



Inspectie van het Onderwijs  
*Ministerie van Onderwijs, Cultuur en  
Wetenschap*

**TECHNISCH RAPPORT Peil.Rekenen en Wiskunde  
Einde leerjaar 2 voortgezet onderwijs**

2021-2022

Februari 2024

## INHOUD

<b>Inleiding</b>		<b>3</b>
<b>1</b>	<b>Het onderwijsleerproces in Rekenen en Wiskunde</b>	<b>4</b>
1.1	Databronnen	4
1.2	Werkwijze	5
1.2.1	Nieuwe variabelen	5
1.2.2	Resultaten en toetsing	6
<b>2</b>	<b>Attituden en achtergrondkenmerken</b>	<b>12</b>
2.1	Databronnen	12
2.2	Werkwijze	12
2.2.1	Nieuwe variabelen	12
2.2.2	Resultaten en toetsing	13
<b>3</b>	<b>Prestaties rekenen en wiskunde</b>	<b>18</b>
3.1	Databronnen	18
3.1.1	Rekenen en wiskunde meten	18
3.1.2	Prestaties rekenen en wiskunde	19
3.2	Werkwijze	20
3.2.1	Significantie toetsing	20
3.2.2	Bepalen van voorbeelditems	25
3.2.3	Globale vergelijking met (s)bo	26
<b>4</b>	<b>Verschillen in reken- en wiskundeprestaties</b>	<b>27</b>
<b>5</b>	<b>Referenties</b>	<b>28</b>
<b>Colofon</b>		<b>29</b>

## Inleiding

Dit is het technisch rapport bij het peilingsonderzoek Rekenen en Wiskunde einde tweede leerjaar van het voortgezet onderwijs (vo). De peiling betreft het schooljaar 2021-2022. Dit peilingsonderzoek is in opdracht van het Nationaal Regieorgaan Onderwijsonderzoek en de Inspectie van het Onderwijs (verder: inspectie) uitgevoerd door een consortium bestaande uit Expertisecentrum Nederlands, KBA Nijmegen, Universiteit Twente en Cito B.V. In de technische rapportage van het uitvoerend consortium (Van Langen et al., 2023) is een uitgebreide verantwoording opgenomen van de wijze van dataverzameling, het instrumentarium en de eerste gegevensverwerking en -analyse ter beantwoording van de onderzoeksvragen.

In dit technisch rapport staan de keuzes centraal die gemaakt zijn door de onderzoekers van de inspectie die de publieksrapportage hebben geschreven. De verdere gegevensverwerking en -analyse die ten grondslag liggen aan het publieksrapport worden hier verantwoord.

Het doel van het peilingsonderzoek Rekenen en wiskunde einde leerjaar 2 voortgezet onderwijs is om vast te stellen hoe reken- en wiskundevaardig leerlingen zijn aan het einde van de tweede klas van het praktijkonderwijs (pro), vmbo-b/k, vmbo-g/t en havo/vwo. Wat kennen en kunnen deze leerlingen op het leergebied rekenen en wiskunde? Daarnaast beschrijft het rapport hoe het onderwijsleerproces voor rekenen (pro, vmbo-basisberoeps) en wiskunde (vmbo-kaderberoeps, vmbo-gemengd/theoretisch en havo/vwo) er op de deelnemende scholen uitziet en geeft het een beeld van een aantal aan rekenen en wiskunde gerelateerde kenmerken van leerlingen, docenten en scholen. Tot slot wordt gekeken naar de prestatieverschillen tussen leerlingen en klassen en naar hoe de prestaties samenhangen met kenmerken van leerlingen, leerkrachten en scholen.

De onderzoeksvragen die het uitgangspunt vormen voor het publieksrapport zijn:

- 1) Hoe kan het onderwijsleerproces van rekenen en wiskunde gekarakteriseerd worden?
- 2) Wat zijn de achtergrondkenmerken en attitudes van leerlingen en docenten ten aanzien van rekenen en wiskunde?
- 3) Wat kennen en kunnen de leerlingen op het gebied van rekenen en wiskunde?
- 4) Hoe hangen de reken-wiskundeprestaties van leerlingen samen met kenmerken van de leerlingen, docenten, scholen en het onderwijsleerproces?

De opbouw van dit technisch rapport volgt de hoofdstukindeling van het publieksrapport (Inspectie van het Onderwijs, 2024). We onderbouwen per hoofdstuk uit deel B (de resultaten), de gegevens die we hebben gebruikt en onze werkwijze. Alle analyses zijn uitgevoerd met behulp van R (R Core Team, 2024).

# 1 Het onderwijsleerproces in Rekenen en Wiskunde

In dit hoofdstuk van de publieksrapportage wordt het onderwijsleerproces op het gebied van rekenen en wiskunde van vo scholen uit dit onderzoek in kaart gebracht.

## 1.1 Databronnen

Voor het in kaart brengen van het onderwijsleerproces zijn vragenlijsten afgenomen bij schoolleiders en wiskundedocenten. Het ging om 101<sup>1</sup> schoolleiders van in totaal 130 deelnemende vo-vestigingen (respons 78%) en 100<sup>2</sup> wiskundedocenten van in totaal 133 vo-klassen (respons 77%).

De schoolleidersvragenlijst bestond uit 17 vragen, sommige vragen bestonden uit meerdere items. Op basis van de vraag over de aangeboden onderwijssoort is onderscheid gemaakt tussen de verschillende onderwijssoorten in het vo. Op schoolniveau wordt een combinatie van meerdere onderwijssoorten aangeboden (leerwegen vmbo, havo en vwo). Dit noemen we een breed aanbod, dit betrof 47 scholen. Op 30 deelnemende scholen wordt pro/vmbo-onderwijs aangeboden en 22 scholen bieden havo/vwo onderwijs (pro/vmbo 30%; breed 47%; havo/vwo 22%).

De docentvragenlijst bestond uit 32 vragen, sommige vragen bestonden uit meerdere items. Op klasniveau zijn de antwoorden van de wiskundedocenten van de pro-klassen vanwege het kleine aantal ( $n = 5$ ) samengevoegd met de antwoorden van de docenten die lesgeven aan vmbo basisberoeps- en/of kaderberoepsgerichte leerweg (vmbo-b/k:  $n = 23$ ). Bij 28 deelnemende klassen betrof het antwoorden van docenten die ten tijde van de afname lesgeven aan de vmbo gemengde en/of theoretische leerweg en bij 44 klassen om de antwoorden van havo/vwo-docenten. De indeling in onderwijssoorten verschilt voor de schoolleiders en de wiskundedocenten, omdat de schoolleiders de vragen hebben beantwoord over de gehele vestiging en de docenten over de klas die meedeed aan het peilingsonderzoek.

De leerlingvragenlijst bestond uit 16 vragen, sommige vragen bestonden uit meerdere items. De leerlingen in het pro vulden een verkorte versie in met 11 vragen. In de leerlingvragenlijst zijn leerlingen onder andere gevraagd naar de wijze van lesgeven door de wiskundedocent: ordelijkheid van de les, helderheid van de instructie, mate van directe instructie en aandacht voor oplossingsstrategieën. De resultaten van deze vragen zijn ook opgenomen in dit hoofdstuk. Het totaal aantal vo-leerlingen dat de vragenlijst invulde was 2620. 14 leerlingen hebben minder dan de helft van de vragenlijst ingevuld en zijn daarom niet meegenomen in deze rapportage. De indeling van de onderwijssoort van de leerlingen is gebaseerd op de onderwijssoort van de klas van de leerling, zoals aangegeven door de docent in de docentenvragenlijst, omdat de leerlingen aan hun docent waren gekoppeld. Van 604 leerlingen was na meermaals in herinnering brengen geen informatie beschikbaar over de onderwijssoort van hun klas, waardoor zij niet zijn meegenomen in de analyses. Het totale aantal leerlingen dat in de analyses is meegenomen betrof daarom 2079 vo-leerlingen, waarvan 77 pro-leerlingen; 394 vmbo-b/k-leerlingen; 573 vmbo-g/t/-leerlingen; en 1035 havo/vwo-leerlingen. Aan de resultaten van de pro-leerlingen konden vanwege het lage aantal geen conclusies worden verbonden voor het gehele pro in Nederland. De groep pro-leerlingen is te klein vanwege non-respons van steekproefscholen.

---

<sup>1</sup> Het totaal aantal schoolleiders dat deelnam was 130, van wie 106 de vragenlijst invulden. 5 schoolleiders vulden minder dan de helft van de vragenlijst in en zijn daarom niet meegenomen in deze rapportage.

<sup>2</sup> 104 van de 130 docenten hebben een vragenlijst ingevuld. 4 docenten vulden de vragenlijst voor meer dan de helft niet in en zijn daarom niet meegenomen in deze rapportage.

Omdat dit de eerste peiling is van rekenen en wiskunde aan het einde van het tweede leerjaar in het voortgezet onderwijs, was het niet mogelijk om te onderzoeken of het onderwijs in rekenen en wiskunde is veranderd ten opzichte van eerdere peilingen. Bovendien was het door verschil in vakgebied (rekenen versus wiskunde) niet mogelijk om het onderwijsleerproces letterlijk te vergelijken met de beschikbare gegevens in het po.

## 1.2

### **Werkwijze**

De beschrijvingen in het hoofdstuk van het publieksrapport zijn overwegend gebaseerd op frequentieverdelingen, uitgedrukt in percentages. De gegevens zijn apart beschreven voor de verschillende onderwijssoorten en het verschil tussen de onderwijssoorten is alleen getoetst op significantie voor de schaalscores. Er is gekozen om niet per stelling te toetsen om meerdere redenen. Zo zou dit resulteren in een zeer grote hoeveelheid toetsen, waardoor de kans op vals positieve resultaten toeneemt. Daarnaast is het aantal docenten en schoolleiders per antwoordcategorie in veel gevallen te laag om een betrouwbaar toetsresultaat te krijgen.

In alle gevallen is er sprake van schaalscores berekend door het consortium, zoals beschreven in hoofdstuk 6 van de technische rapportage (Van Langen et al., 2023). Ook voor de meerniveau-analyses in hoofdstuk 4 worden deze schaalscores gebruikt. Deze schaalscores worden ook in dit hoofdstuk gebruikt om verschillen tussen de onderwijssoorten te toetsen met ANOVA's. In het publieksrapport is de significantie op de schaal als geheel beschreven.

Voor de significantietoetsing van de schaalscores is een tweezijdig significantieniveau van 0,05 gehanteerd. Bij een significant verschil is een post-hoc analyse uitgevoerd.

### 1.2.1

#### *Nieuwe variabelen*

Het consortium heeft schaalscores berekend op basis van itemscores, zoals beschreven in hoofdstuk 6 van hun technische rapportage (Van Langen et al., 2023). Deze schaalscores worden ook voor dit hoofdstuk gebruikt om verschillen tussen de onderwijssoorten te toetsen. Ook bij de meerniveau analyses in hoofdstuk 4 worden deze schaalscores gebruikt.

Om de scores van docenten op de schalen beter te kunnen duiden, zijn voor dit hoofdstuk ook schaalscores op basis van gemiddeldes berekend. Hiervoor is het gemiddelde berekend van de scores op alle onderliggende items voor docenten die minstens de helft van deze items hebben beantwoord. Sommige items zijn voorafgaand hieraan gehercodeerd (4=1, 3=3, 2=3, 1=4) zodat de antwoorden op het item inhoudelijk overeenstemmen met de richting van de schaal. De resulterende schalen lopen van 1 (laag/negatief) tot en met 4 (hoog/positief), of in het geval van prestatiegerichtheid van het schoolklimaat van 1 (zeer laag) tot 5 (heel hoog).

Tot slot is voor het taartfiguur in de grafische samenvatting '*Verdeling van de domeinen over de gemiddelde onderwijstijd van 2.5 uur*' een gemiddelde genomen van de respons van de drie onderwijssoorten per domein.

## 1.2.2 Resultaten en toetsing

De tabellen met resultaten in dit hoofdstuk volgen de volgorde van het publieksrapport.

### 1.2.2.1 Instructie tijdens de les

#### Direct instructiemodel

Tabel 1.2.2.1a Uitkomsten multivariate variantieanalyse voor directe instructiemodel als afhankelijke variabele

	<i>N</i> <i>leerlingen</i>	<i>F</i>	<i>df</i>	<i>p</i> - waarde
Vmbo-b/k	302			
Vmbo-g/t	453	20,89	2;1906	<0,001
Havo/vwo	919			

Tabel 1.2.2.1b Uitkomsten post-hoc variantieanalyses voor directe instructiemodel

	Verskil in gemiddelde scores	Gecorrigeerde <i>p</i> -waarde
Vmbo-g/t-vmbo-b/k	-0,258	<0,001
Havo/vwo-vmbo-b/k	-0,026	0,827
Havo/vwo-vmbo-g/t	0,232	<0,001

#### Helderheid van de instructie

Tabel 1.2.2.1c Uitkomsten multivariate variantieanalyse voor de helderheid van de instructie als afhankelijke variabele

	<i>N</i> <i>Leerlingen</i>	<i>F</i>	<i>df</i>	<i>p</i> - waarde
Vmbo-b/k	302			
Vmbo-g/t	453	22,42	2;1820	<0,001
Havo/vwo	919			

Tabel 1.2.2.1d Uitkomsten post-hoc variantieanalyses voor de helderheid van de instructie

	Verskil in gemiddelde scores	Gecorrigeerde <i>p</i> -waarde
Vmbo-g/t-vmbo-b/k	-0,159	0,004
Havo/vwo-vmbo-b/k	0,103	0,057
Havo/vwo-vmbo-g/t	0,262	<0,001

### Ordelijkheid van de instructie

Tabel 1.2.2.1e Uitkomsten multivariate variantieanalyse voor de ordelijkheid van de instructie als afhankelijke variabele

	<i>N</i> <i>leerlingen</i>	<i>F</i>	<i>df</i>	<i>p</i> - waarde
Vmbo-b/k	302			
Vmbo-g/t	453	47,55	2;1877	<0,001
Havo/vwo	919			

Tabel 1.2.2.1f Uitkomsten post-hoc variantieanalyses voor de ordelijkheid van de instructie

	Vershil in gemiddelde scores	Gecorrigeerde <i>p</i> -waarde
Vmbo-g/t-vmbo-b/k	0,02	0,907
Havo/vwo-vmbo-b/k	0,322	<0,001
Havo/vwo-vmbo-g/t	0,302	<0,001

### 1.2.2.2 *Leeractiviteiten tijdens de reken-wiskundeles*

#### Docentgestuurde klassikale lesactiviteiten

Tabel 1.2.2.2a Uitkomsten multivariate variantieanalyse voor docentgestuurde klassikale lesactiviteiten als afhankelijke variabele

	<i>N</i> <i>docenten</i>	<i>F</i>	<i>df</i>	<i>p</i> - waarde
Pro, vmbo-b/k	28			
Vmbo-g/t	28	3,998	2;96	0,022
Havo/vwo	44			

Tabel 1.2.2.2b Uitkomsten post-hoc variantieanalyses voor docentgestuurde klassikale lesactiviteiten

	Vershil in gemiddelde scores	Gecorrigeerde <i>p</i> -waarde
Pro, vmbo-b/k-vmbo-g/t	0,283	0,049
Pro, vmbo-b/k-havo/vwo	0,279	0,029
Vmbo-g/t-havo/vwo	-0,005	0,999

### Wiskundig probleemoplossen

Tabel 1.2.2.2c Uitkomsten multivariate variantieanalyse voor wiskundig probleemoplossen als afhankelijke variabele

	<i>N</i> <i>docenten</i>	<i>F</i>	<i>df</i>	<i>p</i> - waarde
Pro, vmbo-b/k	28			
Vmbo-g/t	28	3,467	2;97	0,035
Havo/vwo	44			

Tabel 1.2.2.2d Uitkomsten post-hoc variantieanalyses voor wiskundig probleemoplossen

	Vershil in gemiddelde scores	Gecorrigeerde <i>p</i> -waarde
Pro, vmbo-b/k-vmbo-g/t	-0,321	0,250
Pro, vmbo-b/k-havo/vwo	-0,477	0,027
Vmbo-g/t-havo/vwo	-0,156	0,668

### 1.2.2.3 Zicht op ontwikkeling en differentiatie

#### De evaluatie van wiskundeprestaties en het leerproces

Tabel 1.2.2.3a Uitkomsten multivariate variantieanalyse voor de intensiteit van evaluatie van prestaties en leerproces als afhankelijke variabele

	<i>N</i> <i>docenten</i>	<i>F</i>	<i>df</i>	<i>p</i> - waarde
Pro, vmbo-b/k	28			
Vmbo-g/t	28	2,775	2;97	0,067
Havo/vwo	44			

#### Differentiatie

Tabel 1.2.2.3b Uitkomsten multivariate variantieanalyse voor differentiatie als afhankelijke variabele

	<i>N</i> <i>docenten</i>	<i>F</i>	<i>df</i>	<i>p</i> - waarde
Pro, vmbo-b/k	27			
Vmbo-g/t	28	3,998	2;96	0,022
Havo/vwo	44			



Tabel 1.2.2.3c Uitkomsten post-hoc variantieanalyses voor differentiatie

	Vershil in gemiddelde scores	Gecorrigeerde p-waarde
Pro, vmbo-b/k-vmbo-g/t	0,283	0,049
Pro, vmbo-b/k-havo/vwo	0.279	0,029
Vmbo-g/t-havo/vwo	-0,005	0,999

1.2.2.4 *Context: Veranderingen na coronapandemie schoolleiders*

**Verandering in leergedrag**

Tabel 1.2.2.4a Uitkomsten multivariate variantieanalyse voor verandering in leergedrag als afhankelijke variabele

	<i>N</i> <i>schoolleiders</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Pro/vmbo	18			
Breed	30	1,04	2;58	0,36
Havo/vwo	13			

**Verandering in leerlinggedrag**

Tabel 1.2.2.4b Uitkomsten multivariate variantieanalyse voor verandering in leerprestaties als afhankelijke variabele

	<i>N</i> <i>schoolleiders</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Pro/vmbo	18			
Breed	30	3,936	2;58	<0,025
Havo/vwo	13			

Tabel 1.2.2.4c Uitkomsten post-hoc variantieanalyses voor verandering in leerprestaties

	Vershil in gemiddelde scores	Gecorrigeerde p-waarde
Pro/vmbo-breed	-0,457	0,029
Pro/vmbo-havo/vwo	-0.108	0,867
Havo/vwo-breed	-0.349	0,178

### Verandering in leerprestaties

Tabel 1.2.2.4d Uitkomsten multivariate variantieanalyse voor verandering in leerlinggedrag als afhankelijke variabele

	<i>N</i> <i>schoolleiders</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Pro/vmbo	18			
Breed	30	0,875	2;58	0,422
Havo/vwo	13			

### 1.2.2.5 Context: Veranderingen na coronapandemie docenten

#### Verandering in leergedrag

Tabel 1.2.2.5a Uitkomsten multivariate variantieanalyse voor verandering in leergedrag als afhankelijke variabele

	<i>N</i> <i>docenten</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Pro, vmbo-b/k	15			
Vmbo-g/t	13	0,618	2;52	0,543
Havo/vwo	27			

#### Verandering in leerlinggedrag

Tabel 1.2.2.5b Uitkomsten multivariate variantieanalyse voor verandering in leerlinggedrag als afhankelijke variabele

	<i>N</i> <i>docenten</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Pro, vmbo-b/k	15			
Vmbo-g/t	13	5,127	2;52	0,009
Havo/vwo	27			

Tabel 1.2.2.5c Uitkomsten post-hoc variantieanalyses voor verandering in leerlinggedrag

	Verskil in gemiddelde scores	Gecorrigeerde <i>p</i> -waarde
Pro, vmbo-b/k-vmbo-g/t	-0.008	0,999
Pro, vmbo-b/k-havo/vwo	0.565	0,027
Vmbo-g/t-havo/vwo	0.573	0,034

## Veranderingen in leerprestaties

Tabel 1.2.2.5d Uitkomsten multivariate variantieanalyse voor verandering in leerprestaties als afhankelijke variabele

	<i>N</i> <i>docenten</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Pro, vmbo-b/k	15			
Vmbo-g/t	13	1,37	2;52	0,263
Havo/vwo	27			

## 2 Attituden en achtergrondkenmerken

### 2.1 Databronnen

In de leerlingvragenlijst zijn naast vragen over het onderwijsleerproces ook vragen opgenomen over onder andere plezier in wiskunde, interesse in wiskunde en het belang van wiskunde. Aan docenten zijn vragen gesteld over hun zelfvertrouwen in didactische vaardigheden en het hebben van een fixed of growth mindset. Tot slot zijn schoolleiders bevraagd over de mate van prestatiegerichtheid die ze ervaren bij docenten, leerlingen en ouders.

Meer informatie over de vragenlijst en de respons staat in hoofdstuk 1.1.

### 2.2 Werkwijze

De beschrijvingen in het hoofdstuk van het publieksrapport betreffen overwegend frequentieverdelingen, in percentages en schaalscores. De gegevens zijn apart beschreven voor de verschillende onderwijssoorten en het verschil tussen de onderwijssoorten is getoetst op significantie. In het technisch rapport van het consortium (Van Langen et al., 2023) en in vervolganalyses in hoofdstuk 4 wordt gewerkt met schaalscores. Voor dit hoofdstuk zijn gemiddeldes berekend op basis van de schaalscores. Daarmee kunnen de scores op de schaal beter worden geïd.

Voor de significantietoetsing van de schaalscores is een tweezijdig significantieniveau van 0,05 gehanteerd. Bij een significant verschil is een post-hoc analyse uitgevoerd.

#### 2.2.1 *Nieuwe variabelen*

Het consortium heeft schaalscores berekend op basis van itemscores, zoals beschreven in hoofdstuk 6 van hun technische rapportage (Van Langen et al., 2023). Deze schaalscores worden ook voor dit hoofdstuk gebruikt om verschillen tussen de onderwijssoorten te toetsen. Ook bij het hoofdstuk over verschillen tussen scholen en leerlingen (hoofdstuk 4) worden deze schaalscores gebruikt.

Om de scores van leerlingen op de schalen beter te kunnen duiden, zijn voor dit hoofdstuk ook schaalscores op basis van gemiddeldes berekend. Hiervoor is het gemiddelde berekend van de scores op alle onderliggende items voor leerlingen die minstens de helft van deze items hebben beantwoord. Sommige items zijn voorafgaand hieraan gehercodeerd (4=1, 3=3, 2=3, 1=4) zodat de antwoorden op het item inhoudelijk overeenstemmen met de richting van de schaal. De resulterende schalen lopen van 1 (laag/negatief) tot en met 4 (hoog/positief), of in het geval van prestatiegerichtheid van het schoolklimaat van 1 (zeer laag) tot 5 (heel hoog).

2.2.2 Resultaten en toetsing

2.2.2.1 Leerlingkenmerken

**Plezier in wiskunde**

Tabel 2.2.2.1a Uitkomsten multivariate variantieanalyse voor plezier in wiskunde als afhankelijke variabele

	<i>N</i> <i>leerlingen</i>	<i>F</i>	<i>df</i>	<i>p</i> - waarde
Vmbo-b/k	356			
Vmbo-g/t	527	2,1856	2;1856	0,0398
Havo/vwo	976			

Tabel 2.2.2.1b Uitkomsten post-hoc variantieanalyses voor plezier in wiskunde

	Vershil in gemiddelde scores	Gecorrigeerde <i>p</i> -waarde
Vmbo-b/k- havo/vwo	-0.081	0,170
Vmbo-g/t-vmbo-b/k	-0.008	0,987
Vmbo-g/t-havo/vwo	-0.089	0,062

**Belang van wiskunde**

Tabel 2.2.2.1c Uitkomsten multivariate variantieanalyse voor belang van wiskunde als afhankelijke variabele

	<i>N</i> <i>leerlingen</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Vmbo-b/k	349			
Vmbo-g/t	511	6,214	2;1786	0,002
Havo/vwo	929			

Tabel 2.2.2.1d Uitkomsten post-hoc variantieanalyses voor belang van wiskunde

	Vershil in gemiddelde scores	Gecorrigeerde <i>p</i> -waarde
Vmbo-b/k-havo/vwo	-0,134	0,004
Vmbo-g/t-vmbo-b/k	0.044	0,605
Vmbo-g/t-havo/vwo	-0.09	0,041

**Zelfvertrouwen in wiskunde**

Tabel 2.2.2.1e Uitkomsten multivariate variantieanalyse voor zelfvertrouwen in wiskunde als afhankelijke variabele

	<i>N</i> <i>leerlingen</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Vmbo-b/k	339			
Vmbo-g/t	521	7,132	2;1789	<0,001
Havo/vwo	932			

Tabel 2.2.2.1f Uitkomsten post-hoc variantieanalyses voor zelfvertrouwen in wiskunde

	Verskil in gemiddelde scores	Gecorrigeerde p-waarde
Vmbo-b/k-havo/vwo	-0,177	<0,001
Vmbo-g/t-vmbo-b/k	0.083	0,276
Vmbo-g/t-havo/vwo	-0.094	0,069

### Attributie wiskundevaardigheid

Tabel 2.2.2.1e Uitkomsten multivariate variantieanalyse voor attributie wiskundevaardigheid als afhankelijke variabele

	<i>N</i> <i>leerlingen</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Vmbo-b/k	349			
Vmbo-g/t	511	8,315	2;1936	<0,001
Havo/vwo	929			

Tabel 2.2.2.1f Uitkomsten post-hoc variantieanalyses voor attributie wiskundevaardigheid

	Verskil in gemiddelde scores	Gecorrigeerde p-waarde
Vmbo-b/k-havo/vwo	-0,011	0,96
Vmbo-g/t-vmbo-b/k	-0.128	0,011
Vmbo-g/t-havo/vwo	-0.139	<0,001

### Growth mindset wiskunde

Tabel 2.2.2.1g Uitkomsten multivariate variantieanalyse voor growth mindset als afhankelijke variabele

	<i>N</i> <i>leerlingen</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Vmbo-b/k	364			
Vmbo-g/t	533	11,5	2;1894	<0,001
Havo/vwo	1000			

Tabel 2.2.2.1h Uitkomsten post-hoc variantieanalyses voor growth mindset

	Verskil in gemiddelde scores	Gecorrigeerde p-waarde
Vmbo-b/k-havo/vwo	-0,145	<0,001
Vmbo-g/t-vmbo-b/k	0.0003	0,999
Vmbo-g/t-havo/vwo	-0.144	<0.001

2.2.2.2 Docentkenmerken

**Zelfvertrouwen in didactische vaardigheden**

Tabel 2.2.2.2a Uitkomsten multivariate variantieanalyse voor zelfvertrouwen in didactische vaardigheden als afhankelijke variabele

	<i>N</i> <i>docenten</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Pro, vmbo-b/k	28			
Breed	28	1,945	2;97	0,149
Havo/vwo	44			

**Growth mindset**

Tabel 2.2.2.2b Uitkomsten multivariate variantieanalyse voor growth mindset als afhankelijke variabele

	<i>N</i> <i>docenten</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Pro, vmbo-b/k	27			
Breed	28	0,212	2;96	0,809
Havo/vwo	44			

2.2.2.3 *Prestatiegerichtheid*

**Totaal**

Tabel 2.2.2.3a Uitkomsten multivariate variantieanalyse voor prestatiegerichtheid als afhankelijke variabele

	<i>N</i> <i>schoolleiders</i>	<i>F</i>	<i>df</i>	<i>p</i> - waarde
Pro/vmbo	30			
Breed	47	8,95	2;98	<0,001
Havo/vwo	24			

Tabel 2.2.2.3b Uitkomsten post-hoc variantieanalyses voor prestatiegerichtheid

	Vershil in gemiddelde scores	Gecorrigeerde <i>p</i> -waarde
Pro/vmbo-breed	-0,171	0,066
Pro/vmbo-havo/vwo	-0,375	0,036
Havo/vwo-breed	0,204	<0,001

## Docenten

Tabel 2.2.2.3c Uitkomsten multivariate variantieanalyse voor prestatiegerichtheid van docenten als afhankelijke variabele

	<i>N</i> <i>schoolleiders</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Pro/vmbo	30			
Breed	47	0,914	2;98	0,404
Havo/vwo	24			

## Ouders

Tabel 2.2.2.3d Uitkomsten multivariate variantieanalyse voor prestatiegerichtheid van ouders als afhankelijke variabele

	<i>N</i> <i>schoolleiders</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Pro/vmbo	30			
Breed	47	13,03	2;98	<0,001
Havo/vwo	24			

Tabel 2.2.2.3e Uitkomsten post-hoc variantieanalyses voor prestatiegerichtheid van ouders

	Vershil in gemiddelde scores	Gecorrigeerde <i>p</i> -waarde
Pro/vmbo-breed	-0.387	0.004
Pro/vmbo-havo/vwo	-0.700	0,002
Havo/vwo-breed	0.313	<0,001

## Leerlingen

Tabel 2.2.2.3f Uitkomsten multivariate variantieanalyse voor prestatiegerichtheid van leerlingen als afhankelijke variabele

	<i>N</i> <i>leerlingen</i>	<i>F</i>	<i>df</i>	<i>p</i> -waarde
Pro/vmbo	30			
Breed	47	6,107	2;98	0,00317
Havo/vwo	24			

Tabel 2.2.2.3g Uitkomsten post-hoc variantieanalyses voor prestatiegerichtheid van leerlingen

	Vershil in gemiddelde scores	Gecorrigeerde <i>p</i> -waarde
Pro/vmbo-breed	-0.230	0,062
Pro/vmbo-havo/vwo	-0.406	0,002



	Vershil in gemiddelde scores	Gecorrigeerde p-waarde
Havo/vwo-breed	0.176	0.239

### 3 Prestaties rekenen en wiskunde

In dit hoofdstuk van het publieksrapport beschrijven we wat leerlingen aan het einde van het tweede leerjaar van het voortgezet onderwijs kennen en kunnen op het gebied van Rekenen en wiskunde. We zoomen daarbij in op de verschillende domeinen van reken- en wiskundeprestaties die 'typische' laagvaardige (P10), hoogvaardige (P90), 1F-, 2F-, 3F en 1S-leerlingen kenmerken. We maken een globale vergelijking met de reken- en wiskundeprestaties van po-leerlingen die zijn gemeten aan het einde van het (speciaal) basisonderwijs in het peilingsonderzoek van 2018/2019 (Buisman et al., 2020).

#### 3.1 Databronnen

##### 3.1.1 *Rekenen en wiskunde meten*

Om de vaardigheden rekenen en wiskunde te kunnen meten is een reken-wiskundetoets samengesteld met in totaal 264 opgaven uit beveiligde itembanken voor vier rekendomeinen: getallen, verhoudingen, meten en meetkunde, en verbanden. Bovendien zijn bestaande opgaven over wiskundig probleemoplossen aan de toets toegevoegd. Voor deze eerste rekenen en wiskunde peiling in het vo kon geen tijd worden ingeruimd voor daadwerkelijke toetsontwikkeling. De gebruikte rekenopgaven zijn voornamelijk opgaven die aansluiten bij de referentieniveaus (1F, 2F, 3F en 1S).

Andere wiskundige denk- en werkwijzen zijn nu niet in de toets opgenomen. De reden voor deze beperking is dat er in april 2021 bij de start van Peil.onderwijs vo door de direct belanghebbenden bij de peilingsresultaten is verzocht om de eerste peiling reeds 1 jaar later te laten plaatsvinden (in voorjaar 2022) en niet 2 jaar later in 2023. In het laatste geval zou er voldoende ontwikkeltijd zijn geweest voor de toets, maar zouden de eerste peilingsresultaten niet op tijd beschikbaar zijn voor gebruik door de onderwijspraktijk, het landelijk onderwijsbeleid, het toezicht en andere doelgroepen van de peilingen.

De reken-wiskundetoets bevatte zowel open als gesloten opgaven, die verschillende aspecten van rekenen en wiskunde toetsten voor de vier rekendomeinen en wiskundig probleemoplossen. De 264 opgaven uit de reken- wiskundetoets zijn ondergebracht in 12 toetsversies, die van elkaar verschilden in moeilijkheidsgraad, maar wel deels overlappende opgaven bevatten. Leerlingen kregen een toetsversie toebedeeld op basis van het vooraf bekende onderwijssoort voortgezet onderwijs van de klas van de betreffende leerling. Tabel 3.1.1a geeft een overzicht van het aantal opgaven per subdomein en rekenonderwerp die in de reken- wiskundetoets aan de orde kwamen.

Tabel 3.1.1a Verdeling van de opgaven naar type opgave

	<b>Type opgaven</b>		<b>Totaal aantal opgaven</b>
Domeinen	Rekenen	Getallen	73
		Verhoudingen	72
		Meten en meetkunde	49
		Verbanden	48
	Wiskunde	Wiskundig probleemoplossen	22
Referentieniveaus		1F-niveau	65
		1S-niveau	50
		2F-niveau	107
		3F-niveau	42
Met of zonder rekenmachine		RM	103
		ZR	161
Context versus kaal*		Context	234
		Kaal	29

\*De opdeling context versus kale opgaven bevat 1 item minder, omdat bij 1 opgave niet aangegeven is of dit een contextopgave of kale opgave is

### 3.1.2 *Prestaties rekenen en wiskunde*

#### *Verwachte ruwe scores*

Vanwege de deels verschillende versies van de reken-wiskunde toetsboekjes, zijn de resultaten van de leerlingen niet rechtstreeks te vergelijken. Om vergelijkingen toch mogelijk te maken, zijn de opgaven en de leerlingen op 1 onderliggende meetschaal geplaatst (met behulp van itemresponsetheoriemodellen (IRT), zie Van Langen et al., 2023). Daardoor konden voor alle leerlingen schaalscores worden berekend die onderling wél vergelijkbaar zijn. Op basis van deze schaalscores (ofwel vaardigheidsscores) is geschat welke score leerlingen zouden hebben behaald als ze de gehele set aan opgaven (264 in totaal) hadden gemaakt (getransformeerde WML-scores): de verwachte score.

In het publieksrapport wordt zo veel mogelijk gewerkt met deze verwachte score, die loopt van 12 (minimum score) tot en met 376 (maximum score).

#### *Percentage beheersing*

Om vergelijkingen tussen groepen leerlingen op verschillende rekendomeinen en type opgaven (kaal en context opgaven en opgaven met en zonder rekenmachine) mogelijk te maken, zijn op basis van de verwachte scores ook beheersingsniveaus per onderdeel berekend. Dit is gedaan door de verwachte ruwe score te delen door de maximaal haalbare score op dat onderdeel of met dat type opgaven. Deze 'proportie beheersing' is vervolgens omgezet naar een percentage.

#### *Vaardigheid*

In het rapport laten we per domein gemiddelde beheersingsniveaus zien voor leerlingen aan het einde van het tweede leerjaar van het voortgezet onderwijs.

Omgekeerd kunnen we ook onderzoeken welke vaardigheid vereist is voor het correct beantwoorden van elk van de opgaven uit de reken-wiskundetoets. De linialen in het publieksrapport zijn gebaseerd op de onderliggende vaardigheidsschaal. Voor elke opgave is uitgerekend hoeveel vaardigheid nodig is om 50% of 80% kans te hebben op het correct maken van de opgave. Bij 80% of meer kans om een item juist te maken is er sprake van een erg goede beheersing. In het publieksrapport spreken we van het beheersen van een opgave als een leerling een kans van 50% of meer heeft om deze goed te beantwoorden. We kiezen voor een grens van 50% of meer beheersing (en dus niet van 80% of meer beheersing), omdat het dan mogelijk is om prestaties van verschillende groepen leerlingen goed te onderscheiden.

## **3.2 Werkwijze**

De werkwijze om te komen tot beantwoording van de onderzoeksvragen wordt voor een groot deel beschreven in het technisch rapport van het uitvoerend consortium (Van Langen et al., 2023). In de hiernavolgende paragrafen beschrijven we de aanvullende analyses die door de inspectie zijn gedaan en geven we aanvullingen op of beschrijven we hoe we afwijken van de reeds beschreven werkwijze.

### **3.2.1 *Significantie toetsing***

In het rapport is er getoetst of prestaties van groepen leerlingen (de onderwijssoorten: pro, vmbo-b/k, vmbo-g/t, havo/vwo en de referentieniveaus: <1F, 1F, 2F, 1S, 3F) verschillen tussen a) rekendomeinen en wiskundig probleemoplossen, b) opgaven met en zonder context en c) opgaven met en zonder de rekenmachine. In de hiernavolgende paragrafen wordt beschreven welke analysemethoden zijn gebruikt en worden de uitkomsten van de analyses samengevat.

#### *De vier rekendomeinen en wiskundig probleemoplossen*

In het publieksrapport beschrijven we hoe leerlingen in het tweede leerjaar van het voortgezet onderwijs presteren op de verschillende rekendomeinen: getallen, verhoudingen, meten en meetkunde, verbanden en wiskundig probleemoplossen. Omdat de prestaties op de verschillende domeinen niet rechtstreeks te vergelijken zijn (ze bevatten opgaven die – op andere vlakken dan het domein dat ze toetsen – van moeilijkheidsgraad zouden kunnen verschillen), vergelijken we relatieve beheersingsniveaus (percentage beheersing) tussen groepen leerlingen: pro- versus vmbo-b/k- versus vmbo-g/t- versus havo/vwo-leerlingen en de referentieniveaus: <1F- versus 1F- versus 2F- versus 1S- versus 3F-leerlingen. We hebben hierbij ook de pro-leerlingen als groep meegenomen, maar merk op dat dit een hele kleine, niet representatieve groep betreft.

#### *Beheersing van de vier domeinen binnen elke onderwijssoort*

Om verschillen in relatieve beheersingsniveaus te toetsen, is gebruik gemaakt van een ANOVA voor herhaalde metingen met als 'within-subject variabele' de vier rekendomeinen en wiskundig probleemoplossen en als 'between-subjects variabele' onderwijssoort (pro, vmbo-b/k, vmbo-g/t, havo/vwo).

Tabel 3.2.1a geeft de uitkomst van deze analyse. Het significante interactie-effect laat zien dat het effect van rekendomein (op het percentage beheersing) verschilt tussen de onderwijssoorten. Het is niet mogelijk om via post hoc toetsen rechtstreeks te

achterhalen op welk domein beheersingsniveaus significant van elkaar verschillen. Havo/vwo-leerlingen presteren immers beter dan pro-leerlingen en dat geldt voor alle domeinen. Daarom wordt de beheersing van de verschillende rekendomeinen relatief getoetst: op welk domein zijn leerlingen van verschillende schooltypen het meest vaardig en op welk domein het minst? Voor dit doel werd eerst met een repeated measures ANOVA per onderwijssoort het effect van de 'within subjects variabele' *rekendomein op het beheersingsniveau* geanalyseerd. Voor alle onderwijssoorten geldt dat het hoofdeffect van rekendomein significant is (zie Tabel 3.2.2a, model 1).

Tabel 3.2.1a Interacties domein x onderwijssoort

Model		F	df	p-waarde
1	Rekendomein x onderwijssoort (pro, vmbo-b/k, vmbo-g/t, havo/vwo)	31,96	(12)	<0,001
2	Rekendomein in pro*	2,79	(4)	0,0264
3	Rekendomein in vmbo-bk	47,91	(4)	<0,001
4	Rekendomein in vmbo-gt	25,14	(4)	<0,001
5	Rekendomein in havo/vwo	18,91	(4)	<0,001

\*Het gaat hier om kleine aantallen (n=78)

Daarna werden post hoc t-toetsen binnen elke onderwijssoort uitgevoerd met Bonferroni correctie om te analyseren waar er verschillen in beheersing bestaan. Binnen de onderwijssoorten, verschilde de beheersing op de verschillende rekendomeinen significant (alle p's <0,001) met uitzondering van:

Pro: getallen versus meten en meetkunde; verhoudingen versus verbanden; verhoudingen versus wiskundig probleemoplossen; en meten en meetkunde versus verbanden;

Vmbo-g/t: meten en meetkunde versus verbanden;

Havo/vwo: getallen versus verbanden.

#### *Beheersing van de vier domeinen binnen elke referentieniveaugroep*

Eenzelfde analyse werd uitgevoerd per referentieniveaugroep (<1F, 1F, 2F, 1S en 3F). In tabel 2.2.2b (model 1) is te zien dat er een significant interactie is tussen de rekendomeinen en de referentieniveaugroep. Dit laat zien dat het effect van het rekendomein op het percentage beheersing verschilt per referentieniveaugroep. Voor dit doel werd eerst met een repeated measures ANOVA per referentieniveaugroep het effect van de 'within subjects variabele' *rekendomein op het beheersingsniveau* geanalyseerd. Alleen voor de 3F-groep blijkt dat er geen significant hoofdeffect van het rekendomein op het beheersingsniveau is (zie Tabel 3.2.1b, model 6). Dit betekent dus dat alleen voor de 3F-groep, er geen verschillen zijn tussen de beheersing van rekendomeinen.

Tabel 3.2.1b Interacties domein x referentieniveau

Model		F	df	p-waarde
1	Domein x referentieniveaugroep (onder 1F/1F/2F/3F/1S)	13,24	(16)	<0,001
2	Onder 1F	43,56	(4)	<0,001
3	1F	27,33	(4)	<0,001
4	2F	9,31	(4)	<0,001
5	1S	15,33	(4)	<0,001
6	3F	0,98	(4)	0,416

Vervolgens werden post hoc t-toetsen binnen de andere referentieniveaus (<1F, 1F, 2F en 1S) uitgevoerd met Bonferroni correctie om te analyseren waar er verschillen in beheersing bestaan. Uit deze analyse blijken significante verschillen tussen de beheersing op de rekendomeinen binnen de referentieniveaus (alle p's <0,001) met uitzondering van:

<1F: meten en meetkunde versus verhoudingen;

2F: getallen versus verhoudingen; meten en meetkunde versus wiskundig probleemoplossen; en verbanden versus verhoudingen;

1S: getallen versus verbanden.

Leerlingen die op de totale rekenvaardigheidsschaal het 2F-niveau beheersen, presteren significant beter op het domein getallen dan op het domein verhoudingen, beter op meten en meetkunde dan op wiskundig probleemoplossen en beter op verbanden dan op verhoudingen.

#### *Rekenen en wiskunde in context*

Ook hier geldt dat we de contextopgaven en de kale opgaven niet direct kunnen vergelijken, maar alleen tussen groepen leerlingen. We hebben namelijk niet dezelfde opgaven met en zonder context voorgelegd aan de leerlingen. Daardoor zou de moeilijkheid van de kale- en contextopgaven kunnen verschillen, bijvoorbeeld door de in de opgave gebruikte getallen en het aantal bewerkingen. We toetsen verschillen in relatieve beheersingsniveaus (percentage beheersing) tussen verschillende onderwijssoorten en de verschillende referentieniveaugroepen. We maken hiervoor gebruik van een reeks ANOVA's voor herhaalde metingen met opgavetype (context vs. kaal) als 'within subjects' en de variabele onderwijssoort (pro, vmbo-b/k, vmbo-g/t en havo/vwo) en de variabele referentieniveaugroep (<1F, 1F, 2F, 1S en 3F) als 'between subjects'. Tabellen 3.2.1c en 3.2.1d geven de resultaten van deze analyses.

Tabel 3.2.1c Interacties context van de opgave (context versus kale opgave) X onderwijssoort

Model	Context (kaal versus contextopgave) X onderwijssoort	F	df	p-waarde
1	Context x onderwijssoort (pro, vmbo-bk, vmbo-gt, havo/vwo)	30,19	(3)	<0,001
2	Context in pro*	20,65	(4)	<0,001
3	Context in vmbo-bk	23,20	(4)	<0,001
4	Context in vmbo-gt	45,70	(4)	<0,001
5	Context in havo/vwo	37,82	(4)	<0,001

\*Het gaat hier om kleine aantallen (n=78)

Binnen alle onderwijssoorten is het beheersingsniveau (percentage beheersing) op de kale opgaven hoger dan op de opgaven met context:

Pro: kale opgaven ( $M = 78,77$ ,  $SD = 32,44$ ) versus opgaven met context ( $M = 26,93$ ,  $SD = 15,65$ );

Vmbo-bk: kale opgaven ( $M = 58,37$ ,  $SD = 22,57$ ) versus opgaven met context ( $M = 42,04$ ,  $SD = 17,76$ );

Vmbo-gt: kale opgaven ( $M = 71,12$ ,  $SD = 27,21$ ) versus opgaven met context ( $M = 49,55$ ,  $SD = 17,99$ );

Havo/vwo: kale opgaven ( $M = 75,95$ ,  $SD = 26,70$ ) versus opgaven met context ( $M = 54,11$ ,  $SD = 17,85$ ).

Tabel 3.2.1d Interacties context van de opgave (context versus kale opgave) X referentieniveaugroep

Model	Context (kaal versus contextopgave) X referentieniveaugroep	F	df	p-waarde
1	Context x 1F, 2F, 1S, 3F)	3,23	(4)	0,01
2	Context in <1F groep	117,6	(1)	<0,001
3	Context in 1F groep	37,42	(1)	<0,001
4	Context in 2F groep	13,42	(1)	<0,001
5	Context in 1S groep	54	(1)	<0,001
6	Context in 3F groep	1,72	(1)	0,192

Binnen de 3F groep verschilt het beheersingsniveau (percentage beheersing) van leerlingen niet significant tussen de opgaven met context en kale opgaven. In alle andere niveaugroepen is de beheersing op de kale opgaven hoger dan op de opgaven met context:  
 <1F kale opgaven ( $M = 56,81$ ,  $SD = 25,23$ ) versus opgaven met context ( $M = 33,18$ ,  $SD = 14,77$ );  
 1F kale opgaven ( $M = 69,34$ ,  $SD = 27,21$ ) versus opgaven met context ( $M = 47,12$ ,  $SD = 15,71$ );  
 2F kale opgaven ( $M = 72,60$ ,  $SD = 29,47$ ) versus opgaven met context ( $M = 47,77$ ,  $SD = 15,22$ );  
 1S kale opgaven ( $M = 84,04$ ,  $SD = 14$ ) versus opgaven met context ( $M = 60,21$ ,  $SD = 20,3$ ).

*Rekenen en wiskunde met en zonder rekenmachine*

In het rapport maken we een vergelijking tussen de opgaven met en zonder de rekenmachine. Hiertoe voerden we een reeks ANOVA's voor herhaalde metingen met opgavetype (met vs. zonder rekenmachine) als 'within subjects' en de variabele onderwijssoort (pro, vmbo-bk, vmbo-gt en havo/vwo) als 'between subjects'. Tabel 3.2.1e geeft de resultaten weer.

Tabel 3.2.1e Interacties type opgave (met versus zonder rekenmachine) x onderwijssoort

Model	Type opgave (met versus zonder rekenmachine) X onderwijssoort	F	df	p-waarde
1	Type opgave X onderwijssoort (pro, vmbo-b/k, vmbo-g/t, havo/vwo)	40,68	(3)	<0,001
2	Type opgave in pro*	7,44	(1)	0,007
3	Type opgave in vmbo-b/k	76,60	(1)	<0,001
4	Type opgave in vmbo-g/t	25,96	(1)	<0,001
5	Type opgave in havo/vwo	13,89	(1)	<0,001

\*Het gaat hier om kleine aantallen (n=78)

Binnen alle onderwijssoorten is het beheersingsniveau (percentage beheersing) op de zonder rekenmachine hoger dan op de opgaven met rekenmachine:  
 Pro: opgaven zonder rekenmachine ( $M = 44,56$ ,  $SD = 18,31$ ) versus opgaven met rekenmachine ( $M = 15,32$ ,  $SD = 16,15$ );  
 Vmbo-bk: opgaven zonder rekenmachine ( $M = 54,15$ ,  $SD = 19,21$ ) versus opgaven met rekenmachine ( $M = 29,57$ ,  $SD = 20,11$ );  
 Vmbo-gt: opgaven zonder rekenmachine ( $M = 57,88$ ,  $SD = 17,81$ ) versus opgaven met rekenmachine ( $M = 44,09$ ,  $SD = 21,96$ );  
 Havo/vwo: opgaven zonder rekenmachine ( $M = 58,03$ ,  $SD = 18,52$ ) versus opgaven met rekenmachine ( $M = 48,54$ ,  $SD = 21,43$ ).

Ook kijken we in het rapport naar de prestaties op de opgaven met en zonder rekenmachine per referentieniveaugroep. Het is immers denkbaar dat de samenhang tussen prestaties op de opgaven met en zonder rekenmachine afhankelijk is van de algemene rekenvaardigheid van de leerlingen. Om de prestaties op de opgaven met en zonder rekenmachine te vergelijken tussen de referentieniveaugroepen, voerden we een repeated measures ANOVA uit met type opgave (met of zonder rekenmachine) als 'within subjects' variabele en de onderwijssoort (pro, vmbo-b/k, vmbo-g/t, havo/vwo) en de referentieniveaugroep (<1F, 1F, 2F, 1S en 3F) als 'between subjects' variabelen. De resultaten van deze analyse laten zien dat er interactie is tussen de referentieniveaugroep en het type opgave (met of zonder rekenmachine),  $F(4,2473) = 97,73$ ,  $p < 0,001$  ( zie tabel 3.2.1f, model 1). Dit betekent dat het relatieve beheersingsniveau van de opgaven met rekenmachine (t.o.v. de opgaven zonder rekenmachine) verschilt tussen referentieniveaugroepen. Om deze reden vergelijken we ook het relatieve beheersingsniveau van de opgaven met en zonder rekenmachine apart voor de referentieniveaugroepen. We maken hiervoor gebruik van een repeated measures ANOVA met opgavetype (met vs. zonder rekenmachine) als de 'within subjects' (zie tabel 3.2.1f).

Tabel 3.2.1f Interacties type opgave (met versus zonder rekenmachine opgave) X referentieniveaugroep

Model	Type opgave (met versus zonder rekenmachine) X referentieniveaugroep	F	df	p-waarde
1	Type opgave X referentieniveaugroep	97,73	(4)	<0,001
2	Type opgave X <1F groep	155,60	(1)	<0,001
3	Type opgave X 1F groep	39,92	(1)	<0,001
4	Type opgave X 2F groep	0,975	(1)	0,324*
5	Type opgave X 1S groep	6,716	(1)	0,01
6	Type opgave X 3F groep	0,513	(1)	0,475

\*In het publieksrapport is hier per abuis een asterisk bij geplaatst

Binnen de <1F-, 1F- en 1S-groep is het beheersingsniveau (percentage beheersing) op de opgaven zonder rekenmachine hoger dan op de opgaven met rekenmachine:  
 <1F: opgaven zonder rekenmachine ( $M = 46,35$ ,  $SD = 17,51$ ) versus opgaven met rekenmachine ( $M = 21,18$ ,  $SD = 15,04$ );  
 1F: opgaven zonder rekenmachine ( $M = 54,72$ ,  $SD = 18,88$ ) versus opgaven met rekenmachine ( $M = 41,11$ ,  $SD = 18,18$ );  
 1S: opgaven zonder rekenmachine ( $M = 63,96$ ,  $SD = 14,64$ ) versus opgaven met rekenmachine ( $M = 54,53$ ,  $SD = 19,06$ ).



### 3.2.2

#### *Bepalen van voorbeelditems*

In het rapport kijken we, per rekendomein en wiskundig probleemoplossen, naar de exacte opgaven uit de rekentoetsen die sommige groepen leerlingen nog *nét* wel en net niet meer beheersen. Dit noemen we de 'grensitems'. Op die manier geven de opgaven een beeld van wat verschillende groepen leerlingen kennen en kunnen. We onderscheiden in het rapport verschillende groepen leerlingen:

- de laagvaardige leerling: de 10% laagst presterende leerlingen in het tweede leerjaar vo (P10);
- de <1F leerling: de leerlingen die onder het fundamentele niveau scoren;
- de 1F-leerling: leerlingen die tussen niveau 1F en niveau 2F presteren;
- de 2F-leerlingen: leerlingen die tussen niveau 2F en 1S presteren;
- de 1S-leerlingen<sup>3</sup>: leerlingen die tussen niveau 1S en 3F presteren;
- de 3F-leerlingen: leerlingen die op niveau 3F of hoger presteren;
- de hoogvaardige leerling: de 90% hoogst presterende leerling in het tweede leerjaar vo (P90).

De P in P10 en P90 staat voor percentiel. Een percentiel geeft aan hoeveel procent van de leerlingen in de populatie de betreffende of een lagere vaardigheidsscore heeft. Ter illustratie: percentiel 10 ligt in de onderliggende vaardigheidsschaal op 124,9. Dit betekent dat 10% van de leerlingen een score heeft van 124,9 of lager en dat 90% van de leerlingen dus een hogere vaardigheidsscore heeft. Dit percentiel is bepaald voor de totale groep van 10% laagst scorende leerlingen in het tweede leerjaar van het vo. We zien dus dat de laagvaardige leerling in deze beschrijving vooral de pro en vmbo-bk -leerling weerspiegelt. Datzelfde geldt voor het P90-punt: kijken we naar de P90-leerling in het vo, dan zien we dat dit vooral de havo/vwo leerlingen betreft.

Voorbeelditems zijn geselecteerd door te kijken naar de beheersing op itemniveau, hetgeen bepaald is met behulp van het item respons model (zie paragraaf 7.3 in het technisch rapport Van Langen et al., 2023). Voor elke opgave staat in het technisch rapport beschreven hoeveel vaardigheid nodig is om 50 of 80 procent kans te hebben op het correct maken van die opgave. Vervolgens is gekeken hoe dat zich verhoudt tot de hoeveelheid vaardigheid die laagvaardige en hoogvaardige leerlingen bezitten en de hoeveelheid vaardigheid die nodig is om referentieniveau 1F, 2F, 1S of 3F te beheersen.

Voor items op de grens van 1F, ligt het punt op de vaardigheidsverdeling waarop voor het item geen beheersing (kans op een correct antwoord kleiner dan 50%) overgaat in matige beheersing (minimaal 50% kans op een goed antwoord) net boven het punt waar ook het cesuurpunt voor 1F ligt op de (lineair getransformeerde) latente vaardigheidsschaal (namelijk op 165,81). Voor items op de grens van 1F en 2F ligt het beheersingspunt net boven het cesuurpunt voor 2F op de vaardigheidsschaal (op 222,81). Voor items op de grens van 2F en 1S ligt het beheersingspunt net boven het cesuurpunt voor 1S op de vaardigheidsschaal (op 235,60). En voor items op de grens van 1S en 3F ligt het beheersingspunt net boven het cesuurpunt voor 3F op de vaardigheidsschaal (op 303,02).

Op de vaardigheidsschaal zijn ook het P10 en P90 cesuurpunt berekend (respectievelijk: 124,90 en 284,10), om zo ook voor de groep laagvaardige

---

<sup>3</sup> Niveau 1S wordt niet aangeboden in het vo, wel in het po. Er zaten wel 1S-opgaven in de reken-wiskundetoets van deze peiling. Van Langen et al. (2023) lichten de cesuurbepaling toe van de 4 referentieniveaus 1F, 2F, 3F en 1S. De cesuur voor 1S bleek tussen de cesuur van 2F en 3F te liggen. Dit neemt niet weg, dat 1S inhouden niet in het vo worden onderwezen.

leerlingen (P10) en hoogvaardige leerlingen (P90) grensitems te bepalen. Waar het niet mogelijk was om een grensitem te selecteren (bijvoorbeeld omdat het item behoorde tot de niet openbare referentieset) is voor een nabijgelegen item op de vaardigheidsschaal gekozen.

### 3.2.3 *Globale vergelijking met (s)bo*

In het rapport wordt de reken- en wiskundevaardigheid van leerlingen in het huidige peilingsonderzoek vergeleken met de reken-wiskundevaardigheid zoals gemeten in het peilingsonderzoek rekenen-wiskunde eind groep 8 van het (s)bo (Inspectie van het Onderwijs, 2021). Die vergelijking is alleen globaal mogelijk omdat in de verschillende peilingsonderzoeken slechts deels dezelfde opgaven zijn opgenomen.

De groep 8-leerlingen die in het (s)bo in schooljaar 2018-2019 deelnamen aan het peilingsonderzoek betrof een andere groep leerlingen dan de leerlingen die 3 jaar later in het schooljaar 2021-2022 in het vo aan de peiling hebben deelgenomen. Verder is de normering van de afgenomen toets gebaseerd op de referentieset rekenen voor het po (zie paragraaf 3.1.1 van het publieksrapport). Bovendien moet worden opgemerkt dat we hier te maken hebben met het verschil in aanbod van de onderbouw in het vo en het aanbod aan het einde van het (s)bo. In de vo-onderbouw wordt geen onderwijs gegeven over 1F- (met uitzondering van pro en vmbo-b) en 1S-doelen. In dit peilingsonderzoek is niet expliciet nagegaan of aangeleerde doelen uit po onderhouden worden in het vo.

Voor deze trendanalyse is gebruik gemaakt van de analyses zoals gerapporteerd in Van Langen et al. (2023; zie hoofdstuk 7.2). Eerst zijn de leerlingsscores uit de verschillende peilingen met behulp van Item Response Theorie (IRT) op één meetschaal geplaatst om ze vergelijkbaar te maken. Vervolgens is de reken-wiskundevaardigheidsverdeling van de referentieniveaus van po-leerlingen in 2019 vergeleken met de reken- en wiskundevaardigheidsverdeling van de referentieniveaus van vo-leerlingen in 2022. Deze vergelijking laat zien dat er sprake is van een vergelijkbaar patroon van bereikte referentieniveaus in po en vo: het aandeel leerlingen dat niveau 1F nog niet heeft bereikt aan het einde van leerjaar 2 in het vo is ongeveer even groot als aan het einde van (s)bo (vo 25%; (s)bo 20%).

## 4 Verschillen in reken- en wiskundeprestaties

In dit hoofdstuk van het publieksrapport maken we een koppeling tussen de leerlingprestaties, de kenmerken van het onderwijsleerproces en de (achtergrond)kenmerken van leerlingen, docenten en scholen. Daarmee proberen we een antwoord te vinden op de volgende vragen:

- Hoe groot zijn de verschillen tussen klassen en leerlingen in reken- en wiskunde prestaties?
- Met welke algemene en aan rekenen en wiskunde gerelateerde kenmerken van leerlingen, docenten, het onderwijsleerproces en scholen hangen deze prestatieverschillen samen?

Voor het beantwoorden van bovenstaande vragen, baseren we ons op de analyses die door het uitvoerend consortium zijn uitgevoerd (zie hoofdstuk 8 in Van Langen et al., 2023). In deze analyses zijn gegevens op alle niveaus gekoppeld en vervolgens zijn ontbrekende waarden geïmputeerd met behulp van het pakket Mice binnen het programma R (Van Buuren & Groothuis-Oudshoorn, 2011). Voor de analyses naar samenhangen zijn separate modellen gemaakt per predictorniveau (leerlingen, docenten en school), met de gestandaardiseerde (tbv vergelijkbaarheid) 'weighted maximum likelihood' (WML) schatter als uitkomstmaat. Hoofdstuk 8.5 in Van Langen et al., (2023) geeft een nadere beschrijving van de gehanteerde strategie om tot een compact en spaarzaam totaalmodel te komen.

## 5 Referenties

Buisman, M., Kuijper, S., Hickendorff, M., Kuijpers, R.E., Keuning, J., Walet, L., Ankersmit, M. & Kronenburg, F. van (2020). Rekenvaardigheden in het (S)bo. Technisch rapport Peil.onderwijs Rekenen-Wiskunde einde (speciaal) basisonderwijs 2019. Amsterdam: Kohnstamm Instituut.

Inspectie van het Onderwijs (2021). *Peil.Rekenen-Wiskunde Einde (speciaal) basisonderwijs 2018-2019, eindrapport*. Inspectie van het Onderwijs.

Inspectie van het Onderwijs (2024). *Peil.Rekenen en Wiskunde einde leerjaar 2 voortgezet onderwijs 2021-2022, eindrapport*. Inspectie van het Onderwijs.

Langen, A. van, Jenniskens, T., Gubbels, J., Hemker, B, Meelissen, M., Rekers-Mombarg, L., Swart, N. & Vermeulen, J. (2023). *Rekenen-Wiskunde in het VO Technisch rapport Peil.onderwijs. Meting 2022-leerjaar 2*. KBA/ExpertiseCentrum Nederlands/Cito B.V./Universiteit Twente.

R Core Team (2024). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in R. *Journal of statistical software*, 45, 1-67.

## Colofon

Inspectie van het Onderwijs  
Postbus 2730 | 3500 GS Utrecht  
[www.onderwijsinspectie.nl](http://www.onderwijsinspectie.nl)

jaar-inspectienummer | gratis

Een exemplaar van deze publicatie is te downloaden vanaf de website van de  
Inspectie van het Onderwijs: [www.onderwijsinspectie.nl](http://www.onderwijsinspectie.nl).

© Inspectie van het Onderwijs | februari 2024