als dit niet met een specifiek patroon op leerlingniveau verband houdt dan betreft dat meestal opgaven die meer moeite kosten om te beantwoorden. Dat betreft moeilijke vragen of open vragen waarbij de leerling niet uit voorgegeven antwoorden hoeft te kiezen, maar zelf een antwoord moet opschrijven. Moeilijke open vragen worden daardoor vaker overgeslagen. In de vorige paragraaf zagen we al dat open opgaven duidelijk vaker worden overgeslagen dan meerkeuzeopgaven. In dit geval heeft het overslaan van een opgave wel weer met een itemkenmerk te maken. Welke actie ondernomen moet worden verschilt naar gelang de interpretatie van de oorzaak van de ontbrekende waarde. Wanneer ontbrekende waarden in de analyses meegenomen worden, is het bij kennisvragen gebruikelijk een ontbrekende opgave te interpreteren als het ontbreken van een (deels) correct antwoord. De ontbrekende waarde krijgt dan de waarde van de minimumscore op dat item. In de meeste gevallen wordt dat toegepast wanneer een opgave relatief vaak overgeslagen wordt en de rekenexperts daarvoor de moeilijkheid van de opgave als aannemelijke oorzaak zien.

Zoals aangegeven zijn er ook criteria om te veronderstellen dat het ontbreken van een antwoord niet zozeer het kenmerk van een opgave is, maar van een leerling: (1) wanneer het een specifieke reeks ontbrekende waarden betreft; (2) wanneer het percentage ontbrekende waarden extreem hoog is. Wanneer een dergelijk specifiek patroon van ontbrekende waarden gevonden wordt, is het simpelweg imputeren van een ontbrekend antwoord als zijnde een fout antwoord niet altijd verstandig. Hierbij kan onderscheid gemaakt worden tussen de analyses die gebruikt worden voor de schaling van de opgaven en voor de vaardigheidsbepaling. Als een leerling minder vaardig is, is de kans groter dat deze niet tot

het einde van de toets komt. Het niet beantwoorden van de vraag ligt dan niet aan de opgave maar aan de leerling, daar waar we bij de kalibratie vooral geïnteresseerd zijn in de opgaven. Het gebruiken van deze leerlingen bij de schaling zou tot een overschatting van de moeilijkheid van de opgave leiden. Het niet meenemen van de leerling bij de vaardigheidsbepaling zou leiden tot een overschatting van de vaardigheid. Daarom worden dergelijke leerlingen niet meegenomen bij de schaling van de opgaven, maar wel bij de bepaling van de vaardigheid van de populatie.

Ook een reeks ontbrekende opgaven in het midden van een toets is een indicatie om een leerling niet mee te nemen in de schaling. Het kan betekenen dat een pagina in een toetsboekje overgeslagen is, of dat er een andere reden is die niet direct aan de opgave gerelateerd is waarom de opgave ontbreekt. Deze lokaal stochastische afhankelijkheid tussen opgaven past ook niet bij het toe te passen model. Bij het bepalen van de vaardigheid kan op basis van de evaluatie van de mogelijke oorzaak of oorzaken besloten worden deze leerlingen wel of niet mee te nemen in de analyses. In de meeste gevallen worden deze leerlingen dan wel mee genomen. Wanneer een leerling slechts een zeer beperkt aantal opgaven gemaakt heeft is de vraag relevant of deze leerling wel serieus de toets heeft gemaakt of heeft kunnen maken; bijvoorbeeld in geval van ziekte of het ontbreken van motivatie. De vraag is dan of daarmee in plaats van rekenen-wiskunde niet iets anders gemeten is. Het is goed mogelijk dat een lage motivatie wel samenhangt met de rekenvaardigheid, maar het gebrek aan deze vaardigheid wordt overschat wanneer deze leerling in de analyses meegenomen wordt.

Bij het ontbreken van reeksen van opgaven evalueren we de resultaten per itemset. Binnen peilingen is het gebruikelijk om leerlingen met vijf of meer ontbrekende opgaven op een rij – waar dan ook in de toets – niet mee te nemen met de analyses voor de schaling. Voor wat betreft het minimum aan te maken opgaven kan gesteld worden dat de beschikbare tijd voor het maken van de opgaven dusdanig is dat het in ieder geval mogelijk was om op meer dan de helft van de opgaven in een itemset een antwoord te geven. Wanneer een leerling meer dan 55% van de itemset niet gemaakt heeft of vijf opgaven op een rij niet gemaakt heeft, wordt de itemset als *niet-gemaakt* beschouwd voor de kalibraties. Dit percentage is gebaseerd op ervaringen in het verleden. In deze situatie is het aannemelijk dat er samenhang in het ontbreken is (bijv. door het overslaan van een pagina). Als er geen rekening gehouden wordt met reeksen van ontbrekende waarden lijkt dat een schending op te leveren van lokaal stochastische onafhankelijkheid. Een dergelijke schending zorgt mogelijk voor een overschatting van de parameters, hetgeen onwenselijk is. Let wel, het gaat hierbij over het schatten van de kenmerken van de opgaven. Voor de analyses op populatieniveau en de schattingen op leerlingniveau – het veelvoudig overslaan van opgaven heeft meer met de leerling dan met de opgaven te maken – worden ruimere grenzen gehanteerd. Een itemset wordt dan als niet-gemaakt beschouwd als een leerling zes opgaven op een rij heeft overgeslagen, meer dan 80% niet ingevuld heeft, of een combinatie van de twee (wanneer meer dan 80% ontbreekt betreft dat ook meestal een reeks van zes of meer). Tot slot is er gekeken naar de maximale inclusie: een itemset wordt dan alleen als niet-gemaakt beschouwd wordt, als geen enkel item ervan gemaakt is. Over het algemeen wordt erkend dat bij die keuze te veel niet-serieuze afnamen alsnog meegenomen worden. Zo kunnen bijvoorbeeld ook administratieve fouten (per ongeluk een item ingevuld) als afname tellen. Analyses op die selectie worden daarom niet gebruikt. Het aantal leerlingen dat dit scheelt, ten opzichte van de keuze op 80% of meer, is ook beperkt.

Het voordeel van deze analyses van ontbrekende waarden op het niveau van de itemsets is dat de leerling op zich nog wel in de analyses meegenomen kan worden, alleen niet voor de itemsets die als niet gemaakt beschouwd worden. De items en de leerlingen worden geanalyseerd alsof aan hen de specifieke set opgaven niet is voorgelegd en ze een verkorte versie van de toets gemaakt hebben (minder dan vier itemsets gedaan). Het is dan mogelijk om met item response theorie (IRT) de opgaven over te slaan in de analyses. De toets zonder die opgaven is dus als een nieuwe versie van de toets te

beschouwen. Het gevolg is dat behalve de 12 toetsboekjes die in Tabel 3.2b volgens het design verondersteld worden, er nu meer '(toets)boekjes' (toetsversies) zijn. Hoeveel boekjes, en welke, hangt af van de keuze wanneer een itemset als gemaakt beschouwd wordt. Bij de kalibratiekeuze (itemset is niet gemaakt bij vijf opgaven of meer op een rij, dan wel bij meer dan 55% ontbrekend) levert dat dus andere boekjes op dan bij de populatie-schatterskeuze (zes of meer niet gemaakte opgaven op een rij; meer dan 80% ontbrekend). De 'nieuwe toetsboekjes' zijn dan over het algemeen door weinig personen gemaakt.

In Tabel 5.4 wordt een overzicht gegeven van de verschillende typen boekjes. Dat betreft om te beginnen de 12 toetsen volgens het design. In 11 van deze toetsen betekent dit dat er 4 itemsets als afgenomen beschouwd worden, in T01 zijn dat 2 itemsets. Daarnaast is het mogelijk dat leerlingen 4 itemsets maken, maar niet volgens het design. Dat is in deze afname niet gebeurd en wordt daarom niet weergegeven in dit overzicht. Ook bestaat de mogelijkheid dat er maar drie itemsets als afgenomen worden beschouwd; dus 2 RM en 1HR of 1 RM en 2HR. Ook twee itemsets, bijvoorbeeld als RM volledig is gemaakt maar HR niet of vice versa, of als van beide delen 1 itemset is afgenomen. Tot slot kunnen in de analyses leerlingen opgenomen zijn die één RM- dan wel één HR-itemset gemaakt hebben. De leerlingen bij wie geen enkele itemset als afgenomen beschouwd wordt, vallen buiten de analyses.

Ook als er minder dan vier itemsets afgenomen zijn, komt het in geen enkel geval voor – net zoals bij de combinaties van vier itemsets – dat het een combinatie van itemsets betreft die volgens het design niet mogelijk is. Een combinatie als RMis03 en HRis08 is dus niet gevonden. Aangezien het afhankelijk is van wanneer een itemset als niet-gemaakt beschouwd wordt, hoeveel en welke boekjes er zijn, en hoeveel leerlingen die boekjes maakten, hebben we in Tabel 5.4 definities gegeven aan de drie belangrijkste keuzevarianten per gemaakte itemset. Dat zijn de kalibratie-definitie (Kal: voor item-parameter-schattingen), de populatie-definitie (Pop: voor populatieschattingen), en de maximale definitie (Max). Deze laatste definitie wordt alleen als referentie gebruikt.

Tabel 5.4 – Opdeling leerlingen over boekjes op basis van als gemaakt beschouwde\* itemsets

Aantal itemsets		Aantal (N) leerlingen			Aantal boekjes			N leerlingen per boekje		
Totaal	RM / HR	Kal	Pop	Max	Kal	Pop	Max	Kal	Pop	Max
4 (2)	2/2 (design)**	2492	2601	2657	12	12	12	207,7	216,8	221,4
3	2 / 1	53	52	6	7	6	4	7,6	8,7	1,5
3	1 / 2	75	12	13	19	8	8	3,9	1,5	1,6
2	2 / 0	21	21	21	7	7	7	3,0	3,0	3,0
2	1 / 1	30	8	1	8	5	1	3,8	1,6	1,0
2	0 / 2	33	18	17	11	10	9	3,0	1,8	1,9
1	1 / 0	4	2	2	4	2	2	1,0	1,0	1,0
1	0 / 1	5	1	7	3	1	5	1,7	1,0	1,4
In analyses		<b>2713</b>	<b>2715</b>	<b>2724</b>	<b>71</b>	<b>51</b>	<b>48</b>	<b>38,2</b>	<b>53,2</b>	<b>56,8</b>
<b>0</b>	<b>0 / 0</b>	<b>11</b>	<b>9</b>	<b>0</b>	<b>1</b>	<b>1</b>	<b>0</b>			

\* Zie toelichting in de tekst boven de tabel.

\*\*Bij de meeste toetsen is de combinatie van 2 RM- en 2 HR-itemsets volgens design, behalve bij T01 voor Pro, waarbij 1/1 de designoptie is. Die toets valt hier onder 2/2 (design) omdat deze volgens het design afgenomen is

Het is evident dat wanneer er strengere eisen gesteld worden om te bepalen of een itemset als gemaakt beschouwd wordt, dit minder leerlingen in de 'design-boekjes' oplevert. Ook het aantal leerlingen dat geheel buiten de analyses valt is dan groter, maar het is goed te zien dat het aantal leerlingen dat extra



uitvalt bij de strengere Kal-selectie zeer beperkt is ten opzichte van de milde Pop-selectie. De 9 leerlingen die daar uitvallen hebben in iedere itemset maximaal 1 of 2 items mogelijk iets ingevuld, hetgeen niet als serieuze testafname kan gelden. Deze leerlingen kunnen ook beter buiten de verdere analyses gehouden worden.

Op basis van deze selectie zijn de toetsboekjes gemaakt die zijn gebruikt in verdere analyses. Ook met deze toetsboekjes zijn klassieke analyses uit te voeren. We presenteren hier de toetseigenschappen van de designboekjes die in de kalibratie gebruikt worden (Tabel 5.5). Naast het aantal items (N items) en het aantal leerlingen (N lln) per boekje geven we de gemiddelde p-waarde (GemP), de gemiddelde score (Gem(X)) en de standaardafwijking van de score (Gem(SD)). De score is hier het aantal goed gegeven antwoorden. Tot slot wordt Cronbach's alpha gerapporteerd<sup>11</sup>. Het is duidelijk dat bij de Pop- en Max-selecties de gemiddelde p-waarde en scores iets lager liggen, omdat daarin meer leerlingen met meer ontbrekende waarden die als 0 gescoord worden meegenomen worden. De verschillen zijn echter niet opzienbarend.

Tabel 5.5 – Klassieke toetsresultaten en leerlingenaantallen over designboekjes in de kalibratie

DesBk	Toets	N items	N lln	GemP	Gem(X)	SD(X)	Alpha
Bk 1	T01	22	72	35,5	7,8	3,0	,62
Bk 2	T02	44	155	54,8	24,1	6,7	,85
Bk 3	T03	44	130	40,9	18,0	6,6	,86
Bk 4	T04	44	140	43,2	19,0	6,5	,84
Bk 5	T05	44	153	49,9	22,0	6,8	,85
Bk 6	T06	44	157	48,0	21,1	7,9	,88
Bk 7	T07	44	174	53,4	23,5	7,1	,84
Bk 8	T08	44	177	57,5	25,3	7,3	,85
Bk 9	T09	44	253	64,5	28,4	8,0	,89
Bk 10	T10	44	423	59,2	26,1	6,6	,84
Bk 11	T11	44	264	47,8	21,1	7,0	,85
Bk 12	T12	44	394	50,4	22,2	7,1	,85
Totaal			<b>2492</b>	<b>51,3</b>			

N items: aantal items; N lln: aantal leerlingen; GemP: gemiddelde p-waarde; Gem(X) & SD(X) gemiddelde en standaardafwijking van de (ongewogen) score; Alpha: Cronbach's alpha

De gemiddelde p-waarde van de items in T01 is zeer laag. Het is aannemelijk dat dit vooral te maken heeft met de vaardigheid van de leerlingen die deze toets gemaakt hebben, aangezien al deze items ook voorkomen in T02 en die heeft weer juist een van de hogere gemiddelde p-waarden. Om hier een betere uitspraak over te kunnen doen, worden de IRT-analyses toegepast. Ook de betrouwbaarheid is het laagst in T01. Dat is niet verbazingwekkend aangezien deze toets ook maar uit de helft van de

<sup>11</sup> Bij de alpha zal weinig impact gevonden worden van extra samenhang binnen klassen; in iedere klas worden gemiddeld 4 verschillende boekjes ingezet, met gemiddeld 5 leerlingen van een klas die een en hetzelfde boekje maken. In 95% van de gevallen is het aantal leerlingen binnen een klas dat hetzelfde boekje maakt niet hoger dan 8. Als we per boekje kijken, blijkt dat Boekje 1 (dat alleen in pro ingezet wordt) een uitzondering is, met 5 klassen en gemiddeld 15 leerlingen in de klas dat het boekje maakt. Bij de overige boekjes zien we weer gemiddeld 5 leerlingen per klas per boekje. We zien hier ook een negatieve correlatie: hoe meer leerlingen per klas het boekje maken (dus hoe meer ICC een rol zou kunnen spelen), hoe lager de alpha. Al met al lijkt de invloed van ICC beperkt en lijken de waarden van alpha niet overschat door de impact van ICCs.

opgaven bestaat ten opzicht van de andere toetsen. Volgens de eerder genoemde COTAN-richtlijnen is de betrouwbaarheid van T01 voldoende voor het doel van de toets. De andere toetsen zijn met waarden boven de 0,80 (zeer) goed te noemen, gezien het doel van de toetsen. Desgewenst is voor alle items, ook met medeneming van de ‘nieuwe’ (niet-design) boekjes van een p-waarde te geven<sup>12</sup>. De gemiddelde p-waarde is een fractie hoger dan bij de TIA waarbij geen rekening gehouden was met de verwerking van de leerlingen met extreme patronen wat betreft ontbrekende waarden<sup>13</sup>.

#### 5.4 Over de toepassing van IRT

Voor de schaling maken we gebruik van item respons theorie (IRT). Zoals eerder aangegeven is dat noodzakelijk omdat er in dit peilingsonderzoek sprake is van een onvolledig afnamedesign: niet alle 2724 leerlingen maken alle 264 opgaven. Bovendien zijn itemsets toegewezen op basis van het verwachte vaardigheidsniveau. Kortom, we kunnen de scores van leerlingen niet zomaar met elkaar vergelijken. Met behulp van IRT-equivalering is het echter mogelijk om de scores van de leerlingen op één onderliggende meetschaal te plaatsen. Bij IRT geldt dat (1) de schatting van de vaardigheid van een persoon onafhankelijk gedaan kan worden van de set items die hij of zij heeft gemaakt; (2) itemkenmerken die gerelateerd kunnen worden aan de moeilijkheid en onderscheidend vermogen kunnen onafhankelijk geschat worden van de groep personen die de items hebben gemaakt, dus de itemkenmerken van een item blijven gelijk, ongeacht voor welke steekproef van personen deze berekend zijn; en (3) het mogelijk is om de passing van een IRT-model te onderzoeken, om zo te toetsen of de geobserveerde scores kunnen worden verklaard door één onderliggend construct. In dit peilingsonderzoek doen we dat met behulp van het IRT-metmodel *One Parameter Logistic Model* (OPLM; Verhelst et al., 1993; Verhelst & Glas, 1993). In dit model wordt de kans dat een bepaalde leerling een bepaald item goed beantwoordt, uitgedrukt als functie van de latente vaardigheid van de leerling op het betreffende domein enerzijds, en de moeilijkheid en het onderscheidende (onderscheidende) vermogen van het item anderzijds. In het geval van dichotome items, en in deze peiling komen alleen dergelijke items voor, is het enige alternatief voor goed beantwoorden het fout beantwoorden. Alleen scores 0 en 1 komen voor waarbij score 1 wordt gegeven bij een correct antwoord<sup>14</sup>. De itemresponsfunctie die hierbij hoort, kan bij dichotome items worden weergegeven met behulp van de volgende vergelijking:

$$P(X_i = 1|\theta) = P_i(\theta) = \frac{\exp [a_i(\theta - b_i)]}{1 + \exp [a_i(\theta - b_i)]}$$

waarbij  $P_i(\theta)$  de kans is dat een persoon met vaardigheid  $\theta$  item  $i$  correct beantwoordt. De positie van het item op de schaal wordt aangeduid met  $b_i$ . Deze positie wordt gedefinieerd door de benodigde vaardigheid die nodig is om precies een kans van 0,50 te hebben om het item correct te beantwoorden. Het onderscheidend vermogen van item  $i$  wordt aangeduid met  $a_i$  en geeft aan hoe goed het item onderscheid kan maken tussen personen met verschillende vaardigheidsniveaus. Dit onderscheidend vermogen is optimaal rond het punt  $b_i$ . Hoe hoger het onderscheidend vermogen van een item (i.e., hoe hoger de waarde van de discriminatieparameter  $a_i$  van een item), hoe beter het item dus kan onderscheiden of een leerling een vaardigheid boven of onder de waarde  $b_i$  heeft. In de modelpassing – waarover later meer – bleek dat de a-parameters over de verschillende opgaven variëren, waardoor

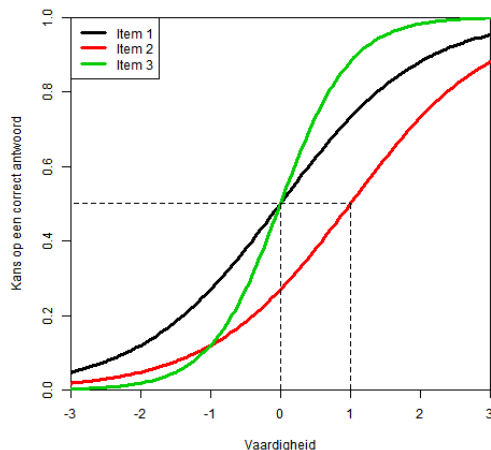
<sup>12</sup> Zie ‘PRV.Alle\_items\_Rekentaken.sav’ en ‘PRV.OPLM.xlsx’, tabblad ‘CTT ps en verschil ML SL en AL’.

<sup>13</sup> Was 50,7; zie ‘220725\_PRV.TIA\_overzicht\_NEW\_JV.xlsx’.

<sup>14</sup> Behalve dichotome items zijn polytome items ook mogelijk, waarbij er meer dan twee scores per item voorkomen. Bij het meten van de vaardigheid zijn deze opgaven niet gebruikt, maar bij de vragenlijsten zijn er wel polytome itemscores.

een Rasch model waarin verondersteld wordt dat de a-parameters voor alle items gelijk zijn, geen goede omschrijving zou geven.

In Figuur 5.1 worden drie verschillende itemresponsfuncties weergegeven voor het OPLM. In de grafiek is de locatie van een item af te lezen aan de positie van de itemcurve op de vaardigheidsschaal, waarbij de locatie is weergegeven met een stippelijijn, en correspondeert met de vaardigheidsscore die hoort bij een kans van 0,50 om het item juist te beantwoorden.



*Figuur 5.1 – Itemrespons functies voor het OPLM*

In de grafiek is te zien dat item 2 een hogere vaardigheid nodig is om een kans van 0,50 te hebben om het item goed te maken dan items 1 en 3. Wanneer de a-parameters identiek zijn (item 1 en 2) zien we dat voor twee personen met dezelfde vaardigheid, en dat voor vaardigheid, geldt dat de kans om item 1 goed te beantwoorden altijd hoger is dan item 2. Volgens dit model is voor iedereen is item 2 dus moeilijker dan item 1. Bij item 3 verschilt het onderscheidend vermogen van dat bij item 1 en 2. De kans het item 3 goed te maken is voor het grootste deel van de vaardigheidsrange groter is dan bij item 2, maar dat dit niet geldt voor de zeer laag-vaardige leerlingen (vaardigheid lager dan -1). Dan is de kans op een goed antwoord bij item 2 groter. Bij item 1 en 3 is te zien dat deze dezelfde locatie  $b=0$  hebben, maar dat voor leerlingen met een lagere vaardigheid item 1 gemakkelijker is, en voor leerlingen met een hogere vaardigheid item 3. Ook dit komt doordat de twee opgaven verschillen in onderscheidend vermogen. Bij item 3 is dat onderscheidend vermogen hoger dan bij item 1: item 3 maakt sterker onderscheid tussen leerlingen met een vaardigheid onder  $b=0$  en boven  $b=0$  dan item 1. Dit komt omdat de kans op het correct beantwoorden van het item sneller toeneemt rond locatie  $b=0$  naarmate de vaardigheid hoger is. Bij item 1 neemt deze kans minder snel toe naarmate de vaardigheid stijgt, en dus discrimineert dit item in vergelijking met item 3 minder goed tussen personen op dat deel van de vaardigheidsschaal. In Figuur 5.1 is te zien dat alle itemresponsfuncties een stijgende lijn laten zien, wat betekent dat het onderscheidend vermogen voor al deze items positief is. Dit betekent dat hoe hoger de vaardigheid, hoe groter de kans om het item correct te beantwoorden.

Naast inhoudelijke redenen, is het onderscheidend vermogen van een item vaak een van de redenen om een item wel of niet uit een schaal te verwijderen. Als items gevonden worden met een negatief onderscheidend vermogen, dan kunnen deze uitgesloten worden van analyses. Deze items hangen namelijk negatief samen met de te meten vaardigheid. Het betreft dan een item vaker correct wordt gemaakt door personen met een lagere vaardigheid dan door personen met een hogere vaardigheid,

wat in strijd is met het gebruikte IRT-model. Een dergelijk item **verlaagt** de betrouwbaarheid van de meting<sup>15</sup>.

Als opgaven een laag onderscheidend vermogen hebben, maken deze over de gehele schaal weinig onderscheid tussen vaardige en minder vaardige leerlingen. De bijdrage aan de betrouwbaarheid van de meting is beperkt. Over het algemeen wil men dergelijk opgaven niet in een toets opnemen omdat bij het meten van individuele leerlingen het opnemen van deze items niet efficiënt is (kost toetstijd, levert weinig informatie). Bij een peiling zoals deze is het verwijderen niet direct nodig zolang het item inhoudelijk relevant is. Het gewicht bij de schatting van de vaardigheid is dan ook beperkt. Aangezien bij een peiling geen beslissingen op individueel niveau worden genomen, maar naar prestaties op groepsniveau is het ook vanuit het oogpunt van transparantie niet direct noodzakelijk om alle opgaven een gelijk gewicht mee te geven. Daarnaast merken we op dat het in een peiling juist ook relevante informatie is als duidelijk is welke items een laag dan wel hoog onderscheidend vermogen hebben. Op basis van de inhoud van het item is het interessant te evalueren bij wat voor soort opgaven vaardiger leerlingen duidelijk beter presteren dan minder vaardige, en bij welke opgaven de mate van vaardigheid maar beperkte invloed heeft op het goed maken van de opgave. Alleen als een opgave negatief samenhangt met het gemeten construct, of als er duidelijke inhoudelijke redenen zijn om een item niet mee te nemen, zou een item echt verwijderd moeten worden. Als een opgave op een belangrijk deel van de vaardigheidsschaal (en dus niet de gehele) een negatieve relatie laat zien tussen de vaardigheid en de kans op het goed maken van een item, kan dat ook tot verwijdering van de opgave leiden. Dit dient in overleg met de inhoudsdeskundige te gebeuren.

Naast de naam van een model is OPLM ook de naam een softwarepakket (Verhelst, Glas & Verstralen, 1993) om het model mee toe te passen. In het model worden de a-parameters als 'gekend' beschouwd. Zodoende is kenmerkend voor het programma OPLM dat de discriminatieparameters van de items niet net als de moeilijkheidsparameters geschat worden, maar *a priori* als constanten worden ingevoerd<sup>16</sup>. Door deze keuze kan de (gewogen) ruwe score van een leerling op de toets direct uit de data berekend worden, zodat het mogelijk wordt om de itemparameters te schatten met *conditional maximum likelihood* (CML). In de analyses voor deze peiling is dat een belangrijk voordeel, omdat CML-schatting van itemparameters niet gepaard gaat met assumpties over de vaardigheidsverdeling in de populatie, over de wijze van steekproeftrekking, of over de toewijzing van toetsvarianten in een incompleet design (Eggen & Verhelst, 2011). Bij de toepassing van het OPLM zijn de waarden van de a-parameters altijd discrete waarden, daar waar bij een twee-parameter logische model (2PL-model) meer variatie mogelijk is. Bij de schattingen van de leerlingvaardigheden functioneren de a-parameters als gewichten waarmee de itemscores gewogen worden om de vaardigheid van de leerling te kunnen bepalen.

OPLM voert tevens een aantal statistische toetsen uit op grond waarvan we kunnen bepalen of het OPLM-model voor een betreffende schaal een adequate beschrijving geeft van de data. Belangrijk zijn de zogenaamde itemgeoriënteerde S-toets en de R1c-toets voor het gehele model. De S-toets is gebaseerd op de verschillen tussen de geobserveerde en verwachte proporties antwoorden in homogene scoregroepen. Als de *p*-waarden (overschrijdingskansen) voor de S-toetsen uniform verdeeld zijn in het interval [0,1] concluderen we dat het model past (Verhelst, Glas, & Verstralen, 1993).

---

<sup>15</sup> De absolute hoogte van de a-parameter is weinig betekenisvol wanneer de standaardafwijking van de verdeling niet gekend is. Deze hoeft echter niet gekend te zijn om de richting van de a-parameter (positief dan wel negatief) te kunnen interpreteren. Als een a-parameter negatief verondersteld is, dan heeft dit item negatieve impact heeft op de betrouwbaarheid. Mits inhoudelijke te verantwoorden, kan een aanpassing van de scoringsregel een opgave dat een negatieve a-parameter had, wel positief onderscheidend maken wanneer de parameters daarna opnieuw geschat worden.

<sup>16</sup> Er zijn kanttekeningen te plaatsen bij de wijze waarop dat in OPLM gebeurt. Duidelijk is echter dat de schatting van de b-parameters los staat van de inschatting van a-parameters.

Per opgave kunnen de geobserveerde en verwachte proporties antwoorden ook visueel onderzocht worden. Als deze porties met elkaar verbonden worden kan deze visuele inspectie gezien worden als de evaluatie van de 'geobserveerde itemresponse functie'<sup>17</sup> die vergeleken wordt met de volgens het model geschatte itemresponsefunctie. Met deze visuele inspectie aan de hand van een programma als WOPLOT kan naast de toetsing van de significantie door middel van de S-statistic ook de ernst van de misfit geïnterpreteerd worden.

De *R1c*-toets voor het gehele model heeft dezelfde onderliggende rationale als de S-toets. De waarde van de *R1c*-toets is zeer gevoelig voor afwijkingen van het model, waardoor ook bij relatief weinig observaties vaak al snel aangeeft dat het IRT statistisch model niet past bij de data. We kunnen ook kijken naar de bijdrage van een item aan de *R1c*-waarde.

Mocht er sprake zijn van statistische misfit is het ook mogelijk om robuustheidsonderzoek uit te voeren om de impact van een mogelijke modelschending te evalueren. Het is in alle gevallen mogelijk dat er opgaven zijn die niet geheel passen in het model. De keuze kan zijn om dergelijke opgaven uit de schaal te verwijderen. Dat kan echter inhoudelijke consequenties hebben omdat hierdoor ook de inhoud van de schaal wordt aangepast. Binnen peilingsonderzoek is het verstandig het verwijderen van opgaven te vermijden. Alle resultaten zijn interessant. Echter, het is dan wel noodzakelijk dat de slechter passende opgaven de schaal niet sterk beïnvloeden en schattingen met behulp van het model wel correct zijn. Om die reden wordt er bij misfit ook gewerkt met robuustheidsonderzoek waarbij de schattingen van het model vergeleken worden met de resultaten van de observaties.

Doordat de toetsen aan specifieke populaties toegewezen worden, kan, als nodig, hier per toets (of een kleine clustering van toetsen) een populatieschatting gemaakt worden. Dat gebeurt met gefixeerde itemparameters en vervolgens het toepassen van het programma OPML. Vervolgens kan met OPLAT (zonder data) met de schaalparameters en de populatieparameters geschat worden wat de klassieke kenmerken van de toets zouden zijn. Als deze overeenkomen met de werkelijk gevonden resultaten bij de klassieke analyses kan met het model gewerkt worden. Bij (sterke) afwijkingen kan er eventueel alsnog voor gekozen worden om opgaven te verwijderen.

## 5.5 Keuze van de te analyseren schalen: toepassing van IRT in dit onderzoek

### 5.5.1 Ter inleiding

Bij het toepassen van IRT-analyses is het van belang duidelijk te hebben welke schalen gevormd moeten worden. De opgaven hiervoor zijn alle op vijf manieren getypeerd, waarbij we het onderscheid tussen openopgaven en meerkeuzeopgaven zeer beperkt hebben beschreven. Dat onderscheid in vraagvorm (open/gesloten) is ook geen criterium om tot verschillende schalen te komen. De overige vier kenmerken zijn dat wel.

Ten eerste kunnen de opgaven gegroepeerd worden op basis van de afnamecondities waarbij wel of niet gebruikgemaakt kan worden van een rekenmachine, hetgeen een schaal RM (met rekenmachine) en een schaal HR (hoofdrekenen; zonder rekenmachine) oplevert. Ten tweede vindt opdeling naar

---

<sup>17</sup> Bij een groep leerlingen met ongeveer dezelfde theta-waarden kan de p-waarde bepaald voor een item bij die groep kandidaten. Door de punten te verbinden van oplopende vaardigheidsgroepen wordt de geobserveerde itemresponse functie verkregen. Door die te vergelijken met de geschatte itemresponse functie is mogelijke misfit te interpreteren. Merk op dat eigenlijk alleen de geobserveerde punten geëvalueerd worden en niet de curve.

schalen plaats op basis van de inhoud van de opgaven. Hiertoe waren een vijftal schalen samengesteld, te weten schaal B (Getallen), schaal C (Verhoudingen), schaal D (Meten en meetkunde), schaal E (Verbanden) en schaal N1 (Probleemoplossen). Het is mogelijk om deze opgaven op basis van hun inhoud verder op te delen naar subcategorieën, maar dat is vooralsnog nagelaten. Bij de conclusies zal beschreven worden waarom het ook niet aannemelijk zal zijn om een dergelijke opdeling verder te onderzoeken. Ten derde is er de opdeling naar referentieniveau met een viertal mogelijke schalen (de 1F, 2F-, 1S- en 3F-schaal). Tot slot is de vorm van de opgaven (met of zonder context) een andere opdeling van de opgaven die tot de schalen Context en Kaal leiden.

Naast de schaling op basis van de hoofdcriteria, is het ook mogelijk om de combinaties van kenmerken mee te nemen als basis voor schaling. De combinatie van inhoud en referentieniveau kan zo tot 20 schalen leiden, van B-1F tot en met N1-3F. Als we daar ook nog het onderscheid HR/RM en Context/Kaal op toepassen zou dat in theorie tot 80 schalen kunnen leiden. Het grote nadeel is echter dat elk van deze schalen bij een totaal aantal van 264 opgaven wel tot zeer weinig opgaven per schaal zou leiden, waardoor de betrouwbaarheid van de meting in het geding komt. Ook blijkt dat bij 32 van de potentieel 80 schalen geen van de 268 items getypeerd zouden zijn als horend bij die schaal (er zijn bijvoorbeeld geen kale opgaven die met een rekenmachine uitgevoerd moeten worden om welke vaardigheid dan ook op het niveau 1F of 1S te meten). Maar ook de 48 overgebleven mogelijke schalen met gemiddeld 5 á 6 items is ook niet zinvol.

Een andere nadeel is dat het design niet geschikt is om dergelijke kleine verzamelingen opgaven als aparte schalen mee te nemen. Bij een schaal van 12 opgaven heeft iedere toets ook gemiddeld maar twee opgaven, waarbij er sprake is van telkens een opgave die overlapt tussen twee toetsen. In dergelijke gevallen is het vaak niet mogelijk de parameters goed te schatten. De schaling is dan ook als eerste uitgevoerd op basis van de opdeling op de hoofdcriteria. De in deze paragraaf beschreven resultaten illustreren dat een verdere opdeling niet heel zinvol blijkt.

### 5.5.2 Schaling op basis van itemkenmerken: losse schalen

Voordat we de schaling evalueren, kijken we eerst nog naar het aantal observaties per opgave. Bij de items in T01 is dat nu 233 observaties per opgave. Dat is niet veel, maar genoeg om schattingen op te doen. De mediaan is 356 observaties per opgave, en het gemiddelde 433. Bij deze schaling kijken we eerst of het Rasch-model een goede beschrijving geeft van de items in de data. Dat houdt in dat we aannemen dat de a-parameters van alle opgaven per schaal aan elkaar gelijk zijn.

Een overzicht van de belangrijkste passingsmaten is gegeven in Tabel 5.6. De resultaten in de tabel worden gegeven per schaal, gesorteerd per opdeling. Dat zijn als eerste de twee schalen op basis van de afnameconditie, daarna op basis van de inhoud, als derde op basis van de referentieniveaus en tot slot op basis van de vorm. Per schaal wordt bij N gegeven van hoeveel opgaven de S-statistics bepaald worden. Het aantal opgaven waar een S-statistic voor bepaald is, is soms lager dan het totaal aantal opgaven op de schaal. Met name wanneer een opgave heel moeilijk of heel makkelijk is, is deze statistics namelijk niet te bepalen. Voor hoeveel opgaven er geen S-statistic beschikbaar is, is gegeven in kolom met de kop 'x'. Van de opgaven waarvan de S-statistic wel te berekenen is, is bepaald of de afwijking significant is. In kolom P01 is het aantal items gegeven met een statistische misfit met  $p < .01$ . Kolom P05 geeft per schaal het aantal items met misfit tussen  $p = .01$  en  $p = .05$ . Kolom P10 geeft het aantal items met een misfit met een  $p$  tussen  $p = .05$  en  $.10$ .

Tabel 5.6 – Overzicht van passingsstatistieken bij de losse schalen onder het Rasch model<sup>18</sup>

Rasch	N	P01	P05	P10	%<.01	%<.05	%<.10	x	R1c*	df	p	f
RM	102	14	11	11	14%	25%	35%	1	986	589	0	1,29
HR	155	19	17	12	12%	23%	31%	6	1622	940	0	1,31
B	69	7	4	5	10%	16%	23%	4	523	355	0	1,21
C	72	11	7	4	15%	25%	31%	0	629	386	0	1,28
D	48	5	2	2	10%	15%	19%	1	430	252	0	1,31
E	47	3	7	4	6%	21%	30%	1	413	234	0	1,33
N1	18	1	0	1	6%	6%	11%	4	49	31	,0220	1,25
1F	61	5	5	11	8%	16%	34%	4	517	337	0	1,24
2F	105	11	7	6	10%	17%	23%	2	985	642	0	1,24
1S	50	8	3	6	16%	22%	34%	0	524	303	0	1,31
3F	42	12	9	1	29%	50%	52%	0	456	219	0	1,44
Context	230	32	23	20	14%	24%	33%	4	2457	1446	0	1,30
Kaal	25	2	2	2	8%	16%	24%	4	134	83	,0003	1,27

Ook de rest van de items worden geëvalueerd, maar we kijken nu vooral naar de opgaven die mogelijk niet passen. Vervolgens wordt gekeken hoeveel procent van de items per schaal een misfit met een  $p < .01$  ( $\%p < .01$ ),  $p < .05$  ( $\%p < .05$ ) of  $p < .10$  ( $\%p < .10$ ) heeft. Op basis van toeval kan dat respectievelijk 1, 5 en 10% zijn. Als die percentages (veel) groter zijn, past het model niet echt goed.

In Tabel 5.6 zien we dat onder het Rasch-model bij de meeste schalen de passing niet goed te noemen is. Het aantal opgaven met een misfit is duidelijk hoger dan te verwachten onder toeval. De resultaten met de R1c-statistics bevestigen het beeld dat het Rasch-model niet past. Op schaal N1 na hebben de schalen een sterk significante misfit. Zoals gezegd, is de R1c-statistic wel een heel gevoelige maat en zal vaak een slechte passing vertonen. Om een inschatting te maken van hoe erg de misfit is, is ooit een heuristisch ontwikkeld waarbij de R1c statistic gedeeld wordt door het aantal vrijheidsgraden (df). De wortel van deze ratio levert de  $\phi$ -coëfficiënt op. De ervaring leert dat als dat een waarde hoger dan 1,225 oplevert dat is het model niet robuust voor een matige passing. Dat is bij de meeste schalen hier het geval. Als we de figuren per item bekijken dan zien we dat de slechte passing er in ligt dat de opgaven per schaal niet een gelijk onderscheidend vermogen hebben en dat een model waarin de a-parameters in waarde mogen variëren beter zou moeten passen. Om die reden gaan we verder met de passing van een OPLM.

Bij de toepassing van OPLM mogen de a-parameters variëren, waarbij de laagste a-parameter gelijk aan  $a=1$  is. Hierin willen we de variatie in a-parameters niet te extreem maken, aangezien de a-parameters ook als itemgewichten werken. De hoogste a-parameter stellen we daarom in op  $a=5$ . De resultaten per schaal worden gegeven in Tabel 5.7. Bij schalen met gemiddeld minder dan 5 opgaven per toets en minder dan 3 overlap was een geheel vrije parametersschatting problematisch. Dat betreft de schaal N1 (probleemoplossen) en Kaal. De koppeling over de toetsversie was te zwak om binnen een

<sup>18</sup> Zie 'PRV.Figuren\_en\_Tabellen.xlsx'; de basis voor deze gegevens zijn de CML-analyses betreffende de OPLM-bestanden 'PRV#\_2.\*', met # als de indicatie voor de schaal (# = RM/HR; sB/sC/sD/sE/sN; 1F/2F/1S/3F; Co/Ka). De analyses waarbij de a-parameters niet gemaximeerd zijn op  $a=5$  zijn te vinden als PRV#\_1.\*. De passing is amper beter.



OPLM met alle parameters vrij de schaling te bepalen. Door nu enige parameters te fixeren was het wel mogelijk.

Tabel 5.7 – Overzicht van passingsstatistieken bij de losse schalen onder het OPLM<sup>19</sup>

OPLM	N	P01	P05	P10	%<.01	%<.05	%<.10	x	R1c*	df	p	f
RM	103	0	7	7	0%	7%	14%	0	663	603	,0447	1,05
HR	155	3	7	3	2%	6%	8%	6	1099	1004	,0186	1,05
B	67	1	1	3	1%	3%	7%	6	492	428	,0165	1,07
C	71	0	5	4	0%	7%	13%	1	424	408	,2784	1,02
D	46	0	2	1	0%	4%	7%	3	336	271	,0037	1,11
E	45	1	0	3	2%	2%	9%	3	344	238	,0000	1,20
N1*	21	0	0	2	0%	0%	10%	1	48	35	,0705	1,17
1F	61	0	2	4	0%	3%	10%	4	406	382	,1894	1,03
2F	104	1	0	10	1%	1%	11%	3	746	673	,0251	1,05
1S	50	1	2	1	2%	6%	8%	0	335	305	,1164	1,05
3F	42	0	4	3	0%	10%	17%	0	260	219	,0284	1,09
Context	229	4	13	18	2%	7%	15%	5	1690	1485	,0001	1,07
Kaal*	28	0	1	0	0%	4%	4%	1	102	97	,3318	1,05

\* Zie opmerking in de tekst

De passing van de opgaven is met dit model bij alle schalen aanzienlijk beter. Bij de S-statistic zien we dat er in sommige schalen bij enkele items een misfit is van  $p < .01$ , maar dat valt binnen de bandbreedte van wat te verwachten valt onder toeval. Dat geldt ook voor  $p < .05$ , en  $p < .10$ , waarbij alleen de schaal voor 3F wat afwijkende resultaten laten zien. Als we verder de visuele inspectie uitvoeren zien we dat de passing weliswaar niet perfect is, maar geen reden om opgaven te verwijderen. Verderop in dit rapport laten we hiervan voorbeelden zien. Geen van de opgaven had ook een negatieve a-parameter, waardoor we ervoor gekozen hebben geen enkel item uit een van de schalen te verwijderen voor de verdere analyses.

Als we kijken naar de R1c-Statistics zien we dat er maar een schaal over is met een sterke misfit (E; Verbanden), maar de heuristiek laat een waarde onder de anderhalf zien. Bij 11 van de schalen is de schatting statistisch niet significant bij  $p < .01$ . Dat is, zo leert de ervaring, bij de R1c-statistics een opmerkelijk goed resultaat. We gaan er vanuit dat we verdere analyses kunnen uitvoeren met deze schalen. Doordat deze schalen zo goed functioneren is er geen reden om aan te nemen om een veel diepere verfijning in de schaling aan te brengen, door schalen verder uit te splitsen naar subschalen. De passing zal waarschijnlijk niet beter worden. Wel is het interessant om verder te kijken naar de samenhang tussen de schalen om te onderzoeken of het mogelijk is om een enkele rekenschaal te maken waarbij alle opgaven op een schaal geplaatst worden.

<sup>19</sup> Zie 'PRV.Figuren\_en\_Tabellen.xlsx'; de basis voor deze gegevens zijn de CML-analyses betreffende de OPLM bestanden PRV#\_2.\*, met # de indicatie voor de schaal (# = RM/HR; sB/sC/sD/sE/sN; 1F/2F/1S/3F; Co/Ka).



### 5.5.3 Samenhang uitkomsten kenmerken van de taakuitvoering

Naast deze vier sets van losse schalen wordt ook een schaal geconstrueerd worden waarbij getracht wordt deze vaardigheden samen te nemen. Om inzicht te krijgen of dat mogelijk is, is de belangrijkste bron van informatie de samenhang tussen de schalen die we hebben gemaakt. We onderzoeken de correlatie tussen de schalen, waarbij we ook rekening houden met de meetprecisie (of meetfout) per schaal.

Dit onderzoek naar de samenhang tussen schalen vindt plaats per type opdeling. De schalen die gemaakt worden op basis van de afnameconditie (RM/HR) worden onderling vergeleken, evenals die op basis van de inhoud (5 schalen), referentieniveau (4 schalen) en vorm (2 schalen). Het is niet zinvol is om schalen van een verschillend type met elkaar te correleren. Iedere opgave zit in principe in vier verschillende schalen: één per type opdeling. Zo zit bijvoorbeeld item 10378 (het eerste OPLM item) in de schaal RM, schaal B, schaal 1F en de Context-schaal. Dit principe geldt ook voor de 263 overige items, waardoor er sprake is van een autocorrelatie tussen deze schalen. Hoe groot die is, hangt af van de mate van overlap in opgaven. Alleen schaal N1 en de schaal Kale opgaven hebben geen overeenkomstige opgaven, hetgeen inhoudelijk te verwachten is.

Met behulp van het programma Cordiml, ontwikkeld door Verhelst (2008), is voor de vier sets van schalen de latente correlatie bepaald tussen de schalen. In dit programma worden per schaal de leerlingprestaties, de parameters en de populatieverdeling meegenomen om deze correlatie te bepalen. Het idee hierachter is gelijk aan de correlatie na attenuatiecorrectie: wat is de correlatie tussen twee schalen, rekening houdend met de betrouwbaarheid van de meting (dus gecorrigeerd voor de meetfout)? Bij een onvolledig design zouden de correlaties na attenuatiecorrectie voor ieder boekje apart uitgerekend kunnen worden. Voor de 12 boekjes van het hoofddesign zou dit tussen de schalen op basis van afnameconditie, inhoud en vorm een twaalfstal verschillende gecorrigeerde correlaties opleveren. Deze 12 verschillende waarden zouden vervolgens samengevat moeten worden.

Naast de latente correlaties geven we voor de volledigheid ook de geobserveerde correlatie. Als we die op basis van de scores zouden geven, dan zou dat per boekje moeten. De scores over de boekjes heen zijn niet te vergelijken gezien de verschillen in moeilijkheid van de toetsen en de toewijzing van de boekjes op basis van de ingeschatte vaardigheden van de leerlingen. Als we correlaties per boekje geven dan levert dat, als we ons alleen tot de 'designboekjes' beperken, in ieder geval 12 (waarschijnlijk) verschillende waarden komen op. Daarnaast is de spreiding in vaardigheid per boekje beperkt, vanwege werkwijze in de toewijzing van de toetsen. Om die reden is – om de geobserveerde correlatie over alle boekjes heen te bepalen – de correlatie tussen de WML-schattingen van de leerlingen gebruikt. Hoewel de WML de latente trek betreft, wordt dit toch gezien als de geobserveerde correlatie aangezien deze nog meetfout bevat. Bij de latente correlatie corrigeren we daarvoor, zoals dat ook na attenuatiecorrectie het geval is.

Wanneer Cordiml toegepast wordt, kan binnen de toepassing van OPLM de correlatie direct voor de latente schalen bepaald worden. We maken hierbij gebruik van de kalibratieschalen die in de vorige paragraaf beschreven zijn. In Tabel 5.8 zijn de resultaten van deze analyses te vinden voor de schalen op basis van afnameconditie en vorm.

Tabel 5.8 – Samenhang tussen de twee afname-conditieschalen en de twee afnamevorm-schalen

Correlaties	Geobserveerd	Latent
Schalen Met Rekenmachine (RM) en zonder (HR)	0,86	0,95
Schalen Context-opgaven en Kale-opgaven	0,80	0,97

Deze correlatie zijn zeer hoog te noemen. De geobserveerde correlaties waren overigens ook al hoog (0,86 tussen RM en HR en 0,80 tussen Context en Kaal). Dit zijn de correlaties wanneer niet gecorrigeerd wordt voor meetfout.

Tabel 5.9 – Samenhang tussen inhoudsdomein en schalen<sup>20</sup>

Onder-diagonaal en cursief: Geobserveerde correlatie	Schaal				
	B	C	D	E	N1
<b>schaal B: Getallen</b>		0,97	0,94	0,95	0,96
schaal C: Verhoudingen	0,83		0,95	0,96	0,95
schaal D: Meten en meetkunde	0,78	0,81		0,94	0,95
schaal E: Verbanden	0,78	0,81	0,77		0,95
schaal N1: Probleemoplossen	0,64	0,66	0,65	0,63	

De geobserveerde correlaties van Probleemoplossen lijken niet zo hoog, en geven mogelijk de indruk dat analyses met deze afzonderlijke schaal wellicht nog interessante verschillen met andere schalen zouden kunnen opleveren. Echter, deze schaal is zeer kort, waarbij de meeste leerlingen 3 of 4 opgaven maken. De metingen bevatten daardoor relatief veel meetfout. De correlaties gecorrigeerd voor die meetfout liggen dan ook geheel in lijn met de correlaties van de andere inhoudsdomein-schalen.

Tot slot is er de opdeling van de schalen op basis van het referentieniveau. De schalen zijn anders dan de overige schalen in die zin dat bij de overige schalen alle leerlingen op ieder van de schalen te meten zijn. Door de specifieke opdeling van de opgaven naar verondersteld niveau bevatten niet alle toetsen opgaven van alle vier de niveaus (alleen 2F-items zitten in alle toetsen). Daar waar de correlaties hierboven gebaseerd zijn op alle toetsen, is dat bij opdeling van de schalen naar referentieniveau niet mogelijk. De 3F-opgaven zitten alleen in de laatste drie itemsets, die gemaakt worden door leerlingen van wie ingeschat is dat zij een hoger rekenvaardigheid hebben. De correlaties tussen de 3F-schaal en de 2F- en 1S-schaal kunnen dus alleen gedaan worden voor die leerlingen. De correlaties tussen de 1F- en 3F-schaal is niet te bepalen aangezien deze leerlingen geen 1F-opgaven voorgelegd krijgen: er zijn geen toetsboekjes met zowel 1F als 3F-opgaven<sup>21</sup>.

<sup>20</sup> Meer informatie staat er in 'PRV.LeerlingResultaten.Calibraties.xlsx', tabbladen 'CorrObserved' en 'CorrLatent'.

<sup>21</sup> Zie ook 'PRV.RefNiv.correlaties.xlsx'.

Tabel 5.10 - Samenhang tussen de referentieniveau-schalen

Onder-diagonaal en cursief: Geobserveerde correlatie	Boven-diagonaal: Latente correlatie			
	Schaal			
	1F	2F	1S	3F
Schaal met items op referentieniveau 1F		0,98	0,98	.
Schaal met items op referentieniveau 2F	0,75		0,99	0,93
Schaal met items op referentieniveau 1S	0,56	0,65		0,97
Schaal met items op referentieniveau 3F	.	0,49	0,64	

Het verschil tussen de geobserveerde en de latente correlaties lijkt hier groot. Om die reden is ook op de klassieke wijze per toetsboekje bepaald wat de geobserveerde correlatie tussen de gewogen scores is, en wat de correlatie is na attenuatiecorrectie. Die resultaten<sup>22</sup> liggen in lijn met wat er gevonden wordt met de IRT-gerichte werkwijze.

Als we naar de samenhang tussen al deze typen schalen kijken valt op dat latent de correlaties in alle gevallen hoog zijn. Alle correlaties zijn hoger dan 0,90 en gemiddeld zijn deze 0,96. Dit zijn voor rekenen geen zeer opmerkelijke resultaten aangezien deze bij andere metingen ook gevonden waren<sup>23</sup>. Dit betekent dat de schalen afzonderlijk niet heel veel informatie bevatten die niet ook al beschreven wordt door de algehele rekenschaal. Als we werken met de geschatte waarden per leerling kunnen er weliswaar verschillen over schalen gevonden worden, maar de kans is aanzienlijk dat deze eerder het gevolg van ruis (meetfout) zijn, dan dat dit werkelijke verschillen zijn.

Als aanvullende analyses hebben we ter controle ook gekeken of er op basis van een exploratieve factoranalyse met extractie op basis van principale componenten een reden was die ons van gedachten zou veranderen om alsnog afzonderlijk schalen te gebruiken. Merk op dat deze analyses per toets plaatsvinden. Omdat we met name geïnteresseerd zijn in maximaal vijf onderliggende schalen (inhouden), beperken we het aantal factoren tot zes. Deze zesde factor geeft de mogelijkheid dat als er ook meer dan vijf factoren zijn, daarvan enige indicatie gegeven kan worden. We maken hierbij gebruik van een varimax rotatie, omdat deze vaak meer inzicht geeft in de aanwezige factoren dan het ongeroteerde resultaat. Zonder verder dieper op de resultaten<sup>24</sup> in te gaan kwam hier, zoals ook te verwachten gezien de hoge latente correlaties, bij de FA niet duidelijk een onderscheid naar voren tussen de verschillende items op basis van de diverse itemkenmerken.

Al met al lijkt het aan te raden om ons te beperken tot een enkele rekenschaal voor de metingen waarin alle 264 opgaven zijn opgenomen. De laatste controle is of de opgaven ook daadwerkelijk alle op een schaal passen volgens het OPLM. Wanneer de passing van die schaal vergelijkbaar is met die van de afzonderlijke schalen zijn ook daar geen redenen om met metingen op afzonderlijke schalen te komen.

<sup>22</sup> Zie 'PRV.RefNiv.correlaties.xlsx'. Het grootste verschil dat gevonden wordt is dat de gemiddelde geobserveerde correlatie per toetsboekje tussen 1F en 2F met een waarde van 0,62 iets lager ligt dan die tussen de WML-schatters. Na attenuatiecorrectie komt dit op een gemiddelde latente correctie uit van 0,91 wanneer bij de geschatte correlaties aangehouden wordt dat deze nooit hoger dan 1 kunnen zijn. Dit verschil wordt met name veroorzaakt door de resultaten in boekje 8 waar slechts 4 opgaven op niveau 1F zitten. Zonder die resultaten liggen de geobserveerde (0,66) en latente correlatie (0,94) dicht in de buurt van de IRT-waarden. Ook als de latente correlatie niet per boekje bepaald wordt, maar gewerkt wordt met de gemiddelde correlatie over de boekjes heen en de gemiddelde correctiefactor, dan zijn de verschillen ook zeer klein. Hoe de berekening ook uitgevoerd wordt, alle gemiddelde gecorrigeerde correlaties zijn groter dan 0,90 en de gemiddelde gecorrigeerde correlatie tussen referentieniveau-schalen komt uit op een waarde boven de 0,95.

<sup>23</sup> Zie bijvoorbeeld Hemker, Kordes en Van Weerden (2010).

<sup>24</sup> Tabellen beschikbaar in 'PRV.LeerlingResultaten.Calibraties.xlsx'; tabbladen 'FA items' en 'FA per type itemkenmerk'.

#### 5.5.4 Schaling op basis van een enkele rekenschaal

Net zoals bij de losse schalen starten we hierbij met een Rasch-schaal (alle a-parameters gelijk aan 1), en vergelijken de passing met die wanneer we de a-parameters vrij laten, waarbij een vijftal waarden is toegestaan (1 – 5), net zoals bij losse schalen. De resultaten zijn gegeven in Tabel 5.11 die te vergelijken zijn met Tabel 5.6 (Rasch) en Tabel 5.7 (OPLM).

Tabel 5.11 – Overzicht passing bij een Rasch-model en een OPLM bij een enkele rekenschaal<sup>25</sup>

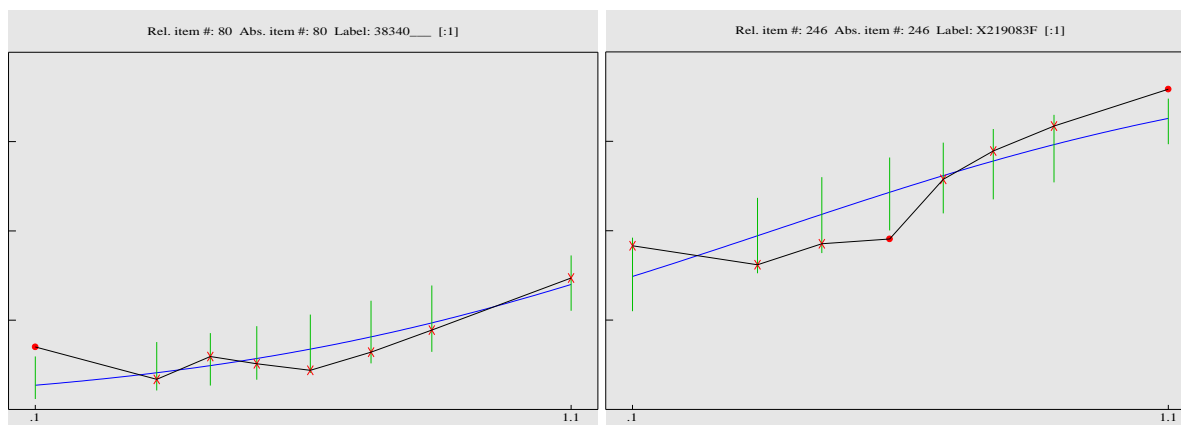
Rekenen	N	P01	P05	P10	%<.01	%<.05	%<.10	x	R1c*	df	p	f
Rasch	258	39	22	26	15%	24%	34%	6	2762	1628	0	1,30
OPLM	258	2	10	17	1%	5%	11%	6	1930	1671	0	1,07

Ook in dit geval zien we dat het OPLM (aanzienlijk) beter past dan het Rasch-model. De passing op basis van de S-statistics is ronduit goed te noemen. Kijken we naar de R1c-statistic dan lijkt de passing van de gehele schaal onder het OPLM slechter dan bij de afzonderlijke schalen. Dit is echter voor een belangrijk deel een vertekening. Bij kortere schalen die zo verspreid zijn over verschillende toetsen is de R1c aanzienlijk minder gevoelig dan wanneer er sprake is van een veel grotere overlap tussen de toetsboekjes, hetgeen we bij de totale schaal mee te maken hebben. Als we naar de bijdrage aan de R1c kijken van de individuele items, lijkt dat sterk op wat we ook vinden bij de resultaten voor S-statistic, namelijk 1% misfit op  $p < .01$  uit, 3% op  $p < .05$  en 11% op  $p < .10$ . Als we de eerder gebruikte heuristiek bekijken en de R1c-statistic delen door het aantal vrijheidsgraden is de waarde 1,16, hetgeen exact gelijk is aan de gemiddelde waarde van die heuristiek van alle afzonderlijke schalen. Ook als we op andere manieren de samenvatting van de losse schalen vergelijken met die op de schaal als geheel, kan gesteld worden dat de gehele rekenschaal zeer vergelijkbare statistieken laat zien als bij de losse schalen.

Zoals aangegeven zouden we hier kijken naar de passing van de individuele opgaven<sup>26</sup>. De twee slechtst passende opgaven wat betreft de S-statistic zijn opgaven 38340 (OPLM nummer 80; RM, E, 3F, Context) en 20211908\_3F (OPLM nummer 246; HR, N1, 3F, Context). Beide items zijn weergegeven in Figuur 5.2.

<sup>25</sup> Zie 'PRV.Figuren\_en\_Tabellen.xlsx'; de basis hiervoor zijn de CML analyses betreffende de OPLM bestanden PRVAL\_\_\$.\*, met \$=0 voor Rasch en \$= 2 voor 2PL, met a-parameters gemaximeerd. Alternatieve analyses waren \$=1, 2PL, a-parameters niet gemaximeerd; \$=3 is de schaal met 3 items uit (OPLMnrs 80, 238 en 246); \$=4 met 5 extra items (8 totaal) uit ( 58, 87, 91, 190, 193). Passing niet dusdanig beter dat dit het verwijderen rechtvaardigt. De passing bij \$=2 is zeker goed genoeg.

<sup>26</sup> Zie 'PRV.OPLM.xlsx'; tabblad 'per item OPLM'.



Figuur 5.2 – Visuele weergave van de passing van item 38340 (links) en 20211908\_3F (rechts)

Bij het linkerdeel van de figuur zien we dat de misfit plaatsvindt aan de onderkant van de vaardigheidsschaal. De groep minst vaardige leerlingen heeft hier een hogere kans om het item goed te beantwoorden dan verwacht onder het model. Dit wordt bij de losse schalen (hier dus RM, E, 3F en Context) ook gevonden. Bij de RM-, E- en de 3F-schaal is dat resultaat niet significant, maar bij de Contextschaal wel. Dat de Contextschaal resultaten oplevert die lijken op de totale rekenschaal valt te verwachten, aangezien de Contextschaal voor bijna 90% overeenkomt met de totale schaal. Opgave 38340 was een meerkeuzeopgave met een zeer aantrekkelijk, maar fout alternatief (de leerlingen die daarvoor kozen waren ook minder vaardig). Alleen de leerlingen die puur gingen gokken hadden daardoor een iets hogere kans. De impact op de gehele schaling voor deze laagvaardige groep is zeer beperkt, en aangezien de opgave voor de rest van de schaal goed functioneert is er geen reden dit item uit de schaal weg te laten.

Bij het rechterdeel van de figuur zien we ook bij het lagere helft van de vaardigheidsschaal misfit. Met de groep (relatief) laagvaardige leerlingen die beter presteert dan verwacht zou dit ook gokgedrag kunnen zijn, maar aangezien dit een open opgave is, is dat minder waarschijnlijk. In alle losse schalen is te zien dat op de onderste helft van de vaardigheidsschaal deze opgave iets minder onderscheidend is dan verwacht, er halverwege een zeer goed onderscheid gemaakt wordt en daarna de relatie tussen de vaardigheid en de kans op een goed antwoord weer iets minder sterk is. Bij N1 is dit overigens niet te zien omdat hier de passing niet bepaald kan worden. De afwijkingen zijn ook hier beperkt en geen reden om de opgave te verwijderen.

Naast de passing op de gehele schaal is er ook gekeken naar *Differential Item Functioning* (DIF) op basis van het verwachte rekenniveau. Hierbij is onderscheid gemaakt tussen vier groepen. De eerste is de groep die verondersteld het laagstvaardig te zijn die alleen opgaven op niveau 1F en 2F maken (T01-T05; 650 leerlingen), de middengroep die geen 3F-opgaven maken (T06-T09; 761 leerlingen) en de vaardigste groep die geen 1F-opgaven maken (T10-T12; 1081 leerlingen). Hierbij is geen DIF gevonden<sup>27</sup>: de gezamenlijke opgaven die door meer dan een groep gemaakt waren hadden dezelfde IRF, ongeacht de groep die de opgaven gemaakt had. De opgaven die de groepen overeenkwamen lijken dus voor deze verschillende vaardigheidsgroepen dezelfde betekenis te hebben.

<sup>27</sup> De middelste groep is ook nog opgedeeld in de groep die vooral 1F- en 2F-opgaven maakt en relatief weinig 1S-opgaven (T06-T07; 331 leerlingen) en de groep die geen of zeer weinig 1F-opgaven maakten, en vooral 2F- en 1S-opgaven (T08-T09; 430 leerlingen). Ook hier is geen DIF gevonden.

Het resultaat van de schaling in combinatie met de hoge latente correlaties is ook een indicatie dat hier zeer goed een enkele schaal te gebruiken is. Dat is wellicht zelfs te prefereren omdat het er op lijkt dat, wanneer losse schalen gebruikt zouden worden, verschillen tussen de schalen vooral ‘ruis’ (meetfout) zou kunnen zijn en niet zozeer werkelijke verschillen.

### 5.5.5 Betrouwbaarheid

Bij het bepalen van de schalen is het ook van belang dat de schalen betrouwbaar zijn. Dat is de reden waarom ook niet te veel met te kleine subschalen gewerkt zou moeten worden. We kijken hier nu even naar de betrouwbaarheid van de schalen in de klassieke opzet omdat dat aangeeft hoe betrouwbaar de vaardigheid daadwerkelijk gemeten is binnen die groep. Deze waarden zijn gegeven in Tabel 5.12. Daarin is per toets per subschaal aangegeven wat Cronbach’s alpha is per subschaal per toets<sup>28</sup>. Hierbij gebruiken we nu alleen de toetsboekjes zoals deze volgens het design zijn afgenomen, omdat voor die toetsen voldoende waarnemingen zijn.

Tabel 5.12 – Cronbach’s alpha<sup>29</sup> per toets, voor de gehele schaal en per subschaal

Toets		Conditie		Inhoud				Referentieniveaus					Vorm	
Design	Totaal	RM	HR	B	C	D	E	N1	1F	2F	1S	3F	Context	Kaal
Bk 1	,62	,58	,49	,56	,43	,22	,31		,60	,30			,66	,36
Bk 2	,85	,71	,79	,66	,70	,42	,49	<0	,79	,67			,84	,39
Bk 3	,86	,79	,79	,69	,69	,50	,54	<0	,78	,79			,86	,32
Bk 4	,84	,75	,75	,56	,72	,44	,46	,26	,73	,77			,85	,25
Bk 5	,85	,75	,76	,60	,73	,50	,52	,20	,60	,82			,85	,28
Bk 6	,88	,79	,81	,64	,76	,61	,50	,22	,60	,83	,50		,87	,46
Bk 7	,84	,74	,78	,56	,73	,49	,49	,14	,55	,74	,60		,82	,44
Bk 8	,85	,76	,78	,60	,75	,53	,45	,26	,33	,77	,72		,86	,06
Bk 9	,89	,81	,81	,68	,73	,48	,62	,47		,80	,80		,89	,18
Bk 10	,84	,70	,78	,59	,61	,58	,40	,23		,57	,72	,62	,83	,32
Bk 11	,85	,70	,80	,58	,63	,64	,38	,21		,46	,60	,78	,82	,61
Bk 12	,85	,73	,78	,65	,66	,52	,44	,37		,35	,51	,82	,81	,61
Minimum	,62	,58	,49	,56	,43	,22	,31	<0	,33	,30	,50	,62	,66	,06
Mediaan	,85	,74	,78	,60	,71	,50	,48	,22	,60	,76	,60	,78	,84	,34
Maximum	,89	,81	,81	,69	,76	,64	,62	,47	,79	,83	,80	,82	,89	,61
Gemiddelde	,84	,73	,76	,61	,68	,49	,47	,20	,62	,66	,64	,74	,83	,36

In een aantal gevallen is geen betrouwbaarheid te bepalen, simpelweg omdat niet alle referentieniveaus is alle toetsen worden afgenomen. Bij de schaal N1 is in het geval van toets 1 geen betrouwbaarheid te bepalen omdat er slechts één opgave ‘Probleemoplossen’ in die toets voorkomt. Bij de toetsen T02 en T03 zaten ook zeer weinig opgaven en was de verdeling van scores binnen die schaal zodanig dat er

<sup>28</sup> Zie ook ‘PRV.OPLM.xlsx’, tabblad ‘CTT toetsen subschaal’.

<sup>29</sup> Zie de voetnoot bij Tabel 5.5 over de mogelijke impact van de ICC, waarvan in deze peiling geen effect lijkt te zijn.

een negatief geschatte waarde van alpha gevonden werd. De Rit-waarden waren wel alle positief binnen die schaal en die toets, dus het is geen reden om die boekjes te verwijderen.

Bij de subschalen zien we in alle gevallen dat de betrouwbaarheid regelmatig lager is dan 0,60 (waarden rood afgebeeld in Tabel 5.12), hetgeen ook niet verwonderlijk is gezien de geringe aantallen opgaven per schaal. Alleen de Contextschaal is vergelijkbaar betrouwbaar, maar zoals al gezegd is dat logisch aangezien die bijna 90% van de opgaven van de gehele schaal bevat.

Bij het onderzoeken van de betrouwbaarheid gebruiken we de mogelijkheden die IRT biedt. Er zou gebruikgemaakt kunnen worden van de toetsinformatiefunctie. Een nadeel is echter dat er geen onduidelzinnige criteria beschikbaar zijn om de toetsinformatiefunctie te beoordelen. Daarom berekenen we ook de zogeheten MAcc-coëfficiënt (Verhelst et al., 1993). De MAcc-coëfficiënt functioneert op eenzelfde wijze als de klassieke betrouwbaarheidsmaten. Na definitieve keuze van de selectie van de schaal zal dit verder uitgewerkt worden. Deze zaken worden bepaald wanneer we de met behulp van MML met gefixeerde parameters de populatieschatting hebben uitgevoerd. We komen zodoende bij de betrouwbaarheid terug nadat deze gegeven zijn.

### 5.5.6 Populatieverdelingen en schattingen van leerlingvaardigheid

Met behulp van MML met gefixeerde parameters is het mogelijk om de latente vaardigheidsverdeling te geven van diverse groepen die deze toetsen gemaakt hebben. We maken hierbij voor nu onderscheid op basis van de indeling op basis van de toewijzing van de toetsboekjes. Naast de verdeling van alle leerlingen die we meenemen in de analyses (ALL) kunnen we – zoals beschreven in Paragraaf 5.4 – een onderscheid maken tussen leerlingen die een ‘designboekje’ (DES) gemaakt hebben, en de leerlingen waarvan gesteld kan worden dat bij hen een itemset ontbreekt (N-D). Voor de groep die een design-boekje gemaakt is kan per boekje de vaardigheid bepaald worden. Bij elk van de niet-design boekjes is hiervoor het aantal leerlingen per boekje te beperkt.

Deze design-boekjes kunnen ook weer geclusterd worden, om een idee te krijgen van de resultaten in verschillende vaardigheidsgroepen op basis van wat de docenten verwachten van de leerlingen. Hierbij onderscheiden we 6 groepen waarbij de labels gegeven zijn op basis welke groep dominant vertegenwoordigd is in de boekjes. Wanneer er geen groep dominant is dan is het combinatieboekje leidend. Dat levert de volgende opdeling op: praktijkonderwijs (Pro; bk01), vmbo-basis- en -kaderberoepsonderwijs (BB-KB; bk02-04); vmbo BB-KB en gemengde en theoretische leerweg (Bk05; BB-GT); vmbo GT (GT; bk06-08); vmbo-GT en havo/vwo (GT-HV; bk09); en havo/vwo (HV; bk10-12). Meer dan een indicatie is het op dit moment in de analyses niet. De werkelijke relatie met achtergrondvariabelen volgt pas bij de multilevel analyses (zie Hoofdstuk 8). Daarin zijn ook de vaardigheden van de leerlingen meegenomen van wie we nu stellen dat er gegevens ontbreken.

Naast deze opdeling is bij de beschrijving van de ontbrekende waarnemingen al aangegeven dat de selectie anders is wanneer we de item-parameters schatten en wanneer we geïnteresseerd zijn in de schattingen van leerlingen. In deze analyses worden nu de item-parameterschattingen gebruikt op de ruimere selectie van leerlingen zoals die eerder beschreven is. Ter vergelijking worden ook de resultaten gegeven voor populaties van de schattingen voor de item-parameters en voor de maximale selectie.

Om te weten of de ruimere selectie ook gebruikt kan worden, moet gekeken worden of de data met die selectie ook passen bij de item-parameters zoals we die geschat hebben. Die passing is ook zeer goed te noemen<sup>30</sup> en vergelijkbaar met de passing bij de kalibratie-selectie. Dat is overigens ook niet heel

<sup>30</sup> Zie ‘PRV.Figuren\_en\_Tabellen.xlsx’, tabblad ‘fit’.



verwonderlijk omdat die groepen ook weer niet extreem verschillen. Het is wel goed te weten dat de verschillen die er zijn compatibel zijn met de geschatte itemparameters.

De resultaten voor de drie selecties van leerlingen geven we voor de totale groep, en uitgesplitst naar de leerlingen die de 'designboekjes' gemaakt hebben en de leerlingen die minstens een itemset missen. Deze resultaten staan in Tabel 5.13.

Tabel 5.13 – Verdelingseigenschappen bij verschillende keuzen bij selecties van onderzoekspopulatie

Analyse	Selectie	Groep	N	Gemiddelde	St.Afw	se(m)	se(sd)
MX_F2	Maximaal	ALL	2724	0,174	0,606	0,012	0,009
SL_F2	Populatie	ALL	2715	0,182	0,603	0,011	0,009
AL_F2	Kalibratie	ALL	2713	0,188	0,601	0,011	0,009
MX_V2	Maximaal	Design	2657	0,185	0,597	0,012	0,009
SL_V2	Populatie	Design	2601	0,183	0,600	0,012	0,009
AL_V2	Kalibratie	Design	2492	0,195	0,595	0,012	0,009
MX_V2	Maximaal	Niet-Design	67	-0,328	0,814	0,098	0,084
SL_V2	Populatie	Niet-Design	114	0,150	0,660	0,062	0,048
AL_V2	Kalibratie	Niet-Design	221	0,103	0,668	0,046	0,034

MX: maximum aantal leerlingen ongeacht ontbrekende waarden; SL: leerlingen met (vrijwel) alle waarden ontbrekend uitgesloten; AL: kalibratie selectie, met beperkt aantal ontbrekende waarden toegestaan

Op het niveau van de totaalscores zien we dat de gemiddelden weinig verschillen. Als we de verschillen in termen van effectgrootte zouden evalueren is het een effect van 0,01 tussen de Kalibratie- en Populatieselectie. De aantallen leerlingen verschillen ook niet veel, maar daarbij moet gezegd worden dat de verschillen in de gemiddelden niet veroorzaakt worden door leerlingen die niet meetellen, maar door de wijze waarop de boekjes gevormd worden. Bij de Maximaal- en Populatieselecties tellen meer ontbrekende observaties mee als scores, die bij de selectie voor de Kalibratie als niet aangeboden worden beschouwd.

Zoals gesteld kijken we nu verder naar de Populatieselectie. Als we binnen de boekjes volgens het design kijken, worden de resultaten per boekje gegeven in Tabel 5.14. Bijna bij ieder hoger genummerd boekje is de prestatie beter dan bij een lager genummerd boekje. Bij boekje 3 (DES\_bk03) is lager dan het boekje erna, waarbij de resultaten voor boekjes 2, 3 en 4 erg dicht bij elkaar liggen. Ook is de prestatie in boekje 11 iets lager dan in boekje 10, al is het verschil daartussen ook wel erg klein. Dat juist hier de verschillen beperkt zijn is niet toevallig aangezien daar de toewijzing van de boekjes ook weinig verschilt. Bij de opdeling naar toewijzing op basis van dominantie, zoals hierboven beschreven, komen juist deze boekjes ook bij elkaar in de groep. De resultaten<sup>31</sup> daarvan zijn te vinden in Tabel 5.14.

<sup>31</sup> Naast de resultaten voor de Populatieselectie zijn Tabel 5.14 en 5.15 ook bepaald voor de Kalibratieselectie ('PRV.MML\_OPLM.xlsx', tabblad 'MML Calibratieset'). De verschillen zijn beperkt (tabblad 'MML all') en verschillen weinig per boekje.



Tabel 5.14 – Resultaten per boekje (N>20) bij de Populatieselectie van boekjes<sup>32</sup>

Groep	N	Gemiddelde	St.Afw	se(m)	se(sd)
DES_bk01	77	-1,02	0,24	0,03	0,03
DES_bk02	162	-0,51	0,29	0,03	0,02
DES_bk03	143	-0,60	0,33	0,03	0,02
DES_bk04	153	-0,51	0,32	0,03	0,02
DES_bk05	158	-0,30	0,30	0,03	0,02
DES_bk06	162	-0,14	0,34	0,03	0,02
DES_bk07	177	0,01	0,29	0,02	0,02
DES_bk08	178	0,17	0,29	0,02	0,02
DES_bk09	258	0,44	0,33	0,02	0,02
DES_bk10	432	0,59	0,30	0,02	0,01
DES_bk11	275	0,58	0,31	0,02	0,02
DES_bk12	426	0,74	0,30	0,02	0,01

Van deze groepen in Tabel 5.14 is duidelijk dat de 77 leerlingen die boekje 1 gemaakt hebben en als minst vaardige leerlingen (in praktijkonderwijs) geïdentificeerd kunnen worden, het laagste presteren. De standaardafwijking binnen deze groep is ook relatief klein. Het is een kleine, duidelijk onderscheiden groep. Kijken we naar de overige groepen, dan zien we dat deze alle een vergelijkbare standaardafwijking hebben. De olopende groepen verschillen alle min of meer op een standaardafwijking van elkaar wat betreft hun vaardigheid. De kleinste stap is die tussen de GT-HV en de HV groep.

Tabel 5.15 – Resultaten per dominante toewijzingsgroep<sup>33</sup> (Designboekjes; Populatieselectie)

Groep	Boekje(s)	N	Gemiddelde	St.Afw	se(m)	se(sd)
Pro	1	77	-1,02	0,24	0,03	0,03
Pro-BB-KB	2-3	458	-0,53	0,32	0,02	0,01
BB-KB-GT	4	158	-0,30	0,30	0,03	0,02
GT	5-7	517	0,02	0,33	0,02	0,01
GT-HV	8	258	0,44	0,33	0,02	0,02
HV	9-12	1133	0,64	0,31	0,01	0,01

Naast de populatieverdelingen zijn ook schattingen gedaan van de vaardigheid op het niveau van de leerling. Dat is op verschillende manieren uitgevoerd. De eerste werkwijze was door middel van WML-schattingen, gebruik makend van de parameters van de items. Deze werkwijze heeft als nadeel dat hierbij de standaardmeetfout ook een rol speelt. Bij de bepaling van de individuele vaardigheid wordt geen rekening gehouden met de onbetrouwbaarheid van de meting. Dat heeft als nadeel bij metingen op systeemniveau dat door deze meetfout de spreiding van de vaardigheid groter lijkt dan dat deze in de populatie is. Om dit tegen te gaan kan gewerkt worden met plausibele waarden (*plausible values*), die dat wel doen. We hebben hier 10 trekkingen uitgevoerd en gebruik gemaakt van de gemiddelde

<sup>32</sup> 'PRV.MML\_OPLM.xlsx', tabblad 'MML Populatieset' en 'MML all'; zie analyse *PRVSL\_F2*.

<sup>33</sup> 'PRV.MML\_OPLM.xlsx', tabblad 'MML Populatieset' en 'MML all'; zie analyse *PRVSL\_W2*.

waarde<sup>34</sup>. In dit geval waren de verschillen tussen de WML-schatters en de (gemiddelde) plausibele waarden te verwaarlozen<sup>35</sup> en is er verder gewerkt met de WML-schatters.

## 5.6 Referentieniveaus

In deze peiling worden vier referentieniveaus onderscheiden. We hadden bij de schaling gezien dat binnen deze populatie de opgaven die gerelateerd zijn aan de verschillende referentieniveaus onderling dusdanig samenhangen dat het mogelijk is een enkele schaal rekenen te gebruiken. Op deze ene schaal moeten nu de cesuurpunten bepaald worden die de rekenschaal in vijf delen opdeelt. Uit eerder onderzoek kennen we volgorde van deze cesuurpunten. Uiteraard is de vaardigheid die nodig is om 1F, 2F dan wel 3F te behalen (oplopend). De vaardigheid die nodig om het niveau 1S aan te tonen zit - volgens de psychometrische analyses - tussen de vaardigheid van 2F en 3F in. Dit levert zodoende de volgende categorieën leerlingen op:

- Onder niveau 1F
- 1F behaald, maar niet 2F
- 2F behaald, maar niet 1S
- 1S behaald, maar niet 3F
- 3F behaald

Hoeveel leerlingen in elk van deze categorieën vallen hangt af van een tweetal zaken: de rekenvaardigheidsverdeling van de onderzochte populatie en de posities van de cesuurpunten op de rekenschaal. Van de peiling is de vaardigheidsverdeling van de steekproef bekend, maar de cesuurpunten waren nog niet bepaald op deze schaal. In deze paragraaf is beschreven hoe dat onderzoek is uitgevoerd, en welke cesuurpunten dat oplevert op de hier gebruikte rekenschaal.

Om de cesuurpunten op de rekenschaal te plaatsen maken we gebruik van de kennis die we al hebben over de referentieniveaus. We kennen niet alleen de volgorde (zoals aangetoond in de psychometrische analyses) maar ook de waarden op referentieniveauschaal zoals deze gemaakt is in het VO. Deze schaal is te maken doordat de openbare sets waarop de referentieniveaus gedefinieerd zijn aan elkaar te relateren zijn. De grenspunten voor de verschillende cesuurpunten op die schaal zijn gegeven in Tabel 5.16.

---

<sup>34</sup> Deze schatting is een EAP (Expected a posteriori ability Estimator; Bock en Mislevy, 1982) die wordt verkregen door het gemiddelde te berekenen van de posterieure verdeling van de vaardigheid, waarbij de laatste wordt ingesteld als de eerdere verdeling maal de waarschijnlijkheidsfunctie. Hier: het gemiddelde van de 10 PV-trekkingen.

<sup>35</sup> De gemiddelde waarde op de vaardigheidsschaal in het geval van de plausibele waarden was 0,187, met een standaardafwijking van 0,600, terwijl bij de WML-schattingen van de persoonsparameters het gemiddelde en de standaardafwijking 0,182 en 0,608 waren. Doordat de metingen met (meestal) 44 rekenopgaven per leerling een behoorlijke betrouwbaarheid hebben, is de meetfout ook gering. De impact van het niet meenemen van die meetfout is zodoende ook beperkt.

Tabel 5.16 – Cesuurpunten op de referentieschaal en een alternatieve schaal

	1F	2F	1S	3F
Referentieschaal-theta	-0,836	0,340	0,604	1,995
Alternatieve schaal	0	42	51	100

De waarden op de referentieschaal zijn bepaald op basis van de gegevens van het referentie-onderzoek zoals deze zijn verkregen binnen het voortgezet onderwijs<sup>36</sup>. Deze waarden zijn niet direct toe te passen op de schaal die we hebben verkregen in ons peilingsonderzoek. Dat komt omdat vaardigheidsschalen geen absoluut nulpunt hebben, en ook de afstanden op de schaal niet op voorhand zijn gedefinieerd. Het betekent ook dat met behulp van een lineaire transformatie de schaal aan te passen is, zonder dat deze aan betekenis verliest<sup>37</sup>. Een voorbeeld hiervan is een herschaling waarbij de afstand tussen het cesuurpunt 1F en 3F op 100 gesteld is door het cesuurpunt 1F op 0 te stellen en dat van 3F op 100. De lineaire transformatie levert zo ook waarden op voor 2F en 1S die gegeven zijn bij 'Alternatieve schaal' in Tabel 5.15. Deze alternatieve schaal geeft gemakkelijk inzicht in de verhouding tussen de cesuurpunten op de schaal in het VO: de afstand tussen 2F en 1S is minder dan 1/10<sup>e</sup> van de afstand tussen 1F en 3F; de afstand tussen 1F en 1S, en tussen 1S en 3F is ongeveer gelijk.

Het kenmerk waarmee we met behulp van een lineaire transformatie de cesuurpunten van een schaal naar een andere schaal kunnen zetten, passen we ook toe bij de equivalering van de cesuurpunten naar de peilingsschaal voor rekenen. Daarbij transformeren we de waarden van de referentieschaal zo dat deze op de peilingsschaal passen. Om dit te kunnen doen, moeten we van twee punten de vertaling hebben van de ene naar de andere schaal. Hierbij kennen we door herschaling ook de andere twee cesuurpunten, zoals hierboven gedaan is om de alternatieve schaal van Tabel 5.15 te kennen. In dit geval doen we dit door gebruik te maken van de peiling in het primair onderwijs uit 2019 waarvan we de cesuurpunten voor 1F en 1S kennen, en waarvan 62 opgaven ook in de VO-peiling zijn gebruikt.

De cesuurpunten 1F en 1S zetten we van de Peilingsschaal PO op de VO peilingsschaal door middel van ware-score-equivalering (True Score Equating; zie bijvoorbeeld Kolen & Brennan, 2004; Lord, 1980; Hemker, 2022). Dat wordt hier toegepast bij de PO-rekenschaal, waarvan bekend is waar de cesuurpunten 1F en 1S liggen<sup>38</sup>. Voor iedere willekeurige verzameling opgaven op die schaal kunnen we bepalen wat de verwachte 'ware' score is op dat punt op de schaal. We weten dus ook voor de 62 opgaven die zowel in de PO-peiling van 2019, als in deze VO-peiling zitten, wat de verwachte score is bij de vaardigheid 1F en de vaardigheid 1S. Dit aantal opgaven is onafhankelijk van de schaal, en vormt de brug tussen de twee schalen, omdat we op de peilingsschaal voor VO kunnen bepalen welke vaardigheid op die schaal nodig is om die verwachte score te halen. Dit is weergegeven in Tabel 5.17.

<sup>36</sup> Deze waarden zijn niet eerder gepubliceerd. Ze zijn bepaald met behulp van de referentie-toetsgegevens, die ook de basis vormden voor de implementatie van referentieniveaus bij de Centrale Eindtoets. De waarden verschillen iets per populatie. De hier gerapporteerde waarden zijn gerelateerd aan VO, waarbij de cesuren voor alle vier de cesuurpunten bekend zijn.

<sup>37</sup> Let wel, als itemparameters bekend zijn, dan moeten deze itemparameters ook getransformeerd worden.

<sup>38</sup> Zie Buisman et al., 2021 (Paragraaf 8.1.1 'Prestatiestandaarden referentieniveaus rekenen').

Tabel 5.17 – Cesuurpunten voor 1F en 1S op drie verschillende schalen

Referentieniveau	1F	1S
Vaardigheid op peilingschaal PO 2019	0,0949	0,6571
Ware score op 62 ankeropgaven	31,0043	48,6749
Vaardigheid op peilingschaal VO 2022	-0,2639	0,4340

De waarden van 1F en 1S op de VO schaal worden gebruikt om de lineaire transformatie te bepalen. Door middel van deze lineaire transformatie met vermenigvuldigingsfactor  $a=0,484653$  en additieve factor  $b=0,141270$  kunnen de waarden van de Referentieschaal-theta omgezet worden naar de vaardigheid op peilingschaal VO 2022. De cesuurpunten voor de vier referentieniveaus op die schaal zijn gegeven in Tabel 5.18.

Tabel 5.18 – Cesuurpunten voor de vier referentieniveaus Rekenen op de VO-peilingschaal<sup>39</sup>

	1F	2F	1S	3F
Vaardigheid op peilingschaal VO 2022	-0,2639	0,3061	0,4340	1,1082

Bij deze werkwijze zijn een aantal opmerkingen te plaatsen. De eerste betreft de equivalentie van de opgaven op de PO-peilingschaal en die op de VO-peilingschaal. Voor deze methode is het niet noodzakelijk dat de items op dezelfde schaal gebracht zijn en exact dezelfde parameters hebben. Wat wel van belang is, is dat de scores een equivalente betekenis hebben. Binnen dit rapport is al uitvoerig beschreven hoe de opgaven die betrekking hebben op de verschillende referentieniveaus goed samen op een schaal passen. Dat was ook het geval in de PO-peiling. Doordat zowel in de PO- als de VO-peilingschaal alle opgaven op een lijn passen, geldt dat ook voor deze 62 ankeropgaven. Hoewel DIF voor het omzetten via een ware-score-equivalering<sup>40</sup> geen ultiem criterium is, geeft het vertrouwen in de equivalentie van de schaal als de volgorde van de items naar moeilijkheid in beide populaties niet fundamenteel verschilt. Daartoe is de correlatie tussen de b-parameters bij een Rasch-schaling geëvalueerd aangezien in dat model de itemparameters echt als moeilijkheid van de items geïnterpreteerd kunnen worden. De correlatie tussen die b-waarden is met een waarde van 0,88 hoog te noemen.

De tweede opmerking heeft te maken met de robuustheid van de resultaten. Zoals gezegd speelt bij de equivalering van referentieniveau door middel van de Ware-Score-Equivalering de evaluatie van DIF een minder centrale rol. Dat is ook de reden dat binnen het referentiekader gewoon over 1F en 1S in het PO gesproken kan worden, 2F en 3F in het mbo, en alle vier de niveaus in het VO. Als we naar de schaalwaarden kijken, ook van de itemparameters, dan zien we daar wel enige verschillen in de verschillende populaties. In dit onderzoek zijn de parameters van de 62 ankeropgaven die voor de equivalering gebruikt zijn ook onderzocht. Wanneer een kalibratie uitgevoerd wordt, met zowel de data van de PO-peiling als die van deze VO-peiling, waarbij alle parameters vrij geschat worden ('concurrent calibration') dat zien we ook dat er vrijwel geen DIF gevonden wordt. Wanneer er beperkingen opgelegd worden door parameters te fixeren wordt deze passing uiteraard slechter. In de gevallen dat subsets van de 62 opgaven niet opgenomen worden, bijvoorbeeld omdat er in deze striktere kalibratie opgaven

<sup>39</sup> Deze waarden worden verkregen door de waarden van de Referentieschaal-theta in Tabel 5.18 te vermenigvuldigen met de in de tekst genoemde waarde voor a, en daarbij de waarde voor b bij op te tellen.

<sup>40</sup> Bij de referentieniveaus is dat ook de reden dat de referentieniveaus eigenlijk gedefinieerd zijn op basis van aantal correcte opgaven van de Openbare Set, en niet direct gedefinieerd (en gecommuniceerd) zijn als cesuurpunten op een vaardigheidschaal. Die cesuurpunten verschillen per doelgroep of populatie. De in dit verslag geleverde referentieschaal-theta-waarden in Tabel 5.18 zijn de cesuur-waarden op de theta-schaal als we de VO-populatie nemen (zonder 6VWO).

mogelijk als DIF vertonen tussen PO en VO, dan zien we dat de resultaten behoorlijk robuust zijn. Diverse selecties zijn toegepast, waarbij de meest radicale een halvering van het aantal opgaven betrof, en de verschillen in verwachte percentages per referentieniveau beperkt<sup>41</sup>. De verschillen lagen qua grootte in de lijn van wat binnen de Eindtoets-equivaleringen als acceptabel gezien wordt (maximaal verschil kleiner dan 4%; merendeel van de verschillen minder dan 2%)<sup>42</sup>.

Gezien de uitlegbaarheid van de procedure, de gevonden equivalentie en de robuustheid van de resultaten, kiezen we ervoor de resultaten zoals gegeven in Tabel 5.18 te gebruiken als de cesuurpunten voor de vier referentieniveaus in de rest van het onderzoek. Een ander voordeel van deze cesuurpunten is dat deze direct gerelateerd kunnen worden aan de WML-schatters die in dit onderzoek zijn opgeleverd. Bij de presentatie van de resultaten aangaande de hoeveelheid leerlingen die een specifiek niveau behalen, wordt daarvan gebruik gemaakt.

---

<sup>41</sup> Deze percentages zijn bepaald door de gegevens over de geschatte verdelingen op een schaal te relateren op de cesuurpunten voor de referentieniveaus op dezelfde schaal. Op die wijze kunnen de resultaten van verschillende schalen direct met elkaar vergeleken worden.

<sup>42</sup> Deze richtlijnen zijn eerder gebruikt door de Expertgroep PO.

## 6 Schaling contextvragenlijsten

### 6.1 Inleiding

In dit hoofdstuk beschrijven en analyseren we de data die met de vragenlijsten voor leerlingen, docenten en schoolleiders zijn verzameld. Voor de leerlingen gaat het om twee vragenlijsten: de versie voor de leerlingen in regulier VO (Paragraaf 6.2) en de verkorte versie voor leerlingen in het praktijkonderwijs (Paragraaf 6.3). Per instrument presenteren we in de volgende paragrafen onze bevindingen aan de hand van een aantal tabellen. Deze staan in dezelfde volgorde als indertijd de betreffende vragen in de vragenlijsten. Ook de items binnen de tabellen staan in dezelfde volgorde als in de vragenlijsten. Bijlage 1, 2 en 3 bij dit rapport bevatten de frequentieverdelingen van de oorspronkelijke variabelen in de vragenlijsten voor leerlingen, docenten en schoolleiders.

### 6.2 Leerlingvragenlijst (versie regulier VO)

#### 6.2.1 Ontbrekende waarden

Leerlingen in regulier VO die meer dan de helft van de vragen in de leerlingvragenlijst niet hebben ingevuld zijn uit de data verwijderd. Het betreft in totaal 14 van de 2620 leerlingen (0,5%). In totaal hebben 2606 VO-leerlingen de vragenlijst helemaal of grotendeels ingevuld (Tabel 6.1).

Tabel 6.1 – Aantal missende waarden in leerlingvragenlijst regulier VO (n=2606, aantal items=78)

Aantal missings	Aantal leerlingen
0	1245
1	417
2	192
3	125
4	138
5	176
6	104
7	55
8	25
9	25
10 t/m 39	104
<b>Totaal</b>	<b>2606</b>

#### 6.2.2 Achtergrondkenmerken

In Tabel 6.2 worden de achtergrondkenmerken van de leerlingen in regulier VO gepresenteerd, in de vorm van een aantal frequentieverdelingen. Omdat de vragenlijst op papier werd ingevuld, konden de leerlingen (sub)vragen relatief gemakkelijk overslaan. In de tabel is daarom per kenmerk vermeld op hoeveel leerlingen de frequentieverdeling betrekking heeft.

Tabel 6.2 – Achtergrondkenmerken<sup>1</sup> leerlingen in regulier vo, in percentages (n maximaal=2606)

Kenmerk	%
<i>Geslacht (n=2573)</i>	
Meisje	47,8
Jongen	49,3
Anders	2,8
<i>Aantal boeken thuis (n=2582)</i>	
Geen of weinig (0–10 boeken)	20,9
Een plank vol (11–25 boeken)	24,1
Een boekenkast vol (26–100 boeken)	30,0
Twee boekenkasten vol (101–200 boeken)	13,4
Drie of meer boekenkasten vol (meer dan 200 boeken)	11,6
<i>Hoogste opleiding moeder (n=2559)</i>	
Mijn moeder is nooit naar school geweest of heeft de basisschool niet afgemaakt	0,5
Basisschool	0,6
Vmbo, vbo of mavo	12,0
Havo	7,7
Vwo	3,1
Mbo	9,1
Hbo	15,4
Universiteit	13,4
Universiteit en gepromoveerd	4,5
Weet ik niet	30,9
Niet van toepassing	2,7
<i>Hoogste opleiding vader (n=2555)</i>	
Mijn vader is nooit naar school geweest of heeft de basisschool niet afgemaakt	0,8
Basisschool	0,8
Vmbo, vbo of mavo	13,2
Havo	6,0
Vwo	3,3
Mbo	8,8
Hbo	11,6
Universiteit	13,2
Universiteit en gepromoveerd	4,7
Weet ik niet	34,3
Niet van toepassing	3,2

Kenmerk (vervolg Tabel 6.2)	%
<i>Moeder geboren in Nederland? (n=2597)</i>	
Ja	81,3
Nee	16,1
Weet ik niet	1,0
Niet van toepassing	1,7
<i>Vader geboren in Nederland? (n=2593)</i>	
Ja	80,3
Nee	16,7
Weet ik niet	1,1
Niet van toepassing	1,9
<i>Zelf geboren in Nederland? (n=2592)</i>	
Ja	93,3
Nee	6,7
<i>Meest gesproken taal thuis (n=2558)</i>	
Nederlands (of Fries of een Nederlands dialect)	87,6
Een andere Europese taal (bijvoorbeeld: Duits, Engels, Pools, Zweeds)	3,6
Een andere niet-Europese taal (bijvoorbeeld: Arabisch, Berber, Chinees, Papiaments, Turks)	8,8
<i>Hoe vaak praat je thuis Nederlands (of Fries of een Nederlands dialect) (n=2588)</i>	
Altijd	73,6
Bijna altijd	16,2
Soms	7,7
Nooit	2,4
<i>Opleidingstype derde leerjaar (n=2577)</i>	
Vmbo	3,8
Havo	12,8
Vwo	18,9
Mbo	23,4
Hbo	27,1
Universiteit	1,0
Weet ik nog niet	13,1
<i>Verwachte opleidingen met diploma (n=2564, meerdere antwoorden mogelijk)</i>	
Vmbo	50,6
Havo	43,4
Vwo	33,2
Mbo	32,4
Hbo	40,1
Universiteit	33,9

<sup>1</sup> Vraag 1.1 en 1.3 tot en met 1.9 van de Leerlingvragenlijst regulier vo



Bij twee achtergrondvragen is het antwoord ‘weet niet’ relatief vaak aangekruist. Het gaat om de vragen over de opleiding van de moeder respectievelijk de vader (31% resp. 34%). Als we dit antwoord samenvoegen met het antwoord ‘niet van toepassing’ (heeft geen moeder / vader) en de ‘echte’ missings (vraag niet beantwoord) kennen we van slechts 66% (moeder) respectievelijk 63% (vader) van de leerlingen het ouderlijk opleidingsniveau. Om die reden zijn deze variabelen niet meegenomen in de verdere analyses. Pro-leerlingen weten het antwoord op deze vragen overigens nog veel vaker niet (ca. 65%; zie Paragraaf 6.3).

### 6.2.3 Domeinspecifieke kenmerken

In deze paragraaf worden de domeinspecifieke kenmerken van de leerlingen in regulier VO gepresenteerd. Deze zijn gemeten aan de hand van een aantal uitspraken over wiskunde en de wiskundelessen; de leerlingen konden aangeven in hoeverre ze het eens of oneens waren met deze uitspraken.

In de tabellen 6.3 tot en met 6.11 worden de gemiddelde scores en standaarddeviatie per item getoond, berekend over alle responderende leerlingen in regulier VO tezamen. Voor de items binnen een vraag die samen een bepaald concept representeren (op basis van andere onderzoeken zoals Peil.Rekenen-Wiskunde in PO, PISA of TIMSS) zijn Cronbach’s alfa en Guttman’s Lambda-2 berekend en vermeld onder de tabel. Bij een waarde van 0,60 of hoger (in de leerlingendata voldoen alle concepten aan die eis) is een nieuwe samengestelde variabele voor dit concept geconstrueerd. Ook het gemiddelde en de standaarddeviatie van dit construct is in de tabel opgenomen, alsmede het aantal leerlingen waarop het gemiddelde betrekking heeft (n). Als één of meer items binnen een construct zijn overgeslagen door een leerling, is geen score berekend.

De oorspronkelijke voorgegeven antwoorden in de leerlingenvragenlijst liepen veelal van positief naar negatief of van meer naar minder (bijvoorbeeld: 1=zeer mee eens, 4=zeer mee oneens). Achteraf zijn deze antwoorden omgecodeerd (1=4 enz.), behalve als het item negatief geformuleerd was. Indien van toepassing staat onder de tabel vermeld hoe de oorspronkelijke antwoorden in de vragenlijst luiden en welke items achteraf omgecodeerd zijn. In het algemeen geldt voor de waardes in de tabellen: **hoe hoger de score, hoe méér / hoe positiever.**

Tabel 6.3 – Plezier in wiskunde<sup>1</sup> van leerlingen in regulier vo, gemiddelden (sd)

	Gemiddelde (sd)
a. Ik vind het leuk om wiskunde te leren	2,1 (1,0)
b. Ik zou willen dat ik geen wiskunde hoefde te leren	2,1 (1,0)
c. Wiskunde is saai	2,1 (1,0)
d. Ik leer veel interessante dingen tijdens de wiskundelessen	2,3 (0,9)
e. Ik vind wiskunde leuk	2,1 (1,0)
f. Ik vind alle schooltaken leuk waarbij ik met getallen moet werken	1,9 (0,8)
g. Ik vind het leuk om wiskundeopgaven op te lossen	2,0 (0,9)
h. Ik verheug mij op de wiskundeles	1,7 (0,9)
i. Wiskunde is één van mijn favoriete vakken	1,8 (1,0)
Construct <i>Plezier in wiskunde</i> <sup>2</sup> (n=2417)	2,0 (0,7)

<sup>1</sup> Vraag 2.1, oorspronkelijke antwoorden: 1=zeer mee eens, 4=zeer mee oneens. Alle items omgecodeerd, m.u.v. item b en c.

<sup>2</sup> Cronbach’s alfa: 0,92 Guttman’s Lambda 2: 0,92

Tabel 6.4 – Zelfvertrouwen in wiskunde<sup>1</sup> van leerlingen in regulier vo, gemiddelden (sd)

	Gemiddelde (sd)
a. Meestal ben ik goed in wiskunde	2,8 (0,9)
b. Wiskunde is voor mij moeilijker dan voor veel van mijn klasgenoten	2,9 (0,9)
c. Ik ben gewoon niet goed in wiskunde	2,8 (1,0)
d. Wiskunde is makkelijk voor mij	2,4 (1,0)
e. Ik ben goed in het oplossen van moeilijke wiskundeopgaven	2,3 (0,9)
f. Ik kan wiskunde goed aan anderen uitleggen	2,3 (0,9)
g. Wiskunde is voor mij moeilijker dan alle andere vakken op school	2,7 (1,0)
h. Ik vind wiskunde moeilijk te snappen	2,7 (1,0)
i. Ik hoef maar weinig tijd aan mijn wiskundehuiswerk te besteden om toch goede cijfers te halen	2,4 (1,1)
Construct <i>Zelfvertrouwen in wiskunde</i> <sup>2</sup> (n=2321)	2,6 (0,8)

<sup>1</sup> Vraag 2.2, oorspronkelijke antwoorden: 1=zeer mee eens, 4=zeer mee oneens. Alle items omgecodeerd, m.u.v. b, c, g, h.

<sup>2</sup> Cronbach's alfa: 0,92 Guttman's Lambda 2: 0,92

Tabel 6.5 – Belang van wiskunde<sup>1</sup> voor leerlingen in regulier vo, gemiddelden (sd)

	Gemiddelde (sd)
a. Ik denk dat wiskunde leren mij helpt in mijn dagelijks leven	2,2 (1,0)
b. Ik heb wiskunde nodig bij het leren van andere schoolvakken	2,2 (0,9)
c. Ik moet goed in wiskunde presteren om later de vervolgstudie te kunnen doen die ik graag wil	2,5 (1,0)
d. Ik moet goed in wiskunde presteren om later de baan te kunnen krijgen die ik graag wil	2,5 (1,0)
e. Ik wil later een baan waarbij je wiskunde gebruikt	2,0 (1,0)
f. Wiskunde leren is belangrijk om in deze wereld vooruit te komen	2,6 (0,9)
g. Door wiskunde te leren krijg ik meer mogelijkheden om later een baan kunnen te vinden	2,8 (0,9)
h. Mijn ouders denken dat het belangrijk is dat ik goed in wiskunde ben	2,8 (0,9)
i. Het is belangrijk om goed in wiskunde te zijn	2,6 (0,9)
Construct <i>Belang van wiskunde</i> <sup>2</sup> (n=2321)	2,4 (0,7)

<sup>1</sup> Vraag 2.3, oorspronkelijke antwoorden: 1=zeer mee eens, 4=zeer mee oneens. Alle items omgecodeerd.

<sup>2</sup> Cronbach's alfa: 0,88 Guttman's Lambda 2: 0,89

Tabel 6.6 – *Attributies positieve wiskundeprestaties*<sup>1</sup> leerlingen in regulier vo, gemiddelden (sd)

	Gemiddelde (sd)
a. Als ik een wiskundeopdracht goed maak komt dat vooral omdat ik goed mijn best heb gedaan	3,1 (0,8)
c. Als ik een wiskundeopdracht goed maak komt dat vooral omdat ik hard gewerkt heb	2,9 (0,9)
f. Als ik een wiskundeopdracht goed maak komt dat vooral omdat ik goed heb opgelet tijdens de les	2,9 (0,9)
Construct <i>Attributies positieve wiskundeprestaties</i> <sup>2</sup> (n=2515)	3,0 (0,7)

<sup>1</sup> Vraag 2.4, oorspronkelijke antwoorden: 1=zeer mee eens, 4=zeer mee oneens. Alle items omgecodeerd.

<sup>2</sup> Cronbach's alfa: 0,70 Guttman's Lambda 2: 0,70

Tabel 6.7 – Growth mindset wiskunde<sup>1</sup> leerlingen in regulier vo, gemiddelden (sd)

	Gemiddelde (sd)
b. Hoe goed je bent in wiskunde ligt vast en je kunt niet veel doen om dat te veranderen	2,8 (0,9)
d. Het maakt niet uit hoe slim je bent, als je wil kun je altijd goed worden in wiskunde	2,9 (0,9)
e. Je kunt niet veranderen hoe slim je in wiskunde bent	2,9 (1,0)
g. Goed zijn in wiskunde is iets waar je mee geboren wordt, je kunt niet veel doen om jezelf beter in wiskunde te maken	3,1 (0,9)
h. Iedereen kan goed zijn in wiskunde	3,0 (0,9)
i. Je kunt zelf veranderen hoe goed je bent in wiskunde	3,1 (0,9)
Construct <i>Growth mindset</i> <sup>2</sup> (n=2461)	3,0 (0,7)

<sup>1</sup> Vraag 2.4, oorspronkelijke antwoorden: 1=zeer mee eens, 4=zeer mee oneens. Alle items omgecodeerd, m.u.v. item b, e, g.

<sup>2</sup> Cronbach's alfa: 0,81 Guttman's Lambda 2: 0,81

Tabel 6.8 – Ordelijkheid wiskundeles<sup>1</sup> volgens leerlingen in regulier vo, gemiddelden (sd)

	Gemiddelde (sd)
a. De leerlingen luisteren niet naar wat de docent zegt	2,6 (0,8)
b. Het is lawaaiig en rommelig in de klas	2,4 (0,9)
c. Het is te onrustig voor de leerlingen om te kunnen werken	2,7 (0,9)
d. De docent moet een lange tijd wachten voordat de leerlingen stil zijn	2,5 (1,0)
e. De leerlingen onderbreken de docent	2,6 (1,0)
f. De docent moet de regels in de klas heel vaak herhalen	2,9 (1,0)
g. Het gedrag van andere leerlingen in de klas maakt het voor mij moeilijk om me te kunnen concentreren.	2,9 (1,0)
h. De leerlingen beginnen pas met werken als de les allang begonnen is	2,5 (1,0)
Construct <i>Ordelijkheid wiskundeles</i> <sup>2</sup> (n=2433)	2,6 (0,7)

<sup>1</sup> Vraag 3.1, oorspronkelijke antwoorden: 1=elke of bijna elke les, 2=ongeveer de helft van de lessen, 3=sommige lessen, 4=nooit. Alle items omgecodeerd.

<sup>2</sup> Cronbach's alfa: 0,89 Guttman's Lambda 2: 0,89

Tabel 6.9 – Helderheid instructie wiskundeles<sup>1</sup> volgens leerlingen in regulier vo, gemiddelden (sd)

	Gemiddelde (sd)
a. Ik weet precies wat ik van mijn docent moet doen	3,1 (0,9)
b. Ik kan mijn docent goed begrijpen	3,0 (0,9)
c. Mijn docent geeft duidelijk antwoord op mijn vragen	3,0 (1,0)
d. Mijn docent kan wiskunde goed uitleggen	3,1 (1,0)
e. Mijn docent doet verschillende dingen om ons te helpen met leren	3,0 (0,9)
f. Tijdens de uitleg van nieuwe leerstof maakt mijn docent gebruik van wat ik al weet	2,9 (0,9)
g. Als wij iets niet snappen dan legt mijn docent het onderwerp nog een keer uit	3,2 (0,9)
h. Mijn docent gaat net zo lang door met uitleggen tot de leerlingen het begrijpen	2,9 (1,0)
i. Als ik iets niet begrijp dan legt mijn docent het op een andere manier nog een keer uit	2,9 (1,0)
Construct <i>Helderheid instructie wiskundeles</i> <sup>2</sup> (n=2375)	3,0 (0,7)

<sup>1</sup> Vraag 3.2, oorspronkelijke antwoorden: 1=zeer mee eens, 4=zeer mee oneens. Alle items omgecodeerd.

<sup>2</sup> Cronbach's alfa: 0,92 Guttman's Lambda 2: 0,92

Tabel 6.10 – Oplossingsstrategieën in wiskundeles<sup>1</sup> volgens leerlingen in regulier vo, gemiddelden (sd)

	Gemiddelde (sd)
j. Mijn docent leert ons verschillende manieren om een opgave op te lossen	2,9 (0,9)
k. Mijn docent geeft ons opgaven waarvoor we zelf een eigen oplossingsmanier moeten bedenken	2,5 (1,0)
l. Ik mag van mijn docent een opgave op mijn eigen manier oplossen	2,6 (1,0)
Construct <i>Oplossingsstrategieën in de wiskundeles</i> <sup>2</sup> (n=2471)	2,7 (0,8)

<sup>1</sup> Vraag 3.2, oorspronkelijke antwoorden: 1=zeer mee eens, 4=zeer mee oneens. Alle items omgecodeerd.

<sup>2</sup> Cronbach's alfa: 0,68 Guttman's Lambda 2: 0,68

Tabel 6.11 – Directe instructie in wiskundeles<sup>1</sup> volgens leerlingen regulier vo, gemiddelden (sd)

	Gemiddelde (sd)
a. De docent stelt duidelijke doelen voor wat we leren	3,1 (0,9)
b. De docent stelt vragen om te controleren of we begrepen hebben wat we hebben besproken	3,1 (1,0)
c. Aan het begin van de les geeft de docent een korte samenvatting van de vorige les	2,5 (1,1)
d. De docent zegt ons wat we moeten leren	3,3 (0,9)
Construct <i>Directe instructie in de wiskundeles</i> <sup>2</sup> (n=2483)	3,0 (0,7)

<sup>1</sup> Vraag 3.3, oorspronkelijke antwoorden: 1=elke of bijna elke les, 2=ongeveer de helft van de lessen, 3=sommige lessen, 4=nooit. Alle items omgecodeerd.

<sup>2</sup> Cronbach's alfa: 0,75 Guttman's Lambda 2: 0,75

In Tabel 6.12 zijn alle leerlingen in regulier VO ingedeeld in drie categorieën, verwijzend naar het onderwijsniveau in de tweede klas zoals aangegeven door de docent in de docentvragenlijst (Tabel 6.12). Onder andere door de non-respons op die vragenlijst kan een deel van de leerlingen niet worden ingedeeld. Bij de multilevel analyses (Hoofdstuk 8) wordt dat gecorrigeerd met behulp van imputatie.

Tabel 6.12 – Leerlingen in regulier VO naar onderwijsniveau, o.b.v. categorisering in docentvragenlijst.

	N	%
Vmbo-basis en/of vmbo-kader	394	15,1%
Vmbo-gemengd en/of -theoretisch	573	22,0%
Havo en/of vwo	1035	39,7%
<i>Missing</i>	604	23,2%
Totaal	2606	100%

Tabel 6.13 toont vervolgens de scores van alle constructvariabelen uit de voorafgaande tabellen, gemiddeld per onderwijsniveau van de leerlingen. Door het overslaan van bepaalde items en vragen varieert het aantal leerlingen (n) per constructvariabele. Tabel 6.13 laat kleine verschillen in gemiddelden tussen de onderwijsniveaus zien; echter niet steeds in dezelfde richting.

Tabel 6.13 – Constructvariabelen: gemiddelde (sd) per onderwijsniveau en totaal, leerlingen regulier vo

Constructvariabelen	Vmbo-bk	Vmbo-gt	Havo/vwo	Totaal
Plezier in wiskunde (n=1859)	2,0 (0,8)	1,9 (0,7)	2,0 (0,7)	2,0 (0,7)
Zelfvertrouwen in wiskunde (n=1792)	2,5 (0,8)	2,5 (0,8)	2,6 (0,8)	2,6 (0,8)
Belang van wiskunde (n=1789)	2,4 (0,7)	2,4 (0,7)	2,5 (0,6)	2,5 (0,7)
Attributies positieve wiskundeprestaties (n=1939)	3,0 (0,7)	2,9 (0,7)	3,1 (0,6)	3,0 (0,7)
<i>Growth mindset</i> (n=1897)	2,9 (0,6)	2,9 (0,7)	3,1 (0,7)	3,0 (0,7)
Ordelijkheid wiskundeles (n=1880)	2,5 (0,7)	2,5 (0,7)	2,8 (0,7)	2,7 (0,7)
Helderheid instructie wiskundeles (n=1823)	3,0 (0,7)	2,9 (0,8)	3,1 (0,7)	3,1 (0,7)
Oplossingsstrategieën in de wiskundeles (n=1900)	2,8 (0,8)	2,6 (0,8)	2,7 (0,7)	2,7 (0,8)
Directe instructie in de wiskundeles (n=1909)	3,1 (0,7)	2,9 (0,8)	3,1 (0,7)	3,0 (0,7)

### 6.3 Leerlingvragenlijst (versie praktijkonderwijs)

#### 6.3.1 Ontbrekende waarden

In totaal zijn hebben 77 van de 78 pro-leerlingen de verkorte leerlingvragenlijst helemaal of grotendeels ingevuld. Eén pro-leerling heeft driekwart van de vragenlijst niet ingevuld (Tabel 6.14).

Tabel 6.14 – Aantal missende waarden in leerlingvragenlijst praktijkonderwijs (n=77, aantal items=45)

Aantal missings	Aantal leerlingen
0	40
1	17
2	5
3	3
4	5
5	2
6	1
7	0
8	1
9	0
10 t/m 22	3
<b>Totaal</b>	<b>77</b>

#### 6.3.2 Achtergrondkenmerken

In Tabel 6.15 worden de achtergrondkenmerken van de pro-leerlingen gepresenteerd, in de vorm van een aantal frequentieverdelingen. In de tabel is per kenmerk vermeld op hoeveel leerlingen de frequentieverdeling betrekking heeft.

Tabel 6.15 – Achtergrondkenmerken<sup>1</sup> leerlingen praktijkonderwijs, in percentages (n maximaal=77)

Kenmerk	%
<b>Geslacht (n=73)</b>	
Meisje	45,2
Jongen	53,4
Anders	1,4
<b>Aantal boeken thuis (n=73)</b>	
Geen of weinig (0–10 boeken)	43,8
Een plank vol (11–25 boeken)	32,9
Een boekenkast vol (26–100 boeken)	15,1
Twee boekenkasten vol (101–200 boeken)	4,1
Drie of meer boekenkasten vol (meer dan 200 boeken)	4,1
<b>Hoogste opleiding moeder (n=72)</b>	
Mijn moeder is nooit naar school geweest of heeft de basisschool niet afgemaakt	1,4
Basisschool	4,2
Vmbo, vbo of mavo	8,3
Havo	5,6
Vwo	0,0
Mbo	2,8
Hbo	2,8
Universiteit	1,4
Universiteit en gepromoveerd	2,8
Weet ik niet	66,7
Niet van toepassing	4,2
<b>Hoogste opleiding vader (n=70)</b>	
Mijn vader is nooit naar school geweest of heeft de basisschool niet afgemaakt	4,3
Basisschool	2,9
Vmbo, vbo of mavo	12,9
Havo	1,4
Vwo	0,0
Mbo	1,4
Hbo	4,3
Universiteit	1,4
Universiteit en gepromoveerd	2,9
Weet ik niet	64,3
Niet van toepassing	4,3

Kenmerk	%
<i>Moeder geboren in Nederland? (n=75)</i>	
Ja	58,7
Nee	33,3
Weet ik niet	6,7
Niet van toepassing	1,3
<i>Vader geboren in Nederland? (n=75)</i>	
Ja	52,0
Nee	44,0
Weet ik niet	2,7
Niet van toepassing	1,3
<i>Zelf geboren in Nederland? (n=76)</i>	
Ja	75,0
Nee	25,0
<i>Meest gesproken taal thuis (n=67)</i>	
Nederlands (of Fries of een Nederlands dialect)	67,2
Een andere Europese taal (bijvoorbeeld: Duits, Engels, Pools, Zweeds)	6,0
Een andere niet-Europese taal (bijvoorbeeld: Arabisch, Berber, Chinees, Papiaments, Turks)	26,9
<i>Hoe vaak praat je thuis Nederlands (of Fries of een Nederlands dialect) (n=76)</i>	
Altijd	52,6
Bijna altijd	23,7
Soms	17,1
Nooit	6,6

<sup>1</sup> Vraag 1.1 en 1.3 tot en met 1.7 van de verkorte leerlingvragenlijst

### 6.3.3 Domeinspecifieke kenmerken

In de tabellen in deze paragraaf worden de domeinspecifieke kenmerken van de leerlingen in het praktijkonderwijs gepresenteerd. Deze zijn gemeten aan de hand van een aantal uitspraken over rekenen en de rekenlessen; de leerlingen konden aangeven in hoeverre ze het eens of oneens waren met deze uitspraken. In de tabellen 6.16 tot en met 6.20 worden de gemiddelde scores en standaarddeviatie per item getoond, berekend over alle responderende leerlingen in het praktijkonderwijs tezamen. Er zijn op basis hiervan geen nieuwe samengestelde variabelen geconstrueerd, vanwege het kleine aantal leerlingen in praktijkonderwijs.

Ook in de verkorte leerlingenvragenlijst liepen de oorspronkelijke voorgegeven antwoorden veelal van positief naar negatief of van meer naar minder (bijv. 1=zeer mee eens, 4=zeer mee oneens). Achteraf zijn deze antwoorden omgecodeerd, behalve als het item negatief geformuleerd was. Indien van toepassing staat onder de tabel vermeld hoe de oorspronkelijke antwoorden in de vragenlijst luiden en welke items omgecodeerd zijn. In het algemeen geldt voor de waardes in de tabellen: **hoe hoger de score, hoe méér / hoe positiever.**

Tabel 6.16 – Plezier in rekenen<sup>1</sup> van pro-leerlingen, gemiddelden (sd); n=71-76

	Gemiddelde (sd)
a. Ik vind het leuk om rekenen te leren	2,8 (1,0)
b. Ik zou willen dat ik geen rekenen hoefde te leren	2,7 (1,1)
c. Rekenen is saai	2,5 (1,1)
d. Ik leer veel interessante dingen tijdens de rekenlessen	2,8 (1,0)
e. Ik vind rekenen leuk	2,5 (1,1)
f. Ik vind alle schooltaken leuk waarbij ik met getallen moet werken	2,3 (1,0)
g. Ik vind het leuk om rekenopgaven op te lossen	2,4 (1,0)
h. Ik verheug mij op de rekenles	2,2 (1,1)
i. Rekenen is één van mijn favoriete vakken	2,0 (1,2)

<sup>1</sup> Vraag 2.1, oorspronkelijke antwoorden: 1=zeer mee eens, 4=zeer mee oneens. Alle items omgecodeerd, m.u.v. item b en c.

Tabel 6.17 – Zelfvertrouwen in rekenen<sup>1</sup> van pro-leerlingen, gemiddelden (sd); n=70-75

	Gemiddelde (sd)
a. Meestal ben ik goed in rekenen	2,8 (0,9)
b. Rekenen is voor mij moeilijker dan voor veel van mijn klasgenoten	2,7 (1,1)
c. Ik ben gewoon niet goed in rekenen	2,8 (1,0)
d. Rekenen is makkelijk voor mij	2,6 (1,0)
e. Ik ben goed in het oplossen van moeilijke rekenopgaven	2,3 (1,0)
f. Ik kan rekenen goed aan anderen uitleggen	2,3 (1,0)
g. Rekenen is voor mij moeilijker dan alle andere vakken op school	2,7 (1,1)
h. Ik vind rekenen moeilijk te snappen	2,7 (1,1)

<sup>1</sup> Vraag 2.2, oorspronkelijke antwoorden: 1=zeer mee eens, 4=zeer mee oneens. Alle items omgecodeerd, m.u.v. b, c, g, h.

Tabel 6.18 – Attributies positieve rekenprestaties<sup>1</sup> pro-leerlingen, gemiddelden (sd); n=74-76

	Gemiddelde (sd)
a. Als ik een rekenopdracht goed maak komt dat vooral omdat ik goed mijn best heb gedaan	3,4 (0,8)
c. Als ik een rekenopdracht goed maak komt dat vooral omdat ik hard gewerkt heb	3,1 (0,9)
f. Als ik een rekenopdracht goed maak komt dat vooral omdat ik goed heb opgelet tijdens de les	3,1 (0,9)

<sup>1</sup> Vraag 2.3, oorspronkelijke antwoorden: 1=zeer mee eens, 4=zeer mee oneens. Alle items omgecodeerd.



Tabel 6.19 – Growth mindset rekenen<sup>1</sup> pro-leerlingen, gemiddelden (sd); n=74-75

	Gemiddelde (sd)
b. Hoe goed je bent in rekenen ligt vast en je kunt niet veel doen om dat te veranderen	2,4 (1,0)
d. Het maakt niet uit hoe slim je bent, als je wil kun je altijd goed worden in rekenen	3,2 (0,9)
e. Je kunt niet veranderen hoe slim je in rekenen bent	2,4 (1,0)
f. Goed zijn in rekenen is iets waar je mee geboren wordt, je kunt niet veel doen om jezelf beter in rekenen te maken	2,8 (1,1)
h. Iedereen kan goed zijn in rekenen	3,0 (1,0)
i. Je kunt zelf veranderen hoe goed je bent in rekenen	3,2 (0,9)

<sup>1</sup> Vraag 2.3, oorspronkelijke antwoorden: 1=zeer mee eens, 4=zeer mee oneens. Alle items omgecodeerd, m.u.v. item b, e, g.

Tabel 6.20 – Ordelijkheid rekenles<sup>1</sup> volgens pro-leerlingen, gemiddelden (sd); n=71-76

	Gemiddelde (sd)
a. De leerlingen luisteren niet naar wat de docent zegt	2,6 (0,8)
b. Het is lawaaig en rommelig in de klas	2,6 (0,8)
c. Het is te onrustig voor de leerlingen om te kunnen werken	2,8 (1,0)
d. De docent moet een lange tijd wachten voordat de leerlingen stil zijn	2,5 (1,0)
e. De leerlingen onderbreken de docent	2,4 (1,0)
f. De docent moet de regels in de klas heel vaak herhalen	2,7 (1,0)
g. Het gedrag van andere leerlingen in de klas maakt het voor mij moeilijk om me te kunnen concentreren.	2,8 (1,1)
h. De leerlingen beginnen pas met werken als de les allang begonnen is	2,9 (1,0)

<sup>1</sup> Vraag 3.1. Antwoorden: 1=elke of bijna elke les, 2=ongeveer de helft van de lessen, 3=sommige lessen, 4=nooit. Alle items omgecodeerd.

## 6.4 Docentvragenlijst

### 6.4.1 Ontbrekende waarden

Er hebben 133 klassen deelgenomen aan het onderzoek en alle docenten die rekenen of wiskunde aan deze klassen gaven zijn uitgenodigd om de docentvragenlijst in te vullen. In totaal zijn 104 docenten daarmee gestart, maar vier van hen hebben (veel) minder dan de helft van de vragenlijst ingevuld. Zij zijn uit het bestand gehaald. Het bestand bevat verder nauwelijks missende waarden, omdat het een digitale vragenlijst betrof en de respondent geen vragen kon overslaan. De docentvragenlijst was gelijk voor alle docenten, ongeacht het onderwijsniveau van de tweede klas waaraan ze les gaven.

### 6.4.2 Achtergrondkenmerken

In deze paragraaf presenteren we de achtergrondkenmerken van de docenten die les gaven in Rekenen en Wiskunde (hierna verdere aangeduid als wiskunde, net als in de vragenlijst) aan de getoetste klassen.

Tabel 6.21 – Onderwijservaring<sup>1</sup> docenten van de getoetste klassen, gemiddelden en range

Kenmerk	Gemiddelde (sd)	Min	Max
<i>Onderwijservaring (n=100)</i>			
In totaal lesgegeven (in jaren)	13,4 (9,3)	1,0	39,0
Lesgegeven in wiskunde (in jaren)	10,8 (8,6)	1,0	36,0

<sup>1</sup> Vraag 1.1a en b van de docentvragenlijst

Tabel 6.22 – Overige achtergrondkenmerken<sup>1</sup> docenten van de getoetste klassen, in percentages

Kenmerk	%
<i>Geslacht (n=100)</i>	
Vrouw	49,0
Man	51,0
Anders	0,0
<i>Leeftijd (n=100)</i>	
Jonger dan 25	5,0
25-29	14,0
30-39	26,0
40-49	27,0
50-59	21,0
60 of ouders	7,0
<i>Hoogste opleiding met diploma afgerond (n=100)</i>	
Voortgezet onderwijs of middelbaar beroepsonderwijs	6,0
Hoger beroepsonderwijs – bachelor	65,0
Hoger beroepsonderwijs – master	8,0
Universiteit kandidaats of bachelor	7,0
Universiteit doctoraal of master	14,0
Universiteit en gepromoveerd	6,0
<i>Lerarenopleiding met een diploma afgerond (n=100)</i>	
Niet afgerond/stage	11,0
Afgerond waarvan (meerdere antwoorden mogelijk, n=89):	89,0
• Hbo-lerarenopleiding voortgezet onderwijs	65,2
• Kopopleiding hbo-bachelor	6,7
• Educatieve minor/module universiteit (beperkte tweedegraads)	2,2
• PABO (beperkte tweedegraads)	21,3
• Academische of universitaire PABO	0,0
• Zij-instroom (2-jarige opleiding)	11,2
• Lerarenopleiding hbo-master	7,9
• Lerarenopleiding universiteit	5,6
• Anders	9,0

Kenmerk	%
<i>Lesbevoegdheid (n=100)</i>	
(Nog) geen lesbevoegdheid/stage	12,0
Beperkte tweedegraads	18,0
Tweedegraads	64,0
Eerstegraads	6,0

<sup>1</sup> Vraag 1.2 tot en met 1.6

### 6.4.3 Domeinspecifieke kenmerken

In deze paragraaf worden de domeinspecifieke kenmerken van de docenten gepresenteerd, aan de hand van tabellen. Voor de items binnen een vraag die samen een bepaald concept representeren (op basis van andere onderzoeken zoals Peil.Rekenen-Wiskunde in PO, PISA of TIMSS) zijn Cronbach's alfa en Guttman's Lambda-2 berekend en vermeld onder de tabel. Indien we voor deze maten een waarde van 0,60 of hoger vinden, is een nieuwe samengestelde variabele voor dit concept geconstrueerd. Ook het gemiddelde van dit construct is in de tabel opgenomen.

Ook in de docentenvragenlijst liepen de oorspronkelijke voorgegeven antwoorden vaak van positief naar negatief of van meer naar minder (bijv. 1=zeer mee eens, 4=zeer mee oneens). Achteraf zijn deze antwoorden omgecodeerd, behalve als het item negatief geformuleerd was. Indien van toepassing staat onder de tabel vermeld hoe de oorspronkelijke antwoorden in de vragenlijst luiden en welke items achteraf omgecodeerd zijn. In het algemeen geldt voor de waardes in de tabellen: **hoe hoger de score, hoe méér / hoe positiever.**

#### Wiskunde in de getoetste klas

Tabel 6.23 – Aantal leerlingen in getoetste klas/minuten wiskunde per week<sup>1</sup>, gemiddelden (sd) en range

Kenmerk	n	Gemiddelde (sd)	Min	Max
Aantal leerlingen getoetste klas	100	22,8 (5,1)	5,0	31,0
Aantal minuten wiskunde per week	99	155,3 (32,7)	80,0	300,0

<sup>1</sup> Vraag 2.1 en 2.3

Tabel 6.24 – Opleidingstype(n)<sup>1</sup> getoetste klas, in percentages

Opleidingstype (n=100)	%
Praktijkonderwijs	3,0
Praktijkonderwijs/vmbo bb/vmbo kb	1,0
Vmbo bb	6,0
Vmbo bb/vmbo kb	8,0
Vmbo bb/vmbo kb/vmbo tl	1,0
Vmbo bb/vmbo kb/vmbo gl/vmbo tl	2,0
Vmbo bb/vmbo kb/vmbo gl/vmbo tl/havo (incl technasium)/vwo	1,0
Vmbo kb	7,0
Vmbo kb /vmbo gl	1,0
Vmbo kb/vmbo tl	1,0
Vmbo kb/vmbo gl/vmbo tl	1,0
Vmbo gl	4,0
Vmbo gl/vmbo tl	2,0
Vmbo tl	18,0
Vmbo tl/havo	4,0
Vmbo tl/havo/vwo	2,0
Havo	13,0
Havo/vwo	5,0
Vwo	13,0
Gymnasium	6,0
Gymnasium/vwo	1,0

<sup>1</sup> Vraag 2.2

Tabel 6.25 – Onderwijstype getoetste klas, gecategoriseerd, in percentages

Onderwijstype (n=100)	%
Pro	4,0
Vmbo basis en/of kader	24,0
Vmbo gemengd en/of theoretisch	28,0
Havo, havo/vwo, havo/vwo/vmbo theoretisch	24,0
Vwo en/of gymnasium	20,0

1=pro, pro+basis/kader / 2=basis, kader, basis+kader, kader+gl, kader+tl, basis+kader+tl / 3=gl, tl, gl+tl, gl+tl+kader, gl+tl+kader+basis, alles minus pro / 4= havo, havo+tl, havo+vwo, havo+tl\_vwo / 5= vwo, gymnasium, vwo+gymnasium

Tabel 6.26 – Tijdsbesteding wiskundedomeinen dit schooljaar<sup>1</sup>, gemiddelde percentages (sd) en range

Domein (n=100)	Gemiddeld % (sd)	% Min	% Max
Getallen	16,4 (9,0)	0,0	55,0
Verhoudingen	16,3 (7,2)	0,0	40,0
Meten en meetkunde	24,2 (9,5)	0,0	60,0
Verbanden en formules	30,4 (10,1)	0,0	60,0
Andere reken-en wiskundeonderwerpen	12,8 (11,1)	0,0	58,0

<sup>1</sup> Vraag 2.4

Tabel 6.27 – Domein dit schooljaar niet behandeld<sup>1</sup>, in percentages

Domein (n=100)	%
Getallen	3,0
Verhoudingen	4,0
Meten en meetkunde	2,0
Verbanden en formules	2,0
Andere reken- en wiskundeonderwerpen	26,0

<sup>1</sup> Vraag 2.4

Tabel 6.28 – Hoofdzakelijk gebruikte reken- en wiskundemethode in getoetste klas<sup>1</sup>, in percentages

Methode (n=100)	%
Getal en Ruimte	58,0
Moderne Wiskunde	33,0
MathPlus	0,0
KERN Wiskunde	1,0
Wageningse methode	0,0
Bettermarks	5,0
SmartWiskunde	0,0
SmartRekenen	0,0
Startrekenen 1F	0,0
Startrekenen 2F	0,0
Op weg naar 1F	0,0
Andere reken- en wiskundemethode: - Snappet - Learnbeat - Wereld in Getallen	3,0

<sup>1</sup> Vraag 2.6

Tabel 6.29 – Geschiktheid wiskundemethode<sup>1</sup>, gemiddelden (sd) en range

Geschikt voor: (n=100)	Gemiddelde (sd)	Min	Max
a. uw eigen instructiemethode	4,1 (0,5)	3,0	5,0
b. uw eigen kennis en vaardigheden	4,3 (0,5)	3,0	5,0
c. uw overtuigingen over wiskundeonderwijs	3,9 (0,7)	1,0	5,0
d. de leerlingen in uw klas	3,8 (0,6)	2,0	5,0
Construct <i>Geschiktheid wiskundemethode</i> <sup>2</sup>	4,0 (0,4)	3,0	5,0

<sup>1</sup> Vraag 2.7, oorspronkelijke antwoorden: 1=zeer goed, 5=zeer slecht. Alle items omgecodeerd.

<sup>2</sup> Construct overgenomen uit Peil.Rekenen-Wiskunde 2019 (PO). Cronbach's alfa: 0,73, Guttman's Lambda 2: 0,74

Tabel 6.30 – Leerlingactiviteiten tijdens de wiskundeles<sup>1</sup>, gemiddelden (sd) en range

Activiteit	Gemiddelde (sd)	Min	Max
a. Luisteren naar mijn uitleg over nieuwe wiskundeleerstof	3,9 (0,4)	2,0	4,0
b. Luisteren naar mijn uitleg over hoe je opgaven kunt oplossen	3,6 (0,6)	2,0	4,0
c. Uit het hoofd leren van regels, procedures en feiten (zoals automatiseren)	2,6 (0,9)	1,0	4,0
d. Het zelfstandig oefenen van procedures	3,8 (0,4)	2,0	4,0
e. Het zelfstandig toepassen van het geleerde bij nieuwe probleemsituaties	3,2 (0,8)	1,0	4,0
f. Klassikaal aan opgaven werken onder mijn directe begeleiding	3,3 (0,8)	1,0	4,0
g. Werken in heterogene vaardigheidsgroepen	2,3 (1,1)	1,0	4,0
h. Werken in homogene vaardigheidsgroepen	2,2 (1,0)	1,0	4,0
i. Uitwerkingen opschrijven op het bord	2,8 (1,1)	1,0	4,0
j. Uitleggen hoe ze op een gegeven antwoord zijn gekomen	3,4 (0,7)	2,0	4,0
k. Na te denken over verschillende oplossingsmanieren voor een wiskundig probleem	2,7 (0,8)	1,0	4,0
l. Wiskundige problemen op te lossen waarbij niet direct duidelijk is hoe de oplossing gevonden kan worden	2,2 (0,9)	1,0	4,0
Construct <i>Leerkrachtgestuurde klassikale lesactiviteiten</i> <sup>2</sup> (a, b, c, d en f)	3,5 (0,4)	2,0	4,0
Construct <i>Probleemoplossen</i> <sup>3</sup> (k en l)	2,4 (0,8)	1,0	4,0

<sup>1</sup> Vraag 2.8, oorspronkelijke antwoorden: 1=alle of bijna alle lessen, 2=ongeveer helft van de lessen, 3=sommige lessen, 4=(bijna) nooit. Alle items omgecodeerd.

<sup>2</sup> Construct *Leerkrachtgestuurde klassikale lesactiviteiten* overgenomen uit Peil.Rekenen-Wiskunde 2019 (PO), maar nu deels andere items (a, b, c, d, f i.p.v. a, b, c, f, i, j). Cronbach's alfa: 0,60 Guttman's Lambda 2: 0,62.

<sup>3</sup> Construct *Probleemoplossen*, toegevoegd door begeleidingscommissie. Cronbach's alfa: 0,73 Guttman's Lambda 2: 0,73. In TIMSS, waar deels deze items ook bevroegd werden (niet meer sinds 2019) is er geen construct van gemaakt.

Tabel 6.31 – Instructiemodel tijdens de wiskundeles<sup>1</sup>, gemiddelden (sd) en range

Activiteit	Gemiddelde (sd)	Min	Max
Directe instructiemodel	3,8 (0,5)	2,0	4,0
Vormen van ontdekkend/onderzoekend leren	2,1 (0,8)	1,0	4,0

<sup>1</sup> Vraag 2.9, oorspronkelijke antwoorden: 1=alle of bijna alle lessen, 2=ongeveer helft van de lessen, 3=sommige lessen, 4=(bijna) nooit. Alle items omgecodeerd.

Tabel 6.32 – Aanbod meerdere oplossingsstrategieën<sup>1</sup>, in percentages

Aanbod (n=100)	%
Nee, ik bied meestal één oplossingsstrategie aan	13,0
Ja, maar alleen voor de <u>goed</u> presterende leerlingen in wiskunde	13,0
Ja, maar alleen voor de <u>gemiddeld</u> en <u>goed</u> presterende leerlingen in wiskunde	13,0
Ja, maar alleen voor de <u>zwak</u> en <u>gemiddeld</u> presterende leerlingen in wiskunde	4,0
Ja, maar alleen voor de <u>zwak</u> presterende leerlingen in wiskunde	1,0
Ja, ik bied alle leerlingen meerdere oplossingsstrategieën	56,0

<sup>1</sup>Vraag 2.10

## ICT-gebruik in de wiskundeles

Tabel 6.33 – Gebruik digitale apparaten in de wiskundeles<sup>1</sup>, in percentages

	%
Gebruik digitale apparaten (n=100)	74,0

<sup>1</sup>Vraag 3.1

Tabel 6.34 – ICT-gebruik in de wiskundeles<sup>1</sup>, gemiddelden (sd) en range

Activiteit (n=100)	Gemiddelde (sd)	Min	Max
a. Het oefenen van opgaven en procedures	2,3 (1,3)	1,0	4,0
b. Het oplossen van complexe problemen of problemen uit het dagelijks leven	1,6 (0,9)	1,0	4,0
c. Het maken van grafieken, tabellen of andere manieren om data te presenteren	1,6 (0,9)	1,0	4,0
d. Het gebruiken van games waarin wiskundeopgaven of -concepten worden gebruikt	1,4 (0,8)	1,0	4,0
e. Het lezen van instructie of het kijken van instructie-video's	2,1 (1,1)	1,0	4,0
f. Het maken van wiskundetoetsen	1,4 (0,7)	1,0	3,0
Construct <i>Intensiteit ICT-gebruik</i> <sup>2</sup>	1,8 (0,7)	1,0	3,8

<sup>1</sup> Vraag 3.2. Oorspronkelijke antwoorden: 1=tenminste 1 keer per week, 2=1 of 2 keer per maand, 3=paar keer per jaar, 4=(bijna) nooit. Alle items omgecodeerd. Docenten die bij vraag 3.1 (Tabel 6.33) aangaven geen ICT te gebruiken zijn ondergebracht bij '(bijna) nooit'.

<sup>2</sup> Construct overgenomen uit Peil.Rekenen-Wiskunde 2023 (PO) Cronbach's alfa: 0,83, Guttman's Lambda 2: 0,84.

Tabel 6.35 – Gebruik rekenmachine tijdens de wiskundeles<sup>1</sup>, in percentages

Gebruik (n=100)	%
Ja, is altijd toegestaan	56,0
Ja is beperkt toegestaan	43,0
Nee	1,0

<sup>1</sup>Vraag 3.3

## Wiskundehuiswerk

Tabel 6.36 – Frequentie huiswerk voor wiskunde<sup>1</sup>, gemiddelden (sd) en range

	Gemiddelde (sd)	Min	Max
Frequentie (n=100)	3,6 (0,8)	1,0	4,0

<sup>1</sup> Vraag 4.1, oorspronkelijke antwoorden: 1=alle of bijna alle lessen, 2=ongeveer helft van de lessen, 3=sommige lessen, 4=(bijna) nooit. Alle items omgecodeerd. 6% geeft (bijna) nooit huiswerk op

Tabel 6.37 – Hoeveelheid huiswerk voor wiskunde per keer<sup>1</sup>, in percentages

Aantal minuten (n=100)	%
Geeft (bijna) nooit huiswerk op	6,0
Minder dan 30 minuten	67,0
31-60 minuten	27,0
61-90 minuten	0,0
Meer dan 90 minuten	0,0

<sup>1</sup>Vraag 4.2, gecombineerd met vraag 4.1

Tabel 6.38 – Feedback huiswerk<sup>1</sup>, gemiddelden (sd) en range

Soorten feedback (n=94)	Gemiddelde (sd)	Min	Max
Het huiswerk corrigeren en feedback geven aan de leerlingen	2,1 (0,7)	1,0	3,0
Het huiswerk bespreken in de klas	2,5 (0,6)	1,0	3,0
Controleren of leerlingen het huiswerk gemaakt hebben	2,3 (0,6)	1,0	3,0
Ik laat leerlingen zelf hun huiswerk nakijken	2,7 (0,6)	1,0	3,0

<sup>1</sup> Vraag 4.3, oorspronkelijke antwoorden: 1= altijd of bijna altijd, 2=soms, 3=(bijna) nooit. Alle items omgecodeerd.



## Kennis en vaardigheden docent

Tabel 6.39 – Gevolgde en gewenste (bij)scholing<sup>1</sup>, in percentages

Onderwerp (n=100)	% afgelopen 2 jaar gevolgd	% behoefte in de toekomst
De inhoud van het wiskundeonderwijs	34,0	35,0
Pedagogiek/didactiek van het wiskundeonderwijs	39,0	43,0
Het reken-wiskundecurriculum (bijv. kerndoelen, referentieniveaus)	28,0	31,0
Integratie van ICT in het wiskundeonderwijs	35,0	54,0
Het verbeteren van kritisch denken en de oplossingsvaardigheden van leerlingen	30,0	56,0
Het meten en beoordelen van leervorderingen in wiskunde	29,0	49,0
Inspelen op de behoeften van individuele leerlingen bij wiskunde	30,0	51,0

<sup>1</sup> Vraag 5.1

Tabel 6.40 – Tijdsbesteding (bij)scholing<sup>1</sup>, in percentages

Tijd (n=100)	%
Geen	29,0
Minder dan 6 uur	18,0
6 -15 uur	15,0
16-35 uur	10,0
Meer dan 35 uur	28,0

<sup>1</sup> Vraag 5.2

Tabel 6.41 – Zelfvertrouwen wiskunde-instructie<sup>1</sup>, gemiddelden (sd) en range

Instructie-onderdeel (n=100)	Gemiddelde (sd)	Min	Max
a. Leerlingen inspireren om wiskunde te leren	2,7 (0,6)	1,0	4,0
b. Leerlingen verschillende oplossingsstrategieën aanreiken	2,8 (0,7)	1,0	4,0
c. Uitdagende wiskundetaken geven aan leerlingen die bovengemiddeld presteren	2,3 (0,7)	1,0	4,0
d. Mijn instructie aanpassen om de betrokkenheid van leerlingen bij de les te vergroten	2,8 (0,7)	1,0	4,0
e. Leerlingen het belang en nut van wiskunde te laten inzien	2,5 (0,7)	1,0	4,0
f. Het evalueren van het begrip van leerlingen in wiskunde	2,6 (0,7)	1,0	4,0
g. Het verbeteren van het begrip van zwakke leerlingen in wiskunde	2,8 (0,7)	2,0	4,0
h. Het ontwikkelen van hogere denkvaardigheden bij leerlingen	2,4 (0,7)	1,0	4,0
Construct Zelfvertrouwen wiskunde-instructie <sup>2</sup>	2,6 (0,4)	1,8	3,6

<sup>1</sup> Vraag 5.3, oorspronkelijke antwoorden: 1=zeer hoog, 2=hoog, 3= gemiddeld, 4=laag. Alle items omgecodeerd.

<sup>2</sup> Construct grotendeels overgenomen uit Peil.Rekenen-Wiskunde 2019 (PO), Cronbach's alfa: 0,73, Guttman's Lambda 2: 0,74

Tabel 6.42 – Evaluatie prestaties en leerproces<sup>1</sup>, gemiddelden (sd) en range

Evaluatiebron (n=100)	Gemiddelde (sd)	Min	Max
a. Ik gebruik scores en/of antwoorden op wiskundetoetsen in het kader van een leerlingvolgsysteem om te evalueren of de leerdoelen bereikt zijn	2,2 (0,7)	1,0	3,0
b. Ik analyseer de antwoorden op methodegebonden wiskundetoetsen om te evalueren of de leerdoelen van dat blok bereikt zijn	2,2 (0,7)	1,0	3,0
c. Ik evalueer regelmatig of alle leerlingen de lesdoelen bereikt hebben op basis van hun ingeleverde werk	2,3 (0,7)	1,0	3,0
d. Ik evalueer of alle leerlingen de lesdoelen bereikt hebben op basis van (informele) observaties tijdens de wiskundeles	2,6 (0,5)	2,0	3,0
e. Ik voer diagnostische gesprekken om te evalueren of specifieke leerlingen de lesdoelen bereikt hebben	3,1 (0,7)	1,0	3,0
f. Ik evalueer of de door mij gekozen manieren van instructie en verwerking effectief waren voor de meerderheid van de leerlingen in de klas	2,5 (0,6)	1,0	3,0
g. Ik evalueer of een specifieke manier van instructie effectief was voor specifieke leerlingen	2,1 (0,6)	1,0	3,0
Construct <i>Intensiteit evaluatie prestaties en leerproces</i> <sup>2</sup>	2,3 (0,4)	1,4	3,0

<sup>1</sup> Vraag 6.1, oorspronkelijke antwoorden: 1=zeer, 2= enigszins, 3=niet. Alle items omgecodeerd.

<sup>2</sup> Construct overgenomen uit Peil.Rekenen-Wiskunde 2019 PO ('DSAQ-schaal evaluatie', 5-puntsschaal), Cronbach's alfa: 0,67, Guttman's Lambda 2: 0,68

Tabel 6.43 – Belang van beoordelingsstrategieën<sup>1</sup>, gemiddelden (sd) en range

Strategie (n=100)	Gemiddelde (sd)	Min	Max
Het observeren van leerlingen als zij aan het werk zijn	2,9 (0,3)	2,0	3,0
Leerlingen vragen laten beantwoorden tijdens de les	2,8 (0,4)	2,0	3,0
Het doornemen van hun dagelijkse werk en/of huiswerk	2,3 (0,6)	1,0	3,0
Korte, tussentijdse schriftelijke toetsen	1,8 (0,7)	1,0	3,0
Methodegebonden toetsen aan het eind van een periode of blok	2,7 (0,6)	1,0	3,0
Genormeerde toetsen in kader van een leerlingvolgsysteem, zoals bijvoorbeeld van Cito	1,6 (0,7)	1,0	3,0
Langlopende wiskundeprojecten	1,4 (0,5)	1,0	3,0

<sup>1</sup> Vraag 6.2, oorspronkelijke antwoorden: 1=veel, 2= enigszins, 3=geen. Alle items omgecodeerd.

Tabel 6.44 – Gebruik typen toetsen<sup>1</sup>, gemiddelden (sd) en range

Type (n=100)	Gemiddelde (sd)	Min	Max
Genormeerde toetsen (bijvoorbeeld van Cito)	1,3 (0,5)	1,0	3,0
Methodegebonden toetsen	2,6 (0,7)	1,0	3,0
Zelfgemaakte toetsen	2,6 (0,6)	1,0	3,0

<sup>1</sup> Vraag 6.3, oorspronkelijke antwoorden: 1=veel, 2=soms, 3=(bijna) nooit. Alle items omgecodeerd.

Tabel 6.45 – Differentiatie<sup>1</sup>, gemiddelden (sd) en range

Mate van toepassing (n=99)	Gemiddelde (sd)	Min	Max
a. Ik benut de mogelijkheden die de methode biedt voor differentiatie voor leerlingen die in mijn klas relatief bovengemiddeld in wiskunde presteren	2,2 (0,6)	1,0	3,0
b. Ik benut de mogelijkheden die de methode biedt voor differentiatie voor de relatief zwakke leerlingen in wiskunde in mijn klas	2,2 (0,6)	1,0	3,0
c. Ik stel bewust open vragen tijdens de klassikale instructie	2,8 (0,4)	2,0	3,0
d. Ik stel bewust vragen van verschillende moeilijkheidsgraad tijdens de klassikale instructie	2,6 (0,6)	1,0	3,0
e. Ik bied leerlingen die in mijn klas relatief bovengemiddeld in wiskunde presteren verrijkingsofdrachten	2,1 (0,6)	1,0	3,0
f. Ik bied leerlingen die in mijn klas relatief zwak in wiskunde presteren extra oefenopdrachten	2,2 (0,6)	1,0	3,0
g. De leerlingen kunnen binnen mijn klas wiskunde op een hoger onderwijsniveau volgen (zoals kbl-wiskunde in een bbl-klas)	1,9 (0,9)	1,0	3,0
Construct <i>Differentiatie</i> <sup>2</sup>	2,1 (0,5)	1,2	3,0

<sup>1</sup> Vraag 6.4, oorspronkelijke antwoorden: 1=zeer, 2=enigszins, 3=niet. Alle items omgecodeerd.

<sup>2</sup> Enkele losse items overgenomen uit verschillende 'DSAQ-schalen' Peil.Rekenen-Wiskunde 2019, en nieuwe items toegevoegd. Cronbach's alfa: 0,74, Guttman's Lambda 2: 0,74 zonder item c en d

Tabel 6.46 – Fixed/Growth mindset<sup>1</sup>, gemiddelden (sd) en range

Stelling (n=99)	Gemiddelde (sd)	Min	Max
a. Het maakt niet uit hoe intelligent leerlingen zijn, als leerlingen willen kunnen ze altijd goed worden in wiskunde	2,9 (0,6)	1,0	4,0
b. Hoe goed een leerling is in wiskunde ligt vast en je kunt niet veel doen om dit te veranderen	3,6 (0,6)	2,0	4,0
c. Je kunt leerlingen wel nieuwe dingen leren over wiskunde, maar je kunt niet veranderen hoe slim ze in wiskunde zijn	3,0 (0,8)	1,0	4,0
d. Elke leerling kan goed zijn in wiskunde	2,9 (0,8)	1,0	4,0
e. Een leerling kan zelf veranderen hoe goed hij/zij kan presteren in wiskunde	3,4 (0,6)	1,0	4,0
f. Aanleg voor wiskunde is iets waar leerlingen mee geboren worden, je kunt niet veel doen om ze beter in wiskunde te maken dan het niveau van de aangeboren aanleg	3,4 (0,7)	2,0	4,0
Construct <i>Growth mindset</i> <sup>2</sup>	3,2 (0,4)	2,2	4,0

<sup>1</sup> Vraag 6.5, oorspronkelijke antwoorden: 1=zeer mee eens, 4=zeer mee oneens. Item a, d, e (Growth) omgecodeerd. Items b,c en f (Fixed): niet omgecodeerd. Hoe hoger de score, hoe meer Growth-gerichte mindset (hier opgevat als tegenhanger van een Fixed'-gerichte mindset).

<sup>2</sup> Cronbach's alfa: 0,68 Guttman's Lambda 2:0,69

Tabel 6.47 – Fixed/Growth mindset<sup>1</sup>, in percentages zeer mee eens

Stelling (n=99)	% zeer mee eens
a. Het maakt niet uit hoe intelligent leerlingen zijn, als leerlingen willen kunnen ze altijd goed worden in wiskunde	15,3
b. Hoe goed een leerling is in wiskunde ligt vast en je kunt niet veel doen om dit te veranderen	0,0
c. Je kunt leerlingen wel nieuwe dingen leren over wiskunde, maar je kunt niet veranderen hoe slim ze in wiskunde zijn	1,0
d. Elke leerling kan goed zijn in wiskunde	23,2
e. Een leerling kan zelf veranderen hoe goed hij/zij kan presteren in wiskunde	45,5
f. Aanleg voor wiskunde is iets waar leerlingen mee geboren worden, je kunt niet veel doen om ze beter in wiskunde te maken dan het niveau van de aangeboren aanleg	0,0

<sup>1</sup> Vraag 6.5

## Leerlingen in het tweede leerjaar en de coronapandemie

Tabel 6.48 – Veranderingen door pandemie<sup>1</sup>, gemiddelden (sd) en range

Als gevolg van de coronapandemie... (n=55)	Gemiddelde (sd)	Min	Max
a. ... hebben leerlingen leervertraging opgelopen op het gebied van wiskunde	3,3 (0,8)	1,0	4,0
b. ...zijn de wiskundeprestaties er op achteruit gegaan	3,2 (0,9)	1,0	4,0
c. ...zijn verschillen tussen leerlingen in wiskundeprestaties groter geworden	3,3 (0,8)	1,0	4,0
d. ...is de motivatie om te leren onder leerlingen afgenomen	3,3 (0,8)	1,0	4,0
e. ...is de concentratie van leerlingen tijdens de wiskundeles (online en/of fysiek) afgenomen	3,4 (0,8)	1,0	4,0
f. ...ondervinden leerlingen meer moeite met het plannen van hun werk (zoals huiswerk of leren voor een toets)	3,3 (0,8)	1,0	4,0
g. ... zijn leerlingen minder zorg aan hun huiswerk gaan besteden	3,2 (0,7)	1,0	4,0
h. ... wordt het huiswerk minder vaak gemaakt	3,1 (0,8)	1,0	4,0
i. ...is het aantal conflicten tussen leerlingen onderling toegenomen	2,6 (1,1)	1,0	4,0
j. ...is het aantal conflicten tussen leerlingen en docenten en/of ander personeel toegenomen	2,5 (1,0)	1,0	4,0
k. ... worden leerlingen sneller boos of verdrietig	2,9 (0,8)	1,0	4,0
l. ... is pesten onder leerlingen toegenomen (fysiek of online)	2,5 (0,9)	1,0	4,0
m. ... is spijbelen toegenomen	2,6 (1,0)	1,0	4,0
n. ... is de bereidheid om zich te houden aan de gedragsregels van de school afgenomen	3,2 (0,7)	1,0	4,0
- Construct <i>Verandering in leerprestaties</i> <sup>2</sup>	3,2 (0,7)	1,0	4,0
- Construct <i>Verandering in leergedrag</i> <sup>2</sup>	3,3 (0,6)	1,0	4,0
- Construct <i>Verandering in leerlinggedrag</i> <sup>2</sup>	2,7 (0,7)	1,0	4,0

<sup>1</sup> Vraag 7.2. 55 van de 99 docenten (56%) gaf ook al wiskunde in het tweede leerjaar op dezelfde school, alleen deze docenten hebben 7.2 beantwoord, oorspronkelijke antwoorden: 1=zeer mee eens, 4=zeer mee oneens. Alle items omgecodeerd.

<sup>2</sup> Cronbach's alfa alle items: 0,91 Guttman's Lambda 2:0,91 (maar: 14 items op 56 respondenten). Inhoudelijk 3 subschalen:  
 1 prestaties (a-c) Cronbach's alfa: 0,83, Guttman's Lambda 2: 0,84 (P corr. met schaal 2=0,78, P corr. met schaal 3=0,49)  
 2 leergedrag (d-h) Cronbach's alfa: 0,86 Guttman's Lambda 2:0,86 (P corr. met schaal 3=0,54)  
 3 leerlinggedrag (i-n) Cronbach's alfa: 0,86 Guttman's Lambda 2:0,87

Tabel 6.49 – Gepercipieerde veranderingen door pandemie<sup>1</sup>, in percentages zeer mee eens

Stelling (n=55)	% zeer mee eens
a. ... hebben leerlingen leervertraging opgelopen op het gebied van wiskunde	41,8
b. ...zijn de wiskundeprestaties er op achteruit gegaan	41,8
c. ...zijn verschillen tussen leerlingen in wiskundeprestaties groter geworden	47,3
d. ...is de motivatie om te leren onder leerlingen afgenomen	45,5
e. ...is de concentratie van leerlingen tijdens de wiskundeles (online en/of fysiek) afgenomen	52,7
f. ...ondervinden leerlingen meer moeite met het plannen van hun werk (zoals huiswerk of leren voor een toets)	49,1
g. ... zijn leerlingen minder zorg aan hun huiswerk gaan besteden	40,0
h. ... wordt het huiswerk minder vaak gemaakt	34,5
i. ...is het aantal conflicten tussen leerlingen onderling toegenomen	25,5
j. ...is het aantal conflicten tussen leerlingen en docenten en/of ander personeel toegenomen	14,5
k. ... worden leerlingen sneller boos of verdrietig	20,0
l. ... is pesten onder leerlingen toegenomen (fysiek of online)	16,4
m. ... is spijbelen toegenomen	20,0
n. ... is de bereidheid om zich te houden aan de gedragsregels van de school afgenomen	36,4

<sup>1</sup>Vraag 7.2

#### Domeinspecifieke kenmerken van het onderwijsleerproces in relatie tot het onderwijsniveau

In Tabel 6.50 zijn alle responderende docenten ingedeeld in drie categorieën op basis van het onderwijsniveau in de tweede klas waaraan ze wiskundeles gegeven (Tabel 6.25, gecompriemd).

Tabel 6.50 – Indeling docenten naar onderwijsniveau tweede klas

	N	%
Pro basis kader	28	28,0%
Gemengd en/of theoretisch	28	28,0%
Havo / vwo	44	44,0%
Totaal	100	100%

In de tabellen hierna laten we zien hoe de domeinspecifieke kenmerken van het onderwijsleerproces (volgens de docenten) samenhangt met dit onderwijsniveau in leerjaar 2. Tabel 6.51 toont eerst de scores van alle constructvariabelen (schaalscores), gemiddeld per onderwijsniveau van de klas van de docenten. De laatste drie constructvariabelen in deze tabel zijn slechts aan een deel van de docenten voorgelegd, vandaar de lagere n.

Tabel 6.51 – Constructvariabelen: gemiddelde (sd) per onderwijsniveau klas en totaal

Constructen	Pro, vmbo-b/k	Vmbo-g/t	Havo/vwo	Totaal
Geschiktheid wiskundemethode (n=100)	4,0 (0,4)	4,0 (0,5)	4,0 (0,4)	4,0 (0,4)
Leerkrachtgestuurde klassikale lesactiviteiten (n=100)	3,5 (0,4)	3,4 (0,5)	3,5 (0,4)	3,5 (0,4)
Probleemoplossen (n=100)	2,1 (0,7)	2,4 (0,8)	2,6 (0,7)	2,4 (0,8)
Intensiteit ICT-gebruik (n=100)	1,9 (0,7)	1,7 (0,7)	1,7 (0,7)	1,8 (0,7)
Zelfvertrouwen wiskunde-instructie (n=100)	2,7 (0,4)	2,6 (0,4)	2,5 (0,3)	2,6 (0,4)
Intensiteit evaluatie prestaties & leerproces (n=100)	2,4 (0,4)	2,2 (0,3)	2,3 (0,4)	2,3 (0,4)
Differentiatie (n=99)	2,3 (0,5)	2,1 (0,4)	2,1 (0,4)	2,1 (0,5)
Growth mindset (n=99)	3,2 (0,4)	3,2 (0,4)	3,2 (0,5)	3,2 (0,4)
Verandering in leerprestaties (n=55)	3,2 (0,6)	3,5 (0,6)	3,1 (0,8)	3,2 (0,7)
Verandering in leergedrag (n=55)	3,3 (0,6)	3,4 (0,6)	3,2 (0,7)	3,3 (0,6)
Verandering in leerlinggedrag (n=55)	3,0 (0,7)	3,0 (0,7)	2,4 (0,6)	2,7 (0,7)

De volgende tabellen bevatten de frequentieverdelingen van de andere domeinspecifieke variabelen betreffende het onderwijsleerproces in de klas.

Tabel 6.52 - Tijdsbesteding domeinen rekenen-wiskunde (gemiddelden, range en s.d)\*

Kenmerken	Pro, vmbo-b/k (n=28)	Vmbo-g/t (n=28)	Havo/vwo (n=44)	Totaal (N=100)
Aantal minuten wiskunde per week	167,0 (80-300; 44,5)	150,4 (100-225; 26,5)	151,0 (90-225; 25,5)	155,4 (80-300; 32,7)
Tijdsbesteding wiskundedomeinen				
- Getallen	20,2 (0-55; 10,7)	17,1 (8-40; 7,5)	13,5 (0-30; 7,8)	16,4 (0-55; 9,0)
- Verhoudingen	19,4 (10-40; 6,6)	17,9 (10-30; 5,8)	13,2 (0-30; 7,3)	16,3 (0-40; 7,2)
- Meten en meetkunde	24,6 (10-40; 8,4)	27,6 (15-60; 10,1)	21,8 (0-40; 9,3)	24,2 (0-60; 9,5)
- Verbanden en formules	23,6 (10-37,5; 5,6)	26,5 (10-40; 7,9)	37,2 (20-60; 9,5)	30,4 (10-60; 10,1)
- Andere reken- en wiskundeonderwerpen	12,2 (0-30; 9,0)	11,0 (0-30; 9,0)	14,2 (0-58; 13,3)	12,8 (0-58; 11,1)

\* Zie ook Tabel 6.23 en 6.26.

Tabel 6.53 - Leermethode en aanvullend materiaal (%)\*

Kenmerken	Pro, vmbo-b/k (n=28)	Vmbo-g/t (n=28)	Havo/vwo (n=44)	Totaal (N=100)
Hoofdzakelijk gebruikte reken- en wiskundemethode:				
- Getal en ruimte	61%	57%	57%	58%
- Moderne wiskunde	29%	39%	32%	33%
- KERN Wiskunde	0%	0%	1%	1%
- Bettermarks	4%	0%	9%	5%
- Andere reken- en wiskundemethode**	7%	4%	0%	3%
Kunnen leerlingen van de getoetste klas tijdens uw wiskundelessen gebruik maken digitale apparaten (computers, tablets en/of smartphones)? - aandeel Ja	86%	68%	71%	74%
Is het voor de leerlingen uit de getoetste klas toegestaan een rekenmachine te gebruiken tijdens de wiskundeles?				
- Ja, is altijd toegestaan	68%	82%	32%	56%
- Ja, is beperkt toegestaan	29%	18%	68%	43%
- Nee	4%	0%	0%	1%

\* Zie ook Tabel 6.28, 6.33 en 6.35.

\*\* Wereld in getallen (1), Snappet (1) en LearnBeat (1). De eerste twee bij pro, vmbo-b/k, de laatste bij vmbo-g/t.

Tabel 6.54 - Werkvormen: overige lesactiviteiten (gemiddelde\*, s.d.)\*\*

Kenmerken	Pro, vmbo-b/k (n=28)	Vmbo-g/t (n=28)	Havo/vwo (n=44)	Totaal (N=100)
Werken in heterogene vaardigheidsgroepen	2,4 (1,1)	2,4 (1,2)	2,3 (1,1)	2,3 (1,1)
Werken in homogene vaardigheidsgroepen	2,4 (1,0)	1,9 (0,9)	2,2 (1,0)	2,2 (1,0)
Uitwerkingen opschrijven op het bord	2,9 (1,1)	2,8 (1,1)	2,8 (1,2)	2,8 (1,1)
Uitleggen hoe ze op een gegeven antwoord zijn gekomen.	3,4 (0,7)	3,5 (0,7)	3,4 (0,8)	3,4 (0,7)

\* 1=(bijna) nooit, 2=sommige lessen, 3=ongeveer de helft van de lessen en 4=alle of bijna alle lessen (na omcodering)

\*\* Zie ook Tabel 6.30, item g, h, i en j.

Tabel 6.55 - Directe instructie (%)\*

Kenmerken	Pro, vmbo-b/k (n=28)	Vmbo-g/t (n=28)	Havo/vwo (n=44)	Totaal (N=100)
Frequentie directe instructiemodel				
- Alle of bijna alle lessen	86%	79%	75%	79%
- Ongeveer de helft van de lessen	14%	18%	20%	18%
- Sommige lessen	0%	4%	5%	3%
- (bijna) nooit	0%	0%	0%	0%
Frequentie vormen van ontdekkend/onderzoeken leren				
- Alle of bijna alle lessen	4%	4%	5%	4%
- Ongeveer de helft van de lessen	21%	29%	23%	24%
- Sommige lessen	57%	50%	50%	52%
- (bijna) nooit	18%	18%	23%	20%

\* Zie ook Tabel 6.31.

Tabel 6.56 - Oplossingsstrategieën (%)\*

Kenmerken	Pro, vmbo-b/k (n=28)	Vmbo-g/t (n=28)	Havo/vwo (n=44)	Totaal (N=100)
Aanbod meerdere oplossingsstrategieën				
- Nee, ik bied meestal één oplossingsstrategie aan	14%	7%	16%	13%
- Ja, maar alleen voor de <u>goed</u> presterende leerlingen in wiskunde	18%	21%	5%	13%
- Ja, maar alleen voor de <u>gemiddeld</u> en <u>goed</u> presterende leerlingen in wiskunde	18%	11%	11%	13%
- Ja, maar alleen voor de <u>zwak</u> en <u>gemiddeld</u> presterende leerlingen in wiskunde	7%	4%	2%	4%
- Ja, maar alleen voor de <u>zwak</u> presterende leerlingen in wiskunde	0%	0%	2%	1%
- Ja, ik bied alle leerlingen meerdere oplossingsstrategieën	43%	57%	64%	56%

\* Zie ook Tabel 6.32.

Tabel 6.57 - Omvang en bespreking huiswerk\*

Kenmerken	Pro, vmbo-b/k (n=28)	Vmbo-g/t (n=28)	Havo/vwo (n=44)	Totaal (N=100)
Frequentie huiswerk (%)				
- Alle of bijna alle lessen	68%	82%	91%	82%
- Ongeveer de helft van de lessen	7%	7%	5%	6%
- Sommige lessen	14%	4%	2%	6%
- (bijna) nooit	11%	7%	2%	6%
Omvang huiswerk (%0				
- Minder dan 30 min	84%	73%	63%	71%
- 31-60 min	16%	27%	37%	29%
Feedback (gemiddelde**, s.d.)				
- Het huiswerk corrigeren en feedback geven aan de leerlingen	2,3 (0,7)	2,2 (0,7)	1,9 (0,6)	2,1 (0,7)
- Het huiswerk bespreken in de klas	2,6 (0,5)	2,5 (0,6)	2,3 (0,7)	2,5 (0,6)
- Controleren of leerlingen het huiswerk gemaakt hebben	2,6 (0,6)	2,2 (0,7)	2,1 (0,6)	2,3 (0,6)
- Ik laat leerlingen zelf hun huiswerk nakijken	2,6 (0,6)	2,5 (0,7)	2,9 (0,4)	2,7 (0,6)

\* Zie ook Tabel 6.36, 6.37 en 6.38.

\*\* 1=(bijna) nooit, 2= soms, 3= altijd of bijna altijd (na omcodering).



Tabel 6.58 - Vormen en doelen van evaluatie prestaties en toetsing (gemiddelden\*, s.d.)\*\*

Kenmerken	Pro, vmbo-b/k (n=28)	Vmbo-g/t (n=28)	Havo/vwo (n=44)	Totaal (N=100)
Belang beoordelingsstrategieën				
- Het observeren van leerlingen als zij aan het werk zijn	2,9 (0,3)	2,8 (0,4)	2,9 (0,3)	2,9 (0,3)
- Leerlingen vragen laten beantwoorden tijdens de les	3,0 (0,2)	2,7 (0,5)	2,9 (0,3)	2,8 (0,4)
- Het doornemen van hun dagelijkse werk en/of huiswerk	2,5 (0,5)	2,3 (0,6)	2,2 (0,6)	2,3 (0,6)
- Korte, tussentijdse schriftelijke toetsen	1,9 (0,7)	1,7 (0,6)	1,9 (0,7)	1,8 (0,7)
- Methodegebonden toetsen aan het eind van een periode of blok	2,8 (0,4)	2,5 (0,6)	2,7 (0,6)	2,7 (0,6)
- Genormeerde toetsen in kader van een leerlingvolgsysteem, zoals bijvoorbeeld van Cito	2,0 (0,7)	1,6 (0,6)	1,4 (0,5)	1,6 (0,7)
- Langlopende wiskundeprojecten	1,4 (0,5)	1,3 (0,5)	1,4 (0,6)	1,4 (0,5)
Gebruik typen toetsen				
- Genormeerde toetsen (bijvoorbeeld van Cito)	1,8 (0,6)	1,2 (0,4)	1,2 (0,4)	1,3 (0,5)
- Methodegebonden toetsen	2,8 (0,5)	2,8 (0,4)	2,3 (0,8)	2,6 (0,7)
- Zelfgemaakte toetsen	2,3 (0,7)	2,4 (0,6)	2,8 (0,4)	2,6 (0,6)

\* 1=geen, 2= enigszins, 3=veel (na omcodering)

\*\* Zie ook Tabel 6.43 en 6.44.

## 6.5 Schoolvragenlijst

### 6.5.1 Ontbrekende waarden

Er hebben 130 schoolvestigingen deelgenomen aan het onderzoek en alle leiders van deze vestigingen zijn uitgenodigd om de school(leiders)vragenlijst in te vullen. In totaal hebben 106 schoolleiders daaraan gehoor gegeven, maar vijf van hen hebben (veel) minder dan de helft van de vragen ingevuld. Zij zijn uit het bestand gehaald; de data van 101 schoolleiders resteren. Het databestand bevat verder nauwelijks ontbrekende waarden, omdat de vragenlijst digitaal was en de respondent geen vragen kon overslaan.

### 6.5.2 Algemene schoolkenmerken

In deze paragraaf presenteren we de algemene kenmerken van de scholen volgens de opgave van de schoolleiders. In de paragraaf daarna volgen de domeinspecifieke kenmerken.

Ook in de schoolvragenlijst liepen de oorspronkelijke voorgegeven antwoorden vaak van positief naar negatief of van meer naar minder (bijv. 1=zeer mee eens, 4=zeer mee oneens). Achteraf zijn deze antwoorden omgecodeerd, behalve als het item negatief geformuleerd was. Indien van toepassing staat onder de tabel vermeld hoe de oorspronkelijke antwoorden in de vragenlijst luiden en welke items achteraf omgecodeerd zijn. In het algemeen geldt voor de waardes in de tabellen: **hoe hoger de score, hoe méér / hoe positiever.**

Tabel 6.59 – Vestigingsgrootte<sup>1</sup>, gemiddelden (sd) en range

	Gemiddelde (sd)	Min	Max
Aantal leerlingen bij deze vestiging ingeschreven (n=101)	989,8 (529,0)	105	2189

<sup>1</sup> Vraag 1.1 van schoolvragenlijst

Tabel 6.60 – Opleidingsaanbod vestiging<sup>1</sup>, in percentages

Opleidingsaanbod (n=101)	%
Pro	3,0
Vmbo, vmbo+pro	26,7
Vmbo +havo, vmbo +havo +vwo	46,5
Havo+vwo	20,8
Vwo	3,0

<sup>1</sup> Vraag 1.2 (samengevat, want er zijn heel veel combinaties)

Tabel 6.61 – Onderwijsconcept<sup>1</sup>, in percentages

Onderwijsconcept (n=101)	%
Algemeen onderwijs/geen specifiek onderwijsconcept	83,2
Dalton	3,0
Jenaplan	0,0
Montessori	1,0
Vrijeschool	1,0
Agora onderwijs	1,0
Big Picture Learning	0,0
Kunskapsskolan	3,0
Competentiegericht leren	0,0
Ander onderwijsconcept, namelijk: Brede brugklas Een eigen doorontwikkeling van Kunskapsskolan Gepersonaliseerd onderwijs GO (Geloof in ontwikkeling?) IB Middle Years Program Ook tto havo en tto vwo Projectonderwijs We proberen leerlingen meer eigenaarschap te geven	7,9

<sup>1</sup> Vraag 1.3

Tabel 6.62 – Kenmerken schoolpopulatie<sup>1</sup>, in gemiddelde percentages

Kenmerk (n=101)	0 tot en met 10%	11 tot en met 25%	26 tot en met 50%	meer dan 50%
Komt uit een economisch achterstandsgezin	45,5%	29,7%	12,9%	11,9%
Komt uit een economisch welgesteld gezin	14,9%	23,8%	28,7%	32,7%

<sup>1</sup> Vraag 1.4

Tabel 6.63 – Aandeel leerlingen Nederlands eerste taal<sup>1</sup>, in percentages

Aandeel (n=101)	%
Meer dan 90%	60,4
76 tot en met 90%	19,8
51 tot en met 75%	13,9
26 tot en met 50%	5,0
25% of minder	1,0

<sup>1</sup> Vraag 1.5

Tabel 6.64 – Directe omgeving school<sup>1</sup>, in percentages

Omschrijving (n=101)	%
Stedelijk met een hoge bevolkingsdichtheid	19,8
Voorstad of buitenwijk van een stedelijk gebied	12,9
Grote of middelgrote stad	29,7
Kleine stad of dorp	29,7
Platteland	7,9

<sup>1</sup> Vraag 1.6

Tabel 6.65 – Prestatiegerichtheid schoolklimaat<sup>1</sup>, gemiddelden (sd) en range

Kenmerk (n=101)	Gemiddelde (sd)	Min	Max
a. De mate waarin docenten de onderwijsdoelen van de school kennen	3,6 (0,6)	2	5
b. De mate waarin docenten succesvol zijn in de realisatie van het curriculum van de school	3,7 (0,5)	2	5
c. Verwachtingen van docenten wat betreft de leerresultaten van de leerlingen	3,4 (0,6)	2	5
d. De mate waarin docenten in staat zijn leerlingen te inspireren	3,2 (0,5)	2	5
e. Betrokkenheid van ouders bij schoolactiviteiten	2,9 (0,8)	1	5
f. De mate waarin ouders ervoor zorgen dat hun kind in staat is deel te nemen aan het leerproces (bijv. voldoende slaap, ontbijt)	3,3 (0,8)	1	5
g. Verwachtingen van ouders wat betreft de leerresultaten van hun kind	3,5 (0,6)	2	5
h. Ondersteuning door ouders bij het leren van hun kind	3,1 (0,7)	2	5
i. De wil van leerlingen om goed te presteren op school	3,0 (0,6)	1	5
j. De mate waarin leerlingen in staat zijn om de leerdoelen van de school te halen	3,3 (0,6)	2	5
k. Respect van leerlingen voor medeleerlingen die zeer goed presteren	3,3 (0,6)	2	5
Construct <i>Prestatiegericht schoolklimaat</i> <sup>2</sup>	3,3 (0,3)	2,3	4,5
Subschaal <i>Prestatiegericht schoolklimaat_docenten</i> <sup>2</sup>	3,5 (0,4)	2,5	4,8
Subschaal <i>Prestatiegericht schoolklimaat_ouders</i> <sup>2</sup>	3,2 (0,6)	2,0	4,5
Subschaal <i>Prestatiegericht schoolklimaat_leerlingen</i> <sup>2</sup>	3,2 (0,5)	2,0	4,7

<sup>1</sup> Vraag 1.7, oorspronkelijke antwoorden: 1=erg hoog, 5=erg laag. Alle items omgecodeerd.

<sup>2</sup> Construct en subschalen overgenomen uit Peil.Rekenen-Wiskunde 2019 (PO) en TIMSS.

Cronbach's alfa totale schaal: 0,76, Guttman's Lambda 2: 0,78.

Cronbach's alfa docenten: 0,58, Guttman's Lambda 2: 0,58.

Cronbach's alfa ouders: 0,75 Guttman's Lambda 2: 0,76.

Cronbach's alfa leerlingen: 0,64 Guttman's Lambda 2: 0,64

Tabel 6.66 – Gebruik leerlingvolgsysteem<sup>1</sup>, in percentages

	%
Gebruikt leerlingvolgsysteem (n=101)	98

<sup>1</sup> Vraag 1.8

Tabel 6.67 Belang leerlingvolgsysteem<sup>1</sup>, gemiddelden (sd) en range

Activiteit (n=99)	Gemiddelde (sd)	Min	Max
Het volgen van de cognitieve ontwikkeling van leerlingen, zoals het bijhouden van toetsresultaten	2,7 (0,5)	1,0	3,0
Het volgen van de sociaal-emotionele ontwikkeling van leerlingen	2,3 (0,6)	1,0	3,0
Het volgen van studievoortgang en schoolloopbaan	2,5 (0,5)	1,0	3,0
Het volgen van leerlinggedrag zoals verzuim of pesten	2,7 (0,5)	1,0	3,0
Het opstellen van handelingsplannen	2,0 (0,6)	1,0	3,0
Het communiceren met leerlingen over bijvoorbeeld toetsresultaten, cijfers, roosters, huiswerk, uitval van lessen	2,6 (0,5)	1,0	3,0
Het communiceren met ouders over bijvoorbeeld leerlingresultaten, roosters, huiswerk, uitval van lessen	2,4 (0,6)	1,0	3,0
Het communiceren tussen docenten over bijvoorbeeld leerlingresultaten, leermaterialen onderwijsactiviteiten en -ontwikkelingen	2,2 (0,8)	1,0	3,0
Leerlingen toegang geven tot lesmateriaal	2,2 (0,8)	1,0	3,0
Construct <i>Belang leerlingvolgsysteem</i> <sup>2</sup>	2,4 (0,3)	1,7	3,0

<sup>1</sup> Vraag 1.9, oorspronkelijke antwoorden: 1=zeer belangrijk, 2=belangrijk, 3=nauwelijks/niet belangrijk. Alle items omgecodeerd.

<sup>2</sup> Cronbach's alfa: 0,72 Guttman's Lambda 2: 0,74

### 6.5.3 Domeinspecifieke schoolkenmerken

Tabel 6.68 – Aanbod extra wiskundelessen buiten schooltijd<sup>1</sup>, in percentages

	%
Wordt aangeboden (n=101):	61,4
<b>Indien aangeboden, aan wie (n=62)</b>	
Verrijking voor leerlingen die extra uitdaging kunnen gebruiken	14,5
Bijles voor leerlingen met een leerachterstand	98,5
Extra lessen voor leerlingen ongeacht hun eerdere prestaties	40,3

<sup>1</sup> Vraag 2.1 en 2.2

Tabel 6.69 – Gevolg pandemie voor aanbod extra wiskundelessen buiten schooltijd<sup>1</sup>, gemiddelden (sd) en range

Omvang aanbod (n=99):	Gemiddelde (sd)	Min	Max
Verrijking voor leerlingen die extra uitdaging kunnen gebruiken	2,0 (0,5)	1,0	3,0
Bijles voor leerlingen met een leerachterstand	2,9 (0,3)	1,0	3,0
Extra lessen voor leerlingen ongeacht hun eerdere prestaties	2,3 (0,5)	1,0	3,0

<sup>1</sup> Vraag 2.3, oorspronkelijke antwoorden: 1=afgenomen, 2=gelijk gebleven, 3=toegenomen. Geen items omgecodeerd.

Tabel 6.70 – Extra activiteiten voor leerlingen in tweede leerjaar<sup>1</sup>, in percentages

Activiteit (n=101)	%
Wiskundeclub	1,0
Wiskundewedstrijden, bijvoorbeeld de Kangoeroewedstrijd of Wiskundeolympiade	47,5
Schaakclub	14,9
Computerclub (bijvoorbeeld leren programmeren)	14,9
Schoolkrant of schoolblad	24,8
Schoolband, orkest of koor	52,5
Schooltoneel of schoolmusical	53,5
Sportclub of sportactiviteiten	76,2
Kunstclub of kunstactiviteiten	60,4
Maatschappelijke stage	45,5

<sup>1</sup> Vraag 2.4

Tabel 6.71 – Bevordering interesse in wiskunde<sup>1</sup>, gemiddelden (sd) en range

Uitspraak (n=100)	Gemiddelde (sd)	Min	Max
a. De school biedt leerlingen informatie over wiskunde-gerelateerde vervolgopleidingen en beroepen	2,9 (0,8)	1,0	4,0
b. De school ontwikkelt initiatieven om de interesse van leerlingen voor wiskunde te stimuleren (zoals wiskundeclub of -wedstrijden)	2,4 (0,9)	1,0	4,0
c. De school stimuleert de professionele ontwikkeling van wiskundedocenten	3,3 (0,6)	2,0	4,0
d. De school biedt leerlingen extra wiskundelessen om leerlingen te helpen (nog) beter te presteren	2,8 (0,9)	1,0	4,0
e. De school biedt speciale activiteiten op gebied van wiskunde voor leerlingen die daarin geïnteresseerd zijn	2,2 (0,9)	1,0	4,0
f. De school heeft het specifieke doel om het wiskundeonderwijs op school te verbeteren	2,7 (0,8)	1,0	4,0
g. De school moedigt leerlingen aan om in de toekomst voor wiskunde te kiezen	2,7 (0,7)	1,0	4,0
h. De wiskundedocenten van deze school besteden extra tijd aan leerlingen die geïnteresseerd zijn in wiskunde	2,7 (0,8)	1,0	4,0
i. De school faciliteert leerlingen in de bovenbouw die examens in wiskunde A én B willen doen	2,1 (1,1)	1,0	4,0
Construct <i>Bevordering interesse wiskunde</i> <sup>2</sup>	2,6 (0,5)	1,2	4,0

<sup>1</sup> Vraag 2.5, oorspronkelijke antwoorden: 1=zeer mee eens 4=zeer mee oneens. Alle items omgecodeerd.

<sup>2</sup> Cronbach's alfa: 0,77, Guttman's Lambda 2: 0,77.

Tabel 6.72 – Aanbod professionele ontwikkeling wiskundedocenten<sup>1</sup>, in percentages

Onderwerp (n=100)	%
De inhoud van het wiskundeonderwijs	57,6
Vakdidactiek van het wiskundeonderwijs	60,6
Pedagogiek van het wiskundeonderwijs	53,5
Het reken- en wiskundecurriculum (kerndoelen, referentieniveaus)	56,6
Integratie van ICT in het wiskundeonderwijs	50,5
Het verbeteren van kritisch denken en de oplossingsvaardigheden van leerlingen	47,5
Het meten en beoordelen van leervorderingen in wiskunde	52,5
Inspelen op de behoeften van individuele leerlingen bij wiskunde	64,6

<sup>1</sup> Vraag 2.6

## Leerlingen op school en de coronapandemie

Tabel 6.73 – Veranderingen door pandemie<sup>1</sup>, gemiddelden (sd) en range

Als gevolg van de coronapandemie... (n=61)	Gemiddelde (sd)	Min	Max
a. ... hebben leerlingen leervertraging opgelopen op het gebied van wiskunde	3,2(0,6)	1,0	4,0
b. ... hebben leerlingen leervertraging opgelopen op het gebied van leesvaardigheid	3,3 (0,7)	1,0	4,0
c. ...zijn de wiskundeprestaties er op achteruit gegaan	3,0 (0,8)	1,0	4,0
d. ...is de leesvaardigheid er op achteruit gegaan	3,0 (0,8)	1,0	4,0
e. ...zijn verschillen tussen leerlingen in wiskundeprestaties groter geworden	2,9 (0,9)	1,0	4,0
f. ...zijn verschillen tussen leerlingen in leesvaardigheid groter geworden	3,0 (0,8)	1,0	4,0
g. ...is de motivatie om te leren onder leerlingen afgenomen	3,6 (0,7)	1,0	4,0
h. ...is de concentratie van leerlingen afgenomen	3,6 (0,6)	2,0	4,0
i. ...is het aantal conflicten tussen leerlingen onderling toegenomen	3,0 (1,0)	1,0	4,0
j. ...is het aantal conflicten tussen leerlingen en docenten en/of ander personeel toegenomen	2,8 (1,0)	1,0	4,0
k. ... worden leerlingen sneller boos of verdrietig	3,1 (0,9)	1,0	4,0
l. ... is pesten onder leerlingen toegenomen (fysiek of online)	2,8 (1,0)	1,0	4,0
m. ... is spijbelen toegenomen	2,7 (0,9)	1,0	4,0
n. ... is de bereidheid om zich te houden aan de gedragsregels van de school afgenomen	3,2 (0,8)	1,0	4,0
Construct Verandering in leerprestaties <sup>2</sup>	3,1 (0,6)	1,0	4,0
Construct Verandering in leergedrag <sup>2</sup>	3,6 (0,6)	1,0	4,0
Construct Verandering in leerlinggedrag <sup>2</sup>	3,0 (0,8)	1,0	4,0

<sup>1</sup> Vraag 3.2. 61 van de 99 schoolleiders (62%) waren al voor de pandemie schoolleider op de betreffende school. Alleen zij hebben vraag 3.2 beantwoord, oorspronkelijke antwoorden: 1=zeer mee eens, 4=zeer mee oneens. Alle items omgecodeerd.

<sup>2</sup> Cronbach's alfa alle items: 0,92 Guttman's Lambda 2: 0,92 (let wel: 14 items op 61 respondenten). Inhoudelijk 3 subschalen:  
 1 prestaties (a-f) Cronbach's alfa: 0,89, Guttman's Lambda 2: 0,90 (P corr. met schaal 2=0,51, P corr. met schaal 3=0,49)  
 2 leergedrag (g+h) Cronbach's alfa: 0,84, Guttman's Lambda 2: 0,85 (P corr. met schaal 3=0,47)  
 3 leerlinggedrag (i-n) Cronbach's alfa: 0,91 Guttman's Lambda 2: 0,92

Tabel 6.74 – Gepercipieerde veranderingen door pandemie<sup>1</sup>, percentages zeer mee eens

Stelling (n=61)	% zeer mee eens
a. ... hebben leerlingen leervertraging opgelopen op het gebied van wiskunde	27,9
b. ... hebben leerlingen leervertraging opgelopen op het gebied van leesvaardigheid	37,7
c. ...zijn de wiskundeprestaties er op achteruit gegaan	26,2
d. ...is de leesvaardigheid er op achteruit gegaan	27,9
e. ...zijn verschillen tussen leerlingen in wiskundeprestaties groter geworden	29,5
f. ...zijn verschillen tussen leerlingen in leesvaardigheid groter geworden	27,9
g. ...is de motivatie om te leren onder leerlingen afgenomen	65,6
h. ...is de concentratie van leerlingen afgenomen	68,9
i. ...is het aantal conflicten tussen leerlingen onderling toegenomen	44,3
j. ...is het aantal conflicten tussen leerlingen en docenten en/of ander personeel toegenomen	27,9
k. ... worden leerlingen sneller boos of verdrietig	42,6
l. ... is pesten onder leerlingen toegenomen (fysiek of online)	27,9
m. ... is spijbelen toegenomen	21,3
n. ... is de bereidheid om zich te houden aan de gedragsregels van de school afgenomen	41,0

<sup>1</sup>Vraag 3.2

#### Domeinspecifieke kenmerken van het onderwijsleerproces in relatie tot het onderwijsaanbod

In Tabel 6.75 zijn alle responderende schoolleiders ingedeeld in drie categorieën op basis van het onderwijsaanbod op hun vestiging.

Tabel 6.75 – Indeling schoolleiders naar onderwijsaanbod op school

	N	%
Pro/vmbo	30	29,7%
Breed (vmbo/havo/vwo)	47	46,5%
Havo/vwo	24	23,8%
Totaal	101	100%

Tabel 6.76 toont vervolgens de scores van alle constructvariabelen uit de voorafgaande tabellen, gemiddeld per onderwijsaanbod op de school van de schoolleiders. De laatste drie constructvariabelen zijn slechts aan een deel van de schoolleiders voorgelegd, vandaar de lagere n.

Tabel 6.76 – Indeling schoolleiders naar onderwijsaanbod op school

Constructen	Pro/vmbo	Breed	Havo/vwo	Totaal
Prestatiegericht schoolklimaat (n=101)	3,1 (0,4)	3,3 (0,3)	3,5 (0,3)	3,3 (0,3)
Prestatiegericht schoolklimaat_docenten (n=101)	3,5 (0,3)	3,4 (0,3)	3,6 (0,5)	3,5 (0,4)
Prestatiegericht schoolklimaat_ouders (n=101)	2,8 (0,6)	3,2 (0,5)	3,5 (0,5)	3,2 (0,6)
Prestatiegericht schoolklimaat_leerlingen (n=101)	3,0 (0,4)	3,2 (0,4)	3,4 (0,4)	3,2 (0,5)
Belang leerlingvolgsysteem (n=99)	2,5 (0,3)	2,4 (0,4)	2,4 (0,3)	2,4 (0,3)
Bevordering interesse wiskunde (n=100)	2,3 (0,5)	2,8 (0,4)	2,8 (0,4)	2,6 (0,5)
Verandering in leerprestaties (n=61)	2,8 (0,7)	3,3 (0,5)	2,9 (0,6)	3,1 (0,6)
Verandering in leergedrag (n=61)	3,4 (0,8)	3,7 (0,5)	3,7 (0,5)	3,6 (0,6)
Verandering in leerlinggedrag (n=61)	3,0 (0,8)	3,1 (0,6)	2,7 (1,0)	3,0 (0,8)



## 7 Leerlingprestaties Rekenen-Wiskunde

### 7.1 Inleiding

In het huidige peilingsonderzoek was één van de doelen om de prestaties van leerlingen te relateren aan de vastgestelde referentieniveaus. Een tweede doel van deze peiling is om na te gaan hoe de rekenvaardigheid van leerlingen in het VO in 2022 zich verhoudt tot de meest recente peilingen in het PO. In Paragraaf 7.2 bespreken we de behaalde referentieniveaus van leerlingen in het voortgezet onderwijs (onderzoeksvraag 1). Hierbij vergelijken we de leerlingprestaties in het huidige peilingsonderzoek met de prestaties van leerlingen zoals bepaald bij eerdere peilingsonderzoeken in het PO. Daarbij wijzen we erop dat dit de eerste peiling is in het VO waardoor een jaar-vergelijking binnen deze onderzoeksgroep niet mogelijk is. Vervolgens typeren we in Paragraaf 7.3 de rekenprestaties van leerlingen per domein aan de hand van verschillende typen opgaven en afnamecondities (hoofdrekenen/rekenmachine).

### 7.2 Leerlingprestaties Rekenvaardigheid (onderzoeksvraag 1)

Na het bepalen van de prestatiestandaarden is in kaart gebracht op welk referentieniveau de leerlingen in VO functioneren op het gebied van rekenvaardigheid. In Hoofdstuk 2 is beschreven dat verwacht wordt dat de gemiddelde leerling in jaar 2 zal functioneren tussen 1F en 2F. Meer specifiek werden per deelpopulatie de volgende niveaus verwacht:

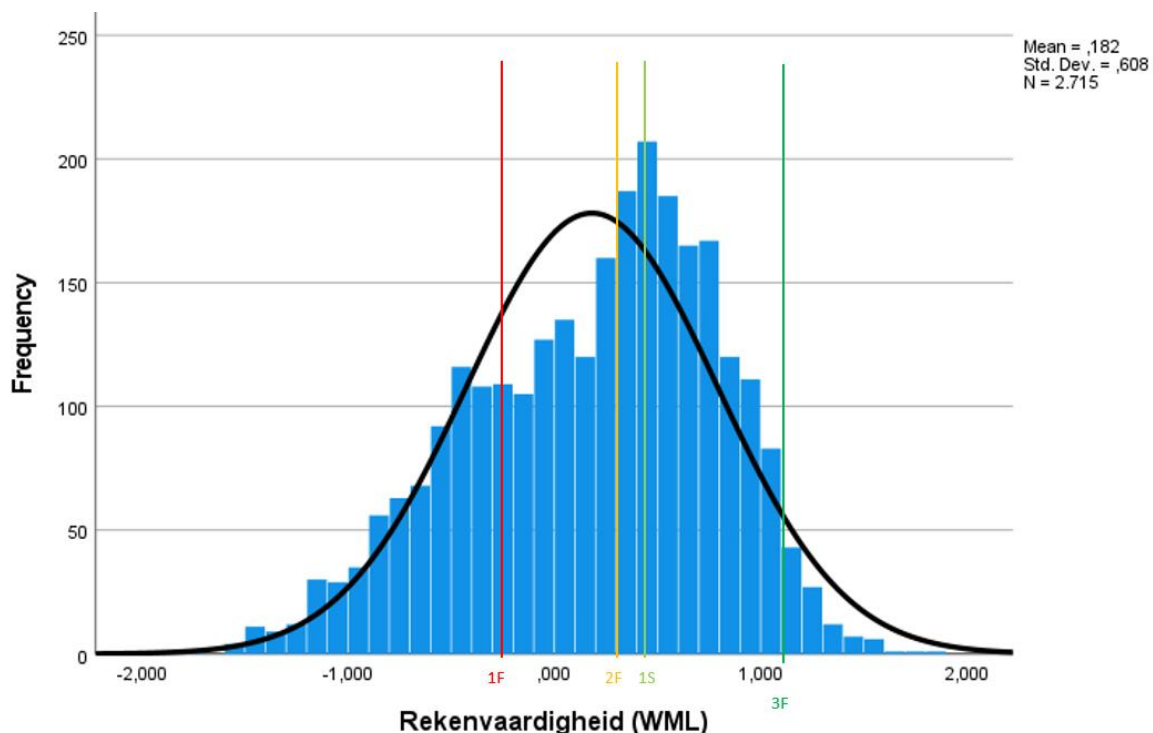
- Vmbo-b/k: tussen 1F en 2F
- Vmbo-g/t :1S behaald en functioneren tussen 2F en 3F
- Havo: 1S behaald en functioneren tussen 2F en 3F
- Vwo: 1S behaald en functioneren tussen 2F en 3F

Hierbij moet opgemerkt worden dat vmbo-leerlingen voorheen in het PO mogelijk geen onderwijsaanbod hebben gehad op niveau 1S omdat dit uitstroomniveau niet haalbaar bleek. In Hoofdstuk 5 is toegelicht hoe vier cesuren (1F, 2F, 1S en 3F) zijn bepaald en waarom dat gedaan is op een enkele rekenschaal. In dat hoofdstuk is ook aangegeven waarom daarvoor gekozen is in plaats van een cesuur op een viertal verschillende schalen (één per referentieniveau), met op iedere schaal een cesuur.

De verdeling van de niveaus in de onderzochte leerlingensteekproef kan op twee manieren bepaald worden. Het eerste is een inschatting op basis van de MML geschatte vaardigheidsverdeling van de steekproef (met gefixeerde itemparameters), onder aanname van een normaalverdeling. Met de gegevens van de PO-peiling van 2019 is dan direct de vergelijking te maken met het (speciaal) basisonderwijs. De andere wijze waarop de verdeling van de referentieniveau bepaald kan worden is door middel van de schattingen per leerling. We gebruiken hiervoor de WML-schattingen. Het voordeel daarvan is dat er geen normaalverdeling verondersteld wordt. In Figuur 7.1 zijn de normaalverdeling (zwarte lijn) en de frequentieverdeling (blauwe staafdiagram) gegeven.

Overigens kan opgemerkt worden dat het gerapporteerde gemiddelde en de standaardafwijking van de MML-schatting weinig verschillen met die van de WML-schattingen zoals die ook gegeven zijn in Figuur 7.1. Het gemiddelde bij de populatieschatting met OPML met gefixeerde parameters is identiek aan dat van de WML-schattingen (0,182) en de standaardafwijking is slechts een fractie kleiner (0,602). Dit kleine verschil in standaardafwijking heeft te maken met de hoge betrouwbaarheid van de metingen per

leerling. Deze is onder meer hoog doordat hier gerapporteerd wordt over een enkele rekenschaal. Hierdoor kunnen we alle opgaven van een leerling gebruiken om de vaardigheid te meten, hetgeen de betrouwbaarheid verhoogt en de meetfout van de vaardigheidsschatting aanzienlijk beperkt.



Figuur 7.1 – Vaardigheidsverdeling Rekenen in de steekproef (zwarte lijn: normaalverdeling; blauwe staafdiagram: frequentieverdeling)

In figuur 7.1 is te zien dat de normaalverdeling en de frequentieverdeling niet geheel overeenkomen. Deze verschillen vooral sterk rond de cesuurpunten voor 1F (rode lijn) en 1S (lichtgroene lijn). Het percentage 1F- en 1S-leerlingen verschilt dan ook het meest, afhankelijk van de werkwijze. Ook is te zien dat de rechterstaart van de frequentieverdeling duidelijk onder die van de normaalverdeling ligt, hetgeen ook duidelijk invloed heeft op de resultaten. De percentages bij de normaalverdeling en de frequentieverdeling zijn gegeven in Tabel 7.1.

Tabel 7.1 – Percentages per referentieniveau en cumulatief in de steekproef bij de normaalverdeling (MML) en de frequentieverdeling (WML)

Percentages (%)	<1F	1F	2F	1S	(≥)3F	≥1S	≥2F	≥1F
Normaalverdeling (MML)	22,9	35,2	8,1	27,6	6,2	33,8	41,9	77,1
Frequentieverdeling (WML)	25,0	26,6	9,1	36,0	3,4	39,4	48,5	75,1

We beschouwen de resultaten van de frequentieverdeling als het meest indicatief voor de werkelijke verdeling. De frequentieverdeling suggereert dat de normaalverdeling niet een optimale beschrijving geeft. Dan kunnen we beter de geobserveerde verdeling gebruiken.

Het gebruik van de WML-scores voor de verdeling van de referentieniveaus heeft ook twee praktische voordelen. Ten eerste zijn deze ook gebruikt bij de peiling Rekenen-Wiskunde in het PO ((speciaal)

basisonderwijs) in 2019. De percentages die daar gevonden waren, zijn te reproduceren. De gegevens van de onderhavige VO-peiling staan samen met die van de PO-peiling in Tabel 7.2, voor de categorieën die in beide peilingen gebruikt zijn.

Tabel 7.2 – Percentage leerlingen in VO (2022) en PO (2019, ook uitgesplitst naar BaO en SBO) in overeenkomstige referentieniveau-categorieën

WML	<1F	1F	≥1S
VO (2022)	25%	36%	39%
PO (2019)	20%	48%	32%
BaO (2019)	18%	50%	33%
SBO (2019)	85%	13%	2%

Voor VO is bij het percentage 1F- nu ook het percentage 2F-leerlingen opgeteld. De categorie 2F wordt niet gebruikt in PO, en ligt op de vaardigheidsschaal onder het niveau 1S. Bij het genoemde percentage 1S zijn ook de leerlingen gerekend die op het niveau 3F presteren. Ook de categorie 3F wordt niet in PO gebruikt. De percentages in de tabel achter 'PO' zijn een (gewogen) combinatie van de gegevens uit regulier en speciaal basisonderwijs. De VO-percentages zijn het beste te vergelijken met deze PO-percentages, aangezien leerlingen uit zowel regulier als speciaal basisonderwijs in het VO terecht komen.

De weging is op basis van gegevens uit het Technisch rapport Peil. onderwijs Rekenen-Wiskunde einde (speciaal) basisonderwijs 2019 (Buisman et al., 2021), waarbij gemeld is (p.13) dat voor het BaO de populatie uit 179.295 leerlingen bestond en voor het SBO uit 7172 leerlingen. Dit zijn allen leerlingen in hun laatste jaar van het PO, dus kort voor de overgang naar het VO. In deze vergelijking is te zien dat er leerlingen lijken te zijn die in het PO eerder 1F haalden en dat in het VO niet meer laten zien, maar ook dat er meer leerlingen op een hoger (tenminste 1S) niveau presteren.

In plaats van de verdeling van de referentieniveaus zijn overigens ook de MML-verdelingen te vergelijken. We hadden al laten zien dat de MML-populatieschatting en de WML-schattingen amper van elkaar verschillen wat betreft gemiddelde en standaardafwijking. Als we dan de MML-gemiddelden van PO (0,172) met die van VO (0,182) vergelijken is het verschil niet significant ( $\alpha=0,05$ ). De effectgrootte van 0,02 is dan ook verwaarloosbaar. De gemiddelde rekenvaardigheid lijkt dus niet vooruit of achteruit te zijn gegaan in twee jaar voortgezet onderwijs.

Een tweede voordeel is dat de WML-waarden ook gebruikt zijn voor de verdere analyses, zoals de multilevel analyses (zie Hoofdstuk 8). De waarden zijn ook gebruikt om in de terugkoppeling aan de scholen te laten zien welke referentieniveaus de leerlingen behaald hebben. Zoals gemeld zijn deze door de hoge betrouwbaarheid van de meting goed te gebruiken, en is de meerwaarde van het werken met plausibele waarden beperkt. De plausibele waarden, die ook berekend zijn, verschillen bovendien niet veel van de WML-schattingen (zie voetnoot 36 in Paragraaf 5.6). Op basis van in dit project gemaakte afspraken is verder gewerkt met de WML-schattingen. De resultaten worden gepresenteerd in Tabel 7.3 en Tabel 7.4.

In Tabel 7.3 zijn de resultaten afgezet tegen de variabele 'Verwacht onderwijsniveau in leerjaar 3 (zie ook Hoofdstuk 4, Tabel 4.9). Met deze variabele is de leerlingensteekproef in vier categorieën ingedeeld: praktijkonderwijs (pro), vmbo-b/k, vmbo-g/t, en havo/vwo. Doordat ook de landelijke verdeling over deze categorieën bekend is, is het mogelijk te wegen. In Paragraaf 4.5 bleek overigens dat dit niet nodig is,

omdat de verdeling van de steekproef niet significant afwijkt van de verdeling in de populatie en weging weinig impact zou hebben. Dat zien we ook terug in Tabel 7.3 waarbij de percentages zoals geobserveerd en zoals verkregen na weging weinig van elkaar verschillen.

Tabel 7.3 – Verdeling referentieniveaus o.b.v. WML-schattingen naar verwacht onderwijsniveau in leerjaar 3 (leerlingensteekproef)

	Aantal leerlingen (n)						Percentage leerlingen (%)					Cumulatief (%)		
	Alle	<1F	1F	2F	1S	≥3F	<1F	1F	2F	1S	≥3F	≥1S	≥2F	≥1F
Pro	79	78	1	0	0	0	99%	1%	0%	0%	0%	0%	0%	1%
Vmbo-b/k	491	356	125	5	5	0	73%	25%	1%	1%	0%	1%	2%	27%
Vmbo-g/t	619	151	342	52	74	0	24%	55%	8%	12%	0%	12%	20%	76%
Havo/vwo	1342	20	209	177	849	87	1%	16%	13%	63%	6%	70%	83%	99%
Onbekend*	184	73	45	12	49	5	40%	24%	7%	27%	3%	29%	36%	60%
<b>Totaal</b>	<b>2715</b>	<b>678</b>	<b>722</b>	<b>246</b>	<b>977</b>	<b>92</b>	<b>25%</b>	<b>27%</b>	<b>9%</b>	<b>36%</b>	<b>3%</b>	<b>39%</b>	<b>48%</b>	<b>75%</b>
Na weging	2715	712	761	240	918	84	26%	28%	9%	34%	3%	37%	46%	74%

\* Zie toelichting boven Tabel 4.9.

Twee van de categorieën in Tabel 7.3 zijn nog verder op te delen naar niveaus, namelijk vmbo-b/k en havo/vwo. De resultaten na verdere opsplitsing staan in Tabel 7.4.

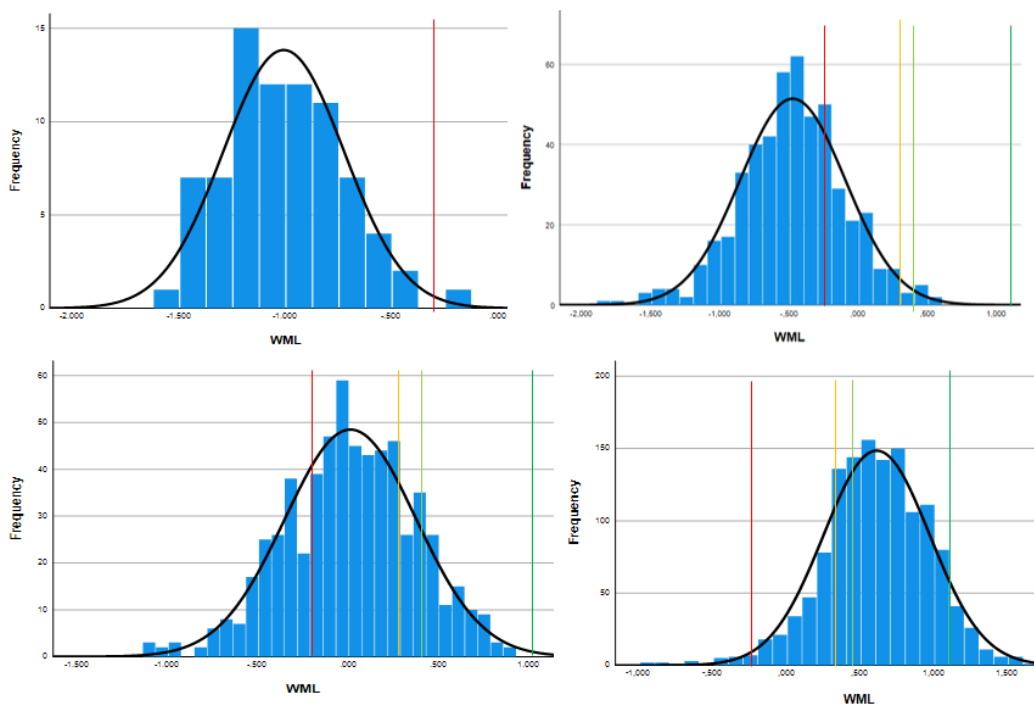
Tabel 7.4 – Verdeling referentieniveaus o.b.v. WML-schattingen, naar verwacht onderwijsniveau in leerjaar 3 (leerlingensteekproef)

	Aantal leerlingen (n)						Percentage leerlingen (%)					Cumulatief (%)		
	Alle	<1F	1F	2F	1S	≥3F	<1F	1F	2F	1S	≥3F	≥1S	≥2F	≥1F
vmbo-b	114	97	15	2	0	0	85%	13%	2%	0%	0%	0%	2%	15%
vmbo-k	377	259	110	3	5	0	69%	29%	1%	1%	0%	1%	2%	31%
havo	634	14	175	120	320	5	2%	28%	19%	50%	1%	51%	70%	98%
vwo	708	6	34	57	529	82	1%	5%	8%	75%	12%	86%	94%	99%

Naast de WML-gegevens is het ook mogelijk de verdelingen te schatten om tot een gewogen MML-verdeling te komen. Hiervoor gebruiken we de vier categorieën uit Tabel 7.3 en bepalen we per groep de verdeling. Dit is zowel uitgevoerd voor de verdeling van de WML-schattingen als voor de latente MML-verdeling.

In Figuur 7.2 is bij de WML-schattingen goed te zien dat wanneer dit per subgroep onderzocht wordt, de verdeling per groep meer een normale verdeling volgt dan als we de verdeling bekijken van het totaal, zoals weergegeven in Figuur 7.1. De gewogen verdeling is bepaald als *mixture distribution*. Het verschil met de eerder gerapporteerde verdelingseigenschappen is zeer beperkt<sup>43</sup>.

<sup>43</sup> Bij het vergelijken van de geschatte gewogen verdeling met de geobserveerde verdeling is de effectgrootte van dat verschil 0,04, hetgeen als niet betekenisvol kan worden geïdentificeerd.



Figuur 7.2 – Vaardigheidsverdeling Rekenen in de steekproef voor pro (linksboven), vmbo-b/k (rechtsboven), vmbo-g/t (linksonder) en havo/vwo (rechtsboven). Zwarte lijn: normaalverdeling; blauwe staafdiagram: frequentieverdeling; verticale lijnen: 1F (rood), 2F (oranje), 1S (lichtgroen), 3F (donkergroen).

### 7.3 Leerlingprestaties per domein en per afnameconditie (onderzoeksvraag 2)

#### 7.3.1 Ter inleiding

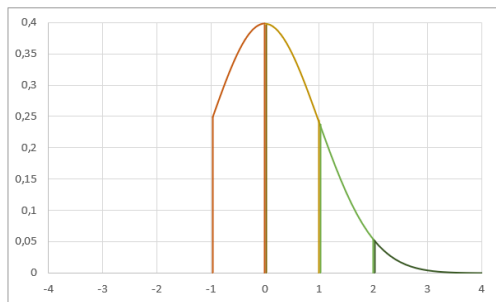
Een belangrijke vraag in dit peilingsonderzoek Rekenen-Wiskunde in het VO betreft de leerlingprestaties per domein, in relatie tot de behaalde referentieniveaus. Deze vraag vereist ten eerste dat de scores op de afgenomen opgaven goed geïnterpreteerd kunnen worden. Daarna worden deze gerelateerd aan de vaardigheid van relevante groepen, of relevante niveaus op de schaal.

Doordat de items en leerlingprestaties in de peiling zijn geanalyseerd binnen het raamwerk van de IRT, is het goed mogelijk om de items en het prestatieniveau van de leerlingen aan elkaar te relateren. Met behulp van het IRT-model is voor ieder item op ieder vaardigheidsniveau te schatten wat de kans op een goed antwoord is. Andersom kan ook worden nagegaan welk niveau van rekenvaardigheid benodigd is om een bepaalde kans te hebben op het correct maken van een item. Van deze twee mogelijkheden wordt gebruikgemaakt voor het representeren van de beheersing van de opgaven. Beheersing wordt voor dichotome items<sup>44</sup> als volgt gedefinieerd:

Geen beheersing: de kans op een correct antwoord is kleiner dan 50 procent  
 Matige beheersing: de kans op een correct antwoord ligt tussen de 50 en 80 procent  
 Goede beheersing: de kans op een correct antwoord is groter dan 80 procent

<sup>44</sup> De definities voor polytome opgaven verschillen iets, maar zijn hier sterk aan verwant. Echter, zoals al aangegeven bevatten deze toetsen alleen dichotome opgaven waardoor we ons in deze beschrijving daartoe beperken.

Voor elk item is dus uitgerekend hoeveel vaardigheid nodig is om 50 of 80 procent kans te hebben op het correct maken van dat item. Deze informatie wordt gebruikt om de opgaven af te beelden op de vaardigheidsschaal. Op deze vaardigheidsschaal kunnen ook populatie-eigenschappen afgebeeld worden. In dit geval is ervoor gekozen de referentieniveaus 1F, 2F, 1S en 3F weer te geven. Bij het interpreteren van de afbeeldingen moet er rekening mee gehouden worden dat de meerderheid van de leerlingen rond niveau 2F of hoger functioneert (zie Paragraaf 7.2). Hierdoor is het lastig weer te geven wat de 'gemiddelde 2F-leerling' kan, en afhankelijk van waar de grens van de volgende referentiegroep precies ligt. De verdelingen in vaardigheid van leerlingen op basis van een opdeling naar behaald referentieniveau zijn altijd afgeknotte verdelingen aangezien die juist gedefinieerd worden door een specifiek grenspunt op de schaal (zie Figuur 7.3 ter illustratie).



*Figuur 7.3 – Fictief voorbeeld van de vier mogelijke populaties op basis van referentieniveaugrenzen*

In iedere figuur wordt een item gerepresenteerd door een balkje in drie kleuren. Het grijze (linker)deel van de balk is het deel van de vaardigheidsschaal waar het item nog helemaal niet beheerst wordt. Binnen dat bereik is de kans om het item goed te beantwoorden 0,50 of lager. Als dat grijze deel kort is, dan betreft het een makkelijk item: er is maar een klein deel op de vaardigheidsschaal waar het item als moeilijk gezien wordt. Het rode deel van de balk geeft de vaardigheidsrange waarin de kans op het goede antwoord ligt tussen 0,50 en 0,80. Op dat deel van de schaal is de opgave uitdagend. Als het rode deel kort is, dan is de opgave goed onderscheidend: het item maakt zeer goed onderscheid tussen leerlingen die het item niet beheersen en leerlingen die het wel beheersen. Het groen gemarkeerde gedeelte (i.e., het rechtergedeelte) in de afbeelding reflecteert het vaardigheidsbereik waarin het item (goed) beheerst wordt (kans van 0,80 of meer het item goed beantwoorden). Als het groene deel van de balk kort is, dan is het item moeilijk. Er is dan een hoge vaardigheid nodig om het item goed te beheersen.

Alle opgaven zijn af te beelden op dezelfde vaardigheidsschaal. De figuren zijn afgebeeld voor iedere rekeninhoud en afnameconditie (HR en RM) afzonderlijk. De eerste figuur die op deze manier is vormgegeven (Figuur 7.4 betreft alle hoofdrekenopgaven van schaal B: Getallen en getalsbewerkingen). De opgaven zijn geordend naar het referentieniveau dat elk item beoogt te meten. Van boven naar beneden zijn dat 3F-, 1S-, 2F- en 1F-opgaven. Tot slot zijn de opgaven geordend naar de positie van de locatieparameter (de b-parameter, of de b50), oftewel de linkerzijde van het rode deel van de balk. Er is wel een relatie tussen het te meten referentieniveau en de b-parameter, maar die is niet perfect. Er zijn opgaven van een 'lager' referentieniveau waarbij de b-parameter hoger is dan van opgaven van een 'hoger' referentieniveau. De ordening van de opgaven naar de grenspunten tussen geen versus een matige beheersing, noch die van de grens tussen een matige en een goede beheersing ligt vast door het gemeten referentieniveau. Kort door de bocht geformuleerd: de moeilijkheid wordt niet geheel bepaald door het referentieniveau dat gemeten wordt.

Zoals hierboven benoemd is, wordt de moeilijkheid van de items niet één op één bepaald door het referentieniveau dat wordt gemeten. In hieronder volgende paragrafen beschrijven we per afnameconditie en per domein de moeilijkheid van de items in relatie tot de referentieniveaus en de niveaus in het VO. Zoals verwacht functioneren de meeste leerlingen in leerjaar 2 boven 1F-niveau (75%). Bijna de helft van de leerlingen behaalt een rekenvaardigheidsniveau tussen 2F en 1S. Zoals weergegeven in Tabel 7.3 zijn het vooral leerlingen uit het praktijkonderwijs, vmbo-b en vmbo-k die niveau 1F nog niet hebben behaald. Van de vmbo-g/t leerlingen behaalt 55% het 1F-niveau. De meeste havo/vwo leerlingen hebben niveau 1S behaald. Zoals eerder opgemerkt krijgen VO-leerlingen geen aanbod op 1S-niveau. Voor een deel van de vmbo-leerlingen zal gelden dat zij vanwege het verwachte uitstroomniveau in het PO geen of een aangepast aanbod hebben gehad op 1S-niveau. Desalniettemin haalt toch een enkele vmbo-leerling uit leerjaar 2 dit niveau (1% vmbo-b/k en 12% vmbo-g/t). Niveau 3F wordt door 6% van de leerlingen uit leerjaar 2 gehaald, zoals te zien is in Tabel 7.4 betreft het voornamelijk vwo-leerlingen (n=87) en enkele havo-leerlingen (n=5). Deze leerlingen behoren tot percentiel 90 (P90). Leerlingen in P10 (90% van de leerlingen presteert beter) hebben het 1F-niveau nog niet gehaald. Leerlingen die 1F-niveau halen komen vooral uit het praktijkonderwijs en uit vmbo-b/k. Bijna een kwart van de vmbo-g/t leerlingen haalt het 1F-niveau niet. Van havo/vwo leerlingen mag verwacht worden dat zij bij het uitstromen uit het PO niveau 1S beheersen en daarmee dus ook niveau 1F hebben behaald. Toch haalt 1% van de havo/vwo leerlingen 1F niet.

### 7.3.2 Domein B – Getallen en bewerkingen

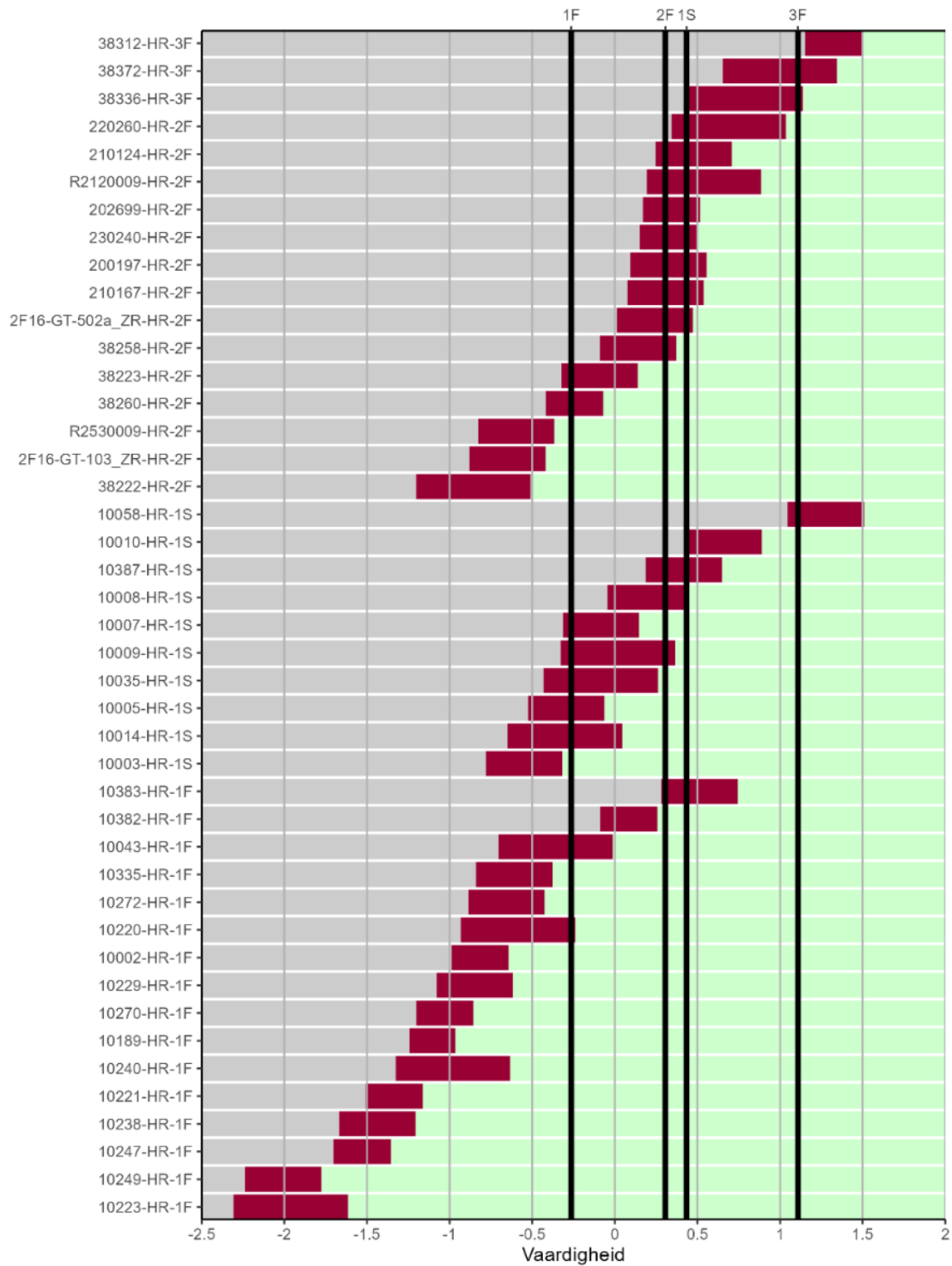
#### Hoofdrekenen

Voor het domein Getallen en bewerkingen laat Figuur 7.4 zien dat er drie moeilijke 1F-items zijn die nog niet worden beheerst door de VO leerlingen die het 1F-niveau beheersen. Het moeilijkste 1F-item betreft opgave 10383 (zie ook Figuur 7.5). Dit item wordt matig beheerst door leerlingen met 2F- of 1S-niveau. Aangezien bijna 50% van de leerlingen 2F of hoger haalt, kan worden gezegd dat dit item voor de leerling in percentiel 50 (P50) nog moeilijk is. Dat wil zeggen dat de meeste vmbo-leerlingen uit jaar 2 deze opgave nog niet beheersen, terwijl havo/vwo-leerlingen deze opgave matig tot goed beheersen. De leerling in P90 beheerst deze opgave goed.

Item 10382 betreft een nagenoeg identiek item met net iets andere getallen. Item 10043 is een 1F/Context item waarbij hele getallen boven de 100 opgeteld moeten worden. Dat item wordt door de meeste leerlingen uit leerjaar 2 matig tot goed beheerst. Alleen leerlingen die 1F nog niet beheersen hebben een kans van minder dan 0,65 om dit item goed te beantwoorden. Leerlingen die 1F beheersen, beheersen één 1S-opgave (10003) waarin gevraagd wordt om een kommagetal op de getallenlijn te positioneren. Kortom, deze opgave wordt door de meeste leerlingen uit leerjaar 2 beheerst. Tevens beheersen deze leerlingen drie 2F-items waarin gevraagd wordt meerdere bedragen op te tellen, miljoenen uit te schrijven in cijfers en het volgende cijfer in de reeks te bepalen. Leerlingen in P10 beheersen deze drie opgaven nog net niet.

Leerlingen die 2F of 1S beheersen presteren beter dan leerlingen in P50, maar slechter dan leerlingen in P80. Zoals Tabel 7.3 laat zien worden niveau 2F en 1S vooral behaald door de meerderheid van de havo/vwo leerlingen. Zoals hierboven genoemd is, wordt het moeilijkste 1F-item ook nog niet beheerst door de VO-leerlingen die niveau 2F beheersen. Hetzelfde geldt voor leerlingen die 1S beheersen, maar die hebben een iets grotere kans om item 10383 goed te maken dan leerlingen die 2F beheersen.

### Schaal B - HR




Figuur 7.4 – Hoofdreken-opgaven domein B – Getallen en bewerkingen<sup>45</sup>

<sup>45</sup> Gewogen vaardigheidsscores per deelpopulatie: pro:-1,02; vmbo-b/k:-0,48; vmbo-g/t: 0,01; havo/vwo: 0,61; totaal: 0,16



De leerlingen die 2F of 1S behalen beheersen zes van de tien 1S-opgaven en zes van de dertien 2F-opgaven, waaronder de drie gemakkelijkste opgaven die ook door de VO-leerlingen met 1F-niveau worden beheerst. Daarnaast beheersen zij opgaven waar één of meer geldbedragen vermenigvuldigd moeten worden met een geheel getal. Ook kunnen zij aftrekken met getallen boven de 10.000. De opgave waarin een getal boven de 100 moet worden gedeeld beheersen leerlingen met 2F-niveau matig ( $p \approx 0,7$ ). Leerlingen met 1S-niveau beheersen deze opgave goed ( $p \approx 0,8$ ). Zowel 2F- als 1S-leerlingen hebben nog moeite met het bepalen van het verschil tussen een positief en een negatief getal en met het vermenigvuldigen van twee negatieve getallen. Ook complexere contextopgaven met meerdere bewerkingen met geld (kommagetallen), uren en/of minuten worden door deze leerlingen matig beheerst. Ook het moeilijkste 2F-item (220260) in dit domein betreft een complexe bewerking met kommagetallen (zie ook Figuur 7.5).

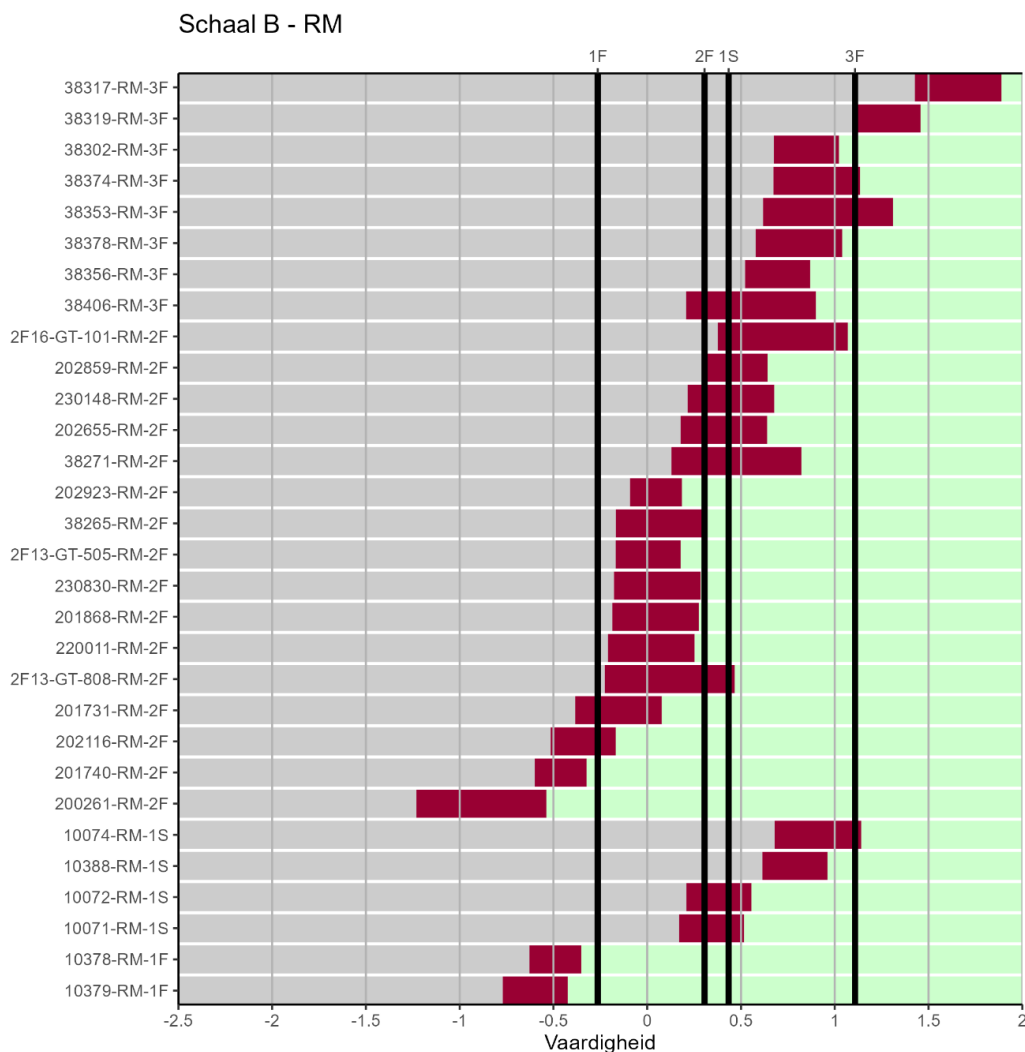
Item en referentieniveau	Item	Beheersing
10383- HR 1F	$\frac{3}{4} \times 4 + 17 =$	Matige beheersing door leerlingen met 2F en 1S. Bijna de helft van de VO-leerlingen (vooral havo/vwo leerlingen) in jaar twee beheerst dit item matig tot goed.
220260- HR- 2F	 <p>Prijs per drie stuks € 6,99</p> <p>Je koopt twaalf glazen.</p> <p><b>Hoeveel moet je betalen?</b></p> <p>€ _____</p>	Dit item wordt nog niet beheerst door de meeste leerlingen uit pro en vmbo. De meeste leerlingen met havo/vwo niveau beheersen deze opgave matig.
10058-HR-1S	<p>De handbalvereniging verzamelt iedere maand oud papier.</p> <p>Vorig jaar verzamelde men 7849 kg papier.</p> <p>Hoeveel kg is dat gemiddeld per maand?</p> <p>Rond je uitkomst af op een heel getal.</p>	Matige beheersing door leerlingen met 3F. Slechts 3% van de leerlingen in VO leerjaar 2 beheerst deze opgave.

Figuur 7.5 – Voorbeelditems Hoofdrekenen domein B – Getallen en Bewerkingen

Slechts een zeer klein percentage van de VO-leerlingen behaalt in hun tweede jaar al het 3F-niveau, respectievelijk 1% van de havo-leerlingen en 12% van de vwo-leerlingen. Wat blijkt in Figuur 7.4 is dat het moeilijkste 1S-item (10058; zie ook Figuur 7.5) slechts matig door deze doelgroep wordt beheerst. De 3F-items in het domein Getallen en bewerkingen betreffen alle drie complexe bewerkingen. De moeilijkste 3F-opgave, die door geen enkele leerling goed wordt beheerst ( $p < 0,5$ ), betreft een complexe opgave met meerdere berekeningen (optellen en delen) met geldbedragen. Opvallend genoeg verschilt deze moeilijkste opgave qua inhoud (bewerkingen en getallen) niet heel duidelijk van de twee andere 3F-items die matig tot goed worden beheerst door leerlingen die het 3F-niveau halen.



## Rekenmachine

Leerlingen die 1F beheersen hebben een goede beheersing van de twee 1F-opgaven en van de twee gemakkelijkste 2F-rekenmachineopgaven in het domein Getallen en bewerkingen. Zoals in de afnameconditie hoofdrekenen hebben deze leerlingen ook in de conditie met rekenmachine moeite met het aftrekken van negatieve getallen. Deze leerlingen hebben een matige beheersing van item 202116 en een slechte beheersing van 201731 (zie Figuur 7.6). Aangezien ongeveer de helft van de onderzochte VO-leerlingen 2F-niveau behaalt, kan gezegd worden dat ongeveer de helft van de VO-leerlingen een matig tot goede beheersing hebben in het aftrekken van negatieve getallen. Gezien de verdeling van de referentieniveaus in de VO-deelpopulaties (zie Tabel 7.3 en 7.4) betreft het vooral vmbo-g/t en havo/vwo leerlingen die deze opgaven matig tot goed beheersen. Daarnaast hebben leerlingen met niveau 2F of 1S een goede beheersing van complexe 1S-opgaven (meerdere bewerkingen) met de rekenmachine, zoals 10071, maar een matig tot slechte beheersing van items zoals 10074 (zie ook Figuur 7.7). Deze moeilijkste 1S-opgave wordt goed beheerst door leerlingen met 3F-niveau. Deze opgave wordt dus alleen beheerst door de hoogvaardige vwo-leerlingen in P90.



Figuur 7.6. Rekenmachine-opgaven domein B – Getallen en bewerkingen<sup>46</sup>

<sup>46</sup> Gewogen vaardigheidsscores per deelpopulatie: pro:-1,02; vmbo-b/k:-0,48; vmbo-g/t: 0,01; havo/vwo: 0,61; totaal: 0,16.

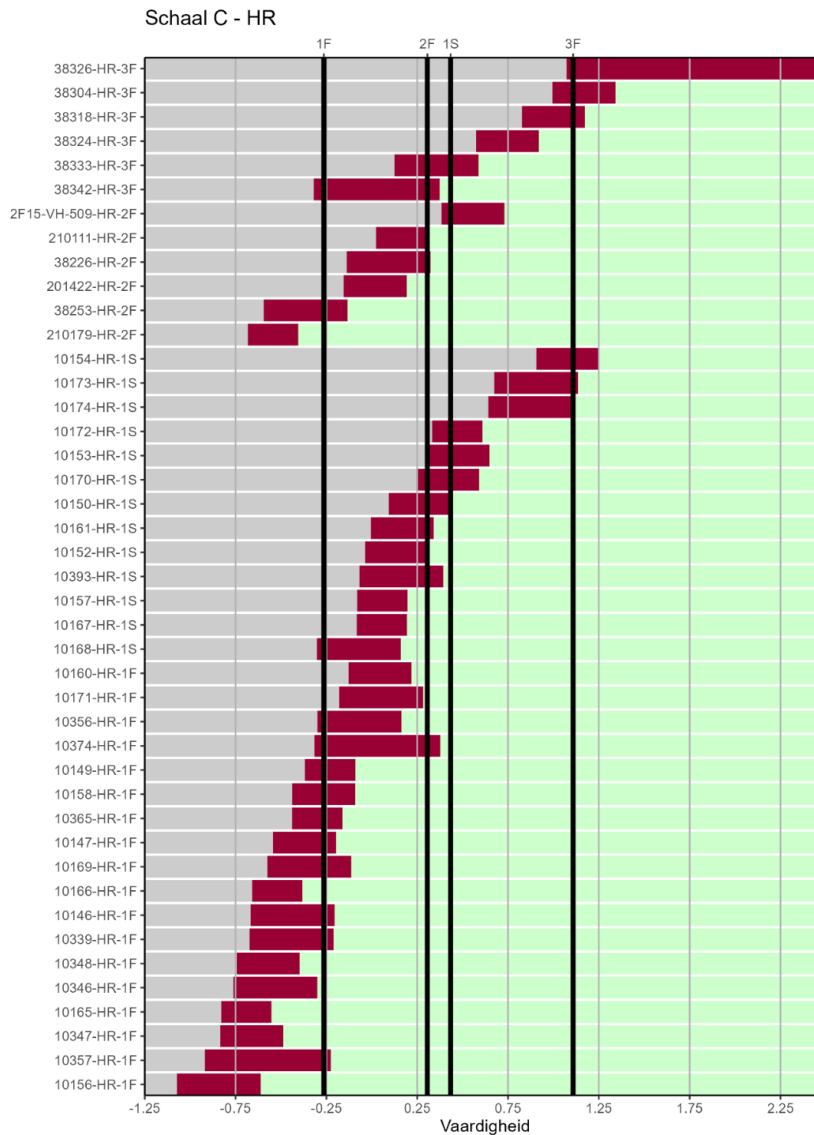
Item en referentieniveau	Item	Beheersing									
202116 – RM-2F	<p>Temperaturen op 10 maart</p> <table border="1"> <thead> <tr> <th></th> <th>°C</th> <th>jaar</th> </tr> </thead> <tbody> <tr> <td>hoogst gemeten temperatuur</td> <td>19,0</td> <td>2014</td> </tr> <tr> <td>laagst gemeten temperatuur</td> <td>-5,3</td> <td>1931</td> </tr> </tbody> </table> <p>Hoeveel °C is het verschil tussen de gemeten temperatuur op 10 maart 1931 en die op 10 maart 2014?</p> <p>_____ °C</p>		°C	jaar	hoogst gemeten temperatuur	19,0	2014	laagst gemeten temperatuur	-5,3	1931	Goede beheersing door de meeste tweedejaars uit havo/vwo. Matig beheerst door leerlingen met 1F-niveau (vmbo-g/t).
	°C	jaar									
hoogst gemeten temperatuur	19,0	2014									
laagst gemeten temperatuur	-5,3	1931									
201731-RM-2F	 <p>Je wilt deze tablet kopen. Je spaart hiervoor € 12,00 per week. Op je bankrekening staat € 259,60.</p> <p>Hoeveel weken moet je nog sparen om de tablet te kunnen kopen?</p>	Goede beheersing door de meeste tweedejaars uit havo/vwo. Matig beheerst door leerlingen met 1F-niveau (vmbo-g/t).									
10071-RM-1S	<p>In het ruim van een schip zit nog 4000 kg graan. Het graan wordt opgezogen en in zakken van 25 kg gedaan. De zakken worden op een heftruck gelegd. Op de heftruck kunnen 12 zakken.</p> <p>Hoe vaak moet de heftruck heen en weer rijden?</p>	Matig tot goede beheersing door de meeste tweedejaars uit vmbo-gt/tl en havo/vwo. Wordt niet beheerst door leerlingen met 1F-niveau.									
10074-RM-1S	 <p>Lea koopt een zak potgrond en 5 potten.</p> <p>Hoeveel euro moet zij in totaal betalen?</p>	Moeilijke opgave. Wordt goed beheerst door leerlingen die 3F hebben behaald. Deze leerlingen vallen in percentiel 90.									

Figuur 7.7 – Voorbeelditems Rekenmachine domein B – Getallen en Bewerkingen

### 7.3.3 Domein C – Verhoudingen

#### Hoofdrekenen

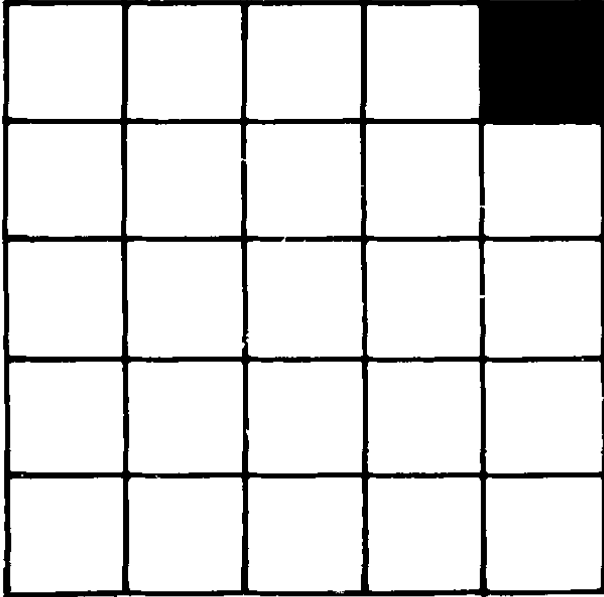
Leerlingen die 1F beheersen hebben nog moeite met twee van de achttien 1F-items in het domein Verhoudingen (Figuur 7.8). Leerlingen met niveau 2F of hoger beheersen deze twee opgaven goed. Leerlingen met 1F-niveau beheersen ook de gemakkelijkste 2F-items zoals 210179 waarin verhouding-taala omgezet moet worden naar procenten. Ook hebben leerlingen die 1F beheersen een matige beheersing van de gemakkelijkste 1S-opgave (waarin wordt gevraagd naar het kortingspercentage en leerlingen moeten rekenen met veelvouden van 100 en 1000) en van 3F-item 38342 (waarin het deel en het percentage zijn gegeven en het totaal moet worden berekend).



Figuur 7.8 – Hoofdreken-opgaven domein C – Verhoudingen<sup>47</sup>

<sup>47</sup> Gewogen vaardigheidsscores per deelpopulatie: pro:-1,02; vmbo-b/k:-0,48; vmbo-g/t: 0,01; havo/vwo: 0,61; totaal: 0,16

Leerlingen met niveau 2F beheersen deze 3F-opgave bijna. Leerlingen die 2F of 1F beheersen, beheersen bijna alle 2F-items (zoals 210111; zie Figuur 7.10) uit het domein Verhoudingen. De moeilijkste 1S-opgaven (zoals 10154, zie Figuur 7.10) beheersen zij nog niet. Deze wordt wel beheerst door leerlingen die het 3F-niveau halen. Leerlingen met 2F-niveau hebben een matige beheersing van de 3F-opgave 38333 waarin gevraagd wordt om twee kortingsacties met elkaar te vergelijken. De moeilijkste 3F-opgave in het domein Verhoudingen betreft een complexe opgave waarin meerdere berekeningen met ongelijknamige breuken uitgevoerd moest worden.

Item en referentieniveau	Item	Beheersing
10171-HR-1F	 <p>Hoeveel procent van dit vierkant is zwart?</p> <p>____%</p>	Deze opgave wordt goed beheerst door leerlingen met 2F-niveau of hoger. Dit betreft vooral havo/vwo en enkele leerlingen uit vmbo-g/t. Leerlingen met 1F-niveau of lager (vooral pro en vmbo-b/k) beheersen deze opgave niet.
10160-HR-1F	<p>Een krant heeft 150 000 abonnees.</p> <p>Hiervan betaalt <math>\frac{2}{3}</math> deel het abonnement automatisch via de bank.</p> <p>Hoeveel mensen betalen niet automatisch?</p>	Deze opgave wordt goed beheerst door leerlingen met 2F-niveau of hoger. Dit betreft vooral leerlingen met vmbo-g/t en havo/vwo. Leerlingen met 1F-niveau of lager (vooral pro en vmbo-b/k) beheersen deze opgave niet.

Figuur 7.9 – Voorbeelditems Hoofdrekenen domein C - Verhoudingen

Item en referentieniveau	Item	Beheersing												
10154-HR-1S	<p>In Hoeland is een <math>\frac{1}{2}</math> % van de inwoners ouder dan 75 jaar.</p> <p>Wat betekent dit?</p> <p><b>A</b> 1 op de 2 inwoners is ouder dan 75 jaar.  <b>B</b> 1 op de 20 inwoners is ouder dan 75 jaar.  <b>C</b> 1 op de 50 inwoners is ouder dan 75 jaar.  <b>D</b> 1 op de 100 inwoners is ouder dan 75 jaar.  <b>E</b> 1 op de 200 inwoners is ouder dan 75 jaar.</p>	<p>Moeilijke opgave voor leerlingen uit leerjaar 2. Deze opgave wordt matig beheerst door leerlingen die 3F hebben behaald. Deze leerlingen doen havo of vwo en vallen in percentiel 90.</p>												
210111-HR-2F	<p>De gemiddelde dagelijkse energiebehoefte van een volwassen man is 2400 kilocalorieën (kcal).</p> <p>Voedingswaarden van voedselproducten</p> <table border="1"> <thead> <tr> <th>Product</th> <th>Aantal kcal</th> </tr> </thead> <tbody> <tr> <td>patat met mayonaise</td> <td>650</td> </tr> <tr> <td>kroket</td> <td>150</td> </tr> <tr> <td>frikadel</td> <td>165</td> </tr> <tr> <td>kipcorn</td> <td>200</td> </tr> <tr> <td>kaassoufflé</td> <td>145</td> </tr> </tbody> </table> <p>Een volwassen man eet 1 patat met mayonaise en een kroket.</p> <p>Hoeveel procent van de gemiddelde dagelijkse energiebehoefte is dat?  Rond je antwoord af op twee decimalen.</p>	Product	Aantal kcal	patat met mayonaise	650	kroket	150	frikadel	165	kipcorn	200	kaassoufflé	145	<p>Leerlingen met 2F of hoger beheersen deze opgave goed. Leerlingen met 1F-niveau of lager beheersen deze opgave niet. De meeste havo/vwo leerlingen beheersen deze opgaven goed.</p>
Product	Aantal kcal													
patat met mayonaise	650													
kroket	150													
frikadel	165													
kipcorn	200													
kaassoufflé	145													

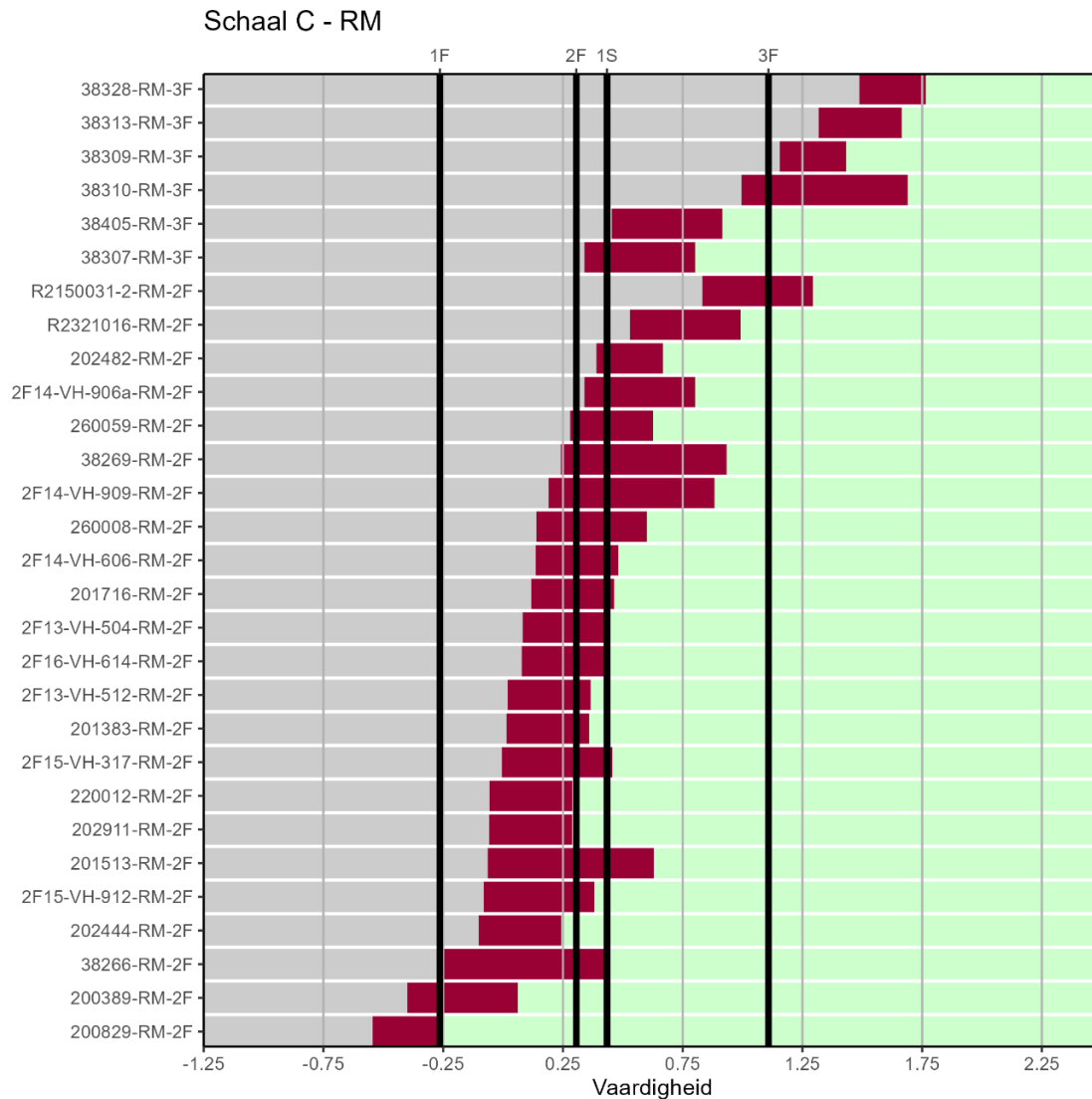
Figuur 7.10 – Meer voorbeelditems Hoofdrekenen domein C – Verhoudingen

### Rekenmachine

Het eerste wat opvalt aan Figuur 7.11 is dat de meeste 2F-opgaven een hoog onderscheidend niveau hebben voor leerlingen die tussen 1F- en 2F-niveau in functioneren. Leerlingen die 1F beheersen, hebben een goede beheersing van 2F-item 200829. In dit item wordt gevraagd om op basis van een recept voor 4 personen uit te rekenen hoeveel er van een bepaald ingrediënt nodig is voor 10 personen. Deze leerlingen hebben een matige beheersing van 2F-opgave 200389 waarin getallen boven de 1000 omgerekend moeten worden naar verhoudingen taal. Leerlingen die 2F of 1S hebben behaald hebben een matige beheersing van een item waarin de werkelijke lengte van een schaalmodel berekend moet worden. Leerlingen met 3F-niveau beheersen deze opgave goed.

Leerlingen met 2F-niveau beheersen niet de opgave waarin bij een deel van een geheel een percentage moet worden berekend met gehele getallen. Leerlingen met 1S-niveau beheersen die opgave wel. Leerlingen met 1S-niveau beheersen de twee moeilijkste 2F-items uit het domein Verhoudingen met rekenmachine nog niet. Het gaat om het rekenen met verhoudingen tussen brandstof en afstand en om het rekenen met procenten aan de hand van gegevens uit een staafgrafiek. Deze moeilijke 2F-opgaven worden respectievelijk goed tot matig beheerst door leerlingen met 3F-niveau. De gemakkelijkste 3F-opgave, die matig wordt beheerst door leerlingen met 1S-niveau, betreft wederom een opgave waarin een recept moet worden omgerekend. Deze opgave is complexer dan 2F-item 200829 doordat leerlingen moeten rekenen met een bepaalde verpakkingsmaat waarin het ingrediënt beschikbaar is.

De drie moeilijkste 3F-items worden nog niet beheerst door leerlingen met 3F-niveau. Het betreft complexe opgaven waarbij meerdere bewerkingen uitgevoerd moeten worden. Zowel item 38310 als 38313 betreffen het rekenen met bedragen per maand en het procentueel vergelijken van dat maandbedrag met een gegeven totaalbedrag. De moeilijkste 3F-opgave betreft het vergelijken van prijzen voor en na btw verhoging.



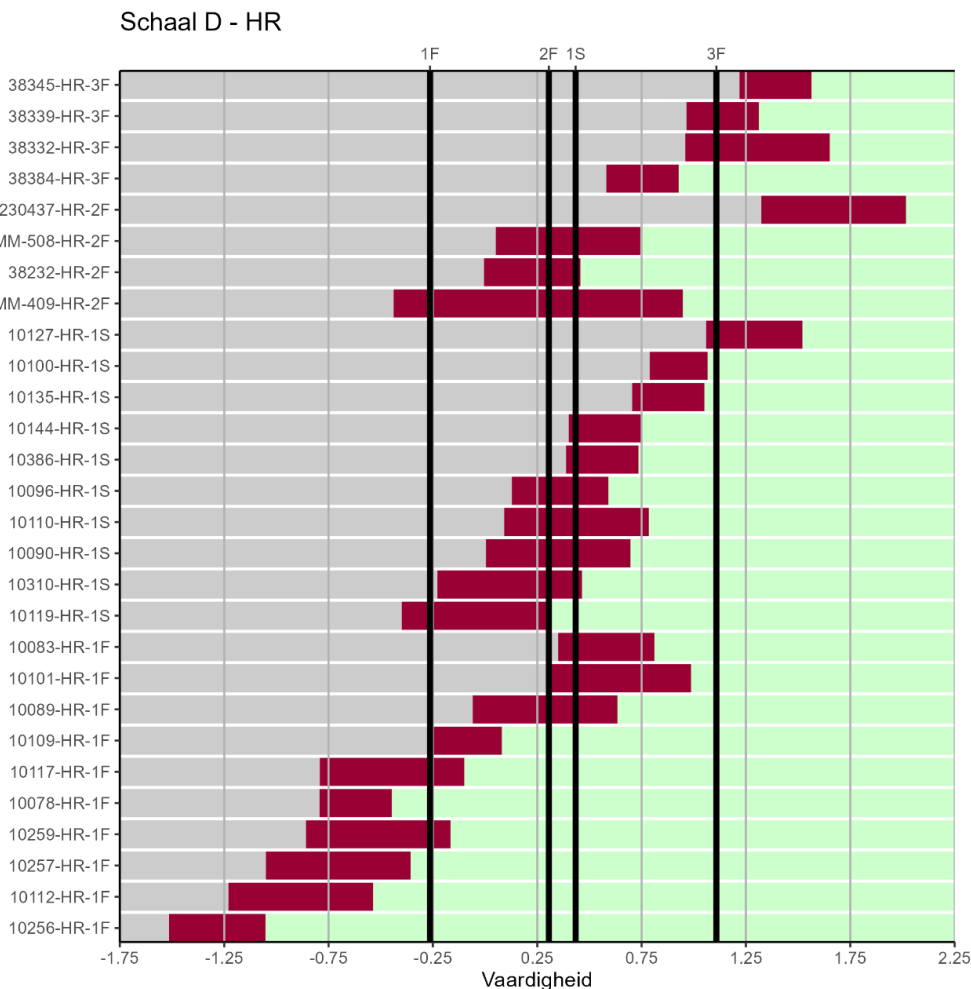
Figuur 7.11 – Rekenmachine-opgaven domein C – Verhoudingen<sup>48</sup>

<sup>48</sup> Gewogen vaardigheidsscores per deelpopulatie: pro:-1,02; vmbo-b/k:-0,48; vmbo-g/t: 0,01; havo/vwo: 0,61; totaal: 0,16

### 7.3.4 Domein D – Meten en Meetkunde

#### Hoofdrekenen

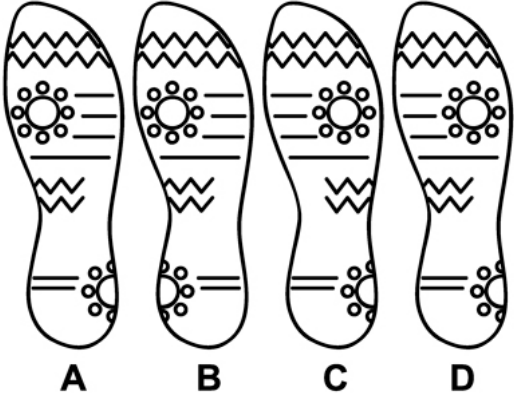

Figuur 7.12 laat de moeilijkheid van de hoofdrekenitems uit domein D – Meten en meetkunde zien. Opvallend is de grote variatie in het onderscheidend vermogen van de items. Zo heeft 2F13-MM-209-HR-2F een relatief laag onderscheidend vermogen. In deze opgave wordt gevraagd naar de wiskundige benaming die hoort bij een ijschoortje. Het gaat hier om een kennisvraag (en dus geen toepassingsvraag) uit het domein meetkunde, wat het logisch maakt dat er leerlingen zijn met 1F-niveau, 2F- of 1S-niveau die dit toevallig al weten. Leerlingen met 1F-niveau hebben een kans van iets hoger dan 0,5 om het item goed te maken, voor leerlingen met 2F- of 1S-niveau ligt dat rond de 0,65. Kortom, de meeste VO-leerlingen in leerjaar 2 beheersen deze opgave matig. Leerlingen met 3F-niveau beheersen deze opgave goed. Ook het gemakkelijkste 1S-item betreft een meetkundeopgave (zie Figuur 7.13), leerlingen met 1S-niveau beheersen de opgave matig, leerlingen vanaf 2F-niveau beheersen deze opgave goed. Leerlingen met niveau 2F of 1S hebben een matig tot goede beheersing van drie van de vier 2F-opgaven. Het gaat bijvoorbeeld om het berekenen van de oppervlakte van een U-vormige kamer. Leerlingen uit het praktijkonderwijs beheersen de drie gemakkelijkste 1F-items matig.



Figuur 7.12 – Hoofdreken-opgaven domein D – Meten en meetkunde<sup>49</sup>

<sup>49</sup> Gewogen vaardigheidsscores per deelpopulatie: pro:-1,02; vmbo-b/k:-0,48; vmbo-g/t: 0,01; havo/vwo: 0,61; totaal: 0,16



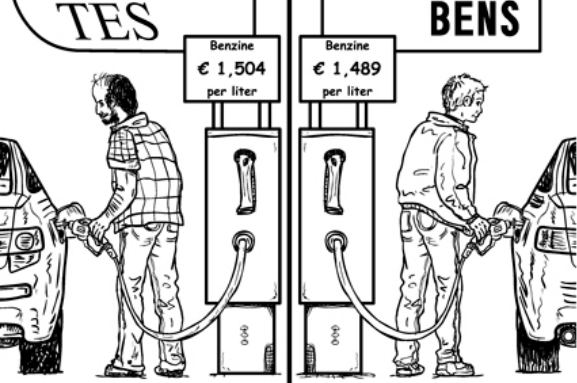
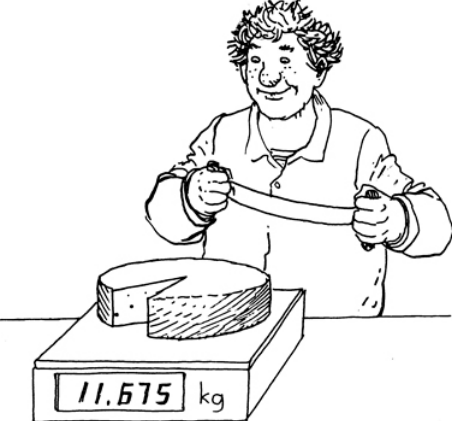
Item en referentieniveau	Item	Beheersing
10119-HR-1S	 <p data-bbox="416 741 799 819">De politie onderzoekt schoenafdrukken. Welke afdruk hoort bij deze schoen?</p> 	<p data-bbox="1027 271 1401 394">Vooral havo/vwo leerlingen (leerlingen met 2F en hoger) beheersen deze opgave goed. De helft van de vmbo-g/t leerlingen halen 1F en die beheersen deze opgave matig.</p>

Figuur 7.13 – Voorbeelditem Hoofdrekenen domein D - Meten en Meetkunde

Het moeilijkste 1F-item gaat over het berekenen van de hoeveelheid draad (c.q. omtrek) die nodig is voor het plaatsen van een hekwerk rondom een niet-rechthoekig landschap. De complexiteit van deze opgave zit in het hekwerk dat bestaat uit twee draden boven elkaar. Leerlingen moeten dus twee keer de omtrek nemen. Deze opgave wordt door leerlingen met 2F-niveau of lager nog niet beheerst. Leerlingen met niveau 1S beheersen dit item matig en leerlingen met 3F-niveau beheersen dit item goed (zie Figuur 7.12).

Leerlingen met 1F-niveau of hoger hebben een goede beheersing van item 10078 waarin leerlingen 0,8 kilometer moeten omrekenen naar meters. Het moeilijkste 1S-item (10127, zie ook Figuur 7.14) wordt door leerlingen met 3F-niveau matig beheerst. Item 10110 wordt door leerlingen met 3F-niveau goed beheerst en door leerlingen met 2F en 1S matig beheerst, dit betekent dat de meeste VO-leerlingen in leerjaar 2 deze opgave matig tot goed beheersen. Het zijn vooral leerlingen op vmbo-b/k die deze opgave nog niet beheersen waarbij opgemerkt moet worden dat vmbo-leerlingen geen onderwijsaanbod op 1S krijgen. In beide opgaven moet gerekend worden met drie decimalen achter de komma, maar in 10110 gaat het (blijkbaar) om eenvoudigere bewerkingen dan in 10127. Leerlingen met 3F-niveau

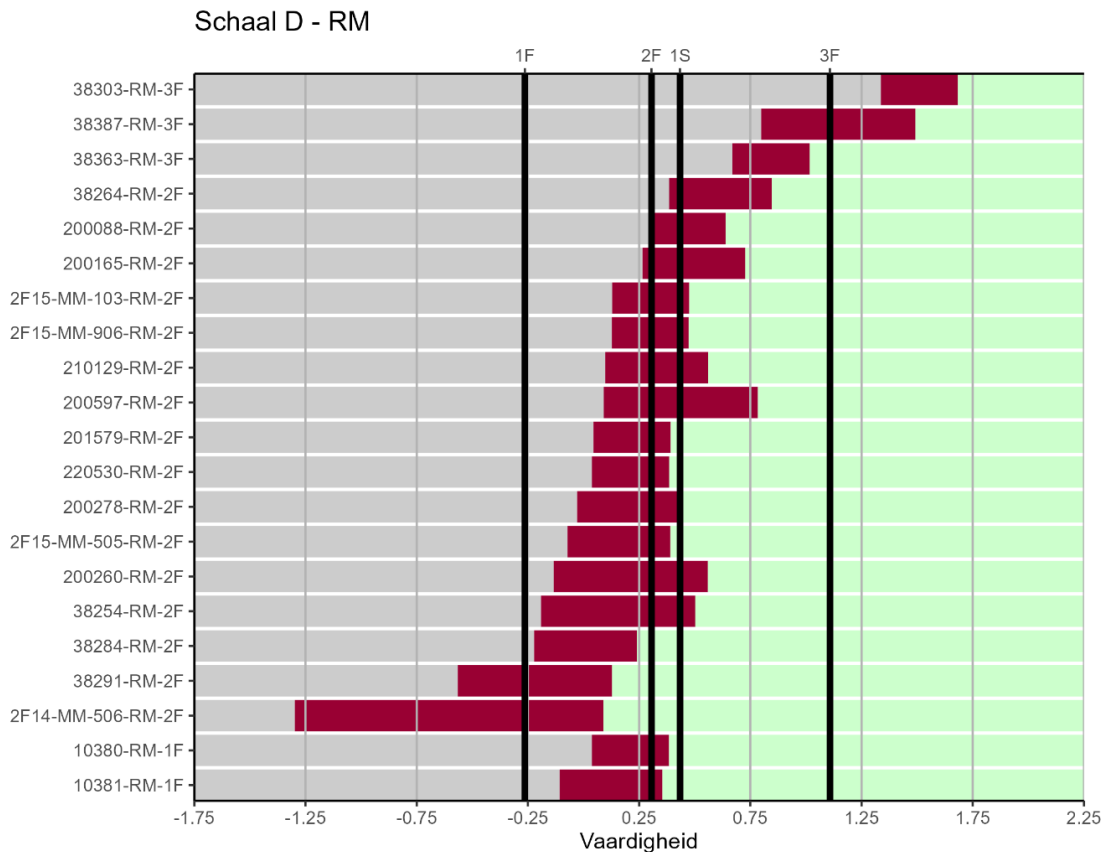
beheersen niet de moeilijkste 2F-opgave (230437), waarin een gewicht in aantal ton omgerekend moet worden naar liters, maar beheersen wel de gemakkelijkste 3F-opgave. In die 3F-opgave (38384) moet gerekend worden met gewicht per vierkante meter. De 3F-opgave waarin gerekend moet worden met land- en zeemijlen wordt door leerlingen met 3F-niveau matig beheerst. Ook de 3F-opgave waarin gerekend moet worden met kommagetallen voor het bepalen van het oppervlakte en de prijs per tegeltje voor het bepalen van de totaal prijs wordt door leerlingen met 3F-niveau matig beheerst.

Item en referentieniveau	Item	Beheersing
10127-HR-1S	 <p>Ali en Martin tanken elk 40 liter benzine. Ali tankt bij Tes en Martin bij Bens.</p> <p>Hoeveel euro moet Ali meer betalen voor 40 liter dan Martin?</p>	<p>Moeilijke opgave voor leerlingen in leerjaar 2 VO. Dit item wordt goed beheerst door leerlingen met 3F-niveau, dus enkel door de havo/vwo leerlingen in P90.</p>
10110-HR-1S	 <p>Kaasman Van der Leiden heeft een stuk kaas van 11,675 kg. Hij wil stukken snijden van ongeveer 400 gram.</p> <p>Hoeveel stukken kan hij uit het grote stuk halen?</p>	<p>Bijna de helft van de VO-leerlingen in leerjaar 2 beheersen deze opgave matig. Leerlingen met 1F-niveau, dus vooral leerlingen uit pro, vmbo-b/k en vmbo-g/t beheersen deze opgave niet.</p>

Figuur 7.14 – Meer voorbeelditems Hoofdrekenen domein D - Meten en Meetkunde

## Rekenmachine

In Figuur 7.15 is te zien dat de twee 1F-items in het domein Meten en meetkunde met rekenmachine (zie ook Figuur 7.16) moeilijker zijn dan ongeveer de helft van de 2F-items. Beide items worden door leerlingen met 2F-niveau matig en met 1S-niveau of hoger goed beheerst. Dat wil zeggen dat de meeste leerlingen in leerjaar 2 van het VO deze items nog niet beheersen.



Figuur 7.15 – Rekenmachine-opgaven domein D – Meten en meetkunde<sup>50</sup>

Het makkelijkste 2F-item betreft een meetkundeopgave waarbij het aantal symmetrieassen uit verschillende figuren gehaald moet worden. Zoals Figuur 7.15 laat zien heeft ook dit meetkunde-item een slechter onderscheidend vermogen dan de meeste items en wordt dit item matig beheerst door leerlingen die het 1F-niveau behalen. Leerlingen met 2F-niveau hebben een matige beheersing van item 2F15-MM-103-RM-2F (zie ook Figuur 7.16) en een goede beheersing van 2F-item 38284 waarbij leerlingen gewichten moeten omrekenen en moeten aftrekken (of aanvullen).

<sup>50</sup> Gewogen vaardigheidsscores per deelpopulatie: pro:-1,02; vmbo-b/k:-0,48; vmbo-g/t: 0,01; havo/vwo: 0,61; totaal: 0,16

Item en referentieniveau	Item	Beheersing
10381-RM-1F	<p>Siem verft de muur van zijn kamer. De muur is 2,4 m hoog en 7,45 m lang.</p> <p>Hoeveel m<sup>2</sup> is de oppervlakte van de muur?</p> <p>Rond af op een heel getal.</p> <p>_____ m<sup>2</sup></p>	Goede beheersing door leerlingen met 2F of hoger. Leerlingen met 1F-niveau beheersen dit item niet. Kortom, dit item wordt vooral beheerst door havo/vwo leerlingen en door een klein deel van de vmbo-g/t leerlingen.
10380-RM-1F	<p>Boerenkaas kost € 24,- per kilogram. Anne koopt 450 gram boerenkaas.</p> <p>Hoeveel euro moet zij betalen?</p> <p>€ _____</p>	Goede beheersing door leerlingen met 2F of hoger. Leerlingen met 1F-niveau beheersen dit item niet. Kortom, dit item wordt vooral beheerst door havo/vwo leerlingen en door een klein deel van de vmbo-g/t leerlingen.
2F15-MM-103-RM-2F	 <p>In de gelukstrekkings wordt 24 uur lang in elk kwartier een prijs van € 100 000 uitgekeerd.</p> <p>Hoeveel wordt zo in totaal in 24 uur uitgekeerd?</p> <p>€ _____</p>	Matige beheersing voor leerlingen met 2F-niveau en goede beheersing voor leerlingen met 1S-niveau.

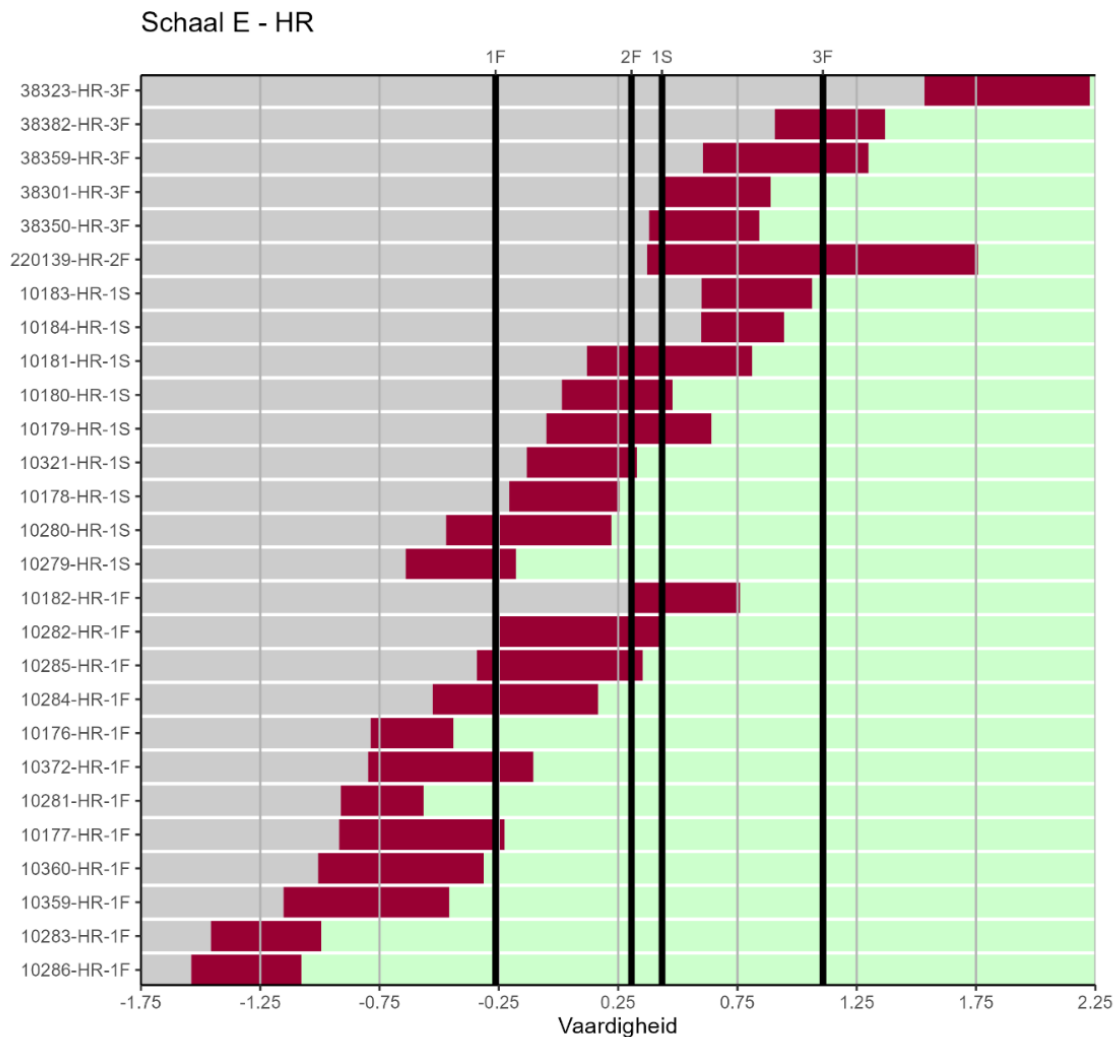
Figuur 7.16 – Voorbeelditems Rekenmachine domein D – Meten en Meetkunde

### 7.3.5 Domein E – Verbanden

#### Hoofdrekenen

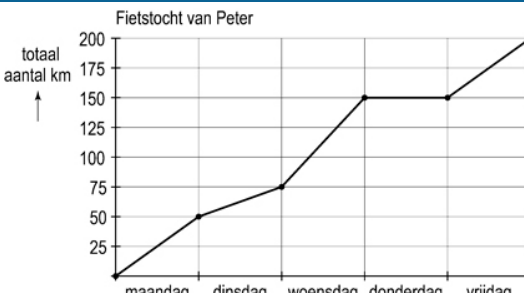
Leerlingen met 1F-niveau beheersen zes van de twaalf 1F-items goed en vijf van de twaalf items matig (zie Figuur 7.17). Deze leerlingen beheersen onder andere het aflezen van een lijngrafiek goed (item 10176) en het aflezen van bus-vertrektijden matig. Een kwart van de vmbo-b/k leerlingen en de helft van de vmbo-g/t leerlingen beheerst deze opgave. Leerlingen uit het praktijkonderwijs functioneren lager dan 1F-niveau en beheersen deze opgave dus niet. Daarnaast beheersen leerlingen met 1F-niveau de twee gemakkelijkste 1S-items (10182 en 10280) in het domein Verbanden matig. Deze twee opgaven worden door leerlingen met 2F-niveau of hoger goed beheerst. Leerlingen met 2F-niveau hebben moeite met de 1S-opgave waarin ze bij een omschrijving de passende grafiek moeten kiezen (item 10181). Leerlingen met 3F-niveau beheersen deze opgave goed. De opgave waarin gegevens uit een staafdiagram omgerekend moet worden naar procenten wordt goed beheerst door leerlingen met 1S-niveau en matig door leerlingen met 2F-niveau. Leerlingen met 1S-niveau of hoger hebben een matige beheersing van het enige 2F-item in het domein Verbanden (zie ook Figuur 7.18). Zoals in Figuur 7.17 is te zien heeft dit item niet zo'n hoog onderscheidend vermogen als de andere items uit dit domein. De drie

gemakkelijkste 3F-items uit het domein Verbanden betreffen drie opgaven over patronen en regelmaat. Leerlingen met 1S beheersen de eerste twee 3F-opgaven matig, leerlingen met 3F beheersen deze twee opgaven matig tot goed. Tot slot hebben leerlingen met niveau 3F een matige beheersing van het maken van berekeningen aan de hand van een afstandstabel. Het maken van complexe berekeningen aan de hand van meerdere tabellen met prijzen beheersen leerlingen in het tweede leerjaar VO nog niet.



Figuur 7.17 – Hoofdreken-opgaven domein E – Verbanden<sup>51</sup>

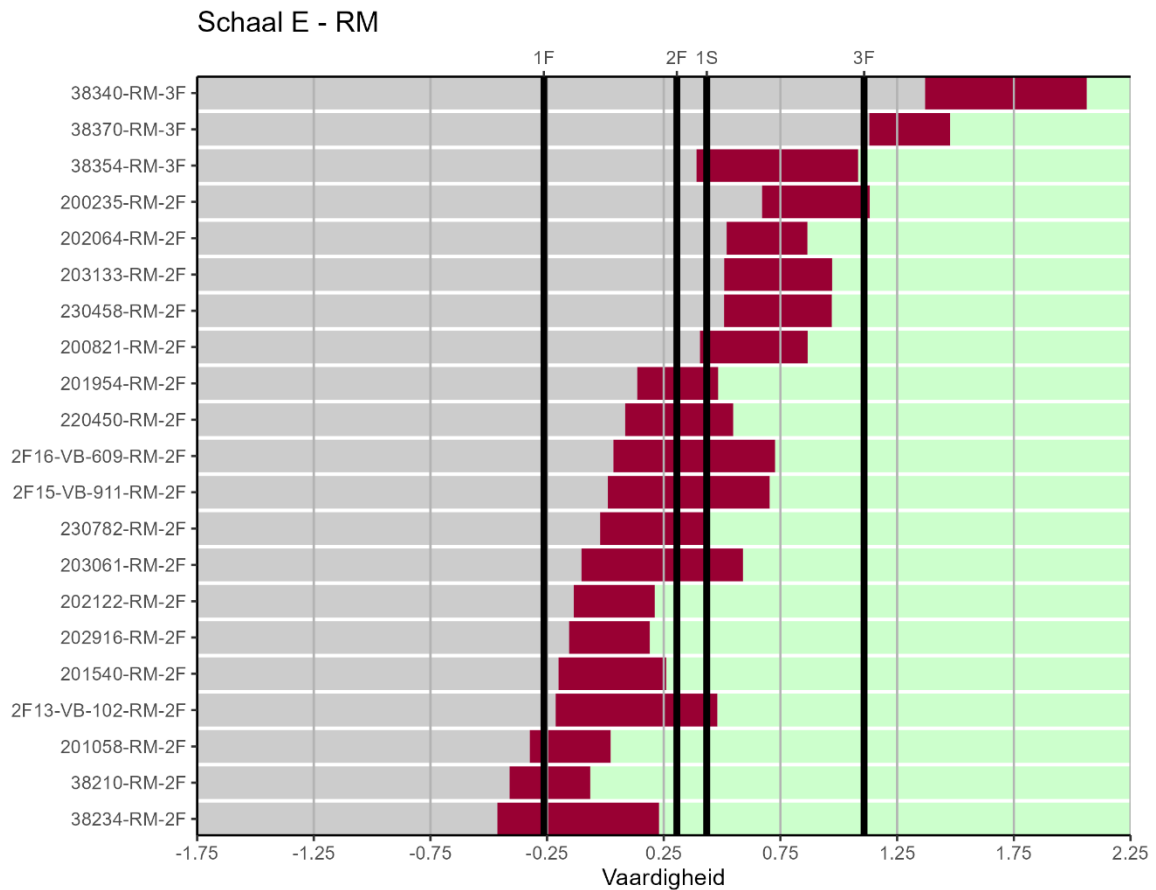
<sup>51</sup> Gewogen vaardigheidsscores per deelpopulatie: pro:-1,02; vmbo-b/k:-0,48; vmbo-g/t: 0,01; havo/vwo: 0,61; totaal: 0,16

Item en referentieniveau	Item	Beheersing																																																
10176-HR-1F	<p>Fietstocht van Peter</p>  <p>Peter heeft een fietstocht door Nederland gemaakt. Hierboven zie je een grafiek van de kilometers die hij heeft gefietst.</p> <p>Op welke dag heeft hij de meeste kilometers gefietst?</p>	Gemakkelijke opgave. Leerlingen met 1F-niveau of hoger beheersen deze opgave goed. Dat wil zeggen dat de meeste vmbo-g/t en havo/vwo leerlingen deze opgave goed beheersen en dat een klein deel (25%) van de vmbo-b/k leerlingen deze opgave goed beheerst.																																																
10279-HR-1S	<table border="1" data-bbox="367 784 1053 1142"> <thead> <tr> <th colspan="6">Ledenaantallen van voetbalverenigingen in Plotterdam</th> </tr> <tr> <th></th> <th>1999</th> <th>2000</th> <th>2001</th> <th>2002</th> <th>2003</th> </tr> </thead> <tbody> <tr> <td><b>S.D.O.L.</b></td> <td>114</td> <td>115</td> <td>108</td> <td>113</td> <td>139</td> </tr> <tr> <td><b>T.I.K. '65</b></td> <td>567</td> <td>567</td> <td>570</td> <td>578</td> <td>578</td> </tr> <tr> <td><b>FC Punten</b></td> <td>317</td> <td>300</td> <td>317</td> <td>327</td> <td>327</td> </tr> <tr> <td><b>FC Vooruit</b></td> <td>292</td> <td>283</td> <td>275</td> <td>266</td> <td>250</td> </tr> <tr> <td><b>P.K.D.O.</b></td> <td>370</td> <td>371</td> <td>375</td> <td>376</td> <td>379</td> </tr> <tr> <td><b>SC HAMONDI</b></td> <td>160</td> <td>148</td> <td>162</td> <td>171</td> <td>173</td> </tr> </tbody> </table> <p>Van welke club is het ledenaantal sinds 1999 elk jaar gestegen?</p>	Ledenaantallen van voetbalverenigingen in Plotterdam							1999	2000	2001	2002	2003	<b>S.D.O.L.</b>	114	115	108	113	139	<b>T.I.K. '65</b>	567	567	570	578	578	<b>FC Punten</b>	317	300	317	327	327	<b>FC Vooruit</b>	292	283	275	266	250	<b>P.K.D.O.</b>	370	371	375	376	379	<b>SC HAMONDI</b>	160	148	162	171	173	Gemakkelijke opgave. Leerlingen met 1F-niveau of hoger beheersen deze opgave goed. Dat wil zeggen dat de meeste vmbo-g/t en havo/vwo leerlingen deze opgave goed beheersen en dat een klein deel (25%) van de vmbo-b/k leerlingen deze opgave goed beheerst.
Ledenaantallen van voetbalverenigingen in Plotterdam																																																		
	1999	2000	2001	2002	2003																																													
<b>S.D.O.L.</b>	114	115	108	113	139																																													
<b>T.I.K. '65</b>	567	567	570	578	578																																													
<b>FC Punten</b>	317	300	317	327	327																																													
<b>FC Vooruit</b>	292	283	275	266	250																																													
<b>P.K.D.O.</b>	370	371	375	376	379																																													
<b>SC HAMONDI</b>	160	148	162	171	173																																													
220139-HR-2F	<p><b>Prijzen tweepersoonskamers hotel La Plagne in euro's per persoon</b></p> <table border="1" data-bbox="367 1321 1133 1724"> <thead> <tr> <th>prijzen per persoon</th> <th>zo</th> <th>ma</th> <th>di</th> <th>wo</th> </tr> </thead> <tbody> <tr> <td>aankomst op</td> <td>20-dec</td> <td>21-dec</td> <td>22-dec</td> <td>23-dec</td> </tr> <tr> <td>aantal nachten</td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>2 nachten</td> <td>282</td> <td>282</td> <td>282</td> <td>282</td> </tr> <tr> <td>3 nachten</td> <td>334</td> <td>334</td> <td>334</td> <td>352</td> </tr> <tr> <td>4 nachten</td> <td>386</td> <td>386</td> <td>399</td> <td>422</td> </tr> <tr> <td>5 nachten</td> <td>472</td> <td>490</td> <td>508</td> <td>525</td> </tr> </tbody> </table> <p>De heer en mevrouw Youssof boeken een tweepersoonskamer in dit hotel voor vier nachten. Ze komen er op 21 december aan.</p> <p>Hoeveel betalen ze in totaal voor de kamer?</p>	prijzen per persoon	zo	ma	di	wo	aankomst op	20-dec	21-dec	22-dec	23-dec	aantal nachten					2 nachten	282	282	282	282	3 nachten	334	334	334	352	4 nachten	386	386	399	422	5 nachten	472	490	508	525	Moeilijke opgave die wordt beheerst door leerlingen met 1S-niveau en matig door leerlingen met 3F-niveau.													
prijzen per persoon	zo	ma	di	wo																																														
aankomst op	20-dec	21-dec	22-dec	23-dec																																														
aantal nachten																																																		
2 nachten	282	282	282	282																																														
3 nachten	334	334	334	352																																														
4 nachten	386	386	399	422																																														
5 nachten	472	490	508	525																																														

Figuur 7.18 – Voorbeelditems Hoofdrekenen domein E - Verbanden

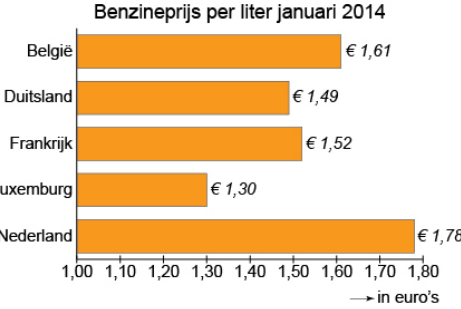
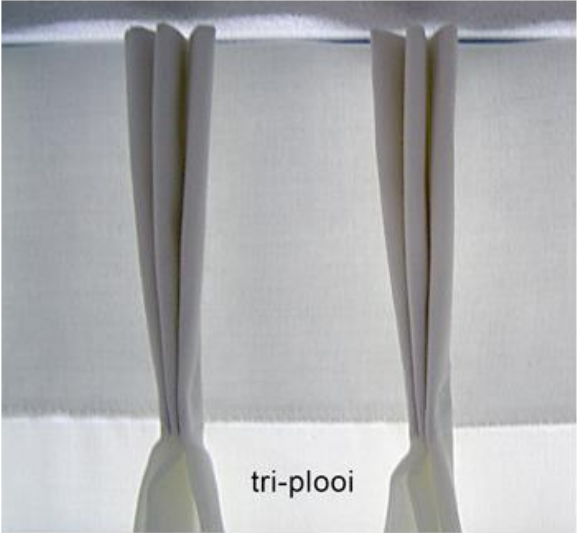
## Rekenmachine

De rekenmachine-items voor het domein Verbanden betreffen voornamelijk 2F-items. Zoals Figuur 7.19 laat zien hebben de meeste items een goed onderscheidend vermogen. Ook valt op dat leerlingen met 1F-niveau een matige beheersing hebben van de drie gemakkelijkste opgaven in dit domein. Van de 21 items beheersen leerlingen met 2F-niveau zes items goed en zeven items matig. Leerlingen met 3F-niveau beheersen alle opgaven goed, behalve de twee moeilijkste opgaven. Het 2F-item 202122 (zie ook Figuur 7.19) wordt door leerlingen met 2F-niveau of hoger goed beheerst. Het 2F-item 230782 (idem) wordt door leerlingen met 2F-niveau matig beheerst en door leerlingen met 1S-niveau of hoger goed beheerst.



Figuur 7.19 – Rekenmachine-opgaven domein E – Verbanden<sup>52</sup>

<sup>52</sup> Gewogen vaardigheidsscores per deelpopulatie: pro:-1,02; vmbo-b/k:-0,48; vmbo-g/t: 0,01; havo/vwo: 0,61; totaal: 0,16

Item en referentieniveau	Item	Beheersing
202122-RM-2F	<p style="text-align: center;"><b>Benzineprijs per liter januari 2014</b></p>  <p>Je rijdt naar je vakantiebestemming. Je tankt 40 liter benzine in Nederland. Onderweg tank je 40 liter in Luxemburg.</p> <p><b>Hoeveel euro betaal je voor die 40 liter in Luxemburg minder dan in Nederland?</b></p>	<p>Leerlingen met 2F of hoger beheersen dit item goed. Kortom, dit item wordt vooral beheerst door havo/vwo leerlingen.</p>
230782-RM-2F	<p>Nora maakt één gordijn met een tri-plooi, voor een raam van 2,20 meter breed.</p>  <p style="text-align: center;">tri-plooi</p> <p><b>Vuistregel gordijn met tri-plooiën:</b>  aantal meters stof = breedte raam in meters x 2,80 + 0,25</p> <p><b>Hoeveel meter stof heeft Nora nodig om het gordijn te maken?</b>  <b>Geef je antwoord in twee decimalen.</b></p>	<p>Leerlingen met 2F-niveau of hoger (havo/vwo leerlingen) beheersen dit item matig. Leerlingen met 3F-niveau (hoogvaardige vwo-leerlingen) beheersen dit item goed.</p>

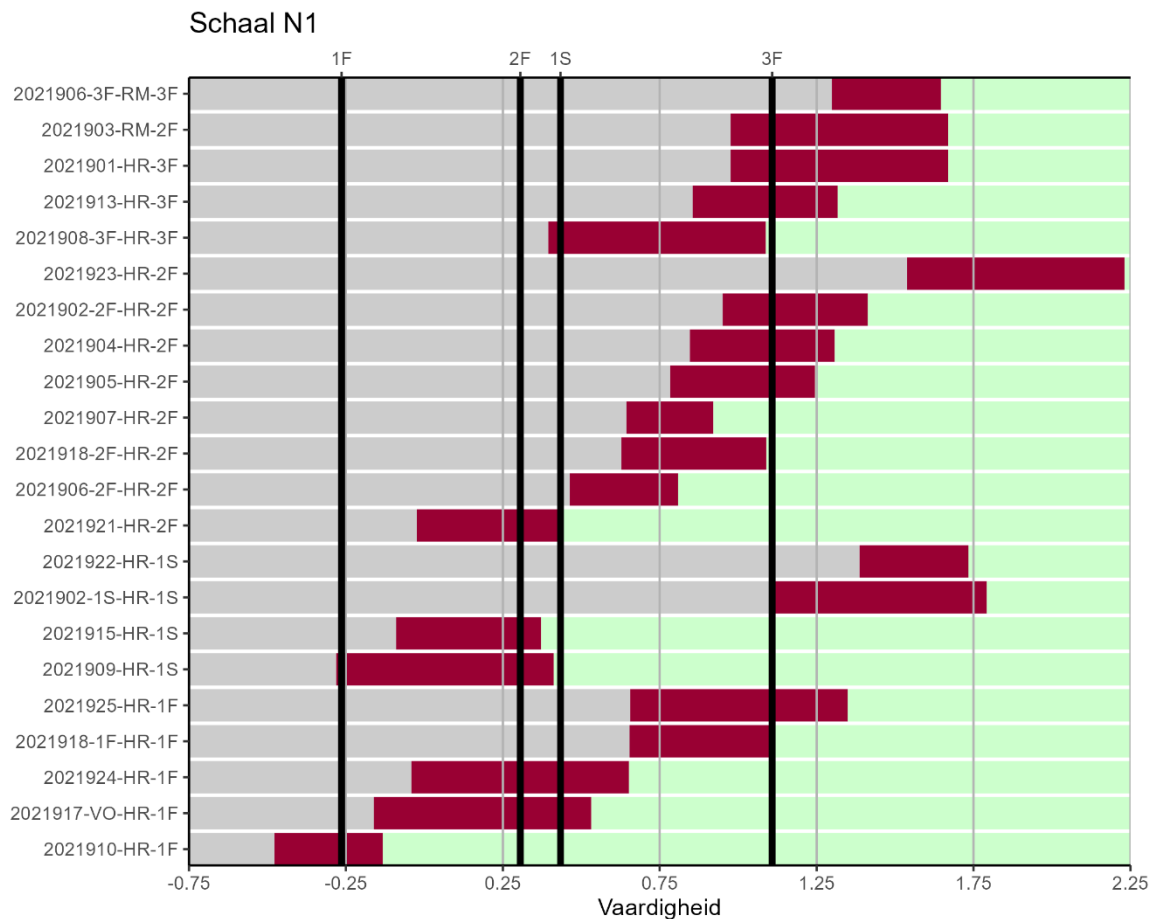
Figuur 7.20 – Voorbeelditems Rekenmachine domein E - Verbanden

### 7.3.6 Domein N1 – Probleemoplossen

Voor Wiskundig Probleemoplossen zijn slechts twee rekenmachine-opgaven afgenomen, daarom zijn deze opgenomen in één gezamenlijke afbeelding met de hoofdteken-opgaven (Figuur 7.21). Het betreft één 3F-item dat door de meeste leerlingen nog niet beheerst wordt en één 2F-item dat matig wordt beheerst door leerlingen met 3F-niveau. Figuur 7.21 laat zien dat de moeilijkheid van de items niet overeenkomt met het referentieniveau waarin ze zijn ingedeeld. Dit is logisch want probleemoplossen is geen onderdeel van het referentiekader, wel van de kerndoelen (zie ook Hoofdstuk 2). Voor toewijzing



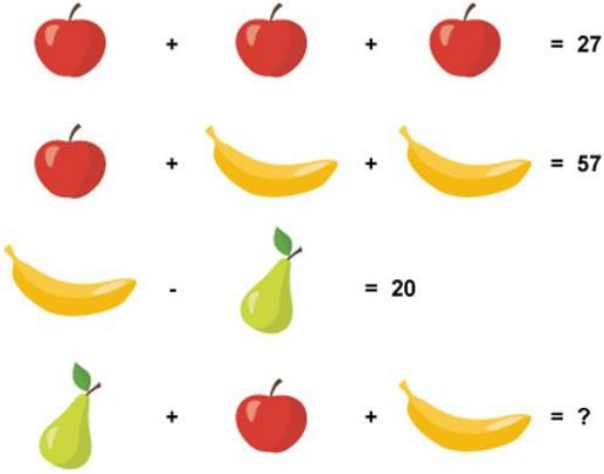
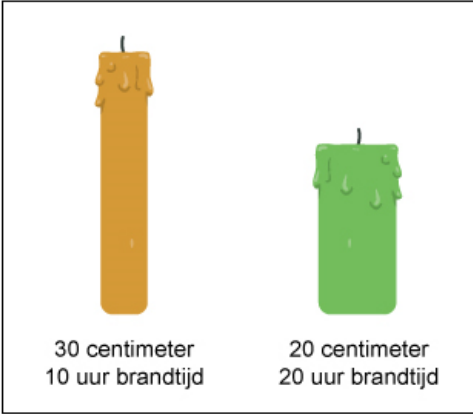
van de items aan boekjes, passend bij het niveau van de leerling, was het echter noodzakelijk om deze items in te delen op referentieniveau. Dit is gedaan door een ervaren toetsdeskundige rekenen-wiskunde die daarbij lette op de bewerkingen en het niveau van redeneren. Zoals in Figuur 7.21 is te zien, bevat deze schaal relatief veel items met een laag onderscheidend vermogen (lange rode balkjes): een hogere vaardigheid levert leerlingen niet direct een hogere kans op om een item goed te maken.



Figuur 7.21 – Rekenmachine- en Hoofdrekenopgaven domein N1 – Probleemoplossen<sup>53</sup>

Leerlingen die 1F beheersen, beheersen geen enkele opgave van het domein Probleemoplossen goed. Alleen het gemakkelijkste 1F-item (2021910, zie Figuur 7.22) en het gemakkelijkste 1S-item (2021909) beheersen zij matig. Daarnaast valt op dat ook leerlingen met 2F- of 1S-niveau weinig opgaven uit dit domein goed beheersen. Het moeilijkste 1F-item (2021925, zie Figuur 7.22) wordt niet beheerst door leerlingen met niveau 1S of lager. Deze leerlingen hebben daarentegen wel een matige beheersing van het makkelijkste 3F-item in dit domein (2021908, zie Figuur 7.22). Leerlingen met 3F-niveau beheersen deze items respectievelijk matig en goed. Kortom, voor de meeste leerlingen in leerjaar 2 van het VO is dit item nog te moeilijk.

<sup>53</sup> Gewogen vaardigheidsscores per deelpopulatie: pro:-1,02; vmbo-b/k:-0,48; vmbo-g/t: 0,01; havo/vwo: 0,61; totaal: 0,16

Item en referentieniveau	Item	Beheersing
2021910-HR-1F	 <p>Elk stuk fruit staat voor een bepaalde waarde. Welke getal hoort op de plaats van het vraagteken te staan?</p>	<p>Leerlingen met 1F-niveau beheersen dit item matig. Leerlingen met 2F-niveau of hoger beheersen dit item goed. Dit betekent dat de meeste vmbo/t en havo/vwo leerlingen dit item matig tot goed beheersen.</p>
2021925-HR-1F	<p>Marloes, Sanne, Petra en Lieke moeten groepjes van 2 maken.</p> <p>Hoeveel verschillende tweetallen kunnen er samengesteld worden uit 4 personen?</p> <p>_____ tweetallen</p>	<p>Zeer moeilijk item. Leerlingen met 3F-niveau (i.e., de hoogvaardige vwo-leerlingen) beheersen dit item matig.</p>
2021908-HR-3F	 <p>Maaïke steekt deze twee kaarsen tegelijk aan.</p> <p>Na hoeveel uur branden zijn de kaarsen even lang?</p> <p>_____ uur</p>	<p>Leerlingen met 1S-niveau of hoger beheersen dit item matig. Dit item wordt door de gemiddelde havo/vwo leerling goed beheerst.</p>

Figuur 7.22 – Voorbeelditems Hoofdrekenen domein N1 – Probleem oplossen

## 8 Samenhang tussen leerlingprestaties en kenmerken van leerlingen, docenten en het onderwijsleerproces

### 8.1 Inleiding

In dit hoofdstuk analyseren we aan de hand van multilevel analyses hoe de rekenvaardigheid van leerlingen samenhangen met kenmerken van leerlingen, docenten, scholen en het onderwijsleerproces (onderzoeksvraag 3, 4, 6 en 7). Daaraan voorafgaande beschrijven we eerst de wijze van imputeren (Paragraaf 8.2) van de data en de opzet van de multilevel analyses (Paragraaf 8.3).

### 8.2 Multiple imputatie

Voorafgaand aan de multilevel analyses zijn de ontbrekende waarden op de predictoren geïmputeerd, waarbij we rekening gehouden hebben met de hiërarchische structuur van de data<sup>54</sup>. We zijn gestart met een dataset van 2606 leerlingen op 125 verschillende scholen. Pro-leerlingen zijn – conform afspraak met de begeleidingscommissie – buiten beschouwing gebleven. De dataset bevat een score voor de rekenvaardigheid van leerlingen met daaraan gekoppeld kenmerken van leerlingen, kenmerken van hun wiskundeleraar of hun -klas en kenmerken van de school die ze bezoeken. De kenmerken hebben betrekking op het onderwijsleerproces, ze zijn domeinspecifiek voor wiskunde of van algemene aard. Ze worden gebruikt als predictoren (onafhankelijke variabelen) in de multilevel modellen. We hebben leerlingen met 75% of meer ontbrekende waarden op de predictoren voor de multilevel analyses uit de dataset verwijderd. Dit resulteert in een dataset van 2334 leerlingen op 112 scholen met een gemiddelde klasgrootte van 20,8 leerlingen (standaarddeviatie=4,9; range 3-30).

Voor het imputeren van de ontbrekende waarden is gebruik gemaakt van het open source programma R en het pakket Mice (van Buuren & Groothuis-Oudshoorn, 2011). Bij de imputatie is uitgegaan van een imputatiemodel met de school als clustervariabele en waarbij de ontbrekende waarden worden geschat door individuele WML-schattingen van de rekenvaardigheid en alle beschikbare kenmerken van leerling, docent/klas en school die deel zullen uitmaken van de hierna volgende multilevel analyses (predictie matrix). De ontbrekende waarden van zowel de predictoren op het niveau van de leerling (niveau 1) als de predictoren op het niveau van de docent/klas/school (niveau 2) zijn geïmputeerd volgens de *Predictive Mean Matching*-methode (PMM): *2l.pmm* voor niveau 1 en *2lonly.pmm* voor niveau 2), gebruikmakend van vijf imputatierondes. Mice gaat uit van de *Fully Conditional Specification* (FCS) methode, ook bekend als *multivariate imputation by chained equations*: elke variabele wordt om de beurt geïmputeerd middels een reeks van imputatiemodellen. PMM houdt rekening met het meetniveau van de variabelen die geïmputeerd dienen te worden, het is een generieke methode<sup>55</sup>. De *random seed* is gesteld op 51000048. Door het vastzetten van de *random seed* is het mogelijk om de imputatie exact te repliceren en daarmee op dezelfde resultaten te komen. De vijf imputatierondes hebben geresulteerd in vijf verschillende geïmputeerde datasets die zijn bewaard als een *mids*-databestand zodat direct gepoolde multilevel analyses uitgevoerd kunnen worden in R.

<sup>54</sup> In Bijlage 4 is de bijbehorende syntax opgenomen.

<sup>55</sup> Zie Tabel 7.2 in <https://stefvanbuuren.name/fimd/sec-multioutcome.html>

### 8.3 Multilevel analyses

Vanwege de hiërarchische structuur van de geïmputeerde datasets zijn multilevel modellen toegepast om de onderzoeksvragen te kunnen beantwoorden. De multilevel modellen betreffen steeds *random intercept modellen* (Snijders & Bosker, 2012; pp. 41). We onderscheiden twee niveaus: de leerlingen vormen niveau 1 en de docent/klas/school is niveau 2. Het schoolniveau valt samen met het docent/klas-niveau omdat per school slechts één klas en één docent heeft deelgenomen. De gehanteerde twee niveaus impliceren ook dat de totale variantie in rekenvaardigheid van de leerlingen wordt opgedeeld in variantie op niveau 1 (tussen leerlingen) en niveau 2 (tussen scholen). De afhankelijke variabele is in alle modellen een *'weighted maximum likelihood (WML) schatter* van de rekenvaardigheid van een individuele leerling. Voor elk gepoold multilevel model is de proportie verklaarde totale variantie berekend volgens de volgende formule (Snijders & Bosker, 2012):

$$R^2_{\text{Totaal}} = 1 - (\sigma^2_{\text{model}} + \tau^2_{\text{model}}) / (\sigma^2_{\text{nulmodel}} + \tau^2_{\text{nulmodel}})$$

Waarbij  $\sigma^2_{\text{model}}$  gelijk staat aan de variantie tussen leerlingen in één van de hieronder beschreven modellen 1 tot 4 en  $\tau^2_{\text{model}}$  is de variantie tussen scholen in deze modellen. De term  $\sigma^2_{\text{nulmodel}}$  is de variantie tussen leerlingen en  $\tau^2_{\text{nulmodel}}$  is de variantie tussen scholen in het nulmodel waarin alleen een intercept is meegenomen. Het berekenen van de verklaarde variantie op leerlingniveau, respectievelijk schoolniveau gaat als volgt:

$$R^2_{\text{leerling}} = 1 - (\sigma^2_{\text{model}} / \sigma^2_{\text{nulmodel}})$$
$$R^2_{\text{school}} = 1 - (\tau^2_{\text{model}} / \tau^2_{\text{nulmodel}})$$

De vijf geïmputeerde datasets maken het toepassen van gepoolde multilevel analyses noodzakelijk voor het correct schatten van de regressiecoëfficiënten, varianties en hun standaardfouten (Snijders & Bosker, 2012, p 135). We hebben gebruik gemaakt van het R-pakket *lmer (linear mixed effects models)* met de toevoegen van *with* om een serie van geneste statistische modellen te fitten. Er wordt vijf keer een model geschat waarna de schattingen van de parameters worden samengenomen. Met *testEstimates(as.mitml.result())* wordt deze gepoolde resultaten van de vijf modellen verkregen.

### 8.4 Overzicht van de multilevel modellen

Om de inzichtelijkheid van de resultaten te bevorderen – en op verzoek van de begeleidingscommissie – hebben we de multilevel analyses uitgevoerd per onderwijstype (gestratificeerd). Informatie over het onderwijstype is afkomstig van de wiskundedocent; in de vragenlijst hebben ze aangegeven tot welk onderwijstype de deelnemende klas behoort. We onderscheiden drie strata, hierna onderwijstypegroepen genoemd:

- vmbo-b/k: 394 leerlingen uit 25 klassen;
- vmbo-gl/tl: 573 leerlingen uit 26 klassen;
- havo-vwo; 1035 leerlingen uit 44 klassen.

Voor 332 leerlingen (17 klassen) was het onderwijstype onbekend; deze leerlingen zijn buiten de gestratificeerde multilevel analyses gelaten.

De onderzoeksvragen zijn beantwoord door vier verschillende modellen op te stellen per onderwijstypegroep. De opzet ervan wordt hieronder beschreven. Voor elk model geldt dat alle predictoren zijn gestandaardiseerd (z-scores), ook de dummyvariabelen, waardoor directe vergelijking van de regressiecoëfficiënten mogelijk wordt. Bovendien hebben we bij elk model via een *backwards*

selectieprocedure één voor één de predictor met de hoogste p-waarde verwijderd totdat er alleen significante predictoren voor de rekenvaardigheid over zijn. Op deze manier werken we van een maximaal model naar een spaarzaam model. Het significantieniveau is gesteld op  $\alpha = .05$ . Resultaten van onderstaande Modellen 0 tot en met 3 vormen de input voor het opstellen van het eindmodel (Model 4).

#### Model 0

Bij Model 0 bevat alleen een random intercept en heeft de rekenvaardigheid van leerlingen (WML score) als afhankelijke variabele. Dit uitgangsmodel geeft weer hoeveel variantie er is op leerlingniveau (niveau 1) en klas/schoolniveau (niveau 2) om volgens hiermee de proportie verklaarde variantie ( $R^2$ ) van de overige modellen per onderwijstypengroep te kunnen schatten.

#### Model 1 (onderzoeksvraag 3)

Model 1 beoogt de samenhang tussen leerlingkenmerken en rekenvaardigheid (WML score) inzichtelijk te maken. In eerste instantie zijn vier kenmerken van het onderwijsleerproces (ordelijkheid wiskundeles, helderheid instructie wiskundeles, oplossingsstrategieën in de wiskundeles en directe instructie in de wiskundeles) van de leerling gebruikt als predictoren. Uit deze set zijn alleen de significante kenmerken van het onderwijsleerproces geselecteerd. De volgende stap behelst uitbreiding met algemene leerlingkenmerken (jongen, vordering, aantal boeken thuis, thuistaal Nederlands en verwachte onderwijstype in leerjaar 3) en met vijf domeinspecifieke leerlingkenmerken (plezier, zelfvertrouwen, belang, attributies positieve wiskundeprestaties en een *Growth mindset* ten opzicht van wiskunde). Opnieuw zijn alleen de significante leerlingkenmerken geselecteerd. Dit heeft geresulteerd in een spaarzame model voor de samenhang van onderwijsleerproces-, algemene en domeinspecifieke leerlingkenmerken met de rekenvaardigheid van leerlingen, per onderwijstypengroep.

#### Model 2 (onderzoeksvraag 4 en 6)

Model 2 toont de samenhang tussen docent- en klaskenmerken en rekenvaardigheid (WML score). De selectieprocedure van predictoren is conform die van Model 1. We zijn gestart met uitbreiden van Model 0 met een set van 23 kenmerken van het onderwijsleerproces afkomstig uit de vragenlijst voor docenten. Het resulteert in het maximale Model 2a met de volgende predictoren:

- aantal minuten wiskunde per week
- intensiteit ICT gebruik
- gebruik rekenmachine tijdens de wiskundeles
- geschiktheid wiskundemethode
- zeven leerlingactiviteiten tijdens de wiskundeles:
  - leerkrachtgestuurde klassikale lesactiviteiten
  - probleemoplossen
  - het zelfstandig toepassen van het geleerde bij nieuwe probleemsituaties
  - werken in heterogene vaardigheidsgroepen
  - werken in homogene vaardigheidsgroepen
  - uitwerkingen opschrijven op het bord
  - uitleggen hoe ze op gegeven antwoord zijn gekomen
- differentiatie
- omvang en bespreking huiswerk
- vier vormen van feedback op huiswerk
  - huiswerk corrigeren en feedback geven aan leerlingen
  - huiswerk bespreken in de klas
  - controleren of leerlingen huiswerk gemaakt hebben
  - leerlingen kijken zelf hun huiswerk na
- intensiteit evaluatie
- drie typen toetsen:

- genormeerde toetsen (bv. Cito)
- methodegebonden toetsen
- zelfgemaakte toetsen
- direct instructie model
- ontdekkend of onderzoekend leren
- alle leerlingen meerdere oplossingsstrategieën aanbieden

Opnieuw zijn alleen de significante kenmerken van het onderwijsleerproces geselecteerd (spaarzame Model 2a). De volgende stap betreft de uitbreiding van het spaarzame Model 2a met vier algemene docent- en klaskenmerken (man, lesbevoegdheid, lesgegeven in wiskunde, aantal leerlingen in getoetste klas) (Model 2b) en drie domeinspecifieke docentkenmerken (tijdsbesteding (bij)scholing, zelfvertrouwen wiskunde-instructie, *Growth mindset* ten opzichte van wiskunde) (Model 2c). De niet-significante kenmerken uit model 2b en 2c zijn vervolgens weer verwijderd. Het resultaat is een spaarzaam model voor de samenhang van docent- en klaskenmerken met de rekenvaardigheid van leerlingen per onderwijstypengroep.

### Model 3 (onderzoeksvraag 4 en 6)

De samenhang tussen schoolkenmerken en rekenvaardigheid (WML score) van leerlingen is onderzocht met Model 3 per onderwijstypengroep. De selectieprocedure correspondeert met die van Model 1 en 2. De eerste stap is weer het uitbreiden van het Model 0 met een set van negen algemene schoolkenmerken (maximale Model 3a):

- vestigingsgrootte
- specifiek onderwijsconcept
- aandeel leerlingen dat afkomstig is uit een economische achterstandsgezin
- aandeel leerlingen dat Nederlands niet als eerste taal heeft
- stedelijkheid
- prestatiegerichte schoolklimaat door docenten
- prestatiegerichte schoolklimaat door ouders
- prestatiegerichte schoolklimaat door leerlingen
- belang leerlingvolgsysteem

Uit deze set zijn alleen de significante schoolkenmerken behouden (spaarzame Model 3a) voor de volgende stap: het uitbreiden met drie domeinspecifieke schoolkenmerken: extra wiskundelessen buiten schooltijd, bevordering interesse in wiskunde en aantal aangeboden professionaliseringsactiviteiten voor wiskundedocenten (Model 3b). Ook nu zijn alleen de significante domeinspecifieke schoolkenmerken geselecteerd voor een spaarzaam model voor de samenhang van algemene en domeinspecifieke schoolkenmerken met de rekenvaardigheid van leerlingen.

### Model 4 (onderzoeksvraag 7)

Het laatste model is het spaarzame eindmodel. Het is opgebouwd uit de voorafgaande drie spaarzame modellen met leerlingkenmerken (Model 1), docent- en klaskenmerken (Model 2) en schoolkenmerken (Model 3). Na verwijdering van de niet-significante predictoren via *backwards selection* is een compact eindmodel met enkel significante voorspellers gevormd, apart voor ieder onderwijstypengroep.

## 8.5 Resultaten multilevel analyses

### 8.5.1 Nulmodel

Tabel 8.1 toont de parameterschattingen, de schatting van de variantie op leerling en schoolniveau en de intraclass correlatie coefficient (ICC) voor de nulmodellen per onderwijstypegroep. De ICC geeft weer welk deel van de totale variantie is geclusterd op het hoogste niveau, in ons geval het school-/klasniveau. De gepresenteerde resultaten zijn verkregen na pooling van de schattingen op basis van vijf geïmputeerde datasets. Het nulmodel is het uitgangsmodel voor de modellen 1 tot en met 4.

Tabel 8.1 – Nulmodel voor de rekenvaardigheid van leerlingen, per onderwijstypegroep

	Model 0: vmbo-b/k			Model 0: vmbo gl/tl			Model 0: havo-vwo		
	$\beta$	SE	p	$\beta$	SE	p	$\beta$	SE	p
<i>Fixed effecten:</i>									
Intercept	-0.520	0.044	0.000	-0.030	0.039	0.000	0.600	0.039	0.000
<i>Random effecten:</i>									
School (niveau 2)	0.042			0.034			0.064		
Leerling (niveau 1)	0.095			0.103			0.080		
ICC	0.306			0.248			0.447		

De totale variantie in rekenvaardigheid van vmbo-b/k leerlingen wordt geschat op 0,137 (Tabel 8.1). De *intraclass correlatie* (ICC) is 0,306. Dit betekent dat 30,6% van de totale variantie in rekenvaardigheidsscores van de leerlingen ter herleiden is naar verschillen tussen scholen (variantie L2=0,042). De rest (69,4%) komt op het conto van verschillen tussen leerlingen (variantie L1=0,095).

De verschillen tussen scholen zijn het grootst voor havo-vwo leerlingen zijn (ICC=44,7%), gevolgd door vmbo-b/k leerlingen (ICC=30,6%) en tenslotte vmbo-g/t leerlingen (ICC=24,8%). Dit heeft consequenties voor de hierna volgende modellen: meer variantie vergemakkelijkt de aantoonbaarheid van relevante predictoren voor rekenvaardigheid van leerlingen.

### 8.5.2 Leerlingkenmerken

We hebben Model 0 uitgebreid met predictoren afkomstig zijn uit de leerlingvragenlijst. Deze predictoren zijn verdeeld in drie sets: domeinspecifieke leerlingkenmerken, kenmerken van het onderwijsleerproces en algemene leerlingkenmerken. Het algemene leerlingkenmerk 'Verwachte onderwijstype in leerjaar 3' hebben we opgesplitst in twee predictoren. De ordinale variabele 'Onderwijstype leerjaar 3 bekend' geeft aan welk onderwijstype de leerling volgend leerjaar denkt te volgen; van laag naar hoog is dat vmbo-b, vmbo-k, vmbo-g/t, havo, vwo. Voor 1972 leerlingen is dit bekend, maar diverse leerlingen (n=337) hebben aangegeven dat ook in leerjaar 3 nog sprake is van gemengde klassen of dat ze op het moment van invullen van de vragenlijst dit nog niet weten. Door het toevoegen van de gestandaardiseerde dummy 'Onderwijstype leerjaar 3 onbekend of gemengd' analyseren we of dit van belang is voor de samenhang met de rekenprestaties van leerlingen.

Per onderwijstypegroep werken we van een maximaal model met alle predictoren naar een spaarzaam model met alleen significante predictoren. Tabellen 8.2, 8.3 en 8.4 tonen links het maximale Model 1 en rechts het spaarzame Model 1.

## vmbo-b/k leerlingen

Uit multilevel analyses met alleen de vier kenmerken van het onderwijsleerproces van vmbo-b/k leerlingen volgt dat ordelijkheid tijdens wiskundeles en oplossingsstrategieën in de wiskundeles al 'voortijdig' zijn afgevallen. Beide kenmerken bleken niet significant geassocieerd met de rekenvaardigheid van vmbo-b/k leerlingen (respectievelijk,  $p=0.691$  en  $p=0.403$ ). Helderheid van de instructie en het toepassen van directe instructie in de wiskunde zijn dat mogelijk wel, aangezien de p-waarde voor deze coëfficiënten  $<0.10$  is (respectievelijk,  $p=0.079$  en  $p=0.043$ ). Alleen deze twee predictoren komen daarom terug in het maximale Model 1 van vmbo-b/k leerlingen. Deze werkwijze hebben we ook toegepast bij de andere twee onderwijstypegroepen. We hebben het model met de twee onderwijsleerprocespredictoren uitgebreid met de twee sets van algemene en domeinspecifieke leerlingkenmerken waardoor het maximale Model 1 is verkregen. De resultaten zijn weergegeven in Tabel 8.2 aan de linkerkant. Diverse predictoren blijken geen onafhankelijke samenhang met rekenvaardigheid te hebben, waardoor ze niet terugkomen in het spaarzame Model 1 dat na *backwards* selectie is verkregen (Tabel 8.2, rechts).

Tabel 8.2 – Model 1 voor samenhang van gestandaardiseerde algemene, domeinspecifieke en onderwijsleerproceskenmerken met de rekenvaardigheid van **vmbo-b/k** leerlingen

	Maximale Model 1			Spaarzame Model 1		
	$\beta$	SE	p	$\beta$	SE	p
<i>Fixed effecten:</i>						
Intercept	-0.444	0.055	0.000	-0.503	0.044	0.000
<b>Onderwijsleerproces*:</b>						
Ordelijkheid wiskundeles	--					
Helderheid instructie wiskundeles	0.002	0.023	0.934			
Oplossingsstrategieën in de wiskundeles	--					
Directe instructie in de wiskundeles	-0.046	0.020	0.025	-0.044	0.017	0.009
<b>Algemene leerlingkenmerken:</b>						
Jongen	0.052	0.018	0.006	0.050	0.017	0.004
Vordering (vertraagd; regulier; versneld)	0.023	0.015	0.134			
Aantal boeken thuis	0.011	0.017	0.510			
Thuis taal Nederlands (nooit; soms; bijna altijd; altijd)	0.007	0.014	0.632			
Onderwijstype leerjaar 3 bekend (vmbo-b; vmbo-k; vmbo-g/t; havo; vwo)	0.033	0.027	0.223			
Onderwijstype leerjaar 3 onbekend of gemengd	-0.016	0.016	0.313			
<b>Domeinspecifieke leerlingkenmerken:</b>						
Plezier in wiskunde	0.025	0.020	0.212			
Zelfvertrouwen in wiskunde	0.088	0.020	0.000	0.100	0.018	0.000
Belang van wiskunde	-0.018	0.022	0.419			
Attributies positieve wiskundeprestaties	0.005	0.016	0.769			
<i>Growth mindset</i> ten opzichte van wiskunde	0.048	0.019	0.011	0.048	0.018	0.007
<i>Random effecten:</i>						
School (niveau 2)	0.038			0.041		
Leerling (niveau 1)	0.080			0.079		
ICC	0.320			0.340		

\* Ordelijkheid wiskundeles en Oplossingsstrategieën in de wiskundeles zijn niet geselecteerd als relevante kenmerken van het onderwijsleerproces van vmbo-b/k leerlingen: beide  $p>0.10$ .



Uit het spaarzame Model 1 voor leerlingkenmerken volgt dat van de vier onderwijsleerproceskenmerken er één van belang is voor de vmbo-b/k leerlingen: naarmate de wiskundedocent vaker directe instructie toepast (volgens de leerlingen) zijn in het algemeen de rekenscores van vmbo-b/k leerlingen lager, een significante negatieve associatie dus. Uit de sets van algemene en domeinspecifieke leerlingkenmerken komen nog drie predictoren naar voren die een onafhankelijke significante samenhang met de rekenvaardigheid hebben. We hebben gevonden dat jongens significant hoger scoren dan meisjes, en dat meer zelfvertrouwen hebben in wiskunde en een *Growth mindset* voor wiskunde samengaan met hogere rekenscores van vmbo-b/k leerlingen.

Als we de regressiecoëfficiënten van de significante predictoren onderling vergelijken, dan kunnen we vaststellen dat van de leerlingkenmerken het zelfvertrouwen in wiskunde de belangrijkste voorspeller is voor rekenvaardigheid van vmbo-b/k leerlingen. Samen verklaren de significante leerlingkenmerken 12,4% van de totale variantie (spaarzame Model 1 vs. Model 0), met 16,8% verklaarde variantie op leerlingniveau en 2,4% op schoolniveau.

#### vmbo-g/t leerlingen

Voor vmbo-g/t leerlingen hebben we dezelfde analysestrategie toegepast. De resultaten worden getoond in Tabel 8.3.

Focussen we ons op het spaarzame Model 1 voor vmbo-g/t leerlingen, dan volgt dat van de vier onderwijsleerproceskenmerken er twee van belang zijn voor deze groep van leerlingen. Naarmate er meer oplossingsstrategieën worden aangeboden in de wiskundeles hebben de vmbo-g/t leerlingen in het algemeen enigzins *lagere* rekenscores; een negatieve verband dus. Daarentegen vonden we een positieve associatie voor directe instructie; naarmate een docent dit meer toepast zijn in het algemeen de rekenprestaties van vmbo-g/t leerlingen hoger.

Uit de sets van algemene en domeinspecifieke leerlingkenmerken komen drie kenmerken naar voren die daarnaast nog een onafhankelijke significante samenhang met rekenvaardigheid van vmbo-g/t leerlingen hebben. We hebben gevonden dat jongens hoger scoren dan meisjes, en dat meer zelfvertrouwen hebben in wiskunde en meer het belang van wiskunde inzien samengaan met hogere rekenscores van vmbo-g/t leerlingen. Van de leerlingkenmerken is zelfvertrouwen in wiskunde de belangrijkste voorspeller voor rekenvaardigheid van deze groep leerlingen (conform vmbo-b/k leerlingen). Samen verklaren de significante predictoren 13,1% van de totale variantie, met 14,6% verklaarde variantie op leerlingniveau en 8,8% op schoolniveau.

Tabel 8.3 – Model 1 voor samenhang van gestandaardiseerde algemene, domeinspecifieke en onderwijsleerproceskenmerken met de rekenvaardigheid van **vmbo-g/t** leerlingen

	Maximale Model 1			Spaarzame Model 1		
	B	SE	p	$\beta$	SE	p
<i>Fixed effecten:</i>						
Intercept	-0.009	0.041	0.824	-0.019	0.037	0.606
<b>Onderwijsleerproces*:</b>						
Ordelijkheid wiskundeles	--					
Helderheid instructie wiskundeles	0.021	0.023	0.364			
Oplossingsstrategieën in de wiskundeles	-0.049	0.016	0.003	-0.041	0.016	0.013
Directe instructie in de wiskundeles	0.025	0.018	0.173	0.034	0.015	0.031
<b>Algemene leerlingkenmerken:</b>						
Jongen	0.032	0.015	0.028	0.029	0.014	0.040
Vordering (vertraagd; regulier; versneld)	0.018	0.014	0.179			
Aantal boeken	0.004	0.015	0.798			
Thuis taal Nederlands (nooit; soms; bijna altijd; altijd)	0.024	0.013	0.072			
Onderwijstype leerjaar 3 bekend (vmbo-b; vmbo-k; vmbo-g/t; havo; vwo)	0.011	0.032	0.745			
Onderwijstype leerjaar 3 onbekend of gemengd	0.005	0.011	0.665			
<b>Domeinspecifieke leerlingkenmerken:</b>						
Plezier in wiskunde	0.009	0.019	0.631			
Zelfvertrouwen in wiskunde	0.104	0.017	0.000	0.112	0.014	0.000
Belang van wiskunde	0.028	0.016	0.084	0.036	0.015	0.015
Attributies positieve wiskunde prestaties	0.005	0.015	0.729			
<i>Growth mindset</i> ten opzichte van wiskunde	0.018	0.014	0.204			
<i>Random effecten:</i>						
School (niveau 2)	0.029			0.031		
Leerling (niveau 1)	0.088			0.088		
ICC	0.249			0.261		

\* Ordelijkheid wiskundeles is niet geselecteerd als relevant kenmerk van het onderwijsleerproces van vmbo-g/t leerlingen:  $p > 0.10$ .

### Havo-vwo leerlingen

Het spaarzame Model 1 voor havo-vwo leerlingen toont dat van de vier onderwijsleerproceskenmerken alleen de helderheid van de instructie van belang is voor de rekenvaardigheid (Tabel 8.4). Het verband is positief. Naarmate meer havo-vwo leerlingen aangeven dat ze de instructie bij de wiskundeles helder vinden, zijn de rekenprestaties in het algemeen hoger. Uit de sets van significante algemene en domeinspecifieke leerlingkenmerken kunnen we afleiden dat jongens opnieuw hoger scoren dan meisjes en dat vaker thuis Nederlands spreken, een hoger onderwijstype verwachten in leerjaar 3 en meer zelfvertrouwen hebben in wiskunde samengaan met hogere rekenscores van havo-vwo leerlingen. Daarentegen is gevonden dat de attributies voor positieve wiskunde prestaties negatief geassocieerd zijn. Dit betekent dat leerlingen die vaker een goed resultaat voor wiskunde toeschrijven aan 'goed je best doen', 'hard werken' en of 'goed opletten', oftewel een goede werkhouding, in het algemeen iets lagere rekenscores behalen.

Van de leerlingkenmerken zijn het verwachte onderwijstype in leerjaar 3 en (wederom) het zelfvertrouwen in wiskunde de belangrijkste voorspellers voor de rekenvaardigheid. De verklaarde variantie is voor

havo-vwo leerlingen relatief hoog. Samen verklaren de significante predictoren 20,1% van de totale variantie, met 13,8% verklaarde variantie op leerlingniveau en 28,1% op schoolniveau.

Tabel 8.4 – Model 1 voor samenhang van gestandaardiseerde algemene, domeinspecifieke en onderwijsleerproces kenmerken met de rekenvaardigheid van **havo-vwo** leerlingen

	Maximale Model 1			Spaarzame Model 1		
	$\beta$	SE	p	$\beta$	SE	p
<i>Fixed effecten:</i>						
Intercept	0.534	0.038	0.000	0.531	0.039	0.000
<b>Onderwijsleerproces*:</b>						
Ordelijkheid wiskundeles	--					
Helderheid instructie wiskundeles	0.029	0.014	0.035	0.028	0.012	0.021
Oplossingsstrategieën in de wiskundeles	--					
Directe instructie in de wiskundeles	-0.013	0.012	0.259			
<b>Algemene leerlingkenmerken:</b>						
Jongen	0.030	0.009	0.001	0.030	0.009	0.001
Vordering (vertraagd; regulier; versneld)	0.002	0.010	0.808			
Aantal boeken	-0.002	0.010	0.868			
Thuis taal Nederlands (nooit; soms; bijna altijd; altijd)	0.026	0.010	0.007	0.025	0.010	0.008
Onderwijstype leerjaar 3 bekend (vmbo-b; vmbo-k; vmbo-g/t; havo; vwo)	0.083	0.029	0.006	0.086	0.029	0.007
Onderwijstype leerjaar 3 onbekend of gemengd	-0.017	0.013	0.201	-0.017	0.013	0.203
<b>Domeinspecifieke leerlingkenmerken:</b>						
Plezier in wiskunde	0.015	0.012	0.229			
Zelfvertrouwen in wiskunde	0.068	0.011	0.000	0.076	0.009	0.000
Belang van wiskunde	0.007	0.011	0.513			
Attributies positieve wiskunde prestaties	-0.033	0.010	0.002	-0.031	0.010	0.002
Growth mindset ten opzichte van wiskunde	0.002	0.010	0.845			
<i>Random effecten:</i>						
School (niveau 2)	0.047			0.046		
Leerling (niveau 1)	0.069			0.069		
ICC	0.402			0.401		

\* Ordelijkheid wiskundeles en Oplossingsstrategieën in de wiskundeles zijn niet geselecteerd als relevante kenmerken van het onderwijsleerproces van havo-vwo leerlingen: beide  $p > 0.10$ .

### 8.5.3 Docentkenmerken

Om de samenhang van docentkenmerken met de rekenvaardigheid van leerlingen te analyseren hebben we Model 0 uitgebreid met sets van predictoren die afkomstig zijn uit de docentvragenlijst. De eerste set betreft docentkenmerken van het onderwijsleerproces (Model 2a), de tweede set algemene docent- en klaskenmerken (Model 2b) en de derde set domeinspecifieke docentkenmerken (Model 2c). We streven wederom naar spaarzame modellen bij iedere set zodat alleen significante predictoren worden meegenomen naar een volgend model.

Tabel 8.5 – Model 2a voor samenhang van gestandaardiseerde kenmerken van het onderwijsleerproces volgens docenten met de rekenvaardigheid van vmbo-b/k leerlingen

	Maximale Model 2a			Spaarzame Model 2a		
	$\beta$	SE	P	$\beta$	SE	p
<i>Fixed effecten:</i>						
Intercept	-3.121	2.406	0.197	-0.502	0.033	0.000
<b>Onderwijsleerproces:</b>						
Aantal minuten wiskunde per week	0.754	1.197	0.530			
Intensiteit ICT-gebruik	-0.443	0.429	0.303	-0.094	0.039	0.017
Gebruik rekenmachine tijdens de wiskundeles	0.554	0.572	0.335			
Geschiktheid wiskundemethode	0.149	0.451	0.741			
Leerlingactiviteiten tijdens de wiskundeles:						
• Leerkrachtgestuurde klassikale lesactiviteiten	1.773	1.435	0.219	0.119	0.034	0.001
• Probleemoplossen	-2.095	2.041	0.307			
1. Het zelfstandig toepassen van het geleerde bij nieuwe probleemsituaties	-2.116	2.244	0.348			
2. Werken in heterogene vaardigheidsgroepen	-0.958	0.800	0.233			
3. Werken in homogene vaardigheidsgroepen	-0.396	0.509	0.438			
4. Uitwerkingen opschrijven op het bord	0.238	0.594	0.690			
5. Uitleggen hoe ze op gegeven antwoord zijn gekomen	1.474	1.592	0.356	0.086	0.036	0.017
Differentiatie	3.059	2.845	0.284			
Omvang en bespreking huiswerk	-0.630	0.481	0.191			
Feedback huiswerk:						
6. Huiswerk corrigeren en feedback geven aan leerlingen	-2.100	2.223	0.347	-0.127	0.038	0.001
7. Huiswerk bespreken in de klas	-0.869	0.751	0.249			
8. Controleren of leerlingen huiswerk gemaakt hebben	-1.779	1.504	0.239			
9. Leerlingen kijken zelf hun huiswerk na	-1.715	1.388	0.219	-0.119	0.042	0.005
Intensiteit evaluatie	-1.355	1.096	0.218			
Gebruik typen toetsen:						
10. Genormeerde toetsen (bv. Cito)	0.827	0.644	0.201			
11. Methodegebonden toetsen	0.504	1.066	0.637			
12. Zelfgemaakte toetsen	0.821	0.736	0.266			
Directe instructiemodel	-0.734	1.003	0.466			
Ontdekkend of onderzoekend leren	0.754	0.839	0.370			
Alle leerlingen meerdere oplossingsstrategieën aanbieden	-2.931	2.586	0.259			
<i>Random effecten:</i>						
School (niveau 2)	0.045			0.017		
Leerling (niveau 1)	0.095			0.095		
ICC	0.323			0.151		

Tabel 8.5 toont de resultaten van de multilevel analyses naar de samenhang van docentkenmerken van het onderwijsleerproces met rekenvaardigheid van vmbo-b/k leerlingen (Model 2a). Het maximale model staat links, het spaarzame model rechts. De tabel laat zien dat van de grote set van 24 docentkenmerken van het onderwijsleerproces van vmbo-b/k leerlingen er slechts vijf aangemerkt kunnen worden als relevant voor hun rekenvaardigheid. De sterkte van de samenhang met rekenvaardigheid is vergelijkbaar voor deze vijf predictoren, maar de richting is verschillend. Leerkrachtgestuurde klassikale lesactiviteiten en de leerlingen laten uitleggen hoe ze op een gegeven antwoord zijn gekomen zijn in het algemeen gunstig voor de rekenvaardigheid van vmbo-b/k leerlingen. Daarentegen is gevonden dat als leraren vaker huiswerk corrigeren en feedback geven aan leerlingen, leerlingen vaker zelf hun huiswerk nakijken en ICT intensiever wordt gebruikt in de wiskundeles, de rekenvaardigheid van vmbo-b/k leerlingen in het algemeen lager is. De verklaarde totale variantie van het spaarzame Model 2a ten opzichte van Model 0 van vmbo-b/k leerlingen is 18,2%. Op schoolniveau wordt 59,5% verklaard (en op leerlingniveau 0%). De ICC is 15,1%.

In Tabel 8.6 worden de uitkomsten van het maximale en spaarzame model getoond voor de samenhang van significante onderwijsleerproceskenmerken en algemene docent- en klaskenmerken met de rekenvaardigheid van vmbo-b/k leerlingen (Model 2b).

*Tabel 8.6 – Model 2b voor samenhang van gestandaardiseerde onderwijsleerproceskenmerken, algemene docent en klaskenmerken met de rekenvaardigheid van **vmbo-b/k** leerlingen*

	Maximale Model 2b			Spaarzame Model 2b		
	$\beta$	SE	p	B	SE	p
<i>Fixed effecten:</i>						
Intercept	-0.455	0.057	0.000	-0.401	0.044	0.000
<b>Onderwijsleerproces:</b>						
Intensiteit ICT-gebruik	-0.076	0.035	0.028	-0.079	0.033	0.016
Leerlingactiviteiten tijdens de wiskundeles:						
• Leerkrachtgestuurde klassikale lesactiviteiten	0.093	0.031	0.002	0.090	0.030	0.003
• Uitleggen hoe ze op gegeven antwoord zijn gekomen	0.084	0.031	0.006	0.097	0.030	0.001
Feedback huiswerk:						
13. Huiswerk corrigeren en feedback geven aan leerlingen	-0.093	0.034	0.007	-0.115	0.032	0.000
14. Leerlingen kijken zelf hun huiswerk na	-0.089	0.036	0.015	-0.100	0.035	0.005
<b>Algemene docent- en klaskenmerken:</b>						
Man (1=ja; 0=nee)	0.060	0.060	0.314			
Lesbevoegdheid	-0.035	0.037	0.339			
Lesgegeven in wiskunde	-0.024	0.029	0.395			
Aantal leerlingen in getoetste klas	0.066	0.030	0.027	0.077	0.027	0.004
<i>Random effecten:</i>						
School (niveau 2)	0.009			0.010		
Leerling (niveau 1)	0.095			0.095		
ICC	0.088			0.092		

Tabel 8.6 laat zien dat als we het spaarzame Model 2a uitbreiden met vier algemene docent- en klaskenmerken (Model 2b), dit niet of nauwelijks van invloed is op de sterkte en richting van de samenhang van docentkenmerken van onderwijsleerproces met de rekenprestaties van vmbo-b/k leerlingen. Van de set met algemene predictoren hangt alleen het aantal leerlingen in de getoetste klas significant samen met de rekenvaardigheid: een grotere klas bij wiskunde gaat in het algemeen samen met hogere

rekenvaardheidsscores van vmbo-b/k leerlingen. De verklaarde totale variantie van het spaarzame Model 2b ten opzichte van Model 0 van vmbo-b/k leerlingen is 23,4%. Op schoolniveau wordt 76,2% van de variantie verklaard (en op leerlingniveau 0%). De ICC is 9,2%.

Model 2c is het model waarbij drie domeinspecifieke docentkenmerken zijn toegevoegd aan het spaarzame model 2b. Het resultaat van model 2c voor vmbo-b/k leerlingen is gepresenteerd in Tabel 8.7.

*Tabel 8.7 – Model 2c voor samenhang van gestandaardiseerde onderwijsleerproceskenmerken, algemene docent- en klaskenmerken en domeinspecifieke docentkenmerken met de rekenvaardigheid van vmbo-b/k leerlingen*

	Maximale Model 2c			Spaarzame Model 2c		
	$\beta$	SE	p	B	SE	p
<i>Fixed effecten:</i>						
Intercept	-0.376	0.042	0.000	-0.374	0.039	0.000
<b>Onderwijsleerproces:</b>						
Intensiteit ICT-gebruik	-0.078	0.031	0.012	-0.075	0.028	0.008
Leerlingactiviteiten tijdens de wiskundeles:						
• Leerkrachtgestuurde klassikale lesactiviteiten	0.065	0.031	0.033	0.062	0.028	0.026
• Uitleggen hoe ze op gegeven antwoord zijn gekomen	0.115	0.029	0.000	0.116	0.027	0.000
Feedback huiswerk:						
15. Huiswerk corrigeren en feedback geven aan leerlingen	-0.100	0.035	0.005	-0.100	0.028	0.000
16. Leerlingen kijken zelf hun huiswerk na	-0.087	0.035	0.013	-0.086	0.031	0.005
<b>Algemene docent- en klaskenmerken:</b>						
Aantal leerlingen in getoetste klas	0.076	0.026	0.004	0.079	0.024	0.001
<b>Domeinspecifieke docentkenmerken:</b>						
Tijdsbesteding (bij)scholing	-0.002	0.031	0.943			
Zelfvertrouwen wiskunde-instructie	-0.007	0.025	0.789			
<i>Growth mindset</i> ten opzichte van wiskunde	-0.080	0.032	0.013	-0.080	0.029	0.005
<i>Random effecten:</i>						
School (niveau 2)	0.007			0.006		
Leerling (niveau 1)	0.095			0.095		
ICC	0.070			0.055		

Als we focussen op het spaarzame model 2c, dan blijkt dat alleen een *Growth mindset* ten opzichte van wiskunde een significante negatieve associatie met rekenvaardigheid van vmbo-b/k leerlingen te hebben. Het impliceert dat naarmate leerkrachten sterker van mening zijn dat elke leerling goed kan zijn of worden in wiskunde, de rekenvaardigheidsscores van de leerlingen lager zijn. De regressiecoëfficiënten van de overige predictoren wijzigen niet of nauwelijks. De verklaarde totale variantie van het spaarzame Model 2c ten opzichte van Model 0 is 26,3% voor vmbo-b/k leerlingen. Op schoolniveau wordt 85,7% van de variantie in rekenvaardigheid van deze groep leerlingen verklaard, de ICC is 5,5%.

#### vmbo-g/t leerlingen

De resultaten van de multilevel analyses naar de samenhang van docentkenmerken van het onderwijsleerproces met de rekenvaardigheid van vmbo-g/t leerlingen zijn weergegeven in Tabel 8.8.

Tabel 8.8. – Model 2a voor samenhang van gestandaardiseerde kenmerken van het onderwijsleerproces volgens docenten met de rekenvaardigheid van **vmbo-g/t** leerlingen

	Maximale Model 2a			Spaarzame Model 2a		
	$\beta$	SE	P	$\beta$	SE	p
<i>Fixed effecten:</i>						
Intercept	-0.150	0.428	0.726	-0.030	0.039	0.441
<b>Onderwijsleerproces:</b>						
Aantal minuten wiskunde per week	-0.375	0.448	0.403			
Intensiteit ICT-gebruik	-0.293	0.360	0.416			
Gebruik rekenmachine tijdens de wiskundeles	-0.177	0.205	0.388			
Geschiktheid wiskundemethode	0.153	0.403	0.707			
Leerlingactiviteiten tijdens de wiskundeles:						
• Leerkrachtgestuurde klassikale lesactiviteiten	-0.113	0.398	0.777			
• Probleemoplossen	-0.189	0.238	0.428			
17. Het zelfstandig toepassen van het geleerde bij nieuwe probleemsituaties	0.208	0.255	0.416			
18. Werken in heterogene vaardigheidsgroepen	0.170	0.316	0.592			
19. Werken in homogene vaardigheidsgroepen	0.044	0.186	0.814			
20. Uitwerkingen opschrijven op het bord	-0.025	0.161	0.878			
21. Uitleggen hoe ze op gegeven antwoord zijn gekomen	0.139	0.272	0.609			
Differentiatie	-0.067	0.235	0.777			
Omvang en bespreking huiswerk	-0.086	0.203	0.672			
Feedback huiswerk:						
22. Huiswerk corrigeren en feedback geven aan leerlingen	0.059	0.437	0.893			
23. Huiswerk bespreken in de klas	-0.025	0.266	0.924			
24. Controleren of leerlingen huiswerk gemaakt hebben	0.140	0.209	0.505			
25. Leerlingen kijken zelf hun huiswerk na	-0.038	0.501	0.940			
Intensiteit evaluatie	0.020	0.295	0.945			
Gebruik typen toetsen:						
26. Genormeerde toetsen (bv. Cito)	0.090	0.665	0.893			
27. Methodegebonden toetsen	0.136	0.356	0.705			
28. Zelfgemaakte toetsen	-0.157	0.292	0.590			
Directe instructiemodel	0.088	0.127	0.488			
Ontdekkend of onderzoekend leren	0.119	0.476	0.804			
Alle leerlingen meerdere oplossingsstrategieën aanbieden	0.010	0.171	0.954			
<i>Random effecten:</i>						
School (niveau 2)	0.092			0.034		
Leerling (niveau 1)	0.103			0.103		
ICC	0.441			0.248		

Het spaarzame model 2a, rechts in Tabel 8.8, laat zien dat geen enkel docentkenmerk van het onderwijsleerproces significant geassocieerd is met de rekenvaardigheid van vmbo-g/t leerlingen.

In Tabel 8.9 worden de uitkomsten van het maximale en spaarzame model getoond voor de samenhang van onderwijsleerproceskenmerken en algemene docent- en klaskenmerken met de rekenvaardigheid van vmbo-g/t leerlingen (Model 2b).

Tabel 8.9 – Model 2b voor samenhang van gestandaardiseerde onderwijsleerproceskenmerken, algemene docent en klaskenmerken met de rekenvaardigheid van **vmbo-g/t** leerlingen

	Maximale Model 2b			Spaarzame Model 2b		
	$\beta$	SE	p	B	SE	p
<i>Fixed effecten:</i>						
Intercept	0.013	0.071	0.857	-0.030	0.039	0.441
<b>Onderwijsleerproces:</b>						
Geen predictor resteert	nvt			nvt		
<b>Algemene docent- en klaskenmerken:</b>						
Man	-0.081	0.086	0.344			
Lesbevoegdheid	-0.037	0.038	0.327			
Lesgegeven in wiskunde	0.057	0.046	0.215			
Aantal leerlingen in getoetste klas	0.035	0.058	0.545			
<i>Random effecten:</i>						
School (niveau 2)	0.036			0.035		
Leerling (niveau 1)	0.103			0.103		
ICC	0.259			0.251		

Tabel 8.9 toont dat als we Model 2a voor vmbo-g/t uitbreiden met vier algemene docent- en klaskenmerken, er opnieuw geen significante predictoren overblijven voor de spaarzame versie van model 2b. Dit impliceert dat we voor geen enkele docent- of klaskenmerk van vmbo-g/t leerlingen een samenhang met rekenvaardigheid kunnen aantonen.

Als laatste stap hebben we de drie domeinspecifieke docentkenmerken toegevoegd aan het spaarzame model 2a/2b van vmbo-g/t leerlingen. Hierdoor verkrijgen we Model 2c. De resultaten zijn weergegeven in Tabel 8.10. Het spaarzame Model 2c in deze tabel laat zien dat naarmate docenten aangeven meer zelfvertrouwen in de wiskunde-instructie te hebben, de rekenvaardigheid van hun vmbo-g/t leerlingen in het algemeen hoger is. Daarmee is zelfvertrouwen in de wiskunde-instructie het enige docent- of klaskenmerk dat een significante associatie heeft met de rekenvaardigheidsscores van vmbo-g/t leerlingen.

De verklaarde totale variantie van het spaarzame Model 2c ten opzichte van Model 0 is 5,1% voor vmbo-g/t leerlingen. Op schoolniveau wordt 20,6% van de variantie in rekenvaardigheidsscores verklaard. De ICC is 20,4%.



Tabel 8.10 – Model 2c voor samenhang van gestandaardiseerde onderwijsleerproceskenmerken, algemene docent- en klaskenmerken en domeinspecifieke docentkenmerken met de rekenvaardigheid van vmbo-g/t leerlingen

	Maximale Model 2c			Sparzame Model 2c		
	$\beta$	SE	p	B	SE	p
<i>Fixed effecten:</i>						
Intercept	-0.021	0.037	0.569	-0.033	0.035	0.346
<b>Onderwijsleerproces:</b>						
Geen predictor resteert	nvt			nvt		
<b>Algemene docent- en klaskenmerken:</b>						
Geen predictor resteert	nvt			nvt		
<b>Domeinspecifieke docentkenmerken:</b>						
Tijdsbesteding (bij)scholing	-0.041	0.037	0.262			
Zelfvertrouwen wiskunde-instructie	0.085	0.034	0.012	0.083	0.031	0.008
<i>Growth mindset</i> ten opzichte van wiskunde	0.001	0.039	0.979			
<i>Random effecten:</i>						
School (niveau 2)	0.027			0.027		
Leerling (niveau 1)	0.103			0.103		
ICC	0.210			0.204		

#### havo-vwo leerlingen

De samenhang van docentkenmerken van het onderwijsleerproces (Model 2a) met de rekenvaardigheid van havo-vwo leerlingen is gepresenteerd in Tabel 8.11. Uit de tabel volgt dat er voor havo-vwo leerlingen zeven docentkenmerken van het onderwijsleerproces aan te wijzen zijn die van belang zijn voor hun rekenvaardigheid. De sterkste samenhang heeft de lesactiviteit het zelfstandig toepassen van het geleerde bij nieuwe probleemsituaties ( $\beta=0.155$ ); dit is bevorderend voor de rekenvaardigheid van deze groep leerlingen. Voor de overige zes predictoren indiceren de regressiecoëfficiënten vergelijkbare sterkte, maar ze verschillen in richting van het verband. We hebben gevonden dat naarmate wiskunde-docenten vaker huiswerk opgeven en dit bespreken, vaker methodegebonden toetsen toepassen, de rekenvaardigheidsscores van havo-vwo leerlingen in het algemeen gunstiger zijn. Daar tegenover staat dat als de rekenmachine altijd mag worden gebruikt in de wiskundeles dit ongunstig is voor hun rekenvaardigheid. Hetzelfde geldt voor het vaker toepassen van leerkrachtgestuurde klassikale lesactiviteiten, meer aandacht voor probleemoplossen en het werken in homogene groepen; deze hangen negatief samen met de rekenvaardigheid van havo-vwo leerlingen. De verklaarde totale variantie van het spaarzame Model 2a ten opzichte van Model 0 van havo-vwo leerlingen is 22,2%. Op schoolniveau wordt 50,0% verklaard (en op leerlingniveau 0%). De ICC is 28,7%.

Tabel 8.11 – Model 2a voor samenhang van gestandaardiseerde kenmerken van het onderwijsleerproces volgens docenten met de rekenvaardigheid van **havo-vwo** leerlingen

	Maximale Model 2a			Sparzame Model 2a		
	$\beta$	SE	P	$\beta$	SE	p
<i>Fixed effecten:</i>						
Intercept	0.564	0.057	0.000	0.562	0.034	0.000
<b>Onderwijsleerproces:</b>						
Aantal minuten wiskunde per week	0.010	0.063	0.879			
Intensiteit ICT-gebruik	0.061	0.058	0.289			
Gebruik rekenmachine tijdens de wiskundeles	-0.046	0.069	0.507	-0.112	0.032	0.000
Geschiktheid wiskundemethode	0.066	0.054	0.229			
Leerlingactiviteiten tijdens de wiskundeles:						
• Leerkrachtgestuurde klassikale lesactiviteiten	-0.167	0.067	0.012	-0.075	0.035	0.031
• Probleemoplossen	-0.066	0.051	0.196	-0.097	0.038	0.010
• Het zelfstandig toepassen van het geleerde bij nieuwe probleemsituaties	0.131	0.063	0.039	0.155	0.042	0.000
• Werken in heterogene vaardigheidsgroepen	0.020	0.055	0.721			
• Werken in homogene vaardigheidsgroepen	-0.056	0.039	0.156	-0.093	0.031	0.002
• Uitwerkingen opschrijven op het bord	-0.051	0.043	0.240			
• Uitleggen hoe ze op gegeven antwoord zijn gekomen	0.026	0.049	0.596			
Differentiatie	-0.027	0.059	0.652			
Omvang en bespreking huiswerk	0.084	0.063	0.182	0.087	0.040	0.032
Feedback huiswerk:						
29. Huiswerk corrigeren en feedback geven aan leerlingen	0.003	0.072	0.972			
30. Huiswerk bespreken in de klas	-0.014	0.049	0.779			
31. Controleren of leerlingen huiswerk gemaakt hebben	-0.027	0.066	0.681			
32. Leerlingen kijken zelf hun huiswerk na	0.073	0.064	0.253			
Intensiteit evaluatie	-0.030	0.062	0.628			
Gebruik typen toetsen:						
33. Genormeerde toetsen (bv. Cito)	0.026	0.051	0.616			
34. Methodegebonden toetsen	0.087	0.047	0.064	0.104	0.028	0.000
35. Zelfgemaakte toetsen	0.004	0.063	0.949			
Directe instructiemodel						
Ontdekkend of onderzoekend leren	0.031	0.056	0.582			
Alle leerlingen meerdere oplossingsstrategieën aanbieden	0.014	0.047	0.766			
<i>Random effecten:</i>						
School (niveau 2)	0.037			0.032		
Leerling (niveau 1)	0.080			0.080		
ICC	0.319			0.287		

Tabel 8.12 laat de resultaten van de multilevel analyses zien voor de uitbreiding van het spaarzame model 2a van havo-vwo met vier algemene docent- en klaskenmerken.

Tabel 8.12 – Model 2b voor samenhang van gestandaardiseerde kenmerken van het onderwijsleerproces volgens docenten met de rekenvaardigheid van **havo-vwo** leerlingen

	Maximale Model 2b			Spaarzame Model 2b		
	B	SE	P	$\beta$	SE	p
<i>Fixed effecten:</i>						
Intercept	0.544	0.045	0.000	0.562	0.034	0.000
<b>Onderwijsleerproces (spaarzame model 2a):</b>						
Gebruik rekenmachine tijdens de wiskundeles	-0.108	0.035	0.002	-0.112	0.032	0.000
Leerlingactiviteiten tijdens de wiskundeles:						
• Leerkrachtgestuurde klassikale lesactiviteiten	-0.071	0.039	0.068	-0.075	0.035	0.031
• Probleemoplossen	-0.107	0.043	0.013	-0.097	0.038	0.010
36. Het zelfstandig toepassen van het geleerde bij nieuwe probleemsituaties	0.163	0.046	0.000	0.155	0.042	0.000
37. Werken in homogene vaardigheidsgroepen	-0.092	0.032	0.004	-0.093	0.031	0.002
Omvang en bespreking huiswerk	0.081	0.045	0.073	0.087	0.040	0.032
Gebruik typen toetsen:						
38. Methodegebonden toetsen	0.110	0.031	0.000	0.104	0.028	0.000
<b>Algemene docent- en klaskenmerken:</b>						
Man (1=ja; 0=nee)	0.039	0.066	0.555			
Lesbevoegdheid	-0.004	0.039	0.908			
Lesgegeven in wiskunde	-0.000	0.034	0.999			
Aantal leerlingen in getoetste klas	0.020	0.040	0.627			
<i>Random effecten:</i>						
School (niveau 2)	0.036			0.032		
Leerling (niveau 1)	0.080			0.080		
ICC	0.308			0.287		

Uit het spaarzame model 2b volgt dat uitbreiding van het Model 2a met vier algemene docent- en klaskenmerken voor havo-vwo leerlingen, niet leidt tot meer docentkenmerken die samenhangen met de rekenvaardigheid van deze groep leerlingen (Tabel 8.12). Dit impliceert dat we voor geen enkele algemeen docent- of klaskenmerk van havo-vwo leerlingen een significante samenhang met hun rekenvaardigheidsscores hebben kunnen aantonen.

Tabel 8.13 laat de resultaten van de multilevel analyses zien voor de uitbreiding van het spaarzame model 2b van havo-vwo leerlingen met drie domeinspecifieke docentkenmerken.

Tabel 8.13 – Model 2c voor samenhang van gestandaardiseerde kenmerken van het onderwijsleerproces volgens docenten en domeinspecifieke docentkenmerken met de rekenvaardigheid van **havo-vwo** leerlingen

	Maximale Model 2c			Spaarzame Model 2c		
	$\beta$	SE	P	$\beta$	SE	p
<i>Fixed effecten:</i>						
Intercept	0.570	0.037	0.000	0.562	0.034	0.000
<b>Onderwijsleerproces (spaarzame model 2a):</b>						
Gebruik rekenmachine tijdens de wiskundeles	-0.118	0.033	0.000	-0.112	0.032	0.000
Leerlingactiviteiten tijdens de wiskundeles:						
• Leerkrachtgestuurde klassikale lesactiviteiten	-0.063	0.040	0.113	-0.075	0.035	0.031
• Probleemoplossen	-0.106	0.040	0.009	-0.097	0.038	0.010
39. Zelfstandig toepassen van het geleerde bij nieuwe probleemsituaties	0.155	0.043	0.000	0.155	0.042	0.000
40. Werken in homogene vaardigheidsgroepen	-0.091	0.032	0.004	-0.093	0.031	0.002
Omvang en bespreking huiswerk	0.068	0.047	0.152	0.087	0.040	0.032
Gebruik typen toetsen:						
41. Methodegebonden toetsen	0.102	0.030	0.001	0.104	0.028	0.000
<b>Algemene docent- en klaskenmerken:</b>						
Geen predictoren resteren	nvt					
<b>Domeinspecifieke docentkenmerken:</b>						
Tijdsbesteding (bij)scholing	-0.011	0.035	0.764			
Zelfvertrouwen wiskunde-instructie	0.048	0.038	0.203			
Growth mindset ten opzichte van wiskunde	-0.023	0.032	0.468			
<i>Random effecten:</i>						
School (niveau 2)	0.033			0.032		
Leerling (niveau 1)	0.080			0.080		
ICC	0.294			0.287		

Uit het spaarzame model 2c volgt dat geen enkele domeinspecifieke docentkenmerk significant samenhangt met de rekenvaardigheid van havo-vwo leerlingen (Tabel 8.13). Daarmee komt het spaarzame model 2c overeen met het spaarzame model 2b en model 2a voor deze onderwijstypengroep. We kunnen dus alleen voor kenmerken van het onderwijsleerproces een significante samenhang met de rekenvaardigheid van havo-vwo leerlingen vinden. Voor de verklaarde variantie en ICC van Model 2c verwijzen we naar Model 2a.

#### 8.5.4 Schoolkenmerken

De samenhang van schoolkenmerken met de algemene rekenvaardigheid van leerlingen is geanalyseerd door Model 0 uit te breiden met twee sets van predictoren die op schoolniveau zijn gemeten. De eerste set zijn algemene schoolkenmerken (Model 3a), de tweede set domeinspecifieke schoolkenmerken (Model 3b). Ook nu werken we toe naar spaarzame modellen.

## vmbo-b/k leerlingen

Tabel 8.14 toont de resultaten van de multilevel analyses met de eerste set van schoolkenmerken voor vmbo-b/k leerlingen: de algemene schoolkenmerken (Model 3a).

Tabel 8.14 – Model 3a voor samenhang van gestandaardiseerde algemene schoolkenmerken met de rekenvaardigheid van **vmbo-b/k** leerlingen

	Maximale Model 3a			Spaarzame Model 3a		
	$\beta$	SE	p	$\beta$	SE	p
<i>Fixed effecten:</i>						
Intercept	-0.406	0.076	0.000	-0.434	0.052	0.000
<b>Algemene schoolkenmerken:</b>						
Vestigingsgrootte	0.050	0.079	0.525			
Speciaal onderwijsconcept	-0.057	0.198	0.777			
Aandeel leerlingen afkomstig uit een economisch achterstandsgezin	-0.031	0.060	0.609			
Aandeel leerlingen voor wie Nederlands <i>niet</i> eerste taal is	-0.007	0.069	0.922			
Stedelijkheid	-0.013	0.071	0.856			
Prestatiegerichte schoolklimaat docenten	0.028	0.082	0.733			
Prestatiegerichte schoolklimaat ouders	0.126	0.062	0.042	0.118	0.044	0.008
Prestatiegerichte schoolklimaat leerlingen	-0.115	0.088	0.192			
Belang leerlingvolgsysteem	0.094	0.048	0.052			
<i>Random effecten:</i>						
School (niveau 2)	0.025			0.030		
Leerling (niveau 1)	0.095			0.095		
ICC	0.210			0.241		

Het spaarzame Model 3a indiceert dat van de negen algemene schoolkenmerken er slechts één relevant is voor de rekenprestaties. Het gaat om het prestatiegerichte schoolklimaat van ouders: naarmate ouders meer prestatiegericht zijn, zijn ook in het algemeen de rekenvaardigheidsscores van vmbo-b/k leerlingen hoger. De verklaarde totale variantie van het spaarzame Model 3a ten opzichte van Model 0 van vmbo-b/k leerlingen is 8,8%. Op schoolniveau wordt 28,6% van de variantie verklaard; de bijbehorende ICC is 24,1%.

De volgende stap is het spaarzame Model 3a uitbreiden met domeinspecifieke predictoren op schoolniveau. De resultaten van dit model, Model 3b, zijn gepresenteerd in Tabel 8.15.

Tabel 8.15 – Model 3b voor samenhang van gestandaardiseerde algemene en domeinspecifieke schoolkenmerken met de rekenvaardigheid van **vmbo-b/k** leerlingen

	Maximale Model 3b			Spaarzame Model 3b		
	$\beta$	SE	p	$\beta$	SE	p
<i>Fixed effecten:</i>						
Intercept	-0.431	0.071	0.000	-0.434	0.052	0.000
<b>Algemene schoolkenmerken:</b>						
Prestatiegerichte schoolklimaat ouders	0.089	0.052	0.094	0.118	0.044	0.008
<b>Domeinspecifieke schoolkenmerken:</b>						
Extra wiskundelessen buiten schooltijd	0.006	0.052	0.904			
Bevordering interesse in wiskunde	0.039	0.028	0.159			
Aantal aangeboden professionaliserings-activiteiten voor wiskundeleraars	-0.030	0.026	0.261			
<i>Random effecten:</i>						
School (niveau 2)	0.025			0.030		
Leerling (niveau 1)	0.095			0.095		
ICC	0.210			0.241		

Het toevoegen van drie domeinspecifieke schoolkenmerken resulteert niet in aanvullende significante predictoren voor de rekenvaardigheid van vmbo-b/k leerlingen (Tabel 8.15). Het spaarzame model 3b komt geheel overeen met het spaarzame model 3a voor deze onderwijstypengroep. Het impliceert dat de prestatiegerichtheid van ouders het enige schoolkenmerk is waarvoor we een significante associatie met de rekenvaardigheid van vmbo-b/k leerlingen hebben kunnen aantonen. Voor de verklaarde variatie en ICC van Model 3b verwijzen we naar Model 3a.

#### vmbo-g/t leerlingen

Tabel 8.16 geeft de uitkomsten van de multilevel analyses voor de samenhang van algemene schoolkenmerken met de rekenvaardigheid van vmbo-g/t leerlingen (Model 3a).

Uit het spaarzame Model 3a (Tabel 8.16) volgt dat geen van de negen algemene schoolkenmerken significant gerelateerd is aan de rekenvaardigheid van vmbo-g/t leerlingen. Dit betekent dat het spaarzame Model 3a geheel overeen komt met het nulmodel van deze groep leerlingen.

Tabel 8.16 – Model 3a voor samenhang van gestandaardiseerde algemene schoolkenmerken met de rekenvaardigheid van **vmbo-g/t** leerlingen

	Maximale Model 3a			Spaarzame Model 3a		
	$\beta$	SE	p	$\beta$	SE	p
<i>Fixed effecten:</i>						
Intercept	-0.035	0.055	0.520	-0.030	0.039	0.441
<b>Algemene schoolkenmerken:</b>						
Vestigingsgrootte	-0.007	0.061	0.912			
Speciaal onderwijsconcept	0.087	0.167	0.602			
Aandeel leerlingen afkomstig uit een economisch achterstandsgezin	-0.080	0.054	0.144			
Aandeel leerlingen voor wie Nederlands <i>niet</i> eerste taal is	-0.061	0.066	0.358			
Stedelijkheid	-0.033	0.064	0.609			
Prestatiegerichte schoolklimaat docenten	0.062	0.058	0.285			
Prestatiegerichte schoolklimaat ouders	-0.068	0.070	0.332			
Prestatiegerichte schoolklimaat leerlingen	-0.026	0.039	0.503			
Belang leerlingvolgsysteem	0.022	0.048	0.654			
<i>Random effecten:</i>						
School (niveau 2)	0.034			0.034		
Leerling (niveau 1)	0.103			0.103		
ICC	0.247			0.248		

De volgende stap is het spaarzame Model 3a uitbreiden met domeinspecifieke predictoren op schoolniveau. Zo ontstaat Model 3b voor vmbo-g/t leerlingen (Tabel 8.17).

Tabel 8.17 – Model 3b voor samenhang van gestandaardiseerde algemene en domeinspecifieke schoolkenmerken met de rekenvaardigheid van **vmbo-g/t** leerlingen

	Maximale Model 3b			Spaarzame Model 3b		
	$\beta$	SE	p	$\beta$	SE	p
<i>Fixed effecten:</i>						
Intercept	-0.431	0.071	0.000	-0.030	0.039	0.441
<b>Algemene schoolkenmerken:</b>						
Geen predictor resteert	nvt					
<b>Domeinspecifieke schoolkenmerken:</b>						
Extra wiskundelessen buiten schooltijd	0.006	0.052	0.904			
Bevordering interesse in wiskunde	0.039	0.028	0.159			
Aantal aangeboden professionaliseringsactiviteiten voor wiskundedocenten	-0.030	0.026	0.261			
<i>Random effecten:</i>						
School (niveau 2)	0.025			0.034		
Leerling (niveau 1)	0.095			0.103		
ICC	0.210			0.248		

Het spaarzame Model 3b (Tabel 8.17) impliceert dat geen enkel algemeen of domeinspecifiek schoolkenmerk significant gerelateerd is aan de rekenvaardigheid van vmbo-g/t leerlingen. Daarmee is spaarzame Model 3b geheel gelijk aan het nulmodel van deze groep leerlingen.

#### havo-vwo leerlingen

De resultaten van de multilevel analyses voor de samenhang van algemene schoolkenmerken met de rekenvaardigheid van havo-vwo leerlingen (Model 3a) zijn gepresenteerd in Tabel 8.18.

Tabel 8.18 – Model 3a voor samenhang van gestandaardiseerde algemene schoolkenmerken met de rekenvaardigheid van **havo-vwo** leerlingen

	Maximale Model 3a			Spaarzame Model 3a		
	$\beta$	SE	p	$\beta$	SE	p
<i>Fixed effecten:</i>						
Intercept	0.590	0.053	0.000	0.577	0.038	0.000
<b>Algemene schoolkenmerken:</b>						
Vestigingsgrootte	-0.004	0.052	0.931			
Speciaal onderwijsconcept	-0.154	0.133	0.262			
Aandeel leerlingen afkomstig uit een economisch achterstandsgezin	-0.086	0.059	0.145			
Aandeel leerlingen voor wie Nederlands <i>niet</i> eerste taal is	0.025	0.057	0.658			
Stedelijkheid	-0.014	0.050	0.780			
Prestatiegerichte schoolklimaat docenten	-0.010	0.041	0.800			
Prestatiegerichte schoolklimaat ouders	0.080	0.062	0.202	0.099	0.042	0.022
Prestatiegerichte schoolklimaat leerlingen	-0.009	0.055	0.874			
Belang leerlingvolgsysteem	0.096	0.040	0.019	0.091	0.041	0.031
<i>Random effecten:</i>						
School (niveau 2)	0.047			0.047		
Leerling (niveau 1)	0.080			0.080		
ICC	0.371			0.372		

Het spaarzame Model 3a toont dat zowel de prestatiegerichte schoolklimaat ouders als het belang van het leerlingvolgsysteem relevant zijn voor de rekenvaardigheidsscores van havo-vwo leerlingen (Tabel 8.18). Beide regressiecoëfficiënten zijn positief. Als ouders meer prestatiegericht zijn en als de school het leerlingvolgsysteem voor meer monitor- en communicatiedoelen inzet, dan zijn in het algemeen de rekenvaardigheidsscores van havo-vwo leerlingen hoger. De verklaarde totale variantie van het spaarzame Model 3a ten opzichte van Model 0 van havo-vwo leerlingen is 11,8%. Op schoolniveau wordt 26,6% van de variantie verklaard; de bijbehorende ICC is 37,2%.

Als we het spaarzame Model 3a uitbreiden met domeinspecifieke predictoren op schoolniveau dan verkrijgen we Model 3b. De resultaten van de multilevel analyses van Model 3b voor havo-vwo leerlingen worden hieronder getoond (Tabel 8.19).



Tabel 8.19 – Model 3b voor samenhang van gestandaardiseerde algemene en domeinspecifieke schoolkenmerken met de rekenvaardigheid van **havo-vwo** leerlingen

	Maximale Model 3b			Spaarzame Model 3b		
	$\beta$	SE	p	$\beta$	SE	p
<i>Fixed effecten:</i>						
Intercept	0.587	0.061	0.000	0.577	0.038	0.000
<b>Algemene schoolkenmerken:</b>						
Prestatiegerichte schoolklimaat ouders	0.126	0.048	0.011	0.099	0.042	0.022
Belang leerlingvolgsysteem	0.091	0.041	0.031	0.091	0.041	0.031
<b>Domeinspecifieke schoolkenmerken:</b>						
Extra wiskundelessen buiten schooltijd	-0.012	0.078	0.875			
Bevordering interesse in wiskunde	0.016	0.050	0.745			
Aantal aangeboden professionaliseringsactiviteiten voor wiskundedocenten	-0.066	0.046	0.147			
<i>Random effecten:</i>						
School (niveau 2)	0.047			0.047		
Leerling (niveau 1)	0.080			0.080		
ICC	0.371			0.372		

Het spaarzame Model 3b maakt duidelijk dat geen van drie domeinspecifieke schoolkenmerken relevant is voor de rekenvaardigheid van havo-vwo leerlingen (Tabel 8.19). Het model komt geheel overeen met het spaarzame Model 3a. Samenvattend kunnen we stellen dat de prestatiegerichtheid van ouders en het belang dat de school hecht aan het gebruik van het leerlingvolgsysteem, de enige twee schoolkenmerken zijn die significant gerelateerd zijn aan de rekenbaarheid van havo-vwo leerlingen. Voor de verklaarde variantie en ICC van Model 3b zie Model 3a.

### 8.5.5 Eindmodel

Tot slot hebben we de predictoren uit de spaarzame modellen voor leerlingkenmerken (Model 1) voor docent- en klaskenmerken (Model 2c) en schoolkenmerken (Model 3b) samengevoegd voor de constructie van Model 4, het eindmodel.

De resultaten zijn gepresenteerd in Tabel 8.20 voor vmbo-b/k leerlingen, in Tabel 8.21 voor vmbo-g/t leerlingen en in Tabel 8.22 voor havo-vwo leerlingen. Opgemerkt moet worden dat de gevonden significante effectgroottes in het algemeen zeer klein zijn (< 0.10).

Tabel 8.20 – Model 4 voor kenmerken van leerlingen, docent, klas en school die in voorafgaande spaarzame modellen significant samenhangen met de rekenvaardigheid van **vmbo-b/k** leerlingen

	Maximale Model 4			Spaarzame Model 4		
	B	SE	p	B	SE	p
<i>Fixed effecten:</i>						
Intercept	-0.387	0.049	0.000	-0.377	0.043	0.000
<b>Leerlingkenmerken (Model 1):</b>						
Directe instructie in de wiskundeles	-0.037	0.017	0.029	-0.037	0.017	0.026
Jongen	0.046	0.017	0.008	0.045	0.017	0.009
Zelfvertrouwen in wiskunde	0.099	0.017	0.000	0.099	0.017	0.000
<i>Growth mindset</i> ten opzichte van wiskunde	0.052	0.018	0.004	0.053	0.018	0.003
<b>Docent/klaskenmerken (Model 2):</b>						
Intensiteit ICT gebruik	-0.076	0.038	0.046	-0.066	0.032	0.040
Leerlingactiviteiten tijdens de wiskundeles:						
• Leerkrachtgestuurde klassikale lesactiviteiten	0.100	0.037	0.008	0.092	0.030	0.002
• Uitleggen hoe ze op gegeven antwoord zijn gekomen	0.084	0.031	0.006	0.081	0.029	0.006
Feedback huiswerk:						
• Huiswerk corrigeren en feedback geven aan leerlingen	-0.120	0.039	0.002	-0.109	0.031	0.001
• Leerlingen kijken zelf hun huiswerk na	-0.107	0.040	0.008	-0.099	0.035	0.005
Aantal leerlingen in getoetste klas	0.089	0.028	0.001	0.086	0.026	0.001
<b>Schoolkenmerken (Model 3):</b>						
Prestatiegerichte schoolklimaat ouders	-0.022	0.044	0.617			
<i>Random effecten:</i>						
School (niveau 2)	0.010			0.010		
Leerling (niveau 1)	0.080			0.079		
ICC	0.112			0.109		

Het spaarzame Model 4 voor vmbo-b/k leerlingen – gepresenteerd in Tabel 8.20 – laat zien dan van alle significante leerling-, docent/klas- en schoolkenmerken uit de voorafgaande modellen, alleen prestatiegericht schoolklimaat van ouders niet langer een eigen significante associatie heeft met rekenvaardigheid. Verder zijn er nauwelijks veranderingen in de regressiecoëfficiënten van de predictoren geconstateerd.

Gelet op de grootte van de gestandaardiseerde regressiecoëfficiënten kunnen we stellen dat van de leerlingkenmerken het zelfvertrouwen in wiskunde de belangrijkste predictor is, en dat van de docent/klaskenmerken vooral de manier waarop feedback op het huiswerk gegeven wordt een aandachtspunt is bij vmbo-b/k leerlingen. Zowel het huiswerk corrigeren en van feedback voorzien door de docent, als het zelf nakijken van het huiswerk door de leerlingen zijn *niet* bevorderlijk voor hun rekenvaardigheid. Echter, leerkrachtgestuurde klassikale lesactiviteiten (b.v. luisteren naar de uitleg, regels en procedures uit het hoofd leren, werken aan opgaven in de klas) en het laten uitleggen hoe ze op het antwoord komen zijn wel bevorderlijk voor hun rekenvaardigheid. De resultaten indiceren dat de rekenvaardigheid van vmbo-b/k leerlingen het meest gebaat is bij directe sturing en intensieve begeleiding in de klas door de leerkracht. De verklaarde totale variantie van het spaarzame eindmodel is 35,0%. Van de variantie op schoolniveau wordt 76,2% verklaard, en van de variantie op leerlingniveau 16,8%.

Tabel 8.21 laat de resultaten van het maximale en spaarzame eindmodel van vmbo-g/t leerlingen zien.

Tabel 8.21 – Model 4 voor kenmerken van leerlingen, docent, klas en school die in voorafgaande spaarzame modellen significant samenhangen met de rekenvaardigheid van **vmbo-g/t** leerlingen

	Maximale Model 4			Spaarzame Model 4		
	B	SE	p	B	SE	p
<i>Fixed effecten:</i>						
Intercept	-0.019	0.037	0.606	-0.019	0.037	0.606
<b>Leerlingkenmerken (Model 1):</b>						
Oplossingsstrategieën in de wiskundeles	-0.041	0.016	0.013	-0.041	0.016	0.013
Directe instructie in de wiskundeles	0.034	0.015	0.031	0.034	0.015	0.031
Jongen	0.029	0.014	0.040	0.029	0.014	0.040
Zelfvertrouwen in wiskunde	0.112	0.014	0.000	0.112	0.014	0.000
Belang van wiskunde	0.036	0.015	0.015	0.036	0.015	0.015
<b>Docent/klaskenmerken (Model 2):</b>						
Geen predictor resteert	nvt			nvt		
<b>Schoolkenmerken (Model 3):</b>						
Geen predictor resteert	nvt			nvt		
<i>Random effecten:</i>						
School (niveau 2)	0.031			0.031		
Leerling (niveau 1)	0.088			0.088		
ICC	0.261			0.261		

Het spaarzame Model 4 voor vmbo-g/t leerlingen blijkt gelijk te zijn aan het spaarzame Model 1 (Tabel 8.3). Dit impliceert dat er geen docent-, klas- of schoolkenmerken zijn gevonden die samenhangen met de rekenvaardigheid van deze groep leerlingen. Het belangrijkste leerlingkenmerk is het zelfvertrouwen in wiskunde. Het bevorderen van het zelfvertrouwen in wiskunde bij vmbo-g/t leerlingen zou een manier kunnen zijn om tot verbetering van de rekenvaardigheid te komen, maar we kunnen ‘*reversed causality*’ hier niet uitsluiten: leerlingen die goed zijn in wiskunde hebben in het algemeen meer zelfvertrouwen. De verklaarde totale variantie van het spaarzame Model 4 is 13,1%. Van de variantie op schoolniveau wordt 8,8% verklaard, en van de variantie op leerlingniveau 14,6%.

Het spaarzame finale model impliceert dat we voor vmbo-g/t leerlingen geen directe handvatten hebben gevonden waarmee de wiskundedocent of de school gericht aan de slag kunnen gaan om de rekenvaardigheid te verbeteren. Mogelijk zijn er wel indirecte handvatten. Naarmate leerlingen vaker aangeven dat directe instructie wordt gegeven tijdens de wiskundeles en ze minder oplossingsstrategieën krijgen aangeboden of zelf moeten bedenken, zijn hun rekenvaardigheidsscores in het algemeen hoger. Dit duidt erop dat vmbo-g/t leerlingen gebaat zouden kunnen zijn bij duidelijke gestructureerde en niet (te) complexe wiskundelessen.

In Tabel 8.22 worden de resultaten van de twee finale modellen voor havo-vwo leerlingen getoond.

Tabel 8.22 – Model 4 voor kenmerken van leerlingen, docent, klas en school die in voorafgaande spaarzame modellen significant samenhangen met de rekenvaardigheid van **havo-vwo** leerlingen

	Maximale Model 4			Spaarzame Model 4		
	B	SE	p	B	SE	p
<i>Fixed effecten:</i>						
Intercept	0.511	0.034	0.000	0.501	0.033	0.000
<b>Leerlingkenmerken (Model 1):</b>						
Helderheid instructie wiskundeles	0.030	0.012	0.012	0.030	0.012	0.012
Jongen	0.030	0.009	0.001	0.030	0.009	0.001
Thuis taal Nederlands (nooit; soms; bijna altijd; altijd)	0.026	0.010	0.007	0.026	0.010	0.007
Onderwijstypen leerjaar 3 bekend (vmbo-b; vmbo-k; vmbo-g/t; havo; vwo)	0.079	0.026	0.004	0.082	0.027	0.004
Zelfvertrouwen in wiskunde	0.077	0.010	0.000	0.077	0.010	0.000
Attributies positieve wiskunde prestaties	-0.031	0.010	0.002	-0.031	0.010	0.002
<b>Docent/klaskenmerken (Model 2):</b>						
Gebruik rekenmachine tijdens de wiskundeles (niet altijd; altijd)	-0.079	0.036	0.029	-0.102	0.027	0.000
Leerlingactiviteiten tijdens de wiskundeles:						
• Leerkrachtgestuurde klassikale lesactiviteiten	-0.055	0.032	0.085	-0.068	0.030	0.023
• Probleemoplossen	-0.086	0.034	0.012	-0.080	0.033	0.015
• Het zelfstandig toepassen van het geleerde bij nieuwe probleemsituaties	0.128	0.039	0.001	0.132	0.036	0.000
• Werken in homogene vaardigheidsgroepen	-0.080	0.028	0.005	-0.076	0.026	0.004
Omvang en bespreking huiswerk	0.070	0.036	0.056	0.078	0.035	0.024
Gebruik typen toetsen:						
• Methodegebonden toetsen	0.081	0.026	0.002	0.089	0.024	0.000
<b>Schoolkenmerken (Model 3):</b>						
Prestatiegerichte schoolklimaat ouders	0.003	0.032	0.932			
Belang leerlingvolgsysteem	0.043	0.036	0.246			
<i>Random effecten:</i>						
School (niveau 2)	0.022			0.023		
Leerling (niveau 1)	0.069			0.069		
ICC	0.244			0.247		

Het spaarzame Model 4 voor havo-vwo leerlingen laat zien dat van alle voorheen significante predictoren van leerling-, docent/klas- en schoolkenmerken, alleen de twee schoolkenmerken niet meer significant samenhangen met rekenvaardigheid (Tabel 8.22). Er zijn bovendien slechts minimale verschillen in regressiecoëfficiënten geconstateerd ten opzichte van de spaarzame Modellen 1 en 2c. Dit betekent dat qua kenmerken van havo-vwo leerlingen het zelfvertrouwen in wiskunde en het verwachte onderwijstype in leerjaar 3 de belangrijkste predictoren voor de rekenvaardigheid blijven. Wat betreft de docent/klaskenmerken hebben we gevonden dat met name het zelfstandig toepassen van het geleerde bij nieuwe probleemsituaties bevorderlijk kan zijn voor de rekenvaardigheid van havo-vwo leerlingen. Ook positief, maar minder sterk, zijn het huiswerk opgeven en bespreken, en het gebruik van methodegebonden toetsen door de wiskundedocent. Hier staat tegenover dat het altijd mogen gebruiken van de

rekenmachine *niet* bevorderlijk is voor de rekenvaardigheid van havo-vwo leerlingen. Dit laatste geldt ook voor het toepassen van leerkrachtgestuurde klassikale lesactiviteiten, probleemoplossen en werken in homogene groepen als lesactiviteit. De verklaarde totale variantie van het spaarzame Model 4 van havo-vwo leerlingen is 36,1%. Op schoolniveau wordt 64,1% verklaard en op leerlingniveau 13,8%.

Het spaarzame eindmodel impliceert dat we voor havo-vwo leerlingen enkele concrete handvatten hebben gevonden waarmee de wiskundedocent of de school gericht aan de slag kunnen gaan om de rekenvaardigheid te verbeteren. Met uitzondering van het minder gebruik van een rekenmachine tijdens de les zijn de resultaten voor havo-vwo leerlingen lastig te duiden. Het lijkt erop dat deze groep van leerlingen gebaat is bij zelfstandige en uitdagende werkvormen in de klas (*minder* leerkrachtgestuurde klassikale lesactiviteiten, het zelfstandig toepassen van het geleerde bij nieuwe probleemsituaties, werken in niet-homogene groepen), extra oefening (huiswerk) en monitoring (methodegebonden toetsen). Echter deze groep lijkt minder profijt te hebben van het oefenen met oplossen van (nieuwe) wiskundige problemen voor hun rekenvaardigheid.

## 9 Samenvatting en beantwoording onderzoeksvragen

### 9.1 Inleiding en context

Peilonderwijs is het geheel van periodieke peilingsonderzoeken in het primair onderwijs, voortgezet onderwijs en (voortgezet) speciaal onderwijs. De Inspectie van het Onderwijs voert de regie over de peilingsonderzoeken in opdracht van het ministerie van Onderwijs, Cultuur en Wetenschap (OCW). In dit rapport doen we verslag van het eerste peilingsonderzoek in het voortgezet onderwijs, te weten Rekenen-Wiskunde in het VO, dat is uitgevoerd door een consortium van onderzoekers van het Expertisecentrum Nederlands, Cito, de Universiteit Twente en KBA Nijmegen.

De dataverzameling voor dit onderzoek vond plaats in het voorjaar van 2022. Doordat het de eerste peiling in het VO betrof en de aanloop (kwartier maken en aanbestedingsprocedure) relatief veel tijd in beslag nam, was er slechts een voorbereidingstijd van vijf maanden voor deze dataverzameling beschikbaar. Daarom is er (conform de aanbesteding) geen pretest uitgevoerd voor de toets en vragenlijsten. In plaats daarvan is er gekozen voor beproefde toetsopgaven en vragensets uit eerder uitgevoerd peilingsonderzoek. In **Hoofdstuk 2** is deze instrumentontwikkeling beschreven.

In overleg met de Inspectie is in het peilingsonderzoek Rekenen-Wiskunde in het VO gekozen voor een getrapte steekproef waarbij eerst een steekproef van schoolvestigingen werd getrokken door de Inspectie, waarna in tweede instantie door het consortium één tweede klas werd geselecteerd die wiskunde volgde binnen de getrokken en geworven vestiging. In **Hoofdstuk 3** wordt deze steekproefaanpak beschreven. Ook het toetsafnamedesign is in Hoofdstuk 3 toegelicht, dat wil zeggen de wijze waarop de rekenopgaven zijn verdeeld over verschillende toetsversies die vervolgens zijn toegewezen aan leerlingen op basis van hun verwachte rekenniveau om daarmee hun rekenvaardigheid te meten.

In **Hoofdstuk 4** wordt uit de doeken gedaan hoe de procedure is verlopen rond de werving van scholen en de dataverzameling bij leerlingen in de geselecteerde tweede klas, hun wiskundedocent en de schoolleider. Samengevat hebben er 130 scholen voor voortgezet onderwijs meegewerkt aan het onderzoek. Op deze scholen waren 133 tweede klassen met daarin 2724 leerlingen betrokken bij de afname van de toets en leerlingenvragenlijst. Deze leerlingensteekproef is representatief voor de landelijke populatie van leerlingen in de tweede klas van het VO. Daarnaast vulden 100 wiskundedocenten en 101 schoolleiders op deze scholen de vragenlijst in.

**Hoofdstuk 5 gaat over** de schaling van de rekenvaardigheid aan de hand van de resultaten van de afgenomen rekenopgaven. We beschrijven eerst de resultaten van een klassieke toets- en itemanalyse en (de omgang met) de ontbrekende waarden. Met behulp van Item Respons Theorie (IRT-equivalering) is vastgesteld dat de rekenvaardigheid van de getoetste populatie aan de hand van een enkele schaal rekenen kan worden uitgedrukt. Op deze schaal zijn tot slot de cesuurpunten bepaald voor de indeling naar behaald referentieniveau.

### 9.2 Beantwoording onderzoeksvragen

De centrale probleemstelling van dit peilingsonderzoek is: *Welke kennis en vaardigheden hebben leerlingen uit het tweede leerjaar van het voortgezet onderwijs op het inhoudsgebied Rekenen-Wiskunde*

*en in hoeverre hangt dit samen met algemene en domeinspecifieke kenmerken van leerlingen, docenten, scholen en het onderwijsleerproces?*

Deze probleemstelling is uitgewerkt in zeven onderzoeksvragen. De basis voor de beantwoording daarvan is te vinden in de beschreven onderzoeksresultaten in **Hoofdstuk 6, 7 en 8**. Hierna formuleren we per onderzoeksvraag (in de volgorde waarin ze in dit rapport zijn beantwoord) de gevonden antwoorden en vatten daarmee tevens de betreffende hoofdstukken samen.

### 9.2.1 Het onderwijsleerproces op de scholen (Hoofdstuk 6)

In Hoofdstuk 6 zijn de resultaten van de leerling-, docent- en schoolleidersvragenlijsten gepresenteerd. Onderdeel daarvan zijn de kenmerken van het onderwijsleerproces voor het vakgebied Rekenen-wiskunde (hierna: wiskunde<sup>56</sup>) volgens de leerlingen, docenten en schoolleiders. We vatten de belangrijkste uitkomsten daarvan puntsgewijs samen en beantwoorden daarmee onderzoeksvraag 5.

*Onderzoeksvraag 5: Hoe kunnen de deelnemende scholen worden getypeerd naar kenmerken van het onderwijsleerproces op het gebied van Rekenen-Wiskunde?*

#### Kenmerken van het onderwijsleerproces op klasniveau

In onderstaande tekst verwijst de term 'alle klassen' naar alle getoetste klassen in deze peiling (één klas per school/vesting) waarvan de wiskundedocent de docentvragenlijst heeft ingevuld. Er wordt een onderscheid gemaakt in drie typen klassen: 1) pro en vmbo-b/k, 2) vmbo-g/t 3) havo/vwo.

#### Tijdsbesteding wiskunde

- Gemiddeld genomen wordt in het tweede leerjaar 2,6 uur per week lesgegeven in wiskunde. In de pro en vmbo-b/k klassen wordt de meeste tijd aan wiskundelessen besteed (gemiddeld 2,8 uur per week), in vmbo-g/t klassen relatief de minste tijd (gemiddeld 2,5 uur per week).
- Van de vier rekendomeinen gaat relatief de meeste lestijd op aan het domein Verbanden en Formules (voor alle klassen gemiddeld 30% van de totale lestijd). In pro en vmbo wordt echter relatief de meeste tijd besteed aan Meten en Meetkunde (pro en vmbo-b/k 25% van de lestijd en vmbo-g/t 28% van de lestijd). In havo/vwo wordt relatief de meeste lestijd besteed aan Verbanden en Formules (37% van de lestijd).
- In het tweede leerjaar krijgen havo/vwo leerlingen ten opzichte van de overige leerlingen het meeste huiswerk voor wiskunde; in 91% van havo/vwo klassen krijgen leerlingen na (bijna) elke les huiswerk en in slechts 2% van de klassen krijgen leerlingen **nooit** huiswerk. Het huiswerk kost volgens de wiskundedocenten meestal minder dan een half uur (63%) of tussen een half uur en een uur (37%).
- Zowel de frequentie als de hoeveelheid wiskundehuiswerk is in pro en vmbo-b/k klassen aanmerkelijk lager. Zo krijgt 68% van deze leerlingen na (bijna) elke les huiswerk en krijgt in 11% van de klassen de leerlingen **nooit** huiswerk mee. Voor de leerlingen in pro en vmbo-b/k die wel wiskundehuiswerk krijgen, kost dit meestal niet meer dan een half uur (84%), slechts 16% van de wiskundedocenten in dit onderwijstype geeft huiswerk op waarmee de leerlingen tussen een half uur tot een uur bezig zijn.

---

<sup>56</sup> In de docent- en schoolvragenlijst hadden de domeinspecifieke vragen betrekking op 'wiskunde' (met de toelichting dat daarmee het hele vakgebied Rekenen-Wiskunde wordt bedoeld). Ook in de leerlingvragenlijst wordt gesproken over 'wiskunde', alleen in de verkorte vragenlijst voor pro-leerlingen gaan de vragen over 'rekenen'.

### Instructiekenmerken

- In alle onderwijstypen wordt de methode *Getal en Ruimte* het meest gebruikt (58% van alle klassen), gevolgd door *Moderne Wiskunde* (33%). De verschillen tussen de onderwijstypen zijn klein. Alleen in vmbo-g/t klassen wordt *Moderne Wiskunde* relatief vaker gebruikt dan in de overige klassen (39%).
- De wiskundedocenten zijn over het algemeen tevreden met de gebruikte methode. Deze past goed bij de eigen kennis en vaardigheden, instructiemethode, eigen opvattingen over wiskundeonderwijs en leerlingen in hun klas. Hierin zijn geen noemenswaardige verschillen tussen de onderwijstypen.
- In bijna driekwart (74%) van de getoetste klassen kunnen leerlingen tijdens de wiskundeles gebruik maken van digitale apparaten (computer, smartphone of tablet). Als leerlingen hiervan gebruik maken is dit vooral voor het oefenen van opgaven en procedures. Het gebruik van digitale apparaten tijdens de wiskundeles ligt het hoogst in pro en vmbo-b/k klassen (86%) en het laagst in vmbo-g/t klassen (68%).
- In 43% van alle klassen wordt het gebruik van de rekenmachine **beperkt** toegestaan, in meer dan de helft van de klassen (56%) is het gebruik van de rekenmachine tijdens de wiskundelessen **altijd** toegestaan. De verschillen tussen de onderwijstypen zijn groot. In vmbo-t/l wordt het gebruik van een rekenmachine in 86% van de klassen **altijd** toegestaan, in havo/vwo is dit slechts 32% en in pro en vmbo- b/k 68%. Alleen voor pro en vmbo-b/k geldt dat in enkele klassen (4%) de rekenmachine tijdens de wiskundeles **nooit** mag worden gebruikt.
- Het 'directe instructiemodel' wordt volgens de wiskundedocenten veel toegepast in het tweede leerjaar. Dit model wordt relatief het meest toegepast in pro en vmbo-b/k (86% in (bijna) alle lessen) en relatief het minst in havo/vwo (75% in (bijna) alle lessen). Geen enkele docent in dit onderzoek heeft aangegeven dit model nooit toe te passen.
- Dit laatste geldt niet voor 'ontdekkend/onderzoekend leren'; een vijfde van de wiskundedocenten past dit model nooit toe en slechts 4% (bijna) elke les. In iets meer dan de helft van alle klassen (52%) wordt dit model in sommige lessen toegepast.
- Een kleine meerderheid van de wiskundedocenten (56%) biedt aan **alle** leerlingen **meerdere** oplossingsstrategieën aan. In pro en vmbo-b/k klassen ligt dit percentage het laagst (43%), in havo/vwo klassen het hoogst (64%). Verder valt op dat meerdere oplossingsstrategieën zelden **alleen** aan **zwak** presterende leerlingen worden aangeboden (1%). In 13% van alle klassen worden meerdere oplossingsstrategieën voor wiskundeopgaven **alleen** aan **goed** presterende leerlingen aangeboden. Wiskundedocenten die meestal maar één oplossingsstrategie aanbieden, ongeacht het niveau van de leerling, zijn eveneens in de minderheid (13%).
- Aan de wiskundedocenten zijn twee leerlingactiviteiten voorgelegd over probleemoplossen: 1) na te denken over verschillende oplossingsmanieren voor een wiskundig probleem en 2) wiskundige problemen op te lossen waarbij niet direct duidelijk is hoe de oplossing gevonden kan worden. Deze activiteiten zijn samengenomen in het construct 'probleemoplossen'. Het blijkt dat de frequentie waarin deze activiteiten van leerlingen gevraagd worden (1=nooit of bijna nooit, 4= alle lessen of bijna alle lessen) verschilt tussen de onderwijstypen. In pro en vmbo-b/k klassen (gemiddeld 2,1) wordt dit minder vaak aan leerlingen gevraagd dan aan vmbo-g/t (2,4) en havo/vwo (gemiddeld 2,6).
- Verschillende stellingen over het toepassen van vormen van differentiatie in het onderwijs van de docent zijn samengenomen in het construct 'Differentiatie' (1=niet, 2=enigszins, 3=zeer). Gemiddeld genomen worden de vormen van differentiatie in de klassen 'enigszins' toegepast (gemiddelde 2,1). In pro en vmbo-b/k (gemiddelde 2,3) worden er vaker vormen van differentiatie toegepast dan in de overige twee onderwijstypen (gemiddelde 2,1).

### Evaluatie leerlingprestaties

- De prestaties van leerlingen in de getoetste klassen worden vooral beoordeeld op basis van observaties door de docent van de leerlingen als zij aan het werk zijn, door vragen te stellen aan leerlingen tijdens de les en op basis van methodegebonden toetsen aan het eind van een periode of blok. Wiskundedocenten maken gemiddeld genomen nauwelijks gebruik van genormeerde toetsen zoals bijvoorbeeld citotoetsen. In pro en vmbo-k/b klassen worden genormeerde toetsen iets vaker gebruikt dan in de overige twee onderwijstypen.



- Aan de wiskundedocenten zijn verschillende stellingen (1= zeer mee oneens, 4=mee eens) voorgelegd over een *Fixed mindset* of een *Growth mindset* ten aanzien van leerlingprestaties in wiskunde. Een *Fixed mindset* betekent dat de docent er vanuit gaat dat wiskundeprestaties vooral met aanleg te maken hebben terwijl een *Growth mindset* betekent dat de docent gelooft dat elke leerling goed kan worden in wiskunde. De wiskundedocenten in het tweede leerjaar onderschrijven vooral de stellingen over de *Growth mindset* en nauwelijks de stellingen over een *Fixed mindset*, hierin zijn geen verschillen tussen de opleidingstypen.

#### Kenmerken onderwijsleerproces op schoolniveau

Onderstaande uitkomsten zijn gebaseerd op de antwoorden van de schoolleider van de school/vestiging waar de leerlingen zijn getoetst. Het onderwijsaanbod van de school/vestiging is ingedeeld in drie categorieën : 1) pro en/of vmbo, 2) breed (verschillende combinaties van pro tot vwo) en 3) havo en/of vwo.

#### Onderwijsklimaat

- Schoolleiders van de deelnemende scholen/vestigingen hebben aan de hand van stellingen een inschatting gegeven van de prestatiegerichtheid van het onderwijsklimaat op hun school (1=erg laag, 5=erg hoog). De voorgelegde stellingen vormen 3 constructen, namelijk de prestatiegerichtheid van de docenten van de school/vestiging (bijvoorbeeld hun verwachtingen over de leerresultaten van leerlingen); de prestatiegerichtheid van de ouders (bijvoorbeeld verwachtingen van de leerresultaten van het kind of ondersteuning door ouders bij het leren van het kind) en de prestatiegerichtheid van leerlingen (bijvoorbeeld de wil van leerlingen om goed te presteren). Tussen de verschillende onderwijstypen van de scholen zijn er geen noemenswaardige verschillen voor de prestatiegerichtheid van docenten. De prestatiegerichtheid van ouders (volgens de schoolleider) ligt echter op pro- en/of vmbo-scholen lager (gemiddeld 2,8) dan op scholen met een breed onderwijsaanbod (gemiddeld 3,2) en havo/vwo scholen (gemiddeld 3,5). De prestatiegerichtheid van de leerlingen ligt op pro- en/of vmbo-scholen ook iets lager dan de overige scholen, maar de verschillen zijn kleiner (gemiddeld 3,0, 3,2 en 3,4 respectievelijk) dan bij de ouders.

#### Bevordering interesse voor wiskunde

- Aan schoolleiders zijn stellingen voorgelegd over het bevorderen van de interesse in wiskunde op school, zoals het bieden van extra wiskundelessen of andere wiskundeactiviteiten voor leerlingen (1=zeer mee oneens, 4=zeer mee eens). Op pro- en/of vmbo-scholen wordt er minder aandacht besteed aan het bevorderen van de interesse in wiskunde (gemiddeld 2,3) dan op de overige scholen (beide onderwijstypen (gemiddeld 2,8)).

### 9.2.2 De prestaties van de leerlingen (Hoofdstuk 7)

In Hoofdstuk 7 zijn de resultaten van de toetsafname beschreven en geanalyseerd. Daarmee wordt antwoord gegeven op de eerste twee onderzoeksvragen, naar de rekenvaardigheid van de onderzochte leerlingen. We vatten de bevindingen hierna samen.

*Onderzoeksvraag 1: Welk percentage van de leerlingen einde leerjaar 2 van praktijkonderwijs, vmbo-b/k, vmbo-g/t en havo/vwo beheerst in welke mate de referentieniveaus voor Rekenen-Wiskunde?*

De meeste leerlingen in het tweede leerjaar VO behalen een vaardigheidsniveau tussen 1F en 2F, zoals was verwacht (zie ook Hoofdstuk 2). De verschillen in behaalde referentieniveaus tussen de vier onderscheiden VO-deelpopulaties (de indeling is gebaseerd op het verwachte onderwijsniveau in leerjaar 3) zijn eveneens grotendeels conform verwachting, als volgt:

- Van de **leerlingen in praktijkonderwijs** behaalt 99% het 1F-niveau niet.
- Van de **vmbo-b/k leerlingen** behaalt 73% het 1F-niveau niet. Een kwart van de vmbo-bb/kb leerlingen haalt het 1F-niveau wel. De overige 2 procent haalt 2F- of 1S-niveau. Niveau 3F wordt – zoals verwacht – niet door vmbo-leerlingen gehaald.
- Van de **vmbo-g/t** leerlingen behaalt ruim de helft (55%) het 1F-niveau, 8% behaalt niveau 2F en 12% behaalt niveau 1S.
- De helft van de **havo-leerlingen** en driekwart van de **vwo-leerlingen** beheerst in het tweede leerjaar het 1S-niveau. Het 3F-niveau wordt alleen door leerlingen in havo (1%) of vwo (12%) bereikt.

We verwijzen naar Tabel 7.3 en 7.4 voor meer informatie. Overigens ligt het gewenste uitstroomniveau aan het einde van het vmbo op 2F, al zijn leerlingen ook in mbo-2 en mbo-3 nog in de gelegenheid om het algemeen maatschappelijk niveau (2F) te bereiken. Tegelijkertijd is voor deze leerlingen 1F wel het gewenste niveau aan het einde van groep 8. In het PO haalt 20% niveau 1F niet, terwijl dit percentage in het VO op 25% ligt. Het is niet bekend van de leerlingen in dit onderzoek of zij op het moment van doorstromen naar het VO wel het 1F-niveau hadden. Desalniettemin is gemiddeld genomen het rekenvaardigheidsniveau van PO naar VO niet gedaald. Er zijn in vergelijking met het PO in leerjaar 2 van het VO ook meer leerlingen die niveau 1S of hoger behalen (39% tegenover 32%, zie Tabel 7.2).

*Onderzoeksvraag 2: Hoe kunnen de prestaties van leerlingen op de referentieniveaus worden getypeerd naar relevante (sub)domeinen die binnen de referentieniveaus Rekenen-Wiskunde onderscheiden worden?*

In Hoofdstuk 7 is per domein gekeken naar de moeilijkheid van de items en de kans die leerlingen uit de deelpopulatie hebben om een bepaald item goed te maken. Hierbij moet opgemerkt worden dat de moeilijkheid van de items wordt geschat in relatie tot de overige items die zijn meegenomen in het onderzoek. Dat wil zeggen dat er op basis van de uitgelichte voorbeeldopgaven geen generaliseerbare conclusies getrokken kunnen worden over bepaalde opgaven typen of deelvaardigheden. In de typering is *kwalitatief* gekeken naar de opgavenkenmerken om op basis daarvan uitspraken te kunnen doen over wat leerlingen van een bepaald vaardigheidsniveau wel of niet beheersen.

Domeinoverstijgend kan gezegd worden dat leerlingen op 1S-niveau (vooral havo/vwo leerlingen) een betere beheersing hebben van complexere opgaven waarin meerdere bewerkingen uitgevoerd moeten worden. Desalniettemin blijken leerlingen met 1F-niveau ook de 2F-opgave uit het domein Getallen en bewerkingen te beheersen waarin meerdere geldbedragen opgeteld moeten worden. In het domein Getallen en bewerkingen en het domein Verhoudingen blijken moeilijkere opgaven veelal te gaan over breuken en kommagetallen. Leerlingen met 1F-niveau of hoger hebben wel een matig tot goede beheersing van de 1F-opgave waarin gevraagd wordt kommagetallen te positioneren op de getallenlijn. Voor leerlingen met 2F- of 1S-niveau geldt dat zij nog moeite hebben met de 2F-opgaven (met of zonder rekenmachine) waarin negatieve getallen afgetrokken moeten worden (zie Figuur 7.7). Dit is overigens geen verrassing aangezien ze dit 2F-doel pas aan het einde van het vmbo hoeven te beheersen. Ongeveer de helft van de leerlingen uit leerjaar twee beheerst dit doel matig tot goed. In het domein Meten en meetkunde viel op dat vooral meetkundeopgaven (zie bijvoorbeeld Figuur 7.13) een laag onderscheidend vermogen hebben: een hogere vaardigheid betekent niet één op één een hogere kans om het item goed te beantwoorden. Voor wat betreft Meten zien we ook hier dat de opgaven met berekeningen met kommagetallen relatief moeilijker zijn dan opgaven met gehele getallen. In het domein Verbanden is te zien dat het aflezen van grafieken en tabellen door de meeste leerlingen met 1F-niveau of hoger wordt beheerst, terwijl het gebruiken van informatie uit tabellen en grafieken voor (complexe) berekeningen veelal nog moeilijk is voor leerlingen met 1S-niveau (zie Figuur 7.18). In het domein Verbanden

zagen we ook dat de drie gemakkelijkste 3F-opgaven over patronen en regelmaat gaan en dat deze matig worden beheerst door leerlingen met 1S-niveau. Vergelijkbaar met Meetkunde is voor Wiskundig probleemoplossen te zien dat de meeste opgaven een laag onderscheidend vermogen hebben. Daarnaast zijn er weinig opgaven die door leerlingen met 1F-, 2F- of 1S-niveau worden beheerst. Kortom, wiskundige problemen – waarvoor geen routineaanpak geldt – worden door de meeste leerlingen in leerjaar 2 nog niet beheerst.

### 9.2.3 Samenhang tussen prestaties en kenmerken van leerlingen, docenten en het onderwijsleerproces (Hoofdstuk 8)

In Hoofdstuk 8 is aan de hand van multilevel analyses de samenhang tussen de leerlingprestaties en kenmerken van leerlingen, docenten en het onderwijsleerproces onderzocht. Leerlingen in praktijkonderwijs zijn hier buiten beschouwing gelaten, omdat het aantal pro-leerlingen daarvoor te klein was. Hierna vatten we de bevindingen samen en beantwoorden daarmee de onderzoeksvragen 3, 4, 6 en 7. In de voorafgaande paragrafen zijn de onderzoeksvragen 5, 1 en 2 al beantwoord.

*Onderzoeksvraag 3: Welke verschillen in leerlingprestaties Rekenen-Wiskunde zijn er naar algemene en domeinspecifieke leerlingkenmerken en welk deel van de verschillen kunnen door leerlingkenmerken worden verklaard?*

*Onderzoeksvraag 4: Welke verschillen in leerlingprestaties Rekenen-Wiskunde zijn er naar kenmerken van het onderwijsleerproces op het gebied van Rekenen-Wiskunde en welk deel van de verschillen in leerlingprestaties Rekenen-Wiskunde wordt verklaard door variantie in scores op kenmerken van het onderwijsleerproces (uitgevoerd curriculum)?*

Met behulp van het spaarzame multilevel Model 1 en Model 2a (zie Hoofdstuk 8) kunnen we deze vragen gecombineerd beantwoorden:

- Verschillen in rekenvaardigheid tussen leerlingen zijn deels te verklaren door **algemene leerlingkenmerken** (Model 1). Het gaat ten eerste om het geslacht van de leerling. Jongens behalen in het algemene hogere rekenvaardigheidsscores dan meisjes en dit geldt voor alle onderwijstypen. Alleen voor havo-vwo leerlingen zijn nog twee aanvullende relevante predictoren gevonden. Naarmate havo-vwo leerlingen vaker thuis Nederlands spreken en verwachten dat ze een hoger onderwijstype gaan volgen komend leerjaar zijn over het algemeen hun rekenvaardigheidsscores hoger.
- Wat betreft de **domeinspecifieke leerlingkenmerken** hebben we gevonden dat vier predictoren 'er toe doen' (Model 1). Allereerst is gevonden dat meer zelfvertrouwen hebben in wiskunde gunstig is voor de rekenvaardigheid bij alle onderwijstypen. Verder is het beeld wisselend. Een sterkere *Growth mindset* voor wiskunde hangt samen met een hogere rekenvaardigheid van vmbo-b/k leerlingen. Meer het belang van wiskunde inzien is positief geassocieerd met de rekenvaardigheid van vmbo-g/t leerlingen. Daarentegen gaat een goede werkhouding als reden voor goede wiskunde-prestaties ('attributies positieve wiskunde-prestaties') in het algemeen samen met lagere rekenvaardigheidsscores van havo-vwo leerlingen.
- Voor de set van **leerlingkenmerken van het onderwijsleerproces** is het beeld van de drie onderwijstypegroepen divers (Model 1). Voor vmbo-b/k leerlingen geldt dat meer directe instructie tijdens de wiskundeles samengaat met enigszins lagere rekenvaardigheidsscores, terwijl voor vmbo-g/t leerlingen het omgekeerde is gevonden. Voor deze laatste groep geldt tevens dat naarmate er meer oplossingsstrategieën worden aangeboden, de rekenvaardigheid lager is. Voor havo-vwo leerlingen is alleen de helderheid van de instructie van belang: naarmate leerlingen hier positiever over zijn, is de rekenvaardigheid in het algemeen hoger.
- De set van 24 **docentkenmerken van het onderwijsleerproces** resulteert ook in een uiteenlopend beeld voor de drie onderwijstypegroepen (Model 2a). Voor vmbo-b/k leerlingen springen drie

negatieve associaties in het oog. We hebben gevonden dat naarmate ICT vaker wordt gebruikt in de wiskundeles, de leerkracht het huiswerk vaker corrigeert en van feedback voorziet en de leerlingen hun eigen huiswerk nakijken, hun rekenprestaties in het algemeen lager zijn. De overige verbanden zijn positief voor deze groep. Als er meer leerkrachtgestuurde klassikale lesactiviteiten zijn en vmbo-b/k leerlingen vaker uitleggen hoe ze op het antwoord zijn gekomen, zijn in het algemeen de rekenvaardigheidsscores hoger. Bij vmbo-g/t leerlingen hangt geen enkel docentkenmerk van het onderwijsleerproces significant samen met hun rekenvaardigheid. Voor havo-vwo leerlingen zijn wel relevante kenmerken gevonden, maar deze overlappen inhoudelijk nauwelijks met die van vmbo-b/k leerlingen. Als havo-vwo leerlingen altijd de rekenmachine mogen gebruiken, vaker in homogene vaardigheidsgroepen werken, vaker klassikale lesactiviteiten hebben of met probleemoplossen bezig zijn, dan zijn hun rekenprestaties in het algemeen lager. Daar staat tegenover dat naarmate havo-vwo leerlingen vaker op zelfstandige wijze het geleerde toepassen bij nieuwe probleemsituaties, er meer huiswerk wordt gegeven en besproken, en meer methodegebonden toetsen worden gebruikt, de rekenvaardigheidsscores hoger zijn. Dit alles duidt op een sterke verschillen in effectiviteit van onderwijsleerproceskenmerken tussen de drie onderwijstypegroepen. Voor leerkrachtgestuurde klassikale lesactiviteiten is dit zelfs tegengesteld: het werkt voor vmbo-b/k leerlingen, maar niet havo-vwo leerlingen.

- De drie typen leerlingkenmerken verklaren gezamenlijk 12, 13 en 20% van de totale variantie in de algemene rekenvaardigheid van respectievelijk vmbo-b/k leerlingen, vmbo tl/gl leerlingen en havo-vwo leerlingen. Voor de docentkenmerken van het onderwijsleerproces zijn deze percentages respectievelijk 18%, 0% en 22%.

*Onderzoeksvraag 6. Welke verschillen in leerlingprestaties Rekenen-Wiskunde zijn er naar algemene en domeinspecifieke docentkenmerken en algemene schoolkenmerken en welk deel van de verschillen in leerlingprestaties Rekenen-Wiskunde wordt verklaard door variantie in scores op algemene en domeinspecifieke docentkenmerken, klaskenmerken en algemene schoolkenmerken?*

Met de resultaten van de spaarzame Modellen 2b, 2c en 3a kunnen we antwoord geven op deze onderzoeksvraag, onder verwijzing naar de tabellen in Hoofdstuk 8:

- Het belang van **algemene docent- en klaskenmerken** als voorspeller voor de rekenvaardigheid van de leerlingen is gering als we tevens corrigeren voor docentkenmerken van het onderwijsleerproces (Model 2b). Voor vmbo-b/k leerlingen is gevonden dat slechts één van de vier algemene docent- en klaskenmerken significant gerelateerd is aan rekenvaardigheid: naarmate er meer leerlingen in de wiskundeklas zitten, zijn de rekenprestaties hoger. Voor zowel vmbo-g/t als havo-vwo leerlingen blijkt dat geen enkele algemene docent- en klaskenmerken geassocieerd met rekenvaardigheid van leerlingen.
- Voor de drie **domeinspecifieke docentkenmerken** kunnen we stellen dat deze van wisselende relevantie zijn, als we tevens rekening houden met onderwijsleerproceskenmerken en algemene docent- en klaskenmerken (Model 2c). Een *Growth mindset* van de docent ten opzichte van wiskunde hangt negatief samen met de rekenvaardigheid van vmbo-b/k leerlingen. Meer zelfvertrouwen van de docent over de wiskunde-instructie hangt positief samen met de rekenvaardigheid van vmbo-g/t leerlingen. Voor havo-vwo leerlingen hangt geen enkel domeinspecifiek docentkenmerk significant samen met de rekenvaardigheidsscores. Model 2c verklaart 26% van de totale variantie in rekenvaardigheidsscores voor vmbo-b/k leerlingen, 5% voor vmbo tl/gl leerlingen en 22% voor havo-vwo leerlingen.
- Voor **algemene schoolkenmerken** hebben we gevonden dat twee van de negen predictoren relevant zijn voor de rekenvaardigheid van leerlingen (Model 3a). Voor zowel vmbo-b/k als havo-vwo leerlingen geldt dat naarmate ouders vaker een prestatiegericht schoolklimaat ondersteunen, de scores voor rekenen-wiskunde hoger zijn. Voor havo-vwo leerlingen geldt daarnaast ook dat als het

leerlingvolgsysteem vaker en veelzijdiger gebruikt wordt op hun school, de rekenprestaties van de leerlingen in het algemeen hoger zijn. Bij vmbo-g/t is geen enkel algemeen schoolkenmerk van belang. De twee algemene schoolkenmerken verklaren gezamenlijk 9% en 12% van de totale variatie in de rekenvaardigheid van respectievelijk vmbo-b/k en havo-vwo leerlingen.

*Onderzoeksvraag 7. Welke kenmerken van leerlingen, docenten, scholen en het onderwijsleerproces op het gebied van Rekenen-Wiskunde zijn in onderlinge samenhang mogelijke verklarende variabelen voor verschillen in leerlingprestaties?*

Het spaarzame Model 4 – het eindmodel – (zie Hoofdstuk 8) geeft antwoord op deze vraag:

- Welke **leerlingkenmerken van het onderwijsleerproces** geassocieerd zijn met de rekenvaardigheid van leerlingen is afhankelijk van het onderwijstype. De richting van het verband wisselt: voor vmbo-b/k leerlingen is het leerlingkenmerk directe instructie in de wiskundeles negatief geassocieerd met rekenvaardigheid, voor vmbo-g/t is deze samenhang echter positief. Voor havo-vwo leerlingen is niet de directe instructie van belang maar de ervaren helderheid van de instructie; deze is positief geassocieerd met hun rekenvaardigheid.
- Welke **docent/schoolkenmerken van het onderwijsleerproces** verband houden met de rekenvaardigheid van leerlingen is ook sterk wisselend per onderwijstype. Voor vmbo-b/k leerlingen zijn intensiever ICT-gebruik, huiswerk corrigeren en van feedback voorzien, leerlingen zelf hun huiswerk laten nakijken niet bevorderlijk voor de rekenvaardigheid, maar de leerkracht gestuurde klassikale lesactiviteiten, en het laten uitleggen hoe leerlingen op het antwoord komen daarentegen wel. Voor vmbo-g/t leerlingen is geen enkele predictor uit deze set significant geassocieerd met hun rekenvaardigheid. Voor havo-vwo leerlingen hebben we zeven significante predictoren gevonden. Negatief geassocieerd met de rekenvaardigheid van havo-vwo leerlingen zijn het altijd mogen gebruiken van de rekenmachine, leerkrachtgestuurde klassikale lesactiviteiten, probleemoplossen en het werken in homogene groepen als lesactiviteit; positief geassocieerd zijn het zelfstandig toepassen van het geleerde bij nieuwe probleemsituaties, huiswerk opgeven en bespreken, en het gebruik van methodegebonden toetsen door de wiskundedocent.
- Wat betreft **de algemene en domeinspecifieke leerlingkenmerken** hebben we voor alle onderwijstypen gevonden dat jongens hogere scores voor rekenvaardigheid behalen dan meisjes en dat meer zelfvertrouwen in wiskunde samengaat met hogere rekenvaardigheidsscores. Daarnaast zijn er ook weer opvallende verschillen tussen de onderwijstypen naar voren gekomen. Voor vmbo-b/k leerlingen is een *Growth mindset* hebben voor wiskunde en meer leerlingen in de wiskundeklas bevorderlijk voor hun rekenvaardigheid. Voor vmbo-g/t leerlingen is het belang inzien van wiskunde gunstig, maar diversiteit aan oplossingsstrategieën in de wiskundeles is ongunstig voor de rekenvaardigheid. Voor havo-vwo leerlingen hangen vaker thuis Nederlands spreken en een hoger onderwijstype verwachten voor het aankomende schooljaar (leerjaar 3) positief samen met rekenvaardigheid, maar het attribueren van positieve wiskundeprestaties aan een goede werkhouding hangt juist negatief samen.
- De spaarzame eindmodellen van de drie onderwijstypen bevatten geen **algemene of domeinspecifieke schoolkenmerken** die onafhankelijk van de overige predictoren samenhangen met de rekenvaardigheid van de leerlingen. Dergelijke schoolkenmerken vormen dus geen relevante aanvulling op kenmerken van het onderwijsleerproces en algemene en domeinspecifieke leerlingkenmerken als het gaat om de samenhang met rekenvaardigheid van leerlingen van alle onderwijstypen.

Afhankelijk van het onderwijstype komen we tot enkele (voorzichtige) conclusies over factoren die de rekenvaardigheid van de leerlingen mogelijk kunnen bevorderen:

- Vmbo-b/k leerlingen lijken het meest gebaat bij directe sturing en intensieve begeleiding in de klas door de wiskundeleerkracht.

- Voor vmbo-g/t leerlingen zou het bevorderen van het zelfvertrouwen in wiskunde een manier kunnen zijn om tot verbetering van de rekenvaardigheid te komen. Ook zouden zij gebaat kunnen zijn bij duidelijke gestructureerde en niet (te) complexe wiskundelessen.
- Havo-vwo leerlingen lijken het meest gebaat bij zelfstandige en uitdagende werkvormen in de klas (minder leerkrachtgestuurde klassikale lesactiviteiten, het zelfstandig toepassen van het geleerde bij nieuwe probleemsituaties, werken in niet-homogene groepen), extra oefening (huiswerk) en monitoring (methodegebonden toetsen). Deze groep lijkt minder profijt te hebben van het oefenen met oplossen van (nieuwe) wiskundige problemen voor hun rekenvaardigheid.



## Literatuur

- Bock, R. D., and Mislevy, R. J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6, 431-444.  
<https://doi.org/10.1177/014662168200600405>
- Buuren, S. van, & Groothuis-Oudshoorn, C. G. M. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of statistical software*, 45(3). <http://www.jstatsoft.org/v45/i03>
- Buisman, M., Kuijper, S., Hickendorff, M., Kuijpers, R.E., Keuning, J., Walet, L., Meijer, J., Ankersmit, M. & Kronenburg, F. van (2021). Rekenvaardigheden in het (S)BO. *Technisch rapport Peil.onderwijs Rekenen-Wiskunde einde (speciaal) basisonderwijs 2019*. Amsterdam: Kohnstamm Instituut.
- Engen, T. J. H. M., and Verhelst, N. D. (2011). Item calibration in incomplete testing designs. *Psicologica: International Journal of Methodology and Experimental Psychology*, 32, 107–132.
- Evers, E., Lucassen, W., Meijer R.R., and Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. NIP.
- Groenestijn, M. van, Borghouts, C., Janssen, C. (2011). *Protocol Ernstige Reken-Wiskunde problemen en Dyscalculie BAO, SBO, SO*. Van Gorcum.
- Gubbels, J., van Langen, A., Maassen, N. & Meelissen, M. (2019). *Resultaten PISA-2018 in vogelvlucht*. Universiteit Twente. <https://doi.org/10.3990/1.9789036549226>
- Haelermans, C., Jacobs, M., Smeets, C., & van Vugt, L. (2021). *Factsheets leergroei na anderhalf jaar covid19-crisis: Technische toelichting*. Nationaal Cohortonderzoek Onderwijs.
- Hemker, B.T, Kordes, J. en Van Weerden, J.J. (2010). *Peiling van de rekenvaardigheid en de taalvaardigheid in jaargroep 8 en jaargroep 4 in 2010. Jaarlijks Peilingsonderzoek van het Onderwijsniveau - Technische rapportage*. Arnhem: Cito.
- Hemker, B.T. (2022). To a or not to a. In: *Practical Measurement* (Eds. Van der Ark, L.A, Emons, W. & Meijer, R.R.). Cham: Springer-Verlag.
- Inspectie van het Onderwijs (2018). *De Staat van het Onderwijs 2018. Onderwijsverslag over 2016/2017*. Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2020a). *De Staat van het Onderwijs 2020. Onderwijsverslag over 2018/2019*. Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2020b). *COVID-19-monitor: voortgezet onderwijs (derde meting)*. Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2021). *De Staat van het Onderwijs 2021. Onderwijsverslag over 2019/2020*. Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2022a). *De Staat van het Onderwijs 2022. Onderwijsverslag over 2020/2021*. Inspectie van het Onderwijs.
- Inspectie van het Onderwijs (2022b). *Technisch rapport Covid-19*. Utrecht: Inspectie van het Onderwijs.
- Kolen, M.J. & Brennan, R.L. (2004). *Test Equating, Scaling, and Linking: Methods and Practices* (2nd ed.). New York: Springer.
- Lord, F.M. (1980). *Applications of Item Response Theory To Practical Testing Problems*. New York: Routledge.
- Mullis, I. V. S. & Martin, M. O. (Eds.). (2017). *TIMSS 2019 Assessment Frameworks*. Boston College, TIMSS & PIRLS International Study Center. Retrieved from <http://tims-sandpirls.bc.edu/timss2019/frameworks/>
- Sjoers, S, & Schmidt, V. (2021). *Rekenen en wiskunde onderbouw VO. Domeinbeschrijving ten behoeve van peilingsonderzoek*. SLO.
- Slob, A. (2021). *Brief aan scholen over Nationaal Programma Onderwijs*. Geraadpleegd van <https://www.nponderwijs.nl/binaries/nationaalprogrammaonderwijs/documenten/publicaties/2021/>

- 03/23/brief-aan-scholen-over-nationaal-programma-onderwijs/brief-aan-scholen-over-nationaal-programma-onderwijs.pdf
- Snijders, T., & Bosker, R. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publishers.
- Verhelst, N.D., & Glas, C.A.W. (1993). A dynamic generalization of the Rasch model. *Psychometrika*, 58, 395–415.
- Verhelst, N.D., Glas, C.A.W., & Verstralen, H.H.F.M. (1993). *OPLM: One parameter logistic model. Computer program and manual*. Cito.
- Verhelst, N. D. (2008). *CORDIM with Box-Cox transformations*. [interne notitie]. Cito.
- Wiersma, D. (2022). *Kamerbrief Masterplan basisvaardigheden*. Kamerstukken, 2021-2022, 31293 nr. 624. Geraadpleegd van [https://www.rijksoverheid.nl/documenten/kamerstukken/2022/04/04/curriculum-funderend-onderwijs#:~:text=Minister%20Wiersma%20\(Primair%20en%20Voortgezet,\(basisonderwijs%20en%20voortgezet%20onderwijs\)](https://www.rijksoverheid.nl/documenten/kamerstukken/2022/04/04/curriculum-funderend-onderwijs#:~:text=Minister%20Wiersma%20(Primair%20en%20Voortgezet,(basisonderwijs%20en%20voortgezet%20onderwijs).).



# Bijlagen

## Bijlage 1 – Leerlingvragenlijst regulier VO: rechte tellingen

Tabel B1.1 – Plezier in wiskunde

Plezier in wiskunde	Zeer mee eens	Beetje mee eens	Beetje mee oneens	Zeer mee oneens	Missing	Totaal
j. Ik vind het leuk om wiskunde te leren	7,6%	31,2%	27,4%	33,5%	0,3%	2606
k. Ik zou willen dat ik geen wiskunde hoefde te leren	36,0%	30,0%	20,9%	11,9%	1,3%	2606
l. Wiskunde is saai	30,0%	36,3%	21,9%	10,1%	1,7%	2606
m. Ik leer veel interessante dingen tijdens de wiskundelessen	9,3%	32,7%	33,7%	23,1%	1,3%	2606
n. Ik vind wiskunde leuk	9,1%	25,4%	30,0%	34,3%	1,1%	2606
o. Ik vind alle schooltaken leuk waarbij ik met getallen moet werken	4,1%	17,2%	39,8%	38,0%	1,0%	2606
p. Ik vind het leuk om wiskundeopgaven op te lossen	6,8%	23,8%	31,4%	37,5%	0,6%	2606
q. Ik verheug mij op de wiskundeles	4,0%	14,9%	31,5%	48,6%	1,1%	2606
r. Wiskunde is één van mijn favoriete vakken	8,4%	15,2%	21,0%	54,6%	0,8%	2606

Tabel B1.2 – Zelfvertrouwen in wiskunde

Zelfvertrouwen in wiskunde	Zeer mee eens	Beetje mee eens	Beetje mee oneens	Zeer mee oneens	Missing	Totaal
a. Meestal ben ik goed in wiskunde	24,1%	42,2%	21,4%	11,6%	0,8%	2606
b. Wiskunde is voor mij moeilijker dan voor veel van mijn klasgenoten	8,9%	20,8%	37,5%	31,0%	1,8%	2606
c. Ik ben gewoon niet goed in wiskunde	15,1%	19,2%	33,0%	31,2%	1,5%	2606
d. Wiskunde is makkelijk voor mij	12,2%	33,1%	29,0%	22,0%	3,8%	2606
e. Ik ben goed in het oplossen van moeilijke wiskundeopgaven	8,5%	34,4%	29,9%	25,9%	1,3%	2606
f. Ik kan wiskunde goed aan anderen uitleggen	8,6%	35,3%	30,0%	25,1%	1,0%	2606
g. Wiskunde is voor mij moeilijker dan alle andere vakken op school	15,6%	23,1%	30,9%	28,8%	1,7%	2606
h. Ik vind wiskunde moeilijk te snappen	15,0%	25,8%	35,8%	21,5%	1,8%	2606
i. Ik hoef maar weinig tijd aan mijn wiskundehuiswerk te besteden om toch goede cijfers te halen	18,6%	27,0%	26,2%	26,9%	1,3%	2606

Tabel B1.3 – Belang van wiskunde

Belang van wiskunde	Zeer mee eens	Beetje mee eens	Beetje mee oneens	Zeer mee oneens	Missing	Totaal
a. Ik denk dat wiskunde leren mij helpt in mijn dagelijks leven	8,4%	31,3%	29,5%	30,3%	0,6%	2606
b. Ik heb wiskunde nodig bij het leren van andere schoolvakken	7,8%	32,2%	32,4%	26,5%	1,1%	2606
c. Ik moet goed in wiskunde presteren om later de vervolgstudie te kunnen doen die ik graag wil	16,2%	35,1%	26,7%	18,8%	3,2%	2606
d. Ik moet goed in wiskunde presteren om later de baan te kunnen krijgen die ik graag wil	15,3%	32,7%	29,0%	19,6%	3,3%	2606
e. Ik wil later een baan waarbij je wiskunde gebruikt	7,4%	19,5%	31,6%	38,3%	3,3%	2606
f. Wiskunde leren is belangrijk om in deze wereld vooruit te komen	15,3%	43,0%	24,8%	15,1%	1,9%	2606
g. Door wiskunde te leren krijg ik meer mogelijkheden om later een baan kunnen te vinden	21,6%	45,3%	19,3%	11,9%	1,9%	2606
h. Mijn ouders denken dat het belangrijk is dat ik goed in wiskunde ben	23,4%	41,9%	21,6%	9,9%	3,1%	2606
i. Het is belangrijk om goed in wiskunde te zijn	15,6%	43,8%	25,4%	13,3%	2,0%	2606

Tabel B1.4 – Attributie positieve wiskundeprestaties

Attributie positieve wiskundeprestaties	Zeer mee eens	Beetje mee eens	Beetje mee oneens	Zeer Mee oneens	Missing	Totaal
a. Als ik een wiskundeopdracht goed maak komt dat vooral omdat ik goed mijn best heb gedaan	35,1%	47,0%	12,4%	4,3%	1,3%	2606
c. Als ik een wiskundeopdracht goed maak komt dat vooral omdat ik hard gewerkt heb	26,1%	43,6%	22,4%	6,1%	1,8%	2606
f. Als ik een wiskundeopdracht goed maak komt dat vooral omdat ik goed heb opgelet tijdens de les	28,0%	43,8%	18,0%	7,9%	2,2%	2606

Tabel B1.5 – Growth mindset wiskunde

Growth mindset wiskunde	Zeer mee eens	Beetje mee eens	Beetje mee oneens	Zeer mee oneens	Missing	Totaal
b. Hoe goed je bent in wiskunde ligt vast en je kunt niet veel doen om dat te veranderen	9,1%	26,7%	35,3%	27,6%	1,4%	2606
d. Het maakt niet uit hoe slim je bent, als je wil kun je altijd goed worden in wiskunde	29,5%	37,3%	23,3%	8,5%	1,3%	2606
e. Je kunt niet veranderen hoe slim je in wiskunde bent	9,8%	20,5%	36,5%	31,2%	2,0%	2606
g. Goed zijn in wiskunde is iets waar je mee geboren wordt, je kunt niet veel doen om jezelf beter in wiskunde te maken	5,7%	16,8%	37,8%	37,8%	1,9%	2606
h. Iedereen kan goed zijn in wiskunde	35,7%	36,7%	17,8%	8,1%	1,7%	2606
i. Je kunt zelf veranderen hoe goed je bent in wiskunde	34,6%	42,9%	15,2%	5,5%	1,8%	2606

Tabel B1.6 – Ordelijkheid wiskundeles volgens leerlingen

Ordelijkheid wiskundeles	Elke of bijna elke les	Ongeveer de helft van de lessen	Sommige lessen	Nooit	Missing	Totaal
a. De leerlingen luisteren niet naar wat de docent zegt	14,4%	22,6%	53,1%	8,6%	1,3%	2606
b. Het is lawaaig en rommelig in de klas	22,5%	25,6%	43,3%	7,2%	1,5%	2606
c. Het is te onrustig voor de leerlingen om te kunnen werken	11,8%	20,8%	47,1%	18,3%	2,0%	2606
d. De docent moet een lange tijd wachten voordat de leerlingen stil zijn	20,3%	21,5%	43,1%	13,4%	1,9%	2606
e. De leerlingen onderbreken de docent	17,7%	20,5%	44,0%	16,3%	1,6%	2606
f. De docent moet de regels in de klas heel vaak herhalen	13,7%	16,6%	34,9%	32,3%	2,4%	2606
g. Het gedrag van andere leerlingen in de klas maakt het voor mij moeilijk om me te kunnen concentreren.	11,9%	18,5%	34,0%	33,8%	1,8%	2606
h. De leerlingen beginnen pas met werken als de les allang begonnen is	19,2%	24,1%	41,1%	13,5%	2,1%	2606

Tabel B1.7 – Helderheid instructie wiskundedocent

Helderheid instructie wiskundedocent	Zeer mee eens	Beetje mee eens	Beetje mee oneens	Zeer mee oneens	Missing	Totaal
a. Ik weet precies wat ik van mijn docent moet doen	34,8%	42,7%	14,8%	5,8%	1,8%	2606
b. Ik kan mijn docent goed begrijpen	32,0%	40,0%	16,7%	8,8%	2,5%	2606
c. Mijn docent geeft duidelijk antwoord op mijn vragen	35,6%	34,0%	18,1%	8,8%	3,5%	2606
d. Mijn docent kan wiskunde goed uitleggen	43,5%	30,8%	14,2%	8,7%	2,8%	2606
e. Mijn docent doet verschillende dingen om ons te helpen met leren	32,4%	37,6%	18,1%	9,2%	2,7%	2606
f. Tijdens de uitleg van nieuwe leerstof maakt mijn docent gebruik van wat ik al weet	27,1%	43,0%	19,6%	7,4%	2,8%	2606
g. Als wij iets niet snappen dan legt mijn docent het onderwerp nog een keer uit	44,7%	33,6%	13,0%	6,2%	2,5%	2606
h. Mijn docent gaat net zo lang door met uitleggen tot de leerlingen het begrijpen	33,1%	32,6%	21,6%	9,7%	3,0%	2606
i. Als ik iets niet begrijp dan legt mijn docent het op een andere manier nog een keer uit	31,4%	34,7%	20,4%	10,7%	2,8%	2606

Tabel B1.8 – Oplossingsstrategieën in de wiskundeles

Oplossingsstrategieën wiskundeles	Zeer mee eens	Beetje mee eens	Beetje mee oneens	Zeer mee oneens	Missing	Totaal
a. Mijn docent leert ons verschillende manieren om een opgave op te lossen	27,3%	38,1%	22,5%	9,1%	3,0%	2606
b. Mijn docent geeft ons opgaven waarvoor we zelf een eigen oplossingsmanier moeten bedenken	14,9%	32,5%	31,8%	17,9%	2,9%	2606
c. Ik mag van mijn docent een opgave op mijn eigen manier oplossen	21,5%	34,6%	25,3%	15,5%	3,1%	2606

Tabel B1.9 – Directe instructie in de wiskundeles

Directie instructie wiskundeles	Zeer mee eens	Beetje mee eens	Beetje mee oneens	Zeer mee oneens	Missing	Totaal
a. De docent stelt duidelijke doelen voor wat we leren	43,7%	28,5%	18,4%	6,4%	3,0%	2606
b. De docent stelt vragen om te controleren of we begrepen hebben wat we hebben besproken	42,3%	27,9%	19,1%	7,6%	3,1%	2606
c. Aan het begin van de les geeft de docent een korte samenvatting van de vorige les	26,0%	21,5%	28,2%	21,3%	3,0%	2606
d. De docent zegt ons wat we moeten leren	55,2%	23,9%	12,9%	4,8%	3,2%	2606

## Bijlage 2 – Docentvragenlijst: rechte tellingen

### B2.1 – Geschiktheid wiskundemethode (n=100 docenten)

Geschiktheid wiskundemethode voor:	Zeer goed	Goed	Matig	Slecht	Zeer slecht	Totaal
j. uw eigen instructiemethode	21,0%	71,0%	8,0%	0,0%	0,0%	100
k. uw eigen kennis en vaardigheden	31,0%	65,0%	4,0%	0,0%	0,0%	100
l. uw overtuigingen over wiskundeonderwijs	12,0%	66,0%	20,0%	1,0%	1,0%	100
m. de leerlingen in uw klas	8,0%	69,0%	21,0%	2,0%	0,0%	100

### B2.2 – Leerlingactiviteiten tijdens de wiskundeles (n=100 docenten)

Leerlingactiviteiten tijdens de wiskundeles	Alle of bijna alle lessen	Ongeveer de helft van de lessen	Sommige lessen	Nooit of bijna nooit	Totaal
m. Luisteren naar mijn uitleg over nieuwe wiskundeleerstof	91,0%	6,0%	3,0%	0,0%	100
n. Luisteren naar mijn uitleg over hoe je opgaven kunt oplossen	70,0%	23,0%	7,0%	0,0%	100
o. Uit het hoofd leren van regels, procedures en feiten (zoals automatiseren)	20,0%	30,0%	41,0%	9,0%	100
p. Het zelfstandig oefenen van procedures	85,0%	14,0%	1,0%	0,0%	100
q. Het zelfstandig toepassen van het geleerde bij nieuwe probleemsituaties	46,0%	34,0%	18,0%	2,0%	100
r. Klassikaal aan opgaven werken onder mijn directe begeleiding	55,0%	24,0%	20,0%	1,0%	100
s. Werken in heterogene vaardigheidsgroepen	22,0%	16,0%	35,0%	27,0%	100
t. Werken in homogene vaardigheidsgroepen	12,0%	23,0%	38,0%	27,0%	100
u. Uitwerkingen opschrijven op het bord	38,0%	21,0%	25,0%	16,0%	100
v. Uitleggen hoe ze op een gegeven antwoord zijn gekomen	56,0%	30,0%	14,0%		100
w. Na te denken over verschillende oplossingsmanieren voor een wiskundig probleem	15,0%	45,0%	32,0%	8,0%	100
x. Wiskundige problemen op te lossen waarbij niet direct duidelijk is hoe de oplossing gevonden kan worden	10,0%	21,0%	46,0%	23,0%	100

### B2.3 – Instructiemodel tijdens wiskundeles (n=100 docenten)

Instructiemodel tijdens wiskundeles	Alle of bijna alle lessen	Ongeveer de helft van de lessen	Sommige lessen	Nooit of bijna nooit	Totaal
a. Directe instructiemodel	79,0%	18,0%	3,0%	0,0%	100
b. Vormen van ontdekkend/onderzoekend leren	4,0%	24,0%	52,0%	20,0%	100

#### B2.4 – ICT-gebruik tijdens wiskundeles (n=100 docenten)

ICT-gebruik in de wiskundeles	Ten minste één keer per week	Eén of twee keer per maand	Een paar keer per jaar	Nooit of bijna nooit	Missing*	Totaal
a. Het oefenen van opgaven en procedures	27,0%	17,0%	13,0%	17,0%	26,0%	100
b. Het oplossen van complexe problemen of problemen uit het dagelijks leven	7,0%	9,0%	23,0%	35,0%	26,0%	100
c. Het maken van grafieken, tabellen of andere manieren om data te presenteren	7,0%	9,0%	23,0%	35,0%	26,0%	100
d. Het gebruiken van games waarin wiskundeopgaven of -concepten worden gebruikt	5,0%	5,0%	19,0%	45,0%	26,0%	100
e. Het lezen van instructie of het kijken van instructie-video's	17,0%	22,0%	19,0%	16,0%	26,0%	100
f. Het maken van wiskundetoetsen	0,0%	13,0%	18,0%	43,0%	26,0%	100

\* 26 van de 100 docenten hebben aangegeven dat de leerlingen tijdens de wiskundeles géén gebruik maken van digitale apparaten

#### B2.5 – Soorten feedback bij huiswerk (n=100 docenten)

Soorten feedback	Altijd of bijna altijd	Soms	Nooit of bijna nooit	Missing*	Totaal
a. Het huiswerk corrigeren en feedback geven aan de leerlingen	28,0%	47,0%	19,0%	6,0%	100
b. Het huiswerk bespreken in de klas	51,0%	36,0%	7,0%	6,0%	100
c. Controleren of leerlingen het huiswerk gemaakt hebben	34,0%	51,0%	9,0%	6,0%	100
d. Ik laat leerlingen zelf hun huiswerk nakijken	73,0%	16,0%	5,0%	6,0%	100

\* 6 van de 100 docenten hebben aangegeven (bijna) nooit huiswerk voor wiskunde op te geven. Bovenstaande stellingen zijn niet voorgelegd aan deze docenten.

*B2.6 – Zelfvertrouwen bij onderwijsactiviteiten voor wiskunde (n=100 docenten)*

Zelfvertrouwen wiskunde-instructie	Zeer hoog	Hoog	Gemiddeld	Laag	Totaal
i. Leerlingen inspireren om wiskunde te leren	7,0%	58,0%	33,0%	2,0%	100
j. Leerlingen verschillende oplossingsstrategieën aanreiken	16,0%	51,0%	32,0%	1,0%	100
k. Uitdagende wiskundetaken geven aan leerlingen die bovengemiddeld presteren	4,0%	33,0%	55,0%	8,0%	100
l. Mijn instructie aanpassen om de betrokkenheid van leerlingen bij de les te vergroten	14,0%	52,0%	33,0%	1,0%	100
m. Leerlingen het belang en nut van wiskunde te laten inzien	5,0%	43,0%	47,0%	5,0%	100
n. Het evalueren van het begrip van leerlingen in wiskunde	8,0%	40,0%	51,0%	1,0%	100
o. Het verbeteren van het begrip van zwakke leerlingen in wiskunde	13,0%	49,0%	38,0%	0,0%	100
p. Het ontwikkelen van hogere denkvaardigheden bij leerlingen	5,0%	37,0%	53,0%	5,0%	100

*B2.7 – Evaluatie van prestaties en leerproces bij wiskunde (n=100 docenten)*

Evaluatiebron	Zeer	Enigszins	Niet	Totaal
a. Ik gebruik scores en/of antwoorden op wiskundetoetsen in het kader van een leerlingvolgsysteem om te evalueren of de leerdoelen bereikt zijn	39,0%	46,0%	15,0%	100
b. Ik analyseer de antwoorden op methodegebonden wiskundetoetsen om te evalueren of de leerdoelen van dat blok bereikt zijn	38,0%	48,0%	14,0%	100
c. Ik evalueer regelmatig of alle leerlingen de lesdoelen bereikt hebben op basis van hun ingeleverde werk	41,0%	46,0%	13,0%	100
d. Ik evalueer of alle leerlingen de lesdoelen bereikt hebben op basis van (informele) observaties tijdens de wiskundeles	63,0%	37,0%	0,0%	100
e. Ik voer diagnostische gesprekken om te evalueren of specifieke leerlingen de lesdoelen bereikt hebben	32,0%	49,0%	19,0%	100
f. Ik evalueer of de door mij gekozen manieren van instructie en verwerking effectief waren voor de meerderheid van de leerlingen in de klas	58,0%	37,0%	5,0%	100
g. Ik evalueer of een specifieke manier van instructie effectief was voor specifieke leerlingen	24,0%	63,0%	13,0%	100



### B2.8 – Belang van beoordelingsstrategieën (n=100 docenten)

Beoordelingsstrategieën	Zeer	Enigszins	Niet	Totaal
a. Het observeren van leerlingen als zij aan het werk zijn	86,0%	14,0%	0,0%	100
b. Leerlingen vragen laten beantwoorden tijdens de les	84,0%	16,0%	0,0%	100
c. Het doornemen van hun dagelijkse werk en/of huiswerk	35,0%	59,0%	6,0%	100
d. Korte, tussentijdse schriftelijke toetsen	16,0%	52,0%	32,0%	100
e. Methodegebonden toetsen aan het eind van een periode of blok	71,0%	25,0%	4,0%	100
f. Genormeerde toetsen in kader van een leerlingvolgsysteem, zoals bijvoorbeeld van Cito	10,0%	42,0%	48,0%	100
g. Langlopende wiskundeprojecten	2,0%	34,0%	64,0%	100

### B2.9 – Gebruik toetsen bij wiskunde (n=100 docenten)

Type	Zeer	Enigszins	Niet	Totaal
a. Genormeerde toetsen (bijvoorbeeld van Cito)	3,0%	27,0%	70,0%	100
b. Methodegebonden toetsen	67,0%	24,0%	9,0%	100
c. Zelfgemaakte toetsen	60,0%	35,0%	5,0%	100

### B2.10 – Differentiatie bij wiskundeles (n=99 docenten)\*

Differentiatie	Zeer	Enigszins	Niet	Missing*	Totaal
a. Ik benut de mogelijkheden die de methode biedt voor differentiatie voor leerlingen die in mijn klas relatief bovengemiddeld in wiskunde presteren	30,0%	62,0%	7,0%	1,0%	100
b. Ik benut de mogelijkheden die de methode biedt voor differentiatie voor de relatief zwakke leerlingen in wiskunde in mijn klas	30,0%	62,0%	7,0%	1,0%	100
c. Ik stel bewust open vragen tijdens de klassikale instructie	81,0%	18,0%	0,0%	1,0%	100
d. Ik stel bewust vragen van verschillende moeilijkheidsgraad tijdens de klassikale instructie	67,0%	26,0%	6,0%	1,0%	100
e. Ik bied leerlingen die in mijn klas relatief bovengemiddeld in wiskunde presteren verrijkingsopdrachten	20,0%	65,0%	14,0%	1,0%	100
f. Ik bied leerlingen die in mijn klas relatief zwak in wiskunde presteren extra oefenopdrachten	30,0%	56,0%	13,0%	1,0%	100
g. De leerlingen kunnen binnen mijn klas wiskunde op een hoger onderwijsniveau volgen (zoals kbl-wiskunde in een bbl-klas)	28,0%	36,0%	35,0%	1,0%	100

\*Vanaf deze vraag is er 1 respondent helemaal gestopt met invullen.

B2.11 – Fixed/Growth mindset (n=99 docenten)

<i>Fixed/Growth mindset</i>	Zeer mee eens	Enigszins mee eens	Enigszins mee oneens	Zeer mee oneens	Missing	Totaal
a. <i>Het maakt niet uit hoe intelligent leerlingen zijn, als leerlingen willen kunnen ze altijd goed worden in wiskunde</i>	15,0%	64,0%	18,0%	2,0%	1,0%	100
b. <i>Hoe goed een leerling is in wiskunde ligt vast en je kunt niet veel doen om dit te veranderen</i>	0,0%	4,0%	35,0%	60,0%	1,0%	100
c. <i>Je kunt leerlingen wel nieuwe dingen leren over wiskunde, maar je kunt niet veranderen hoe slim ze in wiskunde zijn</i>	1,0%	26,0%	46,0%	26,0%	1,0%	100
d. <i>Elke leerling kan goed zijn in wiskunde</i>	22,0%	51,0%	18,0%	8,0%	1,0%	100
e. <i>Een leerling kan zelf veranderen hoe goed hij/zij kan presteren in wiskunde</i>	45,0%	50,0%	2,0%	2,0%	1,0%	100
f. <i>Aanleg voor wiskunde is iets waar leerlingen mee geboren worden, je kunt niet veel doen om ze beter in wiskunde te maken dan het niveau van de aangeboren aanleg</i>	0,0%	14,0%	36,0%	49,0%	1,0%	100

B2.12 – Gevolgen van coronapandemie (n=55 docenten)\*

Als gevolg van de coronapandemie...	Zeer mee eens	Enigszins mee eens	Enigszins mee oneens	Zeer mee oneens	Missing*	Totaal
o. ... hebben leerlingen leervertraging opgelopen op het gebied van wiskunde	23,0%	26,0%	3,0%	3,0%	45,0%	100
p. ...zijn de wiskundeprestaties er op achteruit gegaan	23,0%	23,0%	5,0%	4,0%	45,0%	100
q. ...zijn verschillen tussen leerlingen in wiskundeprestaties groter geworden	26,0%	20,0%	7,0%	2,0%	45,0%	100
r. ...is de motivatie om te leren onder leerlingen afgenomen	25,0%	23,0%	5,0%	2,0%	45,0%	100
s. ...is de concentratie van leerlingen tijdens de wiskundeles (online en/of fysiek) afgenomen	29,0%	20,0%	4,0%	2,0%	45,0%	100
t. ...ondervinden leerlingen meer moeite met het plannen van hun werk (zoals huiswerk of leren voor een toets)	27,0%	22,0%	3,0%	3,0%	45,0%	100
u. ... zijn leerlingen minder zorg aan hun huiswerk gaan besteden	22,0%	25,0%	7,0%	1,0%	45,0%	100
v. ... wordt het huiswerk minder vaak gemaakt	19,0%	25,0%	9,0%	2,0%	45,0%	100
w. ...is het aantal conflicten tussen leerlingen onderling toegenomen	14,0%	14,0%	16,0%	11,0%	45,0%	100
x. ...is het aantal conflicten tussen leerlingen en docenten en/of ander personeel toegenomen	8,0%	20,0%	16,0%	11,0%	45,0%	100
y. ... worden leerlingen sneller boos of verdrietig	11,0%	28,0%	14,0%	2,0%	45,0%	100
z. ... is pesten onder leerlingen toegenomen (fysiek of online)	9,0%	16,0%	22,0%	8,0%	45,0%	100
aa. ... is spijbelen toegenomen	11,0%	20,0%	15,0%	9,0%	45,0%	100
bb. ... is de bereidheid om zich te houden aan de gedragsregels van de school afgenomen	20,0%	26,0%	8,0%	1,0%	45,0%	100

\* Deze vragen zijn alleen gesteld aan docenten die voor het begin van de coronapandemie al wiskunde aan leerlingen in het tweede leerjaar op de aan het onderzoek deelnemende school gaven. (44 docenten deden dit niet, 1 missing).

## Bijlage 3 – Schoolvragenlijst: rechte tellingen

### B3.1 – Prestatiegerichtheid Schoolklimaat

Kenmerk	Erg hoog	Hoog	Gemiddeld	Laag	Erg laag	Totaal
l. De mate waarin docenten de onderwijsdoelen van de school kennen	5,0%	53,5%	38,6%	3,0%	0,0%	101
m. De mate waarin docenten succesvol zijn in de realisatie van het curriculum van de school	4,0%	66,3%	28,7%	1,0%	0,0%	101
n. Verwachtingen van docenten wat betreft de leerresultaten van de leerlingen	1,0%	40,6%	54,5%	4,0%	0,0%	101
o. De mate waarin docenten in staat zijn leerlingen te inspireren	1,0%	25,7%	69,3%	4,0%	0,0%	101
p. Betrokkenheid van ouders bij schoolactiviteiten	1,0%	19,8%	48,5%	24,8%	5,9	101
q. De mate waarin ouders ervoor zorgen dat hun kind in staat is deel te nemen aan het leerproces (bijv. voldoende slaap, ontbijt)	5,0%	36,6%	46,5%	10,9%	1,0	101
r. Verwachtingen van ouders wat betreft de leerresultaten van hun kind	4,0%	43,6%	49,5%	3,0%	0,0%	101
s. Ondersteuning door ouders bij het leren van hun kind	0,0%	29,7%	48,5%	21,8%	0,0%	101
t. De wil van leerlingen om goed te presteren op school	1,0%	16,8%	69,3%	11,9%	1,0%	101
u. De mate waarin leerlingen in staat zijn om de leerdoelen van de school te halen	1,0%	33,7%	61,4%	4,0%	0,0%	101
v. Respect van leerlingen voor medeleerlingen die zeer goed presteren	1,0%	32,7%	59,4%	6,9%	0,0%	101

### B3.2 – Belang leerlingvolgsysteem

Activiteit (n=99)	Zeer belangrijk	Belangrijk	Nauwelijks of niet belangrijk	Missing*	Totaal
a. Het volgen van de cognitieve ontwikkeling van leerlingen, zoals het bijhouden van toetsresultaten	74,3%	22,8%	1,0%	2,0%	101
b. Het volgen van de sociaal-emotionele ontwikkeling van leerlingen	33,7%	55,4%	8,9%	2,0%	101
c. Het volgen van studievoortgang en schoolloopbaan	54,5%	42,6%	1,0%	2,0%	101
d. Het volgen van leerlinggedrag zoals verzuim of pesten	68,3%	28,7%	1,0%	2,0%	101
e. Het opstellen van handelingsplannen	20,8%	58,4%	18,8%	2,0%	101
f. Het communiceren met leerlingen over bijvoorbeeld toetsresultaten, cijfers, roosters, huiswerk, uitval van lessen	66,3%	28,7%	3,0%	2,0%	101
g. Het communiceren met ouders over bijvoorbeeld leerlingresultaten, roosters, huiswerk, uitval van lessen	46,5%	44,6%	6,9%	2,0%	101
h. Het communiceren tussen docenten over bijvoorbeeld leerlingresultaten, leermaterialen onderwijsactiviteiten en -ontwikkelingen	30,7%	41,6%	25,7%	2,0%	101
i. Leerlingen toegang geven tot lesmateriaal	39,6%	38,6%	19,8%	2,0%	101

\*2 vd 101 schoolleiders hebben aangegeven geen leerlingvolgsysteem te gebruiken.

### B3.3 – Bevordering interesse in wiskunde

Uitspraak	Zeer mee eens	Enigszins mee eens	Enigszins mee oneens	Zeer mee oneens	Missing	Totaal
j. De school biedt leerlingen informatie over wiskunde-gerelateerde vervolgopleidingen en beroepen	18,8%	53,5%	21,8%	5,0%	1,0%	101
k. De school ontwikkelt initiatieven om de interesse van leerlingen voor wiskunde te stimuleren (zoals wiskundeclub of -wedstrijden)	10,9%	34,7%	35,6%	17,8%	1,0%	101
l. De school stimuleert de professionele ontwikkeling van wiskundeleraars	37,6%	53,5%	7,9%	99,0%	1,0%	101
m. De school biedt leerlingen extra wiskundelessen om leerlingen te helpen (nog) beter te presteren	23,8%	40,6%	25,7%	8,9%	1,0%	101
n. De school biedt speciale activiteiten op gebied van wiskunde voor leerlingen die daarin geïnteresseerd zijn	8,9%	28,7%	38,6%	22,8%	1,0%	101
o. De school heeft het specifieke doel om het wiskundeonderwijs op school te verbeteren	14,9%	40,6%	38,6%	5,0%	1,0%	101
p. De school moedigt leerlingen aan om in de toekomst voor wiskunde te kiezen	9,9%	48,5%	37,6%	3,0%	1,0%	101
q. De wiskundeleraars van deze school besteden extra tijd aan leerlingen die geïnteresseerd zijn in wiskunde	11,9%	48,5%	31,7%	6,9%	1,0%	101
r. De school faciliteert leerlingen in de bovenbouw die examen in wiskunde A én B willen doen	12,9%	26,7%	20,8%	38,6%	1,0%	101

### B3.4 – Veranderingen door coronapandemie

Als gevolg van de coronapandemie...	Zeer mee eens	Enigszins mee eens	Enigszins mee oneens	Zeer mee oneens	Missing	Totaal
a. ... hebben leerlingen leervertraging opgelopen op het gebied van wiskunde	16,8%	37,6%	5,0%	1,0%	39,6%	101
b. ... hebben leerlingen leervertraging opgelopen op het gebied van leesvaardigheid	22,8%	32,7%	4,0%	1,0%	39,6%	101
c. ...zijn de wiskundeprestaties er op achteruit gegaan	15,8%	27,7%	14,9%	2,0%	39,6%	101
d. ...is de leesvaardigheid er op achteruit gegaan	16,8%	31,7%	9,9%	2,0%	39,6%	101
e. ...zijn verschillen tussen leerlingen in wiskundeprestaties groter geworden	17,8%	23,8%	15,8%	3,0%	39,6%	101
f. ...zijn verschillen tussen leerlingen in leesvaardigheid groter geworden	16,8%	28,7%	12,9%	2,0%	39,6%	101
g. ...is de motivatie om te leren onder leerlingen afgenomen	39,6%	15,8%	4,0%	1,0%	39,6%	101
h. ...is de concentratie van leerlingen afgenomen	41,6%	14,9%	4,0%	0,0%	39,6%	101
i. ...is het aantal conflicten tussen leerlingen onderling toegenomen	26,7%	14,9%	12,9%	5,9%	39,6%	101
j. ...is het aantal conflicten tussen leerlingen en docenten en/of ander personeel toegenomen	16,8%	18,8%	17,8%	6,9%	39,6%	101
k. ... worden leerlingen sneller boos of verdrietig	25,7%	21,8%	8,9%	4,0%	39,6%	101
l. ... is pesten onder leerlingen toegenomen (fysiek of online)	12,9%	25,7%	14,9%	6,9%	39,6%	101
m. ... is spijbelen toegenomen	12,9%	25,7%	14,9%	6,9%	39,6%	101
n. ... is de bereidheid om zich te houden aan de gedragsregels van de school afgenomen	24,8%	27,7%	5,9%	2,0%	39,6%	101

## Bijlage 4 – Syntax hercodering predictoren

Alle predictoren behalve de dummy's zijn uiteindelijk als Z-score in de modellen toegepast. Het maken van de Z-scores is in het programma R gedaan, voor iedere geïmputeerde dataset apart:

### Vordering

Versneld of vertraagd? Cut-off point is de maand sept: 1-okt-2007 t/m 30-sept-2008 = regulier = 14 of 15 jaar in VO-2. Zie artikel Hans Luyten in PS-2013-5. 2007=3; 2008=4.

```
COMPUTE vordering=9.
```

```
If (VL_vr1_2_a_mk<=9 and VL_vr1_2_b_mk<=3) vordering=-1.
```

```
If (VL_vr1_2_a_mk>=10 and VL_vr1_2_b_mk=3) vordering=0.
```

```
If (VL_vr1_2_a_mk<=9 and VL_vr1_2_b_mk=4) vordering=0.
```

```
If (VL_vr1_2_a_mk>=10 and VL_vr1_2_b_mk>=4) vordering=1.
```

```
VALUE LABELS vordering
```

```
-1 'vertraagd'
```

```
0 'normaal vorderend'
```

```
1 'versneld'
```

```
9 'missing'.
```

```
missing values vordering (9).
```

### Thuis taal Nederlands

```
recode VL_vr1_7_mk (4=1) (3=2) (2=3) (1=4) into thuis taalNL.
```

```
value labels thuis taalNL
```

```
1 'nooit'
```

```
2 'soms'
```

```
3 'bijna altijd'
```

```
4 'altijd'.
```

```
*freq thuis taalnl.
```

### Onderwijstype leerjaar 3 bekend & Onderwijstype leerjaar3 onbekend of gemengd

```
MISSING VALUES VL_vr1_8_mk ().
```

```
RECODE VL_vr1_8_mk (1=1) (2=2) (3=3) (4=4) (5=5) (6=99) (7=99) into owtypeLJ3.
```

```
value labels owtypeLJ3
```

```
1 'vmbo bl'
```

```
2 'vmbo kl'
```

```
3 'vmbo gl/tl'
```

```
4 'havo'
```

```
5 'vwo'
```

```
99 'gemengde brugklas/weet nog niet'.
```

```
if (owtypeLJ3>=0 and owtypeLJ3<=5) owtypeLJ3Onb=0.
```

```
if (owtypeLJ3 =99) owtypeLJ3Onb=1.
```

```
missing values owtypeLJ3 (99).
```

### Gebruik rekenmachine tijdens de wiskundeles (docent)

```
recode q3punt3 (1=1) (2=0) (3=0) into rekenmachine.
```

```
value labels rekenmachine
```

```
0 'beperkt'
```

```
1 'altijd'.
```



#### Alle leerlingen meerdere Oplossingsstrategieën aanbieden (docent)

recode q2punt10 (1=1) (2=2) (3=2) (4=2) (5=2) (6=3) into OplosStrategieDiff.

value labels **OplosStrategieDiff**

- 1 'een strategie ongediff'
- 2 'meerdere gediff'
- 3 'meerdere ongediff'.

#### Man (docent)

recode q1punt2 (1=0) (2=1) into man.

#### Speciaal onderwijsconcept (school)

recode owc (1=0) (2 thru 10=1) into SpecOWconcept.

value labels **SpecOWconcept**

- 0 'algemeen owc'
- 1 'specifiek owc'.

#### Aandeel leerlingen voor wie Nederlands niet de eerste taal is (school)

De oorspronkelijke naam variabele vond ik verwarrend, daarom hernoemd.

compute NLniet1etaal= q1punt5.

VARIABLE LABELS NLniet1etaal 'Percentage leerlingen voor wie NL NIET de eerste taal is'.

value labels **NLniet1etaal**

- 1 '0-9%'
- 2 '10-24%'
- 3 '25-49%'
- 4 '50-74%'
- 5 '75-100%'.

#### Stedelijkheid (school)

recode q1punt6 (1=5) (2=4) (3=3) (4=2) (5=1) into stedelijkheid.

value labels **stedelijkheid**

- 1 'platteland'
- 2 'klein drop of stad'
- 3 'grote of middelgrote stad'
- 4 'voorstad of buitenwijk stedelijk gebied'
- 5 'stedelijkheid met een hoge bevolkingsdichtheid'.

#### Extra wiskundelessen buiten schooltijd (school)

recode q2punt1 (1=1) (2=0) into ExtraWisk.

value labels **ExtraWisk**

- 0 'nee'
- 1 'ja, buiten schooltijd'.

#### Aantal aangeboden professionaliseringsactiviteiten voor wiskundedocenten (school)

recode q2punt6\_SQ001 q2punt6\_SQ002 q2punt6\_SQ003 q2punt6\_SQ004 q2punt6\_SQ005  
q2punt6\_SQ009 (1=1) (2=0).

compute ProfesWisDocent = sum(q2punt6\_SQ001, q2punt6\_SQ002, q2punt6\_SQ003,  
q2punt6\_SQ004, q2punt6\_SQ005, q2punt6\_SQ009).

VARIABLE LABELS **ProfesWisDocent** 'Aantal aangeboden professionaliseringsactiviteiten voor Wiskunde docenten'.



