

Deelrapport 2

Bronnen van externe kwaliteitszorg als hefboom voor onderwijskwaliteit

Auteurs:

Lore Pelgrims, Jonas Dockx en Koen Aesaert

OBPWO 21.01

Mei 2023

Onderzoek uitgevoerd in opdracht van het
Departement Onderwijs & Vorming van de Vlaamse Overheid

Inhoudstafel

Lexicon	7
Inleiding	8
Leeswijzer	9
Module 2.1: Gelijkenissen en verschillen tussen bronnen van externe kwaliteitszorg	11
1. Onderzoeksopzet.....	11
1.1 Geanalyseerde bronnen.....	11
1.2 Analyse kader.....	11
1.3 Aanpak van de documentanalyse	12
1.4 Matrix.....	12
2. Horizontale analyse	13
2.1 Design.....	13
2.2 Standaardbepaling	18
2.3 Analytische methode	19
2.4 Waar komen die gelijkenissen en verschillen vandaan?	21
3. Matrix	24
3.1 De matrix lezen en interpreteren	24
3.2 Wat zijn de mogelijkheden en beperkingen om uitspraken te doen over specifieke aspecten van onderwijskwaliteit?	25
3.3 Besluit	44
Module 2.2: Mate van convergentie of divergentie tussen de resultaten van verschillende bronnen.....	45
Studie 1: PIRLS Repeat.....	45
1. Inleiding	45
2. Onderzoeksopzet.....	45
2.1 PIRLS- en peilingschaal (Vergelijking 1).....	45
2.2 PIRLS, peiling en doorlichtingen (Vergelijking 2)	47
3. Resultaten PIRLS- en peilingschaal (Vergelijking 1).....	48
3.1 Modellen op de PIRLS- en peilingschaal	48
3.2 Schooleffecten volgens de PIRLS- en peilingschaal	52
4. Resultaten PIRLS, peiling en doorlichtingen (Vergelijking 2).....	66
4.1 Schooleffecten (PIRLS- en peilingschaal) en inschaling bij de doorlichtingen.....	66
4.2 Behalen van de eindtermen (peilingschaal) en inschaling bij de doorlichtingen	74
4.3 Posities binnen de benchmarks op de PIRLS-schaal en inschaling bij de doorlichtingen	75

Studie 2: Scholen met meer dan één recente bron van kwaliteitsindicatoren	77
1. Inleiding	77
2. Onderzoeksopzet.....	77
2.1 Doorlichtingen en internationaal vergelijkend onderzoek (Vergelijking 1).....	79
2.2 Doorlichtingen en peilingsonderzoek (Vergelijking 2).....	79
2.3 Internationaal vergelijkend onderzoek en peilingsonderzoek (Vergelijking 3)	80
3. Resultaten.....	80
3.1 Leeswijzer voor de resultaten	80
3.2 Hoofdbevindingen	81
Studie 3: Vergelijking van trends op systeemniveau	91
1. Inleiding	91
2. Onderzoeksopzet.....	91
2.1 Doorlichtingen van de onderwijsinspectie	91
2.2 Trends vergelijken tussen het internationaal vergelijkend onderzoek en het peilingsonderzoek.....	94
3. De doorlichtingen doorheen de tijd	100
4. Vergelijking van trends op basis van het peilingsonderzoek en het internationaal vergelijkend onderzoek.....	102
4.1 Leeswijzer voor de resultaten	102
4.2 Hoofdbevindingen	104
Conclusie over de drie studies heen	107
1. Inleiding	107
2. Mate van convergentie of divergentie op schoolniveau	107
3. Mate van convergentie of divergentie op systeemniveau	108
3.1 De doorlichtingen van de onderwijsinspectie	109
3.2 Het peilingsonderzoek en het internationaal vergelijkend onderzoek	109
Module 2.3: Complementariteit van de verschillende bronnen.....	111
1. Inleiding	111
2. Onderzoeksopzet.....	111
2.1 Dataverzameling	111
2.2 Data-analyse	113
3. Resultaten.....	113
3.1 Verklaringen voor de mate van concordantie tussen de resultaten van verschillende bronnen van externe kwaliteitszorg.....	113
3.2 Resultaten van verschillende bronnen combineren.....	117
3.3 Een toekomst met Vlaamse toetsen.....	124

Referenties	125
Bijlage 1 – Verticale analyse	133
1. Peiling basisonderwijs Nederlands 2018 en wiskunde 2021	133
1.1 Design.....	133
1.2 Standaardbepaling	148
1.3 Analytische methode	149
1.4 Besluitvorming	151
1.5 Praktische proef bij de peiling Nederlands	152
1.6 Het peilingsonderzoek doorheen de tijd	153
2. Peiling wiskunde A-stroom 2018 en B-stroom 2019	155
2.1 Design.....	155
2.2 Standaardbepaling	168
2.3 Analytische methode	168
2.4 Besluitvorming	169
2.5 Het peilingsonderzoek doorheen de tijd	169
3. PIRLS 2016	169
3.1 Design.....	169
3.2 Standaardbepaling	180
3.3 Analytische methode	180
3.4 Besluitvorming	183
3.5 PIRLS doorheen de tijd.....	185
4. TIMSS 2019	185
4.1 Design.....	185
4.2 Standaardbepaling	195
4.3 Analytische methode	196
4.4 Besluitvorming	198
4.5 TIMSS doorheen de tijd	199
5. PISA 2018	200
5.1 Design.....	200
5.2 Standaardbepaling	218
5.3 Analytische methode	220
5.4 Besluitvorming	224
5.5 PISA doorheen de tijd	224
6. Doorlichtingen van de onderwijsinspectie	225
6.1 Design.....	225

6.2	Standaardbepaling	233
6.3	Analytische methode	233
6.4	Besluitvorming	235
6.5	Doorlichtingen doorheen de tijd.....	236
Bijlage 2 – Resultaten van studie 2		237
Vergelijking 1: Internationaal vergelijkend onderzoek en doorlichtingen		237
1.	PIRLS 2016 en de doorlichting: 27 scholen.....	237
1.1	Type 0-schooleffect.....	237
1.2	Type A-schooleffect	238
1.3	Type B-schooleffect	239
2.	TIMSS 2019 en de doorlichting: 21 scholen	240
2.1	Type 0-schooleffect.....	240
2.2	Type A-schooleffect	241
2.3	Type B-schooleffect	242
3.	PISA 2018 leesvaardigheid en de doorlichting: 7 scholen	243
3.1	Type 0-schooleffect.....	243
3.2	Type A-schooleffect	244
3.3	Type B-schooleffect	245
4.	PISA 2018 wiskundige geletterdheid en de doorlichting: 19 scholen	246
4.1	Type 0-schooleffect.....	246
4.2	Type A-schooleffect	247
4.3	Type B-schooleffect	248
Vergelijking 2: Peilingsonderzoek en doorlichtingen		249
1.	Peiling Nederlands 2018 en de doorlichting: 20 scholen	249
1.1	Type 0-schooleffect.....	249
1.2	Type A-schooleffect	250
1.3	Type B-schooleffect	251
2.	Peiling wiskunde basisonderwijs 2021 en de doorlichting: 19 scholen.....	252
2.1	Functies (7 scholen)	252
2.2	Problemen oplossen bij getallen en bewerkingen (7 scholen).....	255
2.3	Problemen oplossen bij meten, ruimte en meetkunde (10 scholen)	258
2.4	Omtrek, oppervlakte en inhoud (8 scholen).....	261
Vergelijking 3: Internationaal vergelijkend onderzoek en peilingsonderzoek		264
1.	Peiling Nederlands basisonderwijs 2018 & PIRLS 2016: 9 scholen	264
1.1	Type 0-schooleffect.....	264

1.2	Type A-schooleffect	265
1.3	Type B-schooleffect	266
2.	Peiling wiskunde A-stroom 2018 & PISA 2018: 11 scholen	267
2.1	Algebraïsering (6 scholen).....	267
2.2	Evenredigheden (6 scholen)	270
2.3	Omgaan met data (6 scholen).....	273
2.4	Meetkundige procedures: constructies (6 scholen)	276
3.	Peiling wiskunde B-stroom 2019 & PISA 2018: 20 scholen	279
3.1	Getalinzicht en hoofdbewerkingen (17 scholen).....	279
3.2	Meetkunde (15 scholen).....	282
3.3	Informatieverwerking en -verwerking (15 scholen)	285
3.4	Metten (16 scholen).....	288
	Bijkomende analyses: Peilingen eerste graad secundair onderwijs & PISA.....	291
1.	Peiling wiskunde A-stroom 2018 & PISA 2018: 11 scholen	291
1.1	Algebraïsering (6 scholen).....	291
1.2	Evenredigheden (6 scholen)	294
1.3	Omgaan met data (6 scholen).....	297
1.4	Meetkundige procedures: constructies (6 scholen)	300
2.	Peiling wiskunde B-stroom 2019 & PISA 2018: 16 scholen	303
2.1	Getalinzicht en hoofdbewerkingen (16 scholen).....	303
2.2	Meetkunde (15 scholen).....	306
2.3	Informatieverwerking en -verwerking (14 scholen)	309
2.4	Metten (15 scholen).....	312
	Bijlage 3 – Resultaten van studie 3	315
1.	Doorlichtingen doorheen de tijd	315
1.1	Onderwijsaanbod (doorlichtingsronde 3) & Afstemming van het aanbod op een gevalideerd doelenkader (Inspectie 2.0)	315
1.2	Evaluatiepraktijk (doorlichtingsronde 3) & Leerlingenevaluatie (Inspectie 2.0)	317
2.	Vergelijking van trends bij de peilingen en het internationaal vergelijkend onderzoek.....	319
2.1	Peiling Nederlands en PIRLS (basisonderwijs)	319
2.2	Peiling wiskunde en TIMSS (basisonderwijs)	320
2.3	Peiling wiskunde in de A- en B-stroom en PISA (secundair onderwijs)	321
2.4	Bijkomende analyses.....	322
	Bijlage 4 – Leidraad focusgroepen	328
	Bijlage 5 – Codeboom	332

Lexicon

In het voorliggend rapport wordt er heel wat terminologie gebruikt. Met het oog op de leesbaarheid volgt hieronder een overzicht van vaak gebruikte begrippen en afkortingen met daarbij de pagina's waarop het begrip of de afkorting toegelicht wordt.

A

adaptieve toets leesvaardigheid, 212
analysekader
 analytische methode, 12
 besluitvorming, 12
 design, 11
 doel, 12
 domein en construct, 12
 modus, 12
 populatie, 12
 standaardbepaling, 12
 steekproef, 12
ankeritems, 149

B

beleidsniveau, 225
benchmarks, 180

C

cesuurbepaling, 148
criteriumgerichte meting, 148
cross-case analysis, 113

G

gestratificeerde steekproef, 143

I

IEA, 169

K

kalibratieonderzoek, 145
kritische kenmerken, 233
kwaliteitsbewaking, 13

L

leeg model, 149

M

model voor schoolfeedback, 150

N

nettomodel, 150

O

OESO, 200
ontwikkelingsniveaus, 233
ontwikkelingsschaal, 233

P

PIRLS, 169
PIRLS Repeat, 45
PISA, 200
plausible values, 181

S

schoolfeedback, 13
STEP, 133
substantief niveau, 225
systeemmonitoring, 13

T

Test-Curriculum Matching Analysis (TCMA), 187
TIMSS, 185
trenditems, 180
types modellen
 type 0-model, 46
 type AA-model, 46
 type A-model, 46
 type B-model, 46
 type X-model, 46

U

U7-schaal leereffecten, 48

V

vaardigheidsniveaus, 218

W

Warm's weighted likelihood estimation, 149
within-case analysis, 113

Inleiding

Hoe staat het met de kwaliteit van het Vlaamse onderwijs? Om op die vraag een antwoord te krijgen, doet de Vlaamse overheid beroep op verschillende bronnen van externe kwaliteitszorg, namelijk de doorlichtingen, het peilingsonderzoek en het internationaal vergelijkend onderzoek. Elk van die bronnen biedt specifieke informatie over de Vlaamse onderwijskwaliteit. Zo brengt de onderwijsinspectie aan de hand van de doorlichtingen voor elke Vlaamse school de kwaliteit van hun school- en klaspraktijk in kaart (Onderwijsinspectie, 2021f). Het internationaal vergelijkend onderzoek (zoals PISA, PIRLS, TIMSS...) vergelijkt op zijn beurt de prestaties van Vlaamse leerlingen met die van leerlingen uit andere onderwijssystemen (Mullis et al., 2017, 2020; OECD, 2019b). Via de peilingen krijgt het onderwijsbeleid inzicht in het aandeel leerlingen dat in Vlaanderen de onderwijsdoelen bereikt (AHOVOKS, 2021). Vanaf het schooljaar 2023-2024 worden de peilingen vervangen door de Vlaamse toetsen. Ook de resultaten van die toetsen – waarbij eveneens leerwinst in kaart gebracht zal worden – kunnen in de toekomst informatie bieden over de kwaliteit van het Vlaamse onderwijs (Vlaams Ministerie van Onderwijs en Vorming, 2022).

Zoals dat in het Vlaamse onderwijs het geval is, is de externe kwaliteitszorg bij voorkeur gebaseerd op verschillende types informatie die afkomstig zijn uit meerdere bronnen (Baker & Linn, 2002; Schafer, 2003). Het gebruik van meerdere bronnen kan immers leiden tot meer kwaliteitsvolle en betrouwbare besluitvorming over de kwaliteit van de scholen als de informatie uit die verschillende bronnen accuraat en consistent is (Chester, 2005; Porter et al., 2004).

Echter, de resultaten van de verschillende bronnen – en de daarmee samenhangende conclusies over onderwijskwaliteit – liggen in Vlaanderen niet steeds in dezelfde lijn. Zo toont de peiling Nederlands van 2013 bijvoorbeeld dat 92% van de Vlaamse leerlingen de eindtermen lezen beheerst op het einde van het basisonderwijs, wat een stijging is ten opzichte van de peiling uit 2007 (Aesaert & Janssen, 2018). Uit de resultaten van PIRLS 2016 blijkt daarentegen dat Vlaanderen de grootste daler van West-Europa is tussen 2006 en 2016 (Dockx et al., 2019). De onderwijsspiegels tonen voor die periode dan weer aan dat het aantal scholen dat volgens de verwachting scoort voor het leergebied Nederlands varieert tussen 86 en 91 procent (Onderwijsinspectie, 2012, 2013, 2014, 2015, 2016, 2017b).

Het eenvoudigweg gebruik maken van verschillende bronnen leidt dus niet automatisch tot eenduidige, betrouwbare en valide besluitvorming over de effectiviteit van de Vlaamse scholen en het Vlaamse onderwijssysteem. De validiteit en betrouwbaarheid van die besluitvorming zijn in grote mate afhankelijk van de manier waarop de informatie uit die verschillende bronnen gecombineerd wordt (Chester, 2003). Tot op heden bestaat er in Vlaanderen – en in andere onderwijssystemen – geen consensus over hoe verschillende bronnen complementair ingezet kunnen worden en hoe de data van die bronnen gecombineerd kunnen worden om zo een nauwkeuriger zicht te krijgen op onderwijskwaliteit (Martínez et al., 2016; Van Der Lans et al., 2019).

In lijn met bovenstaande probleemstelling beantwoorden we in dit rapport de volgende vier onderzoeksvragen:

1. Welke gelijkenissen en verschillen zijn er tussen de doorlichting, het peilingsonderzoek en het internationaal vergelijkend onderzoek inzake de evaluatiedoelen, de getoetste kwaliteitsindicatoren, de gehanteerde maatstaven en rapporteringsniveaus, de gemeten inhoud, de getoetste steekproef en populatie, en de gehanteerde instrumenten?
2. Uitgaande van de resultaten van die verschillende bronnen, in welke mate convergeren of divergeren de uitspraken over onderwijskwaliteit het voorbije decennium?

3. In welke mate duiden de vastgestelde gelijkenissen en verschillen tussen de bronnen (Onderzoeksvraag 1) de convergerende en divergerende uitspraken over onderwijskwaliteit (Onderzoeksvraag 2)?
4. Op welke manier kunnen de verschillende bronnen en de daarbij horende data gecombineerd worden om tot meer betrouwbare en accurate uitspraken over onderwijskwaliteit te komen op systeem- en schoolniveau?

Leeswijzer

Het voorliggend rapport maakt deel uit van het OBPWO-project 'Kwaliteitsmonitoring in het Vlaamse onderwijs: onderzoek naar het samenspel van het OK-kader, de doorlichting en de gestandaardiseerde toetsen'. Dat project bestaat uit twee verschillende deelrapporten. Een eerste deelrapport focust op de processen en effecten van de doorlichting volgens Inspectie 2.0 en het Referentiekader voor Onderwijskwaliteit, terwijl het tweede deelrapport zich richt op de verschillende bronnen van externe kwaliteitszorg. In dit rapport wordt er enkel gerapporteerd over het tweede deelrapport van dit OBPWO-project.

Meer specifiek bestaat dit rapport uit drie verschillende modules die elk andere onderzoeksvragen beantwoorden. Hieronder geven we kort weer wat de focus van elke module is.

- **Module 2.1: Gelijkenissen en verschillen tussen bronnen van externe kwaliteitszorg**
In deze module beschrijven we de gelijkenissen en verschillen tussen de doorlichting, het peilingsonderzoek en het internationaal vergelijkend onderzoek (zie Onderzoeksvraag 1). Daarnaast geven we voor elk bron aan wat de mogelijkheden en beperkingen zijn om betrouwbare en valide uitspraken te doen over specifieke aspecten van onderwijskwaliteit op een specifiek aggregatieniveau.
- **Module 2.2: Mate van convergentie of divergentie tussen de resultaten van verschillende bronnen**
Deze module omvat de resultaten van drie cross-instrumentele studies die nagaan in welke mate de uitspraken over onderwijskwaliteit van verschillende bronnen convergeren of divergeren (zie Onderzoeksvraag 2). Concreet focussen die drie studies op het volgende:
 - In **Studie 1** onderzoeken we de concordantie tussen de resultaten op de PIRLS-schaal begrijpend lezen en peilingschaal lezen aan de hand van de data van PIRLS Repeat. Ook gaan we voor de deelnemende scholen na in welke mate de resultaten op de PIRLS- en peilingschaal convergeren of divergeren met de resultaten uit de doorlichtingen voor het leergebied Nederlands.
 - Studie 2** focust op de concordantie tussen de resultaten van de verschillende bronnen van externe kwaliteitszorg op schoolniveau. Meer specifiek beantwoorden we de volgende vraag: in welke mate komt het resultaat van een bepaalde school op de ene bron overeen met het resultaat van dezelfde school bij een andere bron?
 - Bij **Studie 3** gaan we na of bronnen die een gelijkaardig domein of construct meten ook dezelfde trends op systeemniveau vinden. Meer bepaald maken we in die studie een vergelijking tussen de volgende trends:
 - doorlichtingsronde 3 en Inspectie 2.0;
 - PIRLS en de peiling Nederlands basisonderwijs;
 - TIMSS en de peiling wiskunde basisonderwijs;
 - PISA en de peiling wiskunde A- en B-stroom.

Aansluitend volgt er een conclusie waarin we op basis van de resultaten uit die drie cross-instrumentele studies een antwoord formuleren op Onderzoeksvraag 2.

- **Module 2.3: Complementariteit van de verschillende bronnen**

Deze module heeft een tweeledig doel. Enerzijds onderzoeken we in welke mate de gelijkenissen en verschillen uit Module 2.1 een verklaring vormen voor de resultaten uit Module 2.2 (zie Onderzoeksvraag 3). Anderzijds belichten we aan de hand van focusgroepen hoe de verschillende bronnen van externe kwaliteitszorg gecombineerd kunnen worden (zie Onderzoeksvraag 4).

Module 2.1: Gelijkenissen en verschillen tussen bronnen van externe kwaliteitszorg

1. Onderzoeksopzet

Het doel van de voorliggende module is om de gelijkenissen en verschillen in de onderzoeksprocessen en -instrumenten bij meerdere bronnen van externe kwaliteitszorg in Vlaanderen, i.e., de doorlichtingen, het peilingsonderzoek¹ en het internationaal vergelijkend onderzoek, in kaart te brengen (Onderzoeksvraag 1). Een concreet overzicht van de gelijkenissen en verschillen is noodzakelijk om nadien de concordantie tussen de resultaten van die verschillende bronnen te kunnen na gaan. Elke bron wordt immers ontwikkeld met het oog op een uniek doel en door een specifieke groep van stakeholders (Brookhart, 2009). Dat leidt tot grote verschillen tussen de bronnen op vlak van onder meer focus, context en methodologie (Treviño & Órdenes, 2017; Wagner et al., 2018).

1.1 Geanalyseerde bronnen

Om de bovenstaande onderzoeksvraag te beantwoorden vergelijken we de volgende recente bronnen:

- peiling basisonderwijs Nederlands 2018 en wiskunde 2021;
- peiling wiskunde A-stroom 2018 en B-stroom 2019;
- PIRLS 2016;
- TIMSS 2019;
- PISA 2018;
- doorlichtingen van de onderwijsinspectie volgens de aanpak bij Inspectie 2.0.²

Voor elk van bovenstaande bronnen worden zowel de documenten die het onderzoeksproces beschrijven als de rapporten waarin de resultaten zijn opgenomen geanalyseerd.

1.2 Analyse kader

Bij de documentanalyse van bovenstaande bronnen hanteren we een analysekader dat gebaseerd is op het kader van Treviño en Órdenes (2017). Concreet vormen de volgende vier dimensies het uitgangspunt:

De eerste dimensie is het **design** van het kwaliteitszorginstrument. Het design omvat de genomen beslissingen over het doel, de rationale en het conceptueel kader dat onderzocht wordt. Tabel 1 toont een overzicht van de verschillende criteria waarover binnen het design beslissingen genomen worden (Treviño & Órdenes, 2017; Wagner et al., 2018).

¹ We beperken ons bij het peilingsonderzoek tot de peilingen wiskunde en Nederlands, aangezien enkel die leergebieden aan bod komen in TIMSS, PIRLS en PISA.

² Sinds 2018 is de aanpak bij Inspectie 2.0 geëvolueerd en heeft de onderwijsinspectie de doorlichtingsaanpak verder aangepast. In dit onderzoek baseren we ons op de doorlichtingsaanpak Inspectie 2.0 zoals die bij de start in 2018 vormgegeven werd. We gaan dus niet in op de recente aanpassingen en evoluties die na aanvang geïmplementeerd werden.

Tabel 1*Designcriteria*

Criterium	Omschrijving
Doel	De reden waarvoor het kwaliteitszorginitiatief ontwikkeld is.
Domein en construct	De inhoud en kenmerken die gemeten worden enerzijds en de aard van wat gemeten wordt anderzijds.
Populatie	De specifieke doelgroep waarbij de bron wordt afgenomen.
Steekproef	De selectie van respondenten uit de populatie waarbij de bron afgenomen wordt en de rationale om voor die steekproef te kiezen.
Modus	De manier waarop de data verzameld worden.

Elke bron spreekt een waardeoordeel uit over de mate waarin de onderzochte eenheid (bijvoorbeeld het Vlaamse onderwijs of een specifieke school) zich situeert ten opzichte van de vastgelegde standaard. De tweede dimensie **standaardbepaling** refereert naar de concrete aanpak en standaard die de bron hanteert om dat waardeoordeel uit te spreken. Meer specifiek wijst die dimensie dus naar de procedures om vaardigheids- en beoordelingsniveaus te maken, de cesuur te bepalen, inhoudelijke omschrijvingen van ontwikkelingsniveaus te bepalen enzoverder (Blömeke & Gustafsson, 2017).

De **analytische methode** verwijst als derde dimensie naar de manier waarop de verzamelde data geanalyseerd worden. De beslissingen omtrent de analytische methode hebben een grote invloed op wat er precies gerapporteerd kan worden over de onderzochte populatie (Treviño & Órdenes, 2017).

Besluitvorming is de vierde dimensie en bouwt verder op de voorgaande drie dimensies. Het design, de standaardbepaling en de analytische methode vormen immers de blauwdruk voor de interpretaties en besluiten die op basis van de verzamelde data genomen kunnen worden. Meer specifiek verwijst de dimensie besluitvorming naar de validiteit van de interpretaties en besluiten die op basis van de data gemaakt worden (Treviño & Órdenes, 2017).

1.3 Aanpak van de documentanalyse

In eerste instantie werd per bron van externe kwaliteitszorg een verticale analyse uitgevoerd. Meer specifiek werden de documenten van elke bron afzonderlijk gecodeerd aan de hand van de dimensies uit bovenstaand analysekader. Per bron werden de resultaten uit de verschillende documenten nadien gebundeld in een ankertekst. De ankerteksten van de verticale analyse van elke bron worden omwille van hun omvang weergegeven in Bijlage 1 (p. 133). In een tweede stap hebben we de gelijkenissen en verschillen tussen de bronnen in kaart gebracht via een horizontale analyse. Concreet hebben we voor elke specifieke dimensie uit het analysekader een vergelijking gemaakt tussen de verticale analyses van de verschillende bronnen. De resultaten daarvan worden verderop in deze module weergegeven bij '2 Horizontale analyse'.

1.4 Matrix

Verder bouwend op de documentanalyse – en meer specifiek de resultaten bij de dimensie besluitvorming – ontwikkelden we een matrix. Daarin wordt voor elke bron van externe kwaliteitszorg aangegeven wat de mogelijkheden en beperkingen zijn om betrouwbare en valide uitspraken te doen

over specifieke aspecten van onderwijskwaliteit op een specifiek aggregatieniveau. Concreet hebben we de matrix als volgt opgebouwd:

- **Specifieke aspecten van onderwijskwaliteit**

Het referentiekader voor onderwijskwaliteit (het OK-kader) omvat de minimale verwachtingen voor kwaliteitsvol onderwijs in Vlaanderen (Onderwijsinspectie, 2017a). De rubrieken en kwaliteitsverwachtingen in het OK-kader vormen de basis voor de matrix, aangezien ze aangeven welke aspecten er tot kwaliteitsvol onderwijs behoren.

- **Meerdere aggregatieniveaus**

De bronnen voor externe kwaliteitszorg doen uitspraken op verschillende aggregatieniveaus:

- het systeemniveau of het niveau van het Vlaamse onderwijs;
- het schoolniveau;
- het leerlingniveau.

- **Combinatie van specifieke aspecten van onderwijskwaliteit en meerdere aggregatieniveaus**

De kwaliteitsverwachtingen uit het OK-kader zijn gericht op het aggregatieniveau van de school. Na een analyse van de kwaliteitsverwachtingen bleek dat er twee rubrieken ook relevant zijn op systeemniveau, namelijk 'resultaten en effecten' en 'kwaliteitsontwikkeling'.

Een specifieke uitspraak op systeem- of schoolniveau over een specifiek aspect van onderwijskwaliteit kan niet beschouwd worden als een uitspraak over de onderwijskwaliteit in zijn totaliteit. Wel kan elk van die uitspraken beschouwd worden als een indicator van onderwijskwaliteit.

Op leerlingniveau kunnen er geen uitspraken gedaan worden over onderwijskwaliteit. Voor de volledigheid nemen het niveau van de leerling toch mee op in de matrix. Er zijn namelijk bronnen die uitspraken kunnen doen over de resultaten van specifieke leerlingen. Die uitspraken kunnen echter niet beschouwd worden als uitspraken over onderwijskwaliteit.

Aan de hand van de documentanalyse hebben we de matrix ingevuld met de concrete inhoudelijke aspecten van onderwijskwaliteit waarover elke bron uitspraken doet of kan doen. Daarbij vermelden we ook de beperkingen die samengaan met die uitspraken.

2. Horizontale analyse

2.1 Design

2.1.1 Doel

Uit de verticale analyse blijkt dat de bronnen van externe kwaliteitszorg verschillende doelen vooropstellen om de kwaliteit van het Vlaamse onderwijs te evalueren, bewaken en verbeteren. Concreet onderscheiden we daarbij de volgende doelen:

- Systemmonitoring omvat het opvolgen en evalueren van de Vlaamse onderwijskwaliteit doorheen de tijd.
- Kwaliteitsbewaking gaat over het toezicht om te garanderen dat zowel elke school als het Vlaamse onderwijs in zijn geheel aan de minimale verwachtingen voor kwaliteitsvol onderwijs voldoet.
- Via schoolfeedback trachten verscheidene bronnen de scholen handvatten te bieden om hun interne kwaliteitszorg te verbeteren of te versterken.

Het doel van de internationaal vergelijkende onderzoeken (PIRLS, PISA en TIMSS) is internationaal bepaalde prestaties of vaardigheden van leerlingen in kaart brengen om vervolgens de trends inzake die prestaties of vaardigheden doorheen de tijd na te gaan. Verder onderzoeken ze de samenhang met kenmerken van de leerlingen enerzijds en de thuissituatie, de klas, de school en het onderwijssysteem anderzijds (De Meyer et al., 2019; Faddar et al., 2020; Tielemans et al., 2017). Het internationaal vergelijkend onderzoek is dus gericht op systeemmonitoring.

Het peilingsonderzoek focust daarentegen niet louter op systeemmonitoring, maar ook op kwaliteitsbewaking (AHOVOKS, 2021; Vlaams Ministerie van Onderwijs en Vorming, 2021). De peilingen hebben immers als doel om het aandeel leerlingen dat de onderwijsdoelen bereikt in kaart te brengen. Daarnaast worden – net zoals bij PISA, PIRLS en TIMSS – trends doorheen de tijd opgevolgd via herhalingspeilingen en wordt bij elke peiling de samenhang met achtergrondkenmerken onderzocht (Carpentier et al., 2019a, 2020a; Denis et al., 2019a; Spikic et al., 2022a). Op basis van de peilingen is het echter niet mogelijk om de prestaties van Vlaamse leerlingen internationaal te vergelijken, wat een verschil is met het internationaal vergelijkend onderzoek.

Net zoals de peilingen richten ook de doorlichtingen zich op kwaliteitsbewaking. Het doel van de doorlichtingen is namelijk waarborgen dat alle scholen voldoende kwaliteitsvol onderwijs afleveren door advies uit te brengen over de erkenning van die scholen (Onderwijsinspectie, 2018a).

Opvallend is dat elke bron inzet op schoolfeedback. Bij alle peilingen en internationaal vergelijkende onderzoeken ontvangen de deelnemende scholen een schoolfeedbackrapport (Appels et al., 2020; Centrum voor Schoolfeedback, 2016; Vakgroep Onderwijskunde, 2020; Vlaams Ministerie van Onderwijs en Vorming, 2021). Ook de onderwijsinspectie geeft schoolfeedback via een synthesegesprek waarin de resultaten van de doorlichting teruggekoppeld en besproken worden (Onderwijsinspectie, 2021f).

Kortom, alle bronnen focussen op schoolfeedback. Systeemmonitoring staat centraal bij PISA, TIMSS, PIRLS en de peilingen, terwijl kwaliteitsbewaking als doel voorop staat bij de doorlichtingen en het peilingsonderzoek.

2.1.2 Domein en construct

Het peilingsonderzoek, het internationaal vergelijkend onderzoek en de doorlichtingen richten zich op verschillende domeinen en constructen:

- De peilingen meten kennis, inzicht, vaardigheden en in mindere mate attitudes. De peilingen die opgenomen zijn in de documentanalyse richten zich op de domeinen Nederlands en wiskunde.
- PIRLS en TIMSS richten zich op het meten van vaardigheden. PIRLS focust op begrijpend lezen, terwijl wiskunde en wetenschappen centraal staat bij TIMSS.
- PISA focust op functionele kennis en vaardigheden op drie domeinen: leesvaardigheid, wiskundige geletterdheid en wetenschappelijke geletterdheid.
- De onderwijsinspectie richt zich hoofdzakelijk op proceskenmerken op zowel het beleids- als het substantief niveau.³

³ De onderwijsinspectie onderscheidt proceskenmerken op twee niveaus, namelijk het beleids- en het substantief niveau. Op het beleidsniveau wordt de kwaliteitsontwikkeling van scholen gemeten. Op het substantief niveau – ofwel het niveau van de klaspraktijk – worden de onderwijsleerpraktijk, enkele kwaliteitsgebieden en de bewoonbaarheid, veiligheid en hygiëne gemeten (Onderwijsinspectie, 2021f).

Hoewel het op het eerste zicht kan lijken dat heel wat verschillende bronnen hetzelfde domein meten, zijn er toch verschillen. De peilingen wiskunde, TIMSS en PISA (wiskundige geletterdheid) meten niet noodzakelijk exact hetzelfde domein. Hoewel die drie bronnen allemaal focussen op leerprestaties wiskunde, leggen ze elk hun eigen accenten wat betreft de definitie van wiskunde en dus ook de specifieke vaardigheden en kennis die gemeten wordt. Zo komt het gebruik van een zakrekenmachine bij bewerkingen bijvoorbeeld wel aan bod bij de peiling wiskunde, maar niet in TIMSS (Mullis & Martin, 2017; Van Renterghem et al., 2020). Hetzelfde geldt voor leesvaardigheid bij PISA, begrijpend lezen bij PIRLS en het domein lezen bij de peiling Nederlands. Hoewel elk van die bronnen focust op de leesvaardigheid van leerlingen hebben ze elk hun eigen definitie en conceptueel kader (Aesaert et al., 2018; Mullis & Martin, 2015; OECD, 2019a). Toch is inhoudelijke overlap tussen kaders mogelijk. Zo komt bijvoorbeeld de berekening van oppervlakte bij eenvoudige veelhoeken aan bod bij de peiling wiskunde in het basisonderwijs en bij TIMSS 2019 (Mullis & Martin, 2017; Van Renterghem et al., 2020).

Opvallend is wel dat alle bronnen daarnaast aandacht hebben voor de verzameling van achtergrond- en contextkenmerken (Aesaert et al., 2018; Mullis & Martin, 2015, 2017; OECD, 2019a; Onderwijsinspectie, 2021f; Van Renterghem et al., 2020; Willem, Harnisfeger, et al., 2018b; Willem, Van Nijlen, et al., 2018).

Ook wat betreft de opstelling van de conceptuele kaders zijn er gelijkenissen en verschillen te vinden. Zo vormen curricula het uitgangspunt voor de conceptuele kaders van de peilingen, PIRLS en TIMSS. Echter, waar de Vlaamse eindtermen de basis vormen voor het conceptueel kader van de peilingen, vormt het gemeenschappelijke curriculum van alle deelnemende landen de basis voor het kader van PIRLS en TIMSS. De toetsen van PIRLS en TIMSS kunnen daardoor ook vragen bevatten die niet aansluiten bij de Vlaamse eindtermen. PISA daarentegen houdt minder rekening met de curricula en focust voornamelijk op wat de OESO beschouwt als essentiële vaardigheden om maatschappelijke uitdagingen aan te kunnen. Het referentiekader voor onderwijskwaliteit of kortweg het OK-kader vormt op zijn beurt het uitgangspunt voor de doorlichtingen. Net zoals bij het peilingsonderzoek is er bij het opstellen van dat kader voornamelijk rekening gehouden met de Vlaamse – en in mindere mate met de internationale – context (Aesaert et al., 2018; Mullis et al., 2020; OECD, 2019b; Onderwijsinspectie, 2021f; Tielemans et al., 2017; Van Renterghem et al., 2020; Willem, Harnisfeger, et al., 2018b; Willem, Van Nijlen, et al., 2018).

2.1.3 Populatie

Om de vooropgestelde doelen (zie 2.1.1 Doel) te bereiken, verzamelen de bronnen data voor een bepaalde populatie. Wanneer we die populaties vergelijken, zijn er zowel gelijkenissen als verschillen te vinden. Zo is er een duidelijk verschil tussen de doorlichtingen en de andere databronnen. De populatie bij de doorlichtingen bestaat – in tegenstelling tot het peilingsonderzoek en het internationaal vergelijkend onderzoek – immers uit scholen en niet uit leerlingen (Onderwijsinspectie, 2021f). De populaties van het peilingsonderzoek en het internationaal vergelijkend onderzoek richten zich beiden op leerlingen en komen in dat opzicht wel overeen (Carpentier et al., 2019a, 2020a; De Meyer et al., 2019; Denis et al., 2019a; Faddar et al., 2020; Spikic et al., 2022a; Tielemans et al., 2017).

Ook bij de vergelijking van het peilingsonderzoek en het internationaal vergelijkend onderzoek vinden we gelijkenissen en verschillen. Ten eerste definiëren het peilingsonderzoek, PIRLS en TIMSS hun populatie leerlingen op basis van het leerjaar waarin die leerlingen les volgen (ongeacht hun leeftijd). Bij PISA daarentegen wordt de populatie gedefinieerd op basis van leeftijd en niet op basis van leerjaar (Carpentier et al., 2019a, 2020a; De Meyer et al., 2019; Denis et al., 2019a; Faddar et al., 2020; Spikic et al., 2022a; Tielemans et al., 2017). Dat heeft als gevolg dat de populatie van de peilingen, PIRLS en

TIMSS steeds bestaat uit leerlingen met verschillende leeftijden, terwijl de PISA-populatie leerlingen uit verschillende leerjaren omvat die ongeveer dezelfde leeftijd hebben.

Ten tweede worden leerlingen uit het buitengewoon onderwijs niet opgenomen bij de peilingen. Elk peilingsonderzoek baseert zich immers op de eindtermen terwijl er in het buitengewoon onderwijs gewerkt wordt met ontwikkelingsdoelen (Carpentier et al., 2019a, 2020a; Denis et al., 2019a; Spikic et al., 2022a). Bij elk internationaal vergelijkend onderzoek worden leerlingen uit het buitengewoon onderwijs wel toegelaten (De Meyer et al., 2019; Faddar et al., 2020; Tielemans et al., 2017).⁴ Dat verschil kan mogelijk een invloed hebben op de mate waarin de resultaten van het internationaal vergelijkend onderzoek enerzijds en het peilingsonderzoek anderzijds divergeren of convergeren.

Ten derde worden anderstalige nieuwkomers bij zowel het peilingsonderzoek als het internationaal vergelijkend onderzoek uitgesloten uit de vooropgestelde populatie.⁵ (Carpentier et al., 2019a, 2020a; Denis et al., 2019a; Martin et al., 2017, 2020; OECD, 2021; Spikic et al., 2022a).

Tot slot zijn er enkele bronnen waarbij de populatie volledig overeenkomt. Dat geldt voor de peiling Nederlands en wiskunde in het basisonderwijs (die zich richten op leerlingen uit het zesde leerjaar) enerzijds en voor TIMSS en PIRLS (die focussen op leerlingen uit het vierde leerjaar) anderzijds. Onder de andere databronnen zijn er geen twee databronnen te vinden die hun populatie op exact dezelfde manier definiëren (Carpentier et al., 2019a, 2020a; De Meyer et al., 2019; Denis et al., 2019a; Faddar et al., 2020; Spikic et al., 2022a; Tielemans et al., 2017). Ook dat kan de mate van overeenkomst tussen de resultaten van verschillende databronnen beïnvloeden.

2.1.4 Steekproef

De steekproeftrekking verloopt vrij gelijkaardig bij de peilingsonderzoeken en bij de internationaal vergelijkende onderzoeken. Elk van die databronnen doet beroep op een gestratificeerde steekproef. Al wordt er afhankelijk van de bron rekening gehouden met andere stratificatievariabelen. Daarenboven gebeurt die steekproeftrekking uit de populatie van leerlingen (zie 2.1.3 Populatie) steeds in twee stappen. Eerst wordt er een steekproef van scholen geselecteerd, waarna er in een tweede stap een aantal leerlingen of klassen geselecteerd worden. Die selectie van leerlingen of klassen verloopt wel verschillend van bron tot bron:

- PIRLS selecteert alle klassen in een school die behoren tot de populatie.
- De peilingen selecteren willekeurig minstens één klas per school. In het secundair onderwijs wordt er daarbij rekening gehouden met de verdeling van studierichtingen in de populatie.
- TIMSS selecteert at random twee klassen van het vierde leerjaar in elke school.
- PISA selecteert op elke school willekeurig 35 leerlingen die 15 jaar oud zijn.

⁴ Welke groep leerlingen uit het buitengewoon onderwijs opgenomen wordt in de populatie, varieert tussen de bronnen (en soms ook over cycli van eenzelfde bron heen):

- Bij PIRLS 2016 werden leerlingen opgenomen die leerstof uit het vierde leerjaar onderwezen krijgen en les volgen in de (toenmalige) types 1, 3 en 8 (Tielemans et al., 2017).
- TIMSS 2019 richtte zich op leerlingen die de leerstof uit het vierde leerjaar onderwezen krijgen en les volgen in types basisaanbod, 3 en 9 (Faddar et al., 2020).
- PISA nam leerlingen op in de populatie die les volgen in type basisaanbod (OV3) of type 1 (OV3) (De Meyer et al., 2019).

⁵ Het is enerzijds mogelijk om die groep leerlingen uit te sluiten door ze niet te laten deelnemen aan de toetsen en vragenlijsten. Anderzijds is het ook mogelijk om ze – samen met de rest van de klas – te laten deelnemen en de data van die leerlingen bij de analyses uit te sluiten. Bij de tweede optie kan er wel beslist worden om de resultaten van die leerlingen mee te nemen in het schoolfeedbackrapport (Carpentier et al., 2019a, 2020a; De Meyer et al., 2019; Denis et al., 2019a; Faddar et al., 2020; Spikic et al., 2022a; Tielemans et al., 2017).

Daarnaast wordt er bij het peilingsonderzoek en het internationaal vergelijkend onderzoek ook steeds een reservesteekproef van vervangscholen getrokken (Carpentier et al., 2019a, 2020a; De Meyer et al., 2019; Denis et al., 2019a; Faddar et al., 2020; Spikic et al., 2022a; Tielemans et al., 2017).

Bij de steekproeftrekking houdt de onderwijsinspectie rekening met de voorwaarde dat elke Vlaamse school minstens één keer elke zes jaar doorgelicht moet worden. Er is echter geen garantie dat die steekproef van scholen representatief is voor de Vlaamse populatie van scholen in tegenstelling tot de steekproeven bij de andere bronnen (Onderwijsinspectie, 2021f). Opvallend is wel dat zowel de doorlichtingen als de andere bronnen bijkomende, maar relevante actoren bevragen. Hoewel van bron tot bron verschilt welke actoren bevroegd worden⁶, is er bij geen enkele bron een garantie dat die steekproef van relevante (school)actoren representatief is.

2.1.5 Modus

Vooraleer de data verzameld kunnen worden, moeten er eerst toetsen ontwikkeld worden bij het peilingsonderzoek en het internationaal vergelijkend onderzoek. De toetsontwikkeling verloopt vrij gelijkaardig bij die onderzoeken (zie Tabel 2). Verder is een opvallende gelijkenis dat de peilingen, PIRLS, TIMSS en PISA naast de toetsen ook bijkomende achtergrondvragenlijsten ontwikkelen (Aesaert et al., 2018; Martin et al., 2017, 2020; OECD, 2021; Van Renterghem et al., 2020; Willem, Harnisfeger, et al., 2018b; Willem, Van Nijlen, et al., 2018).

Tabel 2

Overzicht van het toetsontwikkelingsproces bij de peilingen, PIRLS, TIMSS en PISA

Peilingsonderzoek	PIRLS 2016	TIMSS 2019	PISA 2018
<ul style="list-style-type: none"> • Toetsmatrijs opstellen • Opgaven ontwikkelen⁷ • Vooronderzoek en pilootafname • Kalibratieonderzoek • Samenstelling definitieve toetsen 	<ul style="list-style-type: none"> • Conceptueel kader updaten • Teksten selecteren • Items ontwikkelen • Vooronderzoek • Finale teksten en items bepalen • Samenstelling toetsboekjes • Vertaling van toets 	<ul style="list-style-type: none"> • Conceptueel kader updaten • Items ontwikkelen • Vooronderzoek • Finale items bepalen • Samenstelling toetsboekjes • Vertaling van toets 	<ul style="list-style-type: none"> • Conceptueel kader definiëren • Items ontwikkelen • Items selecteren • Vooronderzoek • Finale items bepalen • Samenstelling toetsboekjes • Vertaling van toets

Ook wat betreft de toetsafname zijn er gelijkenissen tussen de peilingen, PIRLS, TIMSS en PISA. Bij die databronnen gebeurt de afname steeds onder begeleiding van een leerkracht en/of toetsbegeleider. Daarnaast worden er gedetailleerde handleidingen voorzien waarin de te volgen procedures stap voor

⁶ Welke (school)actoren juist bevroegd worden, varieert van bron tot bron:

- De peilingen bevragen een leerkracht per klas en de ouders (Carpentier et al., 2019a, 2020a; Denis et al., 2019a; Spikic et al., 2022a).
- PIRLS en TIMSS bevragen de ouders, de leerkracht(en) uit het vierde leerjaar en de directeur (Faddar et al., 2020; Tielemans et al., 2017).
- PISA bevroegt de ouders en de directeur van de school (De Meyer et al., 2019).
- Bij de doorlichting worden enkele leerkrachten, tien leerlingen en zes ouders bevroegd die door de school gekozen zijn (Onderwijsinspectie, 2021a, 2021b).

⁷ In tegenstelling tot bij PIRLS, TIMSS en PISA zijn er geen richtpercentages voor de ontwikkeling van opgaven bij het peilingsonderzoek (Aesaert et al., 2018; Van Renterghem et al., 2020; Willem, Harnisfeger, et al., 2018b; Willem, Van Nijlen, et al., 2018). Het doel van de peilingen is immers om een globale uitspraak te doen over het percentage leerlingen dat het minimumniveau bereikt. Het is niet de bedoeling om een uitspraak per (sub)cluster of subdomein te doen. Daarom worden er geen richtpercentages vooropgesteld bij het peilingsonderzoek. Wel wordt er gestreefd een zo evenwichtig mogelijke peilingstoets.

stap beschreven worden. Er zijn echter ook verschillen tussen die databronnen. Zo zijn de meeste toetsen en vragenlijsten op papier afgenomen. Bij PISA daarentegen worden alle toetsen en vragenlijsten (behalve de oudervragenlijst) digitaal afgenomen.⁸ Bij de peilingen zijn alleen de leerkrachtvragenlijsten gedeeltelijk digitaal afgenomen. Tot slot hanteert PISA in tegenstelling tot de andere databronnen een adaptief toetsdesign bij de toets leesvaardigheid (Carpentier et al., 2019a, 2020a; Denis et al., 2019a; Martin et al., 2017, 2020; OECD, 2021; Spikic et al., 2022a).

De manier van dataverzameling verloopt helemaal anders bij de doorlichtingen dan bij de hierboven vermelde databronnen. Eerst en vooral maakt de onderwijsinspectie gebruik van verschillende dataverzamelmethode (bijvoorbeeld gesprekken, observaties...) om triangulatie mogelijk te maken. Opvallend is wel dat de onderwijsinspectie – net zoals bovenstaande bronnen – gebruik maakt van uitgebreide leidraden en/of instrumentenbundels. Verder gebeurt – in tegenstelling tot bij de peilingen, PIRLS, TIMSS en PISA – de inschaling op de ontwikkelingsschalen tijdens een teamoverleg tussen de inspecteurs. Daarenboven vinden er reflectie- en synthesegesprekken plaats aan het eind van de dataverzameling (Onderwijsinspectie, 2018a).

2.2 Standaardbepaling

De standaardbepaling gebeurt op diverse en uiteenlopende manieren bij het peilingsonderzoek, de doorlichtingen en het internationaal vergelijkend onderzoek. Het peilingsonderzoek stelt een criteriumgerichte meting voorop. Dat houdt in dat de onderwijsdoelen zelf het uitgangpunt vormen voor de cesuur die bepaalt of leerlingen de onderwijsdoelen behalen.⁹ Die cesuur verdeelt de meetschaal immers in twee groepen van opgaven: de groep opgaven die leerlingen moeten beheersen en de groep opgaven die leerlingen (nog) niet moeten beheersen om te eindtermen te behalen. Op diezelfde meetschaal worden ook de leerlingen gesitueerd naargelang hun vaardigheid om zo te bepalen welke leerlingen de eindtermen wel of niet behalen. Bovendien wijzigt de cesuur niet over herhalingspeilingen heen (Carpentier et al., 2019a, 2020a; Denis et al., 2019a; Spikic et al., 2022a)

De standaardbepaling bij het peilingsonderzoek staat in sterk contrast met die bij het internationaal vergelijkend onderzoek. TIMSS, PIRLS en PISA maken geen gebruik van criteriumgerichte metingen. De scores van de leerlingen worden immers niet vergeleken met een vooraf bepaald ‘absoluut’ criterium. Wel wordt er een interpretatie gegeven aan de resultaten door de meetschaal (arbitrair) op te delen in verschillende benchmarks (PIRLS en TIMSS)¹⁰ of vaardigheidsniveaus (PISA)¹¹. De specifieke inhoudelijke beschrijving bij die benchmarks of vaardigheidsniveaus wordt pas na de dataverzameling toegekend en varieert bovendien van cyclus tot cyclus (Martin et al., 2017, 2020; OECD, 2021). Welke concrete leesstrategieën leerlingen kunnen hanteren om het middelmatige niveau van PIRLS te behalen, varieert bijvoorbeeld over cycli heen. Dat is een duidelijk verschil met de peilingen. Het peilingsonderzoek werkt met een cesuur die constant blijft over herhalingspeilingen heen, aangezien de eindtermen inhoudelijk niet veranderen. Het internationaal vergelijkend onderzoek daarentegen

⁸ Ter informatie, de toetsen van PIRLS 2021 en TIMSS 2023 worden ook digitaal afgenomen (TIMSS & PIRLS International Study Center, 2022a, 2022b).

⁹ Om die cesuur te bepalen maakt het peilingsonderzoek gebruik van de bookmark methode. Meer informatie over die exacte methode voor cesuurbepaling is te vinden in de verticale analyse van het peilingsonderzoek.

¹⁰ Bij PIRLS en TIMSS zijn er vier internationale benchmarks: het gevorderd, hoog, middelmatig en laag niveau (Martin et al., 2017, 2020). Meer informatie over de benchmarks is te vinden bij de verticale analyse van PIRLS 2016 en TIMSS 2019.

¹¹ PISA onderscheidt afhankelijk van de vaardigheid die gemeten wordt zes tot acht vaardigheidsniveaus. Niveau 2 vormt daarbij het referentiepunt. Leerlingen die dat niveau halen beschikken over de minimale vaardigheden die nodig zijn om dagdagelijkse problemen op te lossen (De Meyer & Warlop, 2010; OECD, 2021). Meer informatie over de vaardigheidsniveaus is te vinden bij de verticale analyse van PISA 2018.

kiest ervoor om de meetschaal steeds op te delen aan de hand van dezelfde benchmarks of vaardigheidsniveaus, maar wel de bijhorende interpretatie te wijzigen bij elke cyclus.

Hoewel de standaardbepaling van het peilingsonderzoek en het internationaal vergelijkend onderzoek sterk van elkaar verschillen, is er een opvallende gelijkenis: het gebruik van ankeritems (peilingen) of trenditems (internationaal vergelijkend onderzoek). Dat zijn gemeenschappelijke opgaven die in voorgaande cycli of peilingen afgenomen worden. Die opgaven worden gebruikt om de data uit verschillende cycli op eenzelfde meetschaal te plaatsen en zo trends in kaart te brengen. Bovendien zijn die gemeenschappelijke opgaven nodig om de cesuur (peilingen), benchmarks (PIRLS en TIMSS) en vaardigheidsniveaus (PISA) over te nemen (Carpentier et al., 2019a, 2020a; Denis et al., 2019a; Martin et al., 2017, 2020; OECD, 2021; Spikic et al., 2022a)

In tegenstelling tot de andere bronnen maakt de onderwijsinspectie geen gebruik van cesuren, benchmarks of vaardigheidsniveaus, maar wel van ontwikkelingsschalen. Voor elk proceskenmerk dat de onderwijsinspectie nagaat, is er een ontwikkelingsschaal opgesteld die bestaat uit vier ontwikkelingsniveaus.¹² Om scholen te beoordelen op die ontwikkelingsschalen maakt de onderwijsinspectie gebruik van kritische kenmerken. Die kenmerken zijn uitgedrukt in concreet en waarneembaar gedrag en vormen de basis voor de beoordeling bij de doorlichtingen. Door die kenmerken te vergelijken met de verzamelde data bepaalt de onderwijsinspectie het ontwikkelingsniveau van de school (Onderwijsinspectie, 2018a).¹³

2.3 Analytische methode

Wanneer we de analytische methode bij de peilingsonderzoeken en de internationaal vergelijkende onderzoeken bekijken, zijn er zowel gelijkenissen als verschillen. Eerst en vooral maken de peilingen, PIRLS, TIMSS en PISA allemaal gebruik van – weliswaar verschillende¹⁴ – IRT-modellen (Carpentier et al., 2019a, 2020a; Denis et al., 2019a; Martin et al., 2017, 2020; OECD, 2021; Spikic et al., 2022a).

De keuze voor een bepaald IRT-model heeft doorgaans weinig invloed op de ‘rangorde’ van de vaardigheden van leerlingen. Zo zal een leerling die hoog scoort bij één IRT-model ook hoog scoren bij een ander IRT-model. Een complexer IRT-model zorgt in principe wel voor een vaardigheidsschatting met een kleinere meetfout, maar dat is enkel zo wanneer het complexere IRT-model effectief correct is. Een nadeel van de complexere IRT-modellen is dat zij het vooropgesteld evenwicht in het toetskader (bijvoorbeeld, bij PIRLS heeft 50% van de items betrekking op het leesdoel ‘leeservaring’ en 50% op het leesdoel ‘informatieverwerking’) niet respecteren, wat ten koste gaat van de inhoudelijke validiteit (Heine & Robitzsch, 2022; Robitzsch & Lüdtke, 2020, 2022). Bij het peilingsonderzoek wordt voor een minder complex model gekozen waarmee de inhoudelijke validiteit wordt gevrijwaard, en op een meer transparante wijze de cesuur kan worden bepaald (Karantonis & Sireci, 2006; MacCann & Stanley,

¹² Concreet gaat het om de volgende vier ontwikkelingsniveaus: (1) beneden de verwachting, (2) benadert de verwachting, (3) volgens de verwachting en (4) overstijgt de verwachting (Onderwijsinspectie, 2021f).

¹³ In het kader van dit onderzoek kon het onderzoeksteam de kritische kenmerken inkijken. Mede op basis van die kritische kenmerken hebben we de documentanalyse vormgegeven en de matrix opgesteld. De onderwijsinspectie kiest er echter voor om de kritische kenmerken niet vrij te geven aan het brede publiek. Daarom rapporteren we in dit rapport steeds op het niveau van de ontwikkelingsschalen. De ontwikkelingsschalen zijn immers wel toegankelijk voor het brede publiek via de website van de onderwijsinspectie (Onderwijsinspectie, 2018a, 2021f).

¹⁴ Meer specifiek, maken de peilingen gebruik van het One Parameter Logistic Model (Carpentier et al., 2019a, 2020a; Denis et al., 2019a; Spikic et al., 2022a). PIRLS en TIMSS hanteren het 2-parameter logistisch model, het 3-parameter logistisch model en een Partial Credit Model (Martin et al., 2017, 2020). PISA gebruikt een 2-parameter logistisch model en een Generalized Partial Credit Model (OECD, 2021).

2006). Bij PIRLS en TIMSS wordt voor meer complexe modellen gekozen en prioriteren ze zo de vaardigheidsschatting. PISA bevindt zich tussen beiden.

Daarnaast is er een duidelijk verschil wat betreft de schattingsprocedure van de vaardigheid van leerlingen:

- Het peilingsonderzoek schat een vaardigheidsscore voor elke leerling aan de hand van *Warm's weighted likelihood estimation*.
- Het internationaal vergelijkend onderzoek maakt gebruik van vijf (bij PIRLS en TIMSS) of tien (bij PISA) plausible values.¹⁵

Doordat het internationaal vergelijkend onderzoek gebruik maakt van plausible values is de meetfout van de schattingen bij PIRLS, TIMSS en PISA kleiner dan bij het peilingsonderzoek. Dat komt doordat er bij de schatting van de plausible values – en dus de vaardigheid van de leerlingen – ook informatie uit de achtergrondvragenlijsten opgenomen wordt. De plausible values zijn om diezelfde reden niet geschikt om uitspraken op leerlingniveau te doen (Martin et al., 2017, 2020; OECD, 2021). De vaardigheidsscores bij het peilingsonderzoek, daarentegen, laten individuele leerlingenfeedback bij de paralleltoetsen wel toe. Bij de schatting van die scores wordt er namelijk geen gebruik gemaakt van informatie uit achtergrondvragenlijsten.

Ook wat betreft de gebruikte modellen zijn er zowel gelijkenissen als verschillen:

- De peilingen en PIRLS maken gebruik van een leeg model, een nettomodel en een model voor schoolfeedback. Daarbij wordt ook de toegevoegde waarde¹⁶ van scholen in kaart gebracht (Carpentier et al., 2019a, 2020a; Denis et al., 2019a; Spikic et al., 2022a; Tielemans et al., 2017).
- TIMSS en PISA specificeren in hun rapport niet welke achterliggende modellen ze gebruiken bij de Vlaamse rapporten. Wel hanteren ze beiden een model voor schoolfeedback. In tegenstelling tot bij PIRLS en de peilingen wordt bij TIMSS en PISA de toegevoegde waarde van scholen niet geschat en gerapporteerd aan de hand van rupsgrafieken. TIMSS en PISA hebben elk hun eigen aanpak waarbij de vergelijking met gelijkaardige scholen centraal staat.

Zo vergelijkt TIMSS de prestaties van de school in kwestie met de gemiddelde prestaties van scholen die een gelijkaardig profiel hebben enerzijds en met het Vlaams gemiddelde anderzijds (Appels et al., 2020).

PISA kiest ervoor om de ruwe resultaten van een bepaalde school te vergelijken met het Vlaams gemiddelde en een beperkt aantal gelijkaardige scholen die deelnamen aan PISA, namelijk de vijf referentiescholen¹⁷ (Vakgroep Onderwijskunde, 2020).

¹⁵ Meer informatie over *Warm's weighted likelihood estimation* en de schatting van plausible values is te vinden bij het onderdeel 'Analytische methode' in de verticale analyses van de peilingen, PIRLS 2016, TIMSS 2019 en PISA 2018.

¹⁶ In elk van de vermelde onderzoeken wordt de toegevoegde waarde van scholen geschat via een gecontextualiseerd statusmodel. Met andere woorden, er worden geen eerdere prestaties als predictoren toegevoegd aan het model.

¹⁷ Bij de vergelijking van ruwe resultaten wordt er *geen* rekening gehouden met factoren waarover scholen geen controle hebben, maar die wel de resultaten beïnvloeden. Het PISA-schoolfeedbackrapport omvat een uitzondering waarbij er juist wel rekening gehouden wordt met die beïnvloedende factoren. Voor

Wanneer we de analytische methode van bovenstaande bronnen vergelijken met die bij de doorlichtingen zijn er twee opvallende verschillen. Ten eerste maken de doorlichtingen geen gebruik van IRT-modellen, vaardigheidsscores of plausible values, maar wel van ontwikkelingsschalen¹⁸ om tot een advies te komen voor elke school. Ten tweede hanteert de onderwijsinspectie geen modellen, maar vooropgestelde criteria om op basis van de inschalingen tot een advies te komen. Zo zijn er criteria vastgelegd om te bepalen welke combinatie van inschalingen leidt tot een tekort, een ontwikkelkans of een sterk punt in het doorlichtingsverslag van de school.¹⁹ Wel geeft de onderwijsinspectie schoolfeedback door de resultaten van de doorlichting in een synthesegesprek met de doorgelichte school terug te koppelen. Daarnaast rapporteert de onderwijsinspectie jaarlijks in de onderwijspiegel over haar vaststellingen tijdens de doorlichtingen (Onderwijsinspectie, 2018a, 2021f, 2021g, 2021h).

2.4 Waar komen die gelijkenissen en verschillen vandaan?

Bovenstaande verschillen en gelijkenissen zijn niet toevallig tot stand gekomen. Elke bron van externe kwaliteitszorg wordt immers ontwikkeld met het oog op een specifiek doel. Vervolgens worden er in functie van dat doel keuzes gemaakt wat betreft het design, de standaardbepaling en de analytische methode. In wat volgt, verduidelijken we aan de hand van enkele voorbeelden uit de horizontale analyse kort hoe de verschillende doelen leiden tot bepaalde gelijkenissen of verschillen tussen de peilingen, het internationaal vergelijkend onderzoek en de doorlichtingen.

Zo heeft het **peilingsonderzoek** als hoofddoel om het aandeel leerlingen in kaart te brengen dat de eindtermen bereikt. Om dat doel te bereiken, dienen de peilingstoetsen logischerwijs hetzelfde te meten als de inhoud die aan bod komen in de eindtermen. De eindtermen vormen dus zelf het uitgangspunt voor het conceptueel kader bij de peilingen (Carpentier et al., 2019a, 2020a; Denis et al., 2019a; Spikic et al., 2022a). Ook de populatie en steekproef zijn niet willekeurig. De eindtermen moeten bereikt worden op het einde van het gewoon lager onderwijs en op het einde van elke graad in het gewoon secundair onderwijs. Bijgevolg heeft het geen zin om een peiling af te nemen in het eerste of vierde leerjaar van het lager onderwijs.

Niet alleen het design van het peilingsonderzoek vloeit uit het vooropgestelde doel. Ook de standaardbepaling en de analytische methode zijn gekozen met oog op het hoofddoel van de peilingen. Zo is een criteriumgerichte meting noodzakelijk om te bepalen wanneer een leerling de eindtermen bereikt – en dus om het aandeel leerlingen dat de eindtermen bereikt in kaart te brengen. Het peilingsonderzoek gebruikt daarbij de bookmark methode om te bepalen welke items leerlingen

leesvaardigheid (het hoofddomein bij PISA 2018) toont het rapport namelijk of de feitelijke score van de school in kwestie hoger of lager ligt dan de verwachte score op basis van de factoren die in het model voor schoolfeedback opgenomen worden (Vakgroep Onderwijskunde, 2020). Meer informatie over schoolfeedback bij PISA en concrete voorbeelden daarvan zijn te vinden in de verticale analyse van PISA 2018.

¹⁸ Ter informatie, ook binnen eenzelfde doorlichtingsronde kan de onderwijsinspectie inhoudelijke aanpassingen maken aan de ontwikkelingsschalen die ze hanteren bij de doorlichtingen.

¹⁹ Dit is een concreet voorbeeld van dergelijke criteria (Onderwijsinspectie, 2018a):

- Geen enkele inschaling van ‘beneden de verwachting’ of ‘benadert de verwachting’ bij de vier ontwikkelingsschalen van leerlingenbegeleiding leidt automatisch tot een sterk punt voor leerlingenbegeleiding in het doorlichtingsverslag.
- Minstens één inschaling van ‘beneden de verwachting’ of ‘benadert de verwachting’ bij de vier ontwikkelingsschalen van leerlingenbegeleiding leidt automatisch tot een ontwikkelkans voor leerlingenbegeleiding in het doorlichtingsverslag.
- Minstens drie inschalingen van ‘beneden de verwachting’ of ‘benadert de verwachting’ bij de vier ontwikkelingsschalen van leerlingenbegeleiding leidt automatisch tot een tekort voor leerlingenbegeleiding in het doorlichtingsverslag.

correct moeten hebben om een bepaalde (set van) eindterm(en) te bereiken. Tot slot is het doel van de peilingen niet om verklaringen te vinden, maar om vast te stellen welk aandeel van de leerlingen de eindtermen bereikt en die vaststelling te kaderen. Daarvoor maakt het peilingsonderzoek gebruik van multilevel analyses en trendanalyses (Carpentier et al., 2019a, 2020a; Denis et al., 2019a; Spikic et al., 2022a).

Het doel van **PIRLS**, **TIMSS** en **PISA** is om internationaal bepaalde prestaties of vaardigheden van leerlingen in kaart brengen om vervolgens de trends inzake die prestaties of vaardigheden doorheen de tijd na te gaan (De Meyer et al., 2019; Faddar et al., 2020; Tielemans et al., 2017). Om een internationale vergelijking te kunnen maken, moeten dezelfde domeinen en constructen in verschillende landen gemeten worden. Hoewel de aanpak bij PIRLS en TIMSS verschilt van die bij PISA, is het uitgangspunt hetzelfde. De drie bronnen willen namelijk eenzelfde conceptueel kader in verschillende landen meten:

- Bij PIRLS en TIMSS vormen de curricula uit de deelnemende landen de basis om een internationaal curriculum op te stellen. De inhoud van dat internationaal curriculum wordt vervolgens via PIRLS en TIMSS gemeten in alle deelnemende landen (Mullis et al., 2020; Tielemans et al., 2017). Met andere woorden, PIRLS en TIMSS kiezen voor een bottom-up aanpak om tot een internationaal conceptueel kader te komen.
- De OESO wil aan de hand van PISA nagaan in welke mate 15-jarige leerlingen wereldwijd de essentiële vaardigheden bezitten om toekomstige, maatschappelijke uitdagingen aan te kunnen. Om dat internationaal te kunnen nagaan, moet eerst bepaald worden welke vaardigheden noodzakelijk zijn om die uitdagingen aan te kunnen. PISA focust dus niet op de curricula van de deelnemende landen, maar stelt zelf een internationaal conceptueel kader van vaardigheden op dat vervolgens in alle deelnemende landen gemeten worden (OECD, 2019a, 2019b). Kortom, bij PISA wordt er een top-down aanpak gehanteerd om een internationaal conceptueel kader wereldwijd te meten.

De verschillen in de vooropgestelde doelen van de internationaal vergelijkende onderzoeken en het peilingsonderzoek leiden niet alleen tot verschillen in de opstelling van het conceptueel kader, maar bijvoorbeeld ook tot andere methodologische keuzes. Zo wil het internationaal vergelijkend onderzoek niet nagaan of leerlingen een bepaald criterium bereiken, maar resultaten tussen landen vergelijken. Daarom gebruiken ze in tegenstelling tot de peilingen geen cesuur, maar wel benchmarks (PIRLS en TIMSS) of vaardigheidsniveaus (PISA) als inhoudelijk referentiepunt bij die vergelijking (Martin et al., 2017, 2020; OECD, 2021). Verder wordt er bij de internationale onderzoeken gekozen voor complexere IRT-modellen en plausible values als schattingsmethode. Dat moet in principe leiden tot een correctere en betrouwbaardere vergelijking tussen landen (Martin et al., 2017, 2020). Die voordelen gelden echter enkel wanneer de IRT-modellen correct zijn. Een nadeel is dat die complexere IRT-modellen het relatieve belang van ieder item herwegen. Aangezien er bij de peilingen geen vergelijking tussen landen gemaakt wordt en de automatisch herweging van de items ongewenst is voor de cesuurbepaling, wordt er voor een minder complex IRT-model gekozen.

De **onderwijsinspectie** heeft een tweeledig doel: waarborgen dat alle onderwijsinstellingen voldoende kwaliteitsvol onderwijs leveren en feedback geven op basis waarvan scholen hun interne kwaliteitszorg kunnen versterken. Om die twee doelen te kunnen bereiken, is het van belang om naar de kwaliteit van de school in zijn totaliteit te kijken. Dat doet de onderwijsinspectie door bij de doorlichting aandacht te besteden aan het beleidsniveau van de school én aan de werking in de klas. Door daarenboven verschillende dataverzamelmethode te hanteren, krijgt de onderwijsinspectie zicht op de kwaliteit van de school in zijn geheel (Onderwijsinspectie, 2018a).

Ook de steekproef van de doorlichtingen is bijvoorbeeld een logisch gevolg van het vooropgestelde doel. Om te kunnen waarborgen dat *alle* onderwijsinstellingen kwaliteitsvol onderwijs aanbieden, is het immers essentieel om alle scholen regelmatig in de steekproef op te nemen. Zoals eerder vermeld, heeft de onderwijsinspectie als doel om elke school minstens een keer om de zes jaar doorgelicht door de onderwijsinspectie. Bovendien wordt 'voldoende kwaliteitsvol onderwijs leveren' geoperationaliseerd aan de hand van verschillende ontwikkelingsschalen. Op die manier kan de onderwijsinspectie voor elk aspect van onderwijskwaliteit nagaan of de school voldoende kwaliteitsvol onderwijs voorziet. De inschaling op die ontwikkelingsschalen wordt vervolgens op basis van eenduidige criteria vertaald in een advies waaruit blijkt of de school al dan niet voldoende kwaliteitsvol onderwijs levert (Onderwijsinspectie, 2018a).

Tot slot zijn er ook gelijkenissen die het gevolg zijn van de vooropgestelde doelen van de peilingen, het internationaal vergelijkend onderzoek en de onderwijsinspectie. Bijvoorbeeld, zowel het peilingsonderzoek als het internationaal vergelijkend onderzoek zet in op systeemmonitoring (zie 2.1.1 Doel). Om dat doel te bereiken, doen beide bronnen beroep op een gestratificeerde steekproef die representatief is voor het Vlaamse onderwijs. Verder maken beide bronnen gebruik van gemeenschappelijke items om trends over verschillende metingen heen in kaart te kunnen brengen.

3. Matrix

3.1 De matrix lezen en interpreteren

Zoals eerder toegelicht, geeft de matrix aan wat de mogelijkheden en beperkingen zijn van elke bron om betrouwbare en valide uitspraken te doen over specifieke aspecten van onderwijskwaliteit op een specifiek aggregatieniveau. Concreet is de matrix opgesplitst in drie afzonderlijke delen die de uitspraken op respectievelijk systeem-, school- en leerlingniveau omvatten. Elk van die delen is opgebouwd als volgt:

- Elke **rij** in de matrix stelt een concreet aspect van onderwijskwaliteit voor zoals geoperationaliseerd in het OK-kader.
- Elke **kolom** omvat een bepaalde bron van externe kwaliteitszorg: de peilingen, de doorlichtingen, PIRLS, TIMSS of PISA.
- Elke **cel** geeft aan:
 - Welke uitspraken de bron (aangegeven in de kolomtitel) kan doen over een bepaald aspect van onderwijskwaliteit (aangegeven in de rijtitel).
 - Welke beperkingen samengaan met die uitspraken. De beperkingen worden in de matrix *in het donkerblauw en schuingedrukt* aangeduid.

Belangrijk om op te merken is dat er op leerlingniveau geen uitspraken gedaan kunnen worden over onderwijskwaliteit. Aangezien er enkele bronnen zijn die uitspraken (kunnen) doen over de resultaten van specifieke leerlingen, nemen we dat niveau voor de volledigheid toch mee op in de matrix. Die uitspraken kunnen echter niet beschouwd worden als uitspraken over onderwijskwaliteit.

Ook de uitspraken op systeemniveau en schoolniveau kunnen onderwijskwaliteit in zijn totaliteit niet vatten. Wel kan elk van die uitspraken beschouwd worden als een indicator over een specifiek aspect van onderwijskwaliteit.

3.2 Wat zijn de mogelijkheden en beperkingen om uitspraken te doen over specifieke aspecten van onderwijskwaliteit?

3.2.1 Uitspraken op systeemniveau

3.2.1.1 Resultaten en effecten

	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
R1. Het Vlaamse onderwijs bereikt de minimaal gewenste output bij een zo groot mogelijke groep van lerenden.	<p>De peilingen beschrijven (trends in) het aandeel leerlingen dat de eindtermen behaalt. Daarbij vormen de onderwijsdoelen zelf het uitgangspunt en blijft de cesuur gelijk over herhalingsmetingen heen.</p> <p>De peiling Nederlands (basis-onderwijs) en wiskunde A-stroom bevragen bij leerkrachten welke eindtermen wel of niet gezien werden (in het huidige/voorgaande jaren). De peiling wiskunde B-stroom bevraagt welke ontwikkelingsdoelen wel of niet behandeld werden (in het huidige/voorgaande jaren).</p> <p><i>Slechts een leerkracht per klas uit zesde leerjaar of eerste graad secundair onderwijs wordt bevraagd (niet de leerkrachten van voorgaande jaren).</i></p>	<p>PIRLS beschrijft (trends in) prestaties begrijpend lezen en maakt daarbij een internationale vergelijking.</p> <ul style="list-style-type: none"> <i>Het uitgangspunt is een 'internationaal curriculum' (niet de eindtermen). Het gaat bovendien om het vierde leerjaar (eindtermen zijn geformuleerd voor het einde van het zesde leerjaar). Dat maakt het moeilijk om a.d.h.v. PIRLS na te gaan of de minimaal gewenste output bereikt is.</i> <i>Daarenboven wijzigt het conceptueel kader (en dus het construct dat PIRLS meet) over cycli heen.</i> 	<p>TIMSS beschrijft (trends in) prestaties wiskunde en wetenschappen en maakt daarbij een internationale vergelijking.</p> <ul style="list-style-type: none"> <i>De eindtermen vormen niet het uitgangspunt voor TIMSS. Het gaat bovendien om het vierde leerjaar (eindtermen zijn geformuleerd voor het einde van het zesde leerjaar). TIMSS meet ook heel wat zaken die niet aansluiten bij de eindtermen (zie TCMA). Dat maakt het moeilijk om a.d.h.v. TIMSS na te gaan of de minimaal gewenste output bereikt is.</i> <i>Het conceptueel kader (en dus het construct dat TIMSS meet) wijzigt over cycli heen.</i> 	<p>PISA beschrijft (trends in) leesvaardigheid, wiskundige geletterdheid en wetenschappelijke geletterdheid (met internationale vergelijking).</p> <ul style="list-style-type: none"> <i>De eindtermen vormen niet het uitgangspunt voor PISA. Het gaat bovendien om het leeftijd gebaseerde steekproef (terwijl de eindtermen in het secundair onderwijs per graad geformuleerd zijn). Dat maakt het moeilijk om o.b.v. PISA na te gaan of de minimaal gewenste output bereikt is.</i> <i>Daarenboven wijzigt het conceptueel kader (en dus de constructen die PISA meet) over cycli heen.</i> 	<p>Bij de U7-schaal (Onderwijsleerpraktijk) wordt voor elke school in kaart gebracht in welke mate er indicaties zijn van kwaliteitsbewaking. In het secundair onderwijs wordt er ook gekeken naar het percentage leerlingen dat jaarlijks slaagt.</p> <ul style="list-style-type: none"> <i>Geen garantie op representatieve steekproef van scholen.</i> <i>Geen absoluut criterium op systeemniveau om te bepalen wie slaagt. Daardoor is het niet mogelijk om o.b.v. de doorlichtingen zicht te krijgen op het aandeel leerlingen dat de minimaal gewenste output bereikt (slagen is niet noodzakelijk gelijk aan minimale gewenste output bereiken).</i>

	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
R2. Het Vlaamse onderwijs streeft naar welbevinden en betrokkenheid bij alle lerenden en het schoolteam en naar tevredenheid bij ouders en bij andere relevante partners.	De peiling wiskunde (basis-onderwijs) bevroagt bij de ouders de mate waarin ze tevreden zijn over de school. <i>Enkel ouders van leerlingen in het zesde leerjaar bevroagt.</i>	PIRLS 2016 bevroagt bij de leerlingen hun welbevinden en betrokkenheid via de schaal: Students' Sense of School Belonging. <ul style="list-style-type: none"> • <i>Die schaal meet niet het streven naar welbevinden en betrokkenheid, maar meet wel het eigenlijke welbevinden van de Vlaamse leerlingen (die deelnamen aan PIRLS 2016).</i> • <i>Enkel leerlingen uit het vierde leerjaar bevroagt.</i> 	TIMSS 2019 bevroagt bij de leerlingen hun welbevinden en betrokkenheid via de schaal: Students' Sense of School Belonging. <ul style="list-style-type: none"> • <i>Die schaal meet niet het streven naar welbevinden en betrokkenheid, maar meet wel het welbevinden van de Vlaamse leerlingen (die deelnamen aan TIMSS 2019).</i> • <i>Enkel leerlingen uit het vierde leerjaar bevroagt.</i> 	PISA 2018 bevroagt bij leerlingen in welke mate ze zich thuis en op hun gemak voelen op school. <ul style="list-style-type: none"> • <i>De schaal meet niet het streven naar welbevinden en betrokkenheid, maar wel de mate waarin leerlingen (die deelnamen aan PISA 2018) zich thuis voelen op school.</i> • <i>Enkel een beperkt aantal 15-jarige leerlingen bevroagt.</i> PISA 2018 bevroagt bij ouders in welke mate ze tevreden zijn over de school. <i>Enkel ouders van beperkt aantal 15-jarige leerlingen bevroagt.</i>	Bij de U7-schaal (Onderwijs-leerpraktijk) wordt voor elke school in kaart gebracht in welke mate er indicaties zijn van welbevinden bij leerlingen. <ul style="list-style-type: none"> • <i>Geen garantie op representatieve steekproef van scholen.</i> • <i>De doorlichtingen meten niet het streven naar welbevinden, maar brengen wel indicaties van welbevinden bij leerlingen in kaart. Bij het gesprek met leerlingen gaat het niet automatisch om een representatieve steekproef van leerlingen op systeemniveau.</i>
R3. Het Vlaamse onderwijs streeft bij elke lerende naar zoveel mogelijk leerwinst .					(zie R1) <i>De onderwijsinspectie meet zelf geen leerwinst, studievoortgang of lange termijn effecten bij elke leerling in de school. Wel heeft ze aandacht voor de kwaliteitsbewaking van de onderwijsleerpraktijk op de school.</i>
R4. Het Vlaamse onderwijs stimuleert de studievoortgang van elke lerende.					
R5. Het Vlaamse onderwijs waarborgt de toegang tot onderwijs voor elke lerende.					
R6. Het Vlaamse onderwijs streeft naar effecten op langere termijn bij alle lerenden.					

3.2.1.2 Kwaliteitsontwikkeling

	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
K1. Het Vlaamse onderwijs ontwikkelt haar kwaliteit vanuit een gedragen visie die vertaald is in de onderwijsleerpraktijk.					Bij de K1-schaal (Kwaliteitsontwikkeling) gaat de onderwijsinspectie na in welke mate er een schooleigen en gedragen visie aanwezig is die richtinggevend is voor de schoolwerking. <i>Geen garantie op representatieve steekproef van scholen op systeemniveau.</i>
K2. Het Vlaamse onderwijs evalueert haar werking cyclisch, systematisch en betrouwbaar vanuit de resultaten en effecten bij de lerenden.	<i>(Het peilingsonderzoek is een van de bronnen die de Vlaamse overheid gebruikt om op systematische wijze de Vlaamse onderwijskwaliteit te evalueren. Echter, de peilingen zeggen zelf niets over de mate waarin het Vlaams onderwijs haar werking evalueert.)</i>	<i>(PIRLS is een van de bronnen die de Vlaamse overheid gebruikt om op systematische wijze de Vlaamse onderwijskwaliteit te evalueren. Echter, PIRLS zegt zelf niets over de mate waarin het Vlaams onderwijs haar werking evalueert.)</i>	<i>(TIMSS is een van de bronnen die de Vlaamse overheid gebruikt om op systematische wijze de Vlaamse onderwijskwaliteit te evalueren. Echter, TIMSS zegt zelf niets over de mate waarin het Vlaams onderwijs haar werking evalueert.)</i>	PISA 2018 be vraagt in de directievragenlijst of de school aan evaluatie doet op initiatief van de school of omdat het verplicht is. <i>Enkel antwoord van de directeur (per school).</i>	Bij de K4- en K5-schaal (Kwaliteitsontwikkeling) gaat de onderwijsinspectie na in welke mate de school systematisch en betrouwbaar de eigen schoolwerking evalueert. <i>Geen garantie op representatieve steekproef van scholen op systeemniveau.</i>
K3. Het Vlaamse onderwijs borgt en ontwikkelt de kwaliteit van de onderwijsleerpraktijk.					De onderwijsinspectie geeft o.b.v. de doorlichting advies aan de Vlaamse overheid over de erkenning van de instelling. Bij de K6-schaal (Kwaliteitsontwikkeling) gaat de onderwijsinspectie na in welke mate de school bewaart wat kwaliteitsvol is en haar werkpunten verbetert. <i>Geen garantie op representatieve steekproef van scholen op systeemniveau.</i>

3.2.2 Uitspraken op schoolniveau

3.2.2.1 Resultaten en effecten

	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
R1. De school bereikt de minimaal gewenste output bij een zo groot mogelijke groep van lerenden.	<p>Terugkoppeling resultaten school:</p> <ol style="list-style-type: none"> 1. Aandeel leerlingen dat eindtermen behaalt op de school (absolute feedback) 2. Vergelijking met andere scholen (relatieve feedback) <p>Daarbij vormen de onderwijsdoelen zelf het uitgangspunt voor de peilingen (en absolute feedback). <i>De foutenmarges bij de resultaten op schoolniveau zijn relatief groot.</i></p>	<p>Terugkoppeling resultaten school:</p> <ol style="list-style-type: none"> 1. Resultaten toets begrijpend lezen (omvat vergelijking met Vlaams gemiddelde en met scholen met een gelijkaardig leerlingenpubliek). 2. Toegevoegde waarde van de school. 3. Beperkingen van de informatie uit het schoolfeedbackrapport worden expliciet besproken. <ul style="list-style-type: none"> • <i>PIRLS wordt afgenomen in het vierde leerjaar (niet het zesde).</i> • <i>Het uitgangspunt is een 'internationaal curriculum' (niet de eindtermen).</i> • <i>De feedback die scholen krijgen is relatief, niet absoluut. Dat maakt het moeilijk om a.d.h.v. PIRLS na te gaan of de minimaal gewenste output bereikt is.</i> • <i>De foutenmarges bij de resultaten op schoolniveau zijn relatief groot.</i> 	<p>Terugkoppeling resultaten school:</p> <ol style="list-style-type: none"> 1. Resultaten toets wis- kunde en wetenschappen (omvat vergelijking met Vlaams gemiddelde en met scholen met een gelijkaardig leerlingenpubliek). 2. Spreiding van de leerlingprestaties binnen een school i.v.m. alle Vlaamse leerlingen die deelnamen. <ul style="list-style-type: none"> • <i>TIMSS wordt afgenomen in het vierde leerjaar (niet het zesde).</i> • <i>Het uitgangspunt is een 'internationaal curriculum' (niet de eindtermen).</i> • <i>De feedback die scholen krijgen is relatief, niet absoluut. Dat maakt het moeilijk om a.d.h.v. TIMSS na te gaan of de minimaal gewenste output bereikt is.</i> • <i>De foutenmarges bij de resultaten op schoolniveau zijn relatief groot.</i> 	<p>Terugkoppeling resultaten school:</p> <p>Ruwe resultaten op de toetsen leesvaardigheid, wiskundige geletterdheid en wetenschappelijke geletterdheid met daarbij:</p> <ol style="list-style-type: none"> 1. Een vergelijking met Vlaams gemiddelde en vijf referentiescholen. 2. Een opsplitsing per studierichting en per onderwijsvorm 3. De relatie tussen de resultaten en SES <p>Ook de verwachte score voor leesvaardigheid wordt weergegeven.</p> <ul style="list-style-type: none"> • <i>PISA wordt afgenomen bij 15-jarigen (terwijl de eindtermen per graad geformuleerd zijn).</i> • <i>Het uitgangspunt is een door de OESO opgesteld conceptueel kader (niet de eindtermen).</i> • <i>De schoolfeedback is relatief, niet absoluut. Dat maakt het moeilijk om a.d.h.v. PISA na te gaan of de minimaal gewenste output bereikt is.</i> • <i>Er wordt bij de prestatie van ruwe resultaten geen rekening gehouden met factoren die scholen niet kunnen beïnvloeden,</i> 	<p>Terugkoppeling resultaten school: Bij de U7-schaal (Onderwijsleerpraktijk) wordt voor elke school in kaart gebracht in welke mate er kwaliteitsbewaking van de onderwijsleerpraktijk aanwezig is. Daarbij wordt er rekening gehouden met de prestaties van leerlingen. In het secundair onderwijs wordt er ook gekeken naar het percentage leerlingen dat jaarlijks slaagt. <i>Er is geen absoluut criterium (over scholen heen) om te bepalen wie slaagt. Met als gevolg, de cesuur ligt bij elke school op een andere plaats. Daardoor is het niet mogelijk om o.b.v. de doorlichtingen zicht te krijgen op het aandeel leerlingen dat de minimaal gewenste output bereikt (geslaagd is niet noodzakelijk gelijk aan minimale gewenste output bereiken).</i></p>

	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
				<p><i>maar die wel een invloed hebben op de prestaties van de leerlingen.</i></p> <ul style="list-style-type: none"> • <i>De foutenmarges bij de resultaten op schoolniveau zijn relatief groot.</i> 	
<p>R2. De school streeft naar welbevinden en betrokkenheid bij alle lerenden en het schoolteam en naar tevredenheid bij ouders en bij andere relevante partners.</p>	<p>De peiling wiskunde (basis-onderwijs) bevroagt bij de ouders de mate waarin ze tevreden zijn over de school. <i>Enkel ouders van leerlingen in het zesde leerjaar bevroagd.</i></p>	<p>PIRLS 2016 bevroagt bij de leerlingen hun welbevinden en betrokkenheid via de schaal: Students' Sense of School Belonging.</p> <ul style="list-style-type: none"> • <i>Die schaal meet niet het streven naar welbevinden en betrokkenheid, maar meet wel het welbevinden van de Vlaamse leerlingen (die deelnamen aan PIRLS 2016).</i> • <i>Enkel leerlingen uit het vierde leerjaar bevroagd.</i> 	<p>TIMSS 2019 bevroagt bij de leerlingen hun welbevinden en betrokkenheid via de schaal: Students' Sense of School Belonging.</p> <ul style="list-style-type: none"> • <i>Die schaal meet niet het streven naar welbevinden en betrokkenheid, maar meet wel het welbevinden van de Vlaamse leerlingen (die deelnamen aan TIMSS 2019).</i> • <i>Enkel leerlingen uit het vierde leerjaar bevroagd.</i> 	<p>PISA 2018 bevroagt bij leerlingen in welke mate ze zich thuis en op hun gemak voelen op school.</p> <ul style="list-style-type: none"> • <i>De schaal meet niet het streven naar welbevinden en betrokkenheid, maar wel de mate waarin leerlingen (die deelnamen aan PISA 2018) zich thuis voelen op school.</i> • <i>Enkel een beperkt aantal 15-jarige leerlingen bevroagd.</i> <p>PISA 2018 bevroagt bij ouders in welke mate ze tevreden zijn over de school. <i>Enkel ouders van beperkt aantal 15-jarige leerlingen bevroagd.</i></p>	<p>Terugkoppeling resultaten school: Bij de U7-schaal (Onderwijsleerpraktijk) wordt voor elke school in kaart gebracht in welke mate er indicaties van leerlingwelbevinden zijn. <i>Bij het gesprek met leerlingen (waarin gepeild wordt naar het welbevinden van leerlingen) gaat het niet om een representatieve steekproef van leerlingen op schoolniveau.</i></p>
<p>R3. De school streeft bij elke lerende naar zoveel mogelijk leerwinst.</p>					<p>(zie R1) <i>De onderwijsinspectie meet de leerwinst van elke leerling in de school niet, maar heeft aandacht voor de kwaliteitsbewaking van de onderwijsleerpraktijk op de school.</i></p>

	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
R4. De school stimuleert de studievoortgang van elke lerende.					Terugkoppeling resultaten school: Bij de L1-schaal (Leerlingenbegeleiding) wordt in kaart gebracht in welke mate de school acties onderneemt met oog op de onderwijsloopbaan van de leerlingen.
R5. De school waarborgt de toegang tot onderwijs voor elke lerende.					Terugkoppeling resultaten school: Bij de L1-schaal (Leerlingenbegeleiding) gaat de onderwijsinspectie na of de school een transparant en toegankelijk inschrijvingsbeleid hanteert.
R6. De school streeft naar effecten op langere termijn bij alle lerenden.					(zie R1) <i>De onderwijsinspectie meet niet de lange termijn effecten bij elke leerling in de school, maar heeft aandacht voor de kwaliteitsbewaking van de onderwijsleerpraktijk op de school.</i>

3.2.2.2 Kwaliteitsontwikkeling

	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
K1. Het Vlaamse onderwijs ontwikkelt haar kwaliteit vanuit een gedragen visie die vertaald is in de onderwijsleerpraktijk.					Terugkoppeling resultaten school: Bij de K1-schaal (Kwaliteitsontwikkeling) gaat de onderwijsinspectie na in welke mate er een school-eigen en gedragen visie aanwezig is die richtinggevend is voor de schoolwerking.

	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
K2. Het Vlaamse onderwijs evalueert haar werking cyclisch, systematisch en betrouwbaar vanuit de resultaten en effecten bij de lerenden.				PISA 2018 bevraagt in de directievragenlijst of de school aan evaluatie doet op initiatief van de school of omdat het verplicht is. <i>Enkel antwoord van de directeur.</i>	Terugkoppeling resultaten school: Bij de K4- en K5-schaal (Kwaliteitsontwikkeling) gaat de onderwijsinspectie na in welke mate de school systematisch en betrouwbaar de eigen schoolwerking evalueert.
K3. Het Vlaamse onderwijs borgt en ontwikkelt de kwaliteit van de onderwijsleerpraktijk.					Terugkoppeling resultaten school: Bij de K6-schaal (Kwaliteitsontwikkeling) gaat de onderwijsinspectie na in welke mate de school bewaart wat kwaliteitsvol is en haar werkpunten verbetert.

3.2.2.3 Ontwikkeling stimuleren

Doelen					
	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
D1. Het schoolteam realiseert doelgericht een brede en harmonische vorming die betekenisvol is.					Terugkoppeling resultaten school: De onderwijsinspectie brengt bij de U1- en U2-schalen (Onderwijsleerpraktijk) het volgende in kaart: <ol style="list-style-type: none"> 1. De volledigheid en evenwichtigheid van het aanbod in kaart gebracht per school. 2. In welke mate het aanbod betekenisvol is. 3. In welke mate het aanbod gebaseerd is op een gevalideerd doelenkader. 4. In welke mate het aanbod haalbaar en uitdagend is. 5. In welke mate de doelen geëxpliciteerd worden.
D2. Het schoolteam hanteert doelen die sporen met het gevalideerd doelenkader en zorgt voor samenhang tussen de doelen.					
D3. Het schoolteam hanteert uitdagende en haalbare doelen .					
D4. Het schoolteam expliciteert de doelen en de beoordelingscriteria .					

Vormgeving onderwijsleerproces en leef- en leeromgeving					
	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
V1. Het schoolteam en de lerende creëren samen een positief en stimulerend school- en klasklimaat .			TIMSS 2019 be vraagt bij de leerlingen en leerkrachten de schaal: Supportive Classroom Climate. <i>Enkel de leerkracht van en de leerlingen uit het vierde leerjaar be vraagd.</i>	PISA 2018 be vraagt bij leerlingen in welke mate er (1) concurrentie tussen leerlingen op school is en (2) belang gehecht wordt aan samenwerking tussen de leerlingen. <i>Enkel (de perceptie van een) beperkt aantal 15-jarige leerlingen op school.</i>	Terugkoppeling resultaten school: Bij de U3-schaal (Onderwijsleerpraktijk) wordt voor elke school in kaart gebracht in welke mate er respectvolle en stimulerende interactie aanwezig is op de school in kwestie.
V2. Het schoolteam en de lerenden gaan positief om met diversiteit .					Terugkoppeling resultaten school: Bij de D1-schaal (Omgaan met diversiteit) wordt voor elke school in kaart gebracht in welke mate: <ol style="list-style-type: none"> 1. Diversiteit als normaal beschouwd wordt. 2. Het schoolteam een positieve houding ten opzichte van diversiteit toont. 3. Het schoolteam de leerlingen begeleidt bij en aanzet tot de ontwikkeling van een positieve houding ten opzichte van diversiteit.
V3. De leef- en leeromgeving en de onderwijsorganisatie ondersteunen het bereiken van de doelen.					Terugkoppeling resultaten school: Bij de U2-schaal (Onderwijsleerpraktijk) wordt voor elke school in kaart gebracht in welke mate de onderwijsorganisatie het bereiken van de doelen ondersteunt.

	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
V4. Het schoolteam biedt een passend, actief en samenhangend onderwijsaanbod aan.	<p>De peilingen Nederlands en wiskunde basisonderwijs bevragen bij leerkrachten in welke mate er school-afspraken gemaakt worden over het onderwijsaanbod. <i>Enkel de leerkracht uit het zesde leerjaar bevroegd.</i></p> <p>De peilingen wiskunde A-stroom en B-stroom bevragen in de leerkrachtvragenlijst:</p> <ol style="list-style-type: none"> 1. Of er een horizontale en/of verticale vakgroepwerking is voor wiskunde en of de leerkrachten eraan deelnemen. 2. In welke mate er school-afspraken gemaakt worden voor wiskunde en over de leerlijnen. 3. In welke mate er wordt samengewerkt met collega's (van andere vakken). <p><i>Slechts één wiskundeleerkracht per klas uit de eerste graad secundair onderwijs bevroegd.</i></p>	<p>PIRLS 2016 bevroegt in de leerkrachtenvragenlijst hoe vaak er wordt samengewerkt met andere leerkrachten van andere leerjaren om doorlopende leerlijnen te verzekeren. <i>Enkel de leerkracht uit het vierde leerjaar bevroegd.</i></p>			<p>Terugkoppeling resultaten school: Bij de U2-schaal (Onderwijsleerpraktijk) wordt voor elke school in kaart gebracht in welke mate het aanbod samenhangend en activerend is.</p>

Begeleiding					
	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
B1. Het schoolteam geeft de begeleiding vorm vanuit een gedragen visie en systematiek en volgt de effecten van begeleiding op.					<p>Terugkoppeling resultaten school: Bij de L2-schaal (Leerlingenbegeleiding) gaat de onderwijsinspectie na in welke mate de school leerlingen begeleidt vanuit het zorgcontinuüm, effecten opvolgt en waar nodig bijstuurt.</p> <p>Bij de K1-schaal (Kwaliteitsontwikkeling) gaat de onderwijsinspectie na in welke mate er een schooleigen en gedragen visie aanwezig is die richtinggevend is voor de begeleiding van leerlingen.</p>
B2. Het schoolteam biedt begeleiding zowel op vlak van leren, leerloopbaan, sociaal-emotionele ontplooiing als fysiek welzijn .					<p>Terugkoppeling resultaten school: Bij de L1-schaal (Leerlingenbegeleiding) gaat de onderwijsinspectie na in welke mate de school initiatieven neemt om het leren en studeren, de onderwijsloopbaan, het psychisch en sociaal functioneren en de preventieve gezondheid van de leerlingen te stimuleren.</p>

	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
B3. Het schoolteam biedt elke lerende een passende begeleiding met oog op gelijke onderwijskansen.	<p>De peiling Nederlands (basis-onderwijs) be vraagt bij leerkrachten in welke mate er ondersteuning wordt geboden aan leerlingen met problemen gerelateerd aan het Nederlands of taal in het algemeen.</p> <p><i>Enkel de leerkracht uit het zesde leerjaar be vraagd.</i></p> <p>De peilingen Nederlands basisonderwijs, wiskunde basisonderwijs, wiskunde A-stroom en wiskunde B-stroom be vragen in de leerkrachtvragenlijst de mate waarin leerkrachten differentiëren in hun lessen.</p> <p><i>Slechts één leerkracht uit het zesde leerjaar of de eerste graad secundair onderwijs (per klas) be vraagd.</i></p>	<p>PIRLS 2016 be vraagt in de leerkrachtvragenlijst de mate waarin leerkrachten differentiëren tijdens leeslessen.</p> <p><i>Enkel de leerkracht uit het vierde leerjaar be vraagd.</i></p>		<p>PISA 2018 be vraagt in de directievragenlijst of de toetsresultaten van leerlingen gebruikt worden om het onderwijs aan te passen aan de noden van de leerlingen.</p> <p><i>Enkel antwoord van de directeur (per school).</i></p>	<p>Terugkoppeling resultaten school:</p> <p>Bij de U2- en U3-schaal (Onderwijsleerpraktijk) wordt voor elke school in kaart gebracht in welke mate er passende begeleiding geboden wordt met het oog op gelijke onderwijskansen.</p> <p>Bij de D2-schaal (Omgaan met diversiteit) wordt in kaart gebracht in welke mate het schoolteam inspeelt op de talige competenties en noden van de leerlingen.</p> <p>Bij de L2-schaal (Leerlingenbegeleiding) gaat de onderwijsinspectie na in welke mate de school zicht heeft op de noden van de leerlingen en op basis daarvan een passende begeleiding biedt gericht op gelijke kansen.</p>
B4. Het schoolteam geeft de begeleiding vorm samen met de lerende, de ouders/ thuisomgeving en andere relevante partners.	<p>De peilingen Nederlands en wiskunde (basisonderwijs) be vragen in de oudervragenlijst de mate waarin ouders betrokken zijn bij schoolactiviteiten en bij het huiswerk maken.</p> <p><i>Enkel ouders van leerlingen in het zesde leerjaar be vraagd</i></p>	<p>PIRLS 2016 be vraagt in de schoolvragenlijst en leerkrachtvragenlijst de mate waarin ouders betrokken zijn bij schoolactiviteiten.</p> <p><i>Enkel de leerkracht uit het vierde leerjaar en de directeur be vraagd.</i></p>		<p>PISA 2018 be vraagt in de directievragenlijst en in de oudervragenlijst in welke mate ouders betrokken worden om de vooruitgang van hun kind te bespreken of bij de algemene werking van de school.</p> <p><i>Enkel antwoord van directeur en beperkt aantal ouders van 15-jarige leerlingen.</i></p>	<p>Terugkoppeling resultaten school: Bij de L2- en L3-schaal (Leerlingenbegeleiding) gaat de onderwijsinspectie na in welke mate de school de begeleiding samen vormgeeft met de leerlingen, thuisomgeving, CLB, pedagogische begeleidingsdiensten, het ondersteuningsnetwerk en andere relevante partners.</p>

Opvolging					
	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
O1. Het schoolteam geeft de lerende adequate feedback met oog op de voortgang in het leer- en ontwikkelingsproces.					Terugkoppeling resultaten school: Bij de U5-schaal (Onderwijsleerpraktijk) wordt voor elke school in kaart gebracht in welke mate feedback ingebed is in het onderwijsleerproces.
O2. Het schoolteam evalueert op een brede en onderbouwde wijze het onderwijsleerproces en het behalen van de doelen.					Terugkoppeling resultaten school: Bij de U6-schaal (Onderwijsleerpraktijk) wordt voor elke school in kaart gebracht in welke mate het onderwijsleerproces afgestemd wordt op de evaluatiegegevens.
O3. Het schoolteam stuurt het onderwijsleerproces bij op basis van feedback- en evaluatiegegevens.				PISA 2018 bevraagt in de directievragenlijst of de toetsresultaten van de leerlingen gebruikt worden om het leerproces van de leerlingen bij te sturen. <i>Enkel antwoord van de directeur (per school).</i>	

	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
O4. Het schoolteam beslist en rapporteert onderbouwd over het behalen van de doelen bij de lerende.				<p>PISA 2018 bevraagt in de directievragenlijst en in de oudervragenlijst in welke mate ouders betrokken worden om de vooruitgang van hun kind te bespreken of bij de algemene werking van de school.</p> <p><i>Enkel antwoord van de directeur en beperkt aantal ouders van 15-jarige leerlingen.</i></p>	<p>Terugkoppeling resultaten school:</p> <p>Bij de U6-schaal (Onderwijs-leerpraktijk) wordt voor elke school in kaart gebracht in welke mate de evaluatie spooft met de doelen; en de evaluatie transparant en betrouwbaar is.</p> <p>Bij de R1-, R2- en R3-schaal (Rapportering en oriëntering) gaat de onderwijsinspectie na in welke mate:</p> <ol style="list-style-type: none"> 1. De school relevant, breed en transparant rapporteert over het bereiken van de doelen en het leer- en ontwikkelingsproces van de leerlingen. 2. De beslissingen onderbouwd zijn en gebaseerd zijn op het bereiken van de doelen met oog op de onderwijs-loopbaan van de leerlingen. 3. Adviezen m.b.t. de oriëntering onderbouwd zijn en geformuleerd worden in functie van de onderwijsloopbaan, studieresultaten, interesses en de mening van leerling en ouders.

3.2.2.4 Beleid

Beleid					
	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
BL1. De school ontwikkelt en voert een gedragen, geïntegreerd en samenhangend beleid rekening houdend met haar (ped)agogisch project.					Terugkoppeling resultaten school: Bij de K1- en K3-schaal (Kwaliteitsontwikkeling) gaat de onderwijsinspectie na in welke mate: <ol style="list-style-type: none"> 1. Er een schooleigen en gedragen visie aanwezig is die richtinggevend is voor de schoolwerking. 2. Het beleid van de school samenhangend is.
BL2. De school geeft haar organisatie vorm op het vlak van cultuur en structuur .					Terugkoppeling resultaten school: Bij de K2-schaal (Kwaliteitsontwikkeling) gaat de onderwijsinspectie na in welke mate:
BL3. De school werkt participatief en responsief .					<ol style="list-style-type: none"> 1. De school een lerende organisatiecultuur stimuleert. 2. Een responsief beleid ontwikkelt waarin participatie en dialoog centraal staan.
BL4. In de school heerst een innovatieve en lerende organisatiecultuur .					<ol style="list-style-type: none"> 3. De school een innovatieve organisatiecultuur stimuleert.

	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
BL5. De school bouwt samenwerkingsverbanden uit die leren en onderwijzen ten goede komen.				PISA 2018 be vraagt bij leerlingen of er activiteiten zijn waarbij wordt samengewerkt met lokale bibliotheken of kranten. <i>Enkel (beperkt aantal) 15-jarige leerlingen per school be vraagd.</i>	Terugkoppeling resultaten school: Bij de L3- en L4-schaal (Leerlingenbegeleiding) gaat de onderwijsinspectie na: 1. In welke mate de school samenwerkt met het CLB. 2. In welke mate de school beroep doet op pedagogische begeleidingsdiensten, andere externe diensten of op het ondersteuningsnetwerk. Bij de K2-schaal (Kwaliteitsontwikkeling) gaat de onderwijsinspectie na in welke mate de school samenwerkingsverbanden aangaat om zowel de onderwijsleerpraktijk als schoolwerking te versterken.
BL6. De school communiqueert transparant over haar werking met alle betrokkenen.	De peiling wiskunde (basisonderwijs) be vraagt bij de ouders in welke mate de communicatie met de school vlot verloopt. <i>Enkel ouders van leerlingen in het zesde leerjaar be vraagd</i>				Bij de K2-schaal (Kwaliteitsontwikkeling) gaat de onderwijsinspectie na in welke mate de school transparant, doelgericht en tijdig communiqueert over haar werking met alle betrokkenen.

Onderwijskundig beleid					
	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
BL7. De school ontwikkelt en voert een doeltreffend beleid op het vlak van leren en onderwijzen .					Terugkoppeling resultaten school: Bij de K3-schaal (Kwaliteitsontwikkeling) gaat de onderwijsinspectie na in welke mate de school een doelgericht beleid ontwikkelt en voert.

Personeels- en professionaliseringsbeleid					
	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
BL8. De school ontwikkelt en voert een doeltreffend personeelsbeleid dat integraal en samenhangend is.					<p>Terugkoppeling resultaten school:</p> <p>Bij de P1-schaal (Personeelsbeleid en professionalisering) gaat de onderwijsinspectie na in welke mate de school:</p> <ol style="list-style-type: none"> 1. Duidelijke criteria hanteert. 2. Rekening houdt met competenties die de schoolwerking kunnen versterken bij de selectie en aanwerving van personeel. <p>Bij de P2-schaal (Personeelsbeleid en professionalisering) gaat de onderwijsinspectie na in welke mate de school aandacht heeft voor coaching; transparant, motiverend en rechtvaardig evalueren en het inzetten van personeel volgens hun competenties.</p>

	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
BL9. De school ontwikkelt en voert een doeltreffend professionaliseringsbeleid en heeft hierbij specifieke aandacht voor beginnende leraren .	<p>Alle geanalyseerde peilingen bevragen bij leerkrachten op welke gebieden ze zich de afgelopen twee jaar bijgeschoold hebben. <i>Slechts één leerkracht (per klas) bevroegd.</i></p> <p>De peiling Nederlands peilt bij leerkrachten naar de mate waarin inzichten uit nascholingen gedeeld worden met collega's en geïntegreerd worden in het schoolbeleid. <i>Slechts één leerkracht (per klas) bevroegd.</i></p> <p>De peiling wiskunde (basisonderwijs) bevroegt:</p> <ol style="list-style-type: none"> 1. Aan welke professionaliseringsvorm en hoe vaak leerkrachten deelnamen in de voorbije jaren. 2. De mate waarin de inhoud van de professionalisering toegepast kan worden in de klas. 3. Aan welke inhoud en vorm van professionalisering de leerkrachten nood hebben. <p><i>Slechts één leerkracht (per klas) bevroegd.</i></p> <p>De peilingen A- en B- stroom bevragen bij leerkrachten hoeveel uur ze de afgelopen twee jaar besteed hebben aan formele professionaliseringsactiviteiten. <i>Slechts één leerkracht (per klas) bevroegd.</i></p>	<p>PIRLS 2016 bevroegt bij leerkrachten hoeveel tijd ze de afgelopen twee jaar gependeed hebben aan professionele ontwikkeling gerelateerd aan lezen via de schaal: Teacher Time Spent on Professional Development Related to Reading in the Past Two Years. <i>Enkel leerkracht vierde leerjaar bevroegd.</i></p>	<p>TIMSS 2019 bevroegt bij leerkrachten in welke mate ze tijdens de laatste twee jaren participeren in professionele ontwikkeling via de schaal: Teachers' Participation in Professional Development. <i>Enkel leerkracht vierde leerjaar bevroegd.</i></p> <p>TIMSS 2019 bevroegt bij leerkrachten ook de noden voor toekomstige professionele ontwikkeling via de schaal: Teachers' Professional Development Needs. <i>Enkel leerkracht vierde leerjaar bevroegd.</i></p>	<p>PISA 2018 bevroegt in de directievragenlijst</p> <ol style="list-style-type: none"> 1. Of de school inzet op een mentorschap voor leerkrachten. 2. Welk percentage leerkrachten tijdens de laatste drie maanden heeft deelgenomen aan een bijscholingscursus. <p><i>Enkel antwoord van de directeur (per school)</i></p>	<p>Terugkoppeling resultaten school: Bij de P3- en P4-schaal (Personeelsbeleid en professionalisering) gaat de onderwijsinspectie de volgende zaken na:</p> <ol style="list-style-type: none"> 1. In welke mate de school de professionaliseringsnoden van het team in kaart brengt. 2. In welke mate de school professionalisering aanmoedigt en inplant. 3. In welke mate er expertise gedeeld en ontwikkeld wordt. 4. In welke mate het effect van de professionalisering zichtbaar is in het handelen van het schoolteam. 5. In welke mate de school inzet op aanvangsbegeleiding voor beginnende teamleden. <p>Bij de K3-schaal (Kwaliteitsontwikkeling) gaat de onderwijsinspectie na in welke mate de school een doelgericht beleid ontwikkelt en voert met oog op de professionalisering van het schoolteam.</p>

Financieel en materieel beleid					
	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
BL10. De school ontwikkelt en voert een doeltreffend financieel en materieel beleid .					
BL11. De school beheerst de kosten voor alle lerenden.				PISA 2018 bevroagt in de oudervragenlijst hoeveel geld de ouders uitgegeven hebben aan de scholing van hun kind tijdens het afgelopen jaar. <i>Enkel ouders van beperkt aantal 15-jarige leerlingen bevroagd.</i>	

Fysieke en mentale veiligheid van de leef-, leer- en werkomgeving					
	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
BL12. De school ontwikkelt en voert een doeltreffend beleid met oog op de fysieke en mentale veiligheid van de leef-, leer- en werkomgeving .		PIRLS 2016 bevroagt bij leerkrachten de veiligheid via de schaal: Safe and Orderly School. <i>Enkel leerkracht vierde leerjaar bevroagd.</i>	TIMSS 2019 bevroagt bij leerkrachten de veiligheid via de schaal: Safe and Orderly School. <i>Enkel leerkracht vierde leerjaar bevroagd.</i>		Bij de BVH-schalen (BVH1, BVH2, BVH3 en BVH4) wordt voor elke school in kaart gebracht in welke mate ze systemisch, preventief en onderbouwd alle processen en bijhorende acties evalueert en bijstuurt.

3.2.3 Uitspraken op leerlingniveau (geen aspect van onderwijskwaliteit)

3.2.3.1 Resultaten en effecten

	Peilingen	PIRLS	TIMSS	PISA	Doorlichtingen
R1. De school bereikt de minimaal gewenste output bij een zo groot mogelijke groep van lerenden.	Uitspraken over resultaten op leerlingniveau bij paralleltoetsen. <i>Dat leidt echter tot grote betrouwbaarheidsintervallen.</i>	<i>Plausible values zijn niet geschikt voor uitspraken op leerlingniveau.</i>	<i>Plausible values zijn niet geschikt voor uitspraken op leerlingniveau.</i>	<i>Plausible values zijn niet geschikt voor uitspraken op leerlingniveau.</i>	
R2. De school streeft naar welbevinden en betrokkenheid bij alle lerenden en het schoolteam en naar tevredenheid bij ouders en bij andere relevante partners.					
R3. De school streeft bij elke lerende naar zoveel mogelijk leerwinst .					
R4. De school stimuleert de studievoortgang van elke lerende.					
R5. De school waarborgt de toegang tot onderwijs voor elke lerende.					
R6. De school streeft naar effecten op langere termijn bij alle lerenden.					

3.3 Besluit

De matrix omvat heel wat lege cellen en toont dus dat geen enkele bron uitspraken kan doen over alle aspecten van onderwijskwaliteit. Elke bron is namelijk ontwikkeld om uitspraken te doen op een specifiek aggregatieniveau. Daardoor gaan uitspraken op andere aggregatieniveaus vaak samen met beperkingen. Bijvoorbeeld, de internationaal vergelijkende onderzoeken zijn gericht op het systeemniveau en focussen op leerlingen in een bepaald leerjaar of met een specifieke leeftijd. Om correcte uitspraken te doen over die populatie op systeemniveau is het belangrijk dat de steekproef op systeemniveau – en niet op schoolniveau – representatief is. Bijgevolg worden uitspraken op basis van die internationale onderzoeken op schoolniveau steeds gekenmerkt door dezelfde beperking: ze zijn gebaseerd op slechts één leerjaar (TIMSS en PIRLS) of een beperkt aantal 15-jarige leerlingen binnen een school (PISA). De resultaten in de feedbackrapporten kunnen daardoor niet veralgemeend worden naar de hele school. Het voorgaande voorbeeld illustreert bovendien dat de beperkingen in de matrix nooit los staan van keuzes die gemaakt zijn bij het design, de standaardbepaling en de analytische methode (zie 2 Horizontale analyse).

Van belang om op te merken is dat de peilingen, de doorlichtingen, PIRLS, TIMSS en PISA ook de contextfactoren op school- en systeemniveau in kaart brengen. In lijn met het OK-kader zijn die contextfactoren zelf geen aspecten van onderwijskwaliteit. Bijgevolg worden ze niet opgenomen in de matrix. Toch zijn die contextfactoren van belang: rekening houden met de context van een onderwijssysteem of een specifieke school is een voorwaarde om eerlijke uitspraken te kunnen doen over de aspecten van onderwijskwaliteit in de matrix.

Module 2.2: Mate van convergentie of divergentie tussen de resultaten van verschillende bronnen

In deze module onderzoeken we in welke mate de resultaten van verschillende bronnen van externe kwaliteitszorg, i.e., de doorlichtingen, het peilingsonderzoek en het internationaal vergelijkend onderzoek, convergeren of divergeren (Onderzoeksvraag 2). In wat volgt, bespreken we drie cross-instrumentele studies:

- Studie 1: PIRLS Repeat
- Studie 2: Scholen met meer dan één recente bron van kwaliteitsindicatoren
- Studie 3: Vergelijking van trends op systeemniveau

De eerste studie heeft betrekking op zowel het systeem- als schoolniveau. De tweede en derde studie daarentegen richten zich op een specifiek aggregatieniveau. De tweede cross-instrumentele studie focust op het schoolniveau, terwijl de derde studie inzoomt op het systeemniveau.

Studie 1: PIRLS Repeat

1. Inleiding

Uit PIRLS 2016 bleek dat Vlaanderen tussen 2006 en 2016 de grootste daler in begrijpend lezen was van alle participerende West-Europese landen (Mullis et al., 2017). Daarom besloot de Vlaamse overheid om in 2018 een herhalingsmeting te organiseren van PIRLS bij dezelfde leerlingen die op dat moment in het zesde leerjaar zaten. Die herhalingsmeting, genaamd PIRLS Repeat of PIRLS 2018 L6, had verschillende doelen waaronder nagaan hoe de PIRLS-meetschaal voor begrijpend lezen zich verhoudt ten opzichte van de schaal voor lezen uit het Vlaamse peilingsonderzoek. Om dat doel te bereiken, werden de toetsen van PIRLS Repeat samengesteld met items uit zowel PIRLS 2016 als de peiling Nederlands in het basisonderwijs 2018. Door dat uitzonderlijke onderzoeksopzet en toetsdesign is het mogelijk om de data van PIRLS Repeat op twee verschillende schalen te plaatsen, namelijk de peilingschaal en de PIRLS-schaal, en de samenhang tussen die twee schalen te onderzoeken (Dockx et al., 2019).

2. Onderzoeksopzet

Het doel van de huidige studie is om na te gaan of de resultaten van PIRLS Repeat, het peilingsonderzoek en de doorlichtingen convergeren of divergeren.

In een eerste vergelijking wordt de concordantie tussen de PIRLS- en de peilingschaal onderzocht. De resultaten op die schalen kunnen immers met elkaar vergeleken worden via de data van de 126 scholen die deelnamen aan PIRLS Repeat.

In een tweede vergelijking wordt de concordantie tussen de peilingschaal, de PIRLS-schaal en de resultaten van de doorlichtingen onderzocht. Van de 126 scholen die deelnamen aan PIRLS Repeat zijn er 25 scholen die aan de hand van het OK-kader doorgelicht werden voor Nederlands in schooljaar 2018-2019 of 2019-2020. Voor die scholen kunnen de schooleffecten op de PIRLS-schaal en de peilingschaal alsook de resultaten van de doorlichtingen met elkaar vergeleken worden.

2.1 PIRLS- en peilingschaal (Vergelijking 1)

Bij de eerste vergelijking schatten we het schooleffect van de 126 PIRLS Repeat scholen op de PIRLS- en de peilingschaal, waarna we de schooleffecten op die twee schalen vergelijken. Het schooleffect is

de bijdrage die een school levert aan de prestaties van zijn leerlingen. Echter, de prestaties van leerlingen worden niet alleen beïnvloed door de school. Er zijn ook andere kenmerken die een invloed hebben op de prestaties van leerlingen. Zo is het bijvoorbeeld voor school A, waar de overgrote meerderheid van de leerlingen thuis Nederlands spreekt, waarschijnlijk eenvoudiger om een hoge score te behalen op de toets begrijpend lezen dan voor school B, waar de meeste leerlingen thuis een andere taal spreken. Daarom maken we in deze studie gebruik van toegevoegde waarde modellen. Die modellen houden rekening met factoren waarover een school geen controle heeft, maar die wel een invloed hebben op de leerlingprestaties (Timmermans et al., 2011).

In de eerste vergelijking stellen we vijf verschillende modellen op. Elk van die vijf modellen leidt tot een ander schooleffect, omdat ze controleren voor een ander soort factoren waarop scholen geen invloed hebben. Concreet vergelijken we een nulmodel (of type 0-model) en vier toegevoegde waarde modellen, namelijk het type AA-, type A-, type B- en type X-model. Daarbij worden er per type steeds meer controlevariabelen opgenomen (Leckie & Prior, 2022; Timmermans et al., 2011):

Bij een **type 0-model** worden de ruwe gemiddeldes van scholen vergeleken. Daarbij wordt er niet gecontroleerd voor factoren waarop scholen geen invloed hebben. Kortom, een type 0-schooleffect geeft aan hoe goed de leerlingen van een bepaalde school presteren in vergelijking met andere scholen zonder te controleren voor de achtergrondkenmerken van leerlingen of scholen.

Een **type AA-model** is een model waarbij er gecontroleerd wordt voor eerdere prestaties. Na controle voor eerdere prestaties blijft het effect van de school over dat niet te wijten is aan die controlevariabele. Concreet geeft een type AA-schooleffect dus weer hoe goed de leerlingen in een bepaalde school scoren in vergelijking met leerlingen uit andere scholen die een gelijkaardige eerdere prestatie hebben. De unieke dataset van PIRLS Repeat die de prestaties voor begrijpend lezen in zowel het vierde als zesde leerjaar bevat, laat ons toe een type AA-model te schatten.

Een **type A-model** controleert voor eerdere prestaties en leerlingkenmerken. Na controle voor die variabelen blijft het type A-schooleffect over. Dat schooleffect geeft dus aan hoe goed de leerlingen in een bepaalde school scoren in vergelijking met gelijkaardige leerlingen in de andere scholen.

Bij een **type B-model** wordt er gecontroleerd voor eerdere prestaties, leerlingkenmerken en schoolcompositiekenmerken. Het schooleffect is niet te wijten aan die controlevariabelen. Met als gevolg, een type B-schooleffect toont hoe goed de leerlingen in een bepaalde school scoren in vergelijking met gelijkaardige leerlingen in scholen met een gelijkaardige schoolcompositie.

Bij een **type X-model** worden de volgende controlevariabelen opgenomen: eerdere prestaties, leerlingkenmerken, schoolcompositiekenmerken en schoolkenmerken die de school niet kan beïnvloeden (bijvoorbeeld provincie, onderwijsnet...). Na controle voor die variabelen blijft het type X-schooleffect over. Dat schooleffect geeft weer hoe goed de leerlingen in een bepaalde school scoren in vergelijking met gelijkaardige leerlingen in scholen met een gelijkaardige schoolcompositie en gelijkaardige niet-beïnvloedbare schoolkenmerken.

Om het schooleffect van de 126 scholen uit PIRLS Repeat volgens de PIRLS- en peilingschaal te vergelijken, wordt elk van de vijf bovenstaande modellen twee keer geschat: eenmaal met lezen op de

PIRLS-schaal²⁰ en eenmaal met lezen op de peilingschaal. Concreet hebben we – op basis van de controlevariabelen die gehanteerd werden bij de peilingen, PIRLS 2016 en PIRLS 2018 L6 – besloten om in deze studie onderstaande controlevariabelen (zie Tabel 3) op te nemen bij de verschillende modellen.

Tabel 3

Overzicht van opgenomen controlevariabelen per type toegevoegde waarde model

Type 0-model	Type AA-model	Type A-model	Type B-model	Type X-model
Geen controle-variabelen	Eerdere leesprestaties (4 ^e leerjaar)	<ul style="list-style-type: none"> • Eerdere leesprestaties (4^e leerjaar) • Thuistaal • Geslacht • SES • Leeftijd 	<ul style="list-style-type: none"> • Eerdere leesprestaties (4^e leerjaar) • Thuistaal • Geslacht • SES • Leeftijd • Gemiddelde eerdere schoolprestatie voor lezen • Gemiddelde SES van de school 	<ul style="list-style-type: none"> • Eerdere leesprestaties (4^e leerjaar) • Thuistaal • Geslacht • SES • Leeftijd • Gemiddelde eerdere schoolprestatie voor lezen • Gemiddelde SES van de school • Schoolgrootte • Onderwijsnet • Provincie

Noot. De SES van de leerlingen werd gemeten aan de hand van de ‘Home Resources for Learning’-schaal. Die schaal omvat vijf items: (1) aantal boeken thuis, (2) aantal kinderboeken thuis, (3) beschikken over toegang tot het internet en een eigen kamer, (4) opleidingsniveau ouders en (5) tewerkstellingsniveau ouders (Martin et al., 2017).

Zoals eerder aangegeven, blijft na controle voor de bovenstaande variabelen – afhankelijk van het type model – een schooleffect (of residu) over dat niet te wijten is aan die variabelen. Op basis van die residuen en de bijhorende betrouwbaarheidsintervallen kunnen we de scholen indelen in drie groepen: scholen met (1) een bovengemiddeld schooleffect, (2) een gemiddeld schooleffect en (3) een ondergemiddeld schooleffect. Op die manier kunnen we nagaan of er scholen zijn waarbij hun effectiviteit volgens de PIRLS-schaal verschilt van hun effectiviteit volgens de peilingschaal. Tot slot wordt ook de correlatie tussen de residuen op de PIRLS- en peilingschaal voor elk van de vijf modellen vergeleken.

2.2 PIRLS, peiling en doorlichtingen (Vergelijking 2)

In een tweede vergelijking wordt de concordantie tussen de resultaten op de peilingschaal, de PIRLS-schaal en de doorlichtingen onderzocht. Daarvoor bouwen we verder op de resultaten uit de eerste vergelijking. Concreet geven we voor de 25 scholen (die deelnamen aan PIRLS Repeat en doorgelicht zijn) de schooleffecten op de PIRLS- en peilingschaal weer. Die schooleffecten vergelijken we vervolgens met de inschaling die de scholen van de onderwijsinspectie ontvangen hebben op de U7-

²⁰ Zoals eerder vermeld, wordt er bij de PIRLS-schaal gewerkt met vijf plausible values. In de analyses werd er verder gewerkt met die vijf verschillende plausible values. Concreet wordt er bij elk type model voor elk van de vijf plausible values een apart toegevoegde waarde model geschat. Daarna worden die aparte modellen per type samengevoegd aan de hand van de methode voor *multiple imputations* (Graham et al., 2007).

schaal 'leereffecten'²¹ voor het leergebied Nederlands. Bij die inschaling hanteert de onderwijsinspectie vier categorieën: (1) beneden de verwachting, (2) benadert de verwachting, (3) volgens de verwachting en (4) overstijgt de verwachting. Tot slot vergelijken we de inschaling van de onderwijsinspectie met het aandeel leerlingen dat de eindtermen bereikt volgens de peilingschaal enerzijds en met het percentage leerlingen van school dat de benchmarks van de PIRLS-schaal bereikt anderzijds.

3. Resultaten PIRLS- en peilingschaal (Vergelijking 1)

3.1 Modellen op de PIRLS- en peilingschaal

Zoals eerder aangegeven, hebben we elk van de vijf types modellen tweemaal opgesteld. De resultaten daarvan zijn weergegeven in de volgende tabellen:

- Tabel 4 omvat een samenvattend overzicht van de verdeling van de onverklaarde variantie en het percentage verklaarde variantie voor de verschillende types toegevoegde waarde modellen op zowel de PIRLS- als peilingschaal.
- Tabel 5 en Tabel 6 tonen de uitgebreide resultaten van de toegevoegde waarde modellen voor de PIRLS-schaal enerzijds en de peilingschaal anderzijds.

De resultaten van het **type 0-model** (zie Tabel 4) tonen aan dat het grootste deel van de onverklaarde variantie in toetsresultaten zich op leerlingniveau bevindt: 88,80% bij de PIRLS-schaal en 92,28% bij de peilingschaal. De overige variantie – respectievelijk 11,20% bij de PIRLS-schaal en 7,72% bij de peilingschaal – situeert zich op schoolniveau. Dat betekent concreet dat de verschillen in toetsresultaten tussen leerlingen van eenzelfde school vrij groot zijn (variantie op leerlingniveau), maar dat de verschillen tussen scholen relatief beperkt – maar wel significant – zijn (variantie op schoolniveau).

Een deel van die variantie op school- en leerlingniveau (uit het type 0-model, zie Tabel 4) wordt verklaard door de controlevariabelen opgenomen in de opeenvolgende modellen. Bij het **type AA-model** op de PIRLS-schaal verklaren de eerdere prestaties van de leerlingen 59,96% van de eerder gevonden verschillen tussen scholen en 29,32% van de eerder gevonden verschillen tussen leerlingen. Op de peilingschaal verklaren de eerdere prestaties 65,59% en 37,86% van de variantie op respectievelijk school- en leerlingniveau. Uit de modellen op beide schalen (zie Tabel 5 en Tabel 6) blijkt bovendien dat leerlingen met een hogere score op PIRLS 2016 (eerdere prestatie) significant hoger scoren op de toets van PIRLS Repeat 2018.

In het **type A-model** worden eerdere prestaties en enkele bijkomende leerlingkenmerken opgenomen. Doordat er bij het type A-model meer controlevariabelen opgenomen worden dan bij het type AA-model stijgt het percentage variantie dat verklaard wordt door die controlevariabelen. Op de PIRLS-schaal verklaren die variabelen samen 62,95% van de variantie op schoolniveau en 32,71% van de totale variantie op leerlingniveau. De opgenomen controlevariabelen verklaren op de peilingschaal 68,83% en 41,16% van de variantie op respectievelijk school- en leerlingniveau. Daarnaast wijzen Tabel 5 en Tabel 6 op significante effecten bij beide schalen voor eerdere prestaties, geslacht (in het voordeel van de meisjes), veel SES-middelen (in vergelijking met enkele middelen) en schoolse vertraging (in vergelijking met normaalvorderende leerlingen).

²¹ Aan de hand van de U7-schaal leereffecten brengt de onderwijsinspectie bij elke school voor een bepaald leergebied (basisonderwijs) het volgende in kaart: in welke mate tonen de onderwijsleerpraktijk, het welbevinden van de leerlingen en de kwaliteitsbewaking van de onderwijsleerpraktijk aan dat een zo groot mogelijke groep van leerlingen de minimaal gewenste output bereikt (Onderwijsinspectie, 2021c).

Wat betreft **het type B-model** verklaren de opgenomen controlevariabelen 66,93% van de verschillen tussen scholen bij de PIRLS-schaal. Op leerlingniveau wordt er 33,08% van de verschillen verklaard. Op de peilingschaal wordt van de variantie op schoolniveau 71,32% verklaard, terwijl op leerlingniveau 41,16% verklaard wordt. Ook bij het type B-model geven Tabel 5 en Tabel 6 op beide schalen significante effecten aan voor eerdere prestaties, geslacht, veel SES-middelen (in vergelijking met enkele middelen) en schoolse vertraging (in vergelijking met normaalvorderende leerlingen). Daarenboven toont het type B-model op de peilingschaal (zie Tabel 6) ook een significant effect van school SES. Dat wil zeggen dat leerlingen die les volgen in een school waar het leerlingenpubliek gekenmerkt wordt door een hogere SES gemiddeld genomen beter presteren op de toets van PIRLS Repeat.

Bij het **type X-model** op de PIRLS-schaal verklaren alle controlevariabelen samen 72,96% van de totale variantie op schoolniveau en 32,60% van de totale variantie op leerlingniveau. Op de peilingschaal worden 77,06% en 40,70% van de totale variantie op respectievelijk school- en leerlingniveau verklaard. Op zowel de PIRLS-schaal als de peilingschaal (zie Tabel 5 en Tabel 6) vinden we dezelfde significante effecten als bij het type B-model voor die schalen.

Over het algemeen zijn de resultaten van de verschillende types toegevoegde waarde modellen voor de PIRLS- en de peilingschaal gelijkaardig. Al zijn er ook enkele verschillen tussen beide schalen. Zo wordt er bij de peilingschaal bij elk type model een groter percentage variantie verklaard door de opgenomen controlevariabelen dan bij de PIRLS-schaal. Daarnaast is er – zoals hierboven reeds vermeld – geen significant effect van school SES bij de PIRLS-schaal, maar wel bij de peilingschaal.

Tabel 4

Variantieverdeling van de onverklaarde variantie en percentage verklaarde variantie voor de verschillende types toegevoegde waarde modellen op de PIRLS- en peilingschaal

	% Onverklaarde variantie	% Verklaarde variantie t.o.v. type 0-model			
		Type 0	Type AA	Type A	Type B
PIRLS-schaal					
Schoolniveau	11,20	59,96	62,95	66,93	72,96
Leerlingniveau	88,80	29,32	32,71	33,08	32,60
Peilingschaal					
Schoolniveau	7,72	65,59	68,83	71,32	77,06
Leerlingniveau	92,28	37,86	41,16	41,16	40,70

Tabel 5

Schatting nulmodel en toegevoegde waarde modellen op de PIRLS-schaal

	Type 0		Type AA		Type A		Type B		Type X						
	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE					
Fixed effecten															
Intercept	584,581	***	2,011	582,562	***	1,509	586,893	***	1,768	586,376	***	55,435	586,039	***	66,273
Eerdere leesprestatie				0,533	***	0,014	0,492	***	0,018	0,488	***	0,018	0,486	***	0,019
Geslacht (meisjes)							-10,720	***	1,642	-10,757	***	1,634	-10,923	***	1,707
SES (enkele middelen)															
Veel middelen							10,213	***	2,016	9,486	***	2,125	9,224	***	2,448
Weinig middelen							-8,956		8,012	-7,515		8,004	-7,427		7,963
Thuis taal (altijd Nederlands)															
Meestal Nederlands							-1,759		3,213	-1,559		3,204	-1,561		3,541
Soms Nederlands							-2,238		2,331	-1,979		2,422	-1,692		2,664
Nooit Nederlands							-1,036		6,080	-0,928		5,977	0,717		6,667
Leeftijd (op leeftijd)															
Voorsprong							9,239		7,392	8,951		7,379	11,473		7,845
Vertraging							-15,768	***	3,312	-15,510	***	3,313	-16,971	***	3,916
Eerdere schoolprestatie voor lezen										0,055		0,078	0,130		0,228
School SES										-16,424		10,753	-18,222		12,393
Provincie (Antwerpen)															
Brussels Hoofdstedelijk Gewest													-3,290		9,774
Limburg													0,022		4,623
Oost-Vlaanderen													2,966		3,897
Vlaams-Brabant													-1,802		4,758
West-Vlaanderen													-0,702		4,130
Onderwijsnet (Vrij gesubsidieerd onderwijs)															
Gemeenschapsonderwijs														-3,714	5,356
Officieel gesubsidieerd onderwijs														-1,035	3,204
Schoolgrootte														0,008	0,010
Random effecten															
Variantie op schoolniveau	368,591	60,205		147,576	30,557		136,577	29,220		121,909	27,318		99,666	29,441	
Variantie op leerlingniveau	2923,835	82,742		2066,600	68,794		1967,322	73,799		1956,550	77,347		1970,655	76,125	
Model fit															
Aantal leerlingen in model	4615 van 4615		4052 van 4615		3638 van 4615		3638 van 4615		3638 van 4615		3243 van 4615				

Noot. De referentiegroep wordt tussen haakjes vermeld.

Het effect is statistisch significant op een significantieniveau van .05; ** op een significantieniveau van .01; * op een significantieniveau van .001*

Tabel 6

Schatting nulmodel en toegevoegde waarde modellen op de peilingschaal

	Type 0		Type AA		Type A		Type B		Type X						
	Effect	SE	Effect	SE	Effect	SE	Effect	SE	Effect	SE					
Fixed effecten															
Intercept	.38653	***	.00668	.37818	***	.00442	.39694	***	.00598	.50472	***	.05354	.49865	***	.06449
Eerdere leesprestatie				.61719	***	.01231	.57977	***	.01360	.57965	***	.01383	.57798	***	.01466
Geslacht (meisjes)							-.03843	***	.00565	-.03834	***	.00565	-.03773	***	.00600
SES (enkele middelen)															
Veel middelen							.03615	***	.00657	.03307	***	.00672	.03277	***	.00716
Weinig middelen							-.02802		.02584	-.02374		.02590	-.02131		.02695
Thuis taal (altijd Nederlands)															
Meestal Nederlands							-.00950		.00801	-.00916		.00962	-.01137		.01021
Soms Nederlands							-.01152		.00801	-.01188		.00807	-.01090		.00860
Nooit Nederlands							-.00552		.01634	-.00714		.01642	.00322		.01750
Leeftijd (op leeftijd)															
Voorsprong							.03506		.02590	.03374		.02590	.04341		.02694
Vertraging							-.07628	***	.01073	-.07544	***	.01073	-.08092	***	.01152
Eerdere schoolprestatie voor lezen										-.02891		.06000	-.03326		.06561
School SES										-.06005	*	.02832	-.06699	*	.03129
Provincie (Antwerpen)															
Brussels Hoofdstedelijk Gewest													-.00642		.03267
Limburg													.00289		.01463
Oost-Vlaanderen													.01279		.01245
Vlaams-Brabant													-.00858		.01492
West-Vlaanderen													.00212		.01317
Onderwijsnet (Vrij gesubsidieerd onderwijs)															
Gemeenschapsonderwijs													-.01009		.01607
Officieel gesubsidieerd onderwijs													-.00105		.01048
Schoolgrootte													.00004		.00003
Random effecten															
Variantie op schoolniveau	.00401	.00070	.00138	.00030	.00125	.00029	.00115	.00028	.00092	.00026					
Variantie op leerlingniveau	.04791	.00101	.02977	.00067	.02819	.00067	.02819	.00067	.02841	.00072					
Model fit															
Aantal leerlingen in model	4615 van 4615		4052 van 4615		3638 van 4615		3638 van 4615		3243 van 4615						

Noot. De referentiegroep wordt tussen haakjes vermeld.

Het effect is statistisch significant op een significantieniveau van .05; ** op een significantieniveau van .01; * op een significantieniveau van .001*

3.2 Schooleffecten volgens de PIRLS- en peilingschaal

Na de schatting van de verschillende types modellen (zie 3.1 Modellen op de PIRLS- en peilingschaal) kunnen we de verschillende types schooleffecten in kaart brengen voor de deelnemende scholen. Hieronder worden die schooleffecten voor zowel de PIRLS- als de peilingschaal gevisualiseerd aan de hand van rupsgrafieken. Die rupsgrafieken worden als volgt gelezen:

- De nullijn geeft de verwachte score volgens het type model aan. Aangezien we in deze studie gebruik maken van verschillende types, varieert de betekenis van die nullijn:
 - In een nulmodel (type 0-model) duidt de nullijn de gemiddelde score van alle deelnemende scholen aan. Dat is gelijk aan het algemeen Vlaams gemiddelde, aangezien de steekproef representatief is voor de populatie van Vlaamse scholen.
 - Bij de toegevoegde waarde modellen (type AA-, type A-, type B- en type X-model) verwijst de nullijn naar de gemiddelde score die we statistisch gezien verwachten als er rekening gehouden wordt met de opgenomen controlevariabelen.
- Elke school wordt voorgesteld aan de hand van een ▲-teken. Concreet geeft het ▲-teken voor elke school het residu weer. Dat is de mate waarin het gemiddelde van die school afwijkt van de verwachte score volgens het model (de nullijn).
- De haakjes rond elk ▲-teken stellen het 95%-betrouwbaarheidsinterval voor.

Het schooleffect van een bepaalde school wordt in de rupsgrafieken dus visueel voorgesteld als de afstand tussen ▲-teken enerzijds en de nullijn anderzijds. Via die rupsgrafieken worden de deelnemende scholen dus vergeleken met het algemeen Vlaams gemiddelde (nulmodel) of het verwachte gemiddelde gegeven de controlevariabelen per type (toegevoegde waarde modellen):

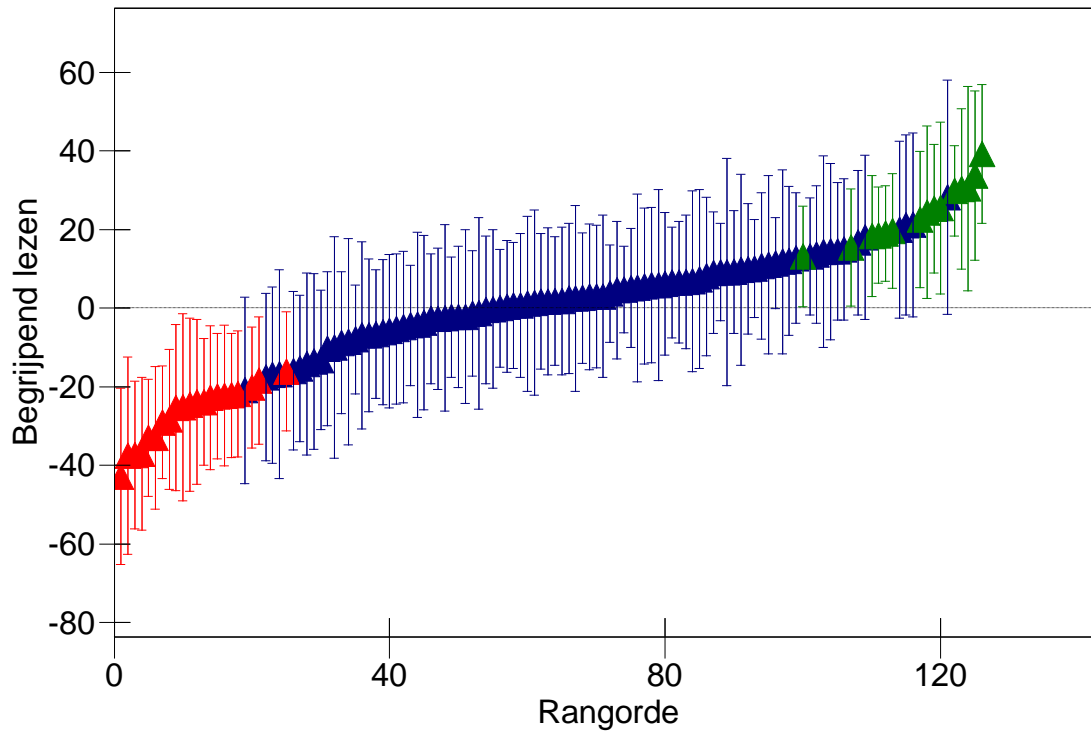
- Wanneer het betrouwbaarheidsinterval van een school volledig boven de nullijn valt, scoort de school in kwestie significant hoger op PIRLS Repeat dan de gemiddelde Vlaamse school (nulmodel) of dan verwacht gegeven de opgenomen controlevariabelen (toegevoegde waarde modellen). Die scholen worden in het **groen** aangeduid.
- Wanneer het betrouwbaarheidsinterval van een school de nullijn omvat, is de score van die school niet significant verschillend van het algemeen Vlaams gemiddelde (nulmodel) of dan verwacht gegeven de opgenomen controlevariabelen (toegevoegde waarde modellen). Die scholen worden in het **blauw** aangeduid.
- Wanneer het betrouwbaarheidsinterval van een school volledig onder de nullijn valt, scoort die specifieke school significant lager op PIRLS Repeat dan de gemiddelde Vlaamse school (nulmodel) of dan verwacht gegeven de opgenomen controlevariabelen (toegevoegde waarde modellen). Die scholen worden in het **rood** aangeduid.

3.2.1 Type 0-schooleffect

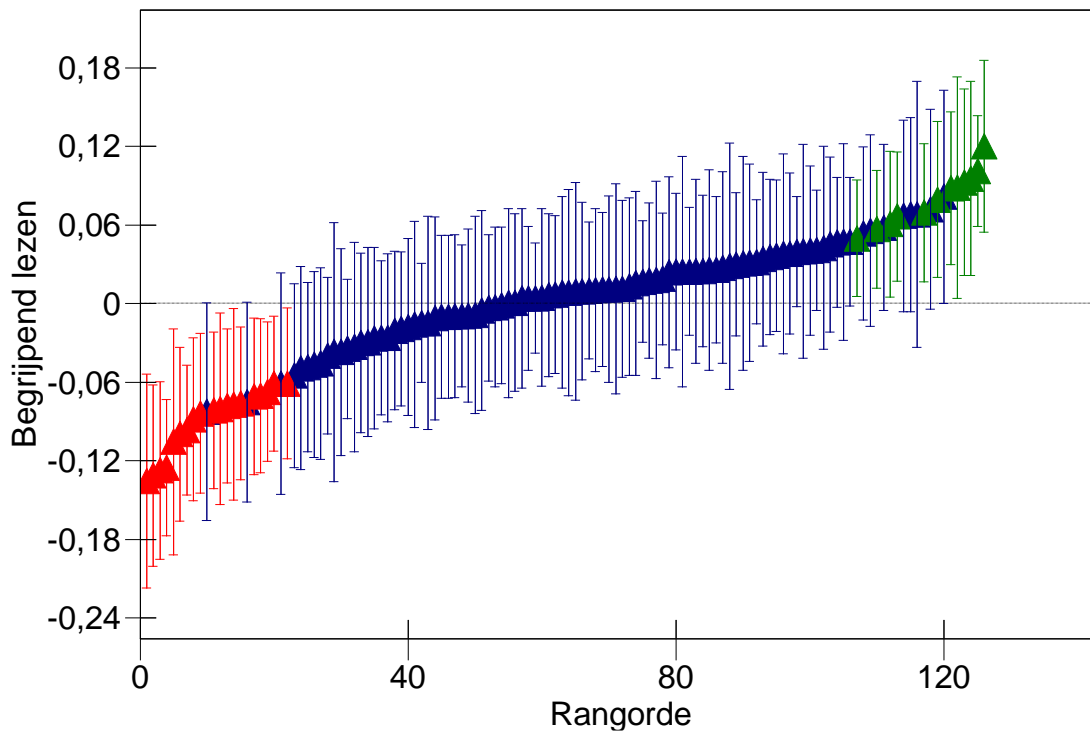
Figuur 1 en Figuur 2 geven de rupsgrafieken voor respectievelijk de PIRLS- en peilingschaal weer. In die grafieken worden de type 0-schooleffecten van alle scholen die deelgenomen hebben aan PIRLS Repeat voorgesteld. De type 0-schooleffecten (het verschil tussen het ▲-teken en de nullijn) geven weer hoe goed de leerlingen van een bepaalde school presteren in vergelijking met andere scholen zonder te controleren voor de achtergrondkenmerken van leerlingen of scholen.

In de rupsgrafieken worden de deelnemende scholen vergeleken met het algemeen Vlaams gemiddelde aan de hand van de betrouwbaarheidsintervallen. In Figuur 1 en Figuur 2 wordt die vergelijking met de volgende kleuren aangeduid:

- De **groene** scholen scoren significant hoger dan de gemiddelde Vlaamse school.
- De **blauwe** scholen scoren niet significant verschillend van het algemeen Vlaams gemiddelde.
- De **rode** scholen scoren significant lager dan de gemiddelde Vlaamse school.



Figuur 1. Rangschikking van scholen volgens het type 0-schooleffect (met 95%-betrouwbaarheidsinterval) op de PIRLS-schaal



Figuur 2. Rangschikking van scholen volgens het type 0-schooleffect (met 95%-betrouwbaarheidsinterval) op de peilingschaal

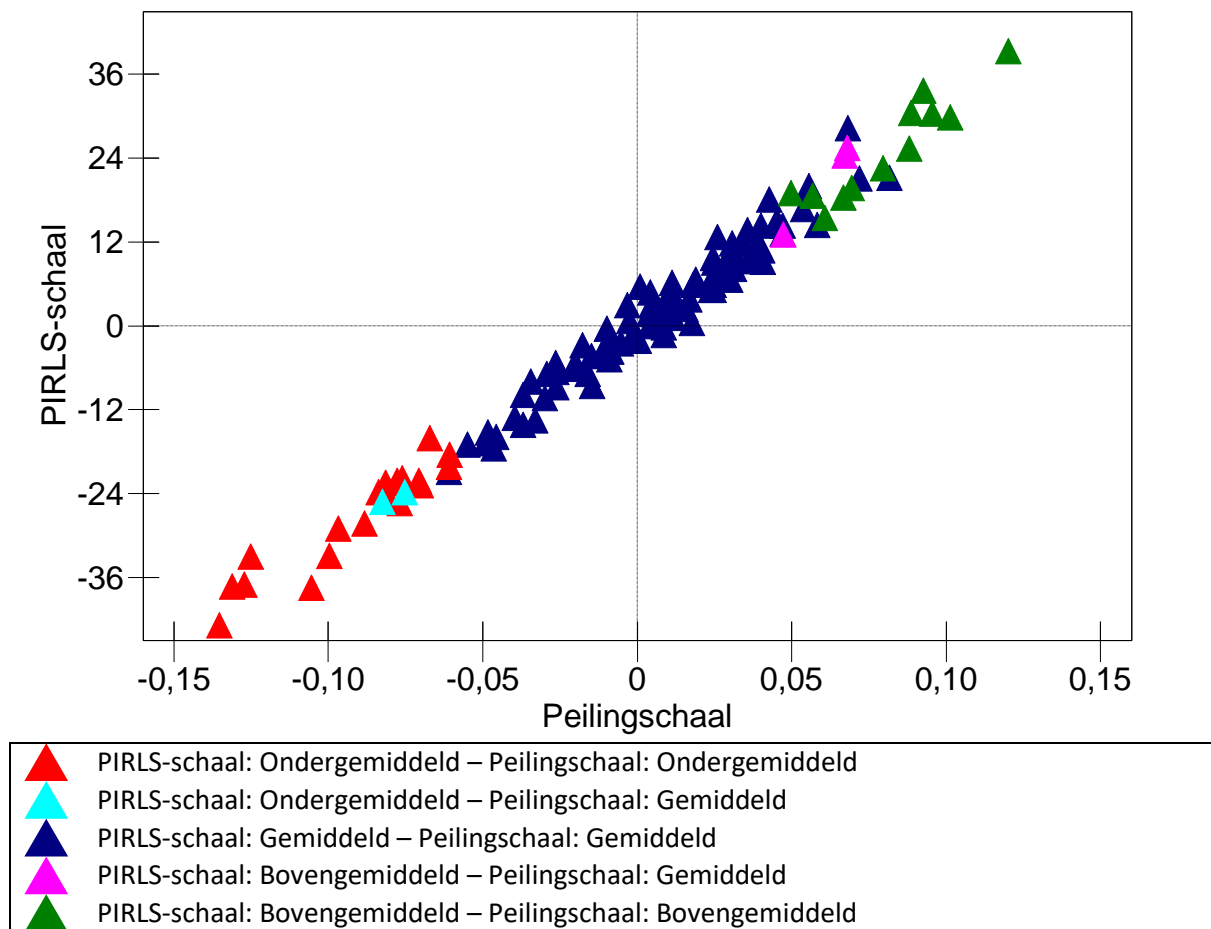
Op basis van Figuur 1 en Figuur 2 is het niet mogelijk om af te leiden in welke mate de resultaten uit de rupsgrafiek van de PIRLS-schaal en de peilingschaal overeenkomen voor de deelnemende scholen. Daarom hebben we de type 0-schooleffecten op beide schalen voor elke school vergeleken. De resultaten van die vergelijking worden getoond in Tabel 7 en visueel voorgesteld in Figuur 3. Uit Tabel 7 blijkt dat bij 96,03% van de scholen het resultaat van hun type 0-schooleffect hetzelfde is op de PIRLS- en peilingschaal. Die scholen worden in Figuur 3 aangeduid in het **rood**, **donkerblauw** en **groen**. Daarnaast toont Tabel 7 dat 3,97% van de scholen een verschillend resultaat heeft op beide schalen. Bij 1,59% van de scholen wordt hun type 0-schooleffect op de peilingschaal geclassificeerd als gemiddeld, terwijl hun schooleffect op de PIRLS-schaal ondergemiddeld is. Die scholen worden in Figuur 3 aangeduid in het **lichtblauw**. Bij 2,38% van de scholen wordt op de peilingschaal hun type 0-schooleffect geclassificeerd als gemiddeld, maar op de PIRLS-schaal als bovengemiddeld is. Dat zijn de **roze** scholen in Figuur 3. Op basis van de eerder besproken resultaten kunnen we de concordantie tussen de type 0-schooleffecten op de PIRLS- en peilingschaal berekenen. Die concordantie bedraagt 90,66% wanneer we rekening houden met de kans om per toeval hetzelfde resultaat te behalen op de PIRLS- en peilingschaal.

Tabel 7

Classificatie van scholen volgens hun type 0-schooleffect op de PIRLS- en peilingschaal

	Peilingschaal			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PIRLS-schaal				
Ondergemiddeld	19 (15,08%)	2 (1,59%)	-	21 (16,67%)
Gemiddeld	-	90 (71,43%)	-	90 (71,43%)
Bovengemiddeld	-	3 (2,38%)	12 (9,52%)	15 (11,90%)
Totaal	19 (15,08%)	95 (75,40%)	12 (9,52%)	126 (100,00%)

Noot. Coëfficiënt kappa²² = 90,66%



Figuur 3. Puntgrafiek van scholen volgens hun type 0-schooleffect op de PIRLS- en peilingschaal

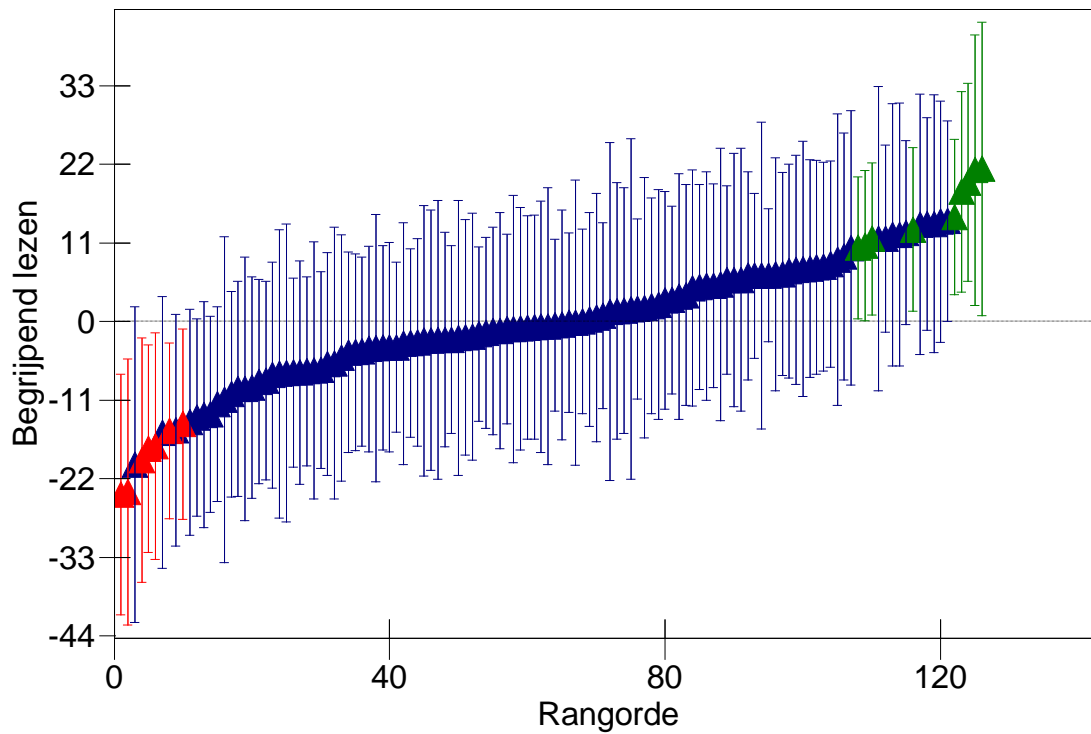
3.2.2 Type AA-schooleffect

In Figuur 4 en Figuur 5 worden de type AA-schooleffecten van alle scholen die deelgenomen hebben aan PIRLS Repeat voorgesteld aan de hand van rupsgrafieken. Die rupsgrafieken worden gelezen zoals beschreven bij onderdeel '3.2 Schooleffecten volgens de PIRLS- en peilingschaal'. Bij onderstaande figuren gaat het om het type AA-schooleffect. De schooleffecten (het verschil tussen het ▲-teken en de nullijn) in Figuur 4 en Figuur 5 tonen dus hoe goed een school het doet in vergelijking met scholen die leerlingen hebben met gelijkaardige eerdere leesprestaties.

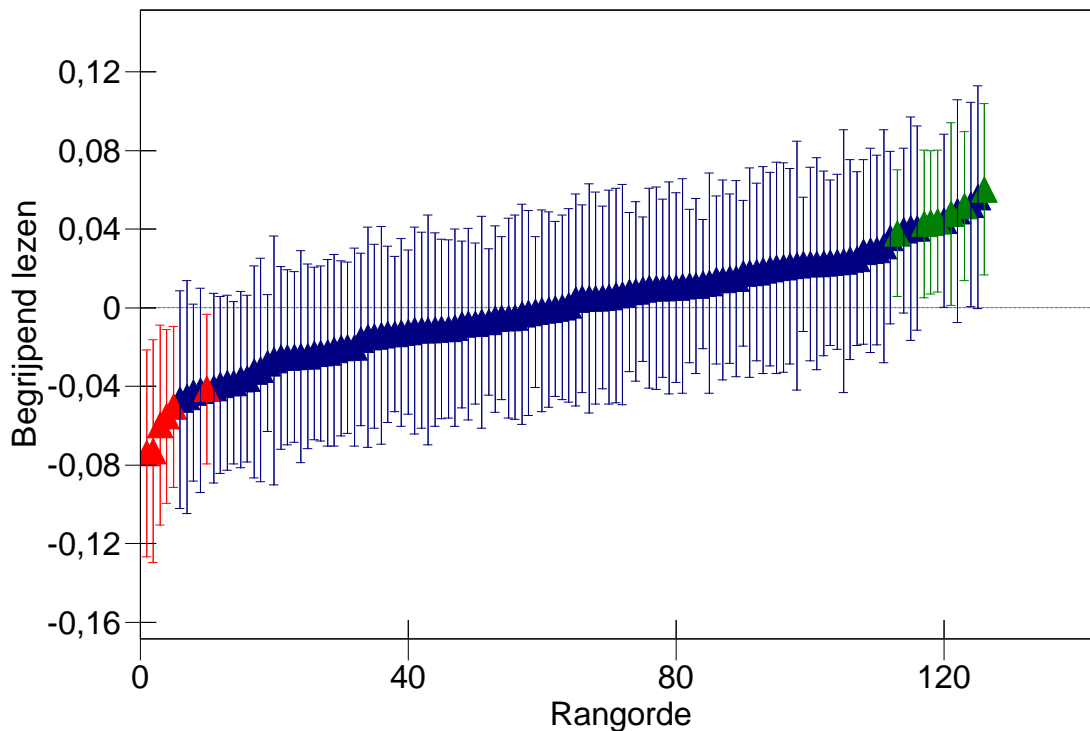
²² Door de concordantie te berekenen aan de hand van coëfficiënt kappa houden we rekening met de kans dat scholen per toeval op zowel de PIRLS- als peilingschaal hetzelfde resultaat behalen (McHugh, 2012).

In Figuur 4 en Figuur 5 worden de schooleffecten van de deelnemende scholen aan de hand van de betrouwbaarheidsintervallen vergeleken:

- De **groene** scholen scoren significant hoger dan we verwachten op basis van de eerdere leesprestaties van de leerlingen.
- De **blauwe** scholen scoren niet significant verschillend van wat we verwachten op basis van de eerdere leesprestaties van de leerlingen.
- De **rode** scholen scoren significant lager dan we verwachten op basis van de eerdere leesprestaties van de leerlingen.



Figuur 4. Rangschikking van scholen volgens het type AA-schooleffect (met 95%-betrouwbaarheidsinterval) op de PIRLS-schaal



Figuur 5. Rangschatting van scholen volgens het type AA-schooleffect (met 95%-betrouwbaarheidsinterval) op de peilingschaal

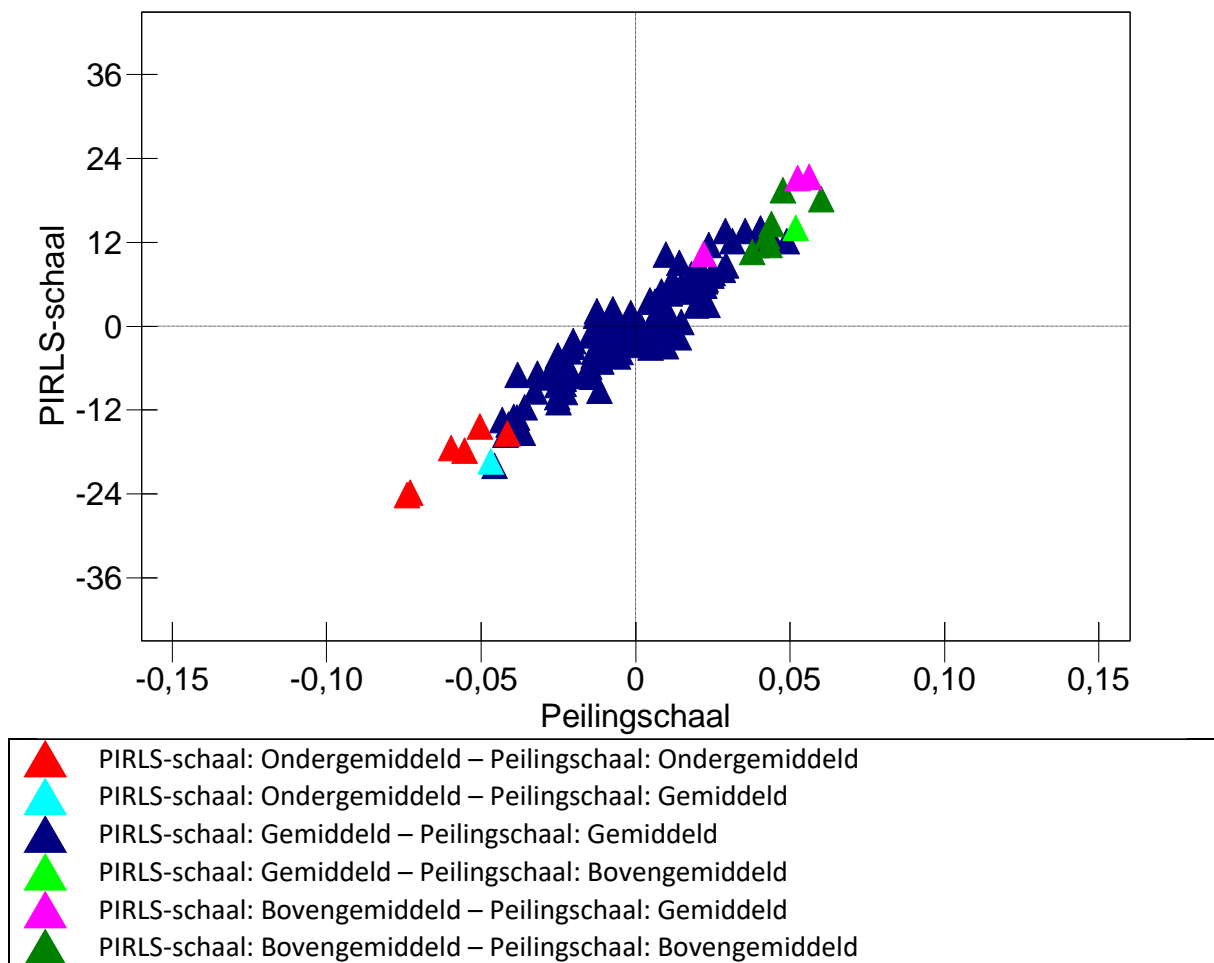
Wanneer we de rupsgrafiek van de PIRLS-schaal vergelijken met die van de peilingschaal kunnen we de concordantie tussen de PIRLS- en peilingschaal nagaan. Die vergelijking resulteert in Tabel 8 en wordt visueel voorgesteld in Figuur 6. Daaruit blijkt dat de PIRLS- en peilingschaal tot hetzelfde resultaat leiden voor het type AA-schooleffect bij 96,03% van de scholen. Dat zijn de **rode**, **donkerblauwe** en **donkergroene** scholen in Figuur 6. Verder geeft Tabel 8 aan dat er het resultaat verschilt tussen beide schalen bij 3,96% van de scholen. Concreet heeft de **lichtblauwe** school (0,79%) een gemiddeld type AA-schooleffect op de peilingschaal, terwijl het schooleffect op de PIRLS-schaal ondergemiddeld is. Bij de **roze** scholen (2,38%) wordt hun type AA-schooleffect op de peilingschaal geclassificeerd als gemiddeld en op de PIRLS-schaal als bovengemiddeld. Omgekeerd, heeft de **lichtgroene** school (0,79%) een gemiddeld schooleffect op de peilingschaal en een bovengemiddeld schooleffect op de PIRLS-schaal. In het algemeen bedraagt de concordantie tussen de type AA-schooleffecten op de PIRLS- en peilingschaal 81,14% wanneer we rekening houden met de kans om per toeval hetzelfde resultaat te behalen op de PIRLS- en peilingschaal.

Tabel 8

Classificatie van scholen volgens hun type AA-schooleffect op de PIRLS- en peilingschaal

	Peilingschaal			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PIRLS-schaal				
Ondergemiddeld	6 (4,76%)	1 (0,79%)	-	7 (5,56%)
Gemiddeld	-	109 (86,51%)	1 (0,79%)	110 (87,30%)
Bovengemiddeld	-	3 (2,38%)	6 (4,76%)	9 (7,14%)
Totaal	6 (4,76%)	113 (89,68%)	7 (5,56%)	126 (100,00%)

Noot. Coëfficiënt kappa = 81,14%



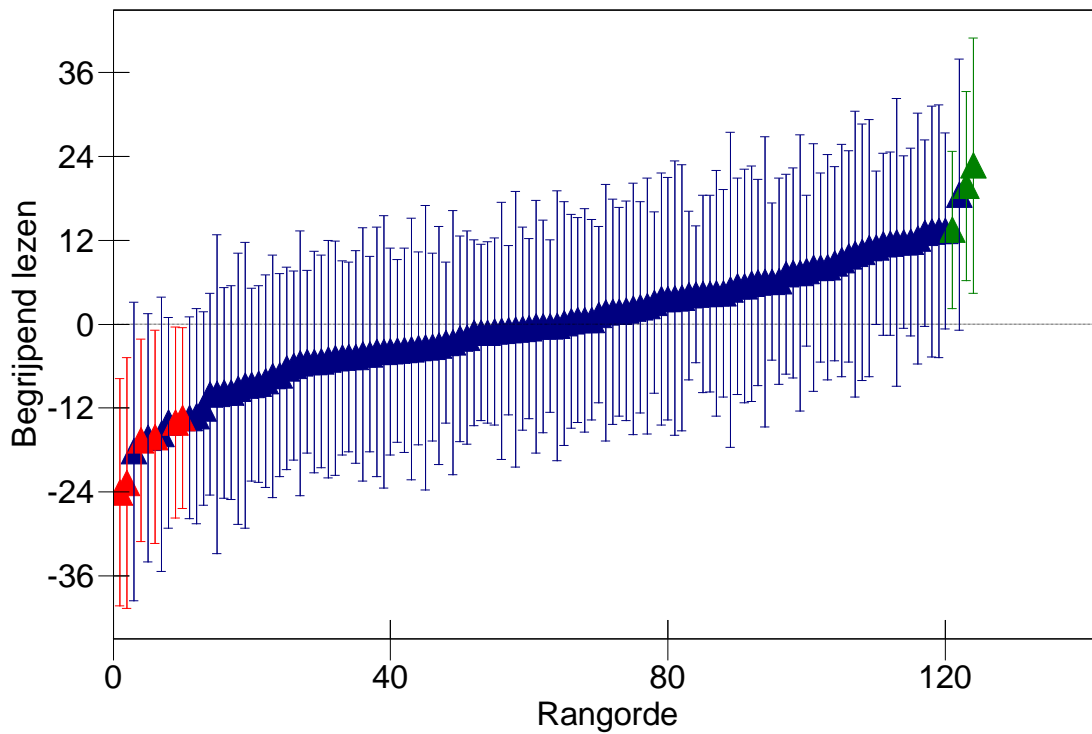
Figuur 6. Puntgrafiek van scholen volgens hun type AA-schooleffect op de PIRLS- en peilingschaal

3.2.3 Type A-schooleffect

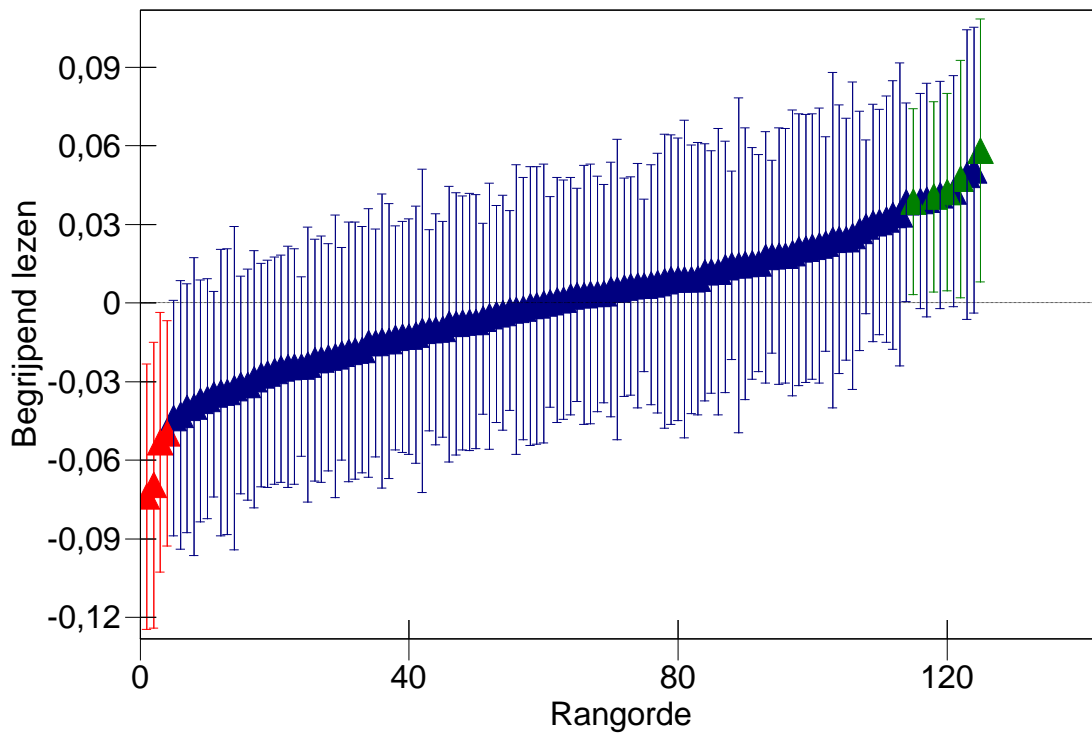
In Figuur 7 en Figuur 8 worden de type A-schooleffecten van alle scholen die deelgenomen hebben aan PIRLS Repeat voorgesteld aan de hand van rupsgrafieken. Die rupsgrafieken worden gelezen zoals eerder vermeld (zie 3.2 Schooleffecten volgens de PIRLS- en peilingschaal). Bij onderstaande figuren gaat het om het type A-schooleffect. Onderstaande schooleffecten (het verschil tussen het ▲-teken en de nullijn in Figuur 7 en Figuur 8) geven dus aan hoe goed de leerlingen in een bepaalde school scoren in vergelijking met gelijkaardige leerlingen (in termen van eerdere prestaties en achtergrondkenmerken) in de andere scholen.

In Figuur 7 en Figuur 8 worden de type A-schooleffecten van deelnemende scholen vergeleken aan de hand van betrouwbaarheidsintervallen:

- De **groene** scholen scoren significant hoger dan we verwachten op basis van de eerdere prestaties en leerlingkenmerken.
- De **blauwe** scholen scoren niet significant verschillend van wat we verwachten op basis van de eerdere prestaties en leerlingkenmerken.
- De **rode** scholen scoren significant lager dan we verwachten op basis van de eerdere prestaties en leerlingkenmerken.



Figuur 7. Rangschrifing van scholen volgens het type A-schooleffect (met 95%-betrouwbaarheidsinterval) op de PIRLS-schaal



Figuur 8. Rangschrifing van scholen volgens het type A-schooleffect (met 95%-betrouwbaarheidsinterval) op de peilingschaal

Wanneer we de rupsgrafiek van de PIRLS-schaal vergelijken met die van de peilingschaal kunnen we de concordantie tussen de PIRLS- en peilingschaal nagaan voor de type A-schooleffecten van alle scholen. Tabel 9 toont de numerieke resultaten en Figuur 9 stelt dezelfde resultaten visueel voor. Uit Tabel 9 blijkt dat 96,77% van de scholen hetzelfde resultaat heeft voor zijn type A-schooleffect op de PIRLS- en peilingschaal. Die scholen worden in Figuur 9 aangeduid in het **rood**, **donkerblauw** en

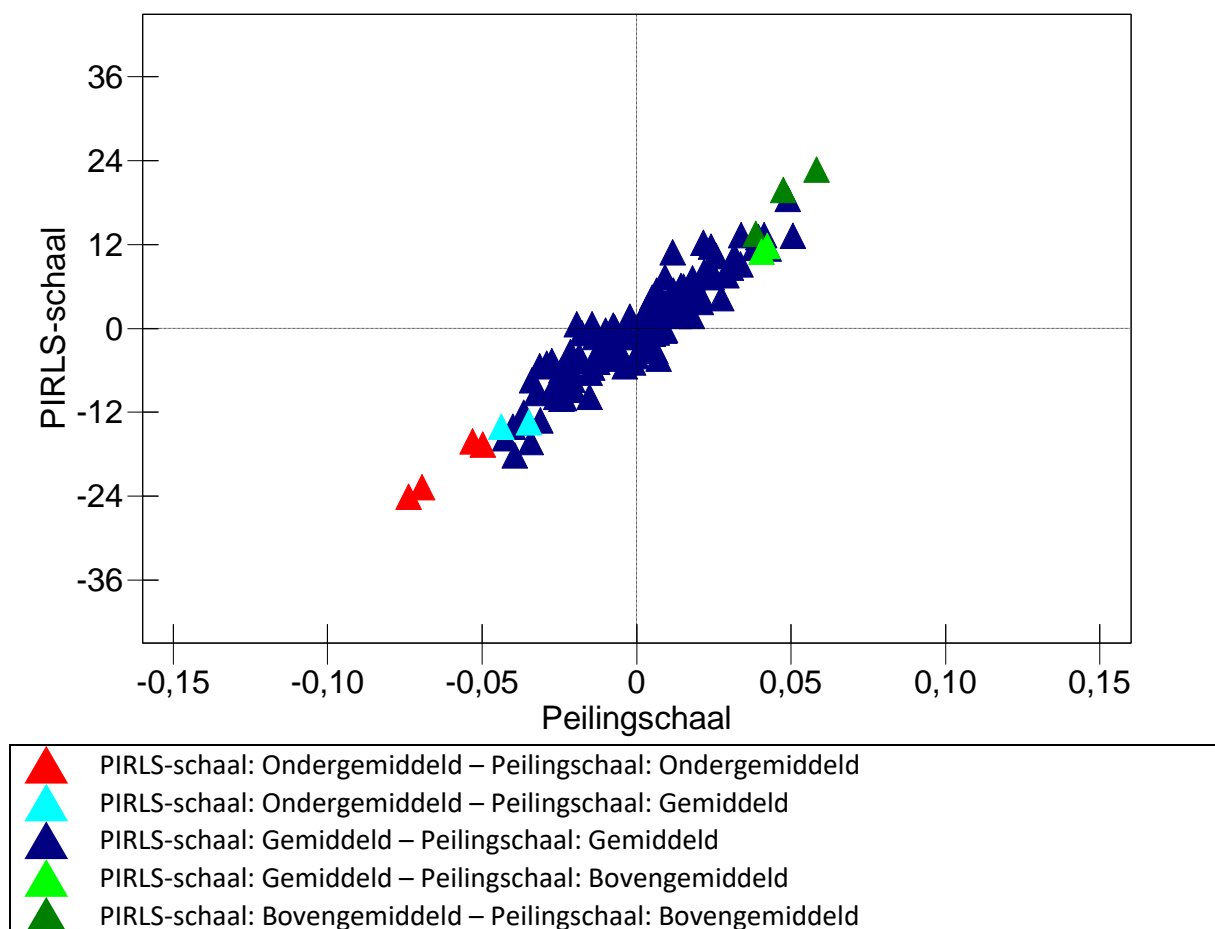
donkergroen. Bij de overige 3,23% van de scholen verschilt het resultaat tussen de PIRLS- en peilingschaal. Bij 1,61% van de scholen wordt hun type A-schooleffect op de peilingschaal geclassificeerd als gemiddeld, terwijl hun schooleffect op de PIRLS-schaal ondergemiddeld is. Dat zijn de **lichtblauwe** scholen. Bij de **lichtgroene** scholen (1,61%) wordt op de peilingschaal hun schooleffect ingedeeld bij bovengemiddeld, maar op de PIRLS-schaal bij gemiddeld. In het algemeen bedraagt de concordantie tussen de type A-schooleffecten op de PIRLS- en peilingschaal 76,52% wanneer er rekening gehouden wordt met de toevalskans om hetzelfde resultaat te behalen op de PIRLS- en peilingschaal.

Tabel 9

Classificatie van scholen volgens hun type A-schooleffect op de PIRLS- en peilingschaal

	Peilingschaal			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PIRLS-schaal				
Ondergemiddeld	4 (3,22%)	2 (1,61%)	-	6 (4,84%)
Gemiddeld	-	113 (91,13%)	2 (1,61%)	115 (92,74%)
Bovengemiddeld	-	-	3 (2,42%)	3 (2,42%)
Totaal	4 (3,22%)	115 (92,74%)	5 (4,03%)	124 (100,00%)

Noot. Coëfficiënt kappa = 76,52%



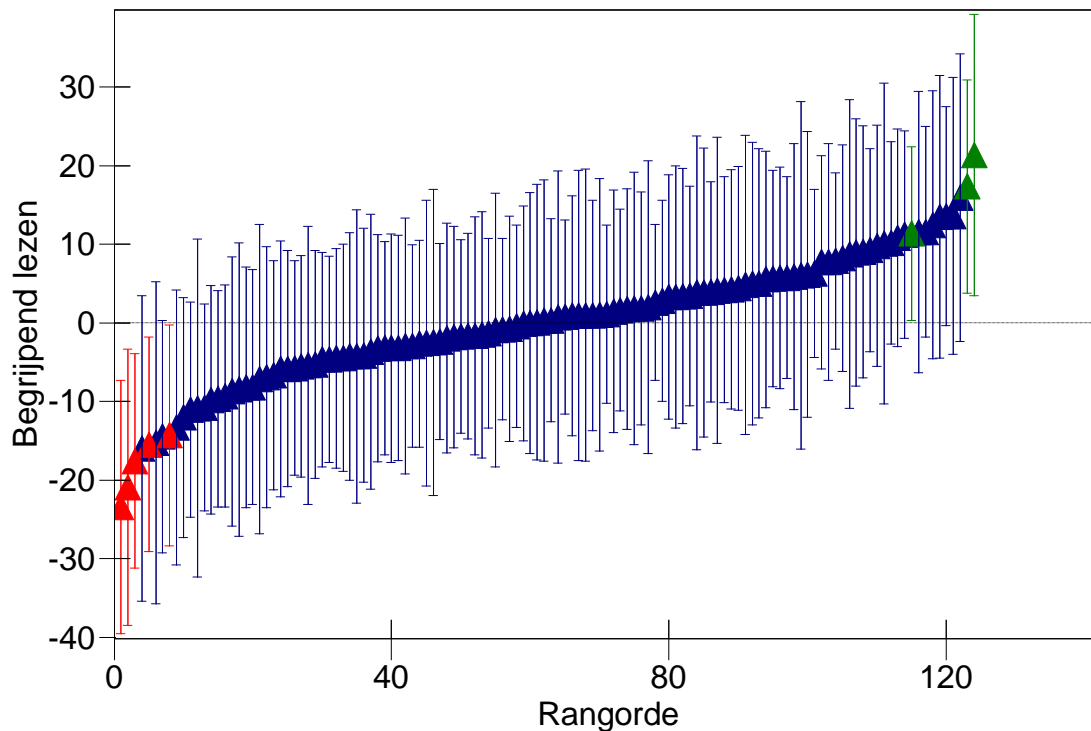
Figuur 9. Puntgrafiek van scholen volgens hun type A-schooleffect op de PIRLS- en peilingschaal

3.2.4 Type B-schooleffect

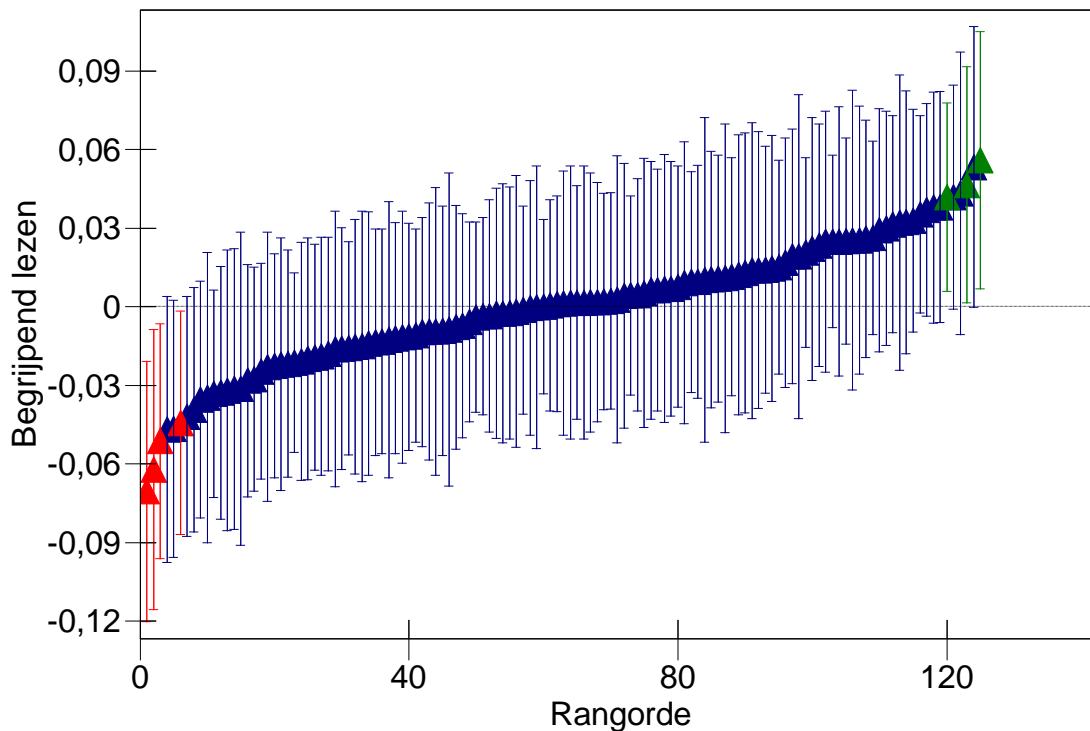
In Figuur 10 en Figuur 11 worden de type B-schooleffecten van alle scholen die deelgenomen hebben aan PIRLS Repeat voorgesteld aan de hand van rupsgrafieken. Die rupsgrafieken worden gelezen zoals eerder omschreven (zie 3.2 Schooleffecten volgens de PIRLS- en peilingschaal). De schooleffecten (het verschil tussen het ▲-teken en de nullijn) in Figuur 10 en Figuur 11 tonen hoe goed de leerlingen in een bepaalde school scoren in vergelijking met scholen die gelijkaardige leerlingen en een gelijkaardige schoolcompositie hebben.

In Figuur 10 en Figuur 11 worden de type B-schooleffecten van deelnemende scholen vergeleken aan de hand van betrouwbaarheidsintervallen:

- De **groene** scholen scoren significant hoger dan verwacht op basis van de eerdere prestaties, leerlingkenmerken en schoolcompositiekenmerken.
- De **blauwe** scholen scoren niet significant verschillend van wat we verwachten op basis van de eerdere prestaties, leerlingkenmerken en schoolcompositiekenmerken.
- De **rode** scholen scoren significant lager dan verwacht op basis van de eerdere prestaties, leerlingkenmerken en schoolcompositiekenmerken.



Figuur 10. Rangschikking van scholen volgens het type B-schooleffect (met 95%-betrouwbaarheidsinterval) op de PIRLS-schaal



Figuur 11. Rangschikking van scholen volgens het type B-schooleffect (met 95%-betrouwbaarheidsinterval) op de peilingschaal

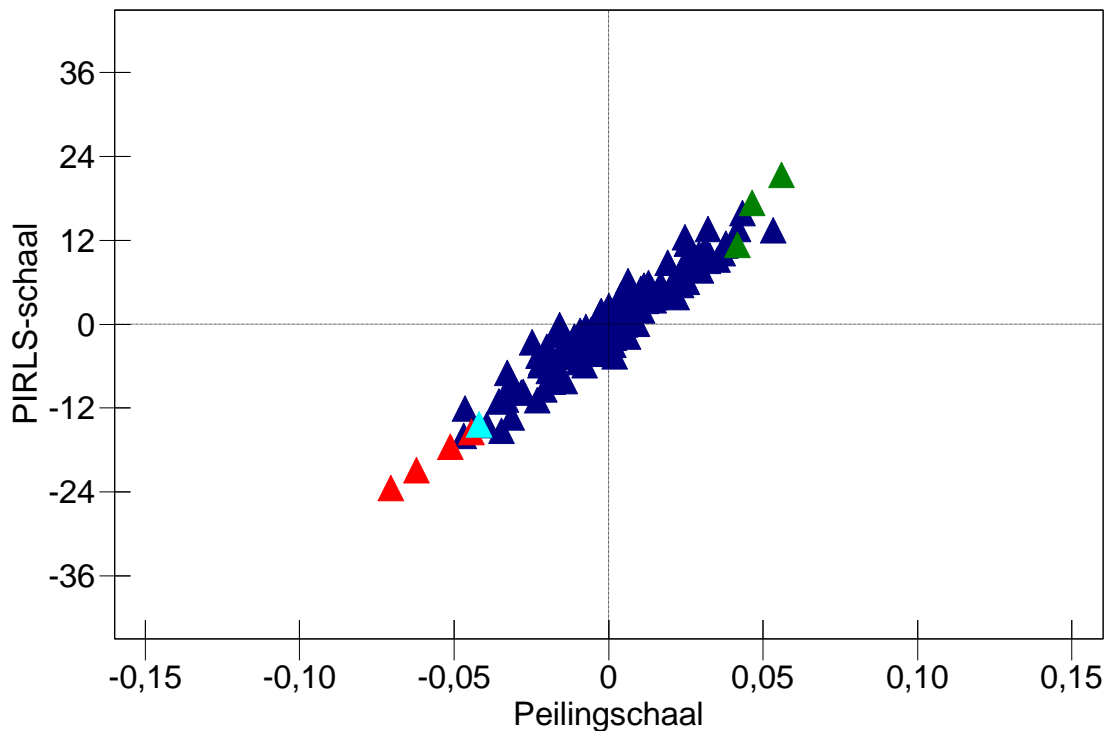
De vergelijking van de bovenstaande rupsgrafieken wordt weergegeven in Tabel 10 en Figuur 12. Voor 99,19% van de deelnemende scholen zijn de resultaten van het type B-schooleffect op de PIRLS- en peilingschaal hetzelfde. Die scholen worden in Figuur 12 aangeduid in een **rode**, **donkerblauwe** en **groene** kleur. Dat is niet het geval bij de **lichtblauwe** school (0,81%). Die school heeft een gemiddeld schooleffect op de peilingschaal en een ondergemiddeld schooleffect op de PIRLS-schaal. De concordantie tussen de type B-schooleffecten op beide schalen bedraagt 93,01% wanneer we rekening houden met de kans om per toeval hetzelfde resultaat te behalen op de PIRLS- en peilingschaal.

Tabel 10

Classificatie van scholen volgens hun type B-schooleffect op de PIRLS- en peilingschaal

	Peilingschaal			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PIRLS-schaal				
Ondergemiddeld	4 (3,22%)	1 (0,81%)	-	5 (4,03%)
Gemiddeld	-	116 (93,55%)	-	116 (93,55%)
Bovengemiddeld	-	-	3 (2,42%)	3 (2,42%)
Totaal	4 (3,22%)	117 (94,35%)	3 (2,42%)	124 (100,00%)

Noot. Coëfficiënt kappa = 93,01%



▲	PIRLS-schaal: Ondergemiddeld – Peilingschaal: Ondergemiddeld
▲	PIRLS-schaal: Ondergemiddeld – Peilingschaal: Gemiddeld
▲	PIRLS-schaal: Gemiddeld – Peilingschaal: Gemiddeld
▲	PIRLS-schaal: Bovengemiddeld – Peilingschaal: Bovengemiddeld

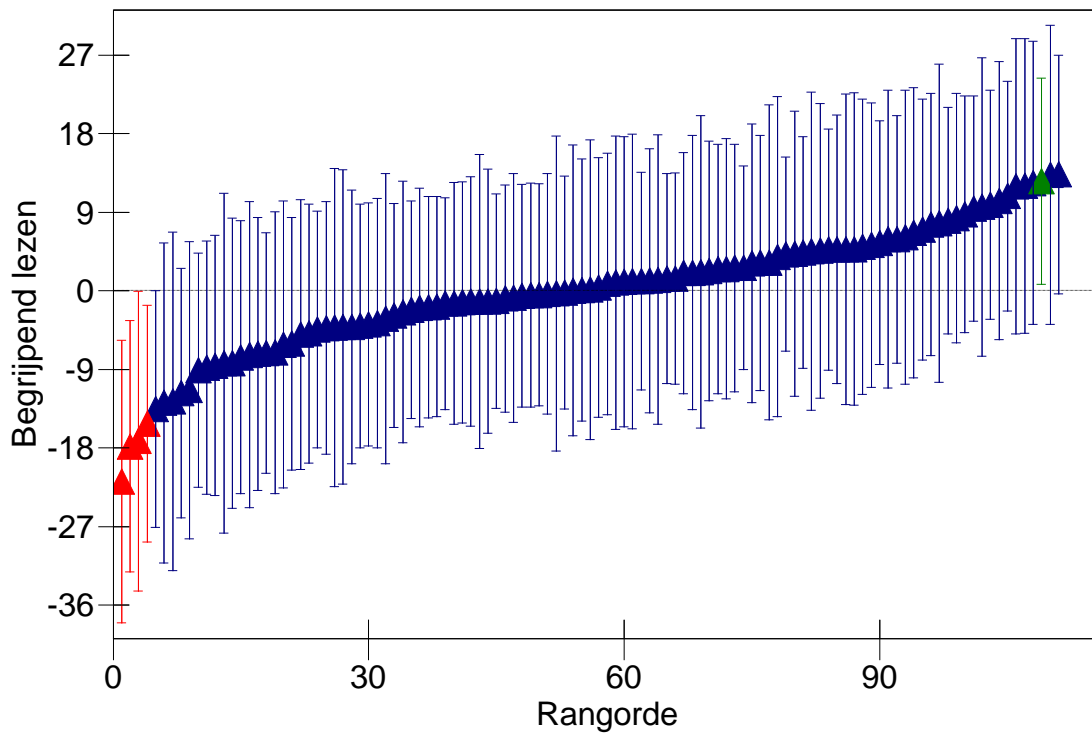
Figuur 12. Puntgrafiek van scholen volgens hun type B-schooleffect op de PIRLS- en peilingschaal

3.2.5 Type X-schooleffect

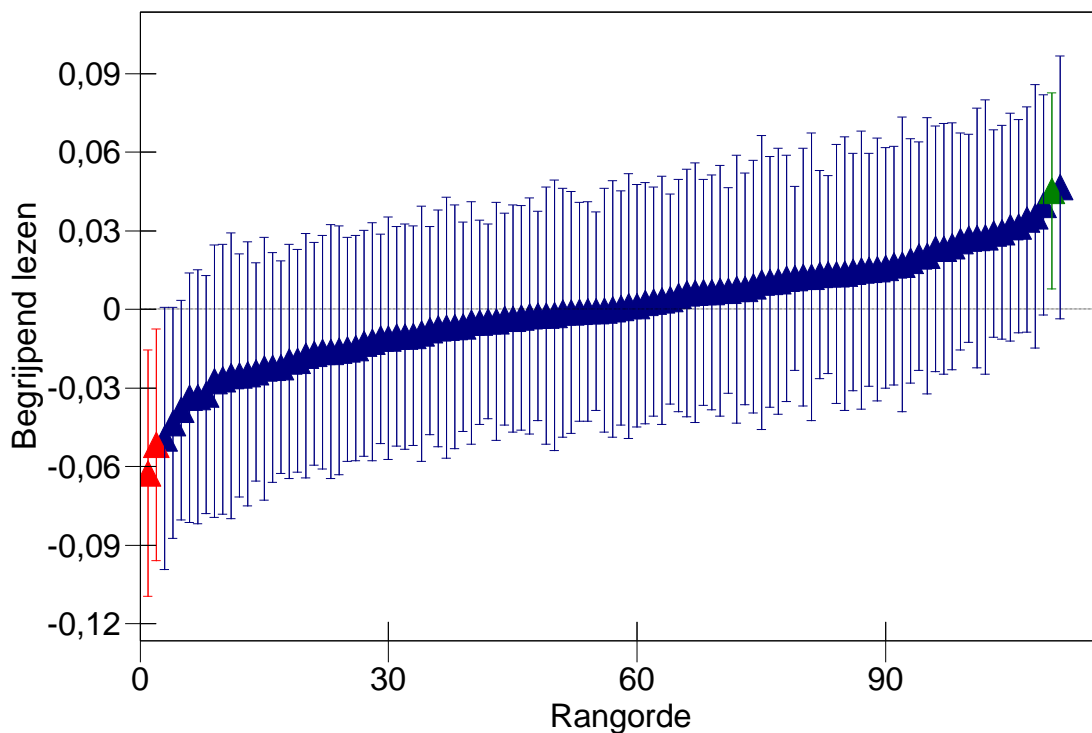
In Figuur 13 en Figuur 14 worden de type X-schooleffecten van alle scholen die deelgenomen hebben aan PIRLS Repeat voorgesteld aan de hand van rupsgrafieken. Die rupsgrafieken worden gelezen zoals eerder besproken (zie 3.2 Schooleffecten volgens de PIRLS- en peilingschaal). Het type X-schooleffect (het verschil tussen het ▲-teken en de nullijn) geeft weer hoe goed de leerlingen in een bepaalde school scoren in vergelijking met gelijkaardige leerlingen in scholen met een gelijkaardige schoolcompositie en gelijkaardige niet-beïnvloedbare schoolkenmerken.

In Figuur 13 en Figuur 14 worden de type B-schooleffecten van de deelnemende scholen vergeleken aan de hand van betrouwbaarheidsintervallen:

- De **groene** scholen scoren significant hoger dan verwacht op basis van de eerdere prestaties, leerlingkenmerken, schoolcompositiekenmerken en niet-beïnvloedbare schoolkenmerken.
- De **blauwe** scholen scoren niet significant verschillend van wat we verwachten op basis van de eerdere prestaties, leerlingkenmerken, schoolcompositiekenmerken en niet-beïnvloedbare schoolkenmerken.
- De **rode** scholen scoren significant lager dan verwacht op basis van de eerdere prestaties, leerlingkenmerken, schoolcompositiekenmerken en niet-beïnvloedbare schoolkenmerken.



Figuur 13. Rangschikking van scholen volgens het type X-schooleffect (met 95%-betrouwbaarheidsinterval) op de PIRLS-schaal



Figuur 14. Rangschikking van scholen volgens het type X-schooleffect (met 95%-betrouwbaarheidsinterval) op de peilingschaal

Ook voor het type X-schooleffect kunnen we de concordantie tussen de PIRLS- en peilingschaal nagaan door bovenstaande rupsgrafieken te vergelijken. Tabel 11 toont de kwantitatieve resultaten en Figuur 15 stelt dezelfde resultaten visueel voor. De PIRLS- en peilingschaal leiden tot hetzelfde resultaat voor het type X-schooleffect bij 98,21% van de scholen. Die scholen worden in Figuur 15 aangeduid in het

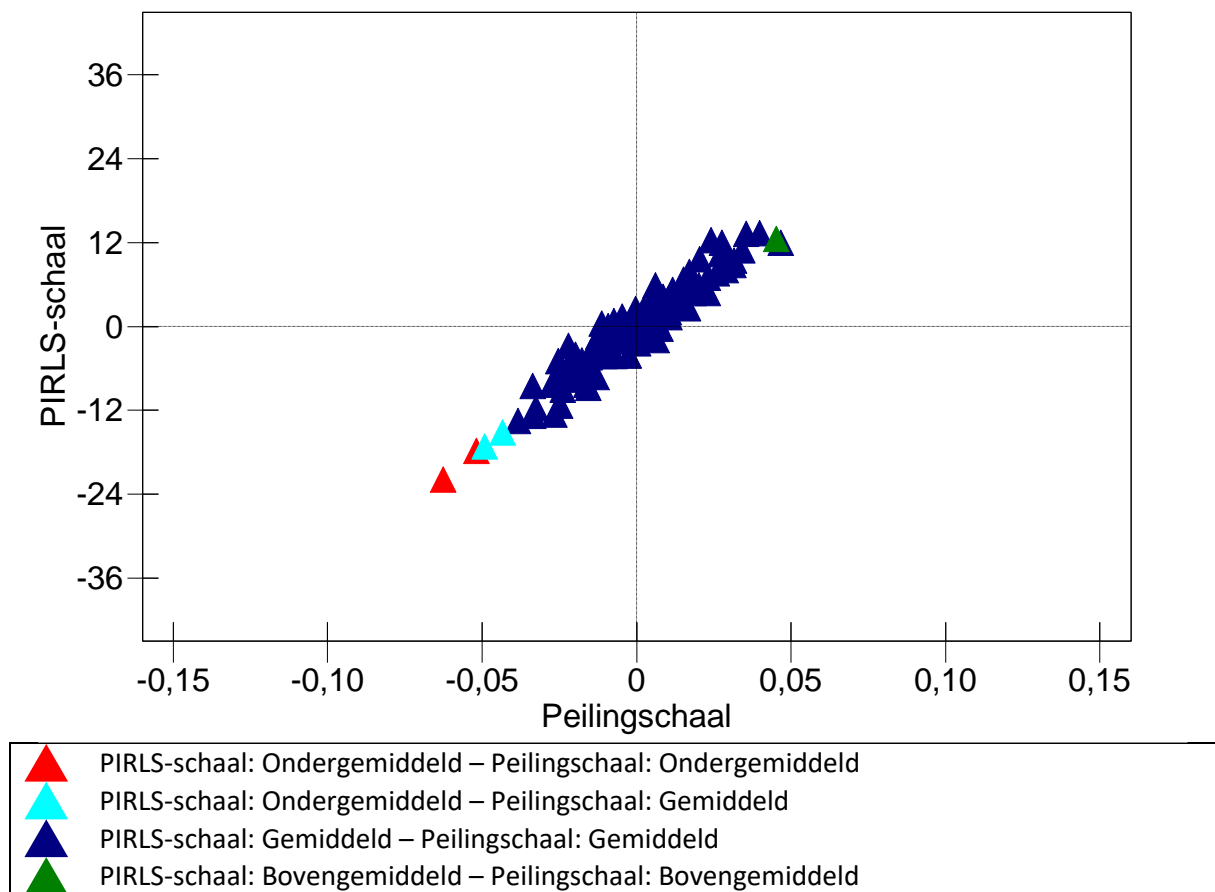
rood, donkerblauw en groen. Bij de overige 1,79% van de scholen is het resultaat verschillend op de PIRLS- en peilingschaal. Bij die scholen wordt hun schooleffect op de peilingschaal ingedeeld bij gemiddeld, maar op de PIRLS-schaal bij ondergemiddeld. In het algemeen bedraagt de concordantie tussen de type X-schooleffecten op de PIRLS- en peilingschaal 74,31% wanneer er rekening gehouden wordt met de toevalskans om hetzelfde resultaat te behalen op de PIRLS- en peilingschaal.

Tabel 11

Classificatie van scholen volgens hun type X-schooleffect op de PIRLS- en peilingschaal

	Peilingschaal			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PIRLS-schaal				
Ondergemiddeld	2 (1,79%)	2 (1,79%)	-	4 (3,57%)
Gemiddeld	-	107 (95,54%)	-	107 (95,54%)
Bovengemiddeld	-	-	1 (0,89%)	1 (0,89%)
Totaal	2 (1,79%)	109 (97,32%)	1 (0,89%)	112 (100,00%)

Noot. Coëfficiënt kappa = 74,31%



Figuur 15. Puntgrafiek van scholen volgens hun type X-schooleffect op de PIRLS- en peilingschaal

3.2.6 Vergelijking over de verschillende types schooleffecten heen: hoofdbevindingen

De correlatie tussen de schooleffecten op de PIRLS- en peilingschaal wordt voor elk van de vijf modellen weergegeven in Tabel 12. Daaruit blijkt dat de correlatie hoog is bij alle types toegevoegde

waarde modellen.²³ Ook de concordantie (berekend op basis van coëfficiënt kappa) tussen de PIRLS- en peilingschaal is bij alle modellen relatief hoog. Bij de type B-modellen vinden we de grootste concordantie: 93,01%.

Tabel 12

Correlatie tussen de residuen op de PIRLS-schaal en op de peilingschaal en de concordantie tussen die schalen

	Type 0	Type AA	Type A	Type B	Type X
Correlatie	0,9897	0,9590	0,9542	0,9650	0,9631
Coëfficiënt kappa	0,9066	0,8114	0,7652	0,9301	0,7431

Bij de overgrote meerderheid van de scholen leveren de PIRLS- en peilingschaal hetzelfde resultaat op. Toch zijn er bij elk van de vijf types modellen scholen waarbij het schooleffect op de PIRLS-schaal verschilt van dat op de peilingschaal. Dat zijn scholen waarvan het schooleffect net wel significant afwijkt ten opzichte van het gemiddelde op de ene schaal en net niet significant afwijkt op de andere schaal. Hoewel het om eerder kleine verschillen – net wel of niet significant – gaat, kunnen die verschillen wel een grote impact hebben voor de scholen in kwestie. Dat zou mogelijk het geval kunnen zijn wanneer er gevolgen (bijvoorbeeld een begeleidingstraject of bijkomende doorlichting) gekoppeld worden aan het resultaat van een school.

4. Resultaten PIRLS, peiling en doorlichtingen (Vergelijking 2)

4.1 Schooleffecten (PIRLS- en peilingschaal) en inschaling bij de doorlichtingen

Zoals eerder aangegeven, bouwen we in een tweede vergelijking verder op de resultaten van de eerste vergelijking. Van de 126 scholen uit de eerste vergelijking zijn er 25 scholen die deelnamen aan PIRLS Repeat én doorgelicht zijn met een focus op Nederlands. Concreet geven we voor die 25 scholen per type model hun schooleffecten op de PIRLS- en peilingschaal (uit Vergelijking 1) weer. Die schooleffecten vergelijken we vervolgens met de inschaling op de schaal ‘leereffecten’ voor het leergebied Nederlands uit het doorlichtingsverslag van die scholen.

4.1.1 Type 0-schooleffect

Figuur 16 toont de type 0-schooleffecten uit Vergelijking 1 (zie Figuur 3) voor de 25 scholen die doorgelicht zijn en deelnamen aan PIRLS Repeat zijn. Daarnaast toont die figuur de inschaling van de onderwijsinspectie bij de doorlichting aan de hand van volgende kleuren:

- De **paarse** scholen (28%) benaderen de verwachting voor de schaal ‘leereffecten’.
- De **blauwe** scholen (72%) scoren volgens de verwachting voor de schaal ‘leereffecten’.

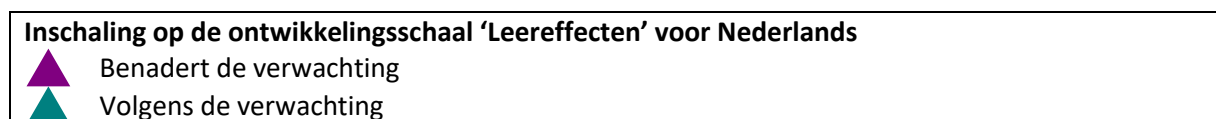
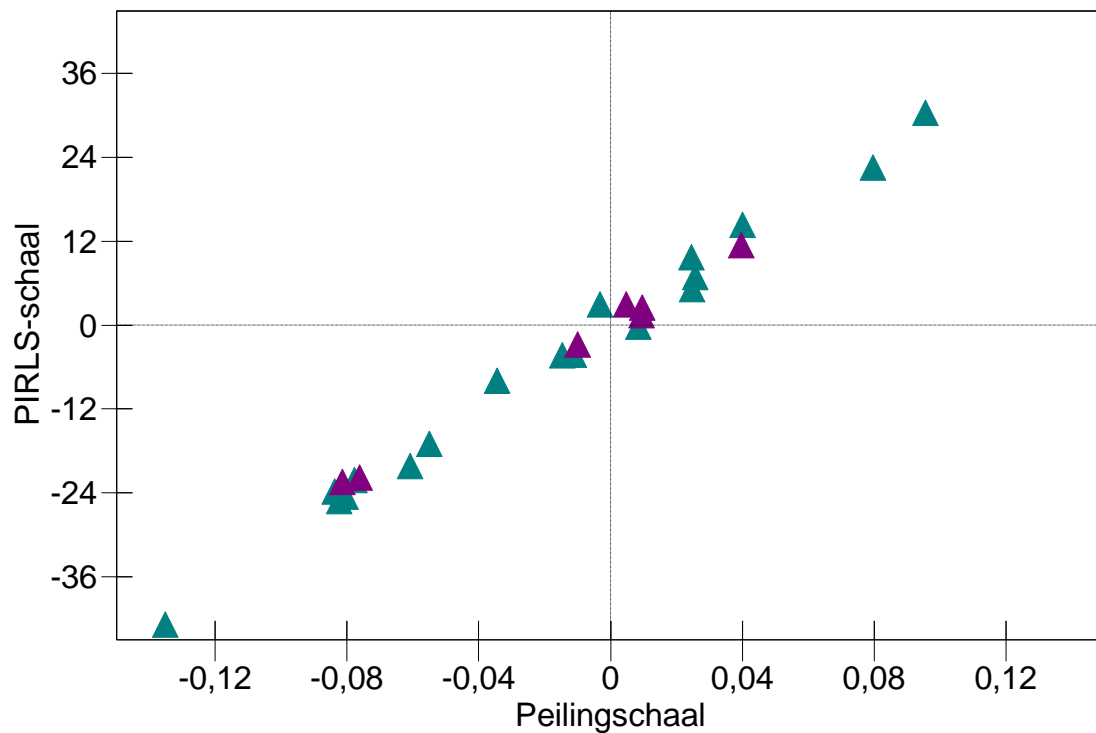
Dezelfde resultaten worden numeriek weergegeven in Tabel 13. De resultaten moeten wel met de nodige voorzichtigheid geïnterpreteerd worden omwille van het beperkte aantal scholen.

Wat betreft het type 0-schooleffect op de **PIRLS-schaal** tonen Figuur 16 en Tabel 13 dat 8% van de scholen een ondergemiddeld schooleffect heeft en volgens de onderwijsinspectie de verwachting benadert. Zo’n 20% van de scholen heeft een gemiddeld schooleffect en benadert de verwachting

²³ Hoewel het om dezelfde data gaat – weliswaar op twee verschillende schalen, is de correlatie niet perfect. Dat is in lijn met eerder onderzoek. Zelfs wanneer eenzelfde toets twee schooljaren na elkaar wordt afgenomen in hetzelfde leerjaar, wordt er geen perfecte correlatie gevonden tussen die twee afnames (Timmermans, 2012). Bij type 0-modellen is die correlatie tussen die twee afnames nog relatief hoog. Naarmate er meer controlevariabelen opgenomen worden, daalt de correlatie tussen beide afnames (Dumay et al., 2013).

volgens de onderwijsinspectie. Verder heeft 24% van de scholen een ondergemiddeld type 0-schooleffect, al scoren die scholen wel volgens de verwachting bij de doorlichting. Nog eens 40% van de scholen scoort gemiddeld op de PIRLS-schaal en volgens de verwachting bij de doorlichting. Tot slot zijn er enkele scholen (8%) met een bovengemiddeld schooleffect op de PIRLS-schaal en een inschaling van volgens de verwachting bij de doorlichting. Er is geen enkele school die 'beneden de verwachting' of 'overstijgt de verwachting' behaalt bij de doorlichtingen.

Op de **peilingschaal** scoort 8% van de scholen ondergemiddeld en benadert de verwachting bij de doorlichting. Net zoals bij de PIRLS-schaal heeft 20% van de scholen een gemiddeld schooleffect en de verwachting benaderd volgens de onderwijsinspectie. Daarnaast scoort nog eens 20% van de scholen ondergemiddeld op de peilingschaal, maar wel volgens de verwachting bij de doorlichting. Bovendien scoort 44% van de scholen gemiddeld op de peilingschaal en volgens de verwachting bij de doorlichting. Tot slot heeft 8% van de scholen een bovengemiddeld schooleffect bij de peilingschaal en een inschaling 'volgens de verwachting' bij de doorlichting. Geen enkele school scoort 'beneden de verwachting' of 'overstijgt de verwachting'.



Figuur 16. Puntgrafiek van scholen volgens hun type 0-schooleffect op de PIRLS- en peilingschaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

Tabel 13

Kruistabel van scholen volgens hun type 0-schooleffect op de PIRLS- of peilingschaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

	Ontwikkelingsschaal leereffecten Nederlands (doorlichtingen)			
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting
PIRLS-schaal				
Ondergemiddeld	-	2 (8%)	6 (24%)	-
Gemiddeld	-	5 (20%)	10 (40%)	-
Bovengemiddeld	-	-	2 (8%)	-
Peilingschaal				
Ondergemiddeld	-	2 (8%)	5 (20%)	-
Gemiddeld	-	5 (20%)	11 (44%)	-
Bovengemiddeld	-	-	2 (8%)	-
Totaal	-	7 (28%)	18 (72%)	-

4.1.2 Type AA-schooleffect

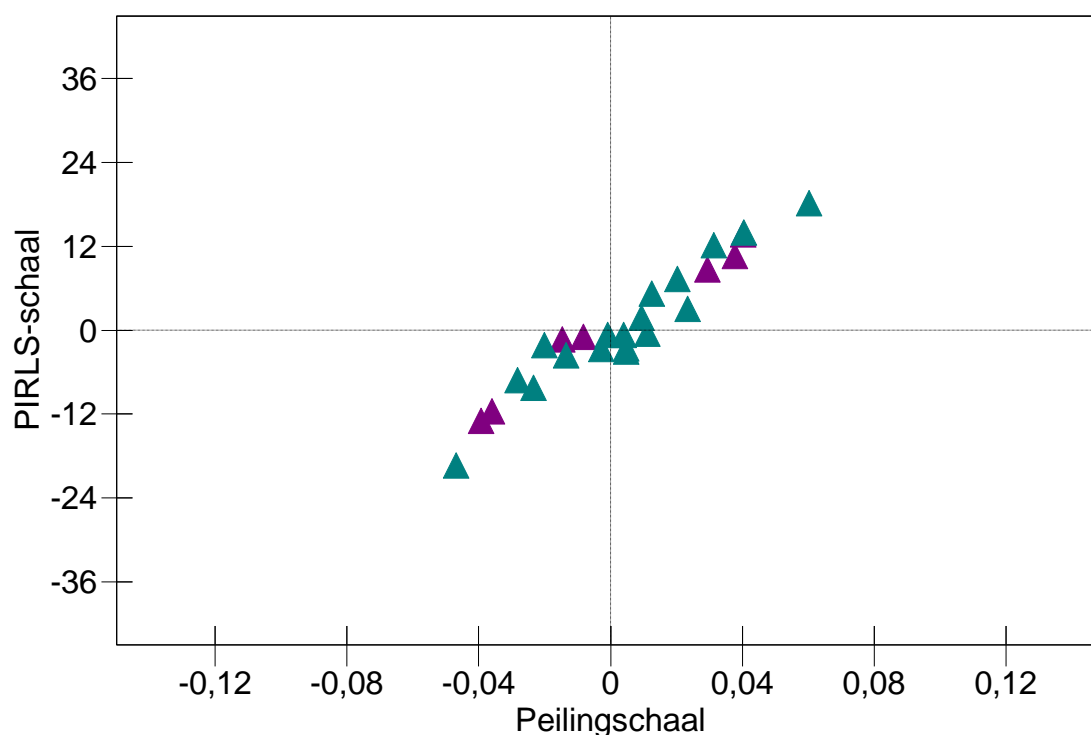
De type AA-schooleffecten (uit Vergelijking 1, zie Figuur 6) van de 25 doorgelichte scholen worden in Figuur 17 weergegeven. Die figuur toont daarnaast de inschaling van de onderwijsinspectie:

- De **paarse** scholen (28%) benaderen de verwachting voor de schaal 'leereffecten'.
- De **blauwe** scholen (72%) scoren volgens de verwachting voor de schaal 'leereffecten'.

Tabel 14 toont dezelfde resultaten, weliswaar niet visueel maar kwantitatief. De resultaten moeten wel met de nodige voorzichtigheid geïnterpreteerd worden omwille van het beperkte aantal scholen

Zo'n 24% van de scholen scoort gemiddeld op de **PIRLS-schaal** en benadert de verwachting volgens de onderwijsinspectie. Daarnaast is er een school (4%) met een bovengemiddeld schooleffect die door de onderwijsinspectie echter ingedeeld wordt als 'benadert de verwachting'. Omgekeerd, is er een andere school (4%) die ondergemiddeld scoort op de PIRLS-schaal, maar wel volgens de verwachting bij de doorlichting. De overgrote meerderheid (64%) scoort gemiddeld op de PIRLS-schaal en volgens de verwachting bij de doorlichting. Ten slotte heeft een school (4%) een bovengemiddeld schooleffect op de PIRLS-schaal en een inschaling van 'volgens de verwachting' bij de doorlichting. Geen enkele school scoort 'beneden de verwachting' of 'overstijgt de verwachting'.

Bij het type AA-schooleffect op de **peilingschaal** heeft 24% van de scholen een gemiddeld schooleffect en de verwachting benaderd volgens de onderwijsinspectie. Net zoals bij de PIRLS-schaal is er een school (4%) met een bovengemiddeld schooleffect die door de onderwijsinspectie nochtans wordt ingedeeld als 'benadert de verwachting'. Verder scoort 68% van de scholen gemiddeld op de peilingschaal en volgens de verwachting bij de doorlichting. Als laatste scoort een school (4%) bovengemiddeld op de peilingschaal en volgens de verwachting bij de doorlichting. Er is geen enkele school die 'beneden de verwachting' of 'overstijgt de verwachting' behaalt bij de doorlichtingen.



Inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

- ▲ Benadert de verwachting
- ▲ Volgens de verwachting

Figuur 17. Puntgrafiek van scholen volgens hun type AA-schooleffect op de PIRLS- en peilingschaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

Tabel 14

Kruistabel van scholen volgens hun type AA-schooleffect op de PIRLS- of peilingschaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

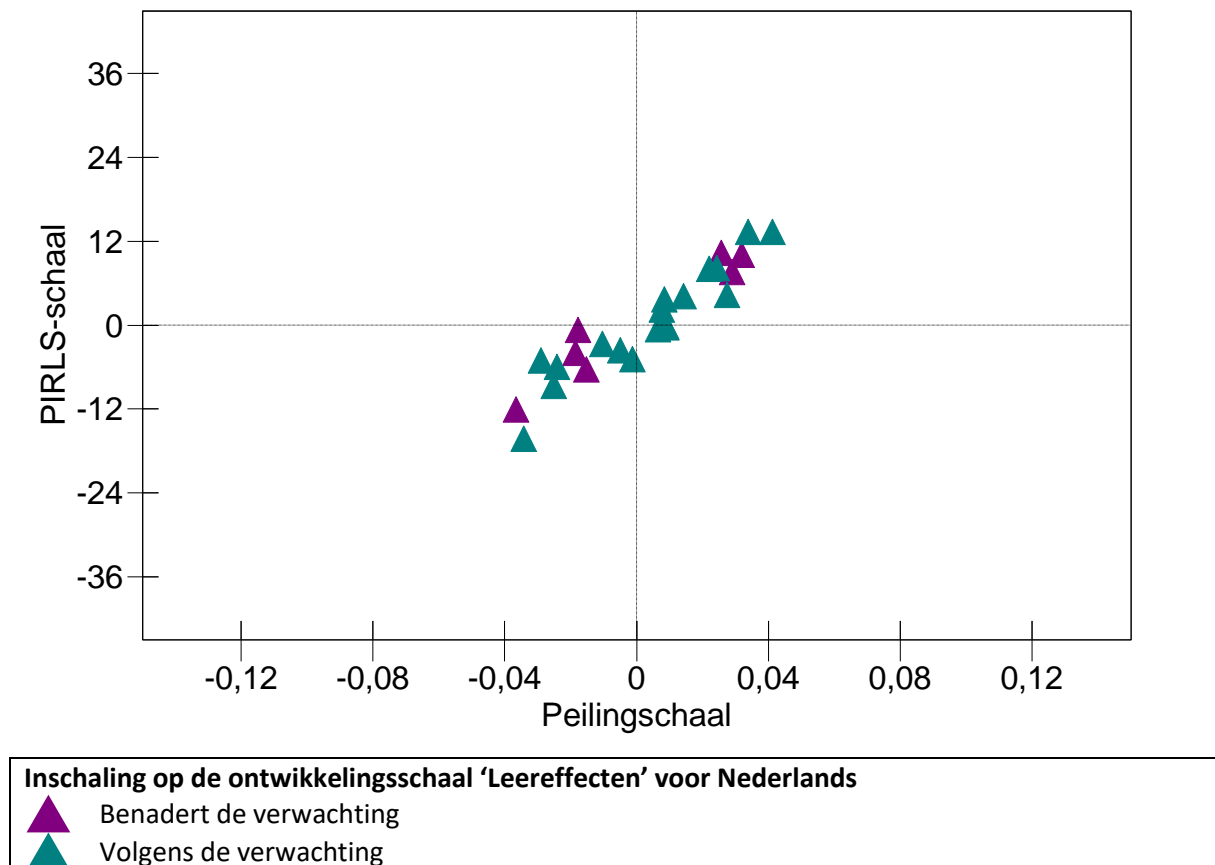
	Ontwikkelingsschaal leereffecten Nederlands (doorlichtingen)			
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting
PIRLS-schaal				
Ondergemiddeld	-	-	1 (4%)	-
Gemiddeld	-	6 (24%)	16 (64%)	-
Bovengemiddeld	-	1 (4%)	1 (4%)	-
Peilingschaal				
Ondergemiddeld	-	-	-	-
Gemiddeld	-	6 (24%)	17 (68%)	-
Bovengemiddeld	-	1 (4%)	1 (4%)	-
Totaal	-	7 (28%)	18 (72%)	-

4.1.3 Type A- schooleffect

Wanneer we de resultaten voor de type A-schooleffecten uit Vergelijking 1 (zie Figuur 9) weergeven voor de 25 scholen die doorgelicht zijn en deelnamen aan PIRLS Repeat zijn, resulteert dat in Figuur 18. Daarnaast toont die figuur de inschaling van de onderwijsinspectie bij de doorlichting. Dezelfde

resultaten worden kwantitatief weergegeven in Tabel 15. De resultaten moeten wel met de nodige voorzichtigheid geïnterpreteerd worden omwille van het beperkte aantal scholen

Uit Figuur 18 en Tabel 15 blijkt dat de resultaten gelijkaardig zijn voor de **PIRLS- en peilingschaal**. Zo'n 28% van de scholen scoort gemiddeld op de PIRLS- of peilingschaal en benadert de verwachting volgens de onderwijsinspectie. Dat zijn de scholen die in het paars zijn aangeduid. De overige 72% scoort ook gemiddeld op de PIRLS- of peilingschaal, maar behaalt de inschaling 'volgens de verwachting' bij de doorlichting. Dat zijn de blauwe scholen in Figuur 18. Er is geen enkele school die 'beneden de verwachting' of 'overstijgt de verwachting' behaalt bij de doorlichtingen.



Figuur 18. Puntgrafiek van scholen volgens hun type A-schooleffect op de PIRLS- en peilingschaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

Tabel 15

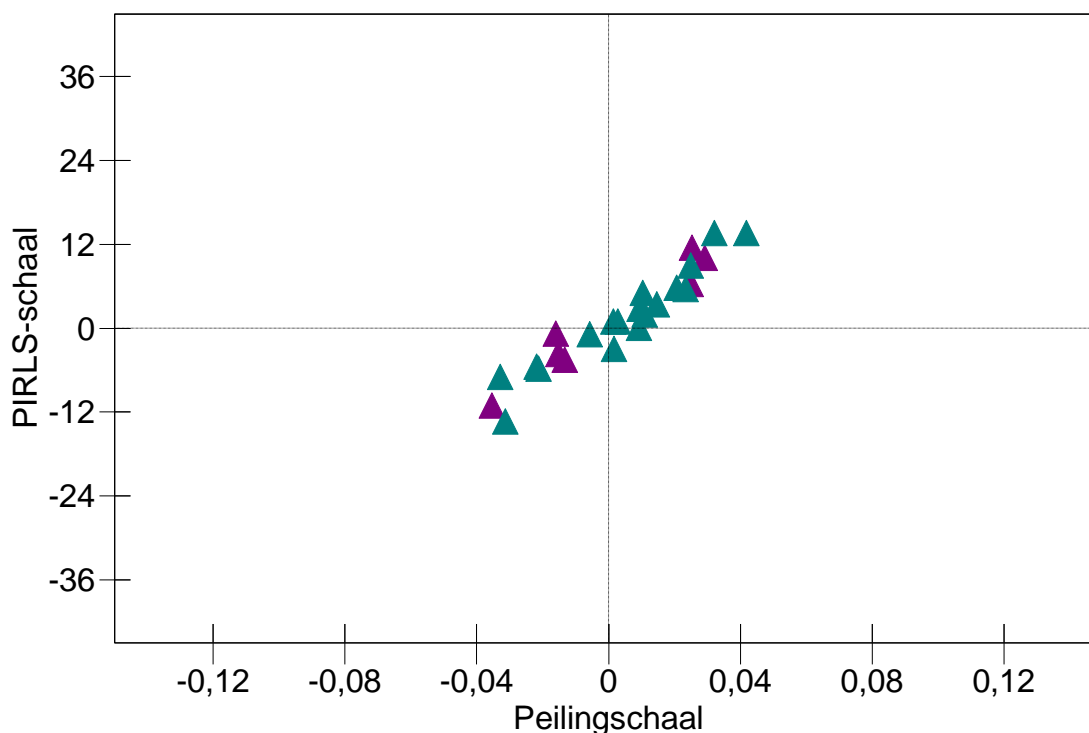
Kruistabel van scholen volgens hun type A-schooleffect op de PIRLS- of peilingschaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

	Ontwikkelingsschaal leereffecten Nederlands (doorlichtingen)			
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting
PIRLS-schaal				
Ondergemiddeld	-	-	-	-
Gemiddeld	-	7 (28%)	18 (72%)	-
Bovengemiddeld	-	-	-	-
Peilingschaal				
Ondergemiddeld	-	-	-	-
Gemiddeld	-	7 (28%)	18 (72%)	-
Bovengemiddeld	-	-	-	-
Totaal	-	7 (28%)	18 (72%)	-

4.1.4 Type B-schooleffect

De type B-schooleffecten (uit Vergelijking 1, zie Figuur 12) van de 25 doorlichte scholen worden in Figuur 19 weergegeven. Die figuur toont daarnaast de inschaling van de onderwijsinspectie. Tabel 16 toont dezelfde resultaten, weliswaar niet visueel maar kwantitatief. De resultaten moeten wel met de nodige voorzichtigheid geïnterpreteerd worden omwille van het beperkte aantal scholen

Daaruit blijkt dat de resultaten gelijkaardig zijn voor het type B-schooleffect op de **PIRLS- en peilingschaal**. De parse scholen (28%) scoren gemiddeld op de PIRLS- of peilingschaal en benaderen de verwachting volgens de onderwijsinspectie. Daarnaast scoort 72% van de scholen (de blauwe scholen) gemiddeld op de PIRLS- of peilingschaal en 'volgens de verwachting' bij de doorlichting. Geen enkele school behaalt 'beneden de verwachting' of 'overstijgt de verwachting' bij de doorlichtingen.



Inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

- ▲ Benadert de verwachting
- ▲ Volgens de verwachting

Figuur 19. Puntgrafiek van scholen volgens hun type B-schooleffect op de PIRLS- en peilingschaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

Tabel 16

Kruistabel van scholen volgens hun type B-schooleffect op de PIRLS- of peilingschaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

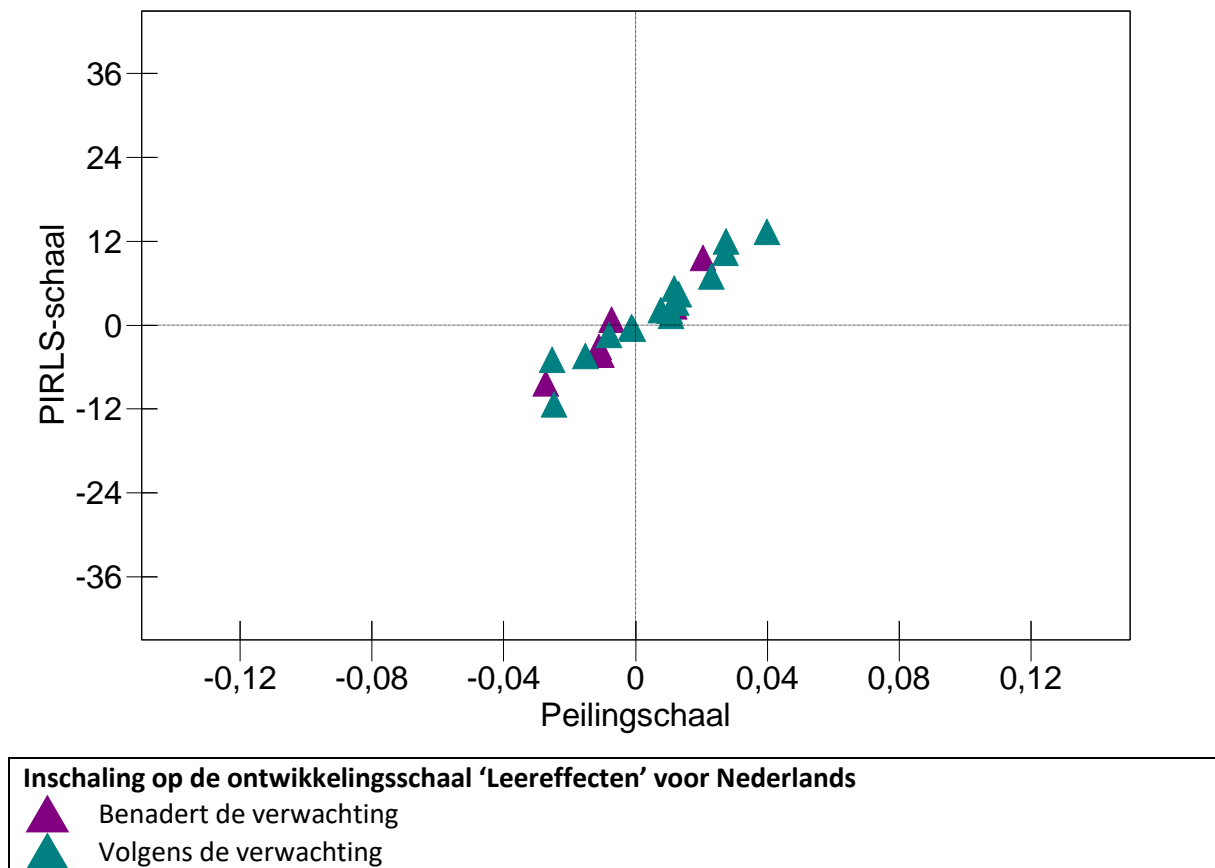
	Ontwikkelingsschaal leereffecten Nederlands (doorlichtingen)			
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting
PIRLS-schaal				
Ondergemiddeld	-	-	-	-
Gemiddeld	-	7 (28%)	18 (72%)	-
Bovengemiddeld	-	-	-	-
Peilingschaal				
Ondergemiddeld	-	-	-	-
Gemiddeld	-	7 (28%)	18 (72%)	-
Bovengemiddeld	-	-	-	-
Totaal	-	7 (28%)	18 (72%)	-

4.1.5 Type X-schooleffect

Wanneer we de resultaten voor de type X-schooleffecten uit Vergelijking 1 (zie Figuur 15) weergeven voor de scholen die doorgelicht zijn en deelnamen aan PIRLS Repeat zijn, leidt dat tot Figuur 20. Daarnaast toont die figuur de inschaling van de onderwijsinspectie bij de doorlichting. Dezelfde resultaten worden – weliswaar kwantitatief in plaats van visueel – weergegeven in Tabel 17. De

resultaten moeten wel met de nodige voorzichtigheid geïnterpreteerd worden omwille van het beperkte aantal scholen

Uit Tabel 17 en Figuur 20 blijkt dat de resultaten gelijkaardig zijn voor de **PIRLS- en peilingschaal**. Bij 27% van de scholen is het type X-schooleffect gemiddeld en wordt de verwachting benaderd volgens de inschaling van de onderwijsinspectie. Dat zijn de paarse scholen in Figuur 20. Zo'n 73% van de scholen scoort gemiddeld op de PIRLS- of peilingschaal en behaalt 'volgens de verwachting' bij de doorlichting. Die scholen worden in het blauw aangeduid in Figuur 20. Geen enkele school behaalt 'beneden de verwachting' of 'overstijgt de verwachting' bij de doorlichtingen.



Figuur 20. Puntgrafiek van scholen volgens hun type X-schooleffect op de PIRLS- en peilingschaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

Tabel 17

Kruistabel van scholen volgens hun type X-schooleffect op de PIRLS- of peilingschaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

	Ontwikkelingsschaal leereffecten Nederlands (doorlichtingen)			
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting
PIRLS-schaal				
Ondergemiddeld	-	-	-	-
Gemiddeld	-	6 (27%)	16 (73%)	-
Bovengemiddeld	-	-	-	-
Peilingschaal				
Ondergemiddeld	-	-	-	-
Gemiddeld	-	6 (27%)	16 (73%)	-
Bovengemiddeld	-	-	-	-
Totaal	-	6 (27%)	16 (73%)	-

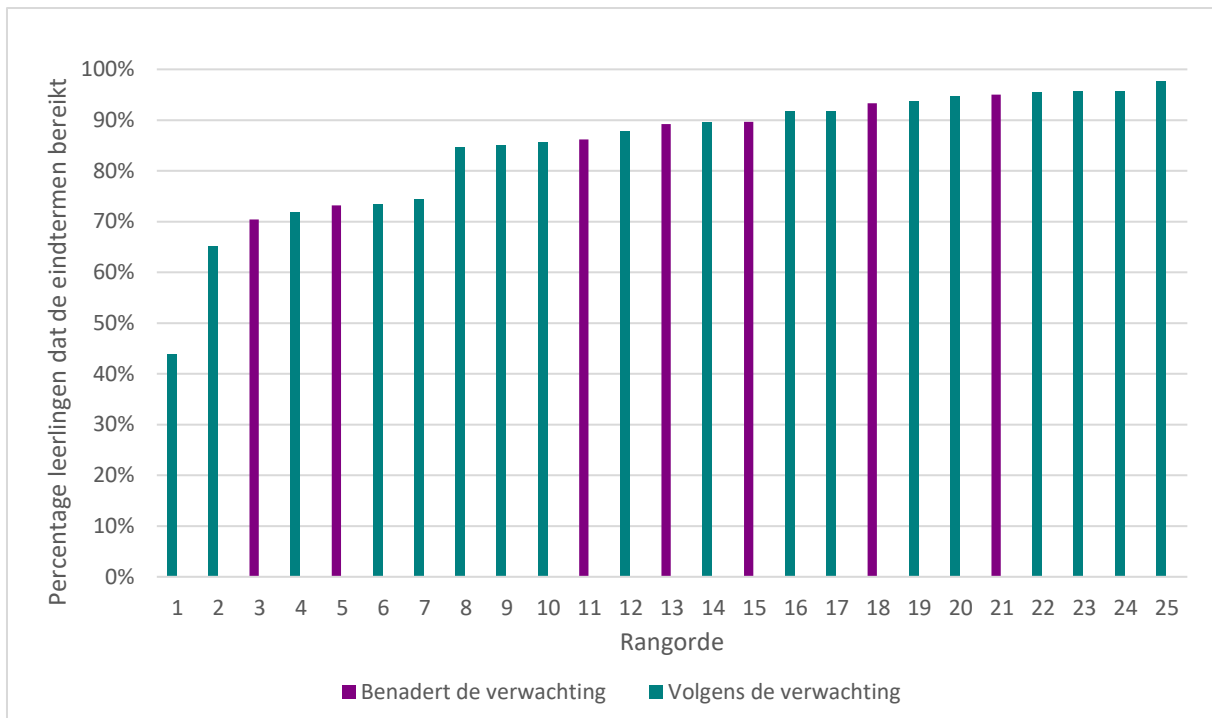
4.1.6 Vergelijking over de verschillende types schooleffecten heen: hoofdbevindingen

Wanneer we de schooleffecten op de PIRLS- en peilingschaal vergelijken met de inschalingen op de schaal 'leereffecten' van de onderwijsinspectie stellen we geen grote afwijkingen vast. Bovendien stijgt de overeenkomst naarmate we meer controlevariabelen toevoegen aan de toegevoegde waarde modellen. Met andere woorden, wanneer er gecontroleerd wordt voor achtergrondkenmerken sluiten de vastgestelde schooleffecten op de PIRLS- en peilingschaal beter aan bij de vaststellingen uit de doorlichtingen. Die resultaten liggen in lijn met het doorlichtingsconcept dat de onderwijsinspectie vooropstelt. De onderwijsinspectie geeft immers zelf aan dat ze bij de doorlichtingen steeds rekening houden met de specifieke context van de school (Onderwijsinspectie, 2018a).

Een belangrijke kanttekening bij bovenstaande vergelijking is dat de onderwijsinspectie werkt met een absoluut criterium (de verwachting volgens het OK-kader), terwijl PIRLS en het peilingsonderzoek gebruik maken van een relatief criterium. Dat heeft als gevolg dat 'volgens de verwachting' scores op de doorlichting niet noodzakelijk overeenkomt met gemiddeld scores op de PIRLS- of peilingschaal. Zo is het zelfs in een scenario waarbij elke school aan de verwachting voldoet nog steeds mogelijk dat bepaalde scholen onder het gemiddelde scores.

4.2 Behalen van de eindtermen (peilingschaal) en inschaling bij de doorlichtingen

Voor de 25 scholen die deelnamen aan PIRLS Repeat én doorgelicht zijn, geven we in Figuur 21 voor elke school het percentage leerlingen weer dat de eindtermen voor lezen bereikt. Daarbij is er geen controle voor achtergrondkenmerken. In die grafiek wordt daarenboven voor elke school de inschaling op de schaal 'leereffecten' voor het leergebied Nederlands getoond.

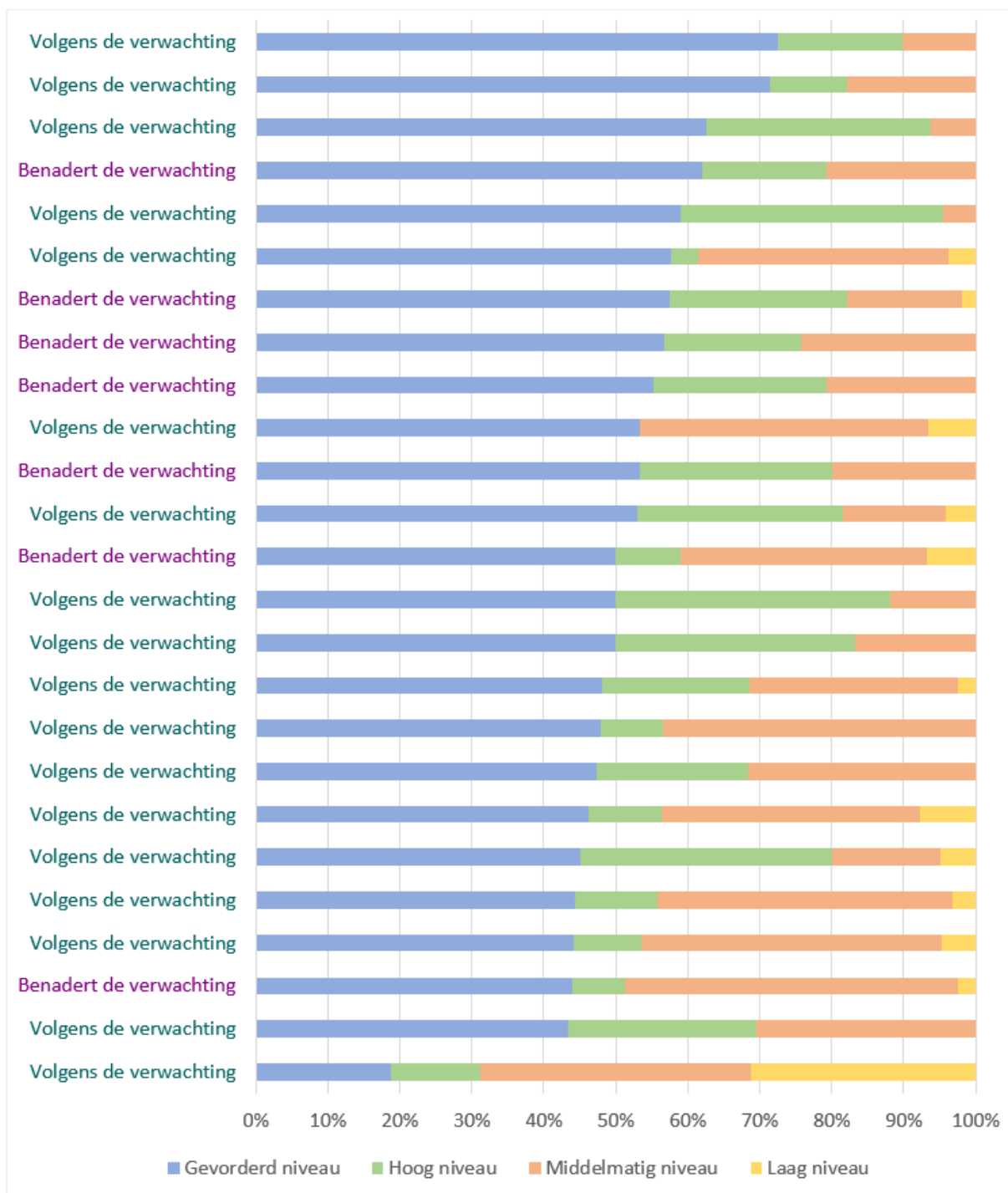


Figuur 21. Grafiek van scholen volgens het percentage leerlingen dat de eindtermen voor lezen bereikt (peilingschaal) en hun inschaling op de ontwikkelingschaal 'Leereffecten' voor Nederlands

Figuur 21 toont dat er weinig concordantie is tussen de inschaling op de schaal 'leereffecten' door de onderwijsinspectie en het percentage leerlingen dat de eindtermen bereikt. Enerzijds zijn er scholen die volgens de verwachting scoren en een laag percentage leerlingen hebben dat de eindtermen bereikt. Anderzijds zijn er eveneens scholen met een hoog percentage leerlingen dat de eindtermen voor lezen bereikt, die volgens de inspectie de verwachtingen voor het leergebied Nederlands benaderen.

4.3 Posities binnen de benchmarks op de PIRLS-schaal en inschaling bij de doorlichtingen

Voor de 25 scholen (uit Vergelijking 2) wordt in Figuur 22 de positie van hun leerlingen binnen de PIRLS-benchmarks weergegeven samen met de inschaling van de onderwijsinspectie voor de schaal 'leereffecten'. Ter verduidelijking, in die grafiek worden de leerlingen gepositioneerd binnen de benchmarks op basis van hun score op PIRLS Repeat (zonder controle voor achtergrondkenmerken). Het gaat dus om scores van leerlingen uit het zesde – en niet uit het vierde – leerjaar.



Figuur 22. Grafiek van scholen volgens het percentage leerlingen dat de benchmarks op de PIRLS-schaal bereikt en hun inschaling op de ontwikkelingschaal 'Leereffecten' voor Nederlands

Figuur 22 geeft een eerder divergerend beeld weer. Meerdere scholen die de verwachting benaderen, blijken scholen te zijn waarin een groot percentage van de leerlingen het gevorderd niveau behaalt. Onderaan in de figuur vinden we voornamelijk scholen die volgens de verwachting scoren. Nochtans zijn dat de scholen waarbij een laag percentage leerlingen uit het zesde leerjaar het gevorderd niveau van PIRLS behaalt.

Studie 2: Scholen met meer dan één recente bron van kwaliteitsindicatoren

1. Inleiding

Heel wat scholen beschikken over meerdere recente bronnen van kwaliteitsindicatoren. Zo zijn er verschillende scholen die recent doorgelicht werden door de onderwijsinspectie én participeerden aan een internationaal vergelijkend onderzoek of peilingsonderzoek. Ook zijn er scholen die deelnamen aan een peilingsonderzoek en een internationaal vergelijkend onderzoek. De vraag is dan: in welke mate komt het resultaat van een bepaalde school op de ene bron overeen met het resultaat van dezelfde school bij een andere bron? Om die vraag te beantwoorden, focussen we in de huidige studie op de concordantie tussen de resultaten op schoolniveau bij de verschillende bronnen van externe kwaliteitszorg.

2. Onderzoeksopzet

In deze studie gaan we na in welke mate de resultaten van de doorlichtingen, het Vlaams peilingsonderzoek en het internationaal vergelijkend onderzoek convergeren of divergeren op schoolniveau. Daarvoor maken we gebruik van de volgende recente bronnen:

- peiling basisonderwijs Nederlands 2018;
- peiling wiskunde 2021;
- peiling wiskunde A-stroom 2018;
- peiling wiskunde B-stroom 2019;
- PIRLS 2016;
- TIMSS 2019;
- PISA 2018;
- doorlichtingen volgens de Inspectie 2.0-aanpak in schooljaren 2018-2019 en 2019-2020.

Concreet hebben we eerst de schooleffecten geschat voor de scholen die deelnamen aan een van bovenstaande kwantitatieve studies én de data van de verschillende bronnen met elkaar gekoppeld:

- **Schatting van de schooleffecten**

We hebben eerst de schooleffecten geschat van alle scholen die deelnamen aan een van bovenstaande peilingen of internationaal vergelijkende onderzoeken. Net zoals bij Studie 1 schatten we voor elke school verschillende types schooleffecten.²⁴ Concreet hebben we de type 0-, type A- en type B-schooleffecten²⁵ geschat rekening houdend met de

²⁴ De verschillende types schooleffecten zijn eerder al toegelicht in de eerste cross-instrumentele studie. Meer informatie daarover is te vinden bij '2.1 PIRLS- en peilingschaal (Vergelijking 1)'.

Zoals eerder aangegeven, wordt er bij PIRLS, TIMSS en PISA bovendien gewerkt met plausible values. Bij de schatting van de schooleffecten werd er verder gewerkt met die plausible values. Concreet wordt er op basis van elk van de plausible values een apart schooleffect geschat, waarna die schooleffecten op het einde samengevoegd worden aan de hand van de methode voor *multiple imputations* (Graham et al., 2007).

²⁵ Zoals aangegeven in de eerste studie, verschilt de betekenis van het schooleffect afhankelijk van het type:

- De type 0-schooleffecten geven weer hoe goed de leerlingen van een bepaalde school presteren in vergelijking met andere scholen zonder te controleren voor de achtergrondkenmerken van leerlingen of scholen.

controlevariabelen in Tabel 18. In lijn met Studie 1 is ervoor gekozen om onderstaande controlevariabelen op te nemen bij de schatting van de schooleffecten.²⁶

Tabel 18

Overzicht van opgenomen controlevariabelen per type model

Type 0-model	Type A-model	Type B-model
Geen controlevariabelen	<ul style="list-style-type: none"> • Thuistaal • Geslacht • SES • Schoolse vordering (op leeftijd, voorsprong of vertraging) 	<ul style="list-style-type: none"> • Thuistaal • Geslacht • SES • Schoolse vordering (op leeftijd, voorsprong of vertraging) • Gemiddelde SES van de school

Noot. De thuistaal en de SES van de leerlingen wordt niet in elke bron op dezelfde manier gemeten (zie Bijlage 1).

Na controle voor bovenstaande variabelen blijft er – afhankelijk van het type model – een schooleffect (of residu) over dat niet te wijten is aan die controlevariabelen. Aan de hand van die residuen en de bijhorende betrouwbaarheidsintervallen kunnen we de scholen indelen in drie groepen: scholen met (1) een bovengemiddeld schooleffect, (2) een gemiddeld schooleffect en (3) een ondergemiddeld schooleffect. Die schooleffecten stellen de resultaten van het peilingsonderzoek en internationaal vergelijkend onderzoek op schoolniveau voor.

- **Data van verschillende bronnen koppelen**

Om de concordantie tussen de resultaten van twee verschillende bronnen in kaart te brengen, dienen we beroep te doen op scholen waarvoor informatie uit die twee databronnen beschikbaar is. Daarom koppelen we – ter voorbereiding – de verschillende databronnen aan de hand van de administratieve codes van de scholen²⁷. Die koppelingen zijn noodzakelijk om de concordantie op schoolniveau in kaart te brengen.

Zodra de schooleffecten geschat zijn en de databronnen gekoppeld zijn, kunnen we de resultaten op schoolniveau vergelijken. Op die manier kunnen bijvoorbeeld de resultaten van de doorlichting van een bepaalde basisschool vergeleken worden met de schooleffecten bij TIMSS 2019 van diezelfde school. Concreet maken we drie vergelijkingen tussen telkens twee verschillende bronnen van externe kwaliteitszorg. In wat volgt, wordt elke vergelijking kort toegelicht.

-
- De type A-schooleffecten geven aan hoe goed de leerlingen in een bepaalde school scoren in vergelijking met leerlingen die gelijkaardige achtergrondkenmerken hebben uit andere scholen.
 - De type B-schooleffecten tonen hoe goed de leerlingen in een bepaalde school scoren in vergelijking met scholen die gelijkaardige leerlingen en een gelijkaardige schoolcompositie hebben.

²⁶ In tegenstelling tot de eerste studie worden de eerdere prestaties in de tweede studie niet opgenomen als controlevariabele. Omwille van het cross-sectionele karakter van de peilingen en het internationaal vergelijkend onderzoek zijn er immers geen data over de eerdere prestaties van de leerlingen. Bijgevolg is het niet mogelijk om de eerdere prestaties op te nemen als controlevariabele bij de schatting. Daarom beperken we ons in deze studie tot nulmodellen (type 0-modellen) en conditionele statusmodellen (type A- en B-modellen) waarbij niet gecontroleerd wordt voor eerdere prestaties (Leckie & Prior, 2022; Lenkeit, 2013).

²⁷ Na de koppeling worden de data van de scholen opnieuw geanonimiseerd.

2.1 Doorlichtingen en internationaal vergelijkend onderzoek (Vergelijking 1)

Uit de koppeling van de verschillende bronnen blijkt dat er heel wat scholen doorgelicht zijn voor een leergebied (lager onderwijs) of vak (secundair onderwijs) dat sterk aansluit bij de focus van een internationaal vergelijkend onderzoek waaraan ze deelnamen:

- 27 scholen namen deel aan PIRLS 2016 en werden doorgelicht voor het leergebied Nederlands.
- 21 scholen namen deel aan TIMSS 2019 en werden doorgelicht voor het leergebied wiskunde.
- 7 scholen namen deel aan PISA 2018 en werden doorgelicht voor het vak Nederlands.
- 19 scholen namen deel aan PISA 2018 en werden doorgelicht voor het vak wiskunde.

Voor elk van die scholen kunnen we hun schooleffecten – geschat op basis van het internationaal vergelijkend onderzoek – vergelijken met de beoordeling op de U7-schaal 'leereffecten'²⁸ voor respectievelijk Nederlands of wiskunde bij de doorlichting. Bij de inschaling van die schaal kent de onderwijsinspectie steeds een van de volgende beoordelingen toe aan de school: (1) beneden de verwachting, (2) benadert de verwachting, (3) volgens de verwachting en (4) overstijgt de verwachting (Onderwijsinspectie, 2021c, 2021d).

2.2 Doorlichtingen en peilingsonderzoek (Vergelijking 2)

Daarnaast zijn er ook scholen die doorgelicht werden voor een leergebied (lager onderwijs) of vak (secundair onderwijs) dat sterk aansluit bij de focus van een peilingsonderzoek waaraan ze deelnamen:

- 20 scholen namen deel aan de peiling Nederlands lezen in 2018 en werden doorgelicht voor het leergebied Nederlands.
- 19 scholen namen deel aan de peiling wiskunde in 2021 en werden doorgelicht voor het leergebied wiskunde.

Er wordt geen vergelijking opgesteld tussen de peiling wiskunde in de A- en B-stroom enerzijds en de doorlichtingen anderzijds omwille van het beperkte aantal scholen (respectievelijk 3 en 2 scholen voor de A- en B-stroom).

Voor bovenstaande scholen die deelnamen aan de peilingen Nederlands en wiskunde vergelijken we hun schooleffecten met de inschaling van de onderwijsinspectie op de schaal leereffecten.

Daarbij is het belangrijk om op te merken dat de peiling wiskunde uit 2021 opgebouwd is aan de hand van 20 verschillende toetsen, waarvan elke leerling er slechts een beperkt aantal aflegt. De peilingen leiden daardoor niet tot een algemene score per leerling voor de hele peiling wiskunde. Wel is er een algemene score per leerling beschikbaar voor elke toets waaraan die leerling deelnam. In de huidige studie hebben we voor vier schalen exemplarisch een vergelijking gemaakt met de inschaling bij de doorlichting. Concreet gaat het om de schalen (1) Functies, (2) Omtrek, oppervlakte en inhoud, (3) Problemen oplossen bij getallen en bewerkingen en (4) Problemen oplossen bij meten, ruimte en meetkunde. Die schalen sluiten inhoudelijk het meest aan bij het conceptueel kader van TIMSS en lenen zich dus het best voor een vergelijking voor wiskunde in het basisonderwijs tussen de doorlichtingen, het peilingsonderzoek en het internationaal vergelijkend onderzoek.

²⁸ Aan de hand van de U7-schaal leereffecten brengt de onderwijsinspectie bij elke school voor een bepaald leergebied (basisonderwijs) of vak (secundair onderwijs) het volgende in kaart: in welke mate tonen de onderwijsleerpraktijk, het welbevinden van de leerlingen en de kwaliteitsbewaking van de onderwijsleerpraktijk aan dat een zo groot mogelijke groep van leerlingen de minimaal gewenste output bereikt (Onderwijsinspectie, 2021c, 2021d).

2.3 Internationaal vergelijkend onderzoek en peilingsonderzoek (Vergelijking 3)

In een derde en laatste vergelijking gaan we de concordantie na bij scholen die recent deelgenomen hebben aan zowel een peilingsonderzoek als een internationaal vergelijkend onderzoek. Meer specifiek zijn we geïnteresseerd in de volgende scholen:

- 9 scholen namen deel aan PIRLS 2016 en de peiling Nederlands lezen in 2018.
- 11 scholen namen deel aan PISA 2018 en de peiling wiskunde A-stroom in 2018.
- 20 scholen namen deel aan PISA 2018 en de peiling wiskunde B-stroom in 2019.

Aangezien slechts een beperkt aantal scholen deelnam aan TIMSS 2019 en de peiling wiskunde in 2021, maken we daarvoor geen vergelijking.

Net zoals bij de vorige vergelijkingen worden ook hier de resultaten op schoolniveau vergeleken om de concordantie tussen de resultaten van beide bronnen te vergelijken.

Ook de peilingen wiskunde in de A- en B-stroom zijn opgebouwd aan de hand van verschillende toetsen, waarvan elke leerling er slechts een beperkt aantal aflegt. Met als gevolg, er zijn enkel scores beschikbaar voor een specifieke leerling bij de toetsen waaraan die leerling deelnam. In de huidige studie hebben we voor de volgende schalen – die inhoudelijk aansluiten bij het conceptueel kader van PISA 2018 – exemplarisch een vergelijking gemaakt met wiskundige geletterdheid bij PISA:

- A-stroom: (1) Algebraïsering, (2) Evenredigheden, (3) Omgaan met data en (4) Meetkundige procedures: constructies.
- B-stroom: (1) Getalinzicht en hoofdbewerkingen, (2) Meetkunde, (3) Informatieverwerking en -verwerking en (4) Meten.

3. Resultaten

Omwille van de omvang van de resultaten worden de grafieken en tabellen van elk van bovenstaande vergelijkingen weergegeven in bijlage en bespreken we hierna exemplarisch een beperkt aantal vergelijkingen om de hoofdbevindingen uit deze studie te illustreren.

3.1 Leeswijzer voor de resultaten

Hieronder lichten we per vergelijking kort toe waar de grafieken en tabellen van de verschillende vergelijkingen te vinden zijn en hoe die resultaten gelezen kunnen worden:

- **Doorlichtingen en internationaal vergelijkend onderzoek (Vergelijking 1)**

De resultaten van de scholen die deelgenomen hebben aan een internationaal vergelijkend onderzoek én doorgelicht zijn voor een leergebied of vak dat sterk aansluit bij de focus van dat onderzoek worden weergegeven in Bijlage 2 – Vergelijking 1 (p. 237).

In die resultaten worden de verschillende types schooleffecten – die geschat zijn op basis van de data uit het internationaal vergelijkend onderzoek – gevisualiseerd aan de hand van een rupsgrafiek (zie ‘3.2 Schooleffecten volgens de PIRLS- en peilingschaal’ voor een korte leeswijzer omtrent rupsgrafieken). Elke school in de rupsgrafiek wordt vervolgens aangeduid met een kleur die de inschaling op de schaal ‘leereffecten’ bij de doorlichting weergeeft. De bijhorende tabellen geven dezelfde resultaten kwantitatief weer.

- **Doorlichtingen en peilingsonderzoek (Vergelijking 2)**

Bijlage 2 – Vergelijking 2 (p. 249) omvat de resultaten van de scholen die deelgenomen hebben aan een peilingsonderzoek én doorgelicht zijn voor een leergebied of vak dat sterk aansluit bij de focus van die peiling. De resultaten worden net zoals bij de eerste vergelijking weergegeven

aan de hand van tabellen en rupsgrafieken waarin elke school aangeduid wordt met de kleur die overeenstemt met het resultaat bij de doorlichtingen.

- **Internationaal vergelijkend onderzoek en peilingsonderzoek (Vergelijking 3)**

De resultaten van de scholen die deelgenomen hebben aan een internationaal vergelijkend onderzoek én een peilingsonderzoek met een gelijkaardige focus zijn opgenomen in Bijlage 2 – Vergelijking 3 (p. 264).

Daarin worden de resultaten getoond aan de hand van een puntgrafiek. In die grafiek worden de schooleffecten die geschat zijn op basis van de data uit het internationaal vergelijkend onderzoek (weergegeven op de y-as) en het peilingsonderzoek (weergegeven op de x-as) met elkaar vergeleken. Elke school in de grafiek wordt bovendien – voor zowel het internationaal vergelijkend onderzoek als het peilingsonderzoek – gecategoriseerd als een school met een (1) bovengemiddeld, (2) gemiddeld of (3) ondergemiddeld schooleffect. Aan de hand van die categorieën kunnen we de concordantie op schoolniveau tussen beide bronnen nagaan. We kunnen immers in kaart brengen in welke mate de schooleffecten van elke school volgens de peiling en het internationaal vergelijkend onderzoek ingedeeld worden in dezelfde categorie. De bijhorende tabellen geven dezelfde resultaten kwantitatief weer.

Bijlage 2 – Bijkomende analyses (p. 291) omvat enkele bijkomende analyses bij de vergelijking tussen PISA en de peilingen wiskunde in de A- en B-stroom. Concreet toont die bijlage de resultaten van:

- Een vergelijking tussen PISA 2018 en de peiling A-stroom wiskunde uit 2018, waarbij er in de analyses enkel leerlingen opgenomen worden die A-stroom, ASO, TSO of KSO volgen en deelnamen aan PISA. De leerlingen die B-stroom of BSO volgen en deelnamen aan PISA zijn uit de dataset verwijderd.
- Een vergelijking tussen PISA 2018 en de peiling B-stroom wiskunde uit 2019, waarbij er in de analyses enkel leerlingen opgenomen worden die B-stroom of BSO volgen en deelnamen aan PISA. De leerlingen die A-stroom, ASO, TSO of KSO volgen en deelnamen aan PISA zijn uit de dataset verwijderd.

Net zoals bij Studie 1 worden er bij elk van bovenstaande vergelijkingen meerdere types schooleffecten weergegeven. Zoals eerder vermeld, beperken we ons in deze studie tot de type 0-, type A- en type B-schooleffecten.

3.2 Hoofdbevindingen

DE OVEREENKOMST MET DE DOORLICHTINGEN WORDT GROTER ALS ER GECONTROLEERD WORDT VOOR KENMERKEN VAN DE LEERLINGEN EN DE SCHOOLCOMPOSITIE (VERGELIJKING 1 & 2).

Bij de eerste en tweede vergelijking stellen we het minst overeenkomst vast bij de type 0-schooleffecten – die geschat zijn op basis van het internationaal vergelijkend onderzoek of het peilingsonderzoek – en de doorlichtingen. Wanneer we de type A- of type B-schooleffecten vergelijken met de doorlichtingen stijgt die overeenkomst aanzienlijk. Kortom, naarmate er gecontroleerd wordt voor kenmerken van de leerlingen en de schoolcompositie stijgt de concordantie met de inschalingen van de doorlichtingen.

Ter illustratie lichten we kort een voorbeeld uit de eerste vergelijking toe: de schooleffecten van TIMSS 2019 en de inschaling bij de doorlichtingen. Dezelfde bevindingen zijn echter ook te vinden bij de doorlichtingen en het peilingsonderzoek (vergelijking 2).

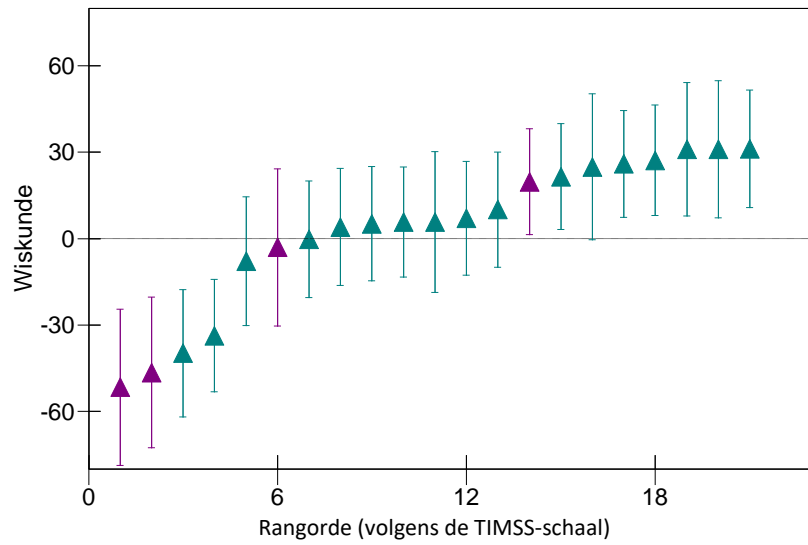
Figuur 23 toont de rupsgrafieken van de verschillende types schooleffecten die geschat zijn op basis van de data van TIMSS 2019. De **paarse** scholen in die rupsgrafieken zijn de scholen die volgens de onderwijsinspectie de verwachting benaderen, terwijl de **blauwe** scholen de inschaling 'volgens de verwachting' behaald hebben bij de doorlichting voor wiskunde.

In de eerste rupsgrafiek (bij de type 0-schooleffecten) zijn er twee scholen die volgens de verwachting scoren bij de onderwijsinspectie, maar wel een ondergemiddeld schooleffect hebben volgens TIMSS. Dat zijn de blauwe scholen die zich onder de nullijn bevinden. Ook is er een paarse school die zich boven de nullijn bevindt. Die school heeft een bovengemiddeld schooleffect volgens TIMSS en de inschaling 'benadert de verwachting' gekregen bij de doorlichting. Voor die scholen is de overeenkomst beperkt tussen het resultaat van de school volgens TIMSS – de schooleffecten – enerzijds en het doorlichtingsverslag dat ze ontvingen van de onderwijsinspectie anderzijds. Voor de andere scholen in de eerste rupsgrafiek is er meer overeenkomst tussen die twee bronnen.

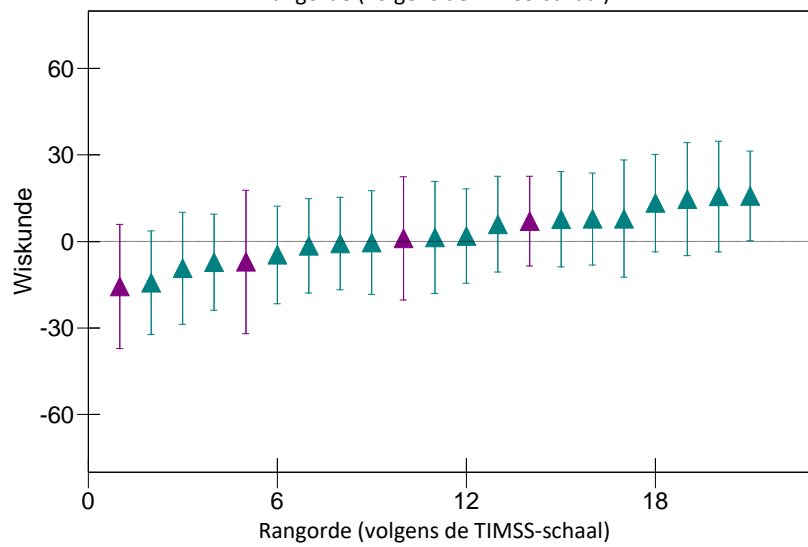
Wanneer we de tweede rupsgrafiek (de type A-schooleffecten) vergelijken met de eerste (type 0-schooleffecten), stijgt de overeenkomst. In tegenstelling tot bij de eerste rupsgrafiek hebben alle scholen in de tweede rupsgrafiek – op de meest rechtse school na – een gemiddeld schooleffect volgens de TIMSS-schaal. De kleuren in de grafiek geven bovendien aan dat diezelfde scholen de verwachting bij de doorlichting benaderen of eraan voldoen. Voor die scholen is er een grote mate van overeenkomst tussen het resultaat van de school volgens TIMSS – de gemiddelde schooleffecten – en het doorlichtingsverslag dat ze ontvingen van de onderwijsinspectie. De meest rechtse school heeft een bovengemiddeld schooleffect en voldoet aan de verwachting volgens de onderwijsinspectie. Ook voor die school leiden de resultaten van de doorlichting en TIMSS tot een convergerend beeld. De concordantie is dus groot voor elk van de deelnemende scholen.

Bij de derde rupsgrafiek (de type B-schooleffecten) hebben alle scholen een gemiddeld schooleffect. Enkele van die scholen benaderen de verwachting, de overgrote meerderheid voldoet aan de verwachting. Ook in de derde rupsgrafiek is de overeenkomst tussen de schooleffecten en de inschalingen van de doorlichtingen groot. Met andere woorden, als we controleren voor de kenmerken van de leerlingen en de schoolcompositie wordt de overeenkomst tussen de resultaten van TIMSS en de doorlichting voor de deelnemende scholen groter.

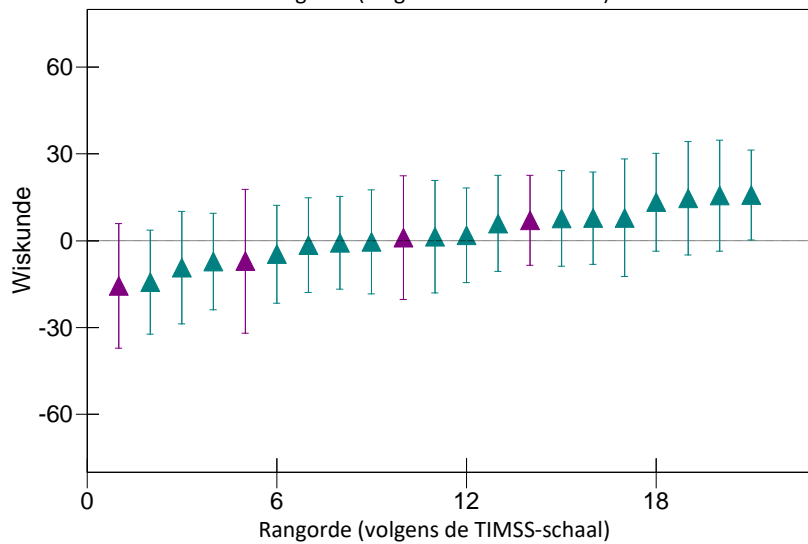
Type 0-schooleffecten



Type A-schooleffecten



Type B-schooleffecten



Inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde
▲ Benadert de verwachting ▲ Volgens de verwachting

Figuur 23. Rupsgrafieken van de vergelijking tussen de schooleffecten geschat op basis van TIMSS 2019 en de inschaling van de doorlichtingen op de ontwikkelingsschaal 'Leereffecten' bij wiskunde

ER ZIJN GEEN GROTE AFWIJINGEN TUSSEN DE RESULTATEN VAN HET INTERNATIONAAL VERGELIJKEND ONDERZOEK EN HET PEILINGSONDERZOEK (VERGELIJKING 3).

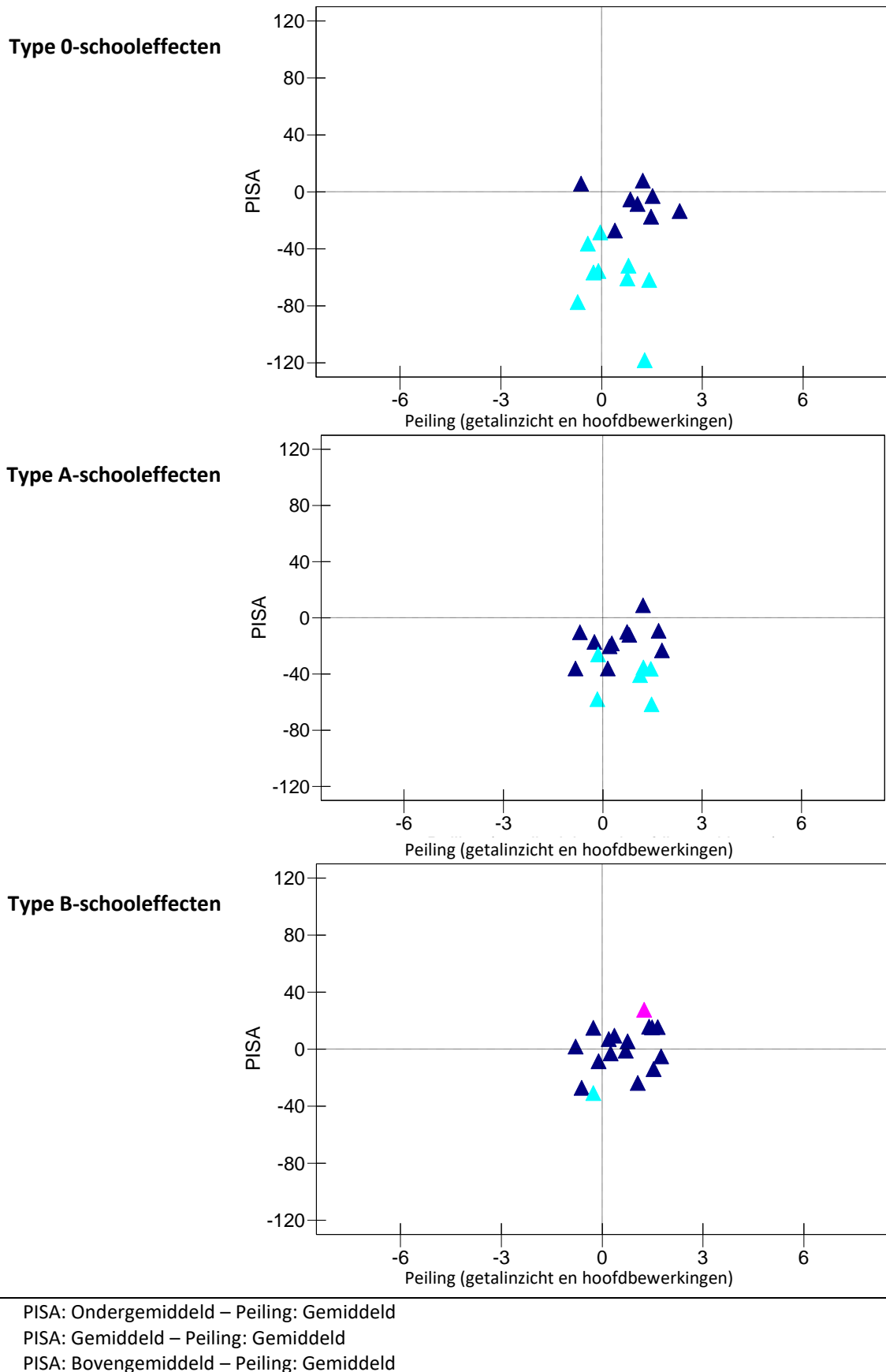
Bij de vergelijking van schooleffecten geschat op basis van internationaal vergelijkend onderzoek enerzijds en schooleffecten geschat op basis van peilingsonderzoek anderzijds stellen we geen grote afwijkingen vast. De peilingen en het internationaal vergelijkend onderzoek leiden bij de meeste scholen immers tot een gelijkaardig resultaat. Wel zijn er bij elke vergelijking scholen waarbij het schooleffect net wel significant afwijkt ten opzichte van het gemiddelde op de ene schaal en net niet significant afwijkt op de andere schaal.

Zo toont Figuur 24 bijvoorbeeld dat er geen grote afwijkingen zijn tussen de peilingschaal getalinzicht en hoofdbewerkingen (peiling wiskunde B-stroom) enerzijds en wiskundige geletterdheid (PISA 2018) anderzijds bij de drie verschillende types schooleffecten. Zo wordt de overgrote meerderheid van de scholen aangeduid in het **donkerblauw**. Die kleur geeft aan dat ze hetzelfde resultaat – namelijk een gemiddeld schooleffect – hebben bij zowel het peilingsonderzoek als het internationaal vergelijkend onderzoek. Wel zijn er ook heel wat **lichtblauwe** scholen die net wel significant afwijken van het gemiddelde op de PISA-schaal – en dus een ondergemiddeld schooleffect hebben, terwijl ze net niet significant afwijken van het gemiddelde op de peilingschaal – en dus een gemiddeld schooleffect hebben. De vergelijkingen van de type 0-, type A- en type B-schooleffecten tonen dus een eerder convergerend beeld.

DE OVEREENKOMST TUSSEN HET INTERNATIONAAL VERGELIJKEND ONDERZOEK EN HET PEILINGSONDERZOEK WORDT GROTER ALS ER GECONTROLEERD WORDT VOOR KENMERKEN VAN DE LEERLINGEN EN DE SCHOOLCOMPOSITIE (VERGELIJKING 3).

Ook bij de derde vergelijking wordt de overeenkomst tussen twee bronnen groter wanneer er gecontroleerd wordt voor kenmerken van de leerlingen en de schoolcompositie. Zoals eerder vermeld stellen we geen grote afwijkingen vast bij de type 0-, type A- en type B-schooleffecten. Toch vermindert het aantal kleine afwijkingen naarmate er gecontroleerd wordt voor kenmerken van de leerlingen en de schoolcompositie. Met andere woorden, de overeenkomst tussen de resultaten van de peiling en het internationaal vergelijkend onderzoek wordt groter.

Ook die bevindingen kunnen geïllustreerd worden aan de hand van Figuur 24. Bij de eerste puntgrafiek (de type 0-schooleffecten) zijn er negen scholen waarbij het schooleffect volgens PISA verschilt van het schooleffect dat geschat is op basis van de peiling. Dat zijn de **lichtblauwe** scholen. In de tweede puntgrafiek (de type A-schooleffecten) is het aantal scholen met een divergerend resultaat kleiner. Er zijn zes **lichtblauwe** scholen met een ondergemiddeld schooleffect volgens PISA en een gemiddeld schooleffect volgens de peiling. De derde en laatste puntgrafiek (de type B-schooleffecten) toont de grootste overeenkomst. Voor de 15 **donkerblauwe** scholen komen PISA en de peiling wiskunde in de B-stroom tot hetzelfde resultaat. Toch zijn er nog steeds twee scholen, namelijk de **lichtblauwe** en **roze** school, waarbij er een – weliswaar klein – verschil is tussen de resultaten volgens PISA en de peiling. Kortom, naarmate er gecontroleerd wordt voor kenmerken van de leerlingen en de schoolcompositie stijgt de overeenkomst tussen het peilingsonderzoek en het internationaal vergelijkend onderzoek.



Figuur 24. Puntgrafieken van de vergelijking tussen de schooleffecten geschat op basis van de schaal 'getalinzicht en hoofdbewerkingen' (peiling wiskunde B-stroom) en PISA 2018

MET WELKE GROEP LEERLINGEN EN SCHOLEN WE VERGELIJKEN, MAAKT EEN VERSCHIL (VERGELIJKING 3: SECUNDAIR ONDERWIJS).

Hoewel het peilingsonderzoek en het internationaal vergelijkend onderzoek elke school met het Vlaams gemiddelde vergelijken, is er een opvallend verschil tussen beide bronnen in het secundair onderwijs. Bij PISA 2018 wordt elke school vergeleken met een groep diverse scholen die niet allemaal dezelfde onderwijsvormen aanbieden. In de steekproef van PISA zitten immers zowel scholen waar leerlingen enkel ASO kunnen volgen als scholen waar de overgrote meerderheid van de leerlingen BSO volgt. Bijgevolg wordt het PISA-gemiddelde berekend op basis van de prestaties van een diverse groep leerlingen, waarna elke school met dat gemiddelde vergeleken wordt. Bij het peilingsonderzoek is dat anders. In de eerste graad secundair onderwijs worden er bijvoorbeeld aparte peilingen georganiseerd voor de A- en de B-stroom. Het peilingsonderzoek maakt dus twee aparte vergelijkingen met het gemiddelde in de A-stroom enerzijds en met het gemiddelde in de B-stroom anderzijds.

Figuur 25 toont de puntgrafieken van het type 0-schooleffecten die geschat zijn op basis van de schaal 'informatieverwerking en -verwerking' bij de peiling wiskunde B-stroom en wiskundige geletterdheid bij PISA 2018. Beide puntgrafieken omvatten enkel scholen met een B-stroom en/of studierichtingen in het BSO, omdat we enkel voor die scholen een vergelijking kunnen maken met de peiling in de B-stroom. Er is echter een belangrijk verschil tussen de eerste en de tweede puntgrafiek:

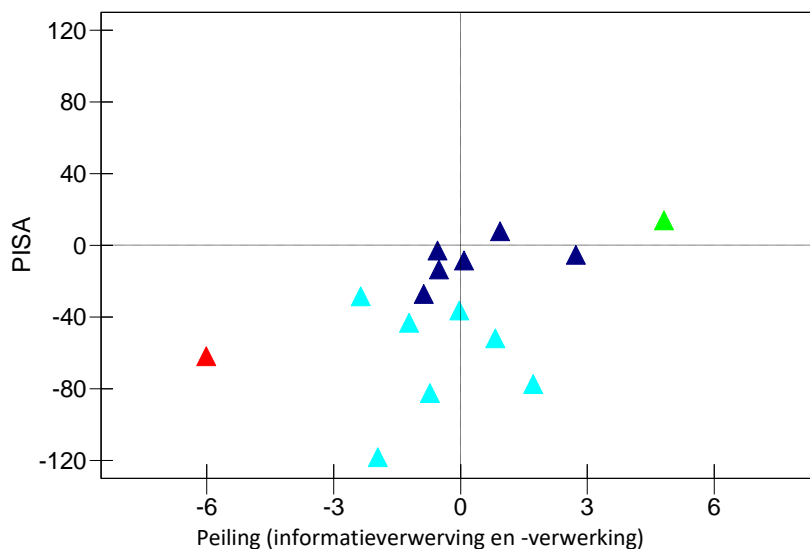
- In de eerste puntgrafiek wordt er een vergelijking gemaakt met alle leerlingen – ongeacht hun onderwijsvorm – in de PISA-dataset.
- In de tweede puntgrafiek wordt er enkel een vergelijking gemaakt bij de leerlingen die de B-stroom of het BSO volgen en deelnamen aan PISA.

In de eerste puntgrafiek zijn er 7 **lichtblauwe** scholen. Dat zijn scholen die een gemiddeld schooleffect hebben op basis van de peiling en een ondergemiddeld schooleffect hebben op basis van PISA. Het is echter niet verrassend dat er heel wat scholen met een B-stroom en/of BSO-studierichtingen ondergemiddeld scoren op PISA. Het gemiddelde van PISA is immers voor alle scholen die deelnemen aan PISA hetzelfde, ongeacht de onderwijsvorm(en) die ze aanbieden. Met andere woorden, die scholen worden ook vergeleken met scholen die enkel ASO en/of A-stroom aanbieden. De resultaten in de eerste puntgrafiek zijn in dat opzicht in lijn met het Vlaamse PISA-rapport dat toont dat leerlingen uit het (D)BSO minder hoog scoren op PISA dan leerlingen uit andere onderwijsvormen (De Meyer et al., 2019).

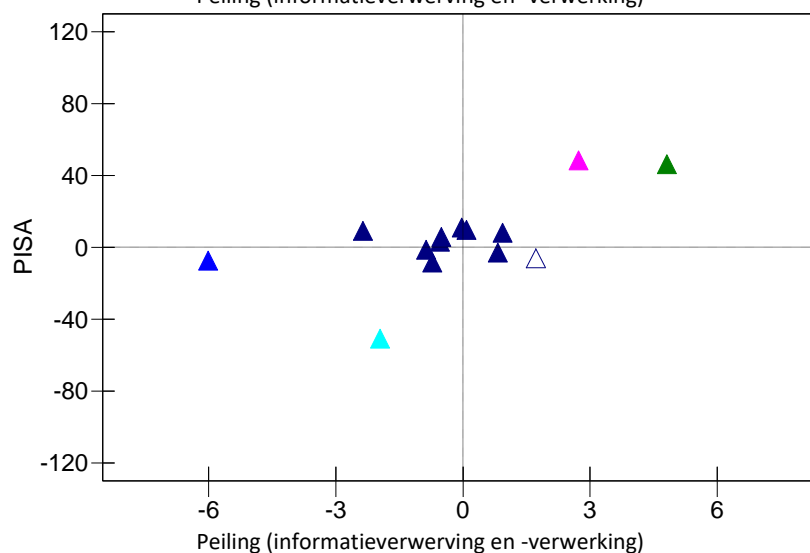
In de tweede grafiek zijn dezelfde scholen opgenomen als in de eerste grafiek. Het enige verschil is dat we het gemiddelde van PISA berekend hebben op basis van de resultaten van leerlingen die B-stroom of BSO volgen en niet op basis van alle leerlingen. In die grafiek vergelijken we de scholen die deelnamen aan PISA én de peiling dus met andere scholen met een B-stroom en/of BSO-studierichtingen die deelnamen aan de peiling enerzijds of PISA 2018 anderzijds. Die puntgrafiek toont een grotere overeenkomst tussen het resultaat van een school op PISA en op de peiling in de B-stroom dan de eerste grafiek. Van de 14 scholen die deelgenomen hebben aan zowel de peiling B-stroom 'informatieverwerking en -verwerking' als PISA zijn er 11 scholen waarbij beide bronnen tot hetzelfde resultaat leiden. Dat zijn de **donkerblauwe** en **donkergroene** scholen.

Hoe het gemiddelde berekend wordt – of anders gezegd, met welke leerlingen en scholen een bepaalde school vergeleken wordt, doet er dus toe. In de eerste puntgrafiek is er voor een relatief groot aantal scholen weinig overeenkomst tussen het resultaat op PISA en de peiling, omdat het gemiddelde bij PISA anders berekend wordt dan bij de peilingen. Wanneer we het gemiddelde op dezelfde manier berekenen bij beide bronnen, is het aantal scholen met een convergerend resultaat groter.

Type 0-schooleffecten
Peiling wiskunde B-stroom
2019 & PISA 2018 (alle
leerlingen)



Type 0-schooleffecten
Peiling wiskunde B-stroom
2019 & PISA 2018 (enkel
leerlingen die B-stroom of
BSO volgen)



▲ (Red)	PISA: Ondergemiddeld – Peiling: Ondergemiddeld
▲ (Cyan)	PISA: Ondergemiddeld – Peiling: Gemiddeld
▲ (Blue)	PISA: Gemiddeld – Peiling: Ondergemiddeld
▲ (Dark Blue)	PISA: Gemiddeld – Peiling: Gemiddeld
▲ (Green)	PISA: Gemiddeld – Peiling: Bovengemiddeld
▲ (Magenta)	PISA: Bovengemiddeld – Peiling: Gemiddeld
▲ (Dark Green)	PISA: Bovengemiddeld – Peiling: Bovengemiddeld
△ (White)	Deze school heeft minder dan 5 leerlingen die B-stroom of BSO volgen.

Figuur 25. Puntgrafieken van de vergelijking tussen de type 0-schooleffecten geschat op basis van de schaal 'informatieverwerving en -verwerking' (peiling wiskunde B-stroom) en PISA 2018²⁹

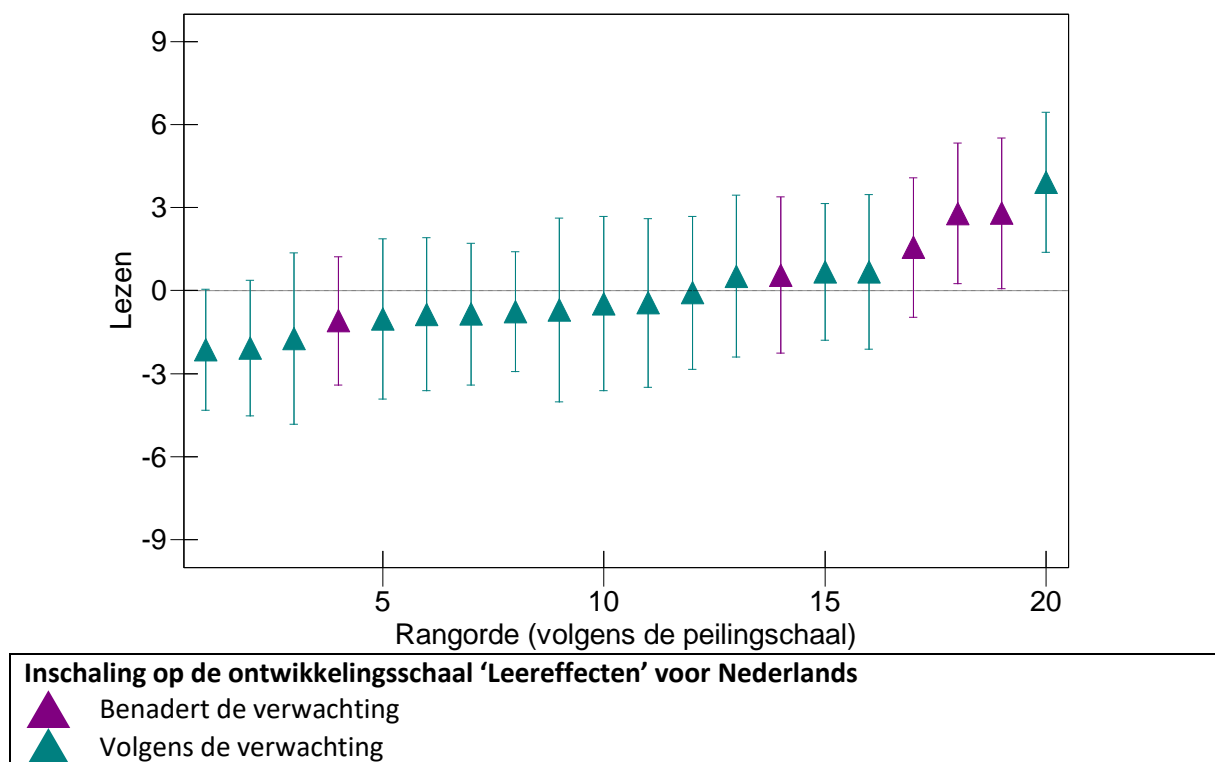
²⁹ Ter verduidelijking, Bijlage 2 – Vergelijking 3 omvat de resultaten van elke vergelijking tussen PISA en de peilingen in de A- en B-stroom, waarin alle deelnemende PISA-scholen en -leerlingen opgenomen worden. In Bijlage 2 – Bijkomende analyses worden de volgende vergelijkingen gemaakt:

- De schoolresultaten van de peiling wiskunde A-stroom worden vergeleken met de schoolresultaten bij PISA. Daarbij wordt er enkel rekening gehouden met de PISA-leerlingen die A-stroom, ASO, TSO of KSO volgen op de school in kwestie.
- De schoolresultaten van de peiling wiskunde B-stroom worden vergeleken met de schoolresultaten bij PISA. Daarbij wordt er enkel rekening gehouden met de PISA-leerlingen die B-stroom of BSO volgen op de school in kwestie.

ZELFS WANNEER MEERDERE BRONNEN TOT GELIJKAARDIGE RESULTATEN LEIDEN BIJ DE MEESTE SCHOLEN, ZIJN ER OOK SCHOLEN WAARBIJ MEERDERE BRONNEN LEIDEN TOT EEN DIVERGENT BEELD (VERGELIJKING 1, 2 & 3).

Hoewel de resultaten over het algemeen aangeven dat meerdere bronnen bij de meeste scholen leiden tot een relatief convergent beeld, zijn er ook steeds scholen waarbij meerdere bronnen leiden tot een eerder divergent beeld. Dat kan om kleine verschillen – net wel of niet significant – gaan zoals bijvoorbeeld bij de vergelijking tussen het peilingonderzoek en het internationaal vergelijkend onderzoek (zie Figuur 24 en bijhorende toelichting). Het kan ook om grotere verschillen gaan.

Zo toont Figuur 26 bijvoorbeeld de vergelijking tussen de peiling lezen (type B-schooleffecten) en de doorlichtingen. In die figuur zijn er twee scholen die volgens de onderwijsinspectie de verwachting benaderen, maar volgens de peiling lezen wel een bovengemiddeld schooleffect hebben. Dat zijn de twee paarse scholen aan de rechterkant van de rupsgrafiek. Voor die scholen is er weinig overeenkomst tussen het resultaat van de school volgens de peiling lezen – een bovengemiddeld type B-schooleffect – enerzijds en het doorlichtingsverslag dat ze ontvingen van de onderwijsinspectie anderzijds – waarin staat dat ze de verwachting benaderen voor het leergebied Nederlands.



Figuur 26. Rangschikking van scholen volgens hun type B-schooleffect op de peilingschaal lezen en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

Bij elke vergelijking zijn er dus scholen waarbij meerdere bronnen leiden tot divergerende resultaten. Nochtans kunnen die divergerende resultaten een grote impact hebben voor de scholen in kwestie. Dat zou bijvoorbeeld het geval kunnen zijn als er gevolgen gekoppeld worden (bijvoorbeeld een begeleidingstraject, een bijkomende doorlichting...) aan het resultaat van een school. Die impact zou bovendien nog groter kunnen worden als de resultaten van slechts één bron bepalend zijn voor de toekenning van dergelijke gevolgen.

EEN ABSOLUUT CRITERIUM VERGELIJKEN MET EEN RELATIEF CRITERIUM: VOLGENS DE VERWACHTING IS NIET AUTOMATISCH HETZELFDE ALS EEN GEMIDDELD SCHOOLEFFECT (VERGELIJKING 1 & 2).

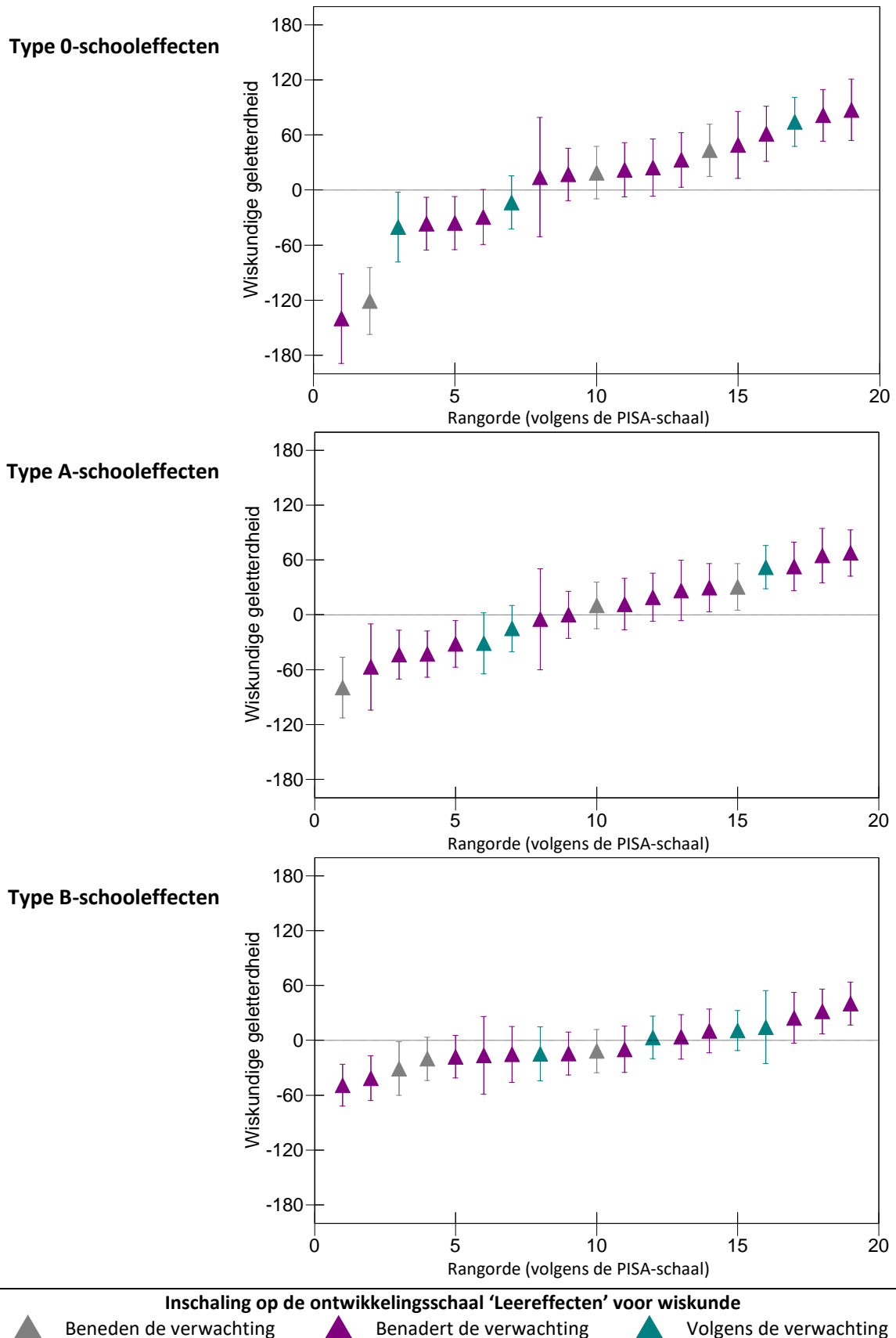
De onderwijsinspectie werkt met een absoluut criterium, namelijk de verwachtingen volgens het OK-kader. Een school die aan die verwachtingen voldoet, krijgt bij de doorlichting minstens de inschaling 'volgens de verwachting'. Scholen die niet aan die verwachtingen voldoen, krijgen de inschaling 'beneden de verwachting' of 'benadert de verwachting'. Echter, het peilingsonderzoek en het internationaal vergelijkend onderzoek maken gebruik van een relatief criterium. Dat wil zeggen dat ze het resultaat van een school steeds vergelijken met het Vlaams gemiddelde.

Hoe het absoluut criterium bij de doorlichtingen – de verwachting – zich juist verhoudt ten opzichte van het relatief criterium bij de verschillende peilingen en internationale onderzoeken – het gemiddelde – is op dit moment onduidelijk. Wel illustreert deze studie dat voldoen aan de verwachting bij de doorlichting niet noodzakelijk overeenkomt met het Vlaams gemiddelde.

Figuur 27 toont bijvoorbeeld de rupsgrafieken van de verschillende types schooleffecten die geschat zijn op basis van 'wiskundige geletterdheid' bij PISA 2018. Opnieuw zijn de **paarse** scholen in de rupsgrafieken de scholen die volgens de onderwijsinspectie de verwachting benaderen. De **blauwe** scholen hebben de inschaling 'volgens de verwachting' behaald bij de doorlichting voor wiskunde. In die rupsgrafieken zijn er ook enkele **grijze** scholen. Zij hebben op de schaal leereffecten de inschaling 'beneden de verwachting' gekregen van de onderwijsinspectie.

Opvallend in die rupsgrafieken is het aantal **grijze** en **paarse** scholen. Van de 19 scholen die doorgelicht zijn voor wiskunde én deelgenomen hebben aan PISA, zijn er 16 scholen die niet aan de verwachting voldoen voor de schaal leereffecten. Slechts drie scholen voldoen aan de verwachtingen volgens de onderwijsinspectie. Ondanks dat de meeste scholen niet aan de verwachting voldoen, zijn er toch nog steeds scholen met een bovengemiddeld schooleffect volgens PISA. De schatting van de schooleffecten is immers een vergelijking met een relatief criterium. Ook al scoren de meeste scholen onder de verwachting, er zijn nog steeds scholen die het beter doen dan het gemiddelde – en dus een bovengemiddeld schooleffect behalen. Zelfs in het scenario waarin geen enkele school voldoet aan de verwachting, is het mogelijk dat bepaalde scholen – die niet voldoen aan de verwachting – toch een bovengemiddeld schooleffect hebben.

Daarbij is het van belang om op te merken dat scholen die deelnemen aan een peilingsonderzoek of internationaal onderzoek geen zicht hebben op die relatie tussen de doorlichtingen en de andere bronnen van externe kwaliteitszorg. Op basis van Figuur 27 is het mogelijk om af te leiden dat de inschaling 'volgens de verwachting' helemaal niet overeenkomt met het Vlaamse gemiddelde volgens PISA. De deelnemende scholen beschikken echter niet over die informatie. Bijvoorbeeld, de meest rechtse school bij de eerste rupsgrafiek ontvangt de inschaling 'benadert de verwachting' voor wiskunde in het doorlichtingsverslag. Op basis van het schoolfeedbackrapport van PISA leidt de school af dat ze een bovengemiddelde schoolprestatie behaald heeft op wiskundige geletterdheid bij PISA. Voor de school in kwestie levert dat een divergerend beeld op. Nochtans is het in deze specifieke situatie niet vreemd om een bovengemiddeld schooleffect te hebben en de verwachting te benaderen. De school in kwestie kan dat echter niet besluiten op basis van de informatie waarover zij beschikt.



Figuur 27. Rupsgrafieken van de vergelijking tussen de schooleffecten geschat op basis van de schaal 'wiskundige geletterdheid' (PISA 2018) en de inschaling van de doorlichtingen op de ontwikkelingsschaal 'leereffecten' bij wiskunde

Studie 3: Vergelijking van trends op systeemniveau

1. Inleiding

De onderwijsinspectie, het internationaal vergelijkend onderzoek en het peilingsonderzoek doen allemaal uitspraken over hoe de kwaliteit van het Vlaamse onderwijs evolueert doorheen de tijd. Zo kunnen we via PIRLS, PISA en TIMSS de trends inzake de gemiddelde score van Vlaanderen – ten opzichte van andere onderwijssystemen – opvolgen. De peilingen beschrijven op hun beurt het aandeel leerlingen dat de onderwijsdoelen bereikt en vergelijken dat aandeel met de voorgaande peilingen. Tot slot doet de onderwijsinspectie via de Onderwijsspiegel elk jaar een uitspraak op systeemniveau, daarnaast is het mogelijk om de beoordelingen in elk doorlichtingsverslag te aggregeren naar het systeemniveau per schooljaar.³⁰

2. Onderzoekopzet

Het doel van de huidige studie is om na te gaan of de bronnen van externe kwaliteitszorg die focussen op een gelijkaardig domein of construct ook dezelfde trends op systeemniveau vinden. In deze studie maken we daarvoor gebruik van de volgende bronnen:

- peiling basisonderwijs Nederlands 2007, 2013 en 2018;
- peiling basisonderwijs wiskunde 2009, 2016 en 2021;
- peiling wiskunde A-stroom 2009 en 2018;
- peiling wiskunde B-stroom 2008 en 2019;
- PIRLS 2006 en 2016;
- TIMSS 2011, 2015 en 2019;
- PISA 2009, 2012, 2015 en 2018;
- doorlichtingen volgens de doorlichtingsronde 3-aanpak van schooljaar 2013-2014 tot en met schooljaar 2016-2017;
- doorlichtingen volgens de Inspectie 2.0-aanpak in schooljaren 2018-2019, 2019-2020 en 2021-2022.

In wat volgt, wordt er eerst gefocust op hoe we de trends op systeemniveau beschrijven bij de doorlichtingen, het internationaal vergelijkend onderzoek en het peilingsonderzoek. Nadien lichten we de concrete aanpak toe die we gebruiken om die trends op systeemniveau te vergelijken.

Om de leesbaarheid van de studie te verhogen, geven we hieronder de technische informatie over de berekening van trends steeds weer in grijze kaders. De lezers die minder geïnteresseerd zijn in die specifieke informatie kunnen de grijze kaders overslaan.

2.1 Doorlichtingen van de onderwijsinspectie

2.1.1 Doorlichtingsronde 3 en Inspectie 2.0

Sinds 2018 worden scholen doorgelicht volgens de doorlichtingsmethode Inspectie 2.0.³¹ In de voorgaande schooljaren werden scholen doorgelicht volgens een andere doorlichtingsmethode,

³⁰ In deze studie maken wij gebruik van die geaggregeerde databestanden en niet van de resultaten uit de Onderwijsspiegel. Kleine verschillen met de Onderwijsspiegel zijn dus mogelijk.

³¹ Ter informatie, de doorlichtingsmethode van Inspectie 2.0 wordt uitgebreid beschreven in Bijlage 1.

namelijk de methode van doorlichtingsronde 3. Er zijn enkele opvallende verschillen tussen die twee doorlichtingsmethodes die van belang zijn voor de huidige studie:

- **Versillen in beoordeelde proceskenmerken**

In tegenstelling tot Inspectie 2.0 is de aanpak van doorlichtingsronde 3 niet gebaseerd op het OK-kader, maar wel op een ander kader: het CIPO-kader³². Dat heeft als gevolg dat de proceskenmerken die de onderwijsinspectie beoordeelde tijdens doorlichtingsronde 3 niet automatisch ook beoordeeld worden bij Inspectie 2.0 (Onderwijsinspectie, 2018a, 2018b). Om trends in kaart te kunnen brengen, is het nochtans essentieel dat bij beide doorlichtingsmethodes een bepaald construct op dezelfde manier gemeten en beoordeeld wordt.

Na overleg met de onderwijsinspectie stelden we vast dat er twee constructen zijn die beoordeeld werden bij zowel doorlichtingsronde 3 als Inspectie 2.0. Concreet gaat het om de volgende proceskenmerken:

- De afstemming van het onderwijsaanbod op de doelen werd in doorlichtingsronde 3 opgenomen als het proceskenmerk ‘onderwijsaanbod’ en in Inspectie 2.0 als de U1- of O1-schaal³³ ‘afstemming van het aanbod op het gevalideerd doelenkader’.
- Evaluatie werd ook beoordeeld bij beide doorlichtingsmethodes. In doorlichtingsronde 3 beoordeelde de onderwijsinspectie het proceskenmerk ‘evaluatiepraktijk’. Bij Inspectie 2.0 worden scholen doorgelicht voor de U6- of O7-schaal³⁴ ‘leerlingenevaluatie’. Hoewel de essentie van de schaal ‘evaluatiepraktijk’ en ‘leerlingenevaluatie’ hetzelfde gebleven is, zijn er volgens de onderwijsinspectie doorheen de tijd meer verwachtingen gekomen met betrekking tot evaluatie. In dat opzicht moeten we voorzichtig zijn bij de interpretatie van de trends van die kenmerken.

- **Versillend aantal inschalingen per proceskenmerk**

Bij doorlichtingsronde 3 werd er gewerkt met kwaliteitsaspecten (zie Figuur 28). Meer specifiek werd een school bij ieder proceskenmerk beoordeeld voor elk van de vier kwaliteitsaspecten. Dat is niet het geval bij Inspectie 2.0. Bijgevolg zijn er dus vier inschalingen – op de vier kwaliteitsaspecten – voor elk proceskenmerk bij doorlichtingsronde 3 en slechts één inschaling per proceskenmerk bij Inspectie 2.0 (Onderwijsinspectie, 2018a, 2018b).

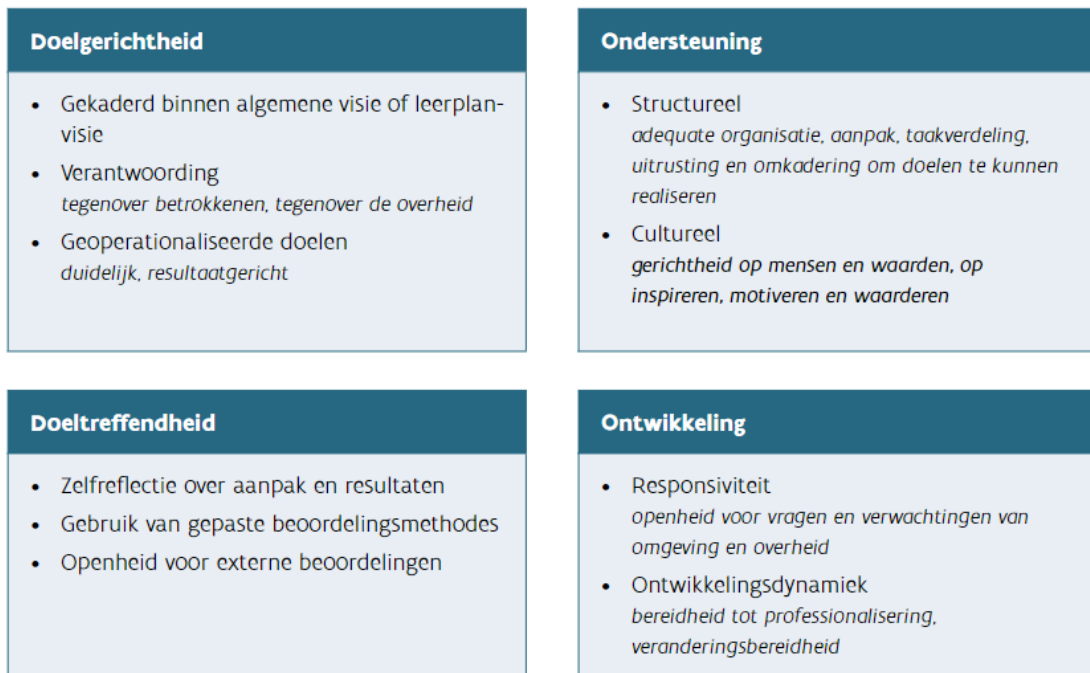
³² CIPO is een acroniem voor context, input, proces en output. Meer informatie over de overgang van het CIPO-naar het OK-kader is te vinden bij Deelrapport 1 van het OBPWO-project.

³³ Sinds het schooljaar 2021-2022 wordt de onderwijsleerpraktijk ingeschaald aan de hand van O-schalen in plaats van U-schalen (Onderwijsinspectie, 2021f):

De U1-schaal werd de O1-schaal ‘afstemming van het aanbod op het gevalideerd doelenkader’.

De U6-schaal werd de O7-schaal ‘leerlingenevaluatie’.

³⁴ Idem.



Figuur 28. Kwaliteitsaspecten op basis waarvan de onderwijsinspectie de proceskenmerken beoordeelde in doorlichtingsronde 3 (Figuur overgenomen uit Onderwijsinspectie, 2018b)

- **Verskillende beoordelingsniveaus**

Bij zowel doorlichtingsronde 3 als Inspectie 2.0 zijn er vier beoordelingsniveaus. Echter, bij doorlichtingsronde 3 wordt een school geclassificeerd als (1) zwak, (2) eerder zwak dan sterk, (3) eerder sterk dan zwak of (4) sterk. Bij Inspectie 2.0 daarentegen zijn er vier ontwikkelingsniveaus, namelijk (1) beneden de verwachting, (2) benadert de verwachting, (3) volgens de verwachting en (4) overstijgt de verwachting (Onderwijsinspectie, 2018a, 2018b).

2.1.2 Aanpak in deze studie

Omwille van bovenstaande verschillen tussen doorlichtingsronde 3 en Inspectie 2.0 is het niet eenvoudig om trends voor de doorlichtingen in kaart te brengen. Daarom kiezen we ervoor om de trends voor doorlichtingsronde 3 en Inspectie 2.0 weer te geven in twee aparte figuren. Concreet geven we in die figuren per schooljaar het percentage scholen weer dat een specifieke inschaling van de onderwijsinspectie ontvangen heeft. Echter, binnen eenzelfde doorlichtingsmethode kan de onderwijsinspectie de inhoudelijke omschrijving van de proceskenmerken (doorlichtingsronde 3) en/of de kritische kenmerken die horen bij de ontwikkelingsschalen (Inspectie 2.0) bijsturen en aanpassen. De figuren van de trends zijn bijgevolg louter indicatief en moeten met enige voorzichtigheid geïnterpreteerd worden.

Bovendien is het niet mogelijk om de trends van de doorlichtingen te vergelijken met trends van het peilingsonderzoek of het internationaal vergelijkend onderzoek. Gelijkaardige proceskenmerken als (1) onderwijsaanbod/afstemming van het aanbod op het gevalideerd doelenkader en (2) evaluatiepraktijk/leerlingenevaluatie worden immers niet steeds op dezelfde manier gemeten bij PIRLS, PISA, TIMSS en de peilingen. Anders gezegd, er zijn gewoonweg geen trends van gelijkaardige proceskenmerken bij de peilingen en de internationale onderzoeken waarmee we de trends van de doorlichtingen zouden kunnen vergelijken.

2.2 Trends vergelijken tussen het internationaal vergelijkend onderzoek en het peilingsonderzoek

2.2.1 Trends bij PIRLS, PISA en TIMSS beschrijven

Aangezien de internationaal vergelijkende onderzoeken herhaaldelijk plaatsvinden, is het mogelijk om trends in gemiddelde leerlingprestaties over de afnames heen te beschrijven. Bij elk van die bronnen worden de toetscores van een nieuwe afnamecyclus immers op dezelfde meetschaal geplaatst als de voorgaande afnamecycli (zie Bijlage 1 voor meer gedetailleerde informatie daarover). Zodra de data van verschillende cycli op dezelfde schaal staan, is het mogelijk om trends in kaart te brengen (Martin et al., 2020; Mullis & Martin, 2015; OECD, 2019a). Afhankelijk van het aantal afnamemomenten hanteren we daarbij een andere methode:

- Bij twee afnames beschrijven we de trends aan de hand van verschilcores.
- Bij meer dan twee afnames beschrijven we de trends aan de hand van lineaire trends.

In wat volgt, lichten we beide methodes kort toe.

2.2.1.1 Trends op basis van twee afnames: verschilcores

Wanneer er twee afnamemomenten zijn – dat is het geval bij PIRLS – kan de trend beschreven worden door het verschil tussen de gemiddelde leerlingprestaties op die twee specifieke afnamemomenten te berekenen (Martin et al., 2020; OECD, 2019b). De schatting van zo'n trend heeft – net als elke vaardigheidsschatting – echter een mate van statistische onzekerheid die wordt uitgedrukt aan de hand van een standaardfout. Ook die standaardfout rapporteren we bij de resultatensectie verderop.

Hoe wordt de standaardfout van een trend (bij verschilcores) berekend?

De standaardfout wordt aan de hand van de volgende formule berekend³⁵ (Martin et al., 2020):

$$SE(t_A - t_B) = \sqrt{SE(t_A)^2 + SE(t_B)^2}$$

Bovenstaande formule beschrijft dat de standaardfout van het verschil tussen de gemiddelde leerlingprestaties van een onderwijssysteem over twee cycli heen gelijk is aan de vierkantswortel van de som van de gekwadrateerde standaardfouten van de twee gemiddelde resultaten.

Voor die berekening hebben we dus enkel de standaardfouten nodig van de gemiddelde leerlingprestaties voor Vlaanderen op twee afnamemomenten zoals ze gerapporteerd zijn door de IEA.

2.2.1.2 Trends op basis van meer dan twee afnames: lineaire trends

Vlaanderen heeft al meer dan twee keer deelgenomen aan TIMSS en PISA. Daardoor is het mogelijk om de trends voor die bronnen aan de hand van meer dan twee afnames te beschrijven (Martin et al., 2017, 2020; OECD, 2019a). Door de trends aan de hand van drie of meer afnames te schatten, is het mogelijk om de statistische onzekerheid te verminderen. De schatting van de trend is namelijk minder gevoelig voor toevallige afwijkingen in gemiddelde prestaties op één bepaald afnamemoment wanneer we gebruik maken van de gemiddelde prestaties op meerdere afnamemomenten.

³⁵ Als men bij twee afnamemomenten van een bepaald onderzoek steeds dezelfde items zou hanteren en de echte itemparameters zou gebruiken, dan kan de standaardfout van de trend aan de hand van de bijhorende formule berekend worden. Er worden echter niet steeds dezelfde items gehanteerd en de echte itemparameters zijn onbekend. Dat brengt heel wat onzekerheid met zich mee, die uitgedrukt wordt in *equating error* of *link(ing) error*. Meer informatie over de *equating error* volgt verderop.

Daarbij moeten we echter een veronderstelling maken over de vorm van de trend. In deze studie veronderstellen we dat de relatie tussen de gemiddelde leerlingprestaties van een onderwijssysteem en de afnamemomenten een rechte lijn is. We beperken ons tot lineaire trends (OECD, 2019b), omdat we nog meer afnamemomenten nodig zouden hebben om complexere vormen van trends te schatten.

Hoe wordt de lineaire trend berekend?

We veronderstellen we dat de relatie tussen de Vlaamse gemiddeldes (op TIMSS of PISA) en de afnamemomenten een rechte lijn is. In de statistiek wordt die rechte traditioneel beschreven door middel van een lineaire regressievergelijking met een intercept en een hellingparameter (Kutner et al., 2005; OECD, 2019b):

$$P_t = \beta_0 + \beta_1 \text{tijd}_t + \varepsilon_t$$

Bovenstaande formule beschrijft dat de gemiddelde prestatie voor Vlaanderen op een bepaald tijdstip (P_t) gelijk is aan de som van het intercept (β_0), het product van de helling (β_1) vermenigvuldigd met het afnamemoment (tijd_t) en het residu (ε_t). Het intercept kan men interpreteren als de gemiddelde prestatie op tijdstip 0 en de helling beschrijft de trend waarin we voornamelijk geïnteresseerd zijn.

Bij de berekening van de lineaire trend op basis van verschillende TIMSS- of PISA-afnames gebruiken we bovenstaande formule om de trend – dat is de helling in de formule – te schatten.

Die helling is een statistische schatting en wordt bijgevolg gekenmerkt door een mate van statistische onzekerheid. Net zoals bij de verschildscores wordt die statistische onzekerheid uitgedrukt aan de hand van een standaardfout. De berekening van die standaardfout gebeurt echter op een andere manier bij TIMSS en PISA. De IEA houdt geen rekening met de zogenaamde *equating error*, terwijl de OESO dat wel doet (Martin et al., 2020; OECD, 2019b).

Wat is equating error?

Bij de afname van meerdere cycli hanteert elk internationaal vergelijkend onderzoek de volgende aanpak: bij elke cyclus van een internationaal vergelijkend onderzoek worden (1) oude items verwijderd, (2) nieuwe items toegevoegd, (3) items uit de vorige cyclus behouden waarvoor er nieuwe itemparameters geschat worden. Die derde groep items worden de trenditems genoemd. Die items worden gebruikt om de toetsscores uit verschillende cycli (bijvoorbeeld PISA 2015 en PISA 2018) op eenzelfde meetschaal te plaatsen.

Die aanpak (meer specifiek: de wijziging van de items die opgenomen worden in de toets én de nieuwe schatting van de itemparameters bij de trenditems) brengt ook heel wat onzekerheid met zich mee, die uitgedrukt wordt in *equating error* of *link(ing) error* (e.g. Wu, 2010). Die *equating error* heeft volgens de IEA en OESO voornamelijk de volgende twee oorzaken:

- De echte parameterwaarden van de trenditems zijn niet gekend. De meetschaal wordt bepaald aan de hand van geschatte parameterwaarden die elk een mate van onzekerheid hebben. Zo werd in PISA 2018 aangegeven dat alle geschatte parameterwaarden van de trenditems uit PISA 2015 tot op zekere hoogte onvolmaakt waren in 2018 (OECD, 2021). Bij TIMSS 2019 (en PIRLS 2016) bleek dan weer dat het lineaire verband tussen de initiële parameters en nieuwe parameters van de trenditems niet exact hetzelfde was. Zowel bij

PISA, TIMSS als PIRLS zijn er dus verschillen tussen de nieuwe en de oude parameterschattingen van trenditems, wat een bron van *equating error* is.

- Er is slechts een beperkt aantal trenditems om de toetscores over twee cycli heen op eenzelfde meetschaal te plaatsen. Een kleiner aantal trenditems – waarbij de parameterschattingen een schattingsfout hebben – leidt tot een grotere *equating error*.

De OESO en de IEA erkennen bovenstaande oorzaken van *equating error* voor trends op landniveau. Ze schatten daarom de *equating error* door te berekenen hoe sterk de gemiddelde vaardigheden van onderwijssystemen verschillen als (1) de parameters van de trenditems uit een oude cyclus gebruikt worden of (2) de parameters van de trenditems van een nieuwe cyclus gebruikt worden.

Toch zijn er ook enkele verschillen tussen PISA enerzijds en PIRLS en TIMSS anderzijds wat betreft de schatting van de *equating error*:

- **Equating error bij PISA**

De OESO schat de *equating error* tussen de verschillende afnamecycli bij elke afnamecyclus opnieuw en rapporteert daarover in de internationale rapporten (OECD, 2019b). Meer specifiek berekent de OESO de *equating error* in PISA over alle onderwijssystemen heen. Er is dus één schatting van de *equating error* tussen meerdere afnamemomenten voor een bepaald domein. Zo is de *equating error* voor een trend in wiskundige geletterdheid op basis van PISA 2009, PISA 2012, PISA 2015 en PISA 2018 gelijk aan 1,2885 (OECD, 2019b).

- **Equating error bij TIMSS en PIRLS**

Tot voor kort werden er standaard bij TIMSS geen *equating errors* berekend. Enkel bij TIMSS 2019 werden er *equating errors* berekend tussen de cycli van TIMSS 2015 en 2019. In tegenstelling tot bij PISA werden daarbij de *equating errors* voor elk onderwijssysteem apart berekend. Voor Vlaanderen werd er een *equating error* van 0,6 berekend tussen de afnamecycli van 2019 en 2015 (Martin et al., 2020). Voor eerdere afnamemomenten werden er voor TIMSS geen *equating errors* berekend.

Bij PIRLS werden er tot nu toe nog geen *equating errors* gerapporteerd in de technische rapporten. Onderzoekers van de IEA hebben de *equating errors* wel onderzocht tussen de PIRLS 2001 en PIRLS 2006. Ook hier werden de *equating errors* voor ieder onderwijssysteem apart berekend, maar Vlaanderen was geen onderdeel van die analyse (Martin et al., 2012).

De IEA berekent de *equating errors* niet standaard bij haar studies, omdat de *equating errors* – volgens het onderzoek van de IEA – een verwaarloosbare rol hebben bij de berekening van standaardfouten bij trends. Uit de IEA-rapporten blijkt inderdaad dat de gemiddelde vaardigheid van Vlaanderen op één afnamemoment stabiel is over twee kalibraties heen en dat de *equating error* dus beperkt is (Martin et al., 2020).

In deze studie kiezen we ervoor om bij de berekening van de trends de aanpak van IEA en OESO te volgen. Dat betekent concreet het volgende:

- Bij TIMSS houden we – net zoals de IEA – geen rekening met de *equating error* bij de berekening van de standaardfout.
- Bij de schatting van de standaardfout houden we bij PISA – in lijn met de aanpak van de OESO – wel rekening met de *equating error* (OECD, 2019b).

Hoe wordt de standaardfout van de helling (bij lineaire trends) berekend?

De standaardfout van de helling wordt bij TIMSS en PISA berekend aan de hand van de volgende formules:

- **Berekening van de standaardfout bij TIMSS**

Aangezien we geen rekening houden met de *equating error*, wordt de onzekerheid van helling (β_1) berekend aan de hand van de volgende formule (Kutner et al., 2005):

$$SE(\beta_1) = \sqrt{\frac{\sum_t (P_t - \hat{P}_t)^2}{(n - 2) * \sum_t (tijd_t - \overline{tijd})^2}}$$

De formule hierboven beschrijft met hoeveel statistische onzekerheid het lineaire regressiemodel de onbekende waarde van de helling schat. De statistische onzekerheid is hier een functie van de geobserveerde gemiddelde prestaties, de voorspelde gemiddelde prestaties en het aantal tijdstippen waarop de helling werd berekend. Voor de interpretatie is hoofdzakelijk de teller van belang. In de teller staat namelijk het gekwadrateerde verschil tussen de geobserveerde gemiddelde prestaties en de voorspelde gemiddelde prestaties over de verschillende afnamemomenten centraal. Indien het lineaire regressiemodel een perfecte voorspelling zou geven, dan zou er geen verschil zijn tussen de geobserveerde en de voorspelde gemiddelde prestaties. De teller zou dan nul zijn, en de statistische onzekerheid zou bijgevolg ook nul zijn. Omgekeerd, hoe de groter verschillen tussen de geobserveerde en de voorspelde gemiddelde prestaties in de teller, hoe het groter de statistische onzekerheid. De logica achter de statistisch onzekerheid van de helling is dus relatief doorzichtig.

- **Berekening van de standaardfout bij PISA**

Bij PISA gebruiken we de volgende formule om de standaardfout van de helling te berekenen:

$$SE(\beta_1) = \sqrt{\sigma_s^2(\beta_1) + \sigma_e^2(\beta_1)}$$

De berekening bestaat nu uit de vierkantswortel van de som van twee bronnen van statistische onzekerheid. De eerste bron van statistische onzekerheid is gewoonweg de gekwadrateerde onzekerheid van de helling, berekend volgens de eerder beschreven formule voor TIMSS. De tweede bron van statistische onzekerheid is dan de *equating error* over meerdere afnamemomenten heen. Die *equating error* werd echter al geschat en gerapporteerd door de OESO. Voor bovenstaande berekening kunnen we dus gebruik maken van de schatting van de *equating error* zoals ze gerapporteerd is door de OESO (OECD, 2019b).

2.2.2 Trends bij het peilingsonderzoek beschrijven

Ook de herhalingspeilingen laten toe om het aandeel leerlingen dat de onderwijsdoelen bereikt doorheen de tijd in kaart te brengen. Echter, die trends in het aandeel leerlingen dat de onderwijsdoelen bereikt, kunnen we niet vergelijken met de trends bij het internationaal vergelijkend onderzoek. Daarom beschrijven we in de huidige studie ook bij de peilingen de trends in gemiddelde leerlingprestaties doorheen de tijd.

Meer specifiek focussen we – net zoals bij Studie 2 – op de volgende peilingen en bijhorende schalen:

- Wiskunde basisonderwijs Functies
 Problemen oplossen bij getallen en bewerkingen
 Problemen oplossen bij meten, ruimte en meetkunde
- Nederlands basisonderwijs Lezen
- Wiskunde A-stroom Algebraïsering
 Evenredigheden
 Omgaan met data
 Meetkundige procedures: constructies
- Wiskunde B-stroom Meetkunde
 Informatieverwerking en -verwerking
 Meten

In tegenstelling tot bij Studie 2 worden (1) de schaal ‘omtrek, oppervlakte en inhoud’ uit de peiling wiskunde basisonderwijs³⁶ en (2) de schaal ‘getalinzicht en hoofdbewerkingen’ uit de peiling wiskunde B-stroom³⁷ niet meegenomen, aangezien het niet mogelijk is om voor die schalen trends in kaart te brengen.

Om trends in kaart te kunnen brengen, is het noodzakelijk dat de leerlingprestaties op dezelfde schaal staan. Voor de peilingen in de A- en B-stroom was dat reeds het geval, voor de peilingen in het basisonderwijs nog niet. Daarom hebben we bij zowel de peiling Nederlands basisonderwijs als de peiling wiskunde basisonderwijs per schaal de data van verschillende afnamemomenten op dezelfde meetschaal geplaatst.

Vervolgens beschrijven we de trends voor elk van bovenstaande schalen op dezelfde manier als de trends van het internationaal vergelijkend onderzoek. Met andere woorden, we maken – afhankelijk van het aantal afnamemomenten – gebruik van verschillcores of lineaire trends (zie sectie 2.2.1.1 en 2.2.1.2).³⁸ In lijn met de aanpak van STEP houden we bij de schatting van de trends geen rekening met de *equating error*.

2.2.3 Aanpak in deze studie

Concreet vergelijken we in de huidige studie de trends van (1) PIRLS en de peiling Nederlands basisonderwijs, (2) TIMSS en de peiling wiskunde basisonderwijs en (3) PISA en de peilingen wiskunde in de A- en B-stroom.

³⁶ De schaal ‘omtrek, oppervlakte en inhoud’ werd afgenomen in 2009 en 2021 en niet in 2016. We beschikken niet over een databestand waarbij de data uit 2009 en 2021 op een meetschaal staan en we kunnen die data niet zelf op een meetschaal plaatsen. Bijgevolg is het niet mogelijk om trends te beschrijven voor die schaal.

³⁷ De schaal ‘getalinzicht en hoofdbewerkingen’ werd anders samengesteld bij de peiling uit 2019 dan bij de peiling uit 2008. In 2008 bestond de schaal uit drie toetsonderdelen: (1) getalinzicht, (2) hoofdbewerkingen en (3) breuken optellen en aftrekken. In 2019 bestond de schaal ‘getalinzicht en hoofdbewerkingen’ enkel uit de eerste twee toetsonderdelen, namelijk (1) getalinzicht en (2) hoofdbewerkingen (Carpentier et al., 2020b; Centrum voor Onderwijseffectiviteit en -evaluatie, 2009). Bijgevolg kunnen we voor die schaal geen trends beschrijven.

³⁸ Meer specifiek hebben we de trends aan de hand van verschillcores geschat bij de toetsen van de peiling wiskunde A-stroom en B-stroom. Daarnaast schatten we lineaire trends voor de toets lezen van de peiling Nederlands basisonderwijs en de toetsen van de peiling wiskunde basisonderwijs.

Bij elk van die drie vergelijkingen doorlopen we telkens de volgende drie stappen:

- **Trends beschrijven**

Eerst beschrijven we de trend voor het Vlaamse onderwijs door het verschil tussen het gemiddelde op twee afnamemomenten te berekenen (i.e., de verschillscores) of een lineaire trends te schatten op basis van het gemiddelde bij drie of meer afnamemomenten.

Vervolgens berekenen we de standaardfout van de trend. Daarbij kiezen we ervoor om bij de berekening van de trends de aanpak van IEA, OESO en STEP te volgen. Dat betekent concreet het volgende: bij PIRLS, TIMSS en de peilingen houden we geen rekening met de *equating error*, bij PISA daarentegen wordt er wel rekening gehouden met de *equating error* (Martin et al., 2020; OECD, 2019b).

- **Herschaling**

Om de trends van het internationaal vergelijkend onderzoek en het peilingsonderzoek te kunnen vergelijken, herschalen we de trend zodat bij elke bron het gemiddelde van het eerste afnamemoment 0 bedraagt en de standaarddeviatie 1 is. De exacte lineaire transformatie die daarvoor nodig is, passen we vervolgens toe op het gemiddelde en de standaarddeviatie van de latere afnamemomenten bij die trend. Die herschaling laat ons toe om later de trends van het internationaal vergelijkend onderzoek en de peilingen met elkaar te vergelijken.

- **Vergelijking van trends**

Tot slot berekenen we voor zowel de trends bij de peilingen als de trends bij het internationaal vergelijkend onderzoek de gemiddelde jaarlijkse trend. Om die manier kunnen we de trend van bijvoorbeeld de peiling wiskunde – die gebaseerd is op afnames in 2009, 2016 en 2021 – vergelijken met de trend van TIMSS – die gebaseerd is op TIMSS 2011, 2015 en 2019.

Hoewel we ons in deze studie baseren op wat de IEA en OESO zelf rapporteren, moeten we voor de volledigheid de volgende twee opmerkingen maken. Ten eerste, de IEA en de OESO onderschatten waarschijnlijk de *equating error* bij internationaal vergelijkende onderzoeken.

PIRLS, PISA en TIMSS houden immers geen rekening met de instabiliteit van de parameterschattingen tussen landen en over de afnamecycli binnen landen (Glas & Jehangir, 2013; Robitzsch & Lüdtke, 2019; Sachse & Haag, 2017). De vergelijking van geschatte trends in verschillende onderwijssystemen heeft daarom een hogere onzekerheid dan traditioneel gedacht wordt. Als het doel is om de trend van slechts één onderwijssysteem te schatten, dan kan die onzekerheid wel verminderd worden door aparte itemparameters te schatten voor dat specifieke onderwijssysteem (Gebhardt & Adams, 2007; Sachse et al., 2016).

Ten tweede, de beschrijving van trends in gemiddelde leerlingprestaties aan de hand van PISA, TIMSS of PIRLS veronderstelt dat de verschillende cycli exact hetzelfde construct meten (e.g. Byrne & De Vijver, 2010; Kane, 2017). Echter, die internationale studies passen hun constructen aan over verschillende cycli heen (zie ook Bijlage 1). Zo wordt in PISA een hoofddomein bij elke afname herbekeken en indien nodig geherdefinieerd (OECD, 2019a). Ook bij TIMSS en PIRLS is er elke afnamecyclus een nieuwe analyse van de curricula van de deelnemende landen (Mullis & Martin, 2015, 2017). Trends op basis van verschillende cycli van PISA, TIMSS of PIRLS zijn dus mogelijk (deels) toe te schrijven aan veranderingen in het gemeten construct.

3. De doorlichtingen doorheen de tijd

Net als bij Studie 2 geven we de uitgebreide resultaten weer in Bijlage 3 (p. 315). Hieronder presenteren we enkel de hoofdbevindingen. Omwille van de eerder aangehaalde verschillen tussen doorlichtingsronde 3 en Inspectie 2.0 omvat Bijlage 3 steeds twee aparte figuren voor de trends bij doorlichtingsronde 3 enerzijds en Inspectie 2.0 anderzijds. Concreet geven we in die figuren per schooljaar het percentage scholen weer dat een specifieke inschaling van de onderwijsinspectie ontvangen heeft. Zoals eerder vermeld, zijn de figuren van de trends bij de doorlichting louter indicatief en moeten ze met enige voorzichtigheid geïnterpreteerd worden.

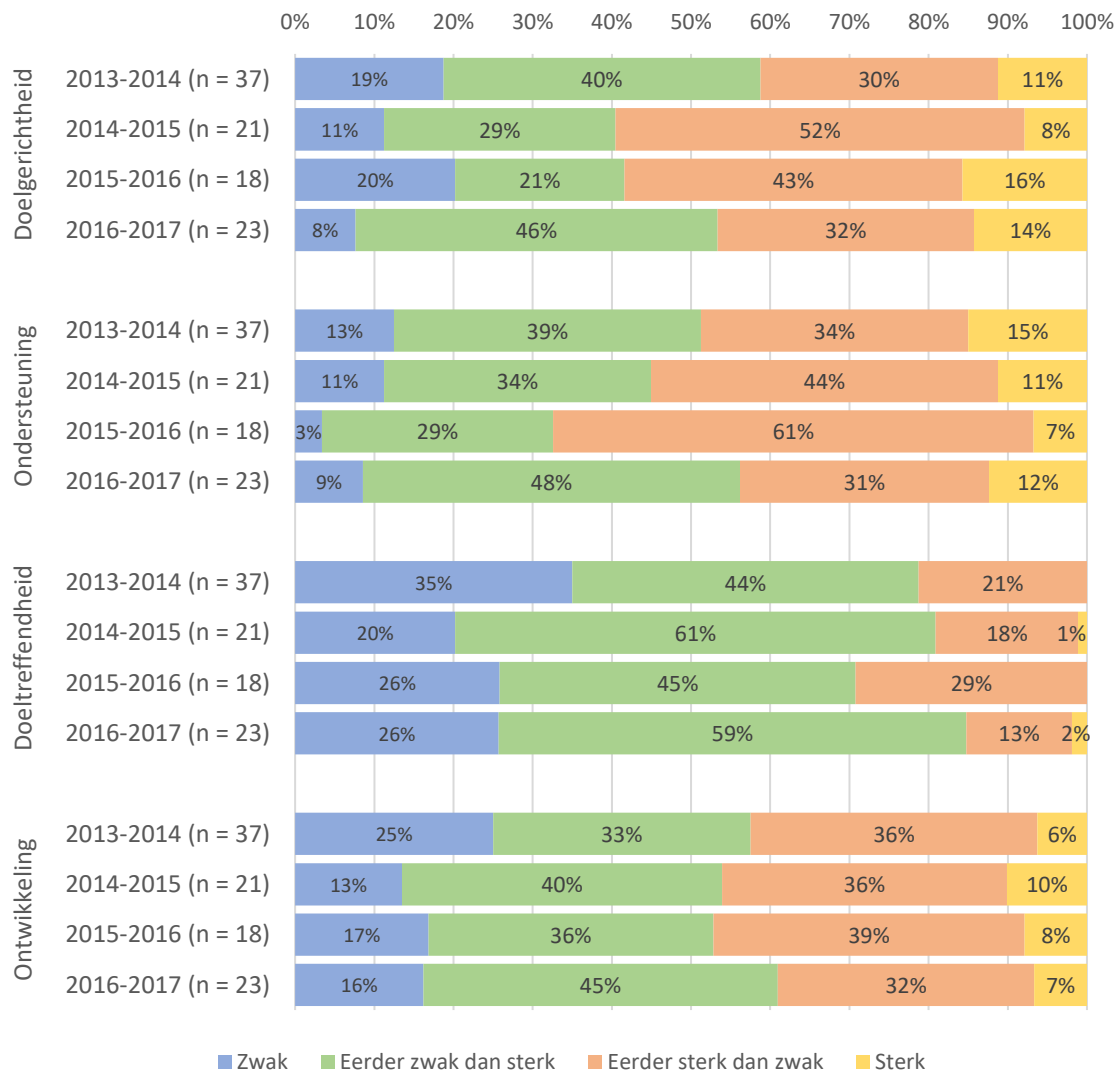
Hoewel een vergelijking met de peilingen, PIRLS, PISA en TIMSS niet mogelijk is, is het ook interessant om naar de trends bij de doorlichtingen op zich te kijken. In wat volgt, illustreren we de voornaamste bevindingen aan de hand van de resultaten voor evaluatie in het secundair onderwijs.

BIJ ZOWEL DOORLICHTINGSRONDE 3 ALS BIJ INSPECTIE 2.0 IS ER RELATIEF VEEL VARIATIE QUA INSCHALINGEN OVER VERSCHILLENDE SCHOOLJAREN HEEN

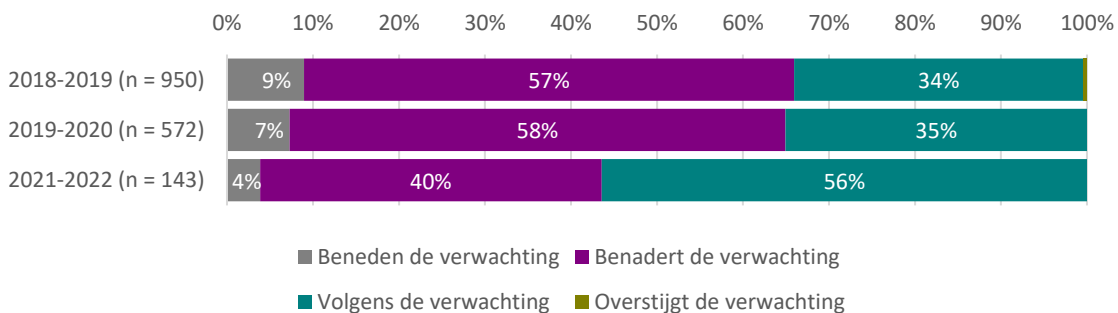
Hoewel er vier inschalingen – op de vier kwaliteitsaspecten – zijn voor elk proceskenmerk bij doorlichtingsronde 3 en slechts één inschaling per proceskenmerk bij Inspectie 2.0, is er een opvallende gelijkens. Bij zowel doorlichtingsronde 3 als Inspectie 2.0 varieert het percentage scholen dat een specifieke inschaling krijgt over schooljaren heen.

Figuur 29 toont het percentage scholen dat een specifieke inschaling van de onderwijsinspectie ontvangen heeft voor het proceskenmerk ‘evaluatiepraktijk’ tijdens doorlichtingsronde 3. Wanneer we het percentage scholen met een bepaald beoordelingsniveau vergelijken van schooljaar 2013-2014 tot en met schooljaar 2016-2017 stellen we heel wat variatie vast. Bijvoorbeeld, het percentage scholen dat na een doorlichting de inschaling **zwak** ontving, varieert in die periode tussen 3 en 35 procent afhankelijk van het kwaliteitsaspect. Bij het beoordelingsniveau **eerder zwak dan sterk** varieert het percentage in diezelfde periode tussen de 21 en 61 procent. Het percentage **eerder sterke dan zwakke** scholen varieert in die periode tussen de 13 en 61 procent en het percentage **sterke** scholen varieert dan weer tussen 0 en 16 procent.

Ook bij de inschaling van de onderwijsinspectie bij de U6- en O7-schaal ‘leerlingenevaluatie’ uit Inspectie 2.0 stellen we variatie vast – weliswaar minder dan bij de schaal ‘evaluatiepraktijk’ uit doorlichtingsronde 3. Uit Figuur 30 blijkt dat tussen schooljaar 2018-2019 en 2021-2022 het percentage scholen dat **beneden de verwachting** scoort, varieert tussen 4 en 9 procent. In diezelfde periode **benadert** afhankelijk van het schooljaar zo’n 40 tot 58 procent van de scholen **de verwachting**. Het percentage scholen dat aan **de verwachting voldoet**, varieert tussen de 34 en 56 procent. Opvallend is wel dat het percentage scholen dat **de verwachting overstijgt** in die periode steeds 0 bedraagt.



Figuur 29. Overzicht van de inschaling op het proceskenmerk 'evaluatiepraktijk' in het secundair onderwijs per schooljaar (doorlichtingsronde 3)



Figuur 30. Overzicht van de inschaling op de U6- en O7-schaal 'Leerlingenevaluatie' in het secundair onderwijs per schooljaar (Inspectie 2.0)

BIJ DOORLICHTINGSRONDE 3 WORDEN DE UITERSTE BEOORDELINGSCATEGORIEËN VAKER GEBRUIKT DAN BIJ INSPECTIE 2.0

Zoals eerder toegelicht (zie 2.1.1 Doorlichtingsronde 3 en Inspectie 2.0), hanteerde de onderwijsinspectie tijdens doorlichtingsronde 3 andere beoordelingsniveaus dan bij Inspectie 2.0. Wanneer we Figuur 29 vergelijken met Figuur 30 valt het op dat het percentage scholen in de uiterste beoordelingsniveaus bij doorlichtingsronde 3 – dat zijn de niveaus ‘zwak’ en ‘sterk’ – hoger ligt dan het percentage scholen dat beoordeeld wordt als ‘beneden de verwachting’ of ‘overstijgt de verwachting’.

Ter illustratie, het percentage scholen dat geclassificeerd wordt als ‘**zwak**’ bedraagt 3 tot 35 procent afhankelijk van het kwaliteitsaspect en het schooljaar. Het percentage scholen dat **beneden de verwachting** scoort bedraagt daarentegen slechts 4 tot 9 procent afhankelijk van het schooljaar. Aan de andere kant van het beoordelingsspectrum wordt 0 tot 16 procent van de scholen beschouwd als **sterk** volgens doorlichtingsronde 3. Bij Inspectie 2.0 bedraagt het percentage scholen dat **de verwachting overstijgt** steeds 0.

Tijdens de focusgroepen (zie Module 2.3) werd er door de participanten van de onderwijsinspectie een verklaring gegeven voor bovenstaande bevinding. Doordat er aanvankelijk meerdere constructen vervat zaten in de schalen van Inspectie 2.0 werden de uiterste beoordelingsniveaus minder gebruikt. Wanneer een school bijvoorbeeld erg goed scoort op het ene construct van de school en minder goed voor het andere construct van die school leidt dat tot een van de middelste beoordelingsniveaus (benadert of volgens de verwachting). Ondertussen zijn de schalen van Inspectie 2.0 meer uitgezuiverd door de onderwijsinspectie.

4. Vergelijking van trends op basis van het peilingsonderzoek en het internationaal vergelijkend onderzoek

Zoals eerder aangegeven, maken we in de huidige studie een vergelijking tussen de trends van (1) PIRLS en de peiling Nederlands basisonderwijs, (2) TIMSS en de peiling wiskunde basisonderwijs en (3) PISA en de peilingen wiskunde in de A- en B-stroom. Omwille van de omvang van de resultaten worden de grafieken en tabellen van elk van die vergelijkingen weergegeven in Bijlage 3.

4.1 Leeswijzer voor de resultaten

Hieronder lichten we de vergelijking tussen PIRLS en de peiling Nederlands gedetailleerd toe. De resultaten in Bijlage 3 van de vergelijking bij (1) TIMSS en de peiling wiskunde basisonderwijs en (2) PISA en de peilingen wiskunde in de A- en B-stroom kunnen op dezelfde manier gelezen worden.

Voordat we de trends van de peiling Nederlands (lezen) en PIRLS kunnen vergelijken, schatten en herschalen we eerst de trends voor beide bronnen afzonderlijk. Tabel 19 geeft de resultaten van die herschaalde trends weer. Hieronder lichten we kort die resultaten toe.

- **Berekening en herschaling van de trend bij PIRLS**

In 2006 nam Vlaanderen deel aan PIRLS en behaalde het Vlaamse onderwijs een gemiddelde van 547 op de PIRLS-schaal. Tien jaar later – in 2016 – nam Vlaanderen voor een tweede keer deel aan PIRLS. Toen bedroeg het Vlaams gemiddelde voor begrijpend lezen 525 (Mullis et al., 2007a, 2017).

Om bovenstaande resultaten op PIRLS 2006 en PIRLS 2016 te vergelijken met de trend voor lezen bij het peilingsonderzoek, herschalen we het gemiddelde bij de afname uit 2006 (dat is 547) naar 0. Ook de standaarddeviatie van PIRLS 2006 herschalen we naar 1. De exacte lineaire transformatie die daarvoor nodig is, passen we vervolgens toe op het gemiddelde en de

standaarddeviatie bij PIRLS 2016. Als resultaat van die transformatie bedraagt het herschaalde gemiddelde van PIRLS 2016 -0,3929.

Wanneer we een trend op basis van twee afnamemomenten willen beschrijven, berekenen we het verschil tussen het gemiddelde op beide afnamemomenten (zie 2.2.1.1 Trends op basis van twee afnames: verschilscores). Bijgevolg is de trend gelijk aan -0,3929. Dat wil zeggen dat de Vlaamse prestatie op PIRLS in tien jaar tijd significant gedaald is met 0,3929. Omgerekend, tussen 2006 en 2016 is de Vlaamse prestatie op PIRLS gemiddeld genomen elk jaar met 0,0393 gedaald.

- **Berekening en herschaling van de trend bij de peiling Nederlands (lezen)**

De afgelopen 15 jaar werden er verschillende peilingen voor lezen georganiseerd in het basisonderwijs. In 2007 behaalde 89% van de leerlingen de eindtermen. Bij de peiling uit 2013 en 2018 ging het om respectievelijk 92% en 84% van de leerlingen die de eindtermen bereikten (Denis et al., 2019b).

Bij het peilingsonderzoek gebruiken we – net zoals bij PIRLS – het gemiddelde om de trend te berekenen en niet het percentage leerlingen dat de eindtermen behaalt. Aan de hand van die gemiddeldes schatten we een lineaire trend, aangezien er meer dan twee afnamemomenten van die peiling zijn (zie 2.2.1.2 Trends op basis van meer dan twee afnames: lineaire trends).

Vervolgens herschalen we het resultaat van die lineaire trend zodat het gemiddelde van de peiling uit 2007 gelijk is aan 0 en de standaarddeviatie 1 is. Diezelfde lineaire transformatie wordt toegepast op het gemiddelde bij de peilingen uit 2013 en 2018.

Het resultaat van die herschaalde trend bedraagt -0,0969. Dat betekent dat de gemiddelde prestatie op de peiling lezen gemiddeld genomen met 0,0969 daalt tussen twee opeenvolgende afnamemomenten (in dit geval: tussen de peiling uit 2007 en 2013 enerzijds en de peiling uit 2013 en 2018 anderzijds). Op basis van die schatting kunnen we de gemiddelde trend per jaar berekenen. Die bedraagt bij de peiling lezen -0,0176. Met andere woorden, tussen 2007 en 2018 is de gemiddelde prestatie op de peiling lezen elk jaar met ongeveer 0,0176 gedaald.

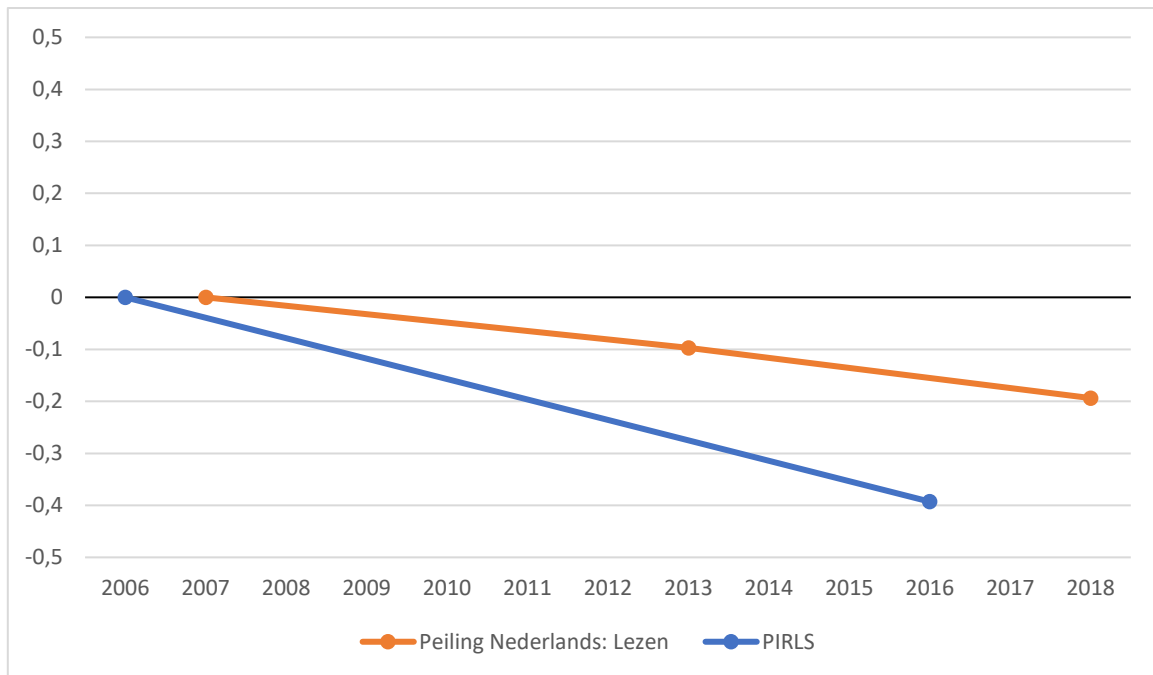
Tabel 19

Trends in (begrijpend) lezen volgens de peiling Nederlands basisonderwijs en PIRLS

	Herschaalde trend		Jaarlijkse trend
	Helling	SE	Helling
PIRLS			
Begrijpend lezen	-0,3929 ***	0,0493	-0,0393
Peiling Nederlands			
Lezen	-0,0969	0,0804	-0,0176

*Noot. *Het effect is statistisch significant op een significantieniveau van .05; ** op een significantieniveau van .01; *** op een significantieniveau van .001*

De herschaalde trends voor zowel PIRLS als de peiling Nederlands lezen (uit Tabel 19) kunnen we vervolgens vergelijken. De resultaten van die vergelijking worden visueel voorgesteld in Figuur 31. De **blauwe** lijn geeft de jaarlijkse trend van PIRLS weer. De jaarlijkse trend volgens de peiling Nederlands lezen wordt weergegeven aan de hand van de **oranje** lijn. Tabel 19 en Figuur 31 tonen een dalende trend bij zowel PIRLS als de peiling Nederlands lezen. De trend bij PIRLS daalt weliswaar sterker dan de trend bij de peiling Nederlands lezen.



Figuur 31. Trends in (begrijpend) lezen volgens de peiling Nederlands basisonderwijs en PIRLS

De trends bij (1) de peiling wiskunde basisonderwijs en TIMSS en (2) de peiling wiskunde in de A- en B-stroom en PISA worden op dezelfde manier geschat, herschaald en visueel voorgesteld in Bijlage 3.

4.2 Hoofdbevindingen

Over het algemeen stellen we een dalende trend vast. De vastgestelde trends bij de meerderheid van de peilingsschalen en alle internationaal vergelijkende onderzoeken dalen immers significant. Bij de schaal omgaan met data (peiling wiskunde A-stroom) vinden we echter een significant stijgende trend. Ook zijn er enkele schalen die stabiel blijven, meer bepaald de schaal algebraïsering (peiling wiskunde A-stroom) en de schaal meten (peiling wiskunde B-stroom).

Hoewel de resultaten een algemeen dalende trend tonen, blijkt de concordantie binnen elk van de drie vergelijkingen beperkt. De figuren en tabellen in Bijlage 3 geven bij elk van de drie vergelijkingen een divers beeld weer. Er zijn enkele opvallende verschillen tussen het peilingsonderzoek en het internationaal vergelijkend onderzoek die mogelijk aan de basis kunnen liggen van die beperkte concordantie:

- **Verschillen in gemeten constructen**

Hoewel de peilingen en het internationaal vergelijkend onderzoek in elke vergelijking hetzelfde domein meten, is er een duidelijk verschil qua gemeten constructen. Bij het peilingsonderzoek wordt er voor specifieke schalen nagegaan in welke mate de leerlingen de eindtermen behalen. Bij het internationaal vergelijkend onderzoek wordt er gefocust op een meer algemene vaardigheid. Bijgevolg resulteren de peilingen in een aparte trend voor elk deeldomein binnen bijvoorbeeld het domein wiskunde. De trends bij het internationaal vergelijkend onderzoek daarentegen geven een eerder algemene trend voor wiskunde weer zonder te focussen op die specifieke deeldomeinen.

Zoals eerder aangegeven, doen de internationale onderzoeken bij elke nieuwe cyclus aanpassingen aan de constructen die ze meten. Het peilingsonderzoek doet dergelijke aanpassingen niet. Bijgevolg zijn trends bij de internationaal vergelijkend onderzoeken dus mogelijk (deels) toe te schrijven aan veranderingen in het gemeten construct.

- **Vershil in het secundair onderwijs: A- en B-stroom afzonderlijk of alle onderwijsvormen samen?**

Uit de bijkomende analyses in Studie 2 bleek dat het een verschil maakt wanneer de A- en B-stroom afzonderlijk of samen worden opgenomen. Ook in de huidige studie moeten we die opmerking maken. Bij PISA worden de trends immers berekend op basis van een diverse groep leerlingen die niet allemaal dezelfde onderwijsvorm volgen. Bij het peilingsonderzoek is dat anders. Er is een aparte peiling georganiseerd voor de A- en B-stroom. Bijgevolg zijn de trends bij de peilingen in de eerste graad secundair onderwijs gebaseerd op de resultaten van leerlingen die allemaal les volgen in ofwel de A-stroom, ofwel de B-stroom. Net zoals in Studie 2 is dat verschil in het secundair onderwijs van belang bij het lezen en interpreteren van de figuren en tabellen in Bijlage 3.

- **Vershil tussen afnamejaar en geboortjaar**

In de figuren in Bijlage 3 worden de trends voorgesteld op basis van het jaar waarin een bepaalde peiling of een specifiek internationaal onderzoek afgenomen is. Echter, het is van belang om bij de interpretatie van die figuren rekening te houden met het geboortjaar van leerlingen die deelnemen aan bepaalde studies, aangezien de populaties van die bronnen verschillen. Bijvoorbeeld, PIRLS 2016 is afgenomen in het vierde leerjaar en de peiling Nederlands 2018 is afgenomen in het zesde leerjaar. Normaalvorderende leerlingen hebben dus deelgenomen aan PIRLS 2016 en de peiling in 2018. Hetzelfde geldt voor de cohorte leerlingen dat deelnam aan TIMSS 2019 en de peiling wiskunde basisonderwijs in 2021.

- **Vershil in schattingsprocedures van de vaardigheid**

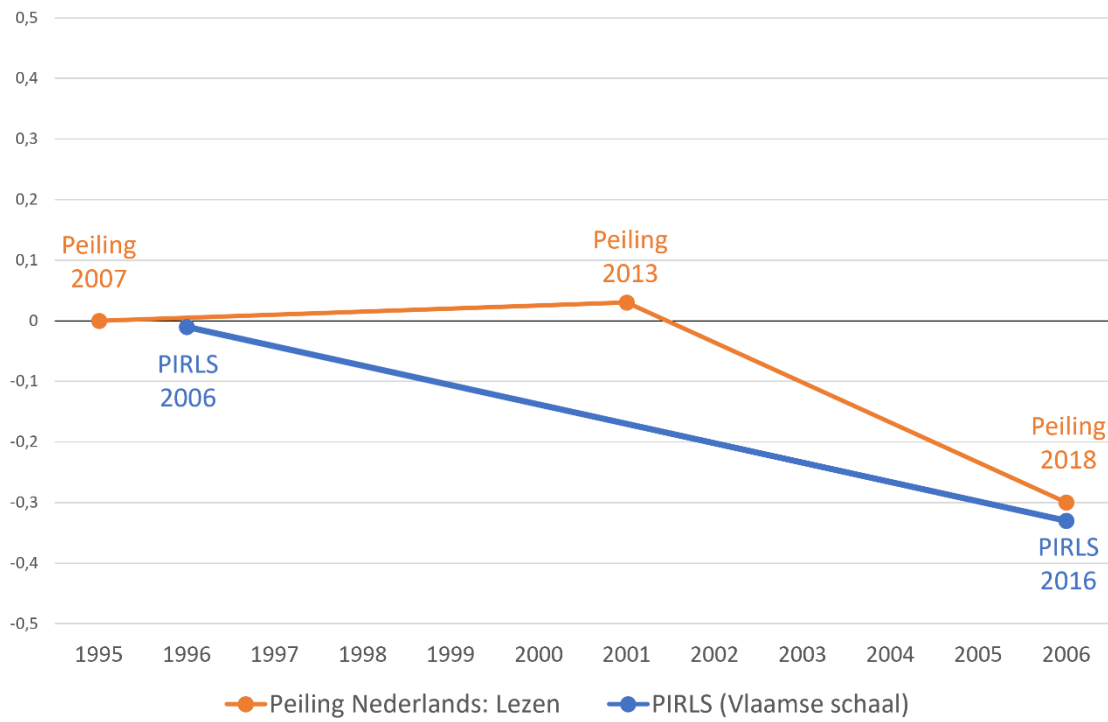
In de huidige studie maken we gebruik van de data zoals die door STEP, IEA en OESO opgeleverd worden. Echter, de internationaal vergelijkende studies houden bij de schatting van de itemparameters rekening met alle landen waar die studie wordt afgenomen. Bij het peilingsonderzoek wordt er bij die schatting logischerwijs enkel rekening gehouden met het Vlaamse onderwijssysteem. Dat verschil kan aan de basis liggen van de beperkte concordantie die we vinden binnen elk van de drie vergelijkingen.

Een mogelijke oplossing is om bij de internationale onderzoeken geen gebruik te maken van de data zoals die zijn opgeleverd door de IEA en OESO en zelf aparte itemparameters te schatten voor Vlaanderen (Gebhardt & Adams, 2007; Sachse et al., 2016). Op basis van die itemparameters kunnen de trends voor Vlaanderen afzonderlijk geschat worden.

Rekening houdend met bovenstaande verschillen hebben we enkele bijkomende analyses uitgevoerd. In die analyses schatten we een algemene trend voor de peilingen (in plaats van aparte trends voor de verschillende deeldomeinen) en houden we rekening met de verschillende onderwijsvormen in het secundair onderwijs. Bovendien schatten we de trends bij de internationale onderzoeken afzonderlijk voor Vlaanderen en brengen we de trends in kaart met oog op het geboortjaar van de leerlingen. De resultaten van die bijkomende analyses zijn te vinden in 'Bijlage 3 - Bijkomende analyses' (p. 322).

De resultaten uit die bijkomende analyses tonen geen significant verschil tussen de trends in het peilingsonderzoek en de trends in het internationaal vergelijkend onderzoek. Zo illustreert Figuur 32 dat er geen significant verschil is in de trend tussen de peiling lezen en Vlaams schaal van PIRLS rekening houdend met het geboortjaar van de leerlingen. De oranje lijn geeft de trends volgens het peilingonderzoek weer en de blauwe lijn de trend volgens de Vlaamse schaal van PIRLS. De x-as geeft telkens het geboortjaar weer van de normaalvorderende leerlingen. Op die manier kunnen we visueel weergeven dat sommige bronnen eenzelfde geboortecohorte meten. Dat is bijvoorbeeld het geval bij

de peiling Nederlands lezen in 2018 en PIRLS in 2016 die beiden gericht waren op de geboortecohorte van 2006.



Figuur 32. Trends in (begrijpend) lezen volgens de peiling Nederlands basisonderwijs en PIRLS

Conclusie over de drie studies heen

1. Inleiding

Aan de hand van drie cross-instrumentele studies hebben we in deze module onderzocht in welke mate de resultaten van de doorlichtingen, het peilingsonderzoek en het internationaal vergelijkend onderzoek in Vlaanderen convergeren of divergeren (Onderzoeksvraag 2). Concreet gaat het om de volgende drie studies die zich focussen op het systeem- en/of schoolniveau:

- Studie 1: PIRLS Repeat (systeem- en schoolniveau);
- Studie 2: Scholen met meer dan één recente bron van kwaliteitsindicatoren (schoolniveau);
- Studie 3: Vergelijking van trends (systeemniveau).

In dit besluit trachten we Onderzoeksvraag 2 te beantwoorden. Daarvoor maken we een onderscheid tussen het systeem- en schoolniveau. Op basis van Studie 1 en Studie 2 kunnen we enkele conclusies formuleren met betrekking tot het schoolniveau. Het besluit over het systeemniveau daarentegen is gebaseerd op de resultaten van Studie 1 en Studie 3.

2. Mate van convergentie of divergentie op schoolniveau

DE CONCORDANTIE TUSSEN DE RESULTATEN VAN DE VERSCHILLENDE BRONNEN VAN EXTERNE KWALITEITSZORG STIJGT WANNEER ER GECONTROLEERD WORDT VOOR ACHTERGRONDKENMERKEN VAN LEERLINGEN EN SCHOLEN.

Uit zowel Studie 1 als Studie 2 blijkt dat de resultaten van de doorlichtingen, het peilingsonderzoek en het internationaal vergelijkend onderzoek over het algemeen bij elkaar aansluiten. Meer specifiek divergeren de resultaten het meest wanneer we een vergelijking maken op basis van type 0-schooleffecten. De resultaten van de verschillende bronnen convergeren dan weer het sterkst wanneer we een vergelijking maken op basis van type B-schooleffecten. Anders gezegd, wanneer we controleren voor de achtergrondkenmerken van leerlingen en scholen – wat het geval is bij een type B-model – sluiten de vaststellingen op basis van de doorlichtingen, het peilingsonderzoek en het internationaal vergelijkend onderzoek beter bij elkaar aan.

BIJ ELKE VERGELIJKING TUSSEN BRONNEN VAN EXTERNE KWALITEITSZORG ZIJN ER SCHOLEN WAARBIJ MEERDERE BRONNEN LEIDEN TOT EEN DIVERGENT BEELD.

Hoewel de resultaten van de verschillende bronnen bij de meeste scholen leiden tot een relatief convergent beeld – dat is zeker het geval wanneer we controleren voor achtergrondkenmerken –, zijn er ook steeds scholen waarbij meerdere bronnen leiden tot een eerder divergent beeld. Het is van belang om te beseffen dat die divergerende resultaten een grote impact kunnen hebben voor de scholen in kwestie. Dat zou bijvoorbeeld het geval kunnen zijn als er gevolgen gekoppeld worden (bijvoorbeeld een begeleidingstraject, een bijkomende doorlichting...) aan het resultaat van een school. Die impact zou bovendien nog groter kunnen worden als de resultaten van slechts één bron bepalend zijn voor de toekenning van dergelijke gevolgen.

KANTTEKENINGEN

Over het algemeen stellen we dus vast dat de resultaten van de verschillende bronnen op schoolniveau voor de meeste – maar niet alle – scholen convergeren wanneer we een eerlijke vergelijking maken door te controleren voor achtergrondkenmerken van de leerlingen en scholen. Bij dat algemene

antwoord op de onderzoeksvraag is het echter van belang om de volgende twee kanttekeningen te maken:

- **Welke groepen leerlingen of scholen we met elkaar vergelijken, maakt een verschil.**
Zoals eerder aangehaald in Module 2.1 richt elke bron zich op een specifieke populatie en zijn er heel wat verschillen in de populaties die de peilingen, de internationale onderzoeken en de doorlichtingen vooropstellen. Wanneer we twee bronnen met elkaar vergelijken, moeten we rekening houden met die verschillen. Zo toont Studie 2 dat het wel degelijk een verschil maakt met welke groep leerlingen we de resultaten van de peiling B-stroom vergelijken. Wanneer we de schooleffecten op basis van de peiling B-stroom vergelijken met de PISA-schooleffecten – die berekend zijn op basis van alle leerlingen ongeacht hun onderwijsvorm – leidt dat voor een groot aantal scholen tot een divergent beeld. Echter, wanneer we een vergelijking maken met enkel de PISA-leerlingen uit de B-stroom en het BSO is er een relatief grote overeenkomst tussen het resultaat op PISA en de peiling wiskunde B-stroom.

- **Een absoluut criterium vergelijken met een relatief criterium**
De onderwijsinspectie werkt met een absoluut criterium. Elke school wordt vergeleken met een criterium, namelijk de verwachting en voldoet bijgevolg wel of niet aan dat criterium. Het peilingsonderzoek en het internationaal vergelijkend onderzoek daarentegen maken (voor het schatten van schooleffecten) gebruik van een relatief criterium. Daarbij wordt elke school vergeleken met het gemiddelde. Dat gemiddelde is echter onderhevig aan fluctuaties in leerprestaties doorheen de tijd, terwijl een absoluut criterium constant blijft doorheen de tijd.

Hoe het absoluut criterium bij de doorlichtingen – de minimale verwachting waaraan scholen zouden moeten voldoen – zich juist verhoudt ten opzichte van het relatief criterium bij de verschillende peilingen en internationale onderzoeken – het gemiddelde – is onduidelijk. Wel toont Studie 2 dat voldoen aan de verwachting bij de doorlichting niet noodzakelijk overeenkomt met het Vlaams gemiddelde. Om te kunnen besluiten of een school voldoet aan de minimale verwachting is een vergelijking met een absoluut criterium dus noodzakelijk. Ook bij het in kaart brengen van het percentage leerlingen dat de eindtermen bereikt, wordt een absoluut criterium gehanteerd. Omgekeerd, een vergelijking met het Vlaamse gemiddelde – het relatieve criterium – is essentieel om te kunnen concluderen of een school (onder/boven)gemiddeld presteert.

Van belang om op te merken is dat scholen zelf geen zicht hebben op die relatie tussen het absolute criterium bij de doorlichtingen en het relatieve criterium bij de andere bronnen van externe kwaliteitszorg. Zo is het zelfs in een scenario waarbij elke school aan de verwachting voldoet nog steeds mogelijk dat een bepaalde school onder het gemiddelde scoort. In dat geval kan het voor de school in kwestie lijken alsof de schoolfeedback die ze ontvangt van de verschillende bronnen niet overeenkomt. Hoewel de combinatie van ‘volgens de verwachting’ en een ondergemiddeld schooleffect in dat scenario niet vreemd is, kan de school in kwestie dat niet besluiten op basis van de informatie waarover zij zelf beschikt.

3. Mate van convergentie of divergentie op systeemniveau

Zoals eerder aangegeven in Studie 3 is het niet mogelijk om de trends van de doorlichtingen te vergelijken met trends van het peilingsonderzoek of het internationaal vergelijkend onderzoek. Er zijn namelijk geen trends van gelijkaardige proceskenmerken bij de peilingen en de internationale onderzoeken waarmee we de trends van de doorlichtingen zouden kunnen vergelijken. De proceskenmerken – die in aanmerking kwamen voor een vergelijking – worden bij het

peilingsonderzoek en het internationaal vergelijkend onderzoek niet steeds op dezelfde manier gemeten. Trends voor proceskenmerken kunnen echter enkel beschreven worden wanneer die proceskenmerken in de achtergrondvragenlijsten steeds op dezelfde manier bevestigd worden.

Bijgevolg wordt er in geen enkele cross-instrumentele studie op systeemniveau een vergelijking gemaakt tussen de resultaten van de doorlichtingen enerzijds en de resultaten van de andere bronnen voor externe kwaliteitszorg anderzijds. Daarom splitsen we de conclusie met betrekking tot het systeemniveau op in twee onderdelen: '3.1 De doorlichtingen van de onderwijsinspectie' en '3.2 Het peilingsonderzoek en het internationaal vergelijkend onderzoek'.

3.1 De doorlichtingen van de onderwijsinspectie

In Studie 3 worden de trends voor doorlichtingsronde 3 vergeleken met de trends voor Inspectie 2.0. Opvallend is dat de vergelijking wijst op zowel een gelijkenis als een verschil tussen doorlichtingsronde 3 en Inspectie 2.0. Ten eerste toont die vergelijking dat het percentage scholen dat een specifieke inschaling ontvangt van de onderwijsinspectie relatief sterk varieert over schooljaren heen. Die vaststelling gaat op voor zowel doorlichtingsronde 3 als Inspectie 2.0. Ten tweede stellen we vast dat de uiterste beoordelingscategorieën vaker gebruikt worden bij doorlichtingsronde 3. Dat is een duidelijk verschil met de resultaten bij Inspectie 2.0 waarbij voornamelijk de middelste twee categorieën – benadert en volgens de verwachting – gebruikt worden.

KANTTEKENINGEN

Hoewel we op schoolniveau vaststellen dat de resultaten van de doorlichtingen meestal aansluiten bij de resultaten van de andere bronnen voor externe kwaliteitszorg (zie 2 Mate van convergentie of divergentie op schoolniveau), stellen we op systeemniveau vast dat er binnen de doorlichtingsresultaten zelf heel wat variatie is in de resultaten over verschillende schooljaren heen. Ook hier is het van belang om enkele kanttekeningen te maken:

- **Doorlichtingsronde 3 is een andere doorlichtingsmethode dan Inspectie 2.0.**
Zoals besproken in Studie 3 zijn er heel wat verschillen tussen de doorlichtingsmethodes 'doorlichtingsronde 3' en 'Inspectie 2.0'. Meer bepaald gaat het om verschillen in (1) beoordeelde proceskenmerken, (2) aantal inschalingen per proceskenmerk en (3) beoordelingsniveaus (zie '2.1.1 Doorlichtingsronde 3 en Inspectie 2.0'). Die verschillen zorgen ervoor dat het niet eenvoudig is om de trends voor de doorlichtingen in kaart te brengen. De resultaten moeten dus met enige voorzichtigheid geïnterpreteerd worden.
- **Er zijn inhoudelijke aanpassingen mogelijk binnen eenzelfde doorlichtingsmethode.**
Op systeemniveau is er heel wat variatie in de resultaten van de doorlichtingen over verschillende schooljaren heen, zelfs binnen eenzelfde doorlichtingsmethode. Die variatie is mogelijk deels te wijten aan de inhoudelijke aanpassingen die de onderwijsinspectie doet binnen eenzelfde doorlichtingsmethode. De onderwijsinspectie kan namelijk de inhoudelijke omschrijving van de proceskenmerken bij doorlichtingsronde 3 en/of de kritische kenmerken die horen bij de ontwikkelingschalen van Inspectie 2.0 bijsturen en aanpassen.

Om zeker te zijn dat die variatie niet te wijten is aan inhoudelijke aanpassingen is het noodzakelijk dat dezelfde constructen ongewijzigd blijven en steeds op dezelfde manier gemeten worden binnen eenzelfde doorlichtingsmethode.

3.2 Het peilingsonderzoek en het internationaal vergelijkend onderzoek

Uit Studie 1 blijkt dat de resultaten voor de PIRLS- en peilingschaal sterk convergeren bij verschillende types toegevoegde waarde modellen. Wanneer we in Studie 3 focussen op de trends bij het

peilingsonderzoek en internationaal vergelijkend onderzoek vinden we over het algemeen een dalende trend. Al zijn er ook enkele peilingschalen die stabiel blijven of stijgen.

Wanneer we een vergelijking maken tussen de oorspronkelijke trends van STEP, IEA en OESO die hetzelfde construct beogen, is er eenzelfde richting in de trends. De sterkte van de trends is wel verschillend. Echter, wanneer we voor zowel de internationale studies als peilingsonderzoeken een herkalibratie doen aan de hand van hetzelfde model op basis van enkel de Vlaamse gegevens, dan wordt ook de sterkte van de trends meer gelijkaardig.

KANTTEKENINGEN

Over het algemeen stellen we dus vast dat de resultaten van het peilingsonderzoek en het internationaal vergelijkend onderzoek op systeemniveau convergeren wanneer we een gelijkaardige schattingsprocedure hanteren. Daarbij is het echter van belang om de volgende opmerkingen te maken:

- **Welke groepen leerlingen of scholen we met elkaar vergelijken, maakt een verschil.**
Net zoals vermeld bij '2 Mate van convergentie of divergentie op schoolniveau' moeten we ook bij de vergelijking van trends rekening houden met die verschillen in populaties. Bijvoorbeeld, bij PISA worden de trends berekend op basis van een diverse groep leerlingen die niet allemaal dezelfde onderwijsvorm volgen. Bij het peilingsonderzoek zijn er aparte peilingen voor de A- en B-stroom. Met als gevolg, wie geïnteresseerd is in afzonderlijke trends voor de A- en B-stroom baseert zich best op de trends bij de peilingen.
- **Er zijn aanpassingen mogelijk aan de gemeten constructen bij het internationaal vergelijkend onderzoek.**
Bij elke nieuwe cyclus doen de internationale onderzoeken aanpassingen aan de constructen die ze meten.³⁹ Bijgevolg zijn de trends bij internationaal vergelijkende onderzoeken mogelijk (deels) toe te schrijven aan veranderingen in het gemeten construct. Om zeker te zijn dat de trends niet te wijten zijn aan veranderingen in het gemeten construct is het essentieel dat de constructen ongewijzigd blijven en steeds op dezelfde manier gemeten worden over cycli heen.

³⁹ Dit zijn enkele voorbeelden van concrete aanpassingen aan de constructen bij PISA, PIRLS en TIMSS:

- De internationale onderzoeken actualiseren hun constructen over afnames heen. Bijvoorbeeld, in PISA wordt in elke cyclus het hoofddomein herbekeken en aangepast door een expertpanel (OECD, 2019a). Bij TIMSS en PIRLS is er elke cyclus een nieuwe analyse van de curricula van de deelnemende landen in die cyclus. Het gewicht dat gegeven wordt aan verschillende facetten van hun constructen wordt op dat moment herbekeken (Mullis & Martin, 2015, 2017).
- Zoals eerder besproken in Studie 3, worden er bij elke cyclus van een internationaal vergelijkend onderzoek (1) oude items verwijderd en (2) nieuwe items toegevoegd. Door die itemverversing kan niet onderzocht worden of de nieuwe items daadwerkelijk hetzelfde construct meten als de items die verwijderd zijn.

Module 2.3: Complementariteit van de verschillende bronnen

1. Inleiding

Om de onderwijskwaliteit van scholen en onderwijssystemen in te schatten wordt het gebruik van meerdere bronnen sterk aanbevolen (Van Der Lans et al., 2019). Nochtans tonen Module 2.1 en 2.2 dat het eenvoudigweg gebruik maken van meerdere bronnen niet automatisch leidt tot eenduidige en betrouwbare besluitvorming over onderwijskwaliteit op systeem- en schoolniveau. Zo illustreert de matrix (p. 24) dat geen enkele bron valide uitspraken kan doen over alle aspecten van onderwijskwaliteit en tonen de resultaten van Module 2.2 (p. 45) dat informatie uit meerdere bronnen ook tot een divergerend beeld kan leiden.

In deze module focussen we daarom op de relaties tussen de verschillende bronnen van externe kwaliteitszorg. In de eerste plaats willen we tot mogelijke verklaringen komen voor de concordantie in de resultaten van verschillende bronnen van externe kwaliteitszorg (Onderzoeksvraag 3). In de tweede plaats willen we een beter beeld krijgen van hoe de verschillende bronnen complementair ingezet kunnen worden (Onderzoeksvraag 4).

2. Onderzoekopzet

Het doel van de voorliggende module is dus tweeledig. Enerzijds onderzoeken we in welke mate de gelijkenissen en verschillen tussen de verschillende bronnen – besproken in Module 2.1 – een verklaring vormen voor de concordantie tussen de resultaten van die bronnen – de resultaten uit Module 2.2 (Onderzoeksvraag 3). Anderzijds willen we een beter beeld krijgen van hoe de verschillende bronnen en de bijhorende data gecombineerd kunnen worden om zo tot meer betrouwbare en accurate uitspraken over onderwijskwaliteit te komen (Onderzoeksvraag 4).

Belangrijk om op te merken is dat verschillende bronnen van externe kwaliteitszorg combineren niet hetzelfde is als één overkoepelende maat voor onderwijskwaliteit genereren. Een overkoepelende maat camoufleert immers de sterke en zwakke punten van een school of onderwijssysteem en bijgevolg ook de domeinen waarop schoolverbetering zich al dan niet moet richten (Hough et al., 2016). Wel staan de relaties tussen verschillende bronnen centraal bij het combineren van die bronnen. Anders gezegd, de vraag naar de combinatie van verschillende bronnen voor externe kwaliteitszorg focust dus op hoe de die bronnen complementair ingezet kunnen worden.

2.1 Dataverzameling

Om bovenstaande onderzoeksvragen te beantwoorden, werden twee focusgroepen georganiseerd. Via de focusgroepen willen we de participanten aanzetten om in interactie met elkaar te reflecteren over (1) de resultaten uit Module 2.1 en Module 2.2 en (2) de complementariteit van de verschillende bronnen. De interactie tussen de participanten is daarbij cruciaal. Mede dankzij die interactie is het mogelijk om verschillende standpunten, meningen en ervaringen te delen en te bespreken, maar ook verder te bouwen op ideeën van anderen. Op die manier is het mogelijk om data te verzamelen die verder gaan dan het individuele standpunt (Kitzinger, 1995; Ravitch & Mittenfelner Carl, 2016). Bovendien laat die sociale interactie tijdens de focusgroepen ons ook toe om te polsen naar de motivering achter bepaalde ideeën en de mate waarin die ideeën beaamd worden door andere deelnemers in de focusgroep (Krueger & Casey, 2015; Morgan, 2013).

Via de focusgroepen willen we zowel een brede als diepgaande analyse maken van mogelijke verklaringen voor de concordantie tussen de resultaten van verschillende bronnen enerzijds en het

complementair gebruik van die bronnen anderzijds. De focusgroepen hebben dus een dubbele insteek in lijn met respectievelijk Onderzoeksvraag 3 en 4.

2.1.1 Participanten

Met het oog op het doel van de focusgroepen, contacteerden en selecteerden we participanten die rechtstreeks in contact komen met de ontwikkeling en rapportage van minstens een van de drie bronnen voor externe kwaliteitszorg, i.e. de doorlichtingen, het peilingsonderzoek en het internationaal vergelijkend onderzoek. Meer specifiek ging het om leden van de onderwijsinspectie (N = 2), het onderwijsbeleid (N = 2), de onderzoeksteams van de internationaal vergelijkende onderzoeken (N = 3; een participant voor respectievelijk PIRLS, TIMSS en PISA), het onderzoeksteam van het peilingsonderzoek (N = 2), directies (N = 2) en de pedagogische begeleidingsdiensten van verschillende onderwijskoepels (N = 4).

In totaal werden twee focusgroepgesprekken afgenomen, waarbij er respectievelijk 6 en 7 participanten deelnamen. Die groepsgrootte is ideaal om voldoende meningen te horen en informatie te verzamelen, maar klein genoeg om te modereren (Krueger & Casey, 2015). Daarnaast kozen we ervoor om de focusgroepen heterogeen en evenwichtig samen te stellen, aangezien we binnen elke focusgroep verschillende standpunten wilden vatten (Cyr, 2019).

Twee participanten hadden initieel bevestigd om deel te nemen aan de focusgroepen, maar lieten in de week voorafgaand aan de focusgroepen weten dat ze niet meer konden deelnemen. Die participanten hebben we een opname van de presentatie en de bijhorende vragen bezorgd. Op basis van die informatie hebben zij ons een schriftelijke reactie bezorgd. Die schriftelijke reacties zijn meegenomen in de analyses van de focusgroepen.

2.1.2 Verloop van de focusgroep

Ter voorbereiding van de focusgroepen werd er een leidraad opgesteld om de discussie te structureren. Meer concreet omvat die leidraad (1) het verloop van de focusgroep en (2) de hoofdvragen en enkele bijkomende vragen die gesteld kunnen worden indien nodig. De leidraad is te vinden in Bijlage 4 (p. 328). Daarnaast werd er een presentatie samengesteld met daarin de bevindingen uit de eerste twee modules. Het doel van die presentatie was om de participanten uitgebreid te informeren over de vaststellingen uit de voorgaande modules. Op die manier kunnen de participanten doelgericht deelnemen aan de focusgroepen. Alle participanten ontvingen een uitnodiging voor een online samenkomst in Microsoft Teams. Zo was het mogelijk om de gesprekken digitaal op te nemen zonder dat de deelnemers toegang kregen tot die opname (Lobe, 2017).

Beide focusgroepen duurden twee tot drie uur en werden online afgenomen in januari 2023. Elke focusgroep startte – zoals de leidraad beschrijft – met een korte inleiding waarin het onderzoek voorgesteld werd en het doel van de focusgroep verduidelijkt werd. Vervolgens werden de bevindingen uit de voorgaande modules uitgebreid toegelicht aan de hand van de eerdergenoemde presentatie. Na een korte pauze volgde de openingsvraag. De participanten mochten zichzelf kort voorstellen en aangegeven op welke manier ze in aanraking komen met minstens een van de verschillende bronnen voor externe kwaliteitszorg. Aangezien de focusgroepen heterogeen samengesteld zijn, is het van belang om te starten met een openingsvraag die toont wat de participanten gemeen hebben met elkaar om van daaruit verder te gaan met het delen van standpunten en ervaringen (Krueger, 2015; Morgan, 2017). Daarna werden de hoofdvragen met betrekking tot de twee doelen van de focusgroep gesteld. De opgestelde leidraad werd daarbij flexibel gehanteerd en waar nodig werd er doorgevraagd om een goed begrip van de verschillende standpunten en ideeën na te streven. Verder werd er – zoals de leidraad aangeeft – ook kort aandacht

besteed aan de toekomstige Vlaamse toetsen. Bij de afronding van de focusgroep werd er aan de participanten gevraagd of zij nog bepaalde zaken kwijt wilden.

2.2 Data-analyse

Op basis van de digitale opnames werden alle online focusgroepgesprekken volledig getranscribeerd. Ook werden er audiochecks uitgevoerd door de opname opnieuw te beluisteren en de transcripten waar nodig te corrigeren of vervolledigen. Vervolgens werden de transcripten gelezen, herlezen en gencodeerd. Concreet werden betekenisvolle tekstfragmenten geselecteerd en kreeg elk fragment één of meerdere codes toegekend die de inhoud van het fragment beschrijven (Vermeir, 2021). De codering gebeurde zowel inductief als deductief. Met andere woorden, de codes kwamen tot stand op basis van het analysekader van de documentanalyse (p. 11) én de inbreng van de participanten. De finale codeboom is te vinden in Bijlage 5 (p. 332).

Eerst werd elke focusgroep afzonderlijk geanalyseerd (within-case analysis). Vervolgens werden twee afzonderlijke analyses systematisch met elkaar vergeleken (cross-case analysis). Doorheen het proces van data-analyse werd er steeds gezocht naar patronen en verbanden tussen de data (Miles & Huberman, 2014).

3. Resultaten

3.1 Verklaringen voor de mate van concordantie tussen de resultaten van verschillende bronnen van externe kwaliteitszorg

Hoewel de resultaten van de bronnen van externe kwaliteitszorg in grote mate convergeren, zullen er ook altijd verschillen worden vastgesteld. Dat kan volgens de leden van de focusgroepen aan allerhande verklaringen liggen. Er is namelijk geen concreet antwoord op wat nu dé oorzaak is van die convergerende en divergerende resultaten. Daarom beperken de focusgroepen zich tot mogelijke verklaringen die een rol kunnen spelen.

Wel zijn de participanten van de focusgroepen ervan overtuigd dat de vastgestelde gelijkenissen en verschillen tussen de bronnen (zie Module 2.1) zeker een bijdrage leveren aan de gevonden resultaten in Module 2.2. Die gelijkenissen en verschillen duiden voor een stuk de gevonden resultaten, maar de leden van de focusgroepen wijzen daarnaast op heel wat andere mogelijke oorzaken. We kunnen dus een onderscheid maken tussen de verklaringen die verband houden met de vastgestelde gelijkenissen en verschillen tussen de bronnen enerzijds (zie Module 2.1) en andere mogelijke oorzaken anderzijds. Voor de duidelijkheid bespreken we hieronder elke mogelijke verklaring afzonderlijk. We benadrukken dat er in werkelijkheid hoogstwaarschijnlijk meerdere verklaringen tegelijkertijd spelen.

3.1.1 De vastgestelde gelijkenissen en verschillen tussen bronnen van externe kwaliteitszorg

Zoals omschreven in Module 2.1, situeren de gelijkenissen en verschillen tussen de bronnen van externe kwaliteitszorg zich op verschillende dimensies. Hieronder geven we voor elke dimensie aan in welke mate de gelijkenissen en verschillen tussen de bronnen (uit Module 2.1) de convergerende en divergerende resultaten (uit Module 2.2) duiden.

- **Doel**

De leden van de focusgroepen bevestigen dat elke bron ontwikkeld wordt met oog op een specifiek doel. Bijgevolg is niet elke bron even geschikt om uitspraken te doen op elk aggregatieniveau (zoals ook blijkt uit de matrix). De toetsen van de peilingen en de internationaal vergelijkende onderzoeken zijn gericht op systeemniveau. Die zijn niet ontworpen om op schoolniveau een betrouwbaar resultaat te geven. Een doorlichting van de onderwijsinspectie daarentegen heeft net als doel om betrouwbare uitspraken op

schoolniveau te doen. Volgens de focusgroepen kunnen die gelijkenissen en verschillen in het beoogde doel aanleiding geven tot convergerende en divergerende resultaten.

- **Domein en construct**

Wat er gemeten wordt, is gewoonweg anders bij de verschillende bronnen van externe kwaliteitszorg. Vooral tussen de doorlichtingen en de andere bronnen is er een groot verschil wat betreft het gemeten domein en construct. Terwijl de onderwijsinspectie zich richt op proceskenmerken, focussen de andere bronnen op outputkenmerken waarbij vooral leerprestaties van leerlingen centraal staan. Al zijn er ook verschillen mogelijk tussen de peilingen en het internationaal vergelijkend onderzoek in wat er gemeten wordt. Dat is bijvoorbeeld het geval wanneer het conceptueel kader van een internationaal onderzoek minder goed afgestemd is op het Vlaamse curriculum. De deelnemers in de focusgroepen veronderstellen meer convergerende resultaten wanneer er meer overeenstemming is in wat er gemeten wordt.

Daarnaast kan het moment waarop bepaalde leerinhouden aan bod komen bepalend zijn. In Vlaanderen zijn er immers verschillende leerplannen. Ook dat kan op systeemniveau in het basisonderwijs een verklarende factor zijn. Zo is het mogelijk dat inhouden die gemeten worden in TIMSS of PIRLS (in het vierde leerjaar) volgens een bepaald leerplan pas in de derde graad aan bod komen, terwijl die inhouden wel al in de tweede graad aan bod komen bij een ander leerplan. In dat geval kan het verschil in leerplannen een mogelijke verklaring zijn voor de mate van concordantie tussen de resultaten van PIRLS/TIMSS en de peilingen.

- **Populatie en steekproef**

Ook de populatie en de bijhorende steekproef kunnen een mogelijke verklaring zijn van zowel convergerende als divergerende resultaten. Ten eerste kan de selectie van (een groep van) specifieke leerlingen voor een stuk de resultaten op **schoolniveau** verklaren. Zo doet het er bijvoorbeeld toe welke klas deelneemt aan een peiling:

“Als net uw klas Latijn-Grieks geselecteerd is, dan is het logisch dat uw resultaat misschien wel beter is dan wanneer een andere studierichting of zo meedoet.”

Dat geldt natuurlijk niet op systeemniveau, want daar wordt er (meestal) gecontroleerd voor de basisopties en studierichtingen van leerlingen. Ook in welk schooljaar een bron wordt afgenomen, speelt een rol. Een voorbeeld: als in dit schooljaar de leerlingen op het einde van de eerste graad erg sterk zijn, haalt die school dit jaar misschien een beter resultaat op de peiling A-stroom of B-stroom dan in andere jaren.

Ten tweede is er op **systeemniveau** een groot verschil wat betreft de verplichting tot deelname. Een doorlichting kunnen scholen niet weigeren, terwijl de vraag tot deelname aan een peilings- of internationaal onderzoek wel geweigerd kan worden. Dat heeft als gevolg dat er voor elke school een doorlichtingsverslag beschikbaar is. Dat is anders bij het peilings- en internationaal onderzoek, aangezien scholen zelf ervoor kunnen kiezen om niet deel te nemen:

“Er zijn de scholen die weigeren om deel te nemen. Soms is dat wel om een toevallige reden van de vaste juf is net in bevallingsverlof, maar vaak merk je toch dat het de scholen zijn die zeggen van ‘Ja wij zitten zo in de penarie op alle mogelijke vlakken. Wij gaan hier nu niet aan deelnemen’.”

De scholen die ervoor kiezen om niet deel te nemen verwachten vaak van zichzelf dat ze niet echt goed gaan scoren. Aangezien we van die groep scholen geen resultaten hebben op de

peilingen en/of het internationaal vergelijkend onderzoek, zijn die scholen niet opgenomen in de vergelijkingen in Module 2.2. Dat kan mogelijks ook een invloed hebben op de gevonden concordantie tussen de resultaten van verschillende bronnen.

- **Modus**

Volgens de focusgroepen speelt ook de manier waarop de data verzameld worden een rol. Een fictief voorbeeld: de onderwijsinspectie en een grootschalig onderzoek meten exact hetzelfde proceskenmerk. Bij de onderwijsinspectie heb je een externe inspecteur die kijkt naar hoe dat proces in de klas verloopt. Bij het grootschalige onderzoek daarentegen wordt dat proceskenmerk bevraagd via een leerkrachtvragenlijst. Eigenlijk vraag je dan aan de leerkracht om het eigen gedrag in de klas te beschrijven. Hoewel de twee bronnen in dit fictieve voorbeeld exact hetzelfde meten, is het mogelijk dat de resultaten divergeren omwille van een verschil in de manier waarop de data verzameld werden. Het is bijvoorbeeld mogelijk dat de leerkracht in de vragenlijst sociaal wenselijke antwoorden invult en dat die antwoorden daardoor sterk verschillen van de beoordeling door de externe inspecteur. Ook de modus kan dus een mogelijke verklaring zijn voor de mate van concordantie.

- **Standaardbepaling**

De standaardbepaling gebeurt op uiteenlopende manieren bij de verschillende bronnen van externe kwaliteitszorg. Ook dat kan bijdragen aan de gevonden resultaten in Module 2.2. In de focusgroep wordt er aangehaald dat het niveau dat volgens eindtermen verwacht wordt – en dus gemeten wordt bij de peilingen – eigenlijk niet zo hoog is in vergelijking met wat er verwacht wordt in internationaal vergelijkend onderzoek. In het internationaal onderzoek zijn de items daardoor wellicht moeilijker dan de items in de peilingstoetsen, wat kan leiden tot divergerende resultaten tussen verschillende bronnen. Dat is in lijn met de resultaten van PIRLS Repeat (Dockx et al., 2019).

- **Analytische methode**

Tot slot is de analytische methode gewoonweg verschillend bij de bronnen van externe kwaliteitszorg. Vooral tussen de doorlichtingen en de andere bronnen is er een groot verschil. Bij de peilingen en de internationaal vergelijkende onderzoeken wordt er gewerkt met scoring, IRT-schalen en verschillende modellen. Dat is iets totaal anders en kan nooit helemaal hetzelfde zijn als de manier waarop de onderwijsinspectie de verzamelde data analyseert. De analyse van kwantitatieve data bij de peilingen en de internationale onderzoeken enerzijds en de analyse van voornamelijk kwalitatieve data bij de doorlichtingen kunnen op zich al aanleiding geven tot convergerende of divergerende resultaten – afhankelijk van welke bronnen je vergelijkt.

3.1.2 Andere mogelijke verklaringen

De vastgestelde gelijkenissen en verschillen tussen de bronnen van externe kwaliteitszorg zijn echter niet de enige mogelijke verklaringen. Er zijn daarnaast heel wat andere elementen die (gedeeltelijk) kunnen bijdragen tot convergerende en divergerende resultaten:

- **De meeste scholen in Vlaanderen presteren gemiddeld (schoolniveau)**

Zoals de resultaten in Module 2.2 aantonen, wijkt de score van de overgrote meerderheid van de scholen in de verschillende studies niet af van het Vlaamse gemiddeld. Ook dat kan een verklaring zijn voor de gevonden convergerende resultaten:

“Dan is het naar mijn gevoel niet moeilijk om als school heel vaak – ook bij herhaalde metingen – in die groep ‘gemiddeld’ te vallen. Dus die donkerblauwe scholen die je daar

had... ja, dat is nu eenmaal de meerderheid in Vlaanderen en dan krijg je al snel het idee van oh ja dat stemt goed overeen.”

- **De achtergrondkenmerken van leerlingen en scholen (schoolniveau)**

Uit de resultaten in Module 2.2 blijkt dat de resultaten het sterkst convergeren wanneer er gecontroleerd wordt voor de achtergrondkenmerken van leerlingen en scholen. Bijgevolg duiden die achtergrondkenmerken van leerlingen en scholen ook een deel van de gevonden resultaten. Ook de participanten van de focusgroepen bevestigen dat. Het leerlingenpubliek van een school maar ook hoe de school met haar leerlingenpubliek omgaat, verklaren voor een stuk de gevonden resultaten op schoolniveau.

- **Verandering is mogelijk (schoolniveau)**

Elke meting door een bron van externe kwaliteitszorg is een momentopname. Bovendien worden die metingen niet op exact hetzelfde moment uitgevoerd. Er kan in een school heel wat veranderen tussen het tijdstip waarop de ene meting afgenomen wordt en het tijdstip waarop de volgende meting plaatsvindt.

Dergelijke verandering kan ook plaatsvinden naar aanleiding van een doorlichting of het ontvangen van een schoolfeedbackrapport. Opvallend is wel dat een doorlichting – in tegenstelling tot de andere bronnen van externe kwaliteitszorg – volgens de leden van de focusgroepen vaker leidt tot verandering:

“Ik denk dat vooral doorlichting zorgt voor verandering op vlak van [...] dit zit goed, dit minder goed. Die peilingen spelen daar minder een rol in dan denk ik. Tenzij ik natuurlijk als directeur zou zien van oei... er is hier iets heel erg aan de hand. [...] Mocht je een hele slechte peiling hebben en zien van we zitten daar serieus onder, ja, dan ga je dat natuurlijk aanpakken. Maar die doorlichting... ja, die haalt wel telkens zaken naar boven.”

Volgens de focusgroep hangt dat samen met het vooropgestelde doel. De doorlichtingen zijn gericht op een specifieke school. Dat gaat over de leraren, de leerlingen en het beleid in jouw school. De peilingen of internationaal vergelijkende onderzoeken bieden minder informatie over wat er op school gebeurt. De impact op een school bij de doorlichtingen is bijgevolg groter dan een internationaal vergelijkend onderzoek of een peiling die als primaire focus het systeemniveau heeft.

Verandering is niet alleen het gevolg van een doorlichting. Er zijn ook heel wat andere factoren, los van de bronnen van externe kwaliteitszorg die kunnen leiden tot verandering op school. Bij de interpretatie van resultaten op schoolniveau moet er volgens de focusgroepen rekening gehouden worden met eventuele veranderingen tussen die twee meetmomenten.

- **Algemene tendensen en veranderingen (systeemniveau)**

Niet alleen op schoolniveau gaat het steeds om momentopnames en zijn veranderingen tussen de tijdstippen van twee metingen mogelijk. Dat geldt ook voor het systeemniveau. Daar spelen er enkele algemene tendensen die een verklaring kunnen zijn voor de mate van concordantie. Dat is bijvoorbeeld het lerarentekort, de coronapandemie... Aangezien dergelijke tendensen een grote impact hebben op alle leerlingen en scholen, kunnen zij een mogelijke verklaring zijn voor de mate van concordantie op systeemniveau.

- **Strategisch gedrag (school- en systeemniveau)**

Scholen die deelnemen aan een peiling, internationaal vergelijkend onderzoek of doorlichting kunnen ook strategisch gedrag vertonen. Zo kan het gebeuren dat een school bepaalde leerlingen niet laat deelnemen aan een internationale toets, omdat ze bang zijn om gezichtsverlies te lijden. Dergelijk strategisch gedrag kan ook een invloed hebben op de resultaten en bijgevolg ook op de mate van concordantie op het einde van de rit. De leden van de focusgroepen geven vanuit hun ervaring echter aan dat het aantal scholen dat strategisch gedrag vertoont op dit moment nog beperkt is.

- **Toetsen meten meer dan de beoogde vaardigheden (school- en systeemniveau)**

Een toets van de peilingen of een internationaal vergelijkend onderzoek meet niet alleen de vaardigheid van de leerlingen, maar ook de houding van de leerlingen ten aanzien van die toets. Dat heeft bijvoorbeeld te maken met de motivatie van de leerlingen om de toets te maken, maar ook met de manier waarop de school het onderzoek voorstelt aan de leerlingen. Volgens de focusgroepen is dat van belang, omdat toetsmotivatie in sterke mate samenhangt met de resultaten die leerlingen behalen. Hoe hoger de leerlingen het belang van de toets percipiëren, hoe meer ze zich gaan inzetten:

“De ervaring leert mij dan toch vooral dat het vaak B-stroomleerlingen zijn die het misschien wel kunnen, maar die gewoon echt geen zin hebben in die toets. En dan zal de doorlichting misschien wel een correcter beeld kunnen geven van wat de school al doet.”

3.2 Resultaten van verschillende bronnen combineren

3.2.1 Elke bron is een stukje van de puzzel

Het werd reeds herhaaldelijk vermeld dat wat de verschillende bronnen van externe kwaliteitszorg meten en hoe ze dat meten, verschilt van bron tot bron. Die diversiteit beschouwen de leden van de focusgroepen net als de sterkte van die verschillende bronnen. Elke bron vormt slechts één stukje van de puzzel. Het is daarom belangrijk dat het Vlaamse onderwijs beroep kan doen op verschillende bronnen met elk hun eigen doel en bijhorend onderzoeksopzet om zo de multidimensionaliteit van onderwijskwaliteit in kaart te brengen:

“Ik denk dat op systeemniveau inderdaad even goed geldt dat dat (de verschillende bronnen voor externe kwaliteitszorg, n.v.d.r.) allemaal puzzelstukjes zijn. In de zin dat we ook als overheid proberen om zoveel mogelijk informatie te verzamelen om ons te kunnen op baseren of toch te kunnen gebruiken in een besluitvormingsproces. Dat is ook de reden waarom we aan verschillende soorten onderzoeken deelnemen, omdat ze wel allemaal een licht of meer verschillend perspectief geven over dat onderwijs. Mocht dat niet zo zijn, dan zouden we er maar eentje moeten uitkiezen en [...] Dan zou je kunnen zeggen, ja we doen mee aan één onderzoek over wiskunde en dat is het dan hè, want dan weten we alles over het wiskundeonderwijs... wat uiteraard niet zo is.”

Anders gezegd, het is net waardevol dat de verschillende bronnen van externe kwaliteitszorg allemaal een verschillend perspectief belichten. Het zou bijgevolg een verkeerd signaal zijn als één specifieke bron doorslaggevend zou zijn of de norm zou worden. Elke bron heeft immers zijn eigen meerwaarde. Dat geldt niet alleen op systeemniveau, maar ook op schoolniveau. Voor scholen is het eveneens

belangrijk dat ze beroep kunnen doen op verschillende bronnen of puzzelstukjes om hun beleid vorm te geven:

“Ik vind het ook belangrijk dat ze blijven die verschillende inputs hebben. Het ene kan het andere niet vervangen, net omdat ze een ander doel hebben.”

3.2.2 De resultaten van verschillende bronnen combineren: drie sleutelfactoren

De analyse van de focusgroepen resulteerde in drie sleutelfactoren die cruciaal zijn om verschillende bronnen van externe kwaliteitszorg te combineren op zowel school- als systeemniveau. Concreet gaat het om: (1) duiding, (2) meerdere bronnen gebruiken en (3) vertaling naar de onderwijspraktijk. Hoewel de sleutelfactoren hieronder afzonderlijk toegelicht worden, hangen ze in realiteit onderling sterk samen. Ze kunnen in de praktijk dus niet strikt van elkaar onderscheiden worden.

3.2.2.1 Sleutelfactor 1: Duiding

Zoals eerder aangehaald, vormt elke bron een stukje van de puzzel. Echter, om de meerwaarde van elke bron te kunnen zien en om de resultaten van verschillende bronnen te kunnen combineren, is duiding essentieel over wat en hoe een bepaalde bron wel of niet meet. Enkel wanneer de gelijkenissen en verschillen in het onderzoeksopzet op voorhand helder zijn, is het mogelijk om de resultaten van verschillende bronnen in verhouding tot elkaar te interpreteren.

Op **schoolniveau** is het belangrijk om heel helder en toegankelijk te communiceren naar de scholen over het onderzoeksopzet van elke bron zodat het voor scholen duidelijk wordt wat ze wel en niet kunnen leren van een bepaalde bron. Die duiding is voor scholen essentieel om te kunnen inschatten wat het statuut van een specifiek schoolfeedbackrapport of doorlichtingsverslag is en om in een volgende stap met die informatie aan de slag te gaan.

“Zodat zij (de scholen, n.v.d.r.) heel helder kunnen zien van oké: TIMSS dat werkt op die manier en daar kan ik dus dat uit leren. En de doorlichting doet dat. De Vlaamse toetsen meten dat. Dus dat ja het palet in verhouding tot elkaar helder is. Dan kan een school daar ook wel iets mee en kan ze die de dingen in de juiste context plaatsen.”

De sleutelfactor ‘duiding’ speelt eveneens een rol op **systeemniveau**. Het juist begrijpen wat iets wel en niet meet, is uiteraard niet alleen relevant op schoolniveau maar ook op systeemniveau.

“Ook op systeemniveau is het belangrijk om op voorhand – voordat de resultaten er zijn – heel helder duidelijk te maken van kijk, [...] deze TIMSS-afname heeft eigenlijk dat tot doel en meet dat. En dat is iets anders dan de peilingstoets doet of dat de Vlaamse toetsen doen. Weet dus dat het resultaat daarop u iets zal zeggen over dit en niet over dat.”

3.2.2.2 Sleutelfactor 2: Meerdere bronnen gebruiken

Eén bron zegt ook niet alles. Elk van die bronnen is immers slechts een puzzelstukje. Bijgevolg is het waardevol om zowel op school- als systeemniveau beroep te doen op meerdere bronnen. De kunst bestaat erin om die verschillende bronnen of puzzelstukjes vanuit het juiste perspectief naast elkaar te leggen. Zowel op school- als systeemniveau is duiding (Sleutelfactor 1) dus essentieel om de resultaten van verschillende bronnen naast elkaar te leggen en in verhouding tot elkaar te interpreteren.

Belangrijk om te beseffen is dat op **schoolniveau** niet alleen de bronnen van externe kwaliteitszorg meegenomen moeten worden. Ook heel wat andere bronnen zoals schoolinterne data, data van smartschool, data van netgebonden toetsen... dienen meegenomen te worden in dat verhaal:

“Want er zijn er inderdaad verschillende: de toetsen in de klas, de leerlingvolgsystemen die de scholen hanteren... Dus ja, er zijn toch wel wat zaken en ik ervaar de moeilijkheid bij de scholen om die vanuit een juist perspectief samen te brengen en samen te leggen.”

Ook eerder onderzoek over het gebruik van data in functie van schoolverbeteringsstrategieën bevestigt dat meerdere databronnen gebruiken een sleutelfactor is. Het gebruik van slechts één bron geeft scholen immers weinig inzicht in mogelijke acties of verbeterstrategieën. Zo kan een peilingstoets wiskunde erop wijzen dat slechts een beperkt aandeel van de leerlingen op een bepaalde school de eindtermen bereikt. Echter, het schoolfeedbackrapport van de peilingen biedt naast de resultaten van de school weinig relevante informatie met oog op verbeteringsstrategieën. Door bijvoorbeeld die peilingsresultaten te combineren met de werkpunten in het doorlichtingsverslag en interne data kan die school mogelijks wel tot concrete actiepunten komen (Sawhill & Williamson, 2001; Schildkamp, 2019; Schildkamp et al., 2018; Schwartz et al., 2011; Sun et al., 2020).

3.2.2.3 Sleutelfactor 3: Vertaling naar de onderwijspraktijk

Hoewel vaststaat dat elke bron zijn meerwaarde heeft, is er geen wonderrecept om die bronnen te combineren op school- of systeemniveau. Eén bron interpreteren is één ding. Meerdere bronnen interpreteren vanuit het juiste perspectief en daaruit afleiden wat dat betekent voor de eigen praktijk is nog iets anders. De derde sleutelfactor verwijst dus naar de vraag: wat betekenen de verzamelde resultaten voor onze school of ons onderwijssysteem?

De vertaling van de informatie uit verschillende bronnen naar de praktijk is allesbehalve evident. Wel belangrijk is om de meerdere bronnen naast elkaar te leggen (Sleutelfactor 2: Meerdere bronnen gebruiken) wetende wat elke bron wel en niet meet (Sleutelfactor 1: Duiding). Op die manier worden die verschillende resultaten een soort informatiebundel waarop scholen en onderwijssystemen beroep kunnen doen om vervolgens te bekijken: Wat kunnen we van die bronnen leren gegeven de specifieke context van onze school of het Vlaamse onderwijssysteem? Wat zijn mogelijke acties die we kunnen ondernemen gegeven die informatiebundel en de prioriteiten van de school of het onderwijssysteem?

Er is geen stappenplan dat scholen of onderwijssystemen kunnen volgen om die vragen te beantwoorden. Om toch tot antwoorden te komen, kan het helpen om in dialoog te gaan over de gecombineerde resultaten. Op die manier kan er samen gezocht worden naar verklaringen voor de resultaten, verbanden tussen de resultaten van verschillende bronnen, de werk- en sterke punten en mogelijke verbeterstrategieën:

“Wij geven zeer expliciet de aanbeveling aan de directeur, zorgcoördinator [...] om over de resultaten in dialoog te gaan. Niet alleen met de collega’s, maar waar mogelijk ook met de leerlingen zelf en de ouders van op school om eens te kijken van ja voelen jullie dat ook zo aan dat we goed bezig zijn of dat er nog werkpunten zijn. En om dat [...] mee te nemen naar de pedagogische begeleidingsdiensten om er echt mee aan de slag te gaan.”

De vertaling naar de onderwijspraktijk is een proces dat tijd nodig heeft. Zo moeten scholen ruimte krijgen om dingen uit te proberen, te evalueren, waar nodig bij te sturen en goede elementen te borgen. Louter de beschikbaarheid van data leidt dus niet meteen tot een sterkere onderwijskwaliteit:

“Het is niet omdat een school een feedbackrapport ontvangt dat we daar onmiddellijk resultaat van kunnen zien.”

3.2.3 Uitdagingen op schoolniveau

Uit de focusgroepen blijkt dat het voor scholen niet evident is om de resultaten van verschillende bronnen te combineren en interpreteren. Er zijn heel wat uitdagingen die scholen tegenkomen wanneer ze aan de slag gaan met de resultaten van verschillende bronnen:

- **De interpretatie van schoolfeedbackrapporten**

Verschillende participanten geven aan dat de interpretatie van schoolfeedbackrapporten niet evident is voor scholen. Opvallend is wel dat het doorlichtingsverslag toegankelijker is dan de schoolfeedbackrapporten bij de peilingen en de internationaal vergelijkende onderzoeken. Vooral de interpretatie van kwantitatieve data is niet eenvoudig voor scholen.

In de focusgroepen wordt een voorbeeld uit de onderwijspraktijk aangehaald: een school heeft deelgenomen aan internationaal vergelijkend onderzoek en ontvangt een schoolfeedbackrapport van dat onderzoek. De resultaten in het schoolfeedbackrapport zijn vrij goed. De directeur is nochtans van mening dat de processen op school onvoldoende kwaliteitsvol zijn. Ze besluit om de resultaten van dat schoolfeedbackrapport niet te delen met de leerkrachten op school, omdat ze denkt dat die resultaten haar gevoel tegenspreken.

Het is een misvatting van de directeur om te denken dat een schoolfeedbackrapport van een internationaal vergelijkend onderzoek hoofdzakelijk iets vertelt over de processen op school. Zo'n schoolfeedbackrapport koppelt terug over de resultaten die leerlingen van haar school behaalden op de toets en zegt weinig tot niets over de processen op school.

Nochtans is die interpretatie essentieel om op schoolniveau aan de slag te kunnen met de resultaten van verschillende bronnen. Als scholen niet weten wat elke bron meet en niet meet – zoals de directeur uit het voorbeeld – , kunnen ze onmogelijk de resultaten van verschillende bronnen combineren en interpreteren vanuit het juiste perspectief:

“Zeker op schoolniveau wil je van die data kunnen gebruik maken om aan kwaliteitsontwikkeling binnen die school te doen. Als je dat dan niet gecontextualiseerd en vanuit een juiste interpretatie kan doen, dan ben je verkeerd bezig.”

Dat wijst nogmaals op het belang van een goede duiding van wat elke bron wel en niet meet (Sleutelfactor 1: Duiding). Om scholen te ondersteunen bij de interpretatie van schoolfeedbackrapporten is het bovendien erg belangrijk dat de rapporten zelf toegankelijk worden opgesteld. Dat betekent dat er toegankelijk en helder gecommuniceerd wordt over het onderzoeksopzet van de bron en hoe je de resultaten van de bron wel en niet kan lezen. Bovendien moeten de resultaten heel bevattelijk en duidelijk worden gepresenteerd. Dat helpt scholen om die bronnen juist te interpreteren en daarbij ook rekening te houden met het specifieke onderzoeksopzet van die bron. Vervolgens kunnen scholen die informatie gericht inzetten met het oog op een rijke kwaliteitsontwikkeling.

- **Tijd en ruimte om te combineren en te vertalen**

Zoals eerder aangehaald, is de vertaling naar de onderwijspraktijk een proces dat tijd nodig heeft (zie 'Sleutelfactor 3: Vertaling naar de onderwijspraktijk'). In de realiteit komen de resultaten van bronnen van externe kwaliteitszorg bovendien op zeer gevarieerde tijdstippen binnen bij een school. Op dat moment wordt vaak niet meteen die puzzel gelegd wegens gebrek aan tijd en ruimte. Nochtans is het erg waardevol om de resultaten van verschillende bronnen te combineren en op basis daarvan een vertaling te maken naar de onderwijspraktijk:

“Meestal is dat (resultaten combineren en vertalen naar de eigen praktijk, n.v.d.r.) iets wat er een beetje tussen de soep en de patatten bij komt, maar door systematisch tijd in te roosteren om dit in gesprek te doen samen met de pedagogische begeleidingsdienst, voelt dat als een absolute meerwaarde.”

- **Omgaan met divergerende resultaten**

Wanneer de resultaten van verschillende bronnen convergeren, is het meestal eenvoudiger om de puzzel te leggen dan wanneer de verschillende puzzelstukjes divergeren of elkaar zelfs tegen spreken. De leden van de focusgroep bevestigen dat het in de praktijk wel degelijk voorkomt dat scholen divergerende of tegenstrijdige resultaten krijgen van verschillende bronnen. Het combineren en vertalen van die resultaten is echter niet evident. Het kan voor scholen heel verwarrend zijn als ze in het ene onderzoek een bepaald resultaat behalen en in een ander onderzoek een totaal ander resultaat behalen.

“En zeker in het geval een school toch tegenstrijdige resultaten zou krijgen uit verschillende bronnen, wat het eens zo moeilijk maakt, hè. Als je convergerende resultaten krijgt, als je ziet in verschillende onderzoeken van ja we vallen precies toch wat uit of we doen het altijd vrij goed, dan denk ik dat een school wel weet wat dat ze wel of niet moeten doen. Maar als die nu... ja tegenstrijdige feedback krijgt, denk ik dat het wel belangrijk is [...] dat die dat kunnen plaatsen en weten van hoe zou dat nu kunnen?”

De vraag is dan: hoe kunnen scholen omgaan met divergerende resultaten? Uit de focusgroepen blijkt dat de sleutelfactoren ook essentieel zijn bij het combineren van divergerende resultaten. Het is enkel mogelijk om die divergerende resultaten te interpreteren als een school zicht heeft op wat elke bron wel en niet meet. Wat meet het ene onderzoek en wat meet het andere? Hoe kan het dan dat die twee onderzoeken tot divergerende resultaten komen? Door de sterke en minder sterke kanten van de bronnen te kennen (Sleutelfactor 1: Duiding) is het mogelijk om meerdere bronnen naast elkaar te leggen en een interpretatie te geven aan die divergerende resultaten (Sleutelfactor 2: Meerdere bronnen):

“Ik denk dat het moeilijk kan zijn voor een school als ze [...] echt tegenstrijdige informatie krijgt. Maar dat je dan wel moeten weten: wat betekent dit onderzoek? Wat betekent dat ander onderzoek? En hoe kan dat dan dat ze elkaar toch tegenspreken, bijvoorbeeld?”

Die interpretatie vormt vervolgens het vertrekpunt om in de context van die school of dat onderwijssysteem naar verklaringen en oplossingen te zoeken (Sleutelfactor 3: Vertaling naar de onderwijspraktijk).

- **De beschikbaarheid van een enorme hoeveelheid data**

De hoeveelheid data die ter beschikking gesteld wordt aan scholen is aanzienlijk. De bronnen van externe kwaliteitszorg zijn namelijk slechts een fractie van de hele berg aan beschikbare

data voor scholen. Die grote hoeveelheid maakt het erg complex voor scholen en bemoeilijkt de interpretatie en combinatie van de resultaten.

De uitdaging bestaat eruit om scholen wegwijs te maken in die enorme berg aan data zodat het voor scholen een overzichtelijk en helder verhaal is – wat op dit moment niet het geval is. Dat kan bijvoorbeeld door op één plaats alle beschikbare data te verzamelen op een ordelijke en begrijpelijke manier en daarbij meteen alle noodzakelijke duiding over die data mee te geven:

“Daar zou een initiatief genomen kunnen worden wat de twee kan combineren, hé... waar dat je de data hebt en waar je ook de juiste informatie en duiding [...] onmiddellijk mee krijgt.”

- **Nood aan begeleiding en datageletterdheid**

Kortom, er zijn heel wat uitdagingen die scholen tegenkomen bij het combineren van verschillende bronnen. Het begrijpen en lezen van de informatie in een schoolfeedbackrapport of doorlichtingsverslag is slechts de eerste stap en vaak loopt het al mis bij die eerste stap. Het is daarom noodzakelijk om te werken aan datageletterdheid bij leraren en schoolleiders.

Hoewel de focusgroepen niet aangaven hoe er concreet gewerkt kan worden aan de datageletterdheid van leraren en schoolleiders, biedt de literatuur hier aanknopingspunten. Een professionaliseringstraject datageletterdheid moet enerzijds inzetten op het aanreiken van concrete handvatten – bijvoorbeeld een gestructureerd stappenplan van een onderzoekscyclus – om aan de slag te gaan met data. Anderzijds is opleiding over statistische terminologie (zoals toegevoegde waarde, significantie...) en het lezen van visueel weergegeven resultaten nodig om de resultaten correct te kunnen interpreteren (Schildkamp et al., 2013).

Daarnaast benadrukken de focusgroepen dat scholen beroep moeten kunnen doen op pedagogische begeleidingsdiensten om hen te ondersteunen bij de interpretatie en combinatie van verschillende bronnen enerzijds en bij de vertaling van de informatie uit verschillende bronnen naar de praktijk anderzijds. Van belang is wel dat de school steeds het eigenaarschap behoudt. Dat betekent dat de school zelf bepaalt hoe ze met de resultaten omgaat. De focusgroepen benadrukken dat scholen het meest bereiken op vlak van kwaliteitszorg wanneer ze eigenaar zijn van hun data en daarmee durven experimenteren. Er is echter ook een keerzijde aan dat verhaal. Er is geen garantie dat scholen ook effectief aan de slag gaan met die verschillende bronnen:

“In sommige scholen gebeurt er ook niets met zo’n rapporten.”

3.2.4 Uitdagingen op systeemniveau

Ook op systeemniveau situeren zich enkele uitdagingen met betrekking tot het combineren van verschillende bronnen van externe kwaliteitszorg:

- **Omgaan met divergerende resultaten**

Niet alleen op schoolniveau, maar ook op systeemniveau zijn divergerende resultaten mogelijk. De uitdaging met betrekking tot divergerende resultaten situeert zich dus zowel op het school- als systeemniveau:

“Elk van die onderzoeken heeft zijn eigen sterke en minder sterke kanten waardoor dat de resultaten ook inderdaad allemaal puzzelstukken zijn die je moet trachten samen te leggen. En soms botst dat bij ons (op systeemniveau, n.v.d.r.) ook, hè. Wij krijgen soms

ook puzzelstukjes waarvan we denken, ja die komen nu toch echt uit een andere doos, hè. [...] Ik denk dat de struggles die je op schoolniveau soms kan voelen van ik heb hier conflicterende informatie... dat op systeemniveau eigenlijk net hetzelfde gebeurt, toch op zijn minst kan gebeuren."

De combinatie van divergerende resultaten op systeemniveau is allesbehalve evident. Door een goed zicht te hebben op het doel en bijhorend onderzoeksopzet van elke bron is het toch mogelijk om die divergerende resultaten te interpreteren. Ook op systeemniveau is het dus essentieel om de sterke en minder sterke kanten van de verschillende bronnen te kennen (Sleutelfactor 1: Duiding).

- **Ontbrekende puzzelstukjes**

De leden van de focusgroep wijzen erop dat er ook een aantal zaken zijn waar we in Vlaanderen voorlopig nog geen zicht op hebben. Meer specifiek gaat het om de stabiliteit van indicatoren enerzijds en de link tussen de toetsen en de processen in de klas anderzijds.

Ten eerste geven de leden van de focusgroep aan dat er op dit moment weinig tot geen Vlaams onderzoek is over de stabiliteit van indicatoren. Er wordt bijvoorbeeld geen peilings- of internationale toets op dezelfde manier bij dezelfde scholen afgenomen tijdens opeenvolgende schooljaren om te gaan kijken in welke mate de resultaten stabiel zijn over tijd heen. Het is niet omdat er het ene jaar een bepaald resultaat gevonden wordt dat hetzelfde resultaat een jaar later ook vastgelegd wordt:

"Ik denk wel dat als we die stabiliteit niet zouden vaststellen, dan zitten we eigenlijk wel met een probleem. Ook in termen van: wat betekent dat dan precies voor opvolging en voor mogelijke remediërings- of begeleidingstrajecten?"

Een tweede ontbrekend puzzelstukje gaat over de link tussen de toetsen van peilingen of internationaal vergelijkend onderzoek met de processen in de klas. De peilingen en internationale vergelijkende onderzoeken focussen hoofdzakelijk op vaardigheden en prestaties van leerlingen, terwijl uitspraken over de processen in de klas op basis van die onderzoeken samengaan met meer beperkingen (zie Matrix). Aan de andere kant zijn er de doorlichtingen die sterk focussen op proceskenmerken, maar de onderwijsinspectie neemt geen toetsen af bij leerlingen:

"Vanuit het perspectief van die internationaal vergelijkende onderzoeken of peilingen bijvoorbeeld op zoek gaan naar die oorzakelijke factoren, dat vergt inzicht in de processen. En wat de onderwijsinspectie doet, geeft wel inzicht in de processen maar niet gelinkt of gelieerd op dezelfde manier aan die outputindicatoren zoals die in die onderzoeken aan bod komen. En die link kunnen we op dit moment natuurlijk niet maken."

Met andere woorden, er is op dit moment geen bron die inzicht geeft in zowel processen als outputkenmerken enerzijds en de link maakt tussen beiden anderzijds.

- **Nood aan datageletterdheid en duiding**

Niet enkel op schoolniveau is het werken aan datageletterdheid nodig. Ook op systeemniveau is dat een uitdaging:

"Los daarvan denk ik dat het niet enkel die onderzoekende houding en die datageletterdheid bij scholen is waaraan gewerkt kan worden, maar zeker ook bij de pedagogische begeleidingsdiensten zelf, ook bij de overheid, ook bij de inspectie... Ik

denk dat dat voor iedereen wel geldt. Dat is niet enkel de blik naar scholen toe, want zij kunnen ook niet alles oplossen.”

Opvallend is ook dat enkele leden van de focusgroepen aangegeven dat ze voorafgaand aan de focusgroep nog nooit eerder een overzicht gekregen hebben van wat elke bron van externe kwaliteitszorg exact meet en hoe ze dat meet. Nochtans blijkt uit de analyses dat duiding over het onderzoeksopzet een cruciaal element is om die verschillende bronnen te kunnen combineren. Bijgevolg is professionalisering die zich daarop richt niet alleen op schoolniveau, maar ook op systeemniveau aangewezen.

3.3 Een toekomst met Vlaamse toetsen

Tijdens de focusgroepen stond de combinatie van resultaten van de doorlichtingen, de peilingen en het internationaal vergelijkend onderzoek centraal. Vanaf schooljaar 2023-2024 wordt het peilingsonderzoek vervangen door de Vlaamse toetsen. Daarom hebben we tijdens de focusgroepen kort aandacht besteed aan een toekomst met Vlaamse toetsen.

Uit de analyse van de focusgroepen blijkt dat bovenstaande resultaten ook relevant zijn voor een toekomst waarin het peilingsonderzoek vervangen wordt door de Vlaamse toetsen:

- Elke bron blijft slechts één stukje van de puzzel.
- De drie sleutelfactoren zijn cruciaal om verschillende bronnen van externe kwaliteitszorg te combineren, ongeacht of het gaat om de resultaten van de peilingen of de resultaten van de Vlaamse toetsen.
- De besproken uitdagingen op school- en systeemniveau blijven ook in een toekomst met Vlaamse toetsen pertinent.

Daarnaast wijzen de leden van de focusgroepen op enkele aandachtspunten. Ten eerste verliezen we ook indicatoren doordat de peilingen verdwijnen. Bij de peilingen werden er meerdere domeinen bevraagd, terwijl de Vlaamse toetsen enkel focussen op wiskunde en Nederlands. Daarom pleiten de focusgroepen ervoor om de puzzel zo rijk en divers mogelijk te houden in de toekomst.

Ten tweede geven de leden van de focusgroepen aan dat duiding (Sleutelfactor 1) nog belangrijker zal worden met de komst van de Vlaamse toetsen:

“Ik denk scholen die het goed doen op de centrale toetsen... die zullen al snel gaan schermen met het idee van kijk, dit is een objectieve mening... dus wat jullie van de inspectie zeggen ja, dat doet er eigenlijk niet meer zo heel hard toe, hè. Ik denk dat dat een groot risico is. [...] Dus ik denk dat nog meer dan nu het heel belangrijk zal zijn om te gaan kaderen wat er in zo'n feedbackrapport bij een centrale toets staat en waar dat wel en niet voor kan dienen.”

Ten derde benadrukken beide focusgroepen dat het belangrijk is om steeds te controleren voor achtergrondkenmerken. Het gaat immers over hoeveel extra leerwinst een school creëert en niet over welke instroom van leerlingen die school heeft.

Tot slot wijzen de focusgroepen op enkele kansen. Zo wordt het in de toekomst eenvoudiger om de concordantie tussen de resultaten van de bronnen van externe kwaliteitszorg in kaart te brengen. De Vlaamse toetsen worden – in tegenstelling tot de peilingen – bij alle leerlingen in alle scholen afgenomen. Dat gaat het eenvoudiger maken om de concordantie met de internationaal vergelijkende onderzoeken en de doorlichtingen te onderzoeken, omdat er altijd gegevens beschikbaar zijn van alle scholen met betrekking tot de Vlaamse toetsen. Dat schept mogelijkheden voor verder onderzoek en verdere samenwerking tussen onderzoekers van de Vlaamse toetsen, de internationaal vergelijkende onderzoeken en de onderwijsinspectie.

Referenties

- Aesaert, K., & Janssen, R. (2018). *Peiling Nederlands (Begrijpend lezen) in het basisonderwijs 2007-2013: rapport trendanalyses*. KU Leuven, Steunpunt Toetsontwikkeling en Peilingen.
- Aesaert, K., Van Nijlen, D., Ameel, E., De Meyst, M., Denis, J., Laenen, I., Becu, A., Detollenaere, K., & Janssen, R. (2017). *De constructie van een peilingsinstrument Nederlands in het basisonderwijs - Verslag vooronderzoeken*. KU Leuven, Steunpunt Toetsontwikkeling en Peilingen.
- Aesaert, K., Van Nijlen, D., Ameel, E., De Meyst, M., Denis, J., Laenen, I., Becu, A., Detollenaere, K., & Janssen, R. (2018). *De constructie van een peilingsinstrument Nederlands in het basisonderwijs - Eindrapport Deel 1*. KU Leuven, Steunpunt Toetsontwikkeling en Peilingen.
- AHOVOKS. (n.d.). *Onderwijsdoelen modernisering*. <https://www.onderwijsdoelen.be/modernisering>
- AHOVOKS. (2021). *Peilingen: Wat en waarom?* <https://eindoelen.be/wat-en-waarom>
- Appels, L., Merckx, B., Faddar, J., Van Petegem, P., De Maeyer, S., Boeve-de Pauw, J., & Delrue, K. (2020). *TIMSS 2019 Schoolfeedbackrapport*. UAntwerpen en Arteveldehogeschool.
- Baker, E. L., & Linn, R. L. (2002). *Validity Issues for Accountability Systems*. National Center for Research on Evaluation, Standards and Student Testing, University of California.
- Blömeke, S., & Gustafsson, J. (2017). *Standard Setting in Education*. Springer International Publishing.
- Blondeel, A., Moens, K., Warlop, N., & De Meyer, I. (2016). *Wetenschappelijke geletterdheid bij 15-jarigen. Vlaams rapport PISA2015*. Vakgroep Onderwijskunde, Universiteit Gent.
- Brookhart, S. (2009). The Many Meanings of "Multiple Measures." *Educational Leadership*, 67(3), 6–12.
- Byrne, B., & De Vijver, F. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, 10(2), 107–132.
- Carpentier, N., Costers, S., Janssen, R., & Willem, L. (2019a). *Peiling wiskunde in de eerste graad van het secundair onderwijs A-stroom - Eindrapport*. KU Leuven, Steunpunt Toetsontwikkeling en Peilingen.
- Carpentier, N., Costers, S., Janssen, R., & Willem, L. (2019b). *Peiling wiskunde in de eerste graad van het secundair onderwijs A-stroom 2018 - Brochure*. KU Leuven, Steunpunt Toetsontwikkeling en Peilingen.
- Carpentier, N., Costers, S., Janssen, R., & Willem, L. (2020a). *Peiling wiskunde in de eerste graad van het secundair onderwijs B-stroom - Eindrapport*. KU Leuven, Steunpunt Toetsontwikkeling en Peilingen.
- Carpentier, N., Costers, S., Janssen, R., & Willem, L. (2020b). *Peiling wiskunde in de eerste graad van het secundair onderwijs B-stroom 2019 - Brochure*. KU Leuven, Steunpunt Toetsontwikkeling en Peilingen.
- Centrum voor Onderwijseffectiviteit en -evaluatie. (2009). *Peiling wiskunde in de eerste graad secundair onderwijs (B-stroom)*. Vlaams Ministerie van Onderwijs en Vorming.
- Centrum voor Schoolfeedback. (2016). *Schoolfeedbackrapport Begrijpend lezen*. KU Leuven.

- Chester, M. D. (2003). Multiple Measures and High-Stakes Decisions: A Framework for Combining Measures. *Educational Measurement: Issues and Practice*, 22(2), 32–41.
- Chester, M. D. (2005). Making Valid and Consistent Inferences About School Effectiveness From Multiple Measures. *Educational Measurement: Issues and Practice*, 24(4), 40–52.
- Codex Secundair Onderwijs. (2010). <https://data-onderwijs.vlaanderen.be/edulex/document.aspx?docid=14289>
- Cyr, J. (2019). *Focus Groups for the Social Science Researcher*. Cambridge University Press.
- De Meyer, I., Janssens, R., Warlop, N., Van Keer, H., De Wever, B., & Valcke, M. (2019). *Leesvaardigheid van 15-jarigen. Vlaams rapport PISA2018*. Vakgroep Onderwijskunde, Universiteit Gent.
- De Meyer, I., & Warlop, N. (2010). *Leesvaardigheid van 15-jarigen in Vlaanderen. De eerste resultaten van PISA 2009*. Vakgroep Onderwijskunde, Universiteit Gent.
- De Meyer, I., Warlop, N., & Van Camp, S. (2013). *Wiskundige geletterdheid bij 15-jarigen. Vlaamse resultaten van PISA2012*. Vakgroep Onderwijskunde.
- Decreet Basisonderwijs. (1997). <https://codex.vlaanderen.be/Zoeken/Document.aspx?DID=1005384¶m=inhoud>
- Decreet betreffende de kwaliteit van onderwijs. (2009). <https://data-onderwijs.vlaanderen.be/edulex/document.aspx?docid=14129>
- Denis, J., Talloen, W., Laenen, I., Janssen, R., & Aesaert, K. (2019a). *Peiling Nederlands in het basisonderwijs - Eindrapport*. KU Leuven en UAntwerpen, Steunpunt Toetsontwikkeling en Peilingen.
- Denis, J., Talloen, W., Laenen, I., Janssen, R., & Aesaert, K. (2019b). *Peiling Nederlands lezen, luisteren en schrijven in het basisonderwijs 2018 - Brochure*. Steunpunt Toetsontwikkeling en Peilingen.
- Dockx, J., Van Landeghem, G., Aesaert, K., Van Damme, J., & De Fraine, B. (2019). *Begrijpend lezen van het vierde naar het zesde leerjaar. Herhalingsmeting van PIRLS in 2018 vergeleken met PIRLS 2016*. Centrum voor Onderwijseffectiviteit en -evaluatie, KU Leuven.
- Dumay, X., Coe, R., & Anumendem, D. N. (2013). Stability over time of different methods of estimating school performance. *School Effectiveness and School Improvement*, 1–19.
- Faddar, J., Appels, L., Merckx, B., Boeve-de-Pauw, J., Delrue, K., De Maeyer, S., & Van Petegem, P. (2020). *Vlaanderen in TIMSS 2019. Wiskunde- en wetenschapsprestaties van het vierde leerjaar in internationaal perspectief en doorheen de tijd*. Universiteit Antwerpen.
- Gebhardt, E., & Adams, R. (2007). The influence of equating methodology on reported trends in PISA. *Journal of Applied Measurement*, 8(3), 305.
- Glas, C., & Jehangir, K. (2013). Modeling Country-Specific Differential Item Functioning. In *Handbook of International large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 97–115). Chapman and Hall/CRC.
- Graham, J. W., Olchowski, A. E., & Gilreath, T. D. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, 8, 206–213.

- Heine, J., & Robitzsch, A. (2022). Evaluating the effects of analytical decisions in large-scale assessments: analyzing PISA mathematics 2003-2012. *Large-Scale Assessments in Education*, 10(1), 1–29.
- Hough, H., Penner, E., & Witte, J. (2016). *Identity Crisis: Multiple Measures and the Identification of Schools under ESSA*. Stanford University, Policy Analysis for California Education.
- Janssen, R., Luyten, B., & Ver Eecke, E. (2006). *Peiling informatieverwerving en -verwerking in de eerste graad secundair onderwijs (A-stroom)*. Vlaamse overheid.
- Janssen, R., & Ver Eecke, E. (2004). *Eerste peiling wiskunde en lezen in het basisonderwijs*. Ministerie van de Vlaamse Gemeenschap.
- Janssens, R., Van Hulle, M., Ver Eecke, E., & Wouters, M. (2007). *Peiling natuur (wereldoriëntatie) in het basisonderwijs*. Vlaamse overheid.
- Kane, M. (2017). Loosening psychometric constraints on educational assessments. *Assessment in Education: Principles, Policy & Practice*, 24(3), 447–453.
- Karantonis, A., & Sireci, S. (2006). The Bookmark standard-setting method: A literature review. *Educational Measurement: Issues and Practice*, 25(1), 4–12.
- Kitzinger, J. (1995). Qualitative Research: Introducing focus groups. *British Medical Journal*, 311(7000), 299–302.
- Krueger, R. (2015). *Moderating Focus Groups*. Sage Publications.
- Krueger, R., & Casey, M. (2015). *Focus Groups: A Practical Guide for Applied Research* (5th Edition). Sage Publications.
- Kutner, H., Nachtsheim, C., & Neter, J. (2005). *Applied linear statistical models* (5th edition). McGraw-Hill/Irwin.
- Leckie, G., & Prior, L. (2022). A Comparison of Value-Added Models for School Accountability. *School Effectiveness and School Improvement*, 1–25.
- Lenkeit, J. (2013). Effectiveness measures for cross-sectional studies: a comparison of value-added models and contextualised attainment models. *School Effectiveness and School Improvement*, 24(1), 1–25.
- Lobe, B. (2017). Best Practices for Synchronous Online Focus Groups. In R. Barbour & D. Morgan (Eds.), *A New Era in Focus Group Research. Challenges, Innovation and Practice* (pp. 227–250). Palgrave Macmillan London.
- MacCann, R., & Stanley, G. (2006). The use of Rasch modeling to improve standard setting. *Practical Assessment, Research, and Evaluation*, 11(1), 2.
- Martin, M., Mullis, I., & Chrostowski, S. (2004). *TIMSS 2003 Technical Report*. TIMSS & PIRLS International Study Center, IEA.
- Martin, M., Mullis, I., & Foy, P. (2012). Estimating linking error in PIRLS. In *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments* (pp. 35–47).
- Martin, M., Mullis, I., & Hooper, M. (2017). *Methods and Procedures in PIRLS 2016*. TIMSS & PIRLS International Study Center, IEA.

- Martin, M., Mullis, I., & Kennedy, A. (2003). *PIRLS 2001 Technical report*. International Study Center, IEA.
- Martin, M., Mullis, I., & Kennedy, A. (2007). *PIRLS 2006 Technical Report*. TIMSS & PIRLS International Study Center, IEA.
- Martin, M., Von Davier, M., & Mullis, I. (2020). *Methods and Procedures: TIMSS 2019 Technical Report*. TIMSS & PIRLS International Study Center, IEA.
- Martínez, J. F., Schweig, J., & Goldschmidt, P. (2016). Approaches for Combining Multiple Measures of Teacher Performance: Reliability, Validity, and Implications for Evaluation Policy. *Educational Evaluation and Policy Analysis, 38*(4), 738–756.
- McHugh, M. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica, 22*(3), 276–282.
- Miles, M., & Huberman, A. (2014). *Qualitative Data Analysis: An Expanded Sourcebook*. Sage Publications.
- Morgan, D. (2013). *The Focus Group Guidebook*. Sage Publications.
- Morgan, D. (2017). Conclusions: A Call for Further Innovations in Focus Groups. In R. Barbour & D. Morgan (Eds.), *A New Era in Focus Group Research. Challenges, Innovation and Practice* (pp. 411–420). Palgrave Macmillan London.
- Mullis, I., Kennedy, A., Martin, M., & Sainsbury, M. (2006). *PIRLS 2006 Assessment Framework and Specifications*. TIMSS & PIRLS International Study Center, IEA.
- Mullis, I., & Martin, M. (2015). *PIRLS 2016 Assessment Framework*. TIMSS & PIRLS International Study Center, IEA.
- Mullis, I., & Martin, M. (2017). *TIMSS 2019 Assessment Frameworks*. TIMSS & PIRLS International Study Center, IEA.
- Mullis, I., Martin, M., Foy, P., & Drucker, K. (2012). *PIRLS 2011 International Results in Reading*. TIMSS & PIRLS International Study Center, IEA.
- Mullis, I., Martin, M., Foy, P., & Hooper, M. (2017). *PIRLS 2016 International Results in Reading*. TIMSS & PIRLS International Study Center, IEA.
- Mullis, I., Martin, M., Foy, P., Kelly, D., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. TIMSS & PIRLS International Study Center, IEA.
- Mullis, I., Martin, M., Kennedy, A., & Foy, P. (2007a). *PIRLS 2006 International Report*. TIMSS & PIRLS International Study Center, IEA.
- Mullis, I., Martin, M., Kennedy, A., & Foy, P. (2007b). *PIRLS 2006 International Report*. TIMSS & PIRLS International Study Center, IEA.
- Mullis, I., Martin, M., & Loveless, T. (2016). *20 Years of TIMSS*. TIMSS & PIRLS International Study Center, IEA.
- NVAO. (2021). *Kwaliteitszorgstelsel Vlaanderen*. <https://www.nvao.net/nl/kwaliteitszorgstelsel-vlaanderen>
- OECD. (n.d.). *About OECD*. <https://www.oecd.org/about/>

OECD. (2015). *PISA 2018 Technical Standards*. OECD Publishing.

OECD. (2019a). *PISA 2018 Assessment and Analytical Framework*. OECD Publishing.

OECD. (2019b). *PISA 2018 Results (Volume I): What Students Know and Can Do*. OECD Publishing.

OECD. (2019c). *PISA 2018 Results (Volume II): Where All Students Can Succeed*. OECD Publishing.

OECD. (2021). *PISA 2018 Technical Report*. OECD Publishing.

Onderwijsinspectie. (n.d.). *Het referentiekader voor onderwijskwaliteit*.
<https://onderwijsinspectie.be/nl/het-ok>

Onderwijsinspectie. (2010). *Onderwijsspiegel 2008-2009*.

Onderwijsinspectie. (2011). *Onderwijsspiegel schooljaar 2009-2010*.

Onderwijsinspectie. (2012). *Onderwijsspiegel 2012*.

Onderwijsinspectie. (2013). *Onderwijsspiegel 2013*.

Onderwijsinspectie. (2014). *Onderwijsspiegel 2014*.

Onderwijsinspectie. (2015). *Onderwijsspiegel 2015*.

Onderwijsinspectie. (2016). *Onderwijsspiegel 2016*.

Onderwijsinspectie. (2017a). *Bronnendocument referentiekader voor onderwijskwaliteit*.
www.mijnschoolisok.be

Onderwijsinspectie. (2017b). *Onderwijsspiegel 2017*.

Onderwijsinspectie. (2018a). *Doorlichten in dialoog. Raamtekst Inspectie 2.0*.

Onderwijsinspectie. (2018b). *Onderwijsspiegel 2018*.

Onderwijsinspectie. (2018c). *Het concept Inspectie 2.0*. <https://onderwijsinspectie.be/nl/het-concept-inspectie-20>

Onderwijsinspectie. (2021a). *Informatie over de doorlichting: gewoon basisonderwijs*.

Onderwijsinspectie. (2021b). *Informatie over de doorlichting: gewoon secundair onderwijs*.

Onderwijsinspectie. (2021c). *Ontwikkelingsschalen onderwijsleerpraktijk: gewoon basisonderwijs*.

Onderwijsinspectie. (2021d). *Ontwikkelingsschalen onderwijsleerpraktijk: gewoon secundair onderwijs*.

Onderwijsinspectie. (2021e). *Beoordelingen hbo5-opleidingen Verpleegkunde*.
<https://www.onderwijsinspectie.be/nl/beoordelingen-hbo5-opleiding-verpleegkunde>

Onderwijsinspectie. (2021f). *Doorlichten*. <https://onderwijsinspectie.be/nl/doorlichten>

Onderwijsinspectie. (2021g). *Het advies*. <https://onderwijsinspectie.be/nl/het-advies>

Onderwijsinspectie. (2021h). *Onderwijsspiegel 2021*.

Onderwijsinspectie. (2021i). *Onderzoek kwaliteitsontwikkeling*.
<https://onderwijsinspectie.be/nl/onderzoek-kwaliteitsontwikkeling>

- Porter, A. C., Chester, M. D., & Schlesinger, M. D. (2004). Framework for an Effective Assessment and Accountability Program: The Philadelphia Example. *Teachers College Record*, 106(6), 1358–1400.
- Ravitch, S., & Mittenfelner Carl, N. (2016). *Qualitative Research. Bridging the Conceptual, Theoretical and Methodological*. Sage Publications.
- Rekenhof. (2011). *Verslag over het toezicht op de kwaliteit van het onderwijs door de inspectie*.
- Robitzsch, A., & Lüdtke, O. (2019). Linking errors in international large-scale assessment: Calculation of standard errors for trend estimation. *Assessment in Education: Principles, Policy & Practice*, 26(4), 444–465.
- Robitzsch, A., & Lüdtke, O. (2020). A review of different scaling approaches under full invariance, partial invariance, and nonvariance for cross-sectional country comparisons in large-scale assessments. *Psychological Test and Assessment Modeling*, 62(2), 233–279.
- Robitzsch, A., & Lüdtke, O. (2022). Some thoughts on analytical choices in the scaling model for test scores in international large-scale assessment studies. *Measurement Instruments for the Social Sciences*, 4(1), 1–20.
- Sachse, K., & Haag, N. (2017). Standard errors for national trends in international large-scale assessments in the case of cross-national differential item functioning. *Applied Measurement in Education*, 30(2), 102–116.
- Sachse, K., Roppelt, A., & Haag, N. (2016). A comparison of linking methods for estimating national trends in international comparative large-scale assessments in the presence of cross-national DIF. *Journal of Educational Measurement*, 53(2), 152–171.
- Sawhill, J. C., & Williamson, D. (2001). Mission Impossible? Measuring Success in Nonprofit Organizations. *Nonprofit Management and Leadership*, 11(3), 371–386.
- Schafer, W. D. (2003). A State Perspective on Multiple Measures in School Accountability. *Educational Measurement: Issues and Practice*, 22(2), 27–31.
- Schildkamp, K. (2019). Data-based decision-making for school improvement: Research insights and gaps. *Educational Research*, 61(3), 257–273.
- Schildkamp, K., Handelzalts, A., Poortman, C., Leusink, H., Meerdink, M., Smit, M., Ebbeler, J., & Hubers, M. (2018). *The Data Team™ Procedure: A Systematic Approach to School Improvement*. Springer Texts in Education.
- Schildkamp, K., Lai, M., & Earl, L. (2013). *Data-based Decision Making in Education. Challenges and Opportunities*. Springer.
- Schwartz, H., Hamilton, L., Stecher, B., & Steele, J. (2011). *Expanded Measures of School Performance*. Rand Corporation.
- Spikic, S., Goos, M., Denis, J., Costers, S., Janssen, R., & Van Renterghem, K. (2022a). *Peiling wiskunde in het basisonderwijs - Eindrapport*. KU Leuven, Steunpunt Toetsontwikkeling en Peilingen.
- Spikic, S., Goos, M., Denis, J., Costers, S., Janssen, R., & Van Renterghem, K. (2022b). *Peiling wiskunde in het basisonderwijs 2021 - Brochure*. KU Leuven, Steunpunt Toetsontwikkeling en Peilingen.
- Steunpunt Toetsontwikkeling en Peilingen. (n.d.). *Projectverloop peilingsonderzoek*. <https://peilingsonderzoek.be/verloop/>

- Steunpunt Toetsontwikkeling en Peilingen. (2019). *Individuele leerlingresultaten paralleltoetsen wetenschappen en techniek basisonderwijs 2019*.
- Steunpunt Toetsontwikkeling en Peilingen. (2021). *Stappenplan voor het gebruik van schoolfeedback van peilingen en paralleltoetsen*. <https://peilingsonderzoek.be/schoolfeedback/>
- Steunpunt Toetsontwikkeling en Peilingen. (2022). *Peilingsonderzoek*. <https://peilingsonderzoek.be/>
- Sun, M., Kennedy, A., & Anderson, E. (2020). The Multidimensionality of School Performance: Using Multiple Measures for School Accountability and Improvement. *Education Policy Analysis Archives*, 28(89).
- Tielemans, K., Vandenbroeck, M., Bellens, K., Van Damme, J., & De Fraine, B. (2017). *Het Vlaams lager onderwijs in PIRLS 2016. Begrijpend lezen in internationaal perspectief en in vergelijking met 2006*. Centrum voor Onderwijseffectiviteit en -evaluatie, KU Leuven.
- Tielemans, K., Vanlaar, G., Van Damme, J., & De Fraine, B. (2019). *Lessen door en voor het Vlaams begrijpend leesonderwijs. 2006 - 2016: tien jaar PIRLS in Vlaanderen*. Centrum voor Onderwijseffectiviteit en -evaluatie, KU Leuven.
- Timmermans, A. (2012). *Value added in educational accountability: possible, fair and useful?* Rijksuniversiteit Groningen.
- Timmermans, A., Doolaard, S., & de Wolf, I. (2011). Conceptual and empirical differences among various value-added models for accountability. *School Effectiveness and School Improvement*, 22(4), 393–413.
- TIMSS & PIRLS International Study Center. (2022a). *PIRLS 2021*. <https://timssandpirls.bc.edu/pirls2021/index.html#>
- TIMSS & PIRLS International Study Center. (2022b). *TIMSS 2023*. <https://timssandpirls.bc.edu/timss2023/>
- Treviño, E., & Órdenes, M. (2017). *Exploring Commonalities and Differences in Regional and International Assessments*. Information Paper No. 48, UNESCO.
- Vakgroep Onderwijskunde. (2020). *PISA2018 - Definitieve schoolfeedback hoofdonderzoek*. UGent.
- Van Der Lans, R., Van De Grift, W., & Van Veen, K. (2019). Same, Similar, or Something Completely Different? Calibrating Student Surveys and Classroom Observations of Teaching Quality Onto a Common Metric. *Educational Measurement: Issues and Practice*, 38(3), 55–64.
- Van Landeghem, G., Dockx, J., Aesaert, K., Van Damme, J., & De Fraine, B. (2019). *PIRLS, de peilingen begrijpend lezen en loopbanen doorheen het lager onderwijs. De impact van alternatieve trajecten op de interpretatie van de prestatiemetingen*. Centrum voor Onderwijseffectiviteit en -evaluatie, KU Leuven.
- Van Renterghem, K., Harnisfeger, J., Verbruggen, K., Vanaenrode, E., Willem, L., Janssen, R., & Aesaert, K. (2020). *De constructie van een peilingsinstrument wiskunde in het basisonderwijs - Eindrapport*. KU Leuven, Steunpunt Toetsontwikkeling en Peilingen.
- Verhelst, N., & Glas, C. (1995). The one parameter logistic model. In G. Fischer & I. Molenaar (Eds.), *Rasch models: Foundations, Recent Developments, and Applications* (pp. 215–237). Springer.

- Vermeir, K. (2021). *Data-analyse van kwalitatieve data*. Centrum voor Onderwijsvernieuwing en de Ontwikkeling van Leraar en School, KU Leuven.
- Vlaams Ministerie van Onderwijs en Vorming. (n.d.). *Kwaliteitszorg in het hoger onderwijs*. <https://www.onderwijs.vlaanderen.be/nl/studenten/vind-een-opleiding-of-instelling/kwaliteitszorg-in-het-hoger-onderwijs>
- Vlaams Ministerie van Onderwijs en Vorming. (2010). *Nederlands Eindtermen: Uitgangspunten*. <https://onderwijsdoelen.be/uitgangspunten/4684>
- Vlaams Ministerie van Onderwijs en Vorming. (2012). *Gelijke onderwijskansen: Ondersteuning van scholen*. <https://www.onderwijs.vlaanderen.be/nl/gelijke-onderwijskansen/ondersteuning-van-scholen>
- Vlaams Ministerie van Onderwijs en Vorming. (2021). *Peilingen*. <https://www.onderwijs.vlaanderen.be/nl/onderwijspersoneel/van-basis-tot-volwassenenonderwijs/lespraktijk/onderwijsdoelen-en-leerplannen/peilingen>
- Vlaams Ministerie van Onderwijs en Vorming. (2022). *Vlaamse toetsen*. <https://onderwijs.vlaanderen.be/nl/beleid/onderwijsregelgeving-in-ontwikkeling/vlaamse-toetsen>
- Vlaamse overheid. (n.d.). *Leerplicht en schoolplicht*. <https://www.vlaanderen.be/leerplicht-en-schoolplicht>
- Wagner, H., Hahn, I., Schöps, K., Ihme, J. M., & Köller, O. (2018). Are the tests scores of the Programme for International Student Assessment (PISA) and the National Educational Panel Study (NEPS) science tests comparable? An assessment of test equivalence in German Schools. *Studies in Educational Evaluation*, 59, 278–287.
- Willem, L., Harnisfeger, J., Carpentier, N., Costers, S., Laenen, I., De Buck, C., & Janssen, R. (2018a). *De constructie van een peilingsinstrument wiskunde in de eerste graad B-stroom van het secundair onderwijs - Verslag vooronderzoeken*. KU Leuven, Steunpunt Toetsontwikkeling en Peilingen.
- Willem, L., Harnisfeger, J., Carpentier, N., Costers, S., Laenen, I., De Buck, C., & Janssen, R. (2018b). *De constructie van een peilingsinstrument wiskunde in de eerste graad van het secundair onderwijs B-stroom - Eindrapport*. KU Leuven, Steunpunt voor Toetsontwikkeling en Peilingen.
- Willem, L., Janssen, R., Luyten, B., & Desmet, P. (2008). *Peiling Frans in de eerste graad van het secundair onderwijs (A-stroom) - Eindrapport*. KU Leuven.
- Willem, L., Van Nijlen, D., De Meyst, M., Bijmens, M., Maesele, L., Aesaert, K., & Janssen, R. (2017). *De constructie van een peilingsinstrument wiskunde in de eerste graad van het secundair onderwijs A-stroom - Verslag vooronderzoeken*. KU Leuven, Steunpunt Toetsontwikkeling en Peilingen.
- Willem, L., Van Nijlen, D., De Meyst, M., Bijmens, M., Maesele, L., Aesaert, K., & Janssen, R. (2018). *De constructie van een peilingsinstrument wiskunde in de eerste graad van het secundair onderwijs A-stroom - Eindrapport*. KU Leuven, Steunpunt Toetsontwikkeling en Peilingen.
- Wu, M. (2010). Measurement, sampling, and equating errors in large-scale assessments. *Educational Measurement: Issues and Practice*, 29(4), 15–27.

Bijlage 1 – Verticale analyse

1. Peiling basisonderwijs Nederlands 2018 en wiskunde 2021

1.1 Design

1.1.1 Doel

Het peilingsonderzoek wordt uitgevoerd door het Steunpunt Toetsontwikkeling en Peilingen (STEP). Het doel van STEP is om het onderwijsbeleid wetenschappelijk te ondersteunen door via de peilingen bij een representatieve steekproef na te gaan “in welke mate de leerlingen de eindtermen beheersen” (Denis et al., 2019b, p. 7). Concreet brengt de peiling Nederlands in kaart welk aandeel van de leerlingen de eindtermen voor lezen, luisteren (en schrijven) beheerst op het einde van het lager onderwijs (Denis et al., 2019a). De peiling wiskunde gaat op het einde van het zesde leerjaar na welk aandeel van de leerlingen de eindtermen voor wiskunde beheerst (Spikic et al., 2022a). Door middel van de peilingen wil de Vlaamse overheid zicht krijgen op eventuele sterke en zwakke punten van het Vlaamse onderwijssysteem om zo de onderwijskwaliteit te evalueren, te bewaken en waar nodig te verbeteren. De Vlaamse overheid zet dus in op het peilingsonderzoek om aan **kwaliteitsbewaking** en **systeemmonitoring** te doen (AHOVOKS, 2021; Vlaams Ministerie van Onderwijs en Vorming, 2021).

Er worden ook regelmatig herhalingspeilingen afgenomen. Bij Nederlands werden de voorgaande peilingen voor lezen afgenomen in 2013, 2007 en 2002, terwijl de peiling luisteren enkel werd georganiseerd in 2013 en 2007 (Denis et al., 2019a). Bij wiskunde werden de vorige peilingen afgenomen in 2016, 2009 en 2002 (Van Renterghem et al., 2020). Die herhalingspeilingen maken het mogelijk om doorheen de tijd het aandeel leerlingen dat de eindtermen bereikt enerzijds en ontwikkelingen in het onderwijsaanbod anderzijds te beschrijven (AHOVOKS, 2021).

Daarnaast laten de peilingen het toe om te onderzoeken in welke mate scholen verschillen van elkaar in de realisatie van de eindtermen (Denis et al., 2019b; Spikic et al., 2022b). De Vlaamse overheid wil ook zo de onderwijskwaliteit bewaken. Kwaliteitsvol onderwijs veronderstelt immers dat er geen grote verschillen zijn tussen leerlingen uit verschillende scholen als leerlingen over dezelfde mogelijkheden beschikken (AHOVOKS, 2021). Bovendien gaan de peilingen na of de prestaties van de leerlingen uit verschillende scholen systematisch samenhangen met achtergrondkenmerken van die scholen, klassen of leerlingen (Denis et al., 2019b; Spikic et al., 2022b). De peilingen kunnen zo een indicatie geven van bepaalde school-, klas- en leerlingkenmerken die samenhangen met minder goede leerlingprestaties. Dat wil echter niet zeggen dat het peilingsonderzoek kan vaststellen waaraan scholen moeten werken opdat meer leerlingen de eindtermen bereiken (Denis et al., 2019a; Spikic et al., 2022a).

Een bijkomende onderzoeksvraag van de peilingen richt zich op het in kaart brengen van het onderwijsaanbod in de Vlaamse lagere scholen. Aansluitend bij die onderzoeksvraag beschrijft het peilingsonderzoek ook het onderwijsaanbod voor respectievelijk Nederlands en wiskunde doorheen de tijd. Om daarover uitspraken te doen baseert het peilingsonderzoek zich op de achtergrondvragenlijsten van de voorgaande en huidige peiling Nederlands in het basisonderwijs (Denis et al., 2019a; Spikic et al., 2022b).

Naast kwaliteitsbewaking en systeemmonitoring heeft het peilingsonderzoek op schoolniveau nog een bijkomend doel, namelijk **schoolfeedback**. Wanneer scholen deelnemen aan een peiling ontvangen ze een schoolfeedbackrapport (Vlaams Ministerie van Onderwijs en Vorming, 2021). Dat rapport toont ten eerste het aandeel deelnemende leerlingen dat de getoetste eindtermen behaald heeft. Ten tweede worden de resultaten van de school in kwestie vergeleken met de andere scholen die

deelgenomen hebben aan de peiling. Ten derde geeft het rapport aan of de school het significant beter of minder goed doet dan scholen met een gelijkaardig leerlingenpubliek. Tot slot bespreekt het rapport hoe scholen concreet aan de slag kunnen gaan met hun schoolfeedback (Denis et al., 2019b; Spikic et al., 2022b). Aansluitend daarbij heeft het STEP ook een stappenplan ontwikkeld voor het gebruik van schoolfeedback (Steunpunt Toetsontwikkeling en Peilingen, 2021). Op die manier kan de terugkoppeling van de resultaten bijdragen aan het versterken van de interne kwaliteitszorg van scholen.

Ook scholen die niet deelnemen aan de peilingen kunnen inzicht krijgen in het aandeel leerlingen van hun school dat de eindtermen bereikt door gebruik te maken van de paralleltoetsen. Die toetsen omvatten andere opgaven dan de peilingstoetsen, maar liggen wel op dezelfde meetschaal. Na de afname van de paralleltoetsen krijgen de scholen eveneens een feedbackrapport (AHOVOKS, 2021; Denis et al., 2019b; Spikic et al., 2022b).

1.1.2 Domein en construct

De eindtermen in het lager onderwijs vormen het uitgangspunt van de peilingsonderzoeken. “Eindtermen voor het gewoon lager onderwijs zijn minimumdoelen die het Vlaams Parlement noodzakelijk en bereikbaar acht voor die leerlingenpopulatie. Met minimumdoelen wordt bedoeld: een minimum aan **kennis, inzicht, vaardigheden en attitudes**, bestemd voor die leerlingenpopulatie” (*Decreet Basisonderwijs*, 1997, Art. 44. §1. 2°).

1.1.2.1 Peiling Nederlands

De eindtermen Nederlands zijn geordend in zeven verschillende domeinen: luisteren, spreken, lezen, schrijven, strategieën, taalbeschouwing en (inter)culturele gerichtheid (Vlaams Ministerie van Onderwijs en Vorming, 2010).⁴⁰ Het peilingsonderzoek uit 2018 focust hoofdzakelijk op de domeinen lezen en luisteren.

Een eindterm voor lezen omvat steeds een specifiek teksttype dat gelinkt wordt aan een van de drie cognitieve verwerkingsniveaus (beschrijven, structureren of beoordelen). Bijna alle eindtermen uit het domein lezen worden opgenomen bij de peiling. Enkel Eindterm 3.3 werd niet geselecteerd. Eindterm 3.5 omvat immers hetzelfde teksttype (tijdschriften), maar is gekoppeld aan een hoger verwerkingsniveau dan Eindterm 3.3 (Aesaert et al., 2018).

Net zoals bij lezen, worden de eindtermen voor luisteren gekoppeld aan één van de drie verwerkingsniveaus (beschrijven, structureren of beoordelen). Bij luisteren wordt Eindterm 1.2 niet opgenomen, omdat er geen informatieve radio-uitzendingen meer bestaan die specifiek gericht zijn op kinderen. Daarnaast worden ook Eindtermen 1.4 en 1.9 niet geselecteerd, aangezien het interactieve karakter van een gesprek moeilijk geoperationaliseerd kan worden in een grootschalig peilingsonderzoek (Aesaert et al., 2018). Tabel 20 toont een overzicht van de geselecteerde eindtermen voor lezen en luisteren die op hun beurt de basis vormen voor de toetsmatrijs (zie 1.1.5.1 Toetsontwikkeling).

⁴⁰ Dit peilingsonderzoek richt zich op de eindtermen Nederlands zoals die golden in 2018.

Tabel 20*De geselecteerde eindtermen Nederlands*

Lezen	
De leerlingen kunnen (verwerkingsniveau = beschrijven) de informatie achterhalen in:	
Eindterm 3.1	voor hen bestemde instructies voor handeling van gevarieerde aard.
Eindterm 3.2	de gegevens in schema's en tabellen ten dienste van het publiek.
De leerlingen kunnen (verwerkingsniveau = structureren) de informatie ordenen die voorkomt in:	
Eindterm 3.4	voor hen bestemde school- en studieteksten en instructies bij schoolopdrachten.
Eindterm 3.5	voor hen bestemde verhalen, kinderromans, dialogen, gedichten, kindertijdschriften en jeugencyclopédieën.
De leerlingen kunnen (verwerkingsniveau = beoordelen) op basis van, hetzij de eigen mening, hetzij informatie uit andere bronnen, informatie beoordelen die voorkomt in:	
Eindterm 3.6	verschillende voor hen bestemde brieven of uitnodigingen.
Eindterm 3.7	reclamateteksten die rechtstreeks verband houden met hun leefwereld.
Luisteren	
De leerlingen kunnen (verwerkingsniveau = beschrijven) de informatie achterhalen in:	
Eindterm 1.1	een voor hen bestemde mededeling met betrekking tot het school- en klasgebeuren.
Eindterm 1.3	een uiteenzetting van een medeleerling, bestemd voor de leerkracht.
De leerlingen kunnen (verwerkingsniveau = structureren) de informatie op een persoonlijke en overzichtelijke wijze ordenen bij:	
Eindterm 1.5	een uiteenzetting of instructie van de leerkracht.
Eindterm 1.6	een voor hen bestemde instructie voor een buitenschoolse situatie.
Eindterm 1.7	een voor hen bestemde informatieve tv-uitzending.
De leerlingen kunnen (verwerkingsniveau = beoordelen) op basis van, hetzij de eigen mening, hetzij informatie uit andere bronnen, informatie beoordelen die voorkomt in:	
Eindterm 1.8	een discussie met bekende leeftijdsgenoten.
Eindterm 1.10	een door leeftijdsgenoten geformuleerde oproep.

1.1.2.2 Peiling wiskunde

In overleg met de Vlaamse overheid werd beslist om alle eindtermen die in voorgaande peilingen getoetst werden op te nemen in de peiling wiskunde van 2021. Concreet gaat het om alle eindtermen wiskunde voor het lager onderwijs behalve de eindtermen 1.9, 1.28, 5.1, 5.3 en 5.4. Die eindtermen worden niet geselecteerd, omdat ze zich niet lenen tot een grootschalige schriftelijke bevraging. Tabel 21 toont een overzicht van de geselecteerde eindtermen wiskunde. Elke eindterm wordt bovendien gelinkt aan een of meerdere cognitieve verwerkingsniveaus: begripsvorming, procedures en toepassen (Van Renterghem et al., 2020).

Tabel 21*De geselecteerde eindtermen wiskunde*

Hoofdrekenen	
Eindterm 1.1	Leerlingen kunnen tellen en terugtellen met eenheden, tweetallen, vijftallen en machten van tien.
Eindterm 1.10*	Leerlingen zijn in staat tot een onmiddellijk geven van correcte resultaten bij optellen en aftrekken tot 10, bij tafels van vermenigvuldiging tot en met de tafels van 10 en de bijhorende deeltafels.
Eindterm 1.13	Leerlingen voeren opgaven uit het hoofd uit waarbij ze een doelmatige oplossingsweg kiezen op basis van inzicht in de eigenschappen van bewerkingen en in de structuur van getallen: <ul style="list-style-type: none"> • optellen en aftrekken tot honderd; • optellen en aftrekken met grote getallen met eindnullen; • vermenigvuldigen met en delen naar analogie met tafels.
Eindterm 1.14	Leerlingen kunnen op concrete wijze de volgende eigenschappen van bewerkingen verwoorden en toepassen: van plaats wisselen, schakelen, splitsen en verdelen.
Kennis van functies en voorstellingswijzen van natuurlijke getallen en andere wiskundige systemen	
Eindterm 1.2	Leerlingen kunnen de verschillende functies van natuurlijke getallen herkennen en verwoorden.
Eindterm 1.7	Leerlingen kunnen door het geven van een paar voorbeelden uit hun eigen leefwereld en in hun leermateriaal aantonen dat doorheen de geschiedenis en ook in niet-westerse culturen andere wiskundige systemen met betrekking tot getallen werden en worden beoefend.
Eindterm 1.8	Leerlingen kunnen gevarieerde hoeveelheidsaanduidingen lezen en interpreteren.
Eindterm 2.4	Leerlingen kunnen de functie van de begrippen “schaal” en “gemiddelde” aan de hand van concrete voorbeelden verwoorden.
Kennis van begrippen en symbolen m.b.t. bewerkingen en inzicht in de relaties tussen de bewerkingen	
Eindterm 1.3	Leerlingen kennen de betekenis van: optellen, aftrekken, vermenigvuldigen, delen, veelvoud, deler, gemeenschappelijke deler, grootste gemeenschappelijke deler, kleinste gemeenschappelijk veelvoud, procent, som, verschil, product, quotiënt en rest. Zij kunnen correcte voorbeelden geven en kunnen verwoorden in welke situatie ze dit handig kunnen gebruiken.
Eindterm 1.6	Leerlingen kunnen volgende symbolen benoemen, noteren en hanteren: =, ≠, <, >, +, -, x, ·, :, /, ÷, % en () in bewerkingen.
Eindterm 1.11	Leerlingen hebben inzicht in de relaties tussen bewerkingen.

Breuken begrijpen en de terminologie hanteren, breuken gelijknamig maken, breuken en kommagetallen zinvol gebruiken

- Eindterm 1.4 Leerlingen kunnen in voorbeelden herkennen dat breuken kunnen uitgelegd worden als: een stuk (deel) van, een verhouding, een verdeling, een deling, een vermenigvuldigingsfactor (operator), een getal (met een plaats op de getallenlijn), weergave van en kans. Leerlingen kunnen volgende terminologie hanteren: stambreuk, teller, noemer, breukstreep, gelijknamig, gelijkwaardig.
- Eindterm 1.22 Leerlingen kunnen eenvoudige breuken gelijknamig maken in functie van het optellen en aftrekken van breuken of in functie van het ordenen en vergelijken van breuken.
- Eindterm 1.23 Leerlingen kunnen in een zinvolle context eenvoudige breuken en kommagetallen optellen en aftrekken. In een zinvolle context kunnen zij eveneens een eenvoudige breuk vermenigvuldigen met een natuurlijk getal.

Natuurlijke getallen, breuken, kommagetallen en procenten lezen, noteren en ordenen; gelijkwaardigheid tussen breuken, kommagetallen en procenten aantonen

- Eindterm 1.5 Leerlingen kunnen natuurlijke getallen van maximaal 10 cijfers en kommagetallen (met 3 decimalen), eenvoudige breuken, eenvoudige procenten lezen, noteren, ordenen en op een getallenlijn plaatsen.
- Eindterm 1.18 Leerlingen kunnen in eenvoudige gevallen de gelijkwaardigheid tussen kommagetallen, breuken en procenten vastleggen en verduidelijken door omzettingen.

Veelvouden en delers vinden en de deelbaarheid toepassen

- Eindterm 1.12 Leerlingen kunnen orde en regelmaat ontdekken in getallenpatronen onder meer om te komen tot de kenmerken van deelbaarheid door 2, 3, 5, 9 en 10 en die te kunnen toepassen.
- Eindterm 1.19 Leerlingen kunnen de delers van een natuurlijk getal (≤ 100) vinden; zij kunnen van twee dergelijke getallen de (grootste) gemeenschappelijke deler(s) vinden.
- Eindterm 1.20 Leerlingen kunnen de veelvouden van een natuurlijk getal (≤ 20) vinden; zij kunnen van twee dergelijke getallen het (kleinste) gemeenschappelijke veelvoud vinden.

Afronden, benaderen en schatten

- Eindterm 1.15 Leerlingen zijn in staat getallen af te ronden. De graad van nauwkeurigheid wordt bepaald door het doel van het afronden en door de context.
- Eindterm 1.16 Leerlingen kunnen de uitkomst van een berekening bij benadering bepalen.
- Eindterm 1.17 Leerlingen kunnen schatprocedures vinden bij niet exact bepaalde of niet exact te bepalen gegevens.
-

Verhoudingen en schaal vaststellen, vergelijken en berekenen

- Eindterm 1.21 Leerlingen zijn in staat in concrete situaties (onder meer tussen grootheden) eenvoudige verhoudingen vast te stellen, te vergelijken, hun gelijkwaardigheid te beoordelen en het ontbrekend verhoudingsgetal te berekenen.
- Eindterm 2.4 Leerlingen kunnen de functie van de begrippen “schaal” en “gemiddelde” aan de hand van concrete voorbeelden verwoorden.

Cijferen met natuurlijke getallen en kommagetallen

- Eindterm 1.24 Leerlingen kennen de cijferalgoritmes. Ze kunnen cijferend vier hoofdbewerkingen uitvoeren met natuurlijke getallen en met kommagetallen.

Rekenen met procenten

- Eindterm 1.25 Leerlingen kunnen eenvoudige procentberekeningen maken met betrekking tot praktische situaties.

Zakrekenmachine gebruiken

- Eindterm 1.26 Leerlingen kunnen de zakrekenmachine doelmatig gebruiken voor hoofdbewerkingen.
- Eindterm 1.27 Leerlingen zijn in staat uitgevoerde bewerkingen te controleren, onder meer met de zakrekenmachine.

Problemen oplossen, zoekstrategieën gebruiken en kritisch nadenken m.b.t. getallen

- Eindterm 1.29 Leerlingen zijn bereid verstandige zoekstrategieën aan te wenden die helpen bij het aanpakken van wiskundige problemen met betrekking tot getallen, meten, ruimtelijke oriëntatie en meetkunde.
- Eindterm 4.1 Leerlingen kunnen met concrete voorbeelden aantonen dat er voor hetzelfde wiskundig probleem met betrekking tot getallen, meten, meetkunde en ruimtelijke oriëntatie, soms meerdere oplossingswegen zijn en soms zelfs meerdere oplossingen mogelijk zijn afhankelijk van de wijze waarop het probleem wordt opgevat.
- Eindterm 4.2 Leerlingen zijn in staat om de geleerde begrippen, inzichten, procedures, met betrekking tot getallen, meten en meetkunde, zoals in de respectievelijke eindtermen vermeld, efficiënt te hanteren in betekenisvolle toepassingssituaties, zowel binnen als buiten de klas.
- Eindterm 4.3 Leerlingen kunnen met concrete voorbeelden uit hun leefwereld aangeven welke de rol en het praktisch nut van wiskunde is in de maatschappij.
- Eindterm 5.2 Leerlingen ontwikkelen een kritische houding ten aanzien van allerlei cijfermateriaal, tabellen, berekeningen waarvan in hun omgeving bewust of onbewust gebruik (misbruik) gemaakt wordt om mensen te informeren, te overtuigen, te misleiden...

Problemen oplossen, zoekstrategieën gebruiken en kritisch nadenken m.b.t. meten, ruimtelijke oriëntatie en meetkunde

- Eindterm 1.29 Leerlingen zijn bereid verstandige zoekstrategieën aan te wenden die helpen bij het aanpakken van wiskundige problemen met betrekking tot getallen, meten, ruimtelijke oriëntatie en meetkunde.
- Eindterm 4.1 Leerlingen kunnen met concrete voorbeelden aantonen dat er voor hetzelfde wiskundig probleem met betrekking tot getallen, meten, meetkunde en ruimtelijke oriëntatie, soms meerdere oplossingswegen zijn en soms zelfs meerdere oplossingen mogelijk zijn afhankelijk van de wijze waarop het probleem wordt opgevat.
- Eindterm 4.2 Leerlingen zijn in staat om de geleerde begrippen, inzichten, procedures, met betrekking tot getallen, meten en meetkunde, zoals in de respectievelijke eindtermen vermeld, efficiënt te hanteren in betekenisvolle toepassingsituaties, zowel binnen als buiten de klas.
- Eindterm 4.3 Leerlingen kunnen met concrete voorbeelden uit hun leefwereld aangeven welke de rol en het praktisch nut van wiskunde is in de maatschappij.
- Eindterm 5.2 Leerlingen ontwikkelen een kritische houding ten aanzien van allerlei cijfermateriaal, tabellen, berekeningen waarvan in hun omgeving bewust of onbewust gebruik (misbruik) gemaakt wordt om mensen te informeren, te overtuigen, te misleiden...

Begrippen, symbolen, notatiewijzen en conventies m.b.t. maateenheden

- Eindterm 2.1 Leerlingen kennen de belangrijkste grootheden en maateenheden met betrekking tot lengte, oppervlakte, inhoud, gewicht (massa), tijd, snelheid, temperatuur en hoekgrootte en ze kunnen daarbij de relatie leggen tussen de grootte en de maateenheid.
- Eindterm 2.2 Leerlingen kennen de symbolen, notatiewijzen en conventies bij de gebruikelijke maateenheden en kunnen meetresultaten op veelzijdige wijze noteren en op verschillende wijze groeperen.
- Eindterm 2.5 Leerlingen weten dat bij temperatuurmeting 0°C het vriespunt is en weten dat de temperaturen beneden het vriespunt met een negatief getal worden aangeduid.

Betekenisvolle herleidingen met grootheden, maatgetallen en maateenheden

- Eindterm 2.6 Leerlingen kunnen allerlei verbanden, patronen en structuren tussen en met grootheden en maatgetallen inzien en ze kunnen betekenisvolle herleidingen uitvoeren.
- Eindterm 2.7 Leerlingen kunnen met gebruikelijke maateenheden betekenisvolle herleidingen uitvoeren.

Maten gebruiken in betekenisvolle situaties en schatten met behulp van referentiepunten

- Eindterm 2.3 Leerlingen kunnen veelvoorkomende maten in verband brengen met betekenisvolle situaties.
- Eindterm 2.8 Leerlingen kunnen schatten met behulp van referentiepunten.
-

Omtrek, oppervlakte en inhoud berekenen

- Eindterm 2.9 Leerlingen kunnen op concrete wijze aangeven hoe ze de oppervlakte en de omtrek van een willekeurige, vlakke figuur en van een veelhoek kunnen bepalen.
- Eindterm 2.10 Leerlingen kunnen concreet aangeven hoe de inhoud van een balk wordt bepaald.

Rekenen met geld en klokkezen

- Eindterm 2.11 Leerlingen kunnen in reële situaties rekenen met geld en geldwaarden.
- Eindterm 2.12 Leerlingen kunnen klokkezen (analoge en digitale klokken). Zij kunnen tijdsintervallen berekenen en zij kennen de samenhang tussen seconden, minuten en uren.

Kennis van begrippen en notaties m.b.t. ruimte en zich ruimtelijk kunnen oriënteren

- Eindterm 3.1 Leerlingen kunnen begrippen en notaties waarmee de ruimte meetkundig wordt bepaald aan de hand van concrete voorbeelden verklaren.
- Eindterm 3.7 Leerlingen zijn in staat:
- zich ruimtelijk te oriënteren op basis van plattegronden, kaarten, foto's en gegevens over afstand en richting;
 - zich in de ruimte mentaal te verplaatsen en te verwoorden wat ze dan zien.

Herkennen, benoemen, classificeren en vormgeven van meetkundige objecten en inzicht in meetkundige relaties

- Eindterm 3.2 Op basis van volgende eigenschappen de volgende meetkundige objecten herkennen en benoemen:
- in het vlak: punten, lijnen, hoeken en vlakke figuren (driehoeken, vierhoeken, cirkels);
 - in de ruimte: veelvlakken (kubus, balk, piramide), bol en cilinder.
- Eindterm 3.3 De symbolen van de loodrechte stand en van de evenwijdigheid lezen en noteren.
- Eindterm 3.4 Leerlingen kunnen de verschillende soorten hoeken classificeren en de verschillende soorten vierhoeken classificeren op grond van zijden en hoeken. Zij kunnen deze ook concreet vorm geven.
- Eindterm 3.5 Leerlingen kunnen met een passer een cirkel tekenen.
- Eindterm 3.6 Leerlingen kunnen de begrippen symmetrie, gelijkvormigheid en gelijkheid ontdekken in de realiteit. Ze kunnen zelf eenvoudige geometrische figuren maken.

*Noot. *Eindterm 1.10 uit de toets hoofdrekenen wordt in een apart onderdeel rond snelrekenen bevraagd.*

1.1.2.3 Achtergrondkenmerken

Naast de eindtermen meet het peilingsonderzoek ook verschillende achtergrondgegevens. Concreet gaat het om: (1) leerlinggegevens, (2) gegevens over de ouders en het gezin, (3) gegevens over de leerkracht en het onderwijsaanbod en (4) vragen over de toets en (5) schoolgegevens. Tabel 22 geeft een overzicht weer van alle bijkomende constructen die het peilingsonderzoek meet (Denis et al., 2019a; Spikic et al., 2022a). Daarbij is het belangrijk om aan te geven dat een bepaalde

achtergrondvariabele niet noodzakelijk op dezelfde manier geoperationaliseerd wordt als bij de Vlaamse overheid. Zo beschrijft het Vlaams Ministerie van Onderwijs en Vorming (2012) de variabele de sociaaleconomische status (kortweg SES) van leerlingen bijvoorbeeld aan de hand van de thuistaal, het ontvangen van een schooltoeslag en het opleidingsniveau van de moeder. In het peilingsonderzoek wordt SES daarentegen gemeten aan de hand van (1) het opleidingsniveau van de ouders, (2) het beroep van de ouders en (3) het gezinsinkomen (Denis et al., 2019a; Spikic et al., 2022a). Bovenstaande constructen vormen niet de focus van het peilingsonderzoek, maar worden hoofdzakelijk gebruikt om verschillen in prestaties tussen leerlingen te duiden enerzijds en het huidige onderwijslandschap in kaart te brengen anderzijds (Denis et al., 2019b; Spikic et al., 2022b).

Tabel 22

Overzicht van achtergrondgegevens bij de peilingen Nederlands en wiskunde in het lager onderwijs

Leerlinggegevens
<ul style="list-style-type: none"> • geslacht en leeftijd; • thuistaal; • leerproblemen; • schoolloopbaan; • hulpaanbod^b; • motivatie^b; • academische zelfconcept; • probleemoplossende vaardigheden^b; • mening over wiskunde of Nederlands.
Gegevens over de ouders en het gezin
<ul style="list-style-type: none"> • opleidingsniveau en tewerkstelling ouders; • SES; • cultureel kapitaal; • migratieachtergrond; • studietoelage^b; • cognitief stimulerend thuisklimaat; • ouderbetrokkenheid; • attitudes ten aanzien van wiskunde of Nederlands; • werkruimte en ICT thuis^b; • communicatie met en tevredenheid over de school en hulp in tijden van corona^b.

Gegevens over de leerkracht en het onderwijsaanbod

- profiel van de leerkracht;
- onderwijsopvattingen over wiskunde^b;
- professionele ontwikkeling;
- inhoudelijke kennis en expertise^b;
- de lessen Nederlands^a:
 - gebruik lesmateriaal Nederlands;
 - belang deelvaardigheden;
 - activiteiten tijdens de lessen Nederlands;
 - evaluatie deelvaardigheden Nederlands;
 - gebruik leerlingvolgsysteem;
 - differentiatie;
- de lessen wiskunde^b:
 - lestijden;
 - gebruik lesmateriaal;
 - vakdidactiek wiskunde;
 - differentiatie;
- schoolbeleid Nederlands of wiskunde;
- wiskundeonderwijs in tijden van corona^b;
- behandelen van de eindtermen^a.

Vragen over de toets

- vragen over de woordenschat in en voorbeelden uit de peilingstoets^b;
- vragen over de moeilijkheidsgraad van de peilingstoets;
- de inzet van de leerlingen;
- het belang van de toets voor de leerlingen.

Schoolgegevens

- schoolgrootte;
- onderwijsnet;
- provincie;
- verstedelijkingsgraad^a;
- percentage GOK-leerlingen.

Noot.

^a Enkel bevroegd bij de peiling Nederlands.

^b Enkel bevroegd bij de peiling wiskunde.

1.1.3 Populatie

Zoals eerder aangehaald, willen deze peilingen onderzoeken welk aandeel van de leerlingen de eindtermen beheerst op het einde van het lager onderwijs. De peilingen worden daarom afgenomen bij leerlingen uit het zesde leerjaar lager gewoon onderwijs. De veronderstelling is immers dat die leerlingen op het moment van de afname (in mei) – op enkele weken na – het volledige curriculum doorlopen hebben (Denis et al., 2019b; Spikic et al., 2022b). Dat heeft als gevolg dat de populatie bestaat uit leerlingen van verschillende leeftijden. Normaalvorderende leerlingen zijn elf of twaalf jaar op het moment dat ze het zesde leerjaar afronden. Leerlingen die voorsprong of vertraging hebben opgelopen doorheen hun schoolloopbaan, kunnen echter jonger of ouder zijn.

Aangezien de peilingen gebaseerd zijn op de eindtermen, werd ervoor gekozen om scholen voor buitengewoon basisonderwijs niet op te nemen in de populatie. In het buitengewoon onderwijs wordt

immers gewerkt met ontwikkelingsdoelen⁴¹ (Denis et al., 2019a). Om dezelfde reden worden er geen leerlingen met een individueel aangepast curriculum (IAC) of anderstalige nieuwkomers opgenomen in de populatie. Daarnaast is het van belang om te beseffen dat niet alle schoolloopbanen in het gewoon basisonderwijs langs het zesde leerjaar verlopen. Zo kunnen leerlingen die vertraging oplopen in het lager onderwijs op twaalfjarige leeftijd rechtstreeks doorstromen naar de B-stroom in het secundair onderwijs. Zij starten in 1B op basis van leeftijd en slaan dus (het vijfde en) het zesde leerjaar lager onderwijs over. Die leerlingen vormen een specifieke en eerder kwetsbare groep in het Vlaamse onderwijs en worden niet gevat door het peilingsonderzoek dat plaatsvindt in het zesde leerjaar. Tot slot is het ook mogelijk dat leerlingen doorheen het lager onderwijs de overstap maken naar het buitengewoon onderwijs (Van Landeghem et al., 2019).

1.1.4 Steekproef

Dit peilingsonderzoek richt zich op leerlingen in het zesde leerjaar gewoon basisonderwijs (zie 1.1.3 Populatie). Het is van belang dat de steekproef gelijkaardig is samengesteld als de Vlaamse populatie om de representativiteit te garanderen. Daarom wordt er een gestratificeerde steekproef van scholen getrokken uit de populatie scholen voor gewoon lager onderwijs. Bij de steekproeftrekking wordt er rekening gehouden met drie stratificatievariabelen, namelijk onderwijsnet, schoolgrootte en provincie (Denis et al., 2019a; Spikic et al., 2022a).

De stratificatievariabele ‘onderwijsnet’ onderscheidt het net van het vrij gesubsidieerd onderwijs, het net van het officieel gesubsidieerd onderwijs en het net van het GO!. Wat betreft de stratificatievariabele ‘schoolgrootte’ worden scholen met minder dan 180 ingeschreven leerlingen beschouwd als kleine scholen. Het peilingsonderzoek omschrijft scholen met minimum 180 ingeschreven leerlingen als grote scholen. Bij de stratificatievariabele ‘provincie’ wordt er een onderscheid gemaakt tussen de provincies Antwerpen, Limburg, Oost-Vlaanderen, Vlaams-Brabant en West-Vlaanderen. Aan de provincie Vlaams-Brabant worden ook de Nederlandstalige scholen uit het Brussels Hoofdstedelijk Gewest toegevoegd (Denis et al., 2019a; Spikic et al., 2022a).

Concreet wordt de Vlaamse populatie van scholen voor gewoon lager onderwijs opgedeeld aan de hand van de twee expliciete stratificatievariabelen, namelijk onderwijsnet en schoolgrootte. Die opdeling leidt tot de volgende zes deelpopulaties of strata:

- kleine scholen behorende tot het vrij gesubsidieerd onderwijs;
- grote scholen behorende tot het vrij gesubsidieerd onderwijs;
- kleine scholen behorende tot het officieel gesubsidieerd onderwijs;
- grote scholen behorende tot het officieel gesubsidieerd onderwijs;
- kleine scholen behorende tot het GO!;
- grote scholen behorende tot het GO!.

De scholen binnen elk stratum worden gesorteerd op basis van de impliciete stratificatievariabele ‘provincie’. At random wordt er een eerste school getrokken binnen elk stratum. Daarna worden de andere scholen binnen het stratum getrokken op basis van een vooraf vastgelegd interval. Dat interval werd berekend door voor elk stratum het totale aantal scholen in dat stratum te delen door het aantal scholen dat nodig was voor de steekproef van dat stratum (Denis et al., 2019a; Spikic et al., 2022a).

⁴¹ Ontwikkelingsdoelen zijn “doelen op vlak van kennis, inzicht, vaardigheden en attitudes die het Vlaams Parlement wenselijk acht voor zoveel mogelijk leerlingen van een bepaalde leerlingenpopulatie” (*Decreet Basisonderwijs*, 1997, Art. 44. §1. 3°). In tegenstelling tot de eindtermen moeten de ontwikkelingsdoelen niet bereikt, maar nagestreefd worden.

Tot slot vond er een trekking van een reservesteekproef plaats, rekening houdend met de respons per stratum uit de eerste steekproef. De feitelijke steekproef voor Nederlands bestond uit 3120 leerlingen uit 119 scholen. Voor wiskunde bestond de steekproef uit 6163 leerlingen uit 230 scholen. Tabel 23 geeft de zes strata of deelpopulaties weer binnen de feitelijke steekproef van het peilingsonderzoek (Denis et al., 2019a; Spikic et al., 2022a).

Tabel 23

Aantal scholen en leerlingen per stratum in de steekproef

Onderwijsnet	Schoolgrootte	
	Klein	Groot
Peiling Nederlands		
Vrij gesubsidieerd onderwijs	9 scholen 108 leerlingen	60 scholen 1797 leerlingen
Officieel gesubsidieerd onderwijs	3 scholen 26 leerlingen	27 scholen 747 leerlingen
Gemeenschapsonderwijs	4 scholen 51 leerlingen	16 scholen 391 leerlingen
Peiling wiskunde		
Vrij gesubsidieerd onderwijs	20 scholen 274 leerlingen	118 scholen 3666 leerlingen
Officieel gesubsidieerd onderwijs	8 scholen 123 leerlingen	46 scholen 1240 leerlingen
Gemeenschapsonderwijs	7 scholen 89 leerlingen	31 scholen 771 leerlingen

Zoals eerder aangegeven, zijn er drie stratificatievariabelen waarvoor de steekproef representatief is. Na de steekproeftrekking wordt bovendien gecontroleerd of de steekproef representatief is voor andere relevante kenmerken. Zo is de steekproef van leerlingen en scholen bij de peiling Nederlands ook representatief voor de Vlaamse populatie op vlak van verstedelijkingsgraad en gemiddeld percentage GOK-leerlingen (Denis et al., 2019a; Spikic et al., 2022a).

Daarnaast be vraagt het peilingsonderzoek ook leerkrachten. Al beperkt het onderzoek zich enkel tot de bevraging van de leerkracht van het zesde leerjaar en wordt er voornamelijk gepeild naar de algemene opvattingen van die leerkracht (Denis et al., 2019a; Spikic et al., 2022a). Nochtans weerspiegelen de opvattingen van de leerkracht niet noodzakelijk de realiteit op school. Bovendien is de leerkracht van het zesde leerjaar niet de enige in de school die verantwoordelijk is voor het bereiken van de eindtermen op het einde van het lager onderwijs. In dat opzicht moet men voorzichtig zijn bij het gebruik van variabelen uit de leerkrachtvragenlijst om uitspraken te doen over het schoolniveau enerzijds en over de samenhang van die variabelen met de leerlingprestaties anderzijds.

1.1.5 Modus

1.1.5.1 Toetsontwikkeling

In overleg met de Vlaamse overheid bepaalt het onderzoeksteam welke onderwijsdoelen getoetst zullen worden (zie 1.1.2 Domein en construct). Nadat de selectie van eindtermen vastgelegd is, volgt de eerste stap in de toetsontwikkeling, namelijk het opstellen van de **toetsmatrijzen** op basis van de geselecteerde eindtermen. Concreet geven die toetsmatrijzen weer voor welke onderwijsdoelen er bepaalde soorten opgaven – en in het geval van de peiling Nederlands bijhorende teksten – ontwikkeld

moeten worden. Daarbij wordt rekening gehouden met de verwerkingsniveaus in de onderwijsdoelen. De toetsmatrijzen vormen het uitgangspunt voor de toetsontwikkeling en verhogen zo de inhoudelijke validiteit van de toetsen (Aesaert et al., 2017, 2018; Van Renterghem et al., 2020).

Vervolgens worden de **opgaven ontwikkeld**. Kenmerkend voor herhalingspeilingen is dat er opgaven uit de voorgaande peilingen worden hergebruikt als ankeritems. Het gebruik van die ankeritems is noodzakelijk om de cesuur (zie 1.2 Standaardbepaling) van de vorige peiling over te nemen (Aesaert et al., 2018; Van Renterghem et al., 2020). Voor de peiling Nederlands gaan ze op basis van de toetsmatrijzen gericht op zoek naar nieuwe teksten voor de toetsen lezen en luisteren. Wanneer alle onderzoekers akkoord gaan met een tekst worden de bijhorende opgaven ontwikkeld en na bespreking aangepast (Aesaert et al., 2018). Ook bij de peiling wiskunde ontwikkelen de toetsontwikkelaars nieuwe opgaven op basis van de toetsmatrijzen (Van Renterghem et al., 2020). Het peilingsonderzoek voor Nederlands en wiskunde maakt voornamelijk gebruik van meerkeuzevragen, maar ook juist-foutvragen, rangschikkingsvragen, verbindingsvragen en een beperkt aantal openvragen komen aan bod (Aesaert et al., 2018; Van Renterghem et al., 2020).

In een volgende stap wordt er een **vooronderzoek** bij experts en een **pilootafname** bij leerlingen georganiseerd om de inhoudelijke validiteit van de toetsen te garanderen en de betrouwbaarheid van de instrumenten na te gaan. Aan verschillende experts wordt gevraagd om eerst de opgaven zelf op te lossen en daarna die opgaven te beoordelen op basis van specifieke beoordelingscriteria die afhankelijk zijn van peiling tot peiling. Verder kunnen de experts ook algemene bedenkingen en hiaten aangeven (Aesaert et al., 2017; Van Renterghem et al., 2020). Er is bovendien een pilootafname georganiseerd bij 174 leerlingen uit zeven scholen voor de peiling Nederlands en bij 575 leerlingen uit 15 scholen voor wiskunde. Na die pilootafname volgen een inhoudelijke analyse van de antwoorden en itemanalyses. Meer specifiek wordt na de pilootafname de moeilijkheidsgraad van de items bekeken, de itemtotaalcorrelatie berekend en de betrouwbaarheid van de gehele toets geëvalueerd aan de hand van Cronbach's alpha. Na feedback van de experts en de analyses van de resultaten uit de pilootafname worden de teksten en toetsopgaven waar nodig herwerkt of geschrapt (Aesaert et al., 2017, 2018; Van Renterghem et al., 2020).

Daarna volgt het **kalibratieonderzoek**⁴² van de toetsen. Met het oog op de psychometrische validiteit van de schriftelijke toetsen, wordt er een grootschalige afname bij een representatieve groep leerlingen georganiseerd om de opgaven te kalibreren. De steekproeftrekking voor het kalibratieonderzoek gebeurt op dezelfde wijze als de steekproeftrekking voor het peilingonderzoek zelf (zie 1.1.4 Steekproef). De feitelijke steekproef voor het kalibratieonderzoek bestond uit 1693 leerlingen in 61 scholen, 68 vestigingsplaatsen en 92 klassen bij de peiling Nederlands. Het kalibratieonderzoek leidt uiteindelijk tot de constructie van een meetschaal voor de toets lezen en luisteren. Door een IRT-model te gebruiken, namelijk het One Parameter Logistic Model (Verhelst & Glas, 1995), kunnen de nieuw ontwikkelde en bestaande items op eenzelfde meetschaal geplaatst worden. Daarbij worden enkel de opgaven die passen bij het gebruikte model behouden (Aesaert et al., 2018; Van Renterghem et al., 2020).

Tot slot volgt de **samenstelling van de definitieve toetsen en de parallelversies**. De opgaven worden opgesplitst in een peilingstoets en een paralleltoets. Doordat die opdeling gebeurt op basis van de

⁴² Omwille van de coronacrisis werd het kalibratieonderzoek bij de peiling wiskunde in het basisonderwijs geannuleerd. Daarom werden de definitieve toetsen samengesteld op basis van de resultaten uit het pilootonderzoek. Bovendien werden er ongeveer 25 items per schaal geselecteerd (in plaats van de vooropgestelde 20 items). Als na de peiling blijkt dat er enkele items minder goed functioneren, kunnen die items uit de toets verwijderd worden (Van Renterghem et al., 2020).

toetsmatrijzen, de discriminatiegraad en de moeilijkheidsgraad van de items, zijn de peilings- en paralleltoetsen zowel op vlak van inhoud als moeilijkheidsgraad equivalent (Aesaert et al., 2017, 2018; Van Renterghem et al., 2020).

Uiteindelijk zijn er voor lezen vier gelijkaardige peilingstoetsen met overlap en een paralleltoets ontwikkeld. Voor luisteren werden twee gelijkaardige toetsen opgesteld, namelijk een peilingstoets en een paralleltoets. Tabel 24 en Tabel 25 geven een overzicht van de teksten en opgaven in de peilingstoetsen voor respectievelijk lezen en luisteren (Aesaert et al., 2017, 2018). Tabel 26 en Tabel 27 tonen het aantal opgaven per toets en de samenstelling van de toetsboekjes bij de peiling wiskunde (Spikic et al., 2022a).

Tabel 24

Overzicht van de teksten en opgaven in de peilingstoets voor lezen

	Aantal teksten	Aantal opgaven per verwerkingsniveau			Totaal aantal opgaven
		Beschrijvend	Structurerend	Beoordelend	
Eindterm 3.1	2	6	3	1	10
Eindterm 3.2	2	6	2	0	8
Eindterm 3.4	1	2	1	1	4
Eindterm 3.5	1	3	2	1	6
Eindterm 3.6	2	2	2	7	11
Eindterm 3.7	2	1	4	3	8
Totaal	10	20	14	13	47

Tabel 25

Overzicht van de teksten en opgaven in de peilingstoets voor luisteren

	Aantal teksten	Aantal opgaven per verwerkingsniveau			Totaal aantal opgaven
		Beschrijvend	Structurerend	Beoordelend	
Eindterm 1.1	1	5	1	0	6
Eindterm 1.3	1	4	0	1	5
Eindterm 1.5	2	5	4	0	9
Eindterm 1.6	1	3	2	1	6
Eindterm 1.7	1	3	2	1	6
Eindterm 1.8	1	1	1	2	4
Eindterm 1.10	1	3	0	1	4
Totaal	8	24	10	6	40

Tabel 26

Overzicht van aantal opgaven per toets bij de peiling wiskunde

Toets	Aantal opgaven
1. Hoofdrekenen	32
2. Functies en voorstellingswijzen	33
3. Bewerkingen	29
4. Breuken (begrip, gebruik en terminologie)	33
5. Getalwaarden en gelijkwaardigheid	64
6. Veelvouden en delers	27
7. Afronden, benaderen en schatten	33
8. Verhoudingen en schaal	30
9. Cijferen	28
10. Procent berekenen	36
11. Zakrekenmachine	21
12. Problemen oplossen bij getallen en bewerkingen	32
13. Problemen oplossen bij meten, ruimte en meetkunde	32
14. Begrippen, symbolen, notatie en conventies m.b.t. maateenheden	36
15. Betekenisvolle herleidingen	57
16. Maten gebruiken in betekenisvolle situaties en schatten met behulp van referentiepunten	29
17. Omtrek, oppervlakte en inhoud	31
18. Rekenen met geld en klokkezen	32
19. Ruimte en ruimtelijke oriëntatie	27
20. Begrippen m.b.t. meetkunde	37
Totaal	679

Tabel 27

Samenstelling toetsboekjes bij de peiling wiskunde

Toets	Toetsboekje							
	1	2	3	4	5	6	7	8
1. Hoofdrekenen	■	■						
2. Functies en voorstellingswijzen					■	■		
3. Bewerkingen			■	■				
4. Breuken (begrip, gebruik en terminologie)				■	■		■	
5. Getalwaarden en gelijkwaardigheid	■			■				
6. Veelvouden en delers			■	■				
7. Afronden, benaderen en schatten						■		■
8. Verhoudingen en schaal					■		■	■
9. Cijferen							■	■
10. Procent berekenen		■	■					
11. Zakrekenmachine			■	■				
12. Problemen oplossen bij getallen en bewerkingen					■	■		
13. Problemen oplossen bij meten, ruimte en meetkunde	■	■						
14. Begrippen, symbolen, notatie en conventies m.b.t. maateenheden						■	■	■
15. Betekenisvolle herleidingen						■	■	
16. Maten gebruiken in betekenisvolle situaties en schatten met behulp van referentiepunten					■			■
17. Omtrek, oppervlakte en inhoud		■		■				
18. Rekenen met geld en klokkezen	■		■	■				
19. Ruimte en ruimtelijke oriëntatie			■	■				
20. Begrippen m.b.t. meetkunde						■		■

Het hele peilingsproject – en dus ook de toetsontwikkeling – wordt daarnaast opgevolgd door een **resonantiegroep**. Die groep bestaat onder andere uit inhoudelijke experts, leerkrachten, pedagogische begeleiders en onderwijsinspecteurs. Zij geven tijdens de verschillende fasen van het peilingsonderzoek input vanuit hun ervaringen in de onderwijspraktijk (Steunpunt Toetsontwikkeling en Peilingen, n.d.).

Naast de toetsen voor lezen en luisteren worden er ook **achtergrondvragenlijsten voor leerlingen, ouders en leerkrachten** ontwikkeld (Denis et al., 2019a; Spikic et al., 2022a). De resonantiegroep en opvolgsgroep (bestaande uit de opdrachtgever en vertegenwoordigers uit het onderwijsveld) bepalen daarbij welke inhoudelijke concepten er bevestigd worden. Het onderzoeksteam van het STEP staat in voor de operationalisering van die concepten en de constructie van de uiteindelijke vragenlijsten. Met het oog op de ontwikkeling van kwaliteitsvolle en betrouwbare instrumenten, worden er factoranalyses uitgevoerd en wordt de betrouwbaarheid van de verschillende schalen geëvalueerd. Wanneer schalen onvoldoende betrouwbaar zijn of inhoudelijk onvoldoende samenhangen, worden ze niet gebruikt in het peilingsonderzoek.

1.1.5.2 Toetsafname

De schriftelijke toetsen voor de peiling Nederlands werden klassikaal afgenomen in de voormiddag van 31 mei 2018. In de eerste twee lesuren losten de leerlingen de schriftelijke toetsen voor lezen en luisteren op. Na een pauze legden leerlingen de praktische proef af en vulden ze de leerlingvragenlijst in. Daarbij werden redelijke aanpassingen toegelaten voor leerlingen met specifieke onderwijsbehoeften (Denis et al., 2019a).

De peiling wiskunde vond plaats op 18 of 20 mei 2021. Tijdens de afname, die maximaal vier lesuren (met pauze) duurde, legden de leerlingen vier, vijf of zes wiskundetoetsen af. Ook bij deze peiling werden redelijke aanpassingen toegelaten voor leerlingen met specifieke onderwijsbehoeften (Spikic et al., 2022b).

Beide afnames gebeurden onder begeleiding van een leerkracht. Met het oog op de standaardisatie werd er een handleiding uitgewerkt voor de leerkrachten waarin gedetailleerd omschreven staat hoe de afname verloopt, wat ze juist moeten doen en welke instructies ze moeten geven aan de leerlingen. Aan de leerkrachten werd ook gevraagd om onregelmatigheden en afwezigheden te noteren in het logboek. Per vestigingsplaats was er daarnaast een toetsassistent aanwezig om de leerkracht bij te staan. Zij keken toe op het correcte verloop van de afname en de geheimhouding van het toetsmateriaal. Ook voor de toetsassistenten was er een gedetailleerde handleiding beschikbaar waarin stapsgewijs wordt omschreven wat precies van hen verwacht wordt. Na de toetsafname verzamelden de toetsassistenten alle documenten, stuurden ze die op naar het onderzoeksteam en brachten ze kort verslag uit over de toetsafname (Denis et al., 2019a; Spikic et al., 2022a).

Bij zowel de peiling Nederlands als de peiling wiskunde werden bovendien vragenlijsten afgenomen bij de leerkrachten en ouders van de leerlingen. In een volgende stap werden de antwoorden van de leerlingen, ouders en leerkrachten ingevoerd en gecontroleerd. Aan de hand van scoringswijzers werden de antwoorden van de leerlingen op de schriftelijke toetsen gescoord (Denis et al., 2019a; Spikic et al., 2022a).

1.2 Standaardbepaling

Het peilingsonderzoek is een criteriumgerichte meting. De eindtermen zelf vormen namelijk het uitgangspunt of criterium om te bepalen welke leerlingen de eindtermen al dan niet bereiken. Meer specifiek maakt het peilingsonderzoek bij de cesuurbepaling gebruik van de bookmark methode. Bij die methode worden de opgaven eerst gerangschikt op de meetschaal van makkelijk naar moeilijk.

Daarna duiden deskundigen op basis van een inhoudelijke analyse van de eindtermen en opgaven een cesuur aan. Die cesuur verdeelt de meetschaal in twee groepen: de groep basisopgaven die leerlingen moeten beheersen en de groep opgaven die leerlingen (nog) niet moeten beheersen om te eindtermen te behalen. Op diezelfde meetschaal kunnen ook de leerlingen gesitueerd worden naargelang hun vaardigheid om zo te bepalen welke leerlingen de eindtermen wel of niet bereiken (Denis et al., 2019a; Spikic et al., 2022a).

Als de eindtermen niet wijzigen, is het mogelijk om de prestaties van leerlingen doorheen te tijd in kaart te brengen met behulp van herhalingspeilingen (zie 1.1.1 Doel). Om correcte uitspraken te doen over die evolutie, is het cruciaal om de cesuur uit de voorgaande peiling over te nemen bij een herhalingspeiling. Op basis van de bookmark methode is immers al eerder bepaald waar de cesuur ligt op de meetschaal van de vorige peiling. Het overnemen van de cesuur vereist het gebruik van gemeenschappelijke opgaven (of ankeritems). Door die opgaven kunnen zowel de leerlingen die deelnamen aan de voorgaande peiling als de leerlingen die deelnamen aan de recente peiling op de nieuwe meetschaal van de recente peiling geplaatst worden. De eerder vastgelegde cesuur kan zo correct overgenomen worden aan de hand van het percentage leerlingen dat bij de vorige peiling de eindtermen behaalde. Ook hier geldt: leerlingen die op de meetschaal boven de cesuur (en dus minimumnorm) gesitueerd zijn, behalen de eindtermen (Denis et al., 2019a).

1.3 Analytische methode

Naast descriptieve analyses worden er in het peilingsonderzoek ook multilevel analyses uitgevoerd. De concrete analysemethodes worden uitgebreid beschreven in de eindrapporten van de peilingen Nederlands en wiskunde in het basisonderwijs (Denis et al., 2019a; Spikic et al., 2022a). In wat volgt, focussen we op schatting van de vaardigheidsscores en de gebruikte modellen bij het peilingsonderzoek.

1.3.1 Schatting van vaardigheidsscores

Hoewel de werkelijke vaardigheid van leerlingen onbekend is, is het wel mogelijk om de vaardigheid van leerlingen te schatten. Het peilingsonderzoek maakt daarbij gebruik van *Warm's weighted likelihood estimation* om de meest waarschijnlijke vaardigheid te schatten gegeven de geobserveerde data en de parameters in het One Parameter Logistic Model. Om de interpretatie van de geschatte vaardigheidsscores te vergemakkelijken, worden de scores na de schatting getransformeerd naar een schaal waarvan het gemiddelde 50 is en de standaarddeviatie 10.

Doorgaans bieden toetsen minder informatie voor leerlingen die extreem laag of hoog scoren, waardoor lage vaardigheden onderschat worden en hoge vaardigheden overschat. De procedure van *Warm's weighted likelihood estimation* corrigeert voor die onder- en overschatting door een correctiefactor te gebruiken bij het schatten van de vaardigheidsscores. Die correctiefactor is groter naarmate de informatie voor een bepaalde vaardigheid beperkter is en kleiner wanneer de informatie voor een specifieke vaardigheid groot is.

1.3.2 De gebruikte modellen

Voor elke toets wordt er een meetschaal geconstrueerd op basis van het One Parameter Logistic Model. Die meetschalen vormen de basis voor de multilevel analyses. Eerst wordt er een **leeg model** opgesteld, waarin de variantie in de vaardigheidsscores voor lezen en luisteren verdeeld wordt over het leerlingniveau en het schoolniveau. Op basis van dat model kan voor elke school het verschil tussen hun schoolgemiddelde en het algemeen Vlaams gemiddelde (het residu) op de toets berekend worden. Door betrouwbaarheidsintervallen te berekenen voor die residuen kan worden nagegaan of het schoolgemiddelde significant hoger of lager ligt dan het algemeen Vlaams gemiddelde. Bovendien kunnen scholen ook onderling vergeleken worden in het leeg model. Om na te gaan of scholen

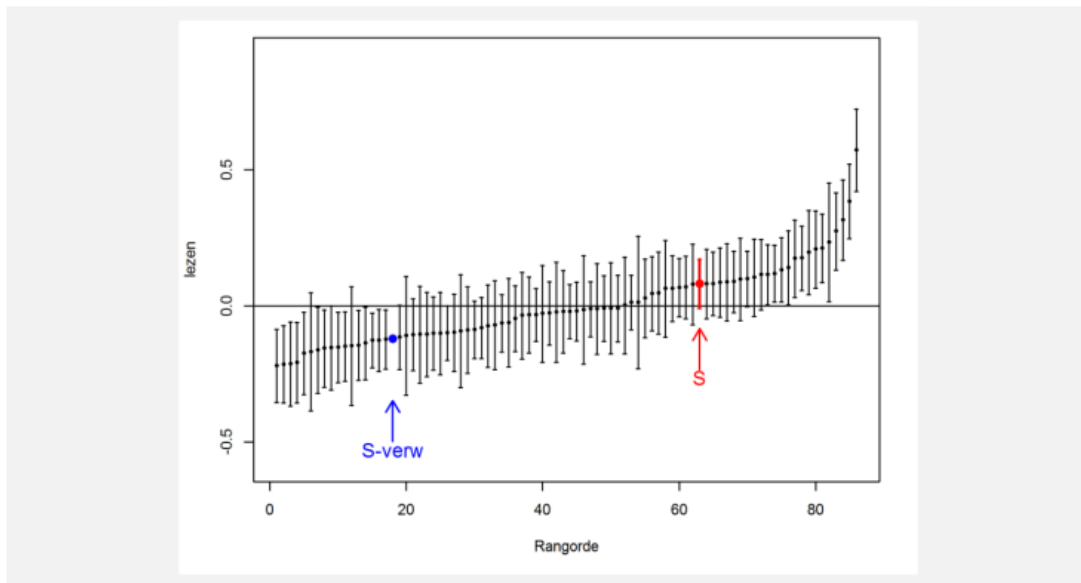
onderling significant van elkaar verschillen, kunnen aangepaste betrouwbaarheidsintervallen berekend worden rond de schoolresiduen (Denis et al., 2019a).

Daarna wordt het **nettomodel** geschat. In dat model wordt er rekening gehouden met de factoren waarop een school geen invloed heeft. Het peilingonderzoek bepaalt a priori welke variabelen worden opgenomen in dat model. Concreet worden op schoolniveau de variabelen (1) provincie, (2) onderwijsnet, (3) schoolgrootte en (4) percentage GOK-leerlingen opgenomen. Op leerlingniveau gaat het om de volgende variabelen: (1) geslacht, (2) thuistaal, (3) schoolse achterstand, (4) SES, (5) aantal boeken thuis en (6) leerproblemen. Voor elke variabele wordt er een parameter geschat die voor een bepaalde groep (bijvoorbeeld: meisjes) het verschil in vaardigheidsscore weergeeft ten opzichte van een referentiegroep (bijvoorbeeld: jongens). Net zoals het leegmodel laat het nettomodel met behulp van residuen en betrouwbaarheidsintervallen toe om scholen te vergelijken met het algemeen Vlaams schoolgemiddelde en met andere scholen. De vergelijking op basis van het nettomodel is eerlijker dan bij het leeg model. Er wordt immers rekening gehouden met factoren die bepalend kunnen zijn voor de resultaten, maar waarop scholen geen invloed hebben (Denis et al., 2019a).

In een volgende stap wordt er onderzocht of er een bijkomend leerling-, klas-, leerkracht- of schoolkenmerk samenhangt met de prestaties van de leerlingen na controle voor de variabelen die reeds opgenomen zijn in het nettomodel. Daarvoor wordt elke bijkomende variabele apart toegevoegd aan het nettomodel (Denis et al., 2019a).

Het **model voor schoolfeedback** maakt gebruik van het leeg model en een model dat sterkt lijkt op het nettomodel. Bij dat laatste model worden alleen de stratificatievariabelen onderwijsnet, schoolgrootte en provincie niet meegenomen op schoolniveau. Die variabelen zijn namelijk niet noodzakelijk om correcte feedback te kunnen geven aan een school. Zoals eerder vermeld, krijgen scholen in het schoolfeedbackrapport informatie over het aantal leerlingen dat de eindtermen bereikt (zie 1.1.1 Doel). Daarnaast wordt op basis van het leeg model een vergelijking gemaakt waarbij de resultaten op de peilingstoets van een bepaalde school vergeleken worden met de andere scholen die deelnamen aan het peilingsonderzoek. Op basis daarvan wordt nagegaan of de school significant verschilt van het Vlaams gemiddelde. Verder wordt er aan de hand van het aangepaste nettomodel een vergelijking gemaakt. Daarbij wordt er rekening gehouden met achtergrondkenmerken die bepalend kunnen zijn voor de resultaten, maar waarop scholen geen invloed hebben. Op die manier is het mogelijk om de resultaten van een specifieke school te vergelijken met scholen die een gelijkaardig leerlingenpubliek hebben (Denis et al., 2019a). Figuur 33 toont een fictief voorbeeld uit het schoolfeedbackrapport bij de peiling Nederlands in het basisonderwijs.

Fictief voorbeeld



- Zwarte stippen: feitelijke gemiddeldes van de scholen die deze toets aflegden in de peiling
- **Rode stip S**: feitelijk gemiddelde van deze school
- **Blauwe stip S-verw**: verwacht gemiddelde van deze school
- Verticale lijn door een stip: betrouwbaarheidsinterval
- Horizontale lijn door 0.0 (nullijn): Vlaams gemiddelde (gemiddelde over alle peilingsscholen heen)
- Verticale as (Y-as, Toets): gemiddelde prestatie per school
- Horizontale as (X-as, Rangorde): volgorde peilingsscholen op basis van hun feitelijke gemiddeldes: scholen met de laagste gemiddeldes uiterst links, scholen met de hoogste gemiddelde uiterst rechts

Figuur 33. Fictief voorbeeld van schoolfeedback over de schoolprestatie in vergelijking met het Vlaams gemiddelde en scholen met een gelijkaardig leerlingenpubliek (verwacht gemiddelde) bij de peiling Nederlands in het basisonderwijs (Figuur overgenomen uit Denis et al., 2019a)

1.4 Besluitvorming

Zoals eerder aangehaald, heeft het peilingsonderzoek als hoofddoel om het aandeel leerlingen dat de eindtermen beheerst op het einde van het lager onderwijs in kaart te brengen. Op basis van het peilingsonderzoek is het mogelijk om valide en betrouwbare uitspraken te doen over het aandeel leerlingen dat de eindtermen (die opgenomen werden in de toetsmatrijzen) beheerst. Door de gehanteerde methode voor cesuurbepaling vormen de eindtermen immers zelf het uitgangspunt om te bepalen welke leerlingen de eindtermen al dan niet bereiken. Bovendien maakt het peilingsonderzoek gebruik van een representatieve steekproef, worden er verscheidene maatregelen genomen om de inhoudelijke validiteit van de toetsen te garanderen en wordt de betrouwbaarheid van de toetsen geëvalueerd.

Daarnaast wordt er onderzocht of de prestaties van de leerlingen samenhangen met kenmerken van scholen, klassen of leerlingen. Op basis van de multilevel analyses is niet mogelijk om verklaringen te bieden voor de gevonden verbanden. Het peilingsonderzoek kan dus geen uitspraken doen over wat werkt in het onderwijs. Bovendien moeten de gevonden verbanden met de nodige voorzichtigheid

geïnterpreteerd worden. Die verbanden zijn eerder indicatief, maar kunnen wel aanleiding geven tot verder onderzoek (Denis et al., 2019b; Spikic et al., 2022b).

Verder beschrijft het peilingsonderzoek de prestaties van de leerlingen en ontwikkelingen in het onderwijsaanbod doorheen de tijd. Door het gebruik van ankeritems kan de cesuur uit de vorige peiling correct overgenomen worden en kunnen de evoluties in leerlingprestaties beschreven worden. Het peilingsonderzoek is echter minder geschikt om het feitelijke onderwijsaanbod doorheen de tijd in kaart te brengen. Aangezien het moeilijk is om de concrete klaspraktijk te vatten in grootschalig onderzoek, focust de leerkrachtvragenlijst eerder op opvattingen van de leerkrachten (Denis et al., 2019b; Spikic et al., 2022b). Daarbij is het belangrijk om op te merken dat enkel de leerkracht uit het zesde leerjaar bevestigd wordt. Nochtans weerspiegelen de opvattingen van die leerkracht niet noodzakelijk de realiteit op school. Daarenboven is de leerkracht van het zesde leerjaar niet de enige op school die verantwoordelijk is voor het bereiken van de eindtermen op het einde van het lager onderwijs.

Een laatste doel van het peilingsonderzoek is schoolfeedback. De schoolfeedbackrapporten kunnen een waardevolle bron van informatie zijn voor scholen en op die manier scholen aanzetten tot zelfevaluatie en reflectie. Aangezien het peilingsonderzoek zich beperkt tot het vaststellen en beschrijven van samenhang, kan het schoolfeedbackrapport niet aangeven waaraan een specifieke school moet werken zodat meer leerlingen de eindtermen bereiken. Bovendien gaat het om een momentopname en worden niet alle eindtermen in het leergebied getoetst. Zoals eerder aangehaald (zie 1.1.1 Doel), vergelijkt het schoolfeedbackrapport de resultaten van een school niet alleen met de resultaten van andere (vergelijkbare) scholen, maar ook met de eindtermen zelf. Het rapport toont immers het aantal leerlingen dat de eindtermen bereikt. Die absolute feedback is mogelijk doordat het peilingsonderzoek een criteriumgerichte meting is (zie 1.2 Standaardbepaling). Toch is het nodig om de schoolfeedback voorzichtig te interpreteren, aangezien de foutenmarge bij uitspraken over de resultaten van individuele scholen groot kan zijn. Het is dus aan de scholen zelf om de informatie uit het schoolfeedbackrapport in de juiste context te plaatsen en te combineren met andere informatiebronnen met het oog op hun interne kwaliteitszorg (Denis et al., 2019b; Spikic et al., 2022b).

1.5 Praktische proef bij de peiling Nederlands

Via een praktische proef werden bij de peiling Nederlands in het basisonderwijs enkele eindtermen uit het domein schrijven getoetst. Een praktische proef bestaat uit een beperkt aantal opdrachten die ingebed zijn in een specifieke context. Concreet werd er in dit peilingsonderzoek aan de leerlingen gevraagd om een of twee korte teksten te schrijven (Denis et al., 2019b). Tabel 28 toont een overzicht van de eindtermen luisteren die exemplarisch getoetst werden op basis van de praktische proef.

Tabel 28*De geselecteerde eindtermen Nederlands schrijven voor de praktische proef*

Schrijven	
De leerlingen kunnen (verwerkingsniveau = kopiëren):	
Eindterm 4.1	overzichten overschrijven.
De leerlingen kunnen (verwerkingsniveau = beschrijven):	
Eindterm 4.2	een oproep, een uitnodiging, een instructie richten tot leeftijdsgenoten.
De leerlingen kunnen (verwerkingsniveau = structureren):	
Eindterm 4.3	een brief schrijven aan een bekende om een persoonlijke boodschap of belevens over te brengen.
Eindterm 4.4	voor een gekend persoon een verslag schrijven van een verhaal, een gebeurtenis, een informatieve tekst.
Eindterm 4.7	voor het realiseren van bovenstaande eindtermen bovendien: <ul style="list-style-type: none">• hun teksten verzorgen rekening houdende met handschrift en lay-out.• spellingsafspraken en -regels toepassen in verband met het schrijven van<ul style="list-style-type: none">○ woorden met vast woordbeeld:<ul style="list-style-type: none">▪ klankzuivere woorden;▪ hoogfrequente niet-klankzuivere woorden;○ woorden met veranderlijk woordbeeld (regelwoorden):<ul style="list-style-type: none">▪ werkwoorden;▪ klinker in open/gesloten lettergreep;▪ verdubbeling medeklinker;▪ niet-klankzuivere eindletter;○ hoofdletters;○ interpunctietekens (. , ? ! :).

De praktische proef is anders van aard dan de eerder besproken toetsen voor lezen en luisteren. Doordat de praktische proef slechts een beperkt aantal opdrachten omvat, is het niet mogelijk om een uitspraak te doen over het aandeel leerlingen dat de eindtermen schrijven bereikt. Toch is een praktische proef waardevol. Niet alle eindtermen kunnen immers getoetst worden door middel van schriftelijke toetsen die hoofdzakelijk bestaan uit meerkeuzevragen. Door het gebruik van een praktische proef is het mogelijk om exemplarisch enkele eindtermen uit het domein schrijven te toetsen (Denis et al., 2019a). Echter, de resultaten van de praktische proef kunnen niet ingezet worden voor systeemmonitoring of kwaliteitsbewaking en zijn louter illustratief.

1.6 Het peilingsonderzoek doorheen de tijd

De eerste peilingen werden afgenomen in 2002. Concreet ging het om een peiling wiskunde en begrijpend lezen in het basisonderwijs. Sindsdien zijn er verschillende peilingen (zie Tabel 29 voor een overzicht) afgenomen in het basis- en secundair onderwijs voor heel wat verschillende domeinen (Steunpunt Toetsontwikkeling en Peilingen, 2022). Bij die eerste peilingen werd er reeds gebruikt gemaakt van multilevel analyses en ontvingen de deelnemende scholen schoolfeedback (Janssen & Ver Eecke, 2004). Al is die schoolfeedback doorheen de jaren weliswaar sterk uitgebreid (Steunpunt Toetsontwikkeling en Peilingen, 2022). Bij de peiling in 2004 werden er voor de eerste keer paralleltoetsen opgesteld en in 2005 werd de eerste praktische proef georganiseerd (Janssen et al., 2006; Janssens et al., 2007). Ook 2007 wordt gekenmerkt door een mijlpaal in het peilingsonderzoek: de eerste gedeeltelijk digitale afname bij de peiling Frans. Daarbij werd er de luistertoets van de

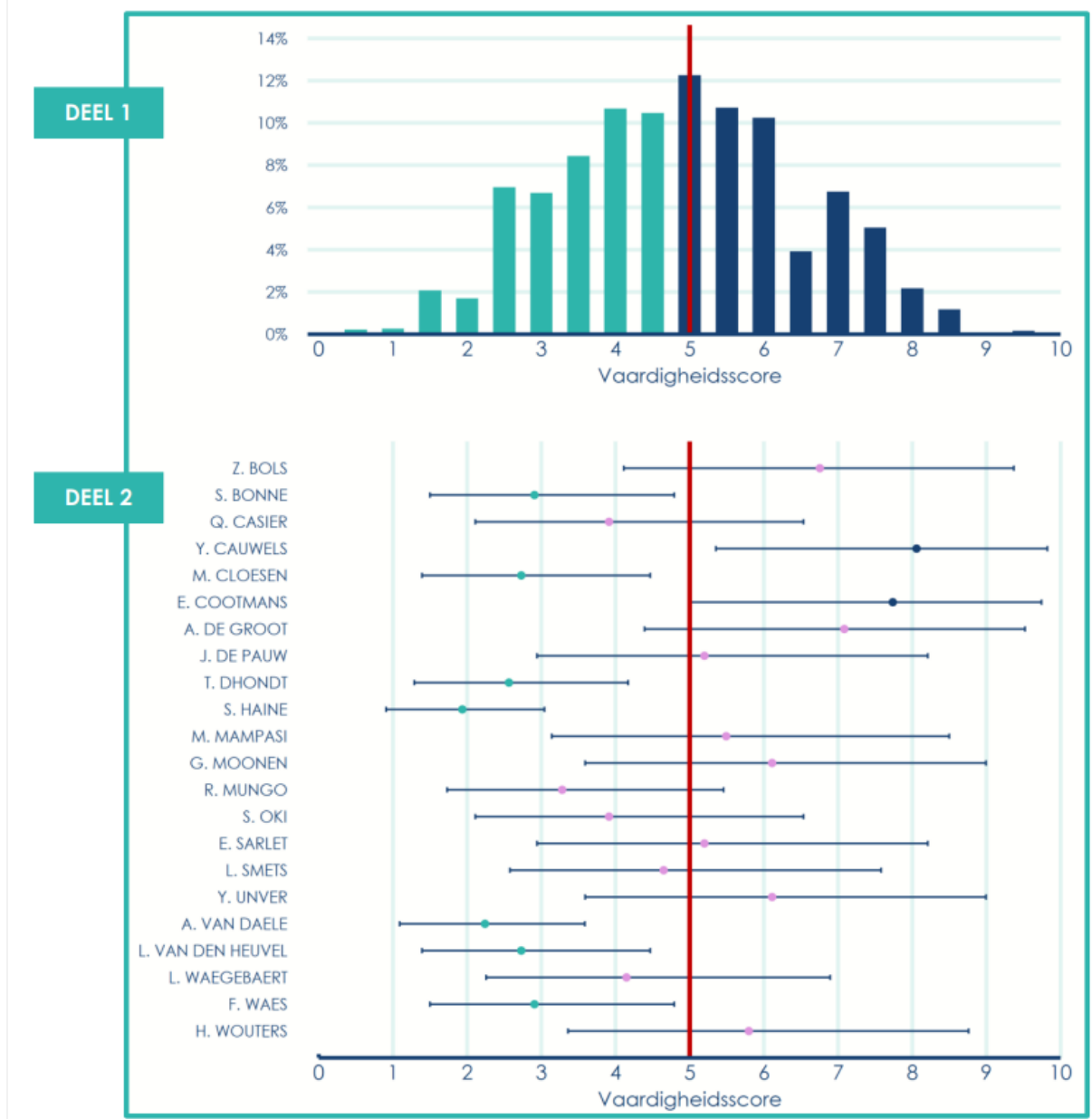
peilingen via de computer afgenomen (Willem et al., 2008). Sindsdien worden er daarenboven twee peilingen per jaar georganiseerd (Steunpunt Toetsontwikkeling en Peilingen, 2022). Sinds 2018 wordt er nog sterker ingezet op kennisdeling met het onderwijsveld en worden de individuele leerlingresultaten opgenomen bij de schoolfeedbackrapporten van de paralleltoetsen (Steunpunt Toetsontwikkeling en Peilingen, 2019). Figuur 34 geeft een voorbeeld weer van feedback op leerlingniveau naar aanleiding van de paralleltoets wetenschappen en techniek voor het basisonderwijs. In 2022 zijn de laatste peilingen afgenomen (Steunpunt Toetsontwikkeling en Peilingen, 2022).

Tabel 29

Peilingskalender

Jaar	Basisonderwijs	Secundair onderwijs
2002	Wiskunde Begrijpend lezen	/
2004	/	Informatieverwerving en -verwerking (1 ^e graad A-stroom)
2005	Wereldoriëntatie	/
2006	/	Biologie (1 ^e graad A-stroom)
2007	Nederlands	Frans (1 ^e graad A-stroom)
2008	Frans	Wiskunde (1 ^e graad B-stroom)
2009	Wiskunde	Wiskunde (1 ^e graad A-stroom)
2010	Wereldoriëntatie	Nederlands (3 ^e graad ASO, KSO en TSO)
2011	/	Informatieverwerving en -verwerking (1 ^e graad A-stroom) Wiskunde (2 ^e graad ASO)
2012	Informatieverwerving en -verwerking	Frans (3 ^e graad ASO, KSO en TSO)
2013	Nederlands	Project algemene vakken (3 ^e graad BSO)
2014	/	Wiskunde (3 ^e graad ASO, KSO en TSO)
2015	Wereldoriëntatie	Natuurwetenschappen (1 ^e graad A-stroom)
2016	Wiskunde	Burgerzin (3 ^e graad ASO, TSO, KSO en BSO)
2017	Frans	Techniek (1 ^e graad A-stroom)
2018	Nederlands	Wiskunde (1 ^e graad A-stroom)
2019	Mens en maatschappij	Wiskunde (1 ^e graad B-stroom)
2020	Informatieverwerving en -verwerking	/
2021	Wiskunde	Project algemene vakken (3 ^e graad BSO) Kritisch denken en mediawijsheid (3 ^e graad ASO, TSO, KSO en BSO)
2022	/	Nederlands (1 ^e graad A- en B-stroom) Wiskunde (1 ^e graad A- en B-stroom)

Fictief voorbeeld:



Figuur 34. Fictief voorbeeld van schoolfeedback over de individuele leerlingresultaten bij de paralleltoets wetenschappen en techniek in het basisonderwijs (Figuur overgenomen uit Steunpunt Toetsontwikkeling en Peilingen, 2019)

2. Peiling wiskunde A-stroom 2018 en B-stroom 2019

2.1 Design

2.1.1 Doel

Het doel van het peilingsonderzoek in het secundair onderwijs is hetzelfde als in het lager onderwijs (zie 1.1.1 Doel voor een uitgebreide toelichting). Concreet brengen de peilingen wiskunde in de A- en B-stroom in kaart welk aandeel van de leerlingen de onderwijsdoelen voor wiskunde bereikt op het einde van de eerste graad secundair onderwijs. De peiling in de A-stroom was een herhaling van de

peiling uit 2009, terwijl de peiling in de B-stroom een herhaling was van de peiling uit 2008 (Carpentier et al., 2019b, 2020b).

2.1.2 Domein en construct

2.1.2.1 Peiling wiskunde in de A-stroom

De eindtermen wiskunde voor de eerste graad secundair onderwijs staan centraal. “Eindtermen zijn minimumdoelen die het Vlaams Parlement noodzakelijk en bereikbaar acht voor een bepaalde leerlingenpopulatie. Met minimumdoelen wordt bedoeld: een minimum aan **kennis, inzicht, vaardigheden en attitudes**, bestemd voor die leerlingenpopulatie” (*Codex Secundair Onderwijs*, 2010, Art. 139. §1.). De geselecteerde eindtermen wiskunde zijn geordend in vier verschillende domeinen: getallenleer, algebra, data en meetkunde. Tabel 30 toont een overzicht van de geselecteerde eindtermen wiskunde (Carpentier et al., 2019a, 2019b). Daarnaast wordt elke eindterm gelinkt aan een of meerdere cognitieve verwerkingsniveaus: feitenkennis, begripsvorming, reproductief toepassen of productief toepassen (Willem, Van Nijlen, et al., 2018).

Tabel 30

De geselecteerde eindtermen wiskunde

Getallenleer	
Getalinzicht	
Eindterm 1	De leerlingen kunnen natuurlijke, gehele en rationale getallen associëren met realistische en betekenisvolle contexten.
Eindterm 4	De leerlingen onderscheiden en begrijpen de verschillende notaties van rationale getallen (breuk- en decimale positie).
Eindterm 10	De leerlingen ordenen getallen en gebruiken de gepaste symbolen (\leq , $<$, \geq , $>$, $=$, \neq).
Eindterm 14	De leerlingen interpreteren een rationaal getal als een getal dat de plaats van een punt op een getallenas bepaalt.

Bewerkingen

Eindterm 2	De leerlingen kennen de tekenregels bij gehele en rationale getallen.
Eindterm 3	De leerlingen weten dat de eigenschappen van de bewerkingen in de verzameling van de natuurlijke getallen geldig blijven en kunnen worden uitgebreid in de verzameling van de gehele en rationale getallen.
Eindterm 5	De leerlingen hanteren de gepaste terminologie in verband met bewerkingen: optelling, som, termen van een som, aftrekking, verschil, vermenigvuldiging, product, factoren van een product, deling, quotiënt, deeltal, deler, rest, percent, kwadraat, vierkantswortel, macht, grondtal, exponent, tegengestelde, omgekeerde, absolute waarde, gemiddelde.
Eindterm 6	De leerlingen passen afspraken in verband met de volgorde van bewerkingen toe.
Eindterm 7	De leerlingen voeren de hoofdbewerkingen (optelling, aftrekking, vermenigvuldiging en deling) correct uit in de verzameling van de natuurlijke, de gehele en de rationale getallen.
Eindterm 8	De leerlingen rekenen handig door gebruik te maken van eigenschappen en rekenregels van bewerkingen.
Eindterm 11	De leerlingen berekenen machten met grondtal 10 en 2 met gehele exponent. Zij passen hierop rekenregels van machten toe.
Eindterm 12	De leerlingen kunnen: <ul style="list-style-type: none">• de uitkomst van een bewerking schatten;• een resultaat oordeelkundig afronden.
Eindterm 13	De leerlingen kunnen procentberekeningen in zinvolle contexten.
Eindterm 15	De leerlingen kunnen het verband uitleggen tussen optellen en aftrekken, vermenigvuldigen en delen.

Algebra

Rekenen met veeltermen

Eindterm 19	De leerlingen kunnen twee- en drietermen optellen en vermenigvuldigen en het resultaat vereenvoudigen.
Eindterm 20	De leerlingen kennen de formules voor de volgende merkwaardige producten: $(a+b)^2$ en $(a+b)(a-b)$; ze kunnen ze verantwoorden en in beide richtingen toepassen.
Eindterm 21	De leerlingen kunnen vergelijkingen van de eerste graad met één onbekende oplossen.

Algebraïsering

Eindterm 18	De leerlingen gebruiken letters als middel om te veralgemenen en als onbekenden.
Eindterm 22	De leerlingen kunnen eenvoudige vraagstukken die te herleiden zijn tot een vergelijking van de eerste graad met één onbekende oplossen.
Eindterm 23	De leerlingen ontdekken regelmaat in eenvoudige patronen en schema's en kunnen ze beschrijven met formules.

Evenredigheden

Eindterm 16	De leerlingen herkennen het recht evenredig en omgekeerd evenredig zijn van twee grootheden in tabellen en in het dagelijks leven.
Eindterm 24	De leerlingen kunnen vanuit tabellen recht evenredige verbanden met formules uitdrukken.
Eindterm 39	De leerlingen stellen recht evenredige verbanden tussen grootheden grafisch voor.

Data

Omgaan met data

Eindterm 17	De leerlingen kunnen vanuit tabellen met cijfergegevens het rekenkundig gemiddelde en de mediaan (voor niet-gegroepeerde gegevens) berekenen en hieruit relevante informatie afleiden.
Eindterm 25	De leerlingen kunnen functioneel gebruik maken van eenvoudige schema's, figuren, tabellen en diagrammen.

Meetkunde

Meetkundige begripsvorming

Eindterm 26	De leerlingen kennen en gebruiken de meetkundige begrippen diagonaal, bissectrice, hoogtelijn, middelloodlijn, straal, middellijn, overstaande hoeken, nevenhoeken, aanliggende hoeken, middelpuntshoeken.
Eindterm 27	De leerlingen herkennen evenwijdige stand, loodrechte stand en symmetrie in vlakke figuren en ze herkennen gelijkvormigheid en congruentie tussen vlakke figuren.
Eindterm 28	De leerlingen herkennen figuren in het vlak, die bekomen zijn door een verschuiving, een spiegeling of een draaiing.
Eindterm 31	De leerlingen kennen meetkundige eigenschappen zoals: de hoekensom in driehoeken en vierhoeken, eigenschappen van gelijkzijdige en gelijkbenige driehoeken, eigenschappen van zijden, hoeken en diagonalen in vierhoeken.
Eindterm 37	De leerlingen beschrijven en classificeren de soorten driehoeken en de soorten vierhoeken aan de hand van eigenschappen.
Eindterm 40	De leerlingen begrijpen een gegeven eenvoudige redenering of argumentatie in verband met eigenschappen van meetkundige figuren.

Meetkundige procedures: rekenen

Eindterm 32	De leerlingen kiezen geschikte eenheden en instrumenten om afstanden en hoeken te meten of te construeren met de gewenste nauwkeurigheid.
Eindterm 33	De leerlingen gebruiken het begrip schaal om afstanden in meetkundige figuren te berekenen.
Eindterm 34	De leerlingen berekenen de omtrek en oppervlakte van driehoek, vierhoek en cirkel en berekenen de oppervlakte en het volume van kubus, balk en cilinder.

Meetkundige procedures: constructies

- Eindterm 35 De leerlingen kunnen:
- het beeld bepalen van een eenvoudige vlakke meetkundige figuur door een verschuiving, spiegeling, draaiing;
 - symmetrieassen van vlakke figuren bepalen;
 - loodlijnen, middelloodlijnen en bissectrices construeren.

Eindterm 38 De leerlingen bepalen punten in het vlak door middel van coördinaten.

Ruimteteekunde

Eindterm 29 De leerlingen weten dat in een tweedimensionale voorstelling van een driedimensionale situatie, informatie verloren gaat.

Eindterm 30 De leerlingen herkennen kubus, balk, recht prisma, cilinder, piramide, kegel en bol aan de hand van een schets, tekening en dergelijke.

Eindterm 36 De leerlingen kunnen zich vanuit diverse vlakke weergaven een beeld vormen van een eenvoudige ruimtelijke figuur met behulp van allerlei concreet materiaal.

2.1.2.2 Peiling wiskunde in de B-stroom

De ontwikkelingsdoelen⁴³ wiskunde voor de B-stroom vormen het uitgangspunt bij deze peiling. “Ontwikkelingsdoelen voor het gewoon voltijds secundair onderwijs zijn minimumdoelen op vlak van **kennis, inzicht, vaardigheden en attitudes** die het Vlaams Parlement wenselijk acht voor een bepaalde leerlingenpopulatie en die de school bij haar leerlingen moet nastreven” (*Codex Secundair Onderwijs*, 2010, Art. 141. §1.). In tegenstelling tot de eindtermen moeten de ontwikkelingsdoelen dus niet bereikt, maar nagestreefd worden. De geselecteerde ontwikkelingsdoelen wiskunde zijn ingedeeld in acht clusters en veertien subclusters. Tabel 31 toont een overzicht van de geselecteerde ontwikkelingsdoelen voor de peiling wiskunde in de B-stroom (Carpentier et al., 2020a, 2020b). Elk ontwikkelingsdoel wordt bovendien gelinkt aan een of meerdere van volgende cognitieve verwerkingsniveaus: begripsvorming, procedures uitvoeren of functioneel toepassen (Willem, Harnisfeger, et al., 2018b).

Tabel 31

De geselecteerde ontwikkelingsdoelen wiskunde

Getalinzicht en hoofdbewerkingen

- | | |
|---------------------|--|
| Ontwikkelingsdoel 6 | De leerlingen hebben inzicht in de relatie tussen breuk, decimaal getal en percent. |
| Ontwikkelingsdoel 7 | De leerlingen kunnen hoofdbewerkingen met natuurlijke getallen maken, met inbegrip van de nulmoeilijkheid. |
| Ontwikkelingsdoel 9 | De leerlingen kunnen hoofdbewerkingen met een decimaal getal en een natuurlijk getal maken. |

⁴³ Dit peilingsonderzoek richt zich op de ontwikkelingsdoelen zoals die golden in schooljaar 2018-2019. Naar aanleiding van de hervorming van het secundair onderwijs zijn er ondertussen nieuwe eindtermen in plaats van ontwikkelingsdoelen ingevoerd in de B-stroom (AHOVOKS, n.d.).

Breuken optellen en aftrekken

Ontwikkelingsdoel 8	De leerlingen kunnen breuken optellen en aftrekken waarbij het resultaat een breuk is met een noemer kleiner dan of gelijk aan 16.
---------------------	--

Geld en functioneel rekenen

Functioneel rekenen in praktische situaties

Ontwikkelingsdoel 10	De leerlingen kunnen de hoofdbewerkingen in verschillende situaties toepassen.
Ontwikkelingsdoel 11	De leerlingen kunnen grootheden en resultaten van bewerkingen schatten en zinvol afronden.
Ontwikkelingsdoel 13	De leerlingen kunnen met verhoudingen en percenten in praktische situaties werken.

Geld

Ontwikkelingsdoel 50	De leerlingen kunnen in reële situaties rekenen met geld.
----------------------	---

Meetkunde

Visualiteit en percepto-motoriek

Ontwikkelingsdoel 2	De leerlingen kunnen figuren herkennen, aanvullen, samenstellen en ordenen.
Ontwikkelingsdoel 3	De leerlingen kunnen een tweedimensionele tekening verkleind, vergroot tekenen met behulp van een raster.
Ontwikkelingsdoel 4	De leerlingen kunnen een tweedimensionele tekening spiegelen om een verticale en horizontale as met behulp van een raster.
Ontwikkelingsdoel 5	De leerlingen kunnen een ontwikkeling maken van een driedimensioneel lichaam.

Lijnen en hoeken

Ontwikkelingsdoel 26	De leerlingen kunnen een lijnstuk tekenen.
Ontwikkelingsdoel 27	De leerlingen kunnen de lengte nauwkeurig meten.
Ontwikkelingsdoel 28	De leerlingen herkennen de onderlinge stand van rechten en kunnen rechten tekenen waarvan de onderlinge stand beschreven is.
Ontwikkelingsdoel 29	De leerlingen kunnen de elementen van een hoek aanduiden en benoemen.
Ontwikkelingsdoel 30	De leerlingen kunnen de hoeken aanduiden en rubriceren (nulhoek, scherpe hoek, stompe hoek, gestrekte hoek, volle hoek).
Ontwikkelingsdoel 31	De leerlingen kunnen hoeken meten en tekenen.

Vlakke figuren en ruimtelijke figuren herkennen, classificeren en tekenen

Ontwikkelingsdoel 32	De leerlingen kunnen figuren indelen in vlakke figuren en ruimtelijke figuren.
Ontwikkelingsdoel 33	De leerlingen kunnen vlakke figuren indelen in veelhoeken en figuren die geen veelhoeken zijn.
Ontwikkelingsdoel 34	De leerlingen kunnen veelhoeken classificeren volgens het aantal hoeken en zijden.
Ontwikkelingsdoel 35	De leerlingen kunnen driehoeken classificeren met als criteria het aantal gelijke zijden of hoeken.
Ontwikkelingsdoel 36	De leerlingen kunnen driehoeken tekenen, waarvan een aantal voorwaarden in verband met gelijkheid van zijden of hoeken gegeven zijn.
Ontwikkelingsdoel 37	De leerlingen kunnen vierhoeken classificeren met als criteria het aantal gelijke zijden, aantal paren evenwijdige zijden, aantal gelijke hoeken, eigenschappen van de diagonalen.
Ontwikkelingsdoel 38	De leerlingen kunnen vierhoeken tekenen, waarvan een aantal voorwaarden in verband met gelijkheid van zijden of hoeken gegeven zijn.
Ontwikkelingsdoel 40	De leerlingen kunnen een cirkel tekenen.
Ontwikkelingsdoel 42	De leerlingen herkennen een kubus en een balk.
Ontwikkelingsdoel 43	De leerlingen herkennen een piramide, cilinder, kegel en bol.

Berekenen van omtrek, oppervlakte en inhoud

Ontwikkelingsdoel 39	De leerlingen kunnen de omtrek en oppervlakte van een driehoek, vierkant en rechthoek berekenen.
Ontwikkelingsdoel 41	De leerlingen kunnen met gegeven formule de omtrek en oppervlakte van een cirkel berekenen.
Ontwikkelingsdoel 44	De leerlingen kunnen met gegeven formule de inhoud van een kubus en een balk berekenen.

Informatieverwerking en -verwerking

Tabellen, grafieken, diagrammen en gemiddelden

Ontwikkelingsdoel 45a	De leerlingen kunnen informatie halen uit grafieken, tabellen.
Ontwikkelingsdoel 48	De leerlingen kunnen een rekenkundig gemiddelde berekenen.

Schaal

Ontwikkelingsdoel 45b	De leerlingen kunnen informatie halen uit kaarten en schaalmodellen.
Ontwikkelingsdoel 46	De leerlingen kunnen met plattegronden en plan werken.
Ontwikkelingsdoel 47	De leerlingen hebben inzicht in het schaalbegrip.
Ontwikkelingsdoel 49	De leerlingen kunnen met tekeningen en modellen op schaal werken.

Meten

Begrijpen en meten van grootheden

- | | |
|-----------------------|---|
| Ontwikkelingsdoel 18 | De leerlingen kunnen twee of meer gelijksoortige objecten vergelijken en ordenen zonder gebruik te maken van een maateenheid. |
| Ontwikkelingsdoel 19 | De leerlingen kennen de begrippen omtrek, oppervlakte, volume, inhoud, massa, tijd, temperatuur en hoekgrootte. |
| Ontwikkelingsdoel 20 | De leerlingen kennen de belangrijkste eenheden en kunnen de symbolen daarvan juist gebruiken. |
| Ontwikkelingsdoel 24a | De leerlingen kunnen grootheden meten. |

Rekenen met grootheden

- | | |
|-----------------------|--|
| Ontwikkelingsdoel 21 | De leerlingen zien het verband tussen de verandering in de eenheid en de verandering bij het maatgetal bij herleidingen. |
| Ontwikkelingsdoel 22 | De leerlingen kunnen eenvoudige vraagstukken in verband met omtrek, oppervlakte, inhoud, massa, tijd, temperatuur en hoekgrootte oplossen. |
| Ontwikkelingsdoel 24b | De leerlingen kunnen grootheden berekenen. |

Bewerkingen met zakrekenmachine

- | | |
|----------------------|--|
| Ontwikkelingsdoel 14 | De leerlingen kunnen met een zakrekenmachine optellen, aftrekken, vermenigvuldigen en delen. |
|----------------------|--|

Procenten met zakrekenmachine

- | | |
|----------------------|---|
| Ontwikkelingsdoel 16 | De leerlingen kunnen met een zakrekenmachine een percent nemen van een getal. |
|----------------------|---|
-

2.1.2.3 Achtergrondkenmerken

Net zoals bij het peilingsonderzoek in het lager onderwijs (zie 1.1.2.3 Achtergrondkenmerken) worden er verschillende achtergrondkenmerken gemeten. Concreet gaat het om: (1) leerlinggegevens, (2) gegevens over de ouders en het gezin, (3) gegevens over de leerkracht en het onderwijsaanbod wiskunde en (4) de administratieve gegevens van de school. Tabel 32 geeft een overzicht weer van alle achtergrondkenmerken die het peilingonderzoek meet (Carpentier et al., 2019a, 2020a). Onderstaande kenmerken vormen niet de focus van het peilingsonderzoek, maar worden hoofdzakelijk gebruikt om verschillen in prestaties tussen leerlingen te duiden enerzijds en het huidige onderwijslandschap in kaart te brengen anderzijds (Carpentier et al., 2019a, 2020a).

Tabel 32

Overzicht van achtergrondgegevens bij de peilingen wiskunde in de A- en B-stroom

Leerlinggegevens
<ul style="list-style-type: none">• basisoptie leerlingen (A-stroom) of beroepenveld leerlingen (B-stroom);• geslacht en leeftijd;• thuistaal en geboorteland;• leerproblemen;• schoolloopbaan;• motivatie;• mening over wiskunde;• vragen over de moeilijkheidsgraad van de peilingstoets, de inzet van de leerlingen en het belang van de toets voor de leerlingen.
Gegevens over de ouders en het gezin
<ul style="list-style-type: none">• tewerkstelling en opleidingsniveau ouders;• SES;• cultureel kapitaal;• ouderbetrokkenheid.
Gegevens over de leerkracht en het onderwijsaanbod wiskunde
<ul style="list-style-type: none">• profiel van de leerkracht;• wiskundelessen:<ul style="list-style-type: none">○ organisatie van de lessen^b;○ teamteaching^b;○ lesmateriaal en ICT;○ activiteiten tijdens de lessen wiskunde;○ domeinen wiskunde;○ evaluatie van wiskunde;○ STEM in de klas^b;• gebruik hulpmiddelen^b;• samenwerking en leerlijnen;• STEM beleid^a;• behandelen van de eindtermen (A-stroom) of ontwikkelingsdoelen (B-stroom).
Administratieve gegevens van de school
<ul style="list-style-type: none">• schooltype;• onderwijsnet;• verstedelijkingsgraad;• percentage GOK-leerlingen.

Noot.

^a Enkel bevroegd bij de peiling in de A-stroom.

^b Enkel bevroegd bij de peiling in de B-stroom.

2.1.3 Populatie

De peilingen wiskunde in de A- en B-stroom willen nagaan welke aandeel van de leerlingen de onderwijsdoelen voor wiskunde beheerst op het einde van de eerste graad secundair onderwijs. Daarom zijn die peilingen afgenomen bij leerlingen uit het tweede leerjaar van de eerste graad A-stroom enerzijds en bij leerlingen uit het beroepsvoorbereidend leerjaar van de eerste graad B-stroom

(Carpentier et al., 2019a, 2020a). Op het moment van de afname (in mei) hebben die leerlingen – op enkele weken na – het volledige curriculum van de eerste graad secundair onderwijs doorlopen. De keuze voor een specifiek leerjaar heeft als gevolg dat de populatie bestaat uit leerlingen van verschillende leeftijden.

2.1.4 Steekproef

De steekproeftrekking in het secundair onderwijs verloopt gelijkaardig aan die in het lager onderwijs (zie 1.1.4 Steekproef), weliswaar wordt er rekening gehouden met andere stratificatievariabelen. De stratificatievariabele ‘schooltype’ onderscheidt (1) autonome middensholen, (2) scholen met een multilaterale bovenbouw, (3) scholen met een ASO-bovenbouw en (4) scholen met een TSO-/BSO-/KSO-bovenbouw. Wat betreft de stratificatievariabele ‘onderwijsnet’ wordt er een onderscheid gemaakt tussen het officieel onderwijs en het vrij onderwijs. De laatste stratificatievariabele verwijst naar de verstedelijkingsgraad van de gemeente waarin de school in kwestie gevestigd is. Die variabelen onderscheidt twee categorieën, namelijk stad en niet-verstedelijkt gebied (Carpentier et al., 2019a, 2020a).

De Vlaamse populatie van scholen voor secundair onderwijs wordt aan de hand van de expliciete stratificatievariabelen, namelijk schooltype en onderwijsnet, opgedeeld in de volgende acht deelpopulaties of strata⁴⁴ (Carpentier et al., 2019a, 2020a):

- autonome middensholen behorende tot het officieel onderwijs;
- autonome middensholen behorende tot het vrij onderwijs;
- scholen met een ASO-bovenbouw behorende tot het officieel onderwijs;
- scholen met een ASO-bovenbouw behorende tot het vrij onderwijs;
- scholen met een multilaterale bovenbouw behorende tot het officieel onderwijs;
- scholen met een multilaterale bovenbouw behorende tot het vrij onderwijs;
- scholen met een TSO-/BSO-/KSO-bovenbouw behorende tot het officieel onderwijs;
- scholen met een TSO-/BSO-/KSO-bovenbouw behorende tot het vrij onderwijs.

Binnen elk stratum worden de scholen gesorteerd op basis van de impliciete stratificatievariabele ‘verstedelijkingsgraad’. At random wordt er een eerste school getrokken binnen elk stratum. Daarna worden de andere scholen binnen het stratum getrokken op basis van een vooraf vastgelegd interval. Dat interval werd berekend door voor elk stratum het totale aantal scholen in dat stratum te delen door het aantal scholen dat nodig was voor de steekproef van dat stratum (Carpentier et al., 2019a, 2020a).

Ook bij het peilingsonderzoek in het secundair onderwijs wordt er een reservesteekproef getrokken rekening houdend met de respons per stratum uit de eerste steekproef. De feitelijke steekproef bestond uit 2985 leerlingen uit 104 scholen in de A-stroom en 3615 leerlingen uit 117 scholen in de B-stroom. Tabel 33 geeft de strata of deelpopulaties weer binnen de feitelijke steekproef. Die steekproef is representatief op vlak van schooltype, onderwijsnet en verstedelijkingsgraad (Carpentier et al., 2019a, 2020a).

⁴⁴ Bij de peiling wiskunde in de B-stroom worden de scholen met een ASO-bovenbouw en multilaterale bovenbouw samengevoegd. Dat heeft als gevolg dat er bij die peiling slechts zes deelpopulaties of strata zijn.

Tabel 33*Aantal scholen en leerlingen per stratum in de steekproef*

Schooltype	Onderwijsnet	
	Officieel	Vrij
Peiling wiskunde in de A-stroom		
Autonome middenscholen	10 scholen 230 leerlingen	33 scholen 1124 leerlingen
ASO-bovenbouw	7 scholen 127 leerlingen	20 scholen 616 leerlingen
Multilaterale bovenbouw	5 scholen 102 leerlingen	6 scholen 222 leerlingen
TSO-/BSO-/KSO-bovenbouw	9 scholen 132 leerlingen	14 scholen 432 leerlingen
Peiling wiskunde in de B-stroom		
Autonome middenscholen	20 scholen 613 leerlingen	23 scholen 816 leerlingen
ASO- of multilaterale bovenbouw	6 scholen 103 leerlingen	8 scholen 162 leerlingen
TSO-/BSO-/KSO-bovenbouw	22 scholen 672 leerlingen	38 scholen 1249 leerlingen

Verder bevaart het peilingonderzoek ook de algemene opvatting van de leerkrachten wiskunde (Carpentier et al., 2019a, 2020a). Net zoals bij het peilingsonderzoek in het lager onderwijs (zie 1.1.4 Steekproef) weerspiegelen de opvattingen van de leerkracht niet noodzakelijk de realiteit op school. In dat opzicht kan men zich vragen stellen bij het gebruik van variabelen uit de leerkrachtvragenlijst om uitspraken te doen over het schoolniveau enerzijds en over de samenhang van die variabelen met de leerlingprestaties anderzijds.

2.1.5 Modus

2.1.5.1 Toetsontwikkeling

De toetsontwikkeling van de peilingsinstrumenten in het secundair onderwijs verloopt op een gelijkaardige manier als bij het peilingsonderzoek in het lager onderwijs (zie 1.1.5.1 Toetsontwikkeling voor een uitgebreide beschrijving van het toetsontwikkelingsproces). Concreet worden er eerst toetsmatrijzen opgesteld op basis van de geselecteerde onderwijsdoelen. Vervolgens worden de nieuwe opgaven ontwikkeld. Daarna volgt het vooronderzoek bij experts enerzijds en de pilootafname bij een beperkte steekproef leerlingen anderzijds. Na het kalibratieonderzoek worden tot slot de definitieve toetsen en parallelversies samengesteld met zowel nieuwe opgaven als ankeropgaven (Willem et al., 2017; Willem, Harnisfeger, et al., 2018a, 2018b; Willem, Van Nijlen, et al., 2018). Tabel 34 en Tabel 35 tonen het aantal opgaven per toets bij de respectievelijk de peiling in de A- en B-stroom. De samenstelling van de toetsboekjes wordt weergegeven in Tabel 36 en Tabel 37 (Carpentier et al., 2019a, 2020a).

Tabel 34*Overzicht van aantal opgaven per toets bij de peiling wiskunde in de A-stroom*

Toets	Aantal opgaven
Getallenleer	
1. Getalinzicht	28
2. Bewerkingen	20
Algebra	
3. Rekenen met veeltermen	21
4. Algebraïsering	28
5. Evenredigheden	23
Data	
6. Omgaan met data	21
Meetkunde	
7. Meetkundige begripsvorming	20
8. Meetkundige procedures: rekenen	22
9. Meetkundige procedures: constructies	20
10. Ruimte meetkunde	22
Totaal	225

Tabel 35*Overzicht van aantal opgaven per toets bij de peiling wiskunde in de B-stroom*

Toets	Aantal opgaven
1. Getalinzicht en hoofdbewerkingen ^a	20
2. Breuken optellen en aftrekken	24
3. Geld en functioneel rekenen	22
4. Meetkunde	20
5. Informatieverwerking en -verwerking	20
6. Meten	20
7. Bewerkingen met zakrekenmachine ^b	20
8. Procenten met zakrekenmachine ^b	17
Totaal	163

Noot.

^a Bij die toets werden in 2008 alleen cijferopgaven aangeboden, terwijl leerlingen in 2019 mochten kiezen tussen de opgave cijferend of uit het hoofd op te lossen. Door die inhoudelijk andere focus kan er geen vergelijking gemaakt worden met de vorige peiling. Dat heeft als gevolg dat er een nieuwe cesuur vastgelegd moest worden (Willem, Harnisfeger, et al., 2018b).

^b Die toetsen bevatten onvoldoende ankeritems uit de peiling van 2008 om de cesuur te kunnen overnemen uit de vorige peiling. Daardoor moest er een nieuwe cesuur vastgelegd worden voor die toets (Willem, Harnisfeger, et al., 2018b).

Tabel 36*Samenstelling toetsboekjes bij de peiling wiskunde in de A-stroom*

Toets	Toetsboekje	
	1	2
1. Getalinzicht		
2. Bewerkingen		
3. Rekenen met veeltermen		
4. Algebraïsering		
5. Evenredigheden		
6. Omgaan met data		
7. Meetkundige begripsvorming		
8. Meetkundige procedures: rekenen		
9. Meetkundige procedures: constructies		
10. Ruimte meetkunde		

Tabel 37*Samenstelling toetsboekjes bij de peiling wiskunde in de B-stroom*

Toets	Toetsboekje				
	1	2	3	4	5
1. Getalinzicht en hoofdbewerkingen					
2. Breuken optellen en aftrekken					
3. Geld en functioneel rekenen					
4. Meetkunde					
5. Informatieverwerving en -verwerking ^a					
6. Meten ^a					
7. Bewerkingen met zakrekenmachine					
8. Procenten met zakrekenmachine					

Noot. ^aDe leerlingen uit toetsboekje vier mochten bij de toetsen 'informatieverwerving en -verwerking' en 'meten' hulpmiddelen gebruiken in tegenstelling tot hun medeleerlingen met andere toetsboekjes. De resultaten op die toetsen met en zonder gebruik van hulpmiddelen werden daarna vergeleken (Carpentier et al., 2020a).

2.1.5.2 Toetsafname

De schriftelijke toetsen voor wiskunde in de A-stroom werden afgenomen op 15, 16 en 17 mei 2018. In totaal kregen de leerlingen vier lesuren verspreid over twee dagen om het toegewezen toetsboekje in te vullen. Bij de overgrote meerderheid van de opgaven mogen leerlingen een zakrekenmachine gebruiken. Daarbij werden redelijke aanpassingen toegelaten voor leerlingen met specifieke onderwijsbehoeften (Carpentier et al., 2019a).

In de B-stroom vond de peiling plaats op 16 mei 2019. Er werden twee lesuren voorzien met daartussen een pauze om een toetsboekje in te vullen. De leerlingen mochten bij alle clusters een zakrekenmachine gebruiken behalve bij de volgende (sub)clusters: (1) getalinzicht en hoofdbewerkingen, (2) breuken optellen en aftrekken en (3) geld. Ook bij deze peiling werden redelijke aanpassingen toegelaten voor leerlingen met specifieke onderwijsbehoeften (Carpentier et al., 2020a).

Net zoals bij het peilingsonderzoek in het lager onderwijs (zie 1.1.5.2 Toetsafname), gebeurde de toetsafname klassikaal onder begeleiding van een leerkracht. Met het oog op de standaardisatie werd er een handleiding uitgewerkt voor de leerkrachten waarin gedetailleerd omschreven staat hoe de afname verloopt, wat ze juist moeten doen en welke instructies ze moeten geven aan de leerlingen. Aan de leerkrachten werd ook gevraagd om onregelmatigheden en afwezigheden te noteren in het

logboek. Per vestigingsplaats was er daarnaast een toetsassistent aanwezig om de leerkracht te ondersteunen. Zij keken toe op het correcte verloop van de afname en de geheimhouding van het toetsmateriaal. Na de toetsafname verzamelden de toetsassistenten alle documenten, stuurden ze die op naar het onderzoeksteam en brachten ze kort verslag uit over de toetsafname (Carpentier et al., 2019b, 2020b).

Naast de toetsen wiskunde worden er achtergrondvragenlijsten afgenomen bij leerlingen, ouders en leerkrachten wiskunde. De afname van de leerlingenvragenlijst verliep anders voor de twee peilingen. Bij de peiling in de A-stroom werd de leerlingenvragenlijst afgenomen samen met de schriftelijke toetsen wiskunde. In de B-stroom werd de leerlingvragenlijst enkele dagen voor de afnamedag klassikaal ingevuld door de deelnemende leerlingen. De afname van de ouder- en leerkrachtvragenlijst verliep wel op dezelfde manier bij beide peilingen. Er werd namelijk aan de ouders gevraagd om een vragenlijst op papier in te vullen, terwijl bij de leerkrachten een online vragenlijst werd afgenomen (Carpentier et al., 2019b, 2020b).

Vervolgens werden de antwoorden van de leerlingen, ouders en leerkrachten wiskunde ingevoerd en gecontroleerd. Aan de hand van scoringswijzers werden de antwoorden van de leerlingen op de schriftelijke toetsen gescoord (Carpentier et al., 2019b, 2020b).

2.2 Standaardbepaling

De standaardbepaling bij het peilingsonderzoek in het secundair onderwijs gebeurt op dezelfde manier als in het lager onderwijs (zie 1.2 Standaardbepaling voor een uitgebreide toelichting). Concreet wordt de cesuur uit de voorgaande peilingen – in dit geval: de peilingen uit 2008 en 2009 – overgenomen door gebruik te maken van ankeritems.⁴⁵ Vervolgens kan er op basis van die cesuur bepaald worden welke leerlingen de onderwijsdoelen behalen. De leerlingen die op de meetschaal boven de cesuur gesitueerd zijn, behalen de onderwijsdoelen (Carpentier et al., 2019a, 2020a).

2.3 Analytische methode

De analytische methode van het peilingsonderzoek in het secundair onderwijs is vergelijkbaar met de methode die gebruikt wordt in het lager onderwijs. Ten eerste worden bij de peiling wiskunde in de A- en B-stroom de vaardigheidsscores ook geschat aan de hand van *Warm's weighted likelihood estimation* (zie 1.3.1 Schatting van vaardigheidsscores voor een uitgebreide beschrijving). Ten tweede worden er gelijkaardige modellen gebruikt. Concreet wordt er net zoals bij het lager onderwijs gebruikt gemaakt van een leeg model, nettomodel en model voor schoolfeedback (zie 1.3.2 De gebruikte modellen voor een algemene toelichting). Ook wordt er onderzocht of bijkomende kenmerken van de leerling, klas, leerkracht of school samenhangen met de prestaties van de leerlingen na controle voor de variabelen die reeds opgenomen zijn in het nettomodel (Carpentier et al., 2019a, 2020a).

Er zijn echter enkele kleine verschillen met de gebruikte modellen uit het lager onderwijs. Ten eerste wordt bij de peiling van de A-stroom in het **leeg model** de basisoptie meegenomen, omdat er bij de data uit de A-stroom een aanzienlijk selectiemechanisme voorkomt. In een specifieke school kunnen immers enkel leerlingen uit bepaalde basisopties (die de school aanbiedt) geselecteerd worden. Ten tweede worden er andere variabelen opgenomen in het **nettomodel** voor de A- en B-stroom in

⁴⁵ Bij het peilingonderzoek wiskunde in de B-stroom werden drie clusters vernieuwd ten opzichte van de peiling in 2008, namelijk (1) getalinzicht en hoofdbewerkingen, (2) bewerkingen met zakrekenmachine en (3) procenten met zakrekenmachine. Die clusters bevatten onvoldoende ankeritems uit de peiling van 2008 of hebben een andere focus ten opzichte van 2008. Daardoor kan er voor die clusters geen vergelijking gemaakt worden met de voorgaande peiling in de B-stroom. Met als gevolg, ook de cesuur kan niet overgenomen worden uit de eerdere peiling. Daarom werd er een nieuwe cesuur vastgelegd voor die clusters door middel van de bookmarkmethode die reeds besproken is bij het peilingonderzoek in het lager onderwijs (Carpentier et al., 2020a).

vergelijking met het lager onderwijs. Concreet worden op schoolniveau de variabelen (1) schooltype, (2) onderwijsnet, (3) verstedelijkingsgraad en (4) percentage GOK-leerlingen opgenomen. Op leerlingniveau gaat het om de volgende variabelen: (1) basisoptie, (2) geslacht, (3) thuistaal, (4) schoolse achterstand, (5) SES, (6) aantal boeken thuis en (7) leerproblemen. Ten derde maakt het **model voor schoolfeedback** gebruik van het leeg model en een model dat sterkt lijkt op het nettomodel waarbij de stratificatievariabelen niet in rekening gebracht worden op het schoolniveau. Aangezien er bij de peiling in de A- en B-stroom andere stratificatievariabelen gebruikt werden (namelijk schooltype, onderwijsnet en verstedelijkingsgraad) dan bij het lager onderwijs, is dat model verschillend van het model voor schoolfeedback in het lager onderwijs (Carpentier et al., 2019a, 2020a).

2.4 Besluitvorming

De besluitvorming op basis van het peilingsonderzoek in het secundair onderwijs kent dezelfde mogelijkheden en uitdagingen als in het lager onderwijs. Die mogelijkheden en uitdagingen worden uitgebreid toegelicht in bij onderdeel '1.4 Besluitvorming'.

2.5 Het peilingsonderzoek doorheen de tijd

De verschillende mijlpalen in het peilingsonderzoek worden uitgebreid toegelicht bij het onderdeel '1.6 Het peilingsonderzoek doorheen de tijd'. Bij die toelichting is er aandacht voor zowel het lager als secundair onderwijs.

3. PIRLS 2016

3.1 Design

3.1.1 Doel

De Progress in International Reading Literacy Study (PIRLS) wordt elke vijf jaar georganiseerd door de International Association for the Evaluation of Educational Achievement (IEA). PIRLS is een internationaal vergelijkend onderzoek dat leerlingprestaties voor begrijpend lezen in verschillende landen in kaart brengt. De IEA wil door middel van PIRLS het onderwijsbeleid informeren over hoe leerlingen presteren in hun onderwijssysteem, de onderwijskwaliteit wereldwijd verbeteren en internationale samenwerking tussen landen bevorderen. PIRLS 2016 werd in Vlaanderen uitgevoerd en gecoördineerd door het Centrum voor Onderwijseffectiviteit en -evaluatie van de KU Leuven (Tielemans et al., 2017).

Hoewel de IEA met PIRLS bovenstaande doelen nastreeft, wordt PIRLS in Vlaanderen voornamelijk ingezet om aan **systeemmonitoring** te doen. Ten eerste laat PIRLS het toe om de gemiddelde prestatie voor begrijpend lezen van leerlingen enerzijds en het aandeel leerlingen per prestatieniveau anderzijds voor het Vlaamse onderwijs te vergelijken met andere landen. Ten tweede brengt PIRLS de prestatieverschillen tussen leerlingengroepen in kaart. Zo worden bijvoorbeeld de verschillen in begrijpend lezen nagaan tussen leerlingen uit het gewoon en buitengewoon lager onderwijs. Ook wordt de samenhang onderzocht tussen kenmerken uit de thuiscontext, schoolcontext en klascontext enerzijds en de gemiddelde prestatie voor begrijpend lezen van een bepaald land anderzijds. Ten derde wordt onderzocht in welke mate de deelnemende scholen in Vlaanderen verschillen van elkaar wat betreft begrijpend lezen. Concreet wordt nagegaan hoe groot de prestatieverschillen voor begrijpend lezen tussen scholen, klassen en leerlingen zijn. Tot slot kunnen trends inzake begrijpend lezen bij vierdeklassers doorheen de tijd opgevolgd worden aan de hand van PIRLS. Vlaanderen nam twee keer deel aan PIRLS, in 2006 en in 2016. Daardoor kan de evolutie in leerlingprestaties tussen 2006 en 2016

in Vlaanderen beschreven worden en vergeleken worden met de evolutie in andere landen (Mullis et al., 2017; Tielemans et al., 2017).

Doordat onderwijssystemen uit verschillende landen steeds ingebed zijn in een specifieke cultuur en context, kunnen de resultaten van diverse onderwijssystemen niet zomaar met elkaar vergeleken worden. Toch kan een vergelijking met andere landen waardevol zijn. Het kan het onderwijsbeleid inzicht geven in de sterke en zwakke punten van het eigen onderwijssysteem of inspireren door na te gaan hoe andere onderwijssystemen omgaan met specifieke uitdagingen. Het overnemen van beleidsmaatregelen uit andere landen heeft echter weinig zin. Er is steeds een vertaalslag nodig, waarbij rekening gehouden wordt met de eigen context, om tot succesvolle maatregelen voor het Vlaamse onderwijs te komen (Tielemans et al., 2019).

Bij PIRLS is het mogelijk om bijkomende onderzoeksvragen op te nemen die specifiek gericht zijn op het Vlaamse onderwijs naast de internationale onderzoeksvragen die de IEA vooropstelt. Zo werd er in PIRLS 2016 voor Vlaanderen onderzocht of er een verschil was tussen de Vlaamse groep leerlingen die deelnam aan PIRLS 2006 en de groep die deelnam aan PIRLS 2016 (Tielemans et al., 2017).

Schoolfeedback wordt niet expliciet vermeld als een doel dat PIRLS vooropstelt. In Vlaanderen wordt ervoor gekozen om alle deelnemende scholen een schoolfeedbackrapport te bezorgen. In dat rapport worden eerst de resultaten op de toets begrijpend lezen voor het vierde leerjaar van een bepaalde school weergegeven. Die resultaten worden vergeleken met het Vlaamse gemiddelde en met scholen die een gelijkaardig leerlingenpubliek hebben. Daarnaast toont het rapport voor bepaalde leerlingenkenmerken in welke mate de school in kwestie verschilt van het gemiddelde leerlingenpubliek in het Vlaamse onderwijs. Ook de samenhang tussen die kenmerken en de prestaties van leerlingen wordt gepresenteerd. Tot slot bespreekt het rapport ook enkele beperkingen van PIRLS. De informatie uit het schoolfeedbackrapport kunnen scholen gebruiken voor zelfevaluatie en -reflectie in het kader van hun interne kwaliteitszorg. Echter, die informatie heeft zijn beperkingen en moet daarom gecombineerd worden met andere informatiebronnen. Bovendien is het van belang om enige voorzichtigheid aan de dag te leggen bij de interpretatie van de schoolfeedback, aangezien de foutenmarge⁴⁶ bij uitspraken over individuele scholen relatief groot is (Centrum voor Schoolfeedback, 2016; Tielemans et al., 2017).

3.1.2 Domein en construct

Begrijpend lezen (Engels: reading literacy) vormt het centrale construct bij PIRLS en wordt door de IEA als volgt gedefinieerd (Mullis & Martin, 2015):

Begrijpend lezen is de **vaardigheid** om geschreven taal te begrijpen en te gebruiken, zoals vereist door de maatschappij en/of het individu. Lezers kunnen betekenis construeren uit teksten van verschillende vormen. Ze lezen om te leren, om deel te nemen aan leesgroepen op school en in het alledaagse leven en voor het plezier (Tielemans et al., 2017, p. 5).

Aansluitend bij bovenstaande definitie wordt begrijpend lezen binnen PIRLS geoperationaliseerd in twee leesdoelen en vier leesprocessen. Er worden twee leesdoelen onderscheiden om de reden voor het lezen van de tekst aan te geven. De vier leesprocessen geven aan hoe leerlingen op basis van een tekst betekenis construeren. Die leesdoelen en -processen functioneren in realiteit niet los van elkaar, maar worden in het conceptueel kader van PIRLS wel apart omschreven. De definities van de leesdoelen en -processen uit het conceptueel kader zijn opgenomen in Tabel 38 (Mullis & Martin, 2015;

⁴⁶ Scholen met minimum vijf leerlingen in het vierde leerjaar kunnen deelnemen aan PIRLS (Martin et al., 2017). Een beperkt aantal leerlingen per school leidt echter tot relatief grote foutenmarges bij resultaten op schoolniveau.

Tielemans et al., 2017). Het conceptueel kader vormt de basis voor de ontwikkeling van de toetsen (zie 3.1.5.1 Toetsontwikkeling).

Tabel 38

Het conceptueel kader van PIRLS 2016

Leesdoelen
Leeservaring De verkenning van (narratieve) tekstfragmenten en lezen voor het plezier.
Informatieverwerking Uit een tekst, lijst, grafiek, afbeelding of diagram informatie halen om die daarna actief te gebruiken.

Leesprocessen
Focus leggen op en vinden van expliciet vermelde informatie Aandacht hebben voor zaken die letterlijk vermeld staan in teksten. Voorbeelden zijn: <ul style="list-style-type: none">• definities van bepaalde woorden of zinnen zoeken;• specifieke ideeën terugvinden in een tekst;• ...
Eenvoudige conclusies trekken Enkele stukken informatie uit een tekst verbinden. Voorbeelden zijn: <ul style="list-style-type: none">• de relatie tussen twee personages beschrijven;• oorzaak en gevolg verbinden aan elkaar;• ...
Tekstinterpretatie Een meer specifiek of compleet begrip van een tekst construeren door persoonlijke kennis en ervaringen te combineren met de betekenis die vervat zit in de tekst. Voorbeelden zijn: <ul style="list-style-type: none">• de algemene boodschap in een tekst vatten;• informatie uit de tekst contrasteren of vergelijken;• ...
Tekstevaluatie Kritisch kijken naar de tekst zelf. Voorbeelden zijn: <ul style="list-style-type: none">• de helderheid en volledigheid van een tekst beoordelen;• het perspectief van de auteur over het centrale onderwerp in de tekst bepalen;• ...

Noot. Elke cyclus wordt het conceptueel kader licht aangepast om zo het kader relevant te houden en eventuele evoluties te integreren. Het conceptueel kader van PIRLS 2016 is dus niet per se gelijk aan het conceptueel kader van andere cycli.

Bij de opstelling en aanpassing van het conceptueel kader worden de curricula uit de verschillende deelnemende landen als uitgangspunt genomen. De IEA bevraagt daarom de inhoud van de curricula uit alle deelnemende landen. Om inzicht te krijgen in de Vlaamse curricula, vulden vertegenwoordigers

van GO!, OVSG, Katholiek Onderwijs Vlaanderen en de Onderwijsinspectie een curriculumvragenlijst in (Tielemans et al., 2017). Daarna stelt een internationale commissie van vakspecialisten een 'internationaal curriculum' op dat voorgelegd wordt aan de deelnemende onderwijssystemen. Onderwerpen die in een groot aantal landen tot het curriculum behoren, worden opgenomen in het conceptueel kader. Onderwerpen die niet of slechts in een klein aantal landen tot het curriculum behoren, worden niet opgenomen in het conceptueel kader van PIRLS. Dat heeft als gevolg dat de PIRLS-toetsen ook vragen kunnen bevatten die niet aansluiten bij het Vlaamse curriculum (Centrum voor Schoolfeedback, 2016).

PIRLS meet niet alleen begrijpend lezen, maar ook achtergrondinformatie uit verschillende contexten: (1) de nationale context, (2) de schoolcontext, (3) de klascontext, (4) de thuiscontext en (5) de kenmerken en attitudes van leerlingen. Tabel 39 geeft de specifieke constructen weer die PIRLS per context meet. Meer gedetailleerde omschrijvingen van die constructen worden gepresenteerd in het PIRLS 2016 Context Assessment Framework (Mullis & Martin, 2015). Daarbij is het belangrijk om aan te geven dat de IEA een bepaalde achtergrondvariabele niet noodzakelijk op dezelfde manier operationaliseert als een bepaald onderwijssysteem. Zo beschrijft het Vlaams Ministerie van Onderwijs en Vorming (2012) de SES van leerlingen bijvoorbeeld aan de hand van de thuistaal, het ontvangen van een schooltoeslag en het opleidingsniveau van de moeder. De IEA daarentegen meet de SES van leerlingen aan de hand van de schaal 'Home Resources for Learning'. Die schaal bestaat uit vijf items: (1) aantal boeken thuis, (2) aantal kinderboeken thuis, (3) beschikken over toegang tot het internet en een eigen kamer, (4) opleidingsniveau ouders en (5) tewerkstellingsniveau ouders (Martin et al., 2017). De gemeten achtergrondvariabelen in PIRLS worden vooral gebruikt om verschillen tussen leerlingengroepen te duiden en om de samenhang met de gemiddelde prestaties voor een bepaald land na te gaan (Tielemans et al., 2017).

Tabel 39

Conceptueel kader voor achtergrondkenmerken bij PIRLS 2016

De nationale context
<ul style="list-style-type: none"> • talen en nadruk op geletterdheid; • economische middelen, demografische gegevens van de populatie en geografische kenmerken; • organisatie en structuur van het onderwijssysteem; • kleuteronderwijs, instapleeftijd lager onderwijs, graad van zittenblijven en groepering bij leesinstructie; • opleiding van de leerkrachten; • monitoring bij de implementatie van het curriculum.
Schoolcontext
<ul style="list-style-type: none"> • locatie van de school; • schoolsamenstelling volgens sociaaleconomische achtergrond van leerlingen • invloed van tekort aan middelen op de instructie; • werkcondities en jobtevredenheid bij leerkrachten; • leiderschap van de directeur; • nadruk van de school op academisch succes; • veilige, ordelijke en gedisciplineerde scholen.

Klascontext

- voorbereiding en ervaring bij leerkrachten;
- beschikbare middelen in de klas;
- instructietijd;
- engagement bij instructie;
- instructie bij online lezen;
- duidelijkheid van instructies;
- klasevaluatie.

Thuiscontext

- SES (via de schaal: Home Resources for Learning);
- thuistaal of -talen;
- verwachtingen van de ouders m.b.t. onderwijs;
- voorschoolse activiteiten ter stimulering van lees- en rekenvaardigheden;
- thuis ondersteuning bij lezen.

Kenmerken en attitudes van leerlingen

- honger en slaapdeprivatie bij leerlingen;
- motivatie van leerlingen;
- zelfconcept van leerlingen;
- gedrag van leerlingen gerelateerd aan hun leesvaardigheid.

Noot. Elke cyclus wordt het conceptueel kader voor achtergrondkenmerken geüpdatet. Bovenstaand conceptueel kader is dus niet per se gelijk aan het conceptueel kader voor achtergrondkenmerken bij andere cycli.

3.1.3 Populatie

PIRLS brengt internationaal de leerlingprestaties voor begrijpend lezen na vier jaar formele scholing in kaart. De IEA kiest voor die specifieke leerlingenpopulatie, omdat de nadruk in het vierde jaar van technisch naar begrijpend lezen verschuift (Mullis et al., 2017). In Vlaanderen komt dat overeen met leerlingen uit het vierde leerjaar lager onderwijs (ongeacht hun leeftijd). Normaalvorderende leerlingen uit het vierde leerjaar lager onderwijs zijn negen of tien jaar. Echter, leerlingen die versneld of vertraagd zijn, kunnen jonger of ouder zijn op het moment dat ze in het vierde leerjaar zitten (Tielemans et al., 2017).

Om het Vlaamse onderwijs zo goed mogelijk te weerspiegelen in PIRLS worden ook leerlingen uit het buitengewoon lager onderwijs opgenomen in de populatie. Concreet gaat het om leerlingen die de leerstof uit het vierde leerjaar onderwezen krijgen en les volgen in de (toenmalige) types 1, 3 en 8 van het buitengewoon lager onderwijs. Leerlingen die niet de leerstof van het vierde leerjaar volgen of tot andere types behoren, maken geen deel uit van de populatie (Tielemans et al., 2017). Internationaal is er beslist om anderstalige nieuwkomers niet op te nemen in de populatie, aangezien het onvoldoende beheersen van de onderwijstaal een te grote belemmering vormt bij de toetsafname (Martin et al., 2017).

3.1.4 Steekproef

De steekproeftrekking gebeurt in elk land op een gelijkaardige manier en daarbij worden dezelfde uitsluitingsnormen gehanteerd (Martin et al., 2017). In totaal namen er 45 landen en 10 regio's (ook wel *benchmarking participants* genoemd) deel aan PIRLS 2016 (Tielemans et al., 2017). Wereldwijd

omvat de PIRLS-steekproef ongeveer 319000 leerlingen, 310000 ouders, 16000 leerkrachten en 12000 scholen (Mullis et al., 2017).

De eerste stap omvat een **steekproeftrekking van scholen** in elk land, weliswaar met andere stratificatievariabelen per land (Martin et al., 2017). In Vlaanderen werd er een gestratificeerde steekproef van 150 scholen getrokken uit alle Vlaamse scholen voor lager onderwijs met een vierde leerjaar. Daarnaast werden er twee vervangscholen geselecteerd voor elk school in de steekproef (Tielemans et al., 2017).

Bij de steekproeftrekking in het Vlaamse onderwijs wordt er rekening gehouden met drie expliciete stratificatievariabelen, namelijk onderwijsnet, provincie en SES. De stratificatievariabele 'onderwijsnet' onderscheidt het officieel en het vrij onderwijs. Bij de stratificatievariabele 'provincie' wordt er een onderscheid gemaakt tussen Antwerpen, Brussels Hoofdstedelijk Gewest, Limburg, Oost-Vlaanderen, Vlaams-Brabant en West-Vlaanderen. De stratificatievariabele 'SES' maakt een onderscheid tussen scholen met een lage, eerder lage, eerder hoge en hoge SES. Scholen worden ingedeeld in één van die vier groepen op basis van het percentage leerlingen dat aantikt op de Onderwijskansarmoede Indicatoren (OKI)⁴⁷ (Tielemans et al., 2017).

Aangezien ook leerlingen uit het buitengewoon onderwijs tot de populatie behoren (zie 3.1.3 Populatie), werden in een apart stratum tien scholen voor buitengewoon lager onderwijs (type 1, 3 en 8) geselecteerd. Slechts zeven scholen, waarvan twee vervangscholen, namen deel aan PIRLS. Daardoor omvat de steekproef slechts 2,58% leerlingen uit het buitengewoon lager onderwijs, terwijl de Vlaamse populatie 3,82% leerlingen uit het buitengewoon lager onderwijs omvat. Om daarvoor te corrigeren, wordt er bij de analyses een groter gewicht toegekend aan de leerlingen uit het buitengewoon onderwijs (Tielemans et al., 2017). Tabel 40 toont een overzicht van het aantal scholen in de steekproef per stratum (Martin et al., 2017).

⁴⁷ De OKI omvat vier indicatoren: (1) moeder met een laag opleidingsniveau, (2) thuistaal niet Nederlands, (3) buurt met veel schoolse vertraging en (4) een schooltoelage ontvangen (Tielemans et al., 2017).

Tabel 40*Aantal scholen in de steekproef per stratum*

Stratum			Aantal scholen
Antwerpen	Lage SES	Officieel onderwijs	8
Antwerpen	Lage SES	Vrij onderwijs	8
Antwerpen	Eerder lage SES		7
Antwerpen	Eerder hoge SES		7
Antwerpen	Hoge SES		8
Brussels Hoofdstedelijk Gewest	Lage SES		7
Limburg	(Eerder) lage SES		8
Limburg	(Eerder) hoge SES		10
Oost-Vlaanderen	Lage SES		8
Oost-Vlaanderen	Eerder lage SES		8
Oost-Vlaanderen	Eerder hoge SES		6
Oost-Vlaanderen	Hoge SES		7
Vlaams-Brabant	(Eerder) lage SES		12
Vlaams-Brabant	(Eerder) hoge SES		11
West-Vlaanderen	(Eerder) lage SES		9
West-Vlaanderen	Eerder hoge SES		9
West-Vlaanderen	Hoge SES		8
Scholen voor buitengewoon onderwijs			7
Totaal			148

Daarbij werden de **internationale uitsluitingsnormen op schoolniveau** toegepast (Martin et al., 2017). Concreet werden in Vlaanderen de Franstalige basisscholen (304 leerlingen) en de scholen met minder dan vijf leerlingen in het vierde leerjaar (186 leerlingen) uit de steekproef verwijderd (Tielemans et al., 2017).

In een tweede stap volgt de **selectie van klassen** binnen de scholen (Martin et al., 2017). In Vlaanderen werden alle klassen uit het vierde leerjaar binnen de geselecteerde scholen opgenomen in de steekproef (Tielemans et al., 2017). Daarna werden de **internationale uitsluitingsnormen op individueel niveau** toegepast (Martin et al., 2017). In de Vlaamse steekproef werden 28 leerlingen uitgesloten, omwille van (1) ernstige fysieke beperkingen, (2) ernstige leer- en leesproblemen en (3) onvoldoende beheersing van de Nederlandse taal (doordat ze minder dan één jaar in Vlaanderen zijn). Om die uitsluitingsnormen te kunnen toepassen, werden de leerkrachten van alle deelnemende klassen bevroegd (Tielemans et al., 2017).

De IEA wil de uitsluitingen op school- en individueel niveau tot een minimum beperken en hanteert daarom de norm dat niet meer dan 5% van de doelpopulatie in een onderwijssysteem uitgesloten mag worden. In 2016 bedroeg het percentage uitsluitingen in Vlaanderen 1,6%. De Vlaamse steekproef blijft zo onder de vooropgestelde norm van de IEA (Martin et al., 2017). De feitelijke steekproef in Vlaanderen bestond uit 5378 leerlingen (waarvan 139 leerlingen uit het buitengewoon lager onderwijs) uit 148 scholen (waarvan 7 scholen voor buitengewoon onderwijs). In lijn met de stratificatievariabelen, is de steekproef van scholen representatief voor de Vlaamse populatie op vlak van onderwijsnet, provincie en SES (Tielemans et al., 2017).

PIRLS bevroegt ook leerkrachten en directeurs uit de deelnemende scholen. Al beperkt die leerkrachtenbevraging zich enkel tot de leerkracht uit het vierde leerjaar. Nochtans is de leerkracht uit het vierde leerjaar niet de enige op school die een invloed heeft gehad op de leesvaardigheden van leerlingen uit het vierde jaar. Bovendien is het van belang om te beseffen dat de opvattingen van de

bevroegde leerkrachten en directeurs niet noodzakelijk de realiteit op school weerspiegelen. We kunnen ons dus vragen stellen bij het gebruik van variabelen uit de leerkracht- en directievragenlijst om uitspraken te doen over de samenhang met de prestaties begrijpend lezen van de leerlingen.

3.1.5 Modus

De dataverzameling behoort tot de verantwoordelijkheid van de onderzoekscentra die instaan voor de coördinatie van PIRLS in de deelnemende landen. Ter ondersteuning van die onderzoekscentra en met oog op de standaardisatie heeft de IEA de procedure voor dataverzameling stapsgewijs beschreven in de PIRLS 2016 Survey Operations Procedures Units. Na de dataverzameling ontvangen de nationale onderzoekscentra een online vragenlijst⁴⁸ over de voorbereiding en uitvoering van de dataverzameling. Daarin wordt gevraagd om het doorlopen proces te omschrijven en eventuele afwijkingen van de vooropgestelde procedures aan te geven (Martin et al., 2017).

3.1.5.1 Toetsontwikkeling

Eerst is het **conceptueel kader van PIRLS geüpdatet** voor de afname in 2016 (zie 3.1.2 Domein en construct). Het geüpdatete conceptueel kader vormt de basis voor de volgende week stappen in de toetsontwikkeling. Ten eerste worden er op basis van de leesdoelen uit het conceptueel kader twaalf teksten opgenomen in de toets. Zes teksten sluiten aan bij het doel leeservaring, de andere teksten vallen onder het doel informatieverwerking. Ten tweede worden er met oog op het conceptueel kader voor elke tekst (met een specifiek leesdoel) bijhorende items ontwikkeld (in lijn met de leesprocessen). Met als gevolg, binnen elk leesdoel worden de vier leesprocessen bevroegd. Het conceptueel kader vormt op die manier dus de basis voor de toetsontwikkeling en draagt zo bij aan de inhoudelijke validiteit van de toetsen (Martin et al., 2017; Tielemans et al., 2017). Tabel 41 toont een overzicht van het vooropgestelde percentage items per leesdoel en -proces. De achterliggende redenen voor die specifieke richtpercentages worden niet vermeld in de rapporten van de IEA (Mullis & Martin, 2015).

Tabel 41

Overzicht vooropgesteld percentage items per leesdoel en -proces

Leesdoelen	%
Leeservaring	50
Informatieverwerking	50
Leesprocessen	%
Focus leggen op en vinden van expliciet vermelde informatie	20
Eenvoudige conclusies trekken	30
Tekstinterpretatie	30
Tekstevaluatie	20

De **selectie van gepaste teksten gebeurde** door een internationaal team van experts. Bij de tekstselectie was er aandacht voor verschillende criteria om zo de inhoudelijke validiteit te garanderen. Ten eerste moeten de teksten geschikt zijn voor leerlingen uit het vierde leerjaar wat betreft leesvaardigheid, inhoud en interesse. Ten tweede moeten de teksten voldoende complex en diepgaand geschreven zijn, zodat er voldoende items per tekst ontwikkeld kunnen worden. Een derde criteria is het vermijden van culturele bias. Teksten die gebaseerd zijn op cultuurspecifieke kennis worden niet opgenomen in de selectie. Ten vierde moet de tekst helder en coherent geschreven zijn. Tot slot moet het mogelijk zijn om op basis van de tekst een bepaald leesdoel en een hele waaier aan

⁴⁸ De online vragenlijst gaat in op de volgende thema's: (1) steekproeftrekking van scholen en klassen, (2) voorbereiding van de toetsinstrumenten, (3) toetsafname, (4) implementatie van de kwaliteitscontroles, (5) voorbereiding van de scoring van antwoorden op open vragen en (6) datasets creëren (Martin et al., 2017).

leesprocessen te evalueren. Op basis van bovenstaande criteria werden twaalf nieuwe teksten geselecteerd voor PIRLS 2016, naast de teksten⁴⁹ die overgenomen worden uit voorgaande afnames (Martin et al., 2017; Mullis & Martin, 2015).

In een volgende stap worden de **items en scorewijzers ontwikkeld** in lijn met het conceptueel kader. De leesprocessen vormen de basis voor de itemontwikkeling. Concreet zijn de items ontwikkeld met de richtinggevendende percentages voor de opdeling in leesprocessen (zie Tabel 41) in het achterhoofd om zo de inhoudelijke validiteit van de toetsen te garanderen. De items bestaan uit zowel open vragen als meerkeuzevragen. Bij de itemontwikkeling wordt er rekening gehouden met de bestaande teksten en de items uit de voorgaande afnames (ook wel trenditems genoemd) die opnieuw gebruikt zullen worden bij de afname in 2016. Na de ontwikkeling van de items worden er bijhorende scorewijzers opgesteld (Martin et al., 2017; Mullis & Martin, 2015).

Vervolgens wordt er een internationaal **vooronderzoek** georganiseerd om de psychometrische validiteit te verhogen en de betrouwbaarheid van de toets te evalueren. De twaalf nieuw ontwikkelde teksten en de bijhorende items werden uitgetest in maart en april 2015. De steekproef voor het vooronderzoek werd op hetzelfde moment en op een gelijkaardige manier getrokken als de steekproef voor het hoofdonderzoek (zie 3.1.4 Steekproef). In Vlaanderen namen 22 scholen deel aan het vooronderzoek. Op basis van de resultaten van het vooronderzoek bij alle deelnemende landen heeft de IEA de kwaliteit van de ontwikkelde items geanalyseerd. Zowel items met een te lage of te hoge moeilijkheidsgraad als items met een lage discriminatiegraad worden verwijderd. Ook items die voor een bepaald land ongebruikelijke statistieken vertonen, worden nader bekeken aangezien die statistieken op vertaalfouten kunnen wijzen. Waar nodig wordt de vertaling aangepast en verbeterd. Verder worden enkel opgaven met een voldoende hoge betrouwbaarheid behouden (Martin et al., 2017; Tielemans et al., 2017).

Op basis van het conceptueel kader, het vooronderzoek en de teksten en trenditems uit voorgaande afnames zijn de **finale teksten en items bepaald**. Van de twaalf nieuw ontwikkelde teksten met bijhorende items zijn er slechts zes teksten opgenomen in de finale PIRLS-toets. Bij het maken van die keuze wordt er rekening gehouden met de relatie tussen de nieuwe teksten en de teksten uit eerdere afnames om zo de algemene coherentie van de afname te garanderen. Daarenboven is er aandacht voor de accuraatheid, helderheid en coherentie van de items en scoringswijzers. Tabel 42 toont een overzicht van het aantal items per leesdoel en -proces dat uiteindelijk werd opgenomen in PIRLS 2016 (Martin et al., 2017).

⁴⁹ In PIRLS 2016 worden zes teksten met bijhorende items overgenomen uit voorgaande PIRLS-cycli. Die teksten en zogenoemde trenditems vormen de basis voor het meten van trends in de tijd. Doordat er bepaalde teksten en items in meerdere cycli afgenomen worden, is het mogelijk om de leerlingprestaties tussen die cycli met elkaar te vergelijken en zo trends doorheen de tijd in kaart te brengen (Martin et al., 2017).

Tabel 42*Overzicht aantal items in PIRLS 2016 per leesdoel en -proces*

	Trend- items	Nieuwe items	Totaal
Leesdoelen	81	94	175
Leeservaring (6 teksten)	44	46	90
Informatieverwerking (6 teksten)	37	48	85
Leesprocessen	81	94	175
Focus leggen op en vinden van expliciet vermelde informatie	18	32	50
Eenvoudige conclusies trekken	28	25	53
Tekstinterpretatie	24	23	47
Tekstevaluatie	11	14	25

Nadat bepaald is welke teksten en items worden opgenomen in PIRLS 2016 worden de **toetsboekjes samengesteld** aan de hand van blokken. Een blok bestaat uit een tekst met bijhorende opgaven. Er zijn dus zes blokken voor leeservaring en zes voor informatieverwerking. In totaal werden er 16 toetsboekjes samengesteld. Met oog op de inhoudelijke validiteit worden de toetsboekjes zo samengesteld dat ze telkens bestaan uit een blok voor leeservaring en een voor informatieverwerking. Dat heeft als gevolg dat elke leerling slechts twee blokken aflegt van de hele PIRLS-toets. Door de overlap van blokken tussen toetsboekjes was het toch mogelijk om de resultaten van de leerlingen op dezelfde meetschaal te plaatsen (Mullis & Martin, 2015; Tielemans et al., 2017). Tabel 43 geeft een overzicht weer van de samenstelling van elk toetsboekje (Mullis & Martin, 2015).

Tabel 43*Samenstelling PIRLS-toetsboekjes*

Toetsboekje	Blok 1	Blok 2
1	PRInf2	PRLit1
2	PRLit3	PRInf2
3	PLInf3	PRLit3
4	PLInf3	PRLit4
5	PRLit4	PRInf1
6	PRLit2	PRInf1
7	PRInf3	PRLit2
8	PLLit3	PRInf3
9	PLLit3	PRInf4
10	PRInf4	PRLit1
11	PRLit3	PRInf1
12	PLInf3	PRLit2
13	PRInf3	PRLit1
14	PLLit3	PRInf2
15	PRInf4	PRLit4
16	PRLit5	PRInf5

Noot. PR = Blok uit PIRLS, PL = Blok uit PIRLS en PIRLS Literacy⁵⁰, Lit = Leeservaring, Inf = Informatieverwerking

⁵⁰ Er is een blok van leeservaring (PLLit3) en een van informatieverwerking (PLInf3) die PIRLS en PIRLS Literacy gemeenschappelijk hebben. PIRLS Literacy is gelijkaardig aan en meet hetzelfde als PIRLS, maar bestaat uit een eenvoudigere toets (Mullis & Martin, 2015).

Ten slotte ontvangen de deelnemende landen de internationale versie van de toets in het Engels. De nationale onderzoekscentra staan in voor de **vertaling van de toets**. Ter ondersteuning voorziet de IEA richtlijnen voor die vertaling. Naast het vertalen van de toets kunnen de onderzoekscentra indien nodig culturele aanpassingen doen die verder gaan dan louter een directe vertaling. Die aanpassingen worden gedocumenteerd door de onderzoekscentra en geëvalueerd door de IEA. Ook de correctheid van de vertaling wordt gecontroleerd. De IEA stelt daarvoor experts aan die de vertaling verifiëren. Het doel van die procedure is om nationale (en regionale) versies van de toets te creëren die rekening houden met de context en taal van dat land zonder dat de internationale vergelijkbaarheid in gedrang komt. Nadat de IEA de nationale versie en de lay-out ervan goedkeurt, worden de toetsen geprint door de nationale onderzoekscentra (Martin et al., 2017).

Naast de toets voor begrijpend lezen worden er ook **achtergrondvragenlijsten voor leerlingen, ouders, leerkrachten en directeuren** ontwikkeld. Eerst wordt het conceptueel kader geüpdatet dat de basis vormt voor achtergrondvragenlijsten (zie 3.1.2 Domein en construct). Dat kader bouwt niet alleen verder op het kader van PIRLS 2011, maar ook op dat van TIMSS 2015. In lijn met het kader voor achtergrondinformatie worden bestaande vragen aangepast en nieuwe vragen toegevoegd. Concreet werd er in 2016 voor gekozen om meer vragen toe te voegen om zo de betrouwbaarheid van de vragenlijsten te verhogen. Daarna volgde het vooronderzoek waarbij de achtergrondvragenlijsten werden afgenomen. Vervolgens werd per schaal de betrouwbaarheid, dimensionaliteit en de correlatie met begrijpend lezen voor elk land geëvalueerd. Op basis van die resultaten werden de vragenlijsten verfijnd tot de uiteindelijke internationale versie. Tot slot konden de nationale onderzoekscentra die versie vertalen en aanpassen tot een nationale versie die daarna werd geëvalueerd door de IEA (Martin et al., 2017).

3.1.5.2 Toetsafname

Er werden verschillende instrumenten afgenomen voor PIRLS tussen maart en mei 2016: een toets begrijpend lezen, een leerlingvragenlijst, een oudervragenlijst, een leerkrachtvragenlijst en een directievragenlijst. Bij de toets begrijpend lezen kregen leerlingen at random een toetsboekje toegewezen. Ze kregen 40 minuten tijd per blok met daartussen een korte pauze. Daarbij werden redelijke aanpassingen toegelaten voor leerlingen met specifieke noden (Tielemans et al., 2017, 2019).

Elke deelnemende klas werd begeleid door een toetsassistent. De taak van de toetsassistent bestond uit (1) het toetsmateriaal uitdelen, (2) instructies geven zoals beschreven in de handleiding, (3) de tijd bewaken en (4) ongebruikelijke zaken tijdens de toetsafname melden. De te volgen procedures werden uitgebreid beschreven in de handleiding voor toetsassistenten (Martin et al., 2017). Met het oog op standaardisatie kregen ook de leerkrachten een handleiding, waarin de toetsafname helder omschreven werd (Tielemans et al., 2019).

Om zowel internationaal als nationaal een kwaliteitsvolle toetsafname te garanderen, bezochten toezichhouders in opdracht van de IEA tijdens de toetsafname 15 scholen in elk land. Ze observeerden en rapporteerden eventuele afwijkingen van de voorgeschreven procedures. De toezichhouders volgden een training en ontvingen een handleiding van de IEA waarin hun rol, verantwoordelijkheden en taken verduidelijkt werden (Martin et al., 2017).

Na de afname worden de antwoorden van de leerlingen op de toets voor begrijpend lezen gescoord. Daarbij wordt niet gekeken naar eventuele taalfouten, aangezien PIRLS begrijpend lezen vooropstelt (Mullis & Martin, 2015). In functie van valide en betrouwbare scoring worden er uitgebreide scoringswijzers voorzien en internationale trainingen over de scoringsprocedure georganiseerd. Bovendien wordt na de scoring van de open vragen de betrouwbaarheid van die scoring in kaart gebracht. In PIRLS worden daarvoor drie indicatoren van betrouwbaarheid onderzocht: de

betrouwbaarheid van de scoring binnen een land, de betrouwbaarheid van de scoring van de trenditems en de betrouwbaarheid van de scoring over landen heen (Martin et al., 2017; Tielemans et al., 2017).

3.2 Standaardbepaling

PIRLS is geen criteriumgerichte meting, aangezien de prestaties voor begrijpend lezen van de leerlingen niet vergeleken worden met een vooraf bepaald 'absoluut' criterium. Wel maakt PIRLS gebruik van internationale benchmarks. Die benchmarks bestaan uit vier specifieke punten op de PIRLS-meetschaal⁵¹ die doorheen de tijd gelijk blijven en functioneren als een referentiepunt om betekenis te geven aan de toetsscores. In PIRLS zijn er vier internationale benchmarks⁵²: het gevorderd niveau (score van minstens 625), het hoog niveau (score van minstens 550), het middelmatig niveau (score van minstens 475) en het laag niveau (score van minstens 400) (Martin et al., 2017; Tielemans et al., 2017).

De inhoudelijke beschrijving die toegekend wordt aan die internationale benchmarks varieert bij elke PIRLS-cyclus (Mullis et al., 2007b, 2012, 2017). Er wordt immers pas een beschrijving toegekend aan elke benchmark nadat de data verzameld zijn. Bij PIRLS 2016 heeft de IEA op basis van de data eerst de items geïdentificeerd die leerlingen hoogstwaarschijnlijk correct beantwoorden als ze scoren op niveau van een specifieke benchmark. Daarna vraagt de IEA aan experts om de inhoud van die items te analyseren om zo te bepalen welke strategieën en vaardigheden leerlingen gebruiken om die items correct te beantwoorden. Die strategieën en vaardigheden vormen op hun beurt de basis voor de inhoudelijke beschrijving van de benchmarks (Martin et al., 2017).

PIRLS is een cross-sectioneel onderzoek met een vijfjaarlijkse iteratie. Om trends doorheen de tijd in kaart te brengen (zie 3.1.1 Doel), maakt PIRLS daarom gebruik van trenditems. Dat zijn items die ook in voorgaande PIRLS-cycli afgenomen werden. Door middel van die trenditems kunnen de data uit PIRLS 2016 op de PIRLS-meetschaal geplaatst worden. Meer specifiek worden de itemparameters voor de trenditems twee keer geschat, een keer op basis van de voorgaande afname en een keer op basis van de huidige afname. Vervolgens worden de plausible values⁵³ van PIRLS 2016 geschat en lineair getransformeerd op de PIRLS-meetschaal. De exacte lineaire transformatie die daarvoor nodig is, bepaalt de IEA op basis van het verschil tussen de geschatte itemparameters voor de voorgaande en de huidige afname. Van zodra de data van PIRLS 2016 op dezelfde meetschaal staan als alle vorige cycli van PIRLS, is het mogelijk om de toetsscores te interpreteren en evoluties in kaart te brengen (Martin et al., 2017).

3.3 Analytische methode

Om de psychometrische validiteit van de toetsen te controleren, worden eerst de itemparameters (moeilijkheidsgraad en discriminatiegraad) in kaart gebracht. Daarbij gaat de IEA ook na of de

⁵¹ De PIRLS-meetschaal werd oorspronkelijk vastgelegd bij PIRLS 2001. Daarbij werd de gemiddelde prestatie van alle deelnemende landen op 500 geplaatst en de standaarddeviatie over alle deelnemende landen op 100. Alle latere PIRLS-cycli werden op die meetschaal geplaatst.

⁵² Bij PIRLS 2001 (en TIMSS 1995 en 1999) legde de IEA de internationale benchmarks vast aan de hand van percentielen. De referentiepunten van de benchmarks kwamen toen overeen met percentiel 90 (gevorderd niveau), 75 (hoog niveau), 50 (middelmatig niveau) en 25 (laag niveau) op respectievelijk de PIRLS- en TIMSS-meetschaal (Martin et al., 2003, 2004). Het gevolg van die methode was echter dat de benchmarks elke cyclus wijzigde. Daarom besloot de IEA in 2004 om voor TIMSS en PIRLS vier vaste punten op de meetschaal te gebruiken voor de internationale benchmarks. De referentiepunten (400, 475, 550 en 625) werden zo gekozen dat ze de percentielmethode benaderde en het interval tussen de referentiepunten gelijk was. Sindsdien zijn de internationale benchmarks hetzelfde voor PIRLS en TIMSS (Martin et al., 2004, 2007).

⁵³ De schatting van de plausible values wordt verderop toegelicht bij het onderdeel 'Analytische methode'.

itemparameters van de trenditems op basis van de afname in 2016 vergelijkbaar zijn met de itemparameters uit voorgaande cycli. Ongebruikelijke statistieken kunnen wijzen op problemen voor een specifiek land. In het geval van een ongewone itemparameter worden de vertaling, eventuele culturele aanpassingen en het toetsboekje onderzocht op mogelijke problemen en kan een item, waar nodig, uit de internationale database verwijderd worden (Martin et al., 2017).

De IEA hanteert verschillende IRT-modellen voor verschillende soorten items. Bij open vragen waarbij het antwoord ofwel juist ofwel fout is, wordt een 2PL-model (2-parameter logistisch model) gehanteerd. Bij de meerkeuzevragen wordt een 3PL-model (3-parameter logistisch model) gebruikt. In dat model wordt er rekening gehouden met de kans dat een leerling met een lage vaardigheid toch het juiste antwoord geeft, omdat hij juist gokt. Tot slot gebruikt de IEA een Partial Credit Model bij open vragen met meer dan twee antwoordmogelijkheden (Martin et al., 2017). Meer gedetailleerde informatie over de IRT-modellen die de basis vormen voor de analyses is te vinden in *Methods and Procedures in PIRLS 2016* (Martin et al., 2017).

3.3.1 Schatting van de plausible values

Aangezien de werkelijke vaardigheid van leerlingen niet gekend is, wordt de vaardigheid van leerlingen geschat aan de hand van de resultaten op de toets begrijpend lezen. Bij PIRLS bestaan de toetsboekjes uit twee blokken, terwijl er in totaal twaalf verschillende blokken opgenomen zijn in PIRLS 2016 (zie 3.1.5.1 Toetsontwikkeling). Doordat leerlingen slechts een deel van de gehele toets afleggen, wordt de onbetrouwbaarheid bij de schatting van de meest waarschijnlijke vaardigheidsscore erg groot. Daarom wordt er geen vaardigheidsscore geschat in PIRLS 2016, maar gebruik gemaakt van plausible values.⁵⁴ Met die methode is het mogelijk om een distributie van de meest waarschijnlijke vaardigheid te schatten voor elke leerling in plaats van één meest waarschijnlijke vaardigheidsscore (Martin et al., 2017).

PIRLS maakt gebruik van conditioneringsvariabelen bij het schatten van de plausible values. Dat betekent dat zowel de antwoorden van de leerlingen op de items als de data uit de achtergrondvragenlijsten gebruikt worden om de plausible values te schatten. Om het aantal conditioneringsvariabelen te reduceren (en niet alle variabelen uit de achtergrondvragenlijsten op te nemen) voert de IEA eerst een principale componentenanalyse uit. Op basis van de informatie uit de achtergrondvragenlijsten werden er in Vlaanderen 259 principale componenten weerhouden die meer dan 90% van de totale variantie in de achtergrondvragenlijsten beschrijven. Die principale componenten worden samen met geslacht opgenomen als conditioneringsvariabelen bij de schatting van de plausible values (Martin et al., 2017).

Concreet wordt er in PIRLS een conditionele distributie van de vaardigheid gemodelleerd, gegeven de antwoorden op de items, de conditioneringsvariabelen en het IRT-model. Daaruit worden vijf plausible values at random getrokken. Die vijf plausible values kunnen beschouwd worden als een betrouwbare schatting van de distributie van de meest waarschijnlijke vaardigheid van een leerling. Na de schatting volgt er een lineaire transformatie (zie 3.2 Standaardbepaling) om de geschatte plausible values op de internationale PIRLS-meetschaal te plaatsen (Martin et al., 2017).

⁵⁴ Ter informatie, in de analyses uit het Vlaamse PIRLS-rapport van 2017 werden er geen plausible values gebruikt. Op het moment van de analyses was de internationale dataset (met plausible values) nog niet beschikbaar. Daarom werd er een vaardigheidsscore geschat voor elke Vlaamse leerling. De vaardigheidsscores hebben een gemiddelde van 50 en een standaarddeviatie van 10. In de analyses voor het eerste Vlaamse PIRLS-rapport werden die vaardigheidsscores gebruikt (Tielemans et al., 2017).

3.3.2 De gebruikte modellen

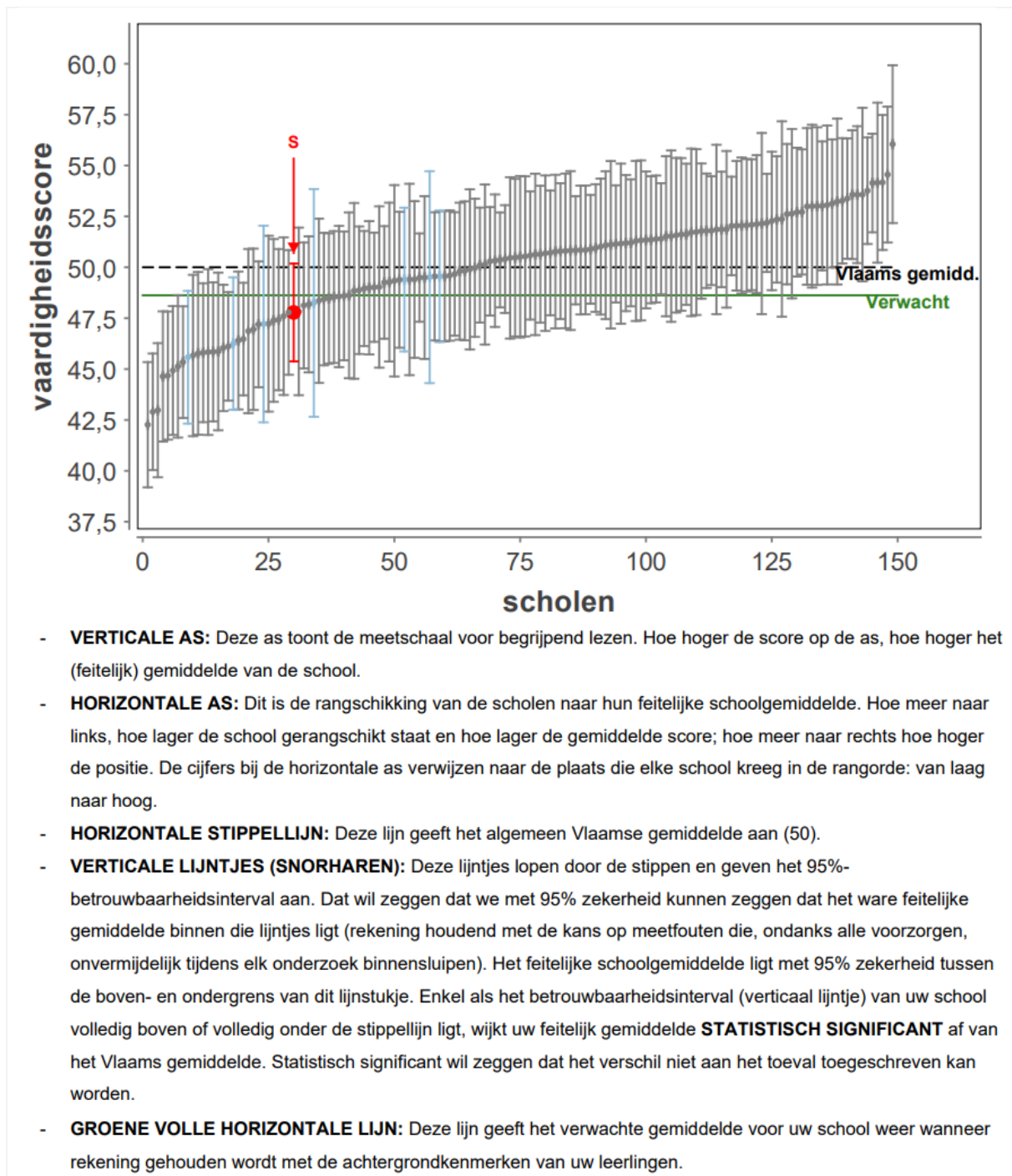
PIRLS brengt de verschillen tussen scholen, tussen klassen (binnen scholen) en tussen leerlingen (binnen klassen) in kaart (zie 3.1.1 Doel). Eerst wordt er een **leeg model** opgesteld, waarin de variantie in de prestaties voor begrijpend lezen verdeeld wordt over het school-, klas- en leerlingniveau. Op basis van dat model kan voor elke school het verschil tussen hun schoolgemiddelde en het algemeen Vlaams gemiddelde (het residu) berekend worden. Door een betrouwbaarheidsinterval op te stellen voor die residuen kan onderzocht worden of scholen significant lager of hoger scoren dan het Vlaams gemiddelde. Bovendien kunnen scholen ook onderling met elkaar vergeleken worden door aangepaste betrouwbaarheidsintervallen te hanteren (Tielemans et al., 2017).

Scholen met elkaar vergelijken op basis van een leeg model is echter niet fair. De resultaten van scholen worden immers gedeeltelijk bepaald door achtergrondkenmerken van de leerlingen. Daarom wordt er in tweede instantie een **nettomodel** gebruikt. In dat model wordt er rekening gehouden met de achtergrondkenmerken van leerlingen, waarover scholen geen controle hebben, maar die wel de prestaties voor begrijpend lezen significant beïnvloeden. Concreet gaat het bij PIRLS om de volgende kenmerken (Tielemans et al., 2017):

- geslacht;
- geboortemaand;
- geboortjaar;
- thuistaal;
- aantal boeken thuis;
- voorschoolse geletterdheid;
- houding van ouders ten opzichte van lezen;
- de mate waarin er thuis activiteiten plaatsvinden die voorschoolse geletterdheid stimuleren;
- opleidingsniveau van de moeder.

Ook bij het nettomodel is het mogelijk om aan de hand van betrouwbaarheidsintervallen aan te geven welke scholen significant afwijken van het Vlaams gemiddelde, rekening houdend met die kenmerken waarop scholen geen invloed hebben. Door middel van aangepaste betrouwbaarheidsintervallen voor de schoolresiduen kunnen scholen ook onderling vergeleken worden (Tielemans et al., 2017).

Alle deelnemende scholen uit PIRLS 2016 krijgen een schoolfeedbackrapport, waarin hun school vergeleken wordt met het Vlaamse gemiddelde en met scholen die een gelijkaardig leerlingenpubliek hebben (zie 3.1.1 Doel). Het **model voor schoolfeedback** maakt gebruik van zowel het leeg model als het nettomodel. Ten eerste wordt op basis van het leegmodel een vergelijking gemaakt waarbij de resultaten op de toets begrijpend lezen voor het vierde leerjaar van een bepaalde school vergeleken worden met de resultaten van de gemiddelde Vlaamse school. Op basis daarvan wordt nagegaan of de school significant verschilt van het Vlaams gemiddelde. Ten tweede wordt er aan de hand van het nettomodel een vergelijking gemaakt. Door rekening te houden met de achtergrondkenmerken die opgenomen zijn in het nettomodel, is het mogelijk om de score van een specifieke school te vergelijken met scholen die een gelijkaardig leerlingenpubliek hebben. Op die manier kunnen scholen op een eerlijkere manier vergeleken worden (Centrum voor Schoolfeedback, 2016; Tielemans et al., 2017). Figuur 35 toont een geanonimiseerd voorbeeld van de schoolfeedback die scholen ontvingen naar aanleiding van PIRLS 2016 (Centrum voor Schoolfeedback, 2016).



Figuur 35. Geanonimiseerd voorbeeld van schoolfeedback over de schoolprestatie in vergelijking met het Vlaams gemiddelde en scholen met een gelijkaardig leerlingenpubliek (verwacht gemiddelde) bij PIRLS 2016 (Figuur overgenomen uit Centrum voor Schoolfeedback, 2016)

3.4 Besluitvorming

De plaats van het Vlaamse onderwijsstelsel in de ranking is eerder een beperkte indicatie van onderwijskwaliteit, aangezien de resultaten van het Vlaams onderwijs moeilijk te vergelijken zijn met de resultaten van andere onderwijsstelsels. Onderwijsstelsels uit verschillende landen zijn immers steeds ingebed in een specifieke cultuur en context en kunnen daardoor niet zomaar met elkaar vergeleken worden. Toch kan een vergelijking met andere landen waardevol zijn. Het kan het onderwijsbeleid inzicht geven in de sterke en zwakke punten van het eigen onderwijsstelsel of

inspireren door na te gaan hoe andere onderwijssystemen omgaan met specifieke uitdagingen. Daarbij moet er specifieke aandacht zijn voor de context waarin de PIRLS-resultaten tot stand zijn gekomen (Tielemans et al., 2019).

PIRLS wordt ook gebruikt om trends doorheen de tijd op te volgen (zie 3.1.1 Doel). Om verschillende PIRLS-afnames met elkaar te vergelijken, doet de IEA beroep op trenditems en een IRT-gebaseerde meetschaal. Ook maakt PIRLS gebruik van een representatieve steekproef, wordt de betrouwbaarheid van de toetsen geëvalueerd en worden er verschillende maatregelen genomen met oog op de inhoudelijke validiteit. Toch kunnen we ons vragen stellen bij het gebruik van PIRLS voor de monitoring van trends. Om trends doorheen de tijd correct in kaart te brengen, moet het construct over cycli heen op dezelfde manier gemeten worden. Dat is bij PIRLS niet altijd het geval, aangezien het conceptueel kader van PIRLS elke cyclus aangepast wordt. Zoals eerder aangegeven, nam Vlaanderen twee keer deel aan PIRLS: in 2006 en 2016. Aangezien het conceptueel kader van PIRLS nauwelijks gewijzigd is tussen 2006 en 2016, kan PIRLS wel valide en betrouwbare uitspraken doen over de evolutie van de Vlaamse prestaties begrijpend lezen doorheen tijd (Mullis et al., 2006; Mullis & Martin, 2015).

Niet alleen het conceptueel kader, maar ook de inhoudelijke beschrijvingen van de benchmarks wijzigen doorheen de tijd. Bij PIRLS wordt er een interpretatie gegeven aan de resultaten door de meetschaal arbitrair op te delen in verschillende internationale benchmarks. Na de dataverzameling wordt er aan elke benchmark een inhoudelijke beschrijving gegeven door experts te raadplegen. De prestaties voor begrijpend lezen van de leerlingen worden dus niet vergeleken met een vooraf bepaald criterium. Pas na de afname wordt er betekenis gegeven aan de verschillende punten op de PIRLS-meetschaal en dus aan de leerlingprestaties voor begrijpend lezen. Bovendien gaat de IEA de adequaatheid van de betekenis die aan de leerlingprestaties gegeven wordt niet expliciet na (Martin et al., 2017).

Naast begrijpend lezen meet PIRLS ook achtergrondkenmerken uit verschillende contexten (zie 3.1.2 Domein en construct). Die kenmerken worden gemeten aan de hand van vragenlijsten voor leerlingen, ouders, leerkrachten en directeurs (Martin et al., 2017; Mullis & Martin, 2015). Belangrijk om op te merken is dat enkel de leerkracht uit het vierde leerjaar bevraagd wordt. Nochtans is de leerkracht uit het vierde leerjaar niet de enige op school die een invloed heeft gehad op de leesvaardigheden van leerlingen uit het vierde jaar. Bovendien weerspiegelen de opvattingen van de bevraagde leerkrachten en directeurs niet noodzakelijk de realiteit op school. De variabelen uit de leerkracht- en directievragenlijst bij PIRLS zijn in dat opzicht in mindere mate geschikt om de concrete klas- en schoolpraktijk in kaart te brengen.

Tot slot beperkt PIRLS zich tot het nagaan van samenhang. Het is daardoor niet mogelijk om de gevonden verbanden te verklaren. Dat geldt ook voor het schoolfeedbackrapport dat de deelnemende scholen ontvangen. De schoolfeedbackrapporten bieden namelijk geen verklaringen of oplossingen aan die de prestaties zouden kunnen verhogen. Daarnaast kent de schoolfeedback nog enkele andere beperkingen. Ten eerste kan de PIRLS-toets ook opgaven bevatten die de leerlingen nog niet gezien hebben. Ten tweede is de informatie uit het rapport enkel gebaseerd op de prestaties van leerlingen uit het vierde leerjaar. De resultaten kunnen daardoor niet veralgemeend worden naar de hele school. Ten derde maakt het schoolfeedbackrapport gebruik van relatieve feedback. Aangezien PIRLS geen criterium vooropstelt om aan te geven wat goed of minder goed is, wordt het resultaat van een school vergeleken met het Vlaams gemiddelde. Dat heeft automatisch als gevolg dat ongeveer de helft van de deelnemende scholen het minder goed doet dan het Vlaams gemiddelde en de andere helft het beter doet. Tot slot is het van belang om enige voorzichtigheid aan de dag te leggen bij de interpretatie van de schoolfeedback, aangezien de foutenmarge bij uitspraken over individuele scholen relatief groot is. Wanneer men zich bewust is van de beperkingen kan de schoolfeedback ingezet worden in

het kader van interne kwaliteitszorg. Scholen kunnen op die manier de informatie uit het rapport combineren met andere informatiebronnen, dat in de juiste context plaatsen en daarover kritisch reflecteren (Centrum voor Schoolfeedback, 2016; Tielemans et al., 2017).

3.5 PIRLS doorheen de tijd

PIRLS werd voor het eerst georganiseerd in 2001 als een vervolg op de Reading Literacy Study uit 1991. Sindsdien wordt PIRLS elke vijf jaar georganiseerd. Doorheen de jaren zijn er enkele fundamentele wijzigingen gebeurd. Zo werden ouders voor het eerst bevraagd in 2001. Sinds 2006 wordt begrijpend lezen geoperationaliseerd in termen van leesdoelen en -processen. Toen werden er vier leesprocessen toegevoegd aan het conceptueel kader. In 2011 werden TIMSS en PIRLS in hetzelfde jaar afgenomen. Naar aanleiding daarvan besloot de IEA om beide onderzoeken verder op elkaar af te stemmen. In 2016 konden landen ook deelnemen aan PIRLS Literacy en ePIRLS.⁵⁵ PIRLS Literacy is gelijkaardig aan PIRLS, maar bestaat uit een eenvoudiger de toets. Die toets meet op een betrouwbaardere manier leesvaardigheden die onderaan de PIRLS-meetschaal gesitueerd zijn. ePIRLS focust dan weer op de meting van online leesvaardigheden (Mullis & Martin, 2015).

4. TIMSS 2019

4.1 Design

4.1.1 Doel

De Trends in International Mathematics and Science Study of kortweg TIMSS wordt elke vier jaar georganiseerd door de IEA. TIMSS is een internationaal vergelijkend onderzoek dat leerlingprestaties voor wiskunde en wetenschappen in kaart brengt. TIMSS bestaat zowel voor het lager als secundair onderwijs, maar Vlaanderen heeft in 2019 enkel deelgenomen met het lager onderwijs. De IEA wil via TIMSS het onderwijsbeleid relevante informatie bieden om het onderwijs voor wiskunde en wetenschappen te verbeteren (Mullis et al., 2020; Mullis & Martin, 2017). TIMSS 2019 werd in Vlaanderen uitgevoerd en gecoördineerd door de onderzoeksgroep Edubron van de Universiteit Antwerpen en het onderzoekscentrum Quest van Arteveldehogeschool (Faddar et al., 2020).

In TIMSS 2019 werden verschillende zaken onderzocht. Ten eerste vergelijkt TIMSS zowel de gemiddelde prestatie van de Vlaamse leerlingenpopulatie als het aandeel Vlaamse leerlingen per prestatieniveau met de prestaties van leerlingen in andere landen. Ten tweede onderzoekt TIMSS de prestatieverschillen tussen leerlingengroepen. Zo worden bijvoorbeeld de verschillen in wiskunde en wetenschappen nagegaan tussen leerlingen die meestal wel Nederlands spreken thuis en leerlingen die meestal een andere taal spreken. Ook wordt de samenhang nagegaan tussen kenmerken uit de thuiscontext, schoolcontext en klascontext enerzijds en de gemiddelde prestatie van de leerlingen uit een bepaald land anderzijds. Ten derde brengt TIMSS in kaart in welke mate leerlingen wiskunde en wetenschappen leuk vinden en hoeveel zelfvertrouwen ze hebben met betrekking tot die twee leergebieden. Tot slot worden trends inzake gemiddelde prestaties voor wetenschappen en wiskunde bij leerlingen uit het vierde leerjaar doorheen de tijd opgevolgd worden aan de hand van TIMSS. Vlaanderen heeft met het lager onderwijs al vier keer deelgenomen aan TIMSS: in 2003, 2011, 2015 en 2019.⁵⁶ Daardoor kunnen tendensen in leerlingprestaties beschreven en vergeleken worden met tendensen in andere landen (Faddar et al., 2020). De resultaten van TIMSS worden in Vlaanderen voornamelijk ingezet om aan **stelselmonitoring** te doen. Bovendien is het mogelijk om bijkomende

⁵⁵ Vlaanderen nam in 2016 enkel deel aan PIRLS en niet aan ePIRLS of PIRLS Literacy (Tielemans et al., 2017).

⁵⁶ Vlaanderen heeft daarnaast met het tweede leerjaar secundair onderwijs deelgenomen aan TIMSS in 1995, 1999 en 2003 (Faddar et al., 2020).

onderzoeksvragen op te nemen – naast de internationale onderzoeksvragen van de IEA – die specifiek gericht zijn op de Vlaamse context.

Doordat onderwijssystemen uit verschillende landen steeds op een bepaalde manier georganiseerd zijn en tegelijkertijd ingebed zijn in een specifieke cultuur en context, kunnen de resultaten van diverse onderwijssystemen niet zomaar met elkaar vergeleken worden. Toch kan een vergelijking met andere landen waardevol zijn. Het kan het onderwijsbeleid nieuwe inzichten of informatie bieden. Cruciaal daarbij is dat de resultaten uit TIMSS steeds geïnterpreteerd worden met oog op de specifieke context waarin die resultaten tot stand kwamen (Faddar et al., 2020).

Net zoals bij PIRLS en het peilingsonderzoek wordt er in Vlaanderen ook voor gekozen om alle scholen die deelnemen aan TIMSS **schoolfeedback** te bezorgen. In het schoolfeedbackrapport worden ten eerste de resultaten van een bepaalde school voor wiskunde en wetenschappen vergeleken met de prestaties van andere scholen in Vlaanderen enerzijds (ruwe scores) en met scholen die een gelijkaardig profiel hebben anderzijds (gecorrigeerde scores). Ten tweede geeft het rapport de spreiding van de leerlingprestaties voor wiskunde en wetenschappen binnen een bepaalde school aan in vergelijking met de spreiding van de Vlaamse leerlingprestaties. Ten derde toont het rapport voor de school in kwestie de verschillen in de prestaties tussen leerlingengroepen in vergelijking met scholen die een gelijkaardig profiel hebben. Ten vierde toont het rapport de motivatie en het vertrouwen m.b.t. wiskunde en wetenschappen bij de leerlingen van de school in kwestie. Ook die resultaten worden vergeleken met het Vlaams gemiddelde. Die schoolfeedback kunnen scholen gebruiken voor zelfevaluatie en -reflectie in het kader van hun interne kwaliteitszorg (Appels et al., 2020). Aangezien de foutenmarge⁵⁷ bij uitspraken over individuele scholen relatief groot is, is het van belang om enige voorzichtigheid aan de dag te leggen bij de interpretatie van de schoolfeedback.

4.1.2 Domein en construct

De IEA beschouwt wiskunde en wetenschappen als essentiële **vaardigheden** in de hedendaagse samenleving (Mullis & Martin, 2017). Zowel wiskunde als wetenschappen worden in TIMSS geoperationaliseerd aan de hand van inhoudsdomeinen en cognitieve domeinen. De inhoudelijke domeinen focussen op de te toetsen onderwerpen, terwijl de cognitieve domeinen aangeven op welke denkprocessen leerlingen beroep moeten doen. De specifieke inhouds- en cognitieve domeinen uit het conceptueel kader en hun omschrijvingen zijn opgenomen in Tabel 44 (Mullis et al., 2020; Mullis & Martin, 2017). Het conceptueel kader vormt de basis voor de ontwikkeling van de toetsen (zie 4.1.5.1 Toetsontwikkeling).

Tabel 44

Het conceptueel kader van TIMSS 2019

Inhoudsdomeinen voor wiskunde
Getallen
<ul style="list-style-type: none">• berekeningen met gehele getallen;• eenvoudige vergelijkingen, uitdrukkingen en relaties tussen grootheden;• bewerkingen met breuken en decimalen.
Geometrische vormen en meten
<ul style="list-style-type: none">• het gebruik van een lat om lengte te meten;• de oppervlakte bij eenvoudige veelhoeken berekenen;

⁵⁷ Scholen met minimum vijf leerlingen in het vierde leerjaar kunnen deelnemen aan TIMSS (Martin et al., 2020). Een beperkt aantal leerlingen per school leidt echter tot relatief grote foutenmarges bij resultaten op schoolniveau.

- het volume bepalen;
- het identificeren van kenmerken bij lijnen, hoeken en een variatie aan twee- en driedimensionale vormen.

Gegevens

- gegevens lezen, interpreteren en weergeven;
- gegevens gebruiken om problemen op te lossen.

Inhoudsdomeinen voor wetenschappen

Biologie

- karakteristieken en levensprocessen van organismen;
- levensloop, erfelijkheid en reproductie;
- organismen in interactie met hun omgeving;
- ecosystemen;
- gezondheid van de mens.

Natuurkunde

- classificatie, eigenschappen en veranderingen van materie;
- energievormen en -transfer;
- krachten en beweging.

Aardrijkskunde

- fysieke kenmerken, bronnen en geschiedenis van de aarde;
- weer en klimaat op aarde;
- de aarde in relatie tot het zonnestelsel.

Cognitieve domeinen voor wiskunde en wetenschappen

Kennen

Feiten, concepten en procedures.

Toepassen

Problemen oplossen door bepaalde kennis en concepten te gebruiken.

Redeneren

Kennis en concepten gebruiken bij onbekende situaties, complexe contexten en meervoudige problemen.

Noot. Elke cyclus wordt het conceptueel kader licht aangepast. Het conceptueel kader van TIMSS 2019 is dus niet per se gelijk aan het conceptueel kader van andere cycli.

Bij de opstelling van het conceptueel kader voor TIMSS 2019 werden de curricula uit de verschillende deelnemende landen als uitgangspunt genomen. Dat wil niet zeggen dat er geen enkel verschil is tussen de toetsen van TIMSS 2019 en de curricula uit de verschillende landen. Integendeel, de toetsen van TIMSS bevatten ook vragen die niet aansluiten bij het Vlaamse curriculum. Om na te gaan in welke mate TIMSS aansluit bij de curricula uit verschillende landen, wordt er een Test-Curriculum Matching Analysis uitgevoerd. Daaruit blijkt dat ongeveer 88% van de vragen uit de toets wiskunde en slechts 41% van de vragen uit de toets wetenschappen aansluiten bij de Vlaamse eindtermen (Mullis et al., 2020).

Daarnaast meet TIMSS ook achtergrondinformatie uit verschillende contexten: (1) de nationale context, (2) de schoolcontext, (3) de klascontext, (4) de thuiscontext en (5) de kenmerken en attitudes

van leerlingen. Tabel 45 toont per context de constructen die TIMSS 2019 meet. Gedetailleerde omschrijvingen van die constructen zijn te vinden in het TIMSS 2019 Context Assessment Framework (Mullis & Martin, 2017). Zoals eerder vermeld bij de bespreking van PIRLS, is het ook hier van belang om aan te geven dat de IEA een bepaalde achtergrondvariabele niet noodzakelijk op dezelfde manier operationaliseert als een onderwijssysteem. De gemeten achtergrondvariabelen worden voornamelijk gebruikt om verschillen tussen leerlingengroepen te duiden en om de samenhang met de leerlingprestaties na te gaan (Faddar et al., 2020).

Tabel 45

Conceptueel kader voor achtergrondkenmerken bij TIMSS 2019

De nationale context
<ul style="list-style-type: none"> • beoogd curriculum; • instructietaal; • kleuteronderwijs, instapleeftijd lager onderwijs, graad van zittenblijven en tracking; • opleiding van de leerkrachten; • nationaal beleid voor certificering van directeurs.
Schoolcontext
<ul style="list-style-type: none"> • kenmerken en demografische gegevens van de school; • invloed van tekort aan middelen op de instructie bij wiskunde en wetenschappen; • nadruk van de school op academisch succes; • de perceptie van ouders over de school; • veilige en ordelijke scholen; • pesten van leerlingen; • schoolse betrokkenheid.
Klascontext
<ul style="list-style-type: none"> • voorbereiding en ervaring bij leerkrachten; • TIMSS onderwerpen die in de les aan bod komen; • instructietijd; • gebruik van specifieke instructiestrategieën bij wiskunde en wetenschappen; • duidelijkheid van instructies; • ondersteunend klasklimaat; • gebruik van technologie bij instructie; • uitdagingen voor leerkrachten.
Thuiscontext
<ul style="list-style-type: none"> • SES (via de schaal: Home Resources for Learning); • thuistaal of -talen; • voorschoolse activiteiten ter stimulering van lees- en rekenvaardigheden; • participatie in het kleuteronderwijs.

Kenmerken en attitudes van leerlingen

- attitude ten opzichte van wiskunde en wetenschappen;
- zelfvertrouwen bij het gebruik van technologie.

Noot. Elke cyclus wordt het conceptueel kader voor achtergrondkenmerken geüpdatet. Bovenstaand conceptueel kader is dus niet per se gelijk aan het conceptueel kader voor achtergrondkenmerken bij andere cycli.

4.1.3 Populatie

TIMSS brengt internationaal de leerlingprestaties voor wiskunde en wetenschappen na zowel vier als acht jaar formele scholing in kaart (ongeacht de leeftijd van de leerlingen). De deelnemende landen kunnen kiezen of ze deelnemen met een of beide populaties (Mullis & Martin, 2017). In Vlaanderen werd ervoor gekozen om TIMSS enkel af te nemen bij leerlingen na vier jaar formele scholing, wat overeenkomt met het vierde leerjaar lager onderwijs. Normaalvorderende leerlingen uit het vierde leerjaar lager onderwijs zijn negen of tien jaar. Echter, leerlingen die versneld of vertraagd zijn, kunnen jonger of ouder zijn op het moment dat ze in het vierde leerjaar zitten (Faddar et al., 2020). In wat volgt, wordt er enkel ingegaan op de TIMSS-afname in het vierde leerjaar lager onderwijs. De TIMSS-afname in het tweede jaar secundair onderwijs wordt niet besproken.

Met het oog op een goede weerspiegeling van het Vlaamse onderwijs worden ook leerlingen uit het buitengewoon lager onderwijs opgenomen in de populatie. Concreet gaat het om leerlingen die de leerstof uit het vierde leerjaar onderwezen krijgen en les volgen in de types basisaanbod, 3 en 9 van het buitengewoon lager onderwijs. Leerlingen die niet de leerstof van het vierde leerjaar volgen of tot andere types behoren, maken geen deel uit van de populatie (Faddar et al., 2020). Internationaal is er beslist om anderstalige nieuwkomers niet op te nemen in de populatie, aangezien het onvoldoende beheersen van de onderwijstaal een te grote belemmering vormt bij de toetsafname (Martin et al., 2020).

4.1.4 Steekproef

De steekproeftrekking gebeurt in elk land op een gelijkaardige manier en daarbij worden dezelfde uitsluitingsnormen gehanteerd. In totaal namen er 58 landen en 8 *benchmarking participants* deel aan TIMSS 2019 (Faddar et al., 2020; Martin et al., 2020). Wereldwijd bestond de steekproef uit ongeveer 330000 leerlingen, 310000 ouders, 30000 leerkrachten en 11000 scholen (Mullis et al., 2020).

De steekproeftrekking van TIMSS gebeurt op een gelijkaardige manier als bij PIRLS en bestaat uit twee stappen (zie 3.1.4 Steekproef). De eerste stap omvat een **steekproeftrekking van scholen** in elk land, weliswaar met andere stratificatievariabelen per land (Martin et al., 2020). In Vlaanderen werd er voor TIMSS een gestratificeerde steekproef van 160 scholen getrokken uit alle Vlaamse scholen voor lager onderwijs met een vierde leerjaar. Daarbij werden er twee vervangscholen geselecteerd voor elk school in de steekproef. Verder werd er bij die steekproeftrekking van scholen meteen rekening gehouden met de **internationale uitsluitingsnormen op schoolniveau**. Concreet werden Franstalige basisscholen en scholen met minder dan vijf leerlingen in het vierde leerjaar vooraf uitgesloten (Faddar et al., 2020; Martin et al., 2020).

Net zoals bij PIRLS wordt er bij de steekproeftrekking in het Vlaamse onderwijs rekening gehouden met drie expliciete stratificatievariabelen, namelijk onderwijsnet, provincie en SES. Het enige verschil met de steekproeftrekking bij PIRLS is de operationalisering van de stratificatievariabele 'SES'. Bij TIMSS worden scholen ingedeeld in drie SES-groepen (bij PIRLS zijn er vier groepen) op basis van het percentage leerlingen dat aantikt op drie indicatoren: (1) laag opleidingsniveau van de moeder, (2) thuistaal niet Nederlands en (3) een schooltoelage ontvangen (Faddar et al., 2020).

Aangezien ook leerlingen uit het buitengewoon onderwijs tot de populatie behoren (zie 4.1.3 Populatie), werden in een apart stratum acht scholen voor buitengewoon lager onderwijs (type basisaanbod, 3 en 9) geselecteerd. Slechts drie scholen, voor buitengewoon onderwijs, waarvan een vervangingschool namen deel aan TIMSS (Faddar et al., 2020). Tabel 46 toont een overzicht van het aantal scholen in de steekproef per stratum (Martin et al., 2020).

Tabel 46

Aantal scholen in de steekproef per stratum

Stratum			Aantal scholen
Antwerpen	Lage SES	Officieel onderwijs	8
Antwerpen	Lage SES	Vrij onderwijs	7
Antwerpen	Middelmatige SES	Officieel onderwijs	8
Antwerpen	Middelmatige SES	Vrij onderwijs	6
Antwerpen	Hoge SES	Officieel onderwijs	8
Antwerpen	Hoge SES	Vrij onderwijs	8
Brussels Hoofdstedelijk Gewest	Lage en middelmatige SES		8
Limburg	Lage SES		7
Limburg	Middelmatige SES		8
Limburg	Hoge SES		7
Oost-Vlaanderen	Lage SES		7
Oost-Vlaanderen	Middelmatige SES		8
Oost-Vlaanderen	Hoge SES		8
Vlaams-Brabant	Lage SES		7
Vlaams-Brabant	Middelmatige SES		8
Vlaams-Brabant	Hoge SES		8
West-Vlaanderen	Lage SES		8
West-Vlaanderen	Middelmatige SES		8
West-Vlaanderen	Hoge SES		7
Scholen voor buitengewoon onderwijs			3
Totaal			147

In een tweede stap volgt de **selectie van klassen** binnen de scholen (Martin et al., 2020). In Vlaanderen werden binnen de geselecteerde scholen willekeurig twee klassen uit het vierde leerjaar geselecteerd. Wanneer er op de geselecteerde school slechts een klas voor het vierde leerjaar was, werd enkel die klas opgenomen in de steekproef (Faddar et al., 2020). Net zoals bij PIRLS werden vervolgens de **internationale uitsluitingsnormen op individueel niveau** toegepast (Martin et al., 2020). In Vlaanderen werden 114 leerlingen uitgesloten, omwille van (1) ernstige fysieke beperkingen, (2) ernstige leer- en leesproblemen en (3) onvoldoende beheersing van de Nederlandse taal (doordat ze minder dan één jaar in Vlaanderen zijn). Om die uitsluitingsnormen te kunnen toepassen, werden de leerkrachten van alle deelnemende klassen bevraagd (Faddar et al., 2020).

De IEA wil de uitsluitingen op school- en individueel niveau tot een minimum beperken en hanteert daarom de norm dat niet meer dan 5% van de doelpopulatie in een onderwijssysteem uitgesloten mag worden. Bij TIMSS 2019 bedroeg het percentage uitsluitingen in Vlaanderen 3%. De Vlaamse steekproef voldoet op die manier aan de vooropgestelde norm van de IEA (Martin et al., 2020). De feitelijke steekproef in Vlaanderen bestond uit 4655 leerlingen (waarvan 36 leerlingen uit het buitengewoon lager onderwijs) uit 147 scholen. De steekproef is voldoende representatief voor de Vlaamse populatie op vlak van onderwijsnet, provincie en SES. Ook wat betreft geslacht, leeftijd, SES en thuistaal wijkt de steekproef niet af van de Vlaamse steekproef voor TIMSS 2015 (Faddar et al., 2020).

TIMSS bevaart ook leerkrachten en directeurs uit de deelnemende scholen. Al beperkt die leerkrachtenbevraging zich enkel tot de leerkracht uit het vierde leerjaar (Faddar et al., 2020). Nochtans is de leerkracht uit het vierde leerjaar niet de enige op school die een invloed heeft gehad op de wiskunde- en wetenschapsvaardigheden van leerlingen uit het vierde leerjaar. Bovendien is het van belang om te beseffen dat de opvattingen van de bevroagde leerkrachten en directeurs niet noodzakelijk de realiteit op school weerspiegelen. We kunnen ons dus vragen stellen bij het gebruik van variabelen uit de leerkracht- en directievragenlijst om uitspraken te doen over de samenhang met de prestaties begrijpend lezen van de leerlingen.

4.1.5 Modus

Niet alleen bij PIRLS, maar ook bij TIMSS behoort de dataverzameling tot de verantwoordelijkheid van de onderzoekscentra die instaan voor de coördinatie in de deelnemende landen. Ter ondersteuning van die onderzoekscentra en met oog op de standaardisatie heeft de IEA de procedure voor dataverzameling stapsgewijs beschreven in de TIMSS 2019 Survey Operations Procedures Units. Na de dataverzameling ontvangen de nationale onderzoekscentra een online vragenlijst⁵⁸ over de voorbereiding en uitvoering van de dataverzameling. Daarin wordt gevraagd om het doorlopen proces te omschrijven en eventuele afwijkingen van de vooropgestelde procedures aan te geven (Martin et al., 2020).

4.1.5.1 Toetsontwikkeling

Eerst wordt het **conceptueel kader van TIMSS geüpdatet** voor de afname in 2019 (zie 4.1.2 Domein en construct). Op basis van dat kader worden de **items en bijhorende scorewijzers ontwikkeld**. Concreet nodigt de IEA de nationale onderzoekscentra uit om samen een hele reeks nieuwe items met bijhorende scorewijzers te ontwikkelen voor zowel wiskunde als wetenschappen. De items omvatten zowel meerkeuzevragen als open vragen. Nadat die items en scorewijzers ontwikkeld zijn, bekijkt en beoordeelt de IEA elk item. Bij die beoordeling wordt er rekening gehouden met verschillende zaken: (1) het conceptueel kader, (2) de richtingspercentages voor items per inhoudsdomein en cognitief domein (zie Tabel 47 voor een overzicht) en (3) de trenditems die opnieuw gebruik worden bij de afname van 2019. Het conceptueel kader vormt dus het uitgangspunt voor de itemontwikkeling en draagt op die manier bij aan de inhoudelijke validiteit van de toetsen wiskunde en wetenschappen (Martin et al., 2020).

⁵⁸ De online vragenlijst gaat in op de volgende thema's: (1) steekproeftrekking van scholen en klassen, (2) toetsinstrumenten vertalen, aanpassen en produceren, (3) toetsafname, (4) implementatie van de kwaliteitscontroles, (5) voorbereiding van de scoring van antwoorden op open vragen en (6) datasets creëren (Martin et al., 2020).

Tabel 47*Overzicht vooropgesteld percentage items per inhoudsdomein en cognitief domein*

Inhoudsdomein wiskunde	%
Getallen	50
Geometrische vormen en meten	30
Gegevens	20
Inhoudsdomein wetenschappen	%
Biologie	45
Natuurkunde	35
Aardrijkskunde	20
Cognitieve domeinen voor wiskunde en wetenschappen	%
Kennen	40
Toepassen	40
Redeneren	20

Daarna wordt er een internationaal **vooronderzoek** georganiseerd om de psychometrische kenmerken van de items en de scoringsprocedures te evalueren. De nieuw ontwikkelde items die positief beoordeeld werden door de IEA zijn uitgetest in een vooronderzoek dat plaatsvond tussen maart en mei 2018. De steekproef daarvoor is op hetzelfde moment en op een gelijkaardige manier getrokken als de steekproef voor het hoofdonderzoek (zie 4.1.4 Steekproef). Op basis van de resultaten van het vooronderzoek bij alle deelnemende landen heeft de IEA de kwaliteit van de ontwikkelde items geanalyseerd. Zowel items met een te lage of te hoge moeilijkheidsgraad als items met een lage discriminatiegraad worden verwijderd. Ook items die voor een bepaald land ongebruikelijke statistieken vertonen, worden nader bekeken aangezien die statistieken op vertaalfouten kunnen wijzen. Waar nodig wordt de vertaling aangepast en verbeterd (Martin et al., 2020).

Op basis van het conceptueel kader, het vooronderzoek en de trenditems uit voorgaande afnames bepaalt de IEA welke **finale items opgenomen worden in TIMSS 2019**. Zo wil de IEA garanderen dat de finale items aansluiten bij de verschillende inhouds- en cognitieve domeinen die TIMSS vooropstelt. Tabel 48 en Tabel 49 tonen een overzicht van het aantal items dat opgenomen is in de TIMSS-toetsen voor respectievelijk wiskunde en wetenschappen (Martin et al., 2020).

Tabel 48*Overzicht aantal items in TIMSS 2019 voor wiskunde per inhoudsdomein en cognitief domein*

	Trend- items	Nieuwe items	Totaal
Inhoudsdomeinen	92	83	175
Getallen	55	29	84
Geometrische vormen en meten	26	27	53
Gegevens	11	27	38
Cognitieve domeinen	92	83	175
Kennen	34	29	63
Toepassen	40	34	74
Redeneren	18	20	38

Tabel 49*Overzicht aantal items in TIMSS 2019 voor wetenschappen per inhoudsdomein en cognitief domein*

	Trend- items	Nieuwe items	Totaal
Inhoudsdomeinen	98	77	175
Biologie	44	34	78
Natuurkunde	36	26	62
Aardrijkskunde	18	17	35
Cognitieve domeinen	98	77	175
Kennen	42	31	73
Toepassen	35	30	65
Redeneren	21	16	37

Nadat bepaald is welke items worden opgenomen in TIMSS 2019 worden de **toetsboekjes samengesteld**. Concreet worden items samengevoegd tot blokken in lijn met de richtingspercentages voor de opdeling in inhouds- en cognitieve domeinen (zie Tabel 47). Bij TIMSS 2019 zijn er 14 blokken voor wiskunde en 14 voor wetenschappen. Elk van die blokken bestaat uit 10 tot 14 items. Met oog op de inhoudelijke validiteit worden er in totaal 14 toetsboekjes samengesteld die telkens bestaan uit twee blokken wiskunde en twee blokken wetenschappen. Elke leerling legt op die manier slechts één toetsboekje af in plaats van de hele TIMSS-toets. Doordat elk blok – en dus ook elk item – is opgenomen in twee verschillende toetsboekjes is het mogelijk om de resultaten van de leerlingen op dezelfde meetschaal te plaatsen. Tabel 50 geeft een overzicht weer van de samenstelling van de toetsboekjes (Mullis & Martin, 2017).

Tabel 50

Samenstelling TIMSS-toetsboekjes

Toetsboekje	Deel 1		Deel 2	
	Blok 1	Blok 2	Blok 1	Blok 2
1	M01	M02	S01	S02
2	S02	S03	M02	M03
3	M03	M04	S03	S04
4	S04	S05	M04	M05
5	M05	M06	S05	S06
6	S06	S07	M06	M07
7	M07	M08	S07	S08
8	S08	S09	M08	M09
9	M09	M10	S09	S10
10	S10	S11	M10	M11
11	M11	M12	S11	S12
12	S12	S13	M12	M13
13	M13	M14	S13	S14
14	S14	S01	M14	M01

Noot. M = Blok wiskunde, S = Blok wetenschappen

Ten slotte ontvangen de deelnemende landen de internationale versie van de toets in het Engels. De nationale onderzoekscentra staan in voor de **vertaling van de toets**. Ter ondersteuning voorziet de IEA richtlijnen voor die vertaling. Naast het vertalen van de toets kunnen de onderzoekscentra indien nodig culturele aanpassingen doen die verder gaan dan louter een directe vertaling. Die aanpassingen worden gedocumenteerd door de onderzoekscentra en geëvalueerd door de IEA. Ook de correctheid van de vertaling wordt gecontroleerd. De IEA stelt daarvoor experts aan die de vertaling verifiëren. Het doel van die procedure is om nationale (en regionale) versies van de toets te creëren die rekening houden met de context en taal van dat land zonder dat de internationale vergelijkbaarheid in gedrang komt. Nadat de IEA de nationale versie en de lay-out ervan goedkeurt, worden de toetsen geprint door de nationale onderzoekscentra (Martin et al., 2020).

Naast de toets voor wiskunde en wetenschappen worden er ook **achtergrondvragenlijsten voor leerlingen, ouders, leerkrachten en directeurs** ontwikkeld. Eerst wordt het conceptueel kader voor achtergrondinformatie geüpdatet door verder te bouwen op het kader van TIMSS 2015. In lijn met het kader voor achtergrondinformatie worden bestaande vragen aangepast en nieuwe vragen toegevoegd. Concreet werd ervoor gekozen om enerzijds meer vragen toe te voegen over de kwaliteit van instructie en het gebruik van technologie bij instructie en evaluatie, en om anderzijds de lengte van de vragenlijst voor leerkrachten te beperken. Daarna volgt het vooronderzoek, waarbij de aangepaste achtergrondvragenlijsten worden afgenomen. Vervolgens wordt per schaal de betrouwbaarheid, unidimensionaliteit en de correlatie met de leerlingprestaties geëvalueerd. Op basis van die resultaten worden de vragenlijsten verfijnd tot de uiteindelijke internationale versie. Tot slot kunnen de nationale onderzoekscentra die versie vertalen en aanpassen tot een nationale versie (Martin et al., 2020).

4.1.5.2 Toetsafname

De verschillende TIMSS-instrumenten werden afgenomen tussen maart en mei 2019: een toets wiskunde en wetenschappen, een leerlingvragenlijst, een oudervragenlijst, een leerkrachtvragenlijst en een directievragenlijst. Bij de toets wiskunde en wetenschappen werd vooraf bepaald welke leerling welk toetsboekje kreeg. Op die manier wou de IEA ervoor zorgen dat elk toetsboekje door proportioneel ongeveer evenveel leerlingen werd ingevuld. Bij de afname kregen de leerlingen 36

minuten voor het eerste deel van de TIMSS-toets, 36 minuten voor het tweede deel en 30 minuten voor de leerlingvragenlijst met daartussen telkens een korte pauze. Daarbij werden redelijke aanpassingen toegelaten voor leerlingen met specifieke noden (Martin et al., 2020; Mullis & Martin, 2017).

Elke deelnemende klas werd begeleid door een toetsassistent. De taak van de toetsassistent bestond uit (1) het toetsmateriaal uitdelen, (2) instructies geven zoals beschreven in de handleiding, (3) de tijd bewaken en (4) ongebruikelijke zaken tijdens de toetsafname melden. Met het oog op standaardisatie werden de te volgen procedures uitgebreid beschreven in de handleiding voor toetsassistenten (Martin et al., 2020).

Om wereldwijd een kwaliteitsvolle toetsafname te garanderen, bezochten toezichthouders in opdracht van de IEA tijdens de toetsafname 15 scholen in elk land. Ze observeerden en rapporteerden eventuele afwijkingen van de voorgeschreven procedures. De toezichthouders volgden een training en ontvingen een handleiding van de IEA waarin hun rol, verantwoordelijkheden en taken verduidelijkt werden (Martin et al., 2020).

Na de afname worden de antwoorden van de leerlingen op de toets voor wiskunde en wetenschappen gescoord. In functie van valide en betrouwbare scoring worden er uitgebreide scoringswijzers voorzien en internationale trainingen over de scoringsprocedure georganiseerd. Bovendien wordt na de scoring de betrouwbaarheid van die scoring in kaart gebracht. Bij TIMSS worden daarvoor drie indicatoren onderzocht: de betrouwbaarheid van de scoring binnen een land, de betrouwbaarheid van de scoring van de trenditems en de betrouwbaarheid van de scoring over landen heen (Martin et al., 2020).

4.2 Standaardbepaling

TIMSS is net zoals PIRLS geen criteriumgerichte meting. De leerlingprestaties voor wiskunde en wetenschappen worden namelijk niet vergeleken met een vooraf bepaald 'absoluut' criterium. Wel maakt TIMSS gebruik van dezelfde internationale benchmarks als PIRLS (zie 3.2 Standaardbepaling): het gevorderd, hoog niveau, middelmatig en laag niveau.⁵⁹ Die benchmarks blijven gelijk doorheen de tijd en functioneren als referentiepunten op de TIMSS-meetschaal⁶⁰ om betekenis te geven aan de toetsscores (Faddar et al., 2020; Martin et al., 2020).

De inhoudelijke beschrijving die gegeven wordt aan die internationale benchmarks wordt bij elke cyclus van TIMSS geüpdatet (Mullis et al., 2020). Die uiteindelijke beschrijving wordt immers pas een toegekend aan een bepaalde benchmark nadat de data verzameld zijn. De procedure om een beschrijving toe te kennen aan de benchmarks verloopt bij TIMSS 2019 op een gelijkaardige manier als bij PIRLS 2016, maar gebeurt apart voor wiskunde en wetenschappen. De IEA identificeert op basis van de data eerst de items die leerlingen hoogstwaarschijnlijk correct beantwoorden als ze scoren op niveau van een specifieke benchmark. Daarna vraagt de IEA aan experts om de inhoud van die items te analyseren om zo te bepalen welke kennis, vaardigheden en cognitieve strategieën leerlingen

⁵⁹ Bij TIMSS 1995 en 1999 (en PIRLS 2001) legde de IEA de internationale benchmarks vast aan de hand van percentielen. De referentiepunten van de benchmarks kwamen toen overeen met percentiel 90 (gevorderd niveau), 75 (hoog niveau), 50 (middelmatig niveau) en 25 (laag niveau) op respectievelijk de PIRLS- en TIMSS-meetschaal (Martin et al., 2003, 2004). Het gevolg van die methode was echter dat de benchmarks elke cyclus wijzigde. Daarom besloot de IEA in 2004 om voor TIMSS en PIRLS vier vaste punten op de meetschaal te gebruiken voor de internationale benchmarks. De referentiepunten (400, 475, 550 en 625) werden zo gekozen dat ze de percentielmethode benaderde en het interval tussen de referentiepunten gelijk was. Sindsdien zijn de internationale benchmarks hetzelfde voor PIRLS en TIMSS (Martin et al., 2004, 2007).

⁶⁰ De TIMSS-meetschaal werd oorspronkelijk vastgelegd in 1995. Daarbij werd de gemiddelde prestatie van alle deelnemende landen op 500 geplaatst en de standaarddeviatie over alle deelnemende landen op 100. Alle latere cycli van TIMSS werden op diezelfde meetschaal geplaatst.

gebruiken om die items correct te beantwoorden. Die kennis, vaardigheden en cognitieve strategieën vormen op hun beurt de basis voor de inhoudelijke beschrijving van de benchmarks (Martin et al., 2020).

TIMSS is net als PIRLS een cross-sectioneel onderzoek. Elke vier jaar worden de TIMSS-toetsen bij andere leerlingen afgenomen. Aan de hand van TIMSS is het echter mogelijk om trends in wiskunde- en wetenschapsprestaties doorheen de tijd in kaart te brengen (zie 4.1.1 Doel). Daarvoor maakt TIMSS gebruik van trenditems. Dat zijn items die ook in voorgaande TIMSS-cycli afgenomen werden. Door middel van die trenditems kunnen de data uit TIMSS 2019 op de TIMSS-meetschaal geplaatst worden. Meer specifiek worden de itemparameters voor de trenditems twee keer geschat, een keer op basis van de voorgaande afname en een keer op basis van de huidige afname. Vervolgens worden de plausible values⁶¹ van TIMSS 2019 geschat en lineair getransformeerd op de TIMSS-meetschaal. De exacte lineaire transformatie die daarvoor nodig is, bepaalt de IEA op basis van het verschil tussen de geschatte itemparameters voor de voorgaande en de huidige afname. Van zodra de data van TIMSS 2019 op dezelfde meetschaal staan als alle vorige cycli van TIMSS, is het mogelijk om de toetsscores te interpreteren en evoluties in kaart te brengen (Martin et al., 2020).

4.3 Analytische methode

De analytische methode bij TIMSS is gelijkaardig aan die van PIRLS. Om de psychometrische validiteit van de toetsen te controleren, worden eerst de itemparameters (moeilijkheidsgraad en discriminatiegraad) in kaart gebracht. Daarbij gaat de IEA ook na of de itemparameters van de trenditems bij TIMSS 2019 vergelijkbaar zijn met de itemparameters uit TIMSS 2015. Ongebruikelijke statistieken kunnen namelijk wijzen op problemen bij een bepaald land. In het geval van een ongewone itemparameter worden de vertaling, eventuele culturele aanpassingen en het toetsboekje onderzocht op mogelijke problemen en kan een item, waar nodig, uit de internationale database verwijderd worden (Martin et al., 2020).

De IEA hanteert verschillende IRT-modellen voor verschillende soorten items. Bij TIMSS worden dezelfde IRT-modellen gehanteerd als bij PIRLS: het 2PL-model, het 3PL-model en het Partial Credit Model (Martin et al., 2020). Meer gedetailleerde informatie over de IRT-modellen die de basis vormen voor de analyses is te vinden in *Methods and Procedures: TIMSS 2019 Technical Report* (Martin et al., 2020).

4.3.1 Schatting van de plausible values

TIMSS werkt net zoals PIRLS met plausible values. Met die methode is het mogelijk om een distributie van de meest waarschijnlijke vaardigheid te schatten voor elke leerling in plaats van één meest waarschijnlijke vaardigheidsscore. De schatting van die plausible values verloopt bij TIMSS op een gelijkaardige manier als beschreven in de bespreking van PIRLS (zie 3.3.1 Schatting van de plausible values). Enkel de conditioneringsvariabelen zijn verschillend in vergelijking met PIRLS. Bij TIMSS werden er op basis van de achtergrondvragenlijst 232 principale componenten weerhouden die meer dan 90% van de totale variantie in de achtergrondvragenlijsten beschrijven. Die principale componenten worden samen met geslacht en de taal van de TIMSS-toets opgenomen als conditioneringsvariabelen bij de schatting van de plausible values (Martin et al., 2020). Ook bij TIMSS volgt er na de schatting van de plausible values een lineaire transformatie (zie 4.2 Standaardbepaling) om de geschatte plausible values op de internationale TIMSS-meetschaal te plaatsen (Martin et al., 2020).

⁶¹ De schatting van de plausible values wordt verderop toegelicht bij het onderdeel 'Analytische methode'.

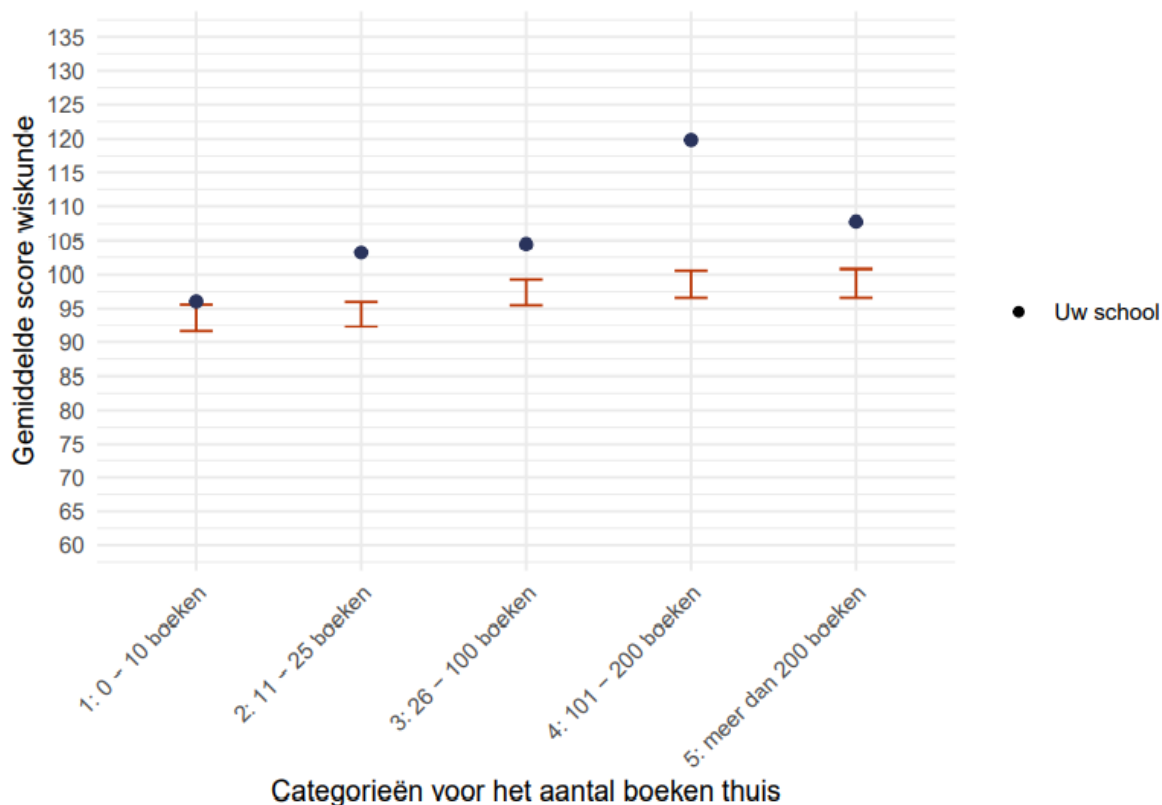
4.3.2 De gebruikte modellen

TIMSS stelt verschillende onderzoeksvragen voorop (zie 4.1.1 Doel), maar beschrijft in de rapporten niet welke achterliggende modellen juist gebruikt worden om die onderzoeksvragen te beantwoorden. Wel maakt TIMSS gebruik van een **model voor schoolfeedback** bij het opstellen van de schoolfeedbackrapporten. De deelnemende scholen vergelijken met het Vlaams gemiddelde leidt niet altijd tot een faire vergelijking. De schoolresultaten worden immers gedeeltelijk bepaald door achtergrondkenmerken van de leerlingen. Daarom wordt er bij de analyses voor de schoolfeedback rekening gehouden met de achtergrondkenmerken van leerlingen, waarover scholen geen controle hebben, maar die wel een invloed hebben op de prestaties. Concreet gaat het om de volgende kenmerken (Appels et al., 2020):

- geslacht;
- geboortemaand;
- geboortjaar;
- thuistaal;
- aantal boeken thuis;
- opleidingsniveau van de ouders;
- voorschoolse wiskundevaardigheden.

Door rekening te houden met bovenstaande kenmerken, is het mogelijk om de score van een specifieke school te vergelijken met scholen die een gelijkaardig profiel hebben. Op die manier kunnen scholen op een eerlijkere manier vergeleken worden. Figuur 36 toont een geanonimiseerd voorbeeld van de schoolfeedback die scholen ontvingen bij TIMSS 2019 (Appels et al., 2020).

Kijkend naar de eerste categorie in de grafiek, ziet u de scores voor de leerlingen die aangaven thuis tussen 0 en 10 boeken te bezitten. De stip is de gemiddelde wiskundescore van deze leerlingen in uw school. Het 'I'-symbool geeft aan hoe de gemiddelde wiskundescore overheen gelijkaardige scholen in Vlaanderen wordt geschat. Als de stip, die de leerlingen in deze categorie voor uw school weergeeft, zich boven het 'I'-symbool bevindt, dan betekent dit dat deze leerlingen significant hoger scoren dan de leerlingen in deze categorie in gelijkaardige Vlaamse scholen. Als de stip zich binnen de randen van het 'I'-symbool bevindt, dan is er geen noemenswaardig verschil in de score van de leerlingen binnen deze categorie voor uw school en die in andere gelijkaardige Vlaamse scholen. Als de stip zich volledig onder het 'I'-symbool bevindt, dan scoren de leerlingen met 0 tot 10 boeken thuis in uw school significant lager dan leerlingen met dit aantal boeken thuis in andere gelijkaardige Vlaamse scholen.



Figuur 36. Geanonimiseerd voorbeeld van schoolfeedback over de gemiddelde prestatie naar aantal boeken thuis voor uw school en Vlaanderen bij TIMSS 2019 (Figuur overgenomen uit Appels et al., 2020)

4.4 Besluitvorming

De plaats van Vlaanderen in de TIMSS-ranking is slechts een beperkte indicatie van onderwijskwaliteit, aangezien de resultaten van onderwijssystemen onderling moeilijk te vergelijken zijn omwille van hun specifieke context. Toch kan een vergelijking met andere landen waardevol zijn. Het kan het Vlaamse onderwijs nieuwe inzichten of informatie bieden. Daarbij moet er echter specifieke aandacht zijn voor de context waarin de TIMSS-resultaten tot stand zijn gekomen (Faddar et al., 2020).

Aan de hand van TIMSS worden trends doorheen de tijd opgevolgd (zie 4.1.1 Doel). Om verschillende TIMSS-cycli met elkaar te vergelijken, maakt de IEA gebruik van trenditems en een IRT-gebaseerde meetschaal. Verder maakt TIMSS gebruik van een representatieve steekproef, wordt de betrouwbaarheid van de toetsen geëvalueerd en worden er verschillende maatregelen genomen met oog op de inhoudelijke validiteit.

Toch kunnen we ons vragen stellen bij het gebruik van TIMSS voor de monitoring van trends. Ten eerste moet het construct over cycli heen op dezelfde manier gemeten worden om trends doorheen de tijd correct in kaart te brengen. Dat is bij TIMSS niet altijd het geval, aangezien het conceptueel kader bij elke cyclus aangepast wordt (Mullis & Martin, 2017). Ten tweede wijzigen ook de inhoudelijke beschrijvingen van de benchmarks doorheen de tijd. Na de TIMSS-dataverzameling wordt er immers aan elke benchmark een inhoudelijke beschrijving gegeven door experts te raadplegen. De prestaties van de leerlingen worden niet vergeleken met een vooraf bepaald criterium. Pas na de afname wordt er betekenis gegeven aan de verschillende punten op de TIMSS-meetschaal en dus aan de leerlingprestaties voor wiskunde en wetenschappen. Bovendien gaat de IEA de adequaatheid van de betekenis die aan de leerlingprestaties gegeven wordt niet expliciet na (Martin et al., 2020).

Daarnaast meet TIMSS achtergrondkenmerken uit verschillende contexten (zie 3.1.2 Domein en construct). Die kenmerken worden gemeten aan de hand van vragenlijsten voor leerlingen, ouders, leerkrachten en directeurs (Martin et al., 2020; Mullis & Martin, 2017). Belangrijk om op te merken is dat enkel de leerkracht uit het vierde leerjaar bevroegd wordt. Nochtans is de leerkracht uit het vierde leerjaar niet de enige op school die een invloed heeft gehad op de prestaties van leerlingen uit het vierde jaar. Bovendien weerspiegelen de opvattingen van de bevroegde leerkrachten en directeurs niet noodzakelijk de realiteit op school. De variabelen uit de leerkracht- en directievragenlijst bij TIMSS zijn in dat opzicht in mindere mate geschikt om de concrete klas- en schoolpraktijk in kaart te brengen.

Tot slot beperkt TIMSS zich tot het nagaan van samenhang. Het is daardoor niet mogelijk om de gevonden verbanden te verklaren (Faddar et al., 2020). Dat geldt ook voor het schoolfeedbackrapport dat de deelnemende scholen ontvangen (Appels et al., 2020). Daarnaast kent de schoolfeedback ook beperkingen. Ten eerste is de informatie uit het rapport enkel gebaseerd op de prestaties van leerlingen uit het vierde leerjaar, waardoor de resultaten niet veralgemeend kunnen worden naar de hele school. Ten tweede kan de toets opgaven bevatten die de leerlingen nog niet gezien hebben. Ten derde gaat het om relatieve schoolfeedback. Er is geen vooropgesteld criterium om aan te geven wat goed of minder goed is. Het resultaat van een specifieke school wordt daarom vergeleken met het Vlaams gemiddelde. Dat heeft automatisch als gevolg dat ongeveer de helft van de deelnemende scholen het minder goed doet dan het Vlaams gemiddelde en de andere helft het beter doet. Tot slot is het van belang om voorzichtig te zijn bij de interpretatie van de schoolfeedback. Zoals eerder aangeven, is de foutenmarge bij uitspraken over individuele scholen relatief groot. Wanneer scholen zich bewust zijn van de beperkingen, kan het schoolfeedbackrapport ingezet worden met het oog op interne kwaliteitszorg. Scholen kunnen de informatie uit het rapport combineren met andere informatiebronnen, het in de juiste context plaatsen en er kritisch over reflecteren.

4.5 TIMSS doorheen de tijd

TIMSS werd voor het eerst georganiseerd in 1995 als een vervolg op eerdere studies van de IEA over wiskunde en wetenschappen. Sindsdien wordt TIMSS elke vier jaar georganiseerd. Doorheen de jaren zijn er enkele wijzigingen gebeurd. Sinds 2007 wordt de TIMSS Encyclopedia bijvoorbeeld opgesteld. Die encyclopedie beschrijft per cyclus voor elk deelnemend land het beleid en curriculum voor wiskunde en wetenschappen (Mullis et al., 2016). In 2011 werden TIMSS en PIRLS in hetzelfde jaar afgenomen. Naar aanleiding daarvan besloot de IEA om beide onderzoeken verder op elkaar af te stemmen. Bij TIMSS 2015 konden landen ook kiezen voor TIMSS Numeracy. Dat is een gelijkaardige, maar eenvoudiger toets dan TIMSS. TIMSS Numeracy meet zo op een betrouwbaardere manier de vaardigheden voor wiskunde die onderaan de TIMSS-meetschaal gesitueerd zijn. In 2019 konden landen kiezen tussen deelnemen aan de papieren versie van TIMSS of aan eTIMSS, een digitale afname van TIMSS. Vlaanderen koos in 2015 voor de klassieke TIMSS-afname en in 2019 voor de papieren versie (Faddar et al., 2020; Mullis et al., 2020; Mullis & Martin, 2017).

5. PISA 2018

5.1 Design

5.1.1 Doel

De Programme for International Student Assessment (PISA) wordt elke drie jaar georganiseerd door de Organisatie voor Economische Samenwerking en Ontwikkeling (OESO of in het Engels: OECD). De OESO wil aan de hand van PISA inzichten verwerven die het onderwijsbeleid en onderwijssystemen kunnen verbeteren (OECD, n.d.). PISA is een internationaal vergelijkend onderzoek dat nagaat in welke mate 15-jarige leerlingen wereldwijd de essentiële vaardigheden bezitten om toekomstige, maatschappelijke uitdagingen aan te kunnen. Daarbij gaat PISA niet alleen na of leerlingen kennis kunnen reproduceren, maar ook of leerlingen wat ze geleerd hebben kunnen toepassen in onbekende situaties. Bij elke cyclus ligt het accent op een van de volgende drie domeinen: leesvaardigheid, wiskundige geletterdheid of wetenschappelijke geletterdheid. In 2018 lag de focus op leesvaardigheid.⁶² PISA 2018 werd in Vlaanderen uitgevoerd en gecoördineerd door vakgroep Onderwijskunde van de Universiteit Gent (De Meyer et al., 2019; OECD, 2019b, 2019a).

De OESO wil de deelnemende landen waardevolle informatie bieden die ingezet kan worden om geïnformeerde beleidsbeslissingen te nemen enerzijds en aan **stelselmonitoring** te doen anderzijds (De Meyer et al., 2019; OECD, 2019b). Concreet brengt PISA, ten eerste, de kennis en vaardigheden van 15-jarige leerlingen in kaart. Die resultaten in Vlaanderen worden ook met de resultaten van andere landen vergeleken. Ten tweede gaat PISA de verschillen na tussen leerlingengroepen met verschillende demografische, economische, sociale en onderwijsgebonden achtergrondkenmerken. Tot slot kunnen trends doorheen de tijd opgevolgd worden aan de hand van PISA. Vlaanderen neemt sinds 2000 elke drie jaar deel aan het PISA-onderzoek (De Meyer et al., 2019).

Bij de vergelijking van vaardigheden van leerlingen tussen onderwijssystemen is het cruciaal om rekening te houden met verschillen in context. Onderwijs is namelijk ingebed in een bepaalde sociale en economische context. Met als gevolg, wanneer een onderwijssysteem een significant hoger gemiddelde heeft dan een ander wil dat niet zeggen dat de scholen van het ene onderwijssysteem automatisch effectiever zijn dan de scholen in het andere. Bij die rankings wordt er immers geen rekening gehouden met de specifieke context of de vooropgestelde onderwijsdoelen die een land nastreeft (OECD, 2019b).

In Vlaanderen wordt ervoor gekozen om alle deelnemende scholen **schoolfeedback** te bezorgen. In dat schoolfeedbackrapport worden de PISA-resultaten van de geselecteerde leerlingen in een bepaalde school vergeleken met de resultaten van geselecteerde leerlingen uit andere scholen die deelnamen. Ook de gemiddelde resultaten van de school in kwestie enerzijds en het Vlaams gemiddelde anderzijds worden per studierichting en per onderwijsvorm gepresenteerd. Verder toont het rapport de samenhang tussen de gemiddelde school SES en de gemiddelde schoolprestatie. Tot slot wordt er ook feedback gegeven over de scores van de leerlingen op de schaal 'plezier in lezen' (Vakgroep Onderwijskunde, 2020).

Daarenboven bespreekt het rapport ook enkele beperkingen van de weergegeven resultaten. Zo zeggen de resultaten uit het feedbackrapport niets over de prestatie van alle leerlingen uit de school, aangezien PISA enkel 15-jarigen bevroegt. Verder wordt er aangegeven dat de PISA-resultaten het niet toelaten om scholen onderling te vergelijken of te rangschikken. Met als gevolg, het is op basis van

⁶² Leesvaardigheid vormde het hoofddomein bij PISA in 2000, 2009 en 2018. Wiskundige geletterdheid stond centraal bij PISA 2003 en PISA 2012, terwijl in 2006 en 2015 het accent lag op wetenschappelijke geletterdheid (De Meyer et al., 2019).

PISA *niet* mogelijk om te stellen dat een school met een hoger PISA-gemiddelde een betere school is dan een school met een lager PISA-gemiddelde. Kortom, de informatie in het schoolfeedbackrapport heeft zijn beperkingen en moet daarom met de nodige voorzichtigheid geïnterpreteerd worden (Vakgroep Onderwijskunde, 2020).

5.1.2 Domein en construct

De drie kerndomeinen, namelijk leesvaardigheid (Engels: reading proficiency), wiskundige geletterdheid (Engels: mathematics proficiency) en wetenschappelijke geletterdheid (Engels: science proficiency) vormen de centrale constructen. Die domeinen worden door de OESO als volgt gedefinieerd (De Meyer et al., 2019):

“Leesvaardigheid is het begrijpen, gebruiken en evalueren van teksten, het reflecteren over teksten en het zich ermee inlaten om een doel te bereiken, om kennis en mogelijkheden te ontwikkelen en om deel te nemen aan de maatschappij” (De Meyer et al., 2019, p. 14).

“Wiskundige geletterdheid is het vermogen van een individu om wiskunde in verschillende contexten te gebruiken, te formuleren en te interpreteren. Dit omvat wiskundig redeneren en het gebruik van wiskundige begrippen, werkwijzen, feiten en hulpmiddelen om fenomenen te beschrijven, te verklaren en te voorspellen” (De Meyer et al., 2019, p. 14).

“Wetenschappelijke geletterdheid is het beheersen van vaardigheden om als kritische burger om te gaan met wetenschappelijke onderwerpen en ideeën. Dit omvat enerzijds de bereidheid om een zinvolle discussie aan te gaan over wetenschap en technologie en anderzijds de vaardigheden om fenomenen wetenschappelijk te verklaren, om wetenschappelijk onderzoek te evalueren en op te zetten en om gegevens en bewijzen wetenschappelijk te interpreteren” (De Meyer et al., 2019, p. 14).

De OESO legt de nadruk bij de definiëring van de kerndomeinen op **functionele kennis en vaardigheden**. Met andere woorden, PISA onderzoekt in welke mate leerlingen de kennis die ze geleerd hebben op school kunnen toepassen om problemen uit het dagelijks leven op te lossen (De Meyer et al., 2019).

Aansluitend bij bovenstaande definities worden er voor PISA conceptuele kaders opgesteld. Bij PISA wordt enkel het conceptueel kader van het hoofddomein herwerkt.⁶³ Bij de afname in 2018 wordt dus het conceptueel kader van leesvaardigheid grondig herwerkt. Een groep van internationale experts herwerkt het kader voor leesvaardigheid, waarna het wordt voorgelegd aan de deelnemende landen. In 2018 vormen wiskundige en wetenschappelijke geletterdheid geen hoofddomeinen. Dat heeft als gevolg dat de conceptuele kaders voor die domeinen overgenomen worden uit de voorgaande afname (OECD, 2019b). In wat volgt, worden de overgenomen kaders voor wiskundige en wetenschappelijke geletterdheid enerzijds en het aangepaste conceptueel kader voor leesvaardigheid anderzijds besproken.

5.1.2.1 Leesvaardigheid

De OESO operationaliseert leesvaardigheid aan de hand van drie dimensies: teksten, cognitieve processen en scenario's. De dimensie 'teksten' verwijst naar het soort tekst dat de leerlingen lezen. De processen geven aan welke cognitieve benaderingen de leerlingen gebruiken wanneer ze met een tekst omgaan, terwijl de scenario's wijzen op het doel of de context voor het lezen van de tekst. PISA

⁶³ Bij de domeinen die niet het hoofddomein van de afname vormen – in 2018 wiskundige en wetenschappelijke geletterdheid – worden er enkel items uit de voorgaande afnames overgenomen. Daardoor verandert het conceptueel kader niet ten opzichte van de voorgaande afname (OECD, 2019b, 2021).

rapporteert niet over de resultaten op basis van de scenario's. Wel worden er in de PISA-test vragen opgenomen die behoren tot verschillende scenario's om zo de brede definitie van leesvaardigheid te toetsen.

Aangezien leesvaardigheid het hoofddomein vormt bij PISA 2018, is de definitie van leesvaardigheid voor deze cyclus geüpdatet. Met oog op die gewijzigde definitie en de uitbreiding naar digitale leesvaardigheden is het conceptueel kader grondig herbekeken en aangepast. Zo wordt er bij PISA 2018 een ruimer aanbod aan digitale teksten opgenomen in het kader van de subdimensie 'indeling en mate van navigatie' bij teksten. Daarnaast is bijvoorbeeld de subdimensie 'medium' bij teksten geschrapt en is 'vlot lezen' als cognitieve proces toegevoegd aan het conceptueel kader. Tabel 51 toont het aangepaste conceptueel kader voor leesvaardigheid met de daarin opgenomen dimensies (De Meyer et al., 2019; OECD, 2019a, 2019b).

Tabel 51

Het conceptueel kader voor leesvaardigheid bij PISA 2018

Teksten	
Bron	De hoeveelheid bronnen van de tekst.
Indeling en mate van navigatie	Er wordt een onderscheid gemaakt tussen statistisch en dynamische teksten. Statische teksten hebben een lineaire structuur, terwijl dynamische teksten meer complexe navigatiemogelijkheden en een ingewikkeldere structuur hebben.
Tekstformaat	De manier waarop de tekst wordt aangeboden: doorlopende tekst, niet doorlopende tekst of een combinatie van beiden.
Teksttype	De reden voor het schrijven van de tekst en de structuur van de tekst.
Processen	
Vlot lezen	De vaardigheid van leerlingen om de woorden automatisch en nauwkeurig te lezen en te verwerken om de betekenis van de woorden en de tekst te begrijpen. Dit proces vormt de basis voor de andere processen uit het conceptueel kader.
Informatie lokaliseren	<ul style="list-style-type: none"> • relevante teksten zoeken en selecteren; • scannen en lokaliseren van relevante informatie.
Begrijpen	<ul style="list-style-type: none"> • letterlijke betekenis van teksten beschrijven; • integreren van informatie en conclusies formuleren.
Evalueren en reflecteren	<ul style="list-style-type: none"> • kwaliteit en geloofwaardigheid beoordelen; • reflecteren over inhoud en vorm van de tekst; • tegenstrijdigheden detecteren in de tekst en ermee omgaan.

Scenario's

Persoonlijk

Omwille van eigen praktische of intellectuele interesses.

Publiek

Omwille van noden of activiteiten van de samenleving.

Educatief

Voor onderwijs of lessen ontwikkeld.

Beroepsmatig

Gekoppeld aan de beroepswereld en de uitvoering van specifieke taken.

Bij de opstelling en aanpassing van de conceptueel kaders houdt PISA weinig rekening met de curricula van de deelnemende landen. Het doel van PISA is immers om na te gaan in welke mate de 15-jarige leerlingen beschikken over de essentiële vaardigheden om toekomstige, maatschappelijke uitdagingen aan te kunnen (zie 5.1.1 Doel). Wel vraagt PISA aan nationale experts in welke mate de PISA-toetsen overeenkomen met de nationale curricula om op die manier inzicht te krijgen in de (minder) sterke punten van een specifiek onderwijssysteem. Uit die inhoudelijke bevraging blijkt dat er in 2018 geen enkel item voor leesvaardigheid was dat niet aansloot bij de Vlaamse eindtermen (OECD, 2019b).

5.1.2.2 Wiskundige geletterdheid

Bij PISA 2018 is wiskundige geletterdheid geen hoofddomein. Dat wil zeggen dat het conceptueel kader en de definitie voor wiskundige geletterdheid overgenomen worden uit de voorgaande cycli. In 2003 en 2012 vormde wiskundige geletterdheid wel het hoofddomein. Met als gevolg, het conceptueel kader voor wiskundige geletterdheid is voor het laatst aangepast bij PISA 2012 (De Meyer et al., 2019). Sindsdien wordt wiskundige geletterdheid geoperationaliseerd aan de hand van drie dimensies: wiskundige inhouden, processen en contexten. De dimensie 'wiskundige inhouden' verwijst naar het inhoudelijke domein waartoe de aangereikte problemen in de PISA-test behoren. De processen geven aan hoe de leerlingen met de aangereikte inhoud moeten omgaan, terwijl de laatste dimensie weergeeft binnen welke context het probleem uit de test zich situeert. Bij de PISA-test zijn de items gelijk verdeeld over de verschillende contexten uit het conceptueel kader. Tabel 52 toont het conceptueel kader en de bijhorende omschrijvingen voor wiskundige geletterdheid (De Meyer et al., 2013; OECD, 2019a, 2019b).

Tabel 52

Het conceptueel kader voor wiskundige geletterdheid bij PISA 2012, PISA 2015 en PISA 2018

Wiskundige inhouden

Relaties en veranderingen

- relaties en veranderingen modelleren aan de hand van functies;
- symbolische en grafische relaties creëren, vergelijken, interpreteren en vertalen.

Vorm en ruimte

- eigenschappen van objecten;
- patronen;
- oriëntaties en posities;
- visuele informatie coderen;
- navigeren;
- ...

Hoeveelheid

- kwalitatieve relaties;
- begrip van grootte, indicatoren, eenheden, patronen, tellen, meten enzoverder.

Onzekerheid en data

- de identificatie en samenvatting van een boodschap uit data die op diverse manieren gepresenteerd wordt;
- variatie in processen herkennen, een idee hebben van de kwantificatie van die variatie, onzekerheid en meetfouten.

Processen

Formuleren

Situaties op wiskundige wijze formuleren.

Gebruiken

Wiskundige feiten, concepten, procedures en redeneringen hanteren.

Interpreteren

De toepassing, interpretatie en evaluatie van wiskundige uitkomsten.

Context

Persoonlijk

Focus op activiteiten in de nabije omgeving van de leerling (familie, leeftijdsgenoten of de leerling zelf).

Beroepsgerelateerd

Gekoppeld aan de beroepswereld en de uitvoering van specifieke taken.

Maatschappelijk

De lokale, nationale of globale context.

Wetenschappelijk

Specifieke verwijzingen naar wiskundige toepassingen uit de wereld van technologie en wetenschap.

Noot. Het conceptueel kader voor wiskundige geletterdheid bij PISA 2012 is gelijk aan het kader bij de PISA-cycli van 2015 en 2018.

5.1.2.3 Wetenschappelijke geletterdheid

Bij PISA 2015 was wetenschappelijke geletterdheid het hoofddomein. Doordat PISA bij de afname in 2015 overgestapt is naar een digitale afname via de computer, werd het conceptueel kader voor wetenschappelijke geletterdheid toen grondig herwerkt met oog op die digitale afname. Bij PISA 2018 wordt het conceptueel kader en de definitie voor wetenschappelijke geletterdheid overgenomen uit de afname in 2015 (De Meyer et al., 2019). In 2015 en 2018 wordt wetenschappelijke geletterdheid geoperationaliseerd aan de hand van vier dimensies: wetenschappelijke inhouden, wetenschappelijke vaardigheden, types kennis en contexten. De dimensie 'wetenschappelijke inhouden' verwijst naar het inhoudelijke domein waarin de aangereikte problemen gesitueerd worden. De wetenschappelijke vaardigheden geven aan hoe de leerlingen met de aangereikte inhoud moeten omgaan. De dimensie 'types kennis' verwijst naar het soort kennis die leerlingen nodig hebben om de vraag op te lossen. De laatste dimensie geeft aan binnen welke context het probleem uit de test zich situeert. Bij de ontwikkeling van de items vormt het conceptueel kader het uitgangspunt. Tabel 53 toont het

conceptueel kader en de bijhorende omschrijvingen voor wetenschappelijke geletterdheid (Blondeel et al., 2016; OECD, 2019a, 2019b).

Tabel 53

Het conceptueel kader voor wetenschappelijke geletterdheid bij PISA 2015 en PISA 2018

Wetenschappelijke inhoud
Fysische systemen
Onderwerpen die sterk aansluiten bij het vak fysica.
Levende systemen
Onderwerpen die sterk aansluiten bij het vak biologie.
Aarde en het heelal
Onderwerpen die sterk aansluiten bij het vak aardrijkskunde.
Competenties
Fenomenen wetenschappelijk verklaren
Verklaringen voor verschillende technologische en natuurlijke fenomenen herkennen, aangeven en evalueren.
Evalueren en opzetten van wetenschappelijk onderzoek
<ul style="list-style-type: none">wetenschappelijk onderzoek beschrijven en beoordelen;voorstel doen om een vraag op een wetenschappelijke manier te onderzoeken.
Data en bewijzen wetenschappelijk interpreteren
<ul style="list-style-type: none">de analyse en evaluatie van data, argumenten en bewerkingen;correcte wetenschappelijke conclusies trekken.
Kennis
Inhoudelijke kennis
Kennis van verklaringen, theorieën en feiten.
Procedurele kennis
Inzicht in de ontwikkeling van bepaalde kennis.
Epistemische kennis
Inzicht in de grondslag van bepaalde kennis.
Context
Persoonlijk
Focus op de nabije omgeving van de leerling (familie, leeftijdsgenoten of de leerling zelf).
Lokaal/nationaal
De gemeenschap waarin de leerling woont.
Globaal
Gerelateerd aan het leven op aarde.

Noot. Het conceptueel kader voor wetenschappelijke geletterdheid bij PISA 2015 is gelijk aan het kader bij PISA 2018.

5.1.2.4 Achtergrondkenmerken

PISA neemt ook verschillende achtergrondvragenlijsten af bij de leerlingen, ouders en directies. Op basis van de achtergrondvragenlijsten meet PISA (1) achtergrondkenmerken van de leerlingen, (2) schoolkenmerken en (3) non-cognitieve en metacognitieve kenmerken van de leerlingen. Aangezien leesvaardigheid het hoofddomein vormt bij de afname in 2018, worden er ook verschillende kenmerken gemeten die gerelateerd zijn aan leesvaardigheid. Tabel 54 toont de specifieke constructen die gemeten worden bij PISA 2018. Meer gedetailleerde omschrijvingen van die constructen en informatie over het belang van die constructen zijn te vinden in het PISA 2018 Assessment and Analytical Framework (OECD, 2019a).

Het is belangrijk om op te merken dat PISA bepaalde achtergrondvariabelen niet noodzakelijk op dezelfde manier operationaliseert als een specifiek onderwijssysteem. Zo beschrijft het Vlaams Ministerie van Onderwijs en Vorming (2012) de SES van leerlingen bijvoorbeeld aan de hand van de thuistaal, het ontvangen van een schooltoeslag en het opleidingsniveau van de moeder. PISA daarentegen operationaliseert SES aan de hand van de PISA index voor economische, sociale en culturele status (ESCS). Die ESCS-index combineert verschillende achtergrondvariabelen van leerlingen: (1) het beroep van de ouders, (2) het onderwijsniveau van de ouders, (3) de economische, culturele en educatieve middelen van het gezin en (4) het aantal boeken thuis (De Meyer et al., 2019; OECD, 2019c). De gemeten achtergrondvariabelen worden vooral gebruikt om verschillen tussen leerlingengroepen te duiden en om een meer genuanceerd beeld te bieden van de prestaties op school- en systeemniveau (OECD, 2019a).

Tabel 54

Conceptueel kader voor achtergrondkenmerken bij PISA 2018

Kenmerken van de leerlingen
<ul style="list-style-type: none">• ESCS-index (SES), achtergrond van familie en thuissituatie;• cultuur en migratie;• schoolloopbaan in de vroege kindertijd;• buitenschoolse leeservaringen.
Kenmerken van de school
Onderwijzen en leren <ul style="list-style-type: none">• kwalificatie van de leerkrachten en professionele ontwikkeling;• onderwijsleerpraktijken voor lezen;• tijd voor leren en curriculum.
Schoolbeleid <ul style="list-style-type: none">• leerklimaat van de school voor lezen;• ouderbetrokkenheid;• schoolklimaat: vertrouwen, interpersoonlijke relaties en verwachtingen;• schoolcontext en middelen.
Schoolbestuur <ul style="list-style-type: none">• evaluatie en verantwoordelijkheid;• allocatie, selectie, keuzes en zittenblijven.

Non-cognitieve en metacognitieve kenmerken

Uitkomsten, motivatie, attitudes en strategieën gerelateerd aan lezen

- leesmotivatie, betrokkenheid en praktijken;
- metacognitie.

Dispositionele en schoolgerelateerde variabelen

- prestatiemotieven;
- competitie;
- wil om hard te werken;
- faalangst;
- growth mindset;
- doorzettingsvermogen;
- subjectief welbevinden;
- ICT-motivatie en praktijken;
- onderwijsvertuigingen en attitude t.o.v. school;
- prestatiedoelen.

Disposities voor globale competenties

- communicatie en relaties onderhouden;
- kennis en interesse in globale ontwikkelingen, uitdagingen en trends;
- flexibiliteit en openheid;
- emotionele sterkte en veerkracht.

Noot. Elke cyclus wordt het conceptueel kader voor achtergrondkenmerken geüpdatet met aandacht voor het hoofddomein van die cyclus. Bovenstaand kader komt dus niet helemaal overeen met de conceptueel kaders voor achtergrondkenmerken bij andere cycli.

5.1.3 Populatie

Bij PISA bestaat de populatie uit alle schoolgaande 15-jarigen uit de landen die deelnemen aan PISA.⁶⁴ Concreet test PISA leerlingen uit het secundair onderwijs met een leeftijd van 15 jaar en 3 maanden tot 16 jaar en 2 maanden op het moment van de afname. Bij PISA 2018 gaat het dus om leerlingen uit het secundair onderwijs die geboren zijn in 2002 ongeacht het leerjaar dat ze volgen. Leerlingen die niet geboren zijn in 2002 behoren niet tot de populatie. De OESO kiest voor die leeftijdsgroep, omdat de school- of leerplicht in verschillende landen eindigt op 16 jaar (De Meyer et al., 2019; OECD, 2019b). Die argumentatie gaat echter niet op voor de Belgische context waar er leerplicht is tot 18 jaar (Vlaamse overheid, n.d.).

Om het Vlaamse onderwijssysteem zo goed mogelijk te weerspiegelen, worden ook leerlingen uit het deeltijds beroepssecundair onderwijs (DBSO) en het buitengewoon secundair onderwijs (BuSO) opgenomen in de populatie.⁶⁵ Concreet gaat het bij het BuSO om leerlingen die les volgen in type basisaanbod (OV3) of type 1 (OV3). Leerlingen die les volgen in andere types of opleidingsvormen, maken geen deel uit van de populatie (De Meyer et al., 2019). Internationaal is beslist om anderstalige

⁶⁴ Er zijn daarnaast ook heel wat 15-jarigen die geen onderwijs volgen. Zij behoren niet tot de PISA-populatie (OECD, 2019b).

⁶⁵ Aangezien de normale toetsboekjes niet aangepast zijn voor leerlingen met specifieke onderwijsnoden, wordt er een aangepast toetsboekje voorzien voor leerlingen uit het DBSO en BuSO, namelijk het UH-boekje (Une Heure-boekje). Dat boekje is een verkorte versie van het normale toetsboekje en duurt daardoor slechts een uur. Ook de achtergrondvragenlijst wordt voor die groepen leerlingen ingekort tot een vragenlijst van 15 minuten (De Meyer et al., 2019). Een aangepast toetsboekje met minder items zou echter kunnen leiden tot minder accurate schattingen van de vaardigheid van leerlingen uit het BuSO en DBSO.

nieuwkomers niet op te nemen in de populatie, aangezien het onvoldoende beheersen van de onderwijstaal een te grote belemmering vormt bij de toetsafname (De Meyer et al., 2019; OECD, 2021).

5.1.4 Steekproef

Aan PISA 2018 namen wereldwijd 612000 15-jarige leerlingen uit 79 verschillende landen deel. Van die landen zijn er 37 landen lid van de OESO. Daarnaast nemen ook 42 niet-lidstaten deel aan PISA (De Meyer et al., 2019). De steekproeftrekking van PISA gebeurt in elk land op een gelijkaardige manier (OECD, 2021).

De eerste stap omvat een **steekproeftrekking van scholen** in elk land, weliswaar met andere stratificatievariabelen per land (OECD, 2021). In Vlaanderen werd er voor PISA 2018 een gestratificeerde steekproef van 188 scholen getrokken uit alle Vlaamse scholen waar leerlingen les volgen die geboren zijn in 2002. Daarnaast werden er twee gelijkaardige reservescholen geselecteerd voor elk school in de steekproef (De Meyer et al., 2019). Bij de steekproeftrekking van scholen in het Vlaamse onderwijs wordt er rekening gehouden met twee expliciete stratificatievariabelen, namelijk onderwijsnet en onderwijsvorm. De stratificatievariabele 'onderwijsnet' onderscheidt het officieel en het vrij onderwijs. Tot slot onderscheidt PISA zes onderwijsvormen (De Meyer et al., 2019):

- algemeen secundair onderwijs (ASO);
- beroepssecundair onderwijs (BSO);
- kunstsecundair onderwijs (KSO);
- technisch secundair onderwijs (TSO);
- buitengewoon secundair onderwijs (BuSO)⁶⁶;
- deeltijds beroepssecundair onderwijs (DBSO).

In een tweede stap volgt de **selectie van leerlingen** binnen de scholen. Voor elke deelnemende school wordt er een lijst opgesteld van leerlingen die geboren zijn in 2002. In die lijst wordt de leerlingen gerangschikt volgens de studierichting en leerjaar, waarna er at random 35 leerlingen geselecteerd worden uit de opgestelde lijst (De Meyer et al., 2019). Daarbij werden de **internationale uitsluitingsnormen op individueel niveau** toegepast. In PISA 2018 werden 24 leerlingen uitgesloten, omwille van (1) ernstige fysieke beperkingen, (2) ernstige mentale beperkingen en (3) onvoldoende beheersing van de Nederlandse taal doordat ze minder dan één jaar Vlaams onderwijs volgen (OECD, 2021).

Bij PISA wordt de norm gehanteerd dat niet meer dan 5% van de populatie in een onderwijssysteem uitgesloten mag worden (De Meyer et al., 2019). Bij PISA 2018 bedroeg het percentage uitsluitingen in Vlaanderen 2,32%. De Vlaamse steekproef voldoet op die manier aan de vooropgestelde norm (OECD, 2019b). De feitelijke PISA-steekproef in Vlaanderen bestond uit 4882 leerlingen uit 147 scholen. De steekproef is representatief met de populatie van Vlaamse scholen op vlak van onderwijsnet, onderwijsvorm en leerjaar. Daarnaast is de steekproef ook representatief voor de Vlaamse leerlingenpopulatie op vlak van onderwijsvorm en leerjaar (De Meyer et al., 2019).

5.1.5 Modus

De dataverzameling behoort tot de verantwoordelijkheid van de onderzoekscentra die instaan voor de coördinatie van PISA in de deelnemende landen. Met oog op de standaardisatie en de

⁶⁶ Internationaal heeft de OESO beslist om scholen die onderwijs bieden aan leerlingen met ernstige mentale en fysieke beperkingen uit te sluiten bij de steekproeftrekking van scholen. Daarom worden er in Vlaanderen enkel scholen voor buitengewoon onderwijs opgenomen in de steekproef die basisaanbod (OV3) en type 1 (OV3) aanbieden (De Meyer et al., 2019; OECD, 2021).

vergelijkbaarheid van de data over landen heen heeft de OESO de procedure voor dataverzameling stapsgewijs omschreven in de PISA 2018 Technical Standards⁶⁷ (De Meyer et al., 2019; OECD, 2015). Na de dataverzameling verzamelt PISA van alle nationale onderzoekscentra informatie over de voorbereiding en uitvoering van de dataverzameling. Op basis van die informatie gaat PISA na in welke mate de nationale onderzoekscentra de PISA 2018 Technical Standards nauwgezet gevolgd hebben om zo de globale kwaliteit van de verzamelde data te garanderen (OECD, 2019b).

5.1.5.1 Toetsontwikkeling

Eerst wordt het **conceptueel kader gedefinieerd** voor PISA 2018. Het conceptueel kader voor leesvaardigheid is door een groep internationale experts grondig herwerkt voor de afname in 2018, terwijl de kaders voor wiskundige geletterdheid en wetenschappelijke geletterdheid overgenomen worden uit PISA 2015 (zie 5.1.2 Domein en construct). Naast het conceptueel kader worden ook de items voor wiskundige en wetenschappelijke geletterdheid overgenomen uit voorgaande cycli. Er worden enkel nieuwe items ontwikkeld voor het hoofddomein leesvaardigheid. Het herwerkte conceptuele kader voor leesvaardigheid vormt daarbij het uitgangspunt en draagt op die manier bij aan de inhoudelijke validiteit van de toets voor leesvaardigheid (OECD, 2019b, 2021).

Op basis van het conceptueel kader heeft het PISA Consortium **nieuwe items ontwikkeld**. Bij die ontwikkeling zijn ook de onderzoekscentra van de deelnemende landen betrokken. Concreet zijn er op die manier 318 nieuwe leesitems ontwikkeld. Die items bestaan uit zowel open vragen als meerkeuzevragen. De ontwikkelde items worden geëvalueerd door de groep internationale experts die het conceptueel kader ontwikkelde om te garanderen dat de items en de test in zijn geheel tegemoetkomen aan de conceptuele kaders (OECD, 2019b, 2021).

In een volgende stap heeft de OESO aan de nationale onderzoekscentra en internationale experts gevraagd om de **geschiktheid van de items te evalueren** aan de hand van twee criteria. Ten eerste moeten de items geschikt zijn voor 15-jarigen wat betreft de inhoud, de context en de cognitieve vaardigheden. Ten tweede moeten de items geschikt zijn voor een internationale toets en relevant zijn voor verschillende culturele contexten. Daarna volgt de **selectie van items**. Bij die selectie werd er rekening gehouden met (1) de evaluatie van nationale onderzoekscentra, (2) de feedback van de internationale experts en (3) het richtpercentage voor het aantal items per dimensie. De geselecteerde items zijn vervolgens vertaald, waarna het PISA consortium de vertaling heeft geverifieerd (OECD, 2019b, 2021). Tabel 55 toont een overzicht van het vooropgesteld percentage items bij de toets leesvaardigheid (OECD, 2019a).

⁶⁷ Wanneer er niet aan de vereisten uit de PISA 2018 Technical Standards voldaan wordt, zal de OESO bepalen of de data opgenomen worden in de internationale dataset van PISA 2018 (OECD, 2015).

Tabel 55*Overzicht vooropgesteld percentage items voor leesvaardigheid per tekstbron en proces*

Tekstbronnen	%
Eén bron	65
Meerdere bronnen	35
Processen	%
Informatie lokaliseren	25
Begrijpen	45
Evalueren en reflecteren	30

Noot. PISA onderscheidt de andere dimensies uit het conceptueel kader niet uitdrukkelijk bij het vooropgesteld percentage items. Wel zorgt PISA ervoor dat de ontwikkelde items ongeveer gelijkmatig verdeeld zijn over die dimensies (OECD, 2019a).

Vervolgens wordt er een internationaal **vooronderzoek** georganiseerd om de haalbaarheid van adaptief toetsen bij leesvaardigheid en de psychometrische validiteit van de toets na te gaan. Voor het vooronderzoek is er een aparte steekproef van scholen getrokken. Daarnaast is er in het vooronderzoek ook een toets op papier afgenomen, zodat PISA het verschil met de digitale toets kon nagaan. Op basis van de resultaten van het vooronderzoek heeft de OESO de kwaliteit en betrouwbaarheid van zowel de nieuw ontwikkelde items als de overgenomen items uit voorgaande cycli geanalyseerd (OECD, 2019b, 2021).

Na het vooronderzoek konden de nationale onderzoekscentra feedback geven op de items die opgenomen waren in het vooronderzoek. Ten eerste konden ze problemen met de vertaling of lay-out melden die opgemerkt waren tijdens het vooronderzoek. Ten tweede werd aan de nationale onderzoekscentra gevraagd om items die niet overeenkwamen met de internationale resultaten te bekijken (OECD, 2019b, 2021).

Tot slot volgde de **finale selectie van items** voor de adaptieve toets leesvaardigheid. Eerst zijn items met een te lage discriminatiegraad, een te lage of hoge moeilijkheidsgraad en een lage samenhang met leesvaardigheid verwijderd. Daarna volgde de selectie van items rekening houdend met de trenditems uit voorgaande afnames en het conceptueel kader. Uiteindelijk worden er voor leesvaardigheid 245 trend- en nieuwe items geselecteerd die samengevoegd worden tot blokken. Die blokken vormen de basis voor het design van de adaptieve toets (zie 5.1.5.2 Opbouw van de adaptieve toets leesvaardigheid). Zoals eerder aangegeven, worden de items voor wiskundige en wetenschappelijke geletterdheid overgenomen uit voorgaande cycli (OECD, 2019b, 2021). Tabel 56 toont een overzicht van het aantal items per domein dat uiteindelijk werd opgenomen in PISA 2018 (OECD, 2021).

Tabel 56*Overzicht aantal items in PISA 2018 per domein*

	Trenditems	Nieuwe items	Totaal
Leesvaardigheid	72	173	245
Wiskundige geletterdheid	82	-	82
Wetenschappelijke geletterdheid	115	-	115

Nadat bepaald is welke items worden opgenomen in PISA 2018 worden de **toetsboekjes samengesteld** aan de hand van clusters. In totaal zijn er 36 verschillende toetsboekjes samengesteld die steeds bestaan uit een adaptieve toets leesvaardigheid en twee clusters. Alle leerlingen hebben de adaptieve toets leesvaardigheid afgelegd. Echter, niet alle leerlingen hebben een cluster van wiskundige en

wetenschappelijke geletterdheid gekregen. Tabel 57 geeft een overzicht weer van de samenstelling van de verschillende toetsboekjes⁶⁸ (OECD, 2021).

Tabel 57

Samenstelling PISA-toetsboekjes

Percentage leerlingen	Toetsboekje	Cluster 1	Cluster 2	Cluster 3	Cluster 4
46%	1		R(adaptive)	M01	M02
	2		R(adaptive)	M02	M03
	3		R(adaptive)	M03	M04
	4		R(adaptive)	M04	M05
	5		R(adaptive)	M05	M06a/b
	6		R(adaptive)	M06a/b	M01
	7	M01	M03	R(adaptive)	
	8	M02	M04	R(adaptive)	
	9	M03	M05	R(adaptive)	
	10	M04	M06a/b	R(adaptive)	
	11	M05	M01	R(adaptive)	
	12	M06a/b	M02	R(adaptive)	
46%	13		R(adaptive)	S01	S02
	14		R(adaptive)	S02	S03
	15		R(adaptive)	S03	S04
	16		R(adaptive)	S04	S05
	17		R(adaptive)	S05	S06
	18		R(adaptive)	S06	S01
	19	S01	S03	R(adaptive)	
	20	S02	S04	R(adaptive)	
	21	S03	S05	R(adaptive)	
	22	S04	S06	R(adaptive)	
	23	S05	S01	R(adaptive)	
	24	S06	S02	R(adaptive)	
8%	25		R(adaptive)	S01	M01
	26		R(adaptive)	M02	S02
	27		R(adaptive)	S03	M03
	28		R(adaptive)	M04	S04
	29		R(adaptive)	S05	M05
	30		R(adaptive)	M06a/b	S06
	31	M01	S01	R(adaptive)	
	32	S02	M02	R(adaptive)	
	33	M03	S03	R(adaptive)	
	34	S04	M04	R(adaptive)	
	35	M05	S05	R(adaptive)	
	36	S06	M06a/b	R(adaptive)	

Noot. R(adaptive) = leesvaardigheid adaptief getoetst, M = wiskundige geletterdheid, S = wetenschappelijke geletterdheid, a = standaard cluster en b = gemakkelijkere cluster.

⁶⁸ De leerlingen uit het DBSO en BuSO krijgen een aangepast toetsboekje, namelijk het UH-boekje (Une Heure-boekje). Dat is een verkorte versie die slechts een uur duurt. Ook de achtergrondvragenlijst wordt voor die groepen leerlingen ingekort tot een vragenlijst van 15 minuten (De Meyer et al., 2019).

Ten slotte ontvangen de deelnemende landen de internationale versie van de toets in het Engels en in het Frans. Op die manier kan elk deelnemend land bij de **vertaling van de toets** een dubbel vertalingsdesign hanteren. Daarbij hebben twee onafhankelijke vertalers de toets vanuit respectievelijk het Engels of het Frans naar het Vlaams vertaald. Die twee vertalingen worden daarna vergeleken en geïntegreerd tot een Vlaamse versie. De correctheid van die uiteindelijke Vlaamse versie wordt geëvalueerd door een team van experts dat aangesteld is door het PISA Consortium (OECD, 2019a, 2021).

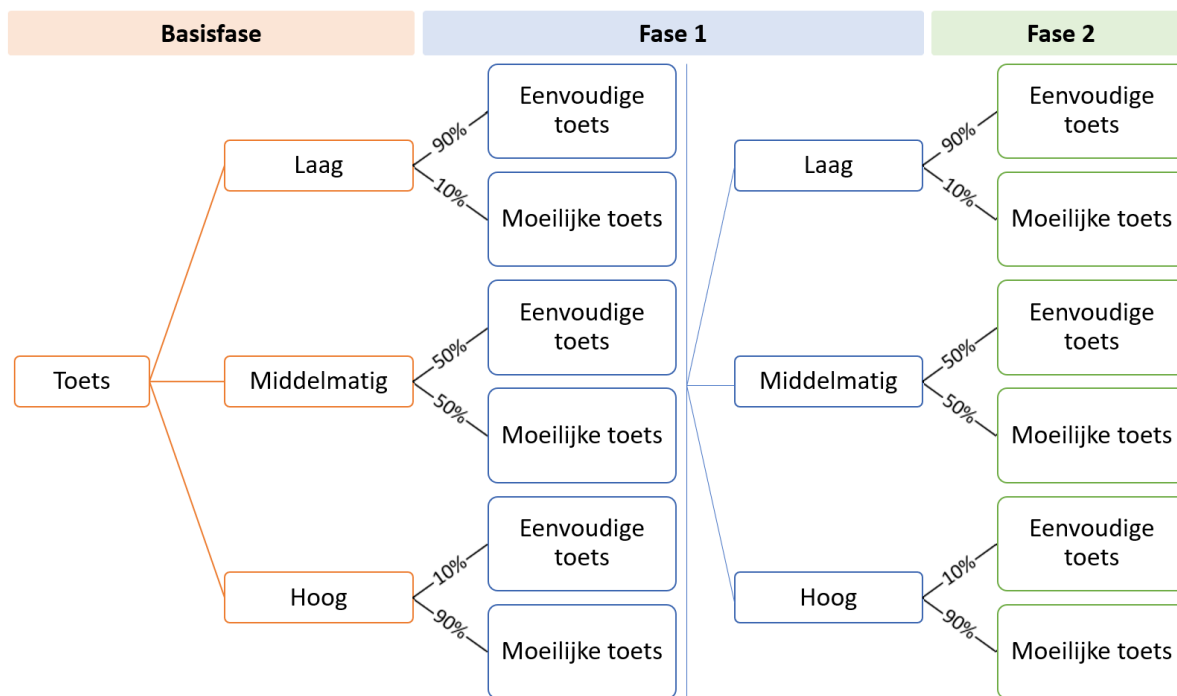
Naast de PISA-toetsen ontwikkelt de OESO ook **achtergrondvragenlijsten voor leerlingen, ouders en directeurs** ontwikkeld. Eerst wordt het conceptueel kader geüpdatet dat de basis vormt voor achtergrondvragenlijsten (zie 5.1.2 Domein en construct). Daarbij wordt er rekening gehouden met het conceptueel kader uit zowel PISA 2015 (de vorige afname) als PISA 2009 (de vorige afname met leesvaardigheid als hoofddomein). In lijn met het herwerkte kader heeft een groep experts de achtergrondvragenlijsten ontworpen. Die achtergrondvragenlijsten vormen de basis voor de nationale versies.⁶⁹ Naast het vertalen van de toets kunnen de nationale onderzoekscentra indien nodig culturele aanpassingen doen die verder gaan dan louter een directe vertaling. Die aanpassingen worden gedocumenteerd door de onderzoekscentra en geëvalueerd door de OESO. Ook de correctheid van de vertaling wordt gecontroleerd. De OESO stelt daarvoor experts aan die de vertaling verifiëren. Wanneer een nationale versie alle kwaliteitschecks doorstaat, wordt die versie opgeladen in het digitale afnamesysteem voor het vooronderzoek. Vervolgens wordt de psychometrische kwaliteit van die vragenlijst geëvalueerd. Op basis van de resultaten en verzamelde feedback uit het vooronderzoek worden de internationale vragenlijsten herwerkt, waarna de nationale versies herwerkt worden voor het hoofdonderzoek bij PISA 2018. Nadat de OESO die finale nationale versies goedgekeurd heeft, volgt de implementatie van de afgewerkte vragenlijsten in het digitale afnameplatform van PISA (OECD, 2021).

De OESO maakt gebruik van bijkomende procedures om de kwaliteit van de achtergrondvragenlijsten te garanderen. Ten eerste worden de nieuw ontwikkelde vragen voorgelegd aan de deelnemende landen voor diepgaande feedback. Daarbij wordt er gevraagd om de relevantie van een vraag te beoordelen voor het onderwijssysteem van dat land. Ook eventuele culturele of politieke gevoeligheden en potentiële taalmoeilijkheden worden opgenomen in de review van alle deelnemende landen. Ten tweede, worden nieuw ontwikkelde vragen uit de leerlingvragenlijst uitgetest in zowel het Frans als het Engels. Daarna wordt alle feedback verzameld en worden de vragen herbekeken en waar nodig aangepast. Ten derde worden de vragen vertaald naar talen uit verschillende taalgroepen om de vergelijkbaarheid te verhogen. Taalexperts geven op basis van die vertalingen aan waar culturele bias of niet-vertaalbaarheid een probleem zou kunnen vormen.

5.1.5.2 Opbouw van de adaptieve toets leesvaardigheid

Leesvaardigheid is in 2018 voor het eerst adaptief getest. Dat wil zeggen dat leerlingen items voorgeschoteld krijgen op basis van hun (geschatte) vaardigheidsniveau. De adaptieve PISA-toets bestaat uit drie fasen: de basisfase, fase 1 en fase 2 die opgebouwd zijn uit zowel nieuwe items als trenditems (De Meyer et al., 2019; OECD, 2021). Figuur 37 toont een schematisch overzicht van het adaptief design voor de toets leesvaardigheid.

⁶⁹ Wanneer er nationale versies beschikbaar zijn van vragenlijstitems uit voorgaande cycli die ook in PISA 2018 opgenomen worden, neemt PISA de nationale versies van die vragenlijstitems automatisch over tenzij er aan de internationale versie wijzigingen gebeuren (OECD, 2015).



Figuur 37. Schematisch overzicht van het adaptief design voor de toets leesbaarheid

Alle leerlingen beginnen met een toets uit de **basisfase**. Ze worden at random toegewezen aan een specifieke versie van de toets uit de basisfase. In totaal zijn er acht verschillende versies waartussen er overlap is. Minstens 80% van de items uit de toets bij de basisfase wordt automatisch gescoord. De leerlingen worden op basis van die score in drie prestatiecategorieën ingedeeld: lage, middelmatig en hoge prestatie (De Meyer et al., 2019; OECD, 2021).

In **fase 1** krijgen de leerlingen een eenvoudige of moeilijke toets op basis van hun geschatte niveau tijdens de basisfase. Concreet gebeurt de toewijzing aan een toets bij fase 1 aan de hand van drie criteria: (1) de toegewezen versie bij de basisfase, (2) de prestatiecategorie waarin de leerling ingedeeld is bij de basisfase en (3) een vooropgestelde kans. Zo hebben leerlingen uit de categorie 'lage prestatie' 90% kans om een eenvoudige test en 10% kans om een moeilijke test te krijgen. Wanneer leerlingen tot de categorie 'middelmatige' prestatie behoren, hebben ze evenveel kans op een eenvoudige als moeilijke toets. Leerlingen uit de categorie 'hoge prestatie' hebben 10% kans om een eenvoudige toets te krijgen en 90% kans op een moeilijke toets. Ook bij de toetsen in fase 1 worden bepaalde items automatisch gescoord. De leerlingen worden op einde van fase 1 opnieuw ingedeeld in drie prestatiecategorieën (De Meyer et al., 2019; OECD, 2021).

Tot slot wordt een van de versies uit **fase 2** toegewezen aan een leerling. Dat gebeurt opnieuw op basis van (1) de toegewezen versie van de toets bij fase 1, (2) de prestatiecategorie waarin leerlingen ingedeeld werden bij de basisfase en fase 1 en (3) de vooropgestelde kansen. Ook bij fase 2 is er overlap tussen de verschillende versies van de toetsen. Tabel 58, Tabel 59 en Tabel 60 tonen de samenstelling van de toetsversies bij respectievelijk de basisfase, fase 1 en fase 2 (De Meyer et al., 2019; OECD, 2021).

Tabel 58*Samenstelling van de toetsversies uit de basisfase*

Versies toets basisfase	Blok 1	Blok 2
RC1	C1 (5 items)	C2 (5 items)
RC2	C1 (5 items)	C4 (3 items)
RC3	C3 (4 items)	C4 (3 items)
RC4	C2 (5 items)	C5 (5 items)
RC5	C4 (3 items)	C5 (5 items)
RC6	C2 (5 items)	C3 (4 items)
RC7	C1 (5 items)	C3 (4 items)
RC8	C1 (5 items)	C5 (5 items)

*Noot. RC = toetsversie uit de basisfase, C = blok uit de basisfase.***Tabel 59***Samenstelling van de toetsversies uit fase 1*

Versies toets fase 1	Blok 1	Blok 2	Blok 3
R11H	S1.01 (5 items)	S1.03 (6 items)	S1.14 (4 items)
R11L	S1.07 (5 items)	S1.08 (5 items)	S1.21 (3 items)
R12H	S1.01 (5 items)	S1.04 (6 items)	S1.05 (3 items)
R12L	S1.10 (5 items)	S1.11 (6 items)	S1.19 (4 items)
R13H	S1.02 (6 items)	S1.06 (4 items)	S1.16 (4 items)
R13L	S1.08 (5 items)	S1.10 (5 items)	S1.12 (4 items)
R14H	S1.03 (6 items)	S1.06 (4 items)	S1.18 (5 items)
R14L	S1.09 (6 items)	S1.12 (4 items)	S1.23 (5 items)
R15H	S1.02 (6 items)	S1.13 (6 items)	S1.15 (3 items)
R15L	S1.09 (6 items)	S1.19 (4 items)	S1.20 (5 items)
R16H	S1.13 (6 items)	S1.16 (4 items)	S1.17 (6 items)
R16L	S1.07 (5 items)	S1.22 (6 items)	S1.23 (5 items)
R17H	S1.04 (6 items)	S1.14 (4 items)	S1.18 (5 items)
R17L	S1.20 (5 items)	S1.22 (6 items)	S1.24 (4 items)
R18H	S1.05 (3 items)	S1.15 (3 items)	S1.17 (6 items)
R18L	S1.11 (6 items)	S1.21 (3 items)	S1.24 (4 items)

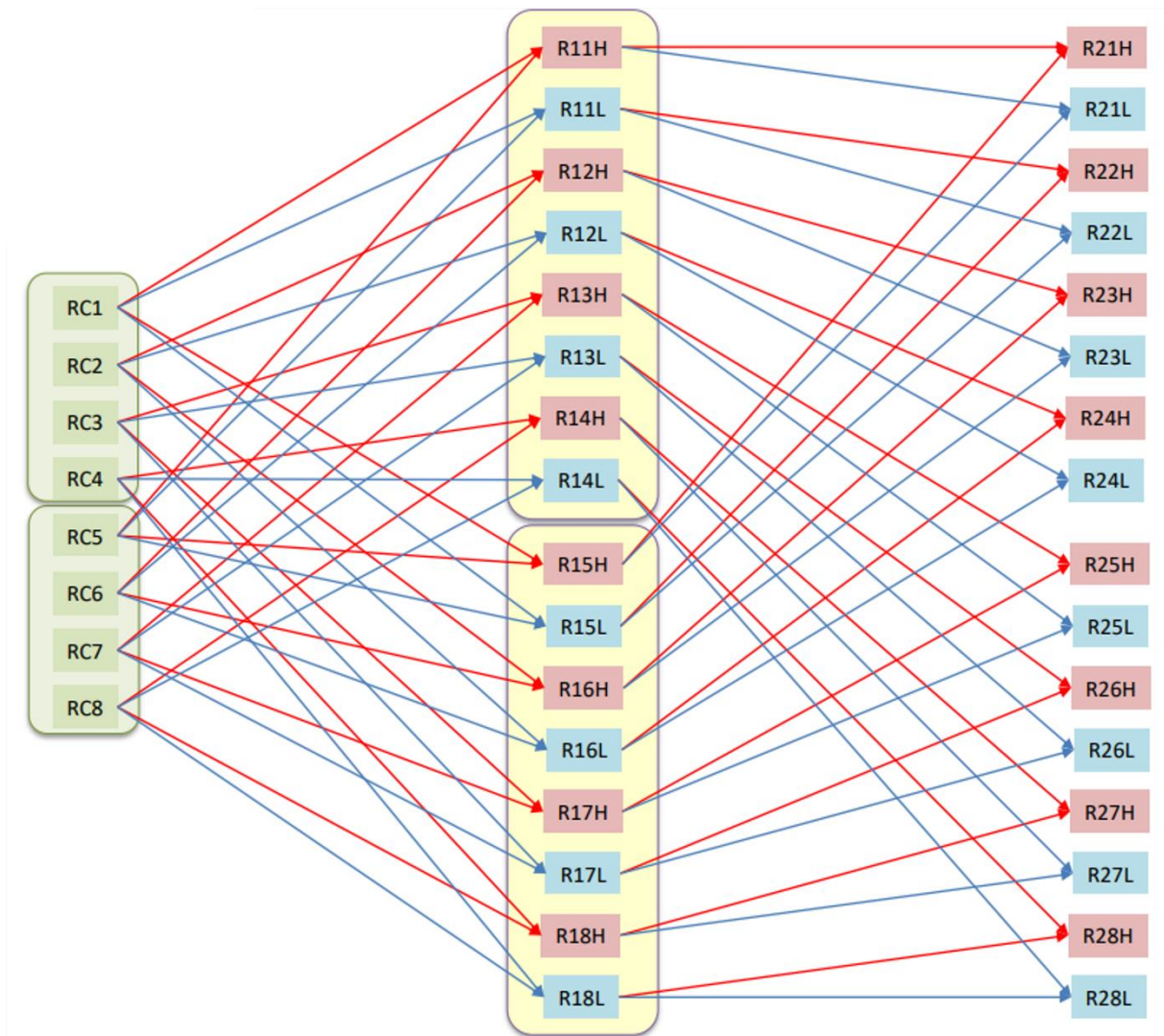
Noot. R1 = toetsversie uit fase 1, H = moeilijke toets, L = eenvoudige toets, S1 = blok uit fase 1.

Tabel 60*Samenstelling van de toetsversies uit fase 2*

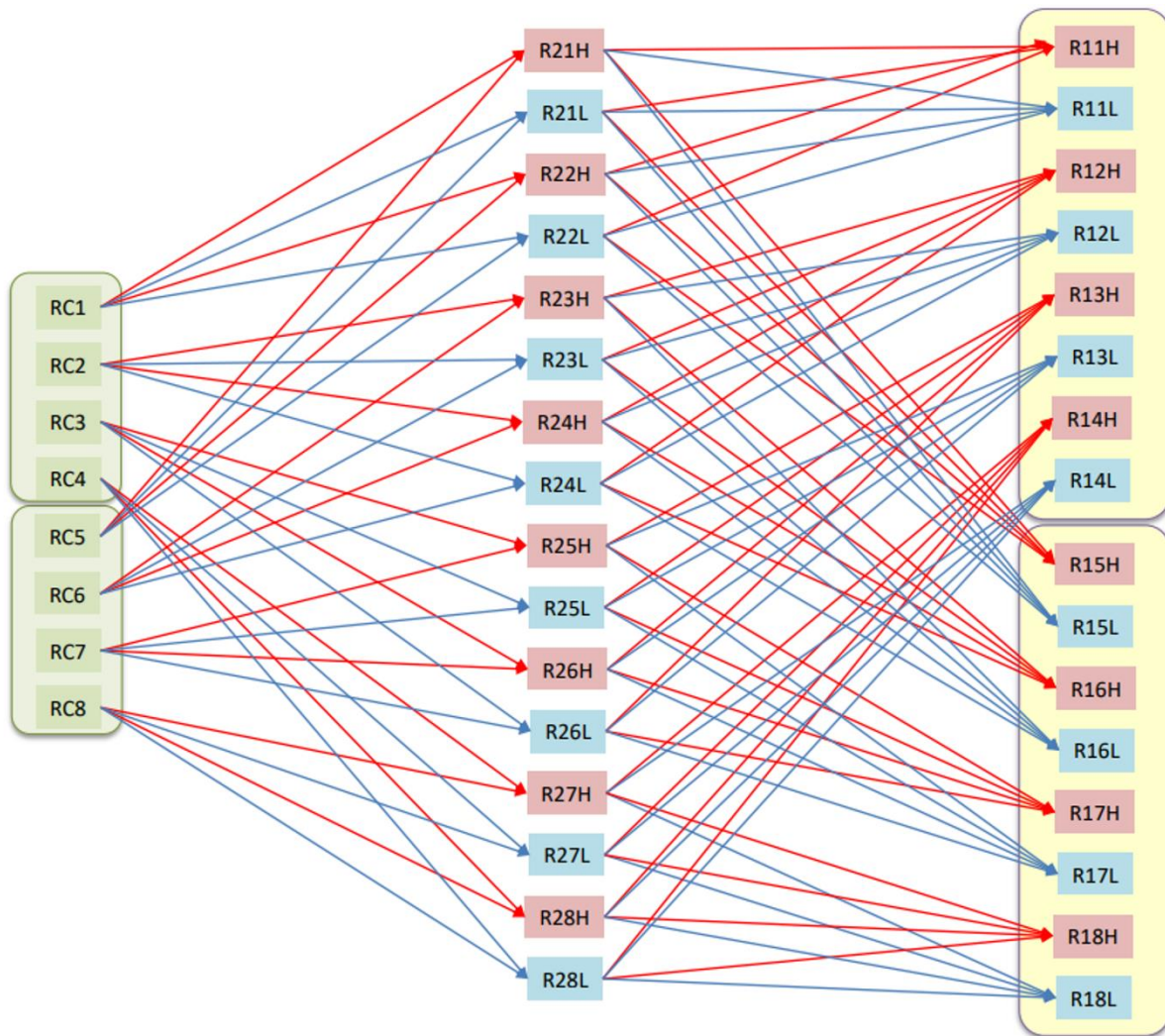
Versies toets fase 2	Blok 1	Blok 2
R21H	S2.05 (7 items)	S2.09 (7 items)
R21L	S2.01 (7 items)	S2.13 (5 items)
R22H	S2.01 (7 items)	S2.15 (5 items)
R22L	S2.07 (8 items)	S2.11 (7 items)
R23H	S2.05 (7 items)	S2.10 (7 items)
R23L	S2.02 (7 items)	S2.15 (5 items)
R24H	S2.02 (7 items)	S2.13 (5 items)
R24L	S2.07 (8 items)	S2.12 (6 items)
R25H	S2.06 (7 items)	S2.09 (7 items)
R25L	S2.03 (7 items)	S2.14 (7 items)
R26H	S2.03 (7 items)	S2.16 (6 items)
R26L	S2.08 (7 items)	S2.11 (7 items)
R27H	S2.06 (7 items)	S2.10 (7 items)
R27L	S2.04 (7 items)	S2.16 (6 items)
R28H	S2.04 (7 items)	S2.14 (7 items)
R28L	S2.08 (7 items)	S2.12 (6 items)

Noot. R2 = toetsversie uit fase 2, H = moeilijke toets, L = eenvoudige toets, S2 = blok uit fase 2.

In het hierboven omschreven design zijn er 64 verschillende paden mogelijk doorheen het toetsmateriaal bij PISA 2018 (zie Figuur 38). Dat design wordt toegewezen aan 75% van de leerlingen. Om de betrouwbaarheid van de schatting te verbeteren, wordt er daarnaast een alternatief design toegevoegd (zie Figuur 39), waarbij na de basisfase eerst fase 2 volgt en daarna fase 1. Dat design werd aan 25% van de leerlingen toegewezen (OECD, 2021).



Figuur 38. Paden doorheen het alternatief design voor de adaptieve toets leesvaardigheid (Figuur overgenomen uit OECD, 2021)



Figuur 39. Paden doorheen het alternatief design voor de adaptieve toets leesvaardigheid (Figuur overgenomen uit OECD, 2021)

5.1.5.3 Toetsafname

Tussen april en mei 2018 werden er in Vlaanderen verschillende instrumenten afgenomen voor PISA 2018. De leerlingen startten met de PISA-test op de computer, waarvoor ze twee uur de tijd kregen. Concreet bestond die test uit een adaptieve toets leesvaardigheid die één uur duurde en twee clusters wiskunde- en/of wetenschapsvragen die elk een halfuur duurden (De Meyer et al., 2019; OECD, 2021).

Daarnaast werd aan de leerlingen gevraagd om drie bijkomende vragenlijsten in te vullen: (1) de achtergrondvragenlijst, (2) de ICT-vragenlijst en (3) de schoolloopbaanvragenlijst. Ze kregen daarvoor in totaal 45 minuten de tijd. Tot slot werd er een ouder- en directievragenlijst⁷⁰ afgenomen (De Meyer et al., 2019). Alle achtergrondvragenlijsten werden digitaal afgenomen, behalve de oudervragenlijst (OECD, 2021).

Elke deelnemende school werd begeleid door een interne testleider (een lid van het schoolteam neemt die rol op zich) of een externe testleider van het nationaal onderzoekscentrum (De Meyer et al., 2019).

⁷⁰ PISA bevaart de directeurs uit de deelnemende scholen via de directievragenlijst. Bij het gebruik van variabelen uit die vragenlijst is het van belang om te beseffen dat de opvattingen van de bevroegde directeurs niet noodzakelijk de feitelijke realiteit op school weerspiegelen.

De te volgen procedures bij de toetsafname werden uitgebreid beschreven in de handleiding voor testleiders. De taak van de testleider bestond uit (1) het toetsmateriaal uitdelen, (2) instructies geven zoals beschreven in de handleiding, (3) ongebruikelijke zaken tijdens de toetsafname melden en (4) de USB-sticks of laptops samen met de achtergrondvragenlijsten terugbezorgen aan het nationaal onderzoekscentrum. Alle interne en externe testleiders kregen vooraf een uitgebreide training. In het kader van standaardisatie kregen ook de schoolcoördinatoren een handleiding waarin de toetsafname helder omschreven werd (De Meyer et al., 2019; OECD, 2019b, 2021).

Om een kwaliteitsvolle toetsafname te garanderen, stelt het PISA Consortium toezichthouders aan die ongeveer 5 Vlaamse scholen bezoeken tijdens de toetsafname. Die toezichthouders observeerden en rapporteerden eventuele afwijkingen van de voorgeschreven procedures. Voorafgaand aan de afname ontvingen zij een training vanuit het PISA Consortium (OECD, 2015, 2019b).

Tijdens de afname werd al een gedeelte van de antwoorden automatisch gescoord. Na de afname werden de antwoorden van de leerlingen op de andere items via de computer manueel of automatisch⁷¹ gescoord (zie Tabel 61 voor een overzicht). Bij de scoring werd niet gekeken naar eventuele schrijffouten, aangezien PISA leesvaardigheid – en geen schrijfvaardigheden (zoals spelling) – vooropstelt. In functie van valide en betrouwbare scoring organiseerde PISA de scoring via een online systeem waarin uitgebreide scoringswijzers en vaste scoringscodes opgenomen waren. Daarnaast werden er zowel internationale als nationale trainingen voorzien over de scoringsprocedures (OECD, 2019b, 2021). Bovendien werd de betrouwbaarheid van de manuele scoring bij open vragen in kaart gebracht op twee niveaus: de betrouwbaarheid van de scoring binnen een land en over verschillende landen heen (OECD, 2019b).

Tabel 61

Overzicht aantal items manueel en automatisch gescoord bij PISA 2018

	Aantal items gescoord		Totaal
	Manueel	Automatisch	
Leesvaardigheid	82	163	245
Wiskundige geletterdheid	21	61	82
Wetenschappelijke geletterdheid	32	83	115

Nadat de Vlaamse data aan de OESO bezorgd zijn, wordt aan het nationaal onderzoekscentrum gevraagd om een vragenlijst in te vullen om feedback te geven op het verloop van het dataverzamelingsproces. Die feedback wordt meegenomen om de kwaliteit van toekomstige PISA-cycli verder te verbeteren (OECD, 2021).

5.2 Standaardbepaling

PISA is geen criteriumgerichte meting, aangezien de resultaten van de leerlingen niet vergeleken worden met een ‘absoluut’ criterium. Wel onderscheidt PISA acht vaardigheidsniveaus⁷² (zie Tabel 62)

⁷¹ Het automatische scoringssysteem van PISA baseert zich op data uit PISA 2015 en het vooronderzoek bij PISA 2018 en hun bijhorende manuele scoring. Het systeem is specifiek voor elke taalgroep binnen een deelnemend land. Concreet zorgt het automatisch scoringssysteem ervoor dat dezelfde antwoorden automatisch dezelfde scoringscode toegewezen krijgen (OECD, 2021).

⁷² De PISA-vaardigheidsniveaus werden ontwikkeld bij PISA 2000. Sindsdien zijn niveau 1b en 1c toegevoegd om de vaardigheden in kaart te brengen van leerlingen die onder niveau 1a scoren. Ook niveau 6 is sinds 2009 toegevoegd aan de PISA-vaardigheidsniveaus (De Meyer et al., 2019; OECD, 2019a).

op de PISA-meetschaal⁷³. Die vaardigheidsniveaus gebruikt PISA om een inhoudelijke betekenis te geven aan de toetscores. Zo vormt Niveau 2 het referentiepunt. Dat wil zeggen dat leerlingen die niveau 2 behalen volgens PISA beschikken over de minimale vaardigheden die nodig zijn om dagdagelijkse problemen op te lossen. Leerlingen die niveau 5 of 6 behalen, omschrijft PISA als toppresterders. Leerlingen die presteren op niveau 1a, 1b en 1c en dus niet het referentiepunt behalen, beschouwt PISA als laagpresteerders (De Meyer et al., 2019; OECD, 2021). Het gebruik van vaardigheidsniveaus vereenvoudigt de communicatie over de PISA-resultaten. De OESO geeft echter zelf aan dat die opdeling in specifieke vaardigheidsniveaus eerder arbitrair is (OECD, 2021).

Tabel 62

Overzicht van de vaardigheidsniveaus per domein bij PISA 2018

	Leesvaardigheid	Wiskundige geletterdheid	Wetenschappelijke geletterdheid
Niveau 6	≥ 698,32	≥ 669,30	≥ 707,93
Niveau 5	≥ 625,61	≥ 606,99	≥ 633,33
Niveau 4	≥ 552,89	≥ 544,68	≥ 558,73
Niveau 3	≥ 480,18	≥ 482,38	≥ 484,14
Niveau 2	≥ 407,47	≥ 420,07	≥ 409,54
Niveau 1a	≥ 334,75	≥ 357,77	≥ 334,94
Niveau 1b	≥ 262,04	-	≥ 260,54
Niveau 1c	≥ 189,33	-	-

Aangezien wiskundige en wetenschappelijke geletterdheid niet de focus vormen van PISA 2018 worden de inhoudelijke beschrijvingen voor de vaardigheidsniveaus overgenomen uit respectievelijk PISA 2012 en PISA 2015. Bij leesvaardigheid worden de inhoudelijke beschrijvingen voor de vaardigheidsniveaus wel aangepast met het oog op het herwerkte conceptueel kader (zie 5.1.2 Domein en construct). Die aanpassing gebeurt nadat de data verzameld zijn. Eerst worden alle items voor leesvaardigheid geclassificeerd volgens het cognitieve proces uit het conceptueel kader. Vervolgens worden alle items op basis de moeilijkheidsgraad (zoals blijkt uit de PISA-afname) ingedeeld in een van de verschillende vaardigheidsniveaus. Op basis van de inhoud en cognitieve processen van alle items behorende tot een specifiek vaardigheidsniveau ontwikkelt PISA een globale inhoudelijke beschrijving voor dat specifieke vaardigheidsniveau (OECD, 2021).

PISA is een cross-sectioneel onderzoek. De vaardigheden van leerlingen worden namelijk slechts een keer gemeten. Via PISA is het echter wel mogelijk om trends van de gemiddelde vaardigheden van 15-jarige leerlingen doorheen de tijd in kaart te brengen (zie 5.1.1 Doel). Daarvoor maakt PISA gebruik van trenditems die ook in voorgaande cycli afgenomen werden. Op basis van die trenditems kan de lineaire transformatie bepaald worden die nodig is om de plausible values⁷⁴ van PISA 2018 op de PISA-meetschaal te plaatsen. Concreet gebruikt de OESO bij PISA 2018 dezelfde transformatiecoëfficiënten als in 2015. De itemparameters uit 2018 werden namelijk op dezelfde schaal geschat als de itemparameters uit 2015. Van zodra de data van PISA 2018 op de internationale PISA-meetschaal staan, is het mogelijk om de resultaten te interpreteren en evoluties in kaart te brengen (OECD, 2021).

⁷³ De PISA-meetschalen voor de drie domeinen werden vastgelegd op het moment dat voor het eerst een meetschaal werd opgesteld voor dat domein. Daarbij werd het gemiddelde van alle deelnemende OESO-landen op 500 geplaatst en de standaarddeviatie van 100. Alle latere cycli van PISA werden op die meetschalen geplaatst (OECD, 2019b).

⁷⁴ De schatting van de plausible values wordt verderop toegelicht bij het onderdeel 'Analytische methode'.

5.3 Analytische methode

Om de psychometrische validiteit van zowel de adaptieve toets leesvaardigheid als de toetsen voor wiskundige en wetenschappelijke geletterdheid na te gaan, worden eerst de itemparameters (moeilijkheidsgraad en discriminatiegraad) in kaart gebracht. Daarnaast werden de itemparameters van de trenditems vergeleken met de itemparameters uit voorgaande PISA-cycli. Die vergelijking werd zowel over alle landen heen als per deelnemend land gemaakt. Items met ongewone parameters werden onderzocht en gecontroleerd op scoringsfouten of andere mogelijke problemen (OECD, 2021).

De IEA hanteert verschillende IRT-modellen voor verschillende soorten items. Bij vragen waarbij het antwoord juist of fout is, wordt een 2PL-model (2-parameter logistisch model) gehanteerd. Verder gebruikt PISA een Generalized Partial Credit Model bij vragen met meer dan twee antwoordmogelijkheden (OECD, 2021). Meer gedetailleerde informatie over de IRT-modellen die de basis vormen voor de analyses is te vinden in het PISA 2018 Technical Report (OECD, 2021).

5.3.1 Schatting van de plausible values

Tijdens de afname leggen leerlingen slechts een beperkt aantal PISA-items af. Wanneer men op basis van dat beperkt aantal items een vaardigheidsscore voor elke leerling zou schatten, leidt dat tot onbetrouwbare schattingen. Daarom maakt PISA net zoals TIMSS en PIRLS geen gebruik van vaardigheidsscores, maar wel van plausible values. Met die methode is het mogelijk om een distributie van de meest waarschijnlijke vaardigheid te schatten voor elke leerling in plaats van één vaardigheidsscore (OECD, 2021).

Bij de schatting van die plausible values maakt PISA gebruik van zowel de antwoorden op de items als de data uit de achtergrondvragenlijsten. Om niet alle variabelen uit de achtergrondvragenlijsten op te nemen bij de schatting, wordt er eerst voor elk deelnemend land een principale componentenanalyse uitgevoerd. Daarbij hanteert PISA het criterium om de principale componenten te weerhouden die nodig zijn om minstens 80% van de totale variantie in de achtergrondvragenlijsten te beschrijven. Op basis van die principale componenten en de antwoorden op de items wordt er een conditionele distributie van de vaardigheid gemodelleerd. Daaruit worden tien plausible values at random getrokken. Die plausible values worden beschouwd als een betrouwbare schatting van de distributie van de meest waarschijnlijke vaardigheid van een leerling. Na de schatting volgt er een lineaire transformatie (zie 3.2 Standaardbepaling) om de geschatte plausible values op de PISA-meetschaal te plaatsen (OECD, 2021).

5.3.2 De gebruikte modellen

PISA wil de deelnemende landen waardevolle informatie bieden met het oog op systeemmonitoring (zie 5.1.1 Doel). In de rapporten vermeldt PISA echter niet welke modellen exact gebruikt worden om die informatie in kaart te brengen (De Meyer et al., 2019).

Wel voorziet PISA **schoolfeedback** voor alle deelnemende scholen. Daarin worden de prestaties van een specifieke school vergeleken met vijf referentiescholen.⁷⁵ Dat zijn scholen die een gelijkaardig

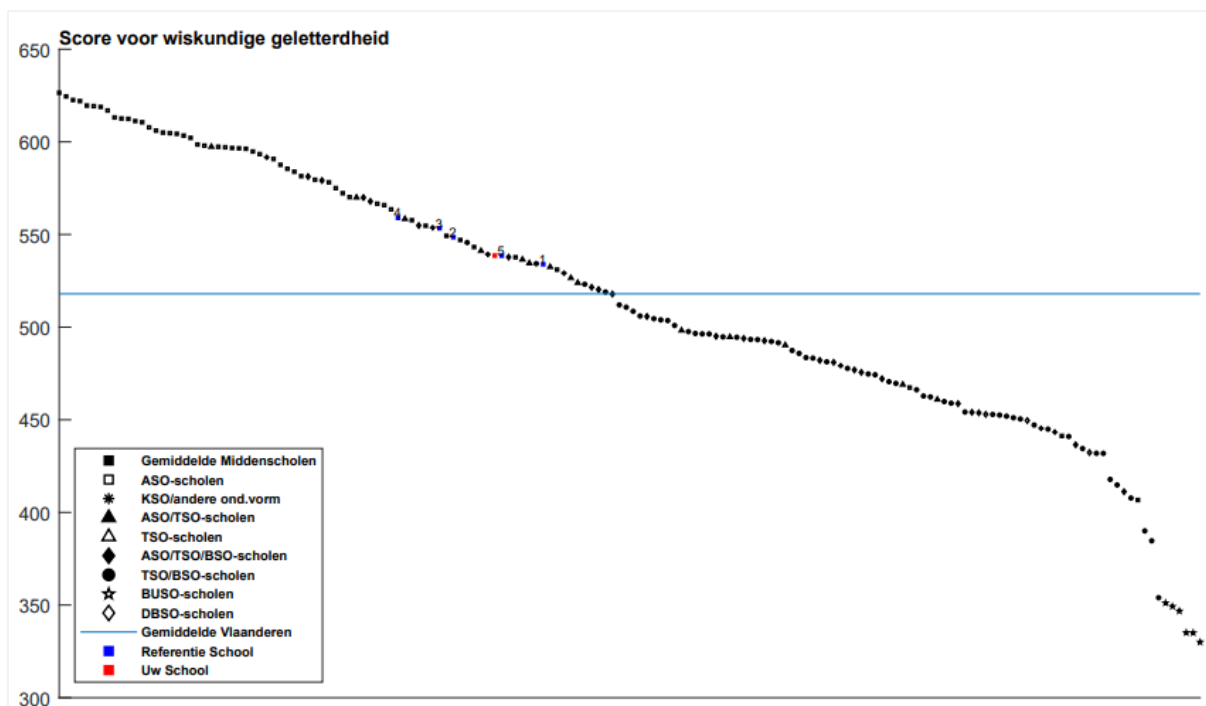
⁷⁵ Bij middensholen (scholen die enkel een eerste graad hebben), DBSO-scholen, scholen met heel specifieke studierichtingen (bijvoorbeeld de studierichting slagerij) of scholen met een specifiek leerlingenpubliek (zoals een school waarbij 75% van de leerlingen anderstalige nieuwkomer is) was het niet evident om referentiescholen te selecteren. Om de vergelijkbaarheid te garanderen werd er – waar mogelijk – vergeleken met minder dan vijf referentiescholen. In de specifieke gevallen waarbij er geen gelijkaardige school gevonden werd, werd er een vergelijking gemaakt referentiescholen die een minder gelijkaardig leerlingenprofiel hadden (Vakgroep Onderwijskunde, 2020).

leerlingenprofiel hebben als de school in kwestie. PISA selecteert die referentiescholen aan de hand van de volgende criteria (Vakgroep Onderwijskunde, 2020):

- De referentieschool behoort tot hetzelfde stratum in de steekproeftrekking.
- De deelnemende leerlingen in de referentieschool volgen les in gelijkaardige studierichtingen.
- De referentieschool heeft een gelijkaardige percentage jongens en meisjes.
- De referentieschool heeft een gelijkaardig percentage leerlingen dat minstens een jaar overgedaan heeft.⁷⁶
- De referentieschool heeft een gelijkaardig percentage leerlingen met een buitenlandse herkomst. Tot die groep behoren zowel leerlingen die zelf niet geboren zijn in België en waarvan de ouders ook niet geboren zijn in België (eerste generatieleerlingen) als leerlingen die zelf wel in België geboren zijn, maar waarvan de ouders dat niet zijn (tweede generatieleerlingen).
- De referentieschool heeft een gelijkaardig percentage leerlingen dat thuis geen Nederlands (of een Vlaams dialect) spreekt.
- De referentieschool heeft een gemiddelde SES die gelijkaardig is aan de gemiddelde SES van de school waarvoor er feedback wordt gegeven.

In het PISA-schoolfeedbackrapport worden voornamelijk ruwe gemiddeldes gerapporteerd. Dat wil zeggen dat er bij die gemiddeldes niet gecontroleerd wordt voor factoren, waarover scholen geen controle hebben, maar die wel een inhoud hebben op de prestaties van leerlingen (Vakgroep Onderwijskunde, 2020). Figuur 40 toont een geanonimiseerd voorbeeld uit het feedbackrapport waarbij de ruwe gemiddeldes van een bepaalde school vergeleken worden met die van de vijf referentiescholen – en bij uitbreiding alle Vlaamse scholen.

⁷⁶ PISA stelt geen vragen over zittenblijven in de leerlingvragenlijst. Daarom gebruiken ze het percentage 15-jarige leerlingen dat les volgt in de eerste graad en deelnam aan PISA op de school in kwestie als indicator voor het percentage zittenblijven (Vakgroep Onderwijskunde, 2020).

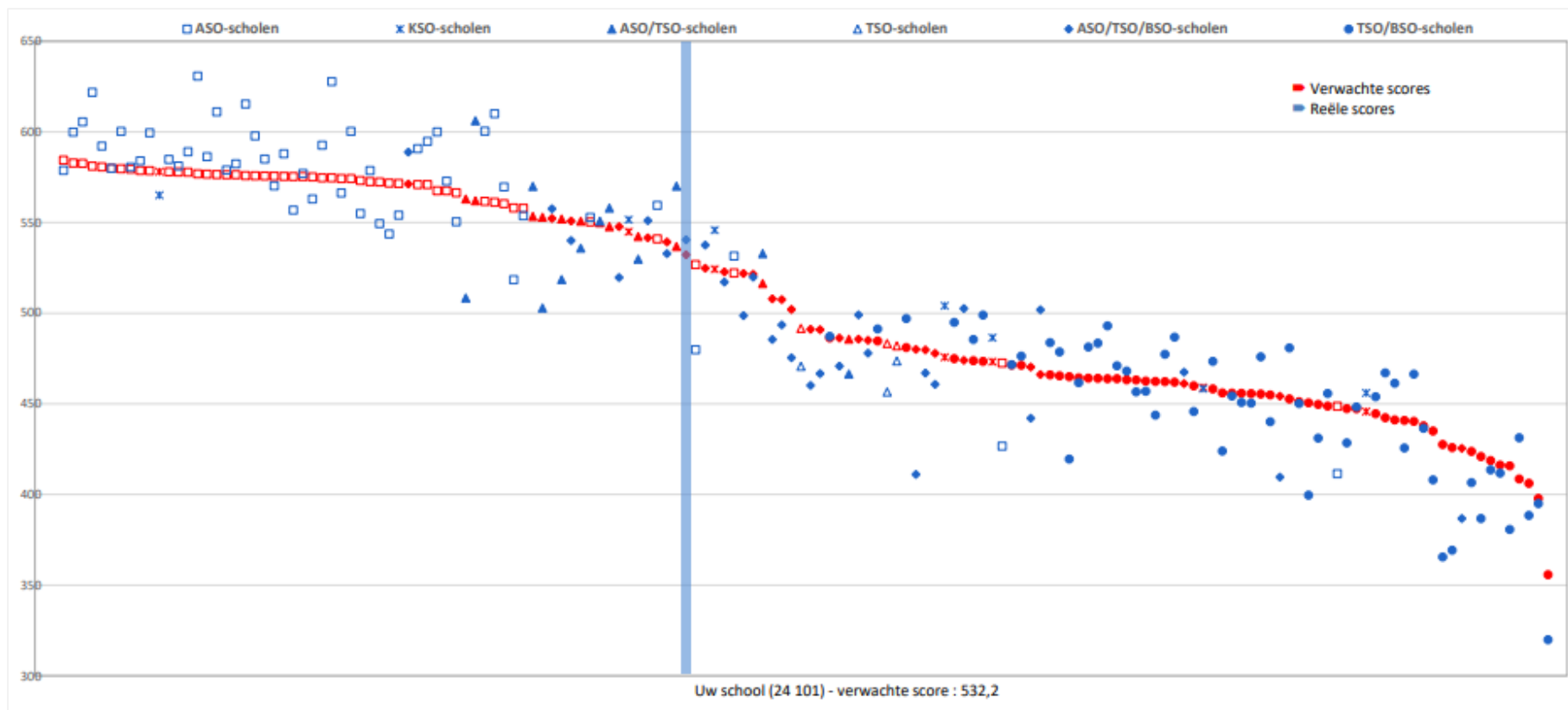


Figuur 40. Geanonimiseerd voorbeeld van schoolfeedback over de gemiddelde prestatie van alle Vlaamse scholen voor wiskundige geletterdheid (Figuur overgenomen uit Vakgroep Onderwijskunde, 2020)

Enkel voor leesvaardigheid (het hoofddomein bij PISA 2018) maakt PISA gebruik van een **model voor schoolfeedback** om de verwachte score te presenteren voor elke school. Bij de berekening van die verwachte scores neemt PISA de volgende kenmerken van de leerlingen op in het model (Vakgroep Onderwijskunde, 2020):

- de onderwijsvorm waarin de leerling les volgt;
- leerlingen die in de eerste graad les volgen;
- leerlingen die in de derde graad les volgen;
- buitenlandse herkomst;
- thuistaal;
- SES;
- geslacht.

Hoewel er nog meer factoren zijn die een grote invloed hebben op de schoolprestaties (bijvoorbeeld studierichting), kiest PISA ervoor om enkel bovenstaande kenmerken op te nemen aangezien die kenmerken een significant effect hebben op de leesvaardigheid (Vakgroep Onderwijskunde, 2020). Figuur 41 toont een geanonimiseerd voorbeeld van de figuur over de verwachte scores voor leesvaardigheid.



Figuur 41. Geanoniseerd voorbeeld van schoolfeedback over de verwachte scores voor leesvaardigheid in uw school en alle Vlaamse scholen (Figuur overgenomen uit Vakgroep Onderwijskunde, 2020)

5.4 Besluitvorming

De plaats van Vlaanderen in de PISA-rankings geeft slechts een beperkte indicatie van de kwaliteit in het Vlaamse onderwijs, aangezien die rankings geen rekening houden met de specifieke context of de onderwijsdoelen die Vlaanderen nastreeft. Toch kan een vergelijking met andere landen waardevol zijn wanneer er rekening gehouden wordt met contextverschillen tussen landen (OECD, 2019b). Die context wordt bij PISA ook bevraagd via de achtergrondvragenlijsten voor leerlingen, ouders en directie (De Meyer et al., 2019; OECD, 2019a). Daarbij is het van belang op te merken dat de perceptie van de bevroegde directeurs niet noodzakelijk een reflectie is van de feitelijke realiteit op de scholen. De variabelen uit de directievragenlijst van PISA zijn daarom minder geschikt om de concrete klas- en schoolpraktijk in kaart te brengen.

Daarnaast wordt PISA ingezet om trends doorheen de tijd op te volgen (zie 5.1.1 Doel). Daarbij maakt PISA gebruik van trenditems en een IRT-gebaseerde meetschaal. Verder is de PISA-steekproef representatief, zijn er verschillende maatregelen om inhoudelijke validiteit te garanderen en wordt de betrouwbaarheid van de PISA-toetsen nagegaan. Toch kunnen er net zoals bij PIRLS en TIMSS vragen gesteld worden bij het gebruik van PISA om trends op te volgen doorheen de tijd. Zo zou het construct over verschillende PISA-cycli heen op eenzelfde manier gemeten moeten worden. Dat is bij PISA niet zo: het conceptueel kader blijft enkel hetzelfde wanneer een domein niet het hoofddomein vormt van de PISA-afname. Bij elke cyclus wordt het conceptueel kader van het hoofddomein en de bijhorende definitie gewijzigd (OECD, 2019b). Bovendien wordt er pas na de afname betekenis gegeven aan de vaardigheden van leerlingen door inhoudelijke beschrijvingen op te stellen voor de PISA-vaardigheidsniveaus. Die opdeling in vaardigheidsniveaus is eerder arbitrair en ook de adequaatheid van die inhoudelijke beschrijvingen wordt door de OESO niet nagegaan (OECD, 2021).

PISA beperkt zich tot samenhang nagaan. Het is op basis van PISA niet mogelijk om causale uitspraken te doen. Dat geldt ook voor de schoolfeedback die de deelnemende scholen ontvangen. PISA kan scholen immers geen verklaringen of oplossingen aanbieden die de prestaties zouden kunnen verhogen. Daarnaast zijn er nog enkele andere beperkingen van de schoolfeedback. Ten eerste is het rapport enkel gebaseerd op de prestaties van 15-jarige leerlingen. De resultaten kunnen daardoor niet veralgemeend worden naar de hele school. Ten tweede vormen de schoolresultaten van PISA geen directe weerspiegeling van de onderwijskwaliteit op schoolniveau. De prestaties van leerlingen op een school worden immers bepaald door verschillende factoren waarop de school niet altijd invloed heeft. Tot slot is het van belang om enige voorzichtigheid aan de dag te leggen bij de interpretatie van de schoolfeedback. Het gaat namelijk om een momentopname. Wanneer men zich bewust is van die beperkingen kan het schoolfeedbackrapport ingezet worden met het oog op interne kwaliteitszorg. Scholen kunnen de informatie uit het schoolfeedbackrapport in de juiste context plaatsen en daarover kritisch reflecteren (De Meyer et al., 2019).

5.5 PISA doorheen de tijd

De OESO organiseerde PISA voor het eerst in 2000. Sindsdien wordt PISA elke drie jaar georganiseerd (De Meyer et al., 2019). Sinds 2009 zijn er heel wat wijzigingen gebeurd. Bij PISA 2009 was het namelijk de eerste keer dat een specifiek domein voor de tweede keer werd afgenomen als hoofddomein. Bijgevolg werd het conceptueel kader voor leesvaardigheid toen grondig herwerkt. Sindsdien wordt er elke cyclus een conceptueel kader aangepast (De Meyer & Warlop, 2010). Bovendien werden bij PISA 2015 de testen voor het eerst via de computer afgenomen. In 2018 werd leesvaardigheid de eerste keer gemeten aan de hand van een adaptieve toets (De Meyer et al., 2019).

6. Doorlichtingen van de onderwijsinspectie

6.1 Design

Het design van de doorlichtingen is gebaseerd op het referentiekader voor onderwijskwaliteit of kortweg het OK-kader. Dat kader weerspiegelt de consensus over wat kwaliteitsvol onderwijs in Vlaanderen op schoolniveau minimaal moet inhouden. Het OK-kader is het resultaat van samenwerking en dialoog tussen verschillende actoren in het onderwijsveld. Concreet bestaat het OK-kader uit 37 kwaliteitsverwachtingen die ingedeeld zijn volgens vier rubrieken: (1) Resultaten en effecten, (2) Kwaliteitsontwikkeling, (3) Ontwikkeling stimuleren en (4) Beleid. Die kwaliteitsverwachtingen staan centraal bij de doorlichtingen van de onderwijsinspectie (Onderwijsinspectie, n.d., 2017a, 2021h).

Sinds 2018 hanteert de onderwijsinspectie een doorlichtingsaanpak 'Inspectie 2.0'. Sindsdien is het concept Inspectie 2.0 verder geëvolueerd en heeft de onderwijsinspectie aanpassingen gemaakt aan de doorlichtingsaanpak. In wat volgt, beschrijven we de aanpak bij Inspectie 2.0 zoals die bij de start in 2018 vormgegeven werd. We gaan dus niet in op de recente aanpassingen en evoluties die na de start van Inspectie 2.0 geïmplementeerd werden.

6.1.1 Doel

De Vlaamse overheid zet in op de doorlichtingen om aan **kwaliteitsbewaking** te doen. Een doorlichting wordt omschreven als "elke vorm van kwaliteitstoezicht door de onderwijsinspectie die resulteert in een verslag en in een advies aan de Vlaamse Regering over de (voorlopige) erkenning van de instelling of van afzonderlijke structuuronderdelen" (*Decreet Betreffende de Kwaliteit van Onderwijs*, 2009, Art. 2, 7°/1). Via die doorlichting waarborgt de onderwijsinspectie dat alle onderwijsinstellingen voldoende kwaliteitsvol onderwijs leveren (Onderwijsinspectie, 2018a). Concreet brengt de onderwijsinspectie bij elke doorlichting de onderwijskwaliteit van een school in kaart op twee niveaus: het systeemniveau (in wat volgt, omschrijven we dat als het beleidsniveau⁷⁷) en het substantief niveau. Op het beleidsniveau staat de volgende onderzoeksvraag centraal: "In welke mate ontwikkelt de onderwijsinstelling haar eigen kwaliteit, met bijzondere aandacht voor de aansturing en de kwaliteitsbewaking van de onderwijsleerpraktijk?" (Onderwijsinspectie, 2018a, p. 6). Daarnaast bekijkt de onderwijsinspectie de kwaliteit van het onderwijs in de klaspraktijk (het substantief niveau). Daarbij wordt er gefocust op het volgende: "In welke mate verstrekt de onderwijsinstelling kwaliteitsvol onderwijs dat tegemoet komt aan de kwaliteitsverwachtingen uit het OK-kader en respecteert ze de regelgeving?" (Onderwijsinspectie, 2018a, p. 7).

Scholen hebben de vrijheid om zelf te bepalen hoe ze de kwaliteit van hun onderwijs systematisch opvolgen en welke pedagogisch-didactische aanpak ze gebruiken. Die schooleigen aanpak en keuzes worden daarom niet beoordeeld tijdens de doorlichting. Wel bekijkt de onderwijsinspectie of een school zijn onderwijskwaliteit systematisch opvolgt en of de gekozen pedagogisch-didactische aanpak tegemoetkomt aan de verwachtingen uit het OK-kader. Op basis daarvan formuleert de onderwijsinspectie een advies over de erkenning van de instelling (Onderwijsinspectie, 2018c, 2021h, 2021i).

De onderwijsinspectie zet niet alleen in op kwaliteitsbewaking vanuit een controlerende rol. Ze wil daarnaast ook een stimulerende rol opnemen. Hoewel scholen in eerste instantie zelf verantwoordelijk zijn voor de realisatie van kwaliteitsvol onderwijs, kunnen de doorlichtingen mogelijk een stimulans

⁷⁷ De onderwijsinspectie onderscheidt het systeem- en substantief niveau. Omwille van conceptuele verwarring met systeemniveau als aggregatieniveau, benoemen we in wat volgt 'het systeemniveau' – zoals gehanteerd bij de doorlichtingen – als 'het beleidsniveau' van een school.

vormen voor de verdere ontwikkeling van de interne kwaliteitszorg van scholen. De onderwijsinspectie wil scholen daarom stimuleren tot reflectie over hun onderwijspraktijk en kwaliteitszorg. Om dat doel te bereiken, reikt de onderwijsinspectie constructieve feedback aan op basis van de doorlichting en gaat ze in dialoog met de school (Onderwijsinspectie, 2018c, 2021f, 2021h). In dat opzicht stelt de onderwijsinspectie naast kwaliteitsbewaking, ook **schoolfeedback** op basis van de resultaten van de doorlichting voorop als doel.

6.1.2 Domein en construct

Om een antwoord te formuleren op de vooropgestelde onderzoeksvragen (zie 6.1.1 Doel) meet de onderwijsinspectie verschillende **proceskenmerken** op zowel het beleids- als het substantief niveau. Op het beleidsniveau wordt de kwaliteitsontwikkeling van scholen gemeten aan de hand van zes proceskenmerken. De zes proceskenmerken zijn afgeleid van de kwaliteitsverwachtingen uit het OK-kader die focussen op systematische kwaliteitsontwikkeling. Die kwaliteitsverwachtingen behoren tot de rubrieken 'Kwaliteitsontwikkeling' en 'Beleid' van het OK-kader (Onderwijsinspectie, 2021f). Tabel 63 toont de proceskenmerken die de onderwijsinspectie meet op niveau van het schoolbeleid.

Tabel 63

Overzicht van de proceskenmerken die gemeten worden op het beleidsniveau

Kwaliteitsontwikkeling
K1. Visie en strategisch beleid
K2. Organisatiebeleid
K3. Onderwijskundig beleid
K4. Cyclische evaluatie van de kwaliteit
K5. Betrouwbare evaluatie van de kwaliteit
K6. Borgen en bijsturen

Op het substantief niveau worden drie aspecten gemeten. Ten eerste brengt de onderwijsinspectie de onderwijsleerpraktijk in kaart aan de hand van acht proceskenmerken. Die proceskenmerken zijn gebaseerd op de kwaliteitsverwachtingen uit de rubrieken 'Ontwikkeling stimuleren' en 'Resultaten en effecten'. Een tweede aspect dat gemeten wordt, zijn een of meerdere kwaliteitsgebieden. De onderwijsinspectie onderscheidt vier kwaliteitsgebieden met elk hun eigen proceskenmerken: (1) rapportering en oriëntering, (2) leerlingenbegeleiding, (3) omgaan met diversiteit en (4) personeelsbeleid en professionalisering. Een kwaliteitsgebied kan gezien worden als een samenhangend geheel van kwaliteitsverwachting uit verschillende rubrieken van het OK-kader. Tot slot wordt de bewoonbaarheid, veiligheid en hygiëne onderzocht aan de hand van vier proceskenmerken (Onderwijsinspectie, 2021f). Tabel 64 geeft een overzicht van de proceskenmerken die gemeten worden op het substantief niveau.

Tabel 64

Overzicht van de proceskenmerken die gemeten worden op substantief niveau

Onderwijsleerpraktijk⁷⁸	
U1. Afstemming van het aanbod op het gevalideerd doelenkader	
U2. Leer- en ontwikkelingsgericht aanbod	
U3. Leer- en leefklimaat	
U4. Materiële leeromgeving	
U5. Feedback	
U6. Leerlingevaluatie	
U7. Leereffecten	
Kwaliteitsgebieden	
Rapportering en oriëntering	
<i>Basisonderwijs</i>	<i>Secundair onderwijs</i>
R1. Rapportering	R1. Rapportering
R2. Studiebekrachtiging en oriëntering	R2. Studiebekrachtiging
	R3. Oriëntering
Leerlingenbegeleiding	
L1. Brede basiszorg	
L2. Passende begeleiding	
L3. Samenwerking met het CLB	
L4. Ondersteuning van de leraren	
Omgaan met diversiteit	
D1. Diversiteitscultuur	
D2. Taalgericht onderwijs	
Personeelsbeleid en professionalisering	
P1. Selectie en aanwerving	
P2. Coaching en beoordeling	
P3. Professionalisering	
P4. Aanvangsbegeleiding	

⁷⁸ Sinds het schooljaar 2021-2022 wordt de onderwijsleerpraktijk ingeschaald aan de hand van O-schalen in plaats van U-schalen. De O-schalen meten de volgende proceskenmerken (Onderwijsinspectie, 2021f):

- O1. Afstemming van het aanbod op het gevalideerd doelenkader
- O2. Afstemming van de begeleiding op de beeldvorming
- O3. Positief en inclusief leef- en leerklimaat
- O4. Leer- en ontwikkelingsgerichte instructie
- O5. Effectieve feedback
- O6. Materiële leef- en leeromgeving
- O7. Leerlingevaluatie
- O8. Leereffecten

Bewoonbaarheid, veiligheid en hygiëne

- BVH1. Planning en uitvoering
 - BVH2. Ondersteuning
 - BVH3. Cyclische en betrouwbare evaluatie
 - BVH4. Borgen en bijsturen
-

Bij bewoonbaarheid, veiligheid en hygiëne zijn er naast de vier proceskenmerken nog negen mogelijke processen die de onderwijsinspectie gebruikt om de bewoonbaarheid, veiligheid en hygiëne te evalueren. Concreet gaat het om de volgende processen: (1) elektriciteit, (2) gebouwen en onderhoud, (3) noodplanning, (4) ongevallen en hulpverlening, (5) onthaal, aankoop en controle arbeidsmiddelen, (6) producten met gevaarlijke eigenschappen, (7) valgevaar en toegankelijkheid, (8) verwarming, ventilatie en verluchting en (9) HACCP en voedselveiligheid. De school kiest een proces en de onderwijsinspectie kiest twee bijkomende processen die de doorlichtingsfocus bepalen. De gekozen processen vormen daarna de basis voor de meting van de proceskenmerken omtrent bewoonbaarheid, veiligheid en hygiëne (Onderwijsinspectie, 2021f).

6.1.3 Populatie

Elke Vlaamse onderwijsinstelling (behalve de instellingen voor hoger onderwijs) en elk centrum voor leerlingenbegeleiding (CLB) wordt minstens één keer elke zes jaar doorgelicht. In tegenstelling tot het peilingsonderzoek en het internationaal vergelijkend onderzoek bestaat de populatie bij de doorlichtingen niet uit leerlingen, maar uit scholen. Concreet omvat de populatie alle scholen voor (buitengewoon) basisonderwijs, (buitengewoon) secundair onderwijs, scholen voor deeltijds kunstonderwijs, centra voor deeltijds onderwijs, centra voor volwassenenonderwijs, centra van Syntra en CLB's (Onderwijsinspectie, 2021f). De instellingen voor hoger onderwijs behoren niet tot de populatie, omdat de externe kwaliteitszorg voor het hoger onderwijs op een andere manier georganiseerd is. Met andere woorden, de onderwijsinspectie is niet bevoegd voor het kwaliteitstoezicht in het hoger onderwijs⁷⁹ (NVAO, 2021; Vlaams Ministerie van Onderwijs en Vorming, n.d.).

Hoewel de populatie bij de doorlichtingen vrij omvangrijk is, beperken we ons in dit rapport voornamelijk tot de doorlichtingen in het basis- en secundair onderwijs. TIMSS, PIRLS, PISA en het Vlaams peilingsonderzoek richten zich immers op die onderwijsniveaus. Het verloop van de doorlichting bij andere onderwijsniveaus en het CLB wordt uitgebreid beschreven op de website van de onderwijsinspectie (Onderwijsinspectie, 2021f).

6.1.4 Steekproef

Elk schooljaar selecteert de onderwijsinspectie een aantal scholen en centra voor de doorlichtingen. Daarbij houdt ze rekening met de voorwaarde dat scholen minstens één keer elke zes jaar doorgelicht moeten worden (zie 6.1.3 Populatie). De steekproef van scholen en centra bij de doorlichting is echter niet noodzakelijk gelijkaardig samengesteld als de Vlaamse populatie en daardoor niet representatief (Onderwijsinspectie, 2021f).

Bovendien worden er heel wat actoren betrokken bij de onderzoeken van de onderwijsinspectie. Concreet gaat het om het beleidsteam, het zorgteam, leerkrachten, ouders, leerlingen, CLB-medewerkers en vertegenwoordigers van de interne dienst voor preventie. Ook op schoolniveau is de

⁷⁹ Er geldt een uitzondering voor de graduaatsopleiding Verpleegkundige (de voormalige HBO5-opleiding Verpleegkunde). Het kwaliteitstoezicht voor die opleiding wordt wel opgenomen door de onderwijsinspectie. De onderwijsinspectie werkt daarvoor samen met enkele externe experts die de Nederlands-Vlaamse Accreditatieorganisatie (NVAO) voorstelt (Onderwijsinspectie, 2021e).

steekproeftrekking van actoren niet noodzakelijk representatief. Zo mag de school bijvoorbeeld zelf bepalen welke leerkrachten er juist deelnemen aan de gesprekken rond een kwaliteitsgebied. Daarnaast selecteert de school tien leerlingen en zes ouders die in gesprek gaan met de onderwijsinspectie. Hoewel de onderwijsinspectie vraagt om een representatieve groep samen te stellen, is er geen garantie dat de door de school samengestelde groepen leerlingen en ouders ook effectief representatief zijn voor de groep leerlingen op school en hun ouders (Onderwijsinspectie, 2021a, 2021b).

6.1.5 Modus

Bij de doorlichtingen staat een triangulatie van dataverzamelmethode en databronnen centraal. De onderwijsinspectie wil namelijk haar bevindingen baseren op de combinatie van minstens drie dataverzamelmethode. Daarenboven verzamelt de onderwijsinspectie zowel kwalitatieve als kwantitatieve data. Figuur 42 toont een schematische voorstelling van de verschillende onderzoeken die bij de doorlichting gehanteerd worden. Die figuur toont dat het onderzoek op beleidsniveau eigenlijk niet los gezien kan worden van het onderzoek op substantief niveau. Zelf maakt de onderwijsinspectie wel een indeling per niveau. De proceswandelings en het onderzoek naar visie, aansturing en systematiek in de kwaliteitszorg beschouwt ze als het onderzoek op beleidsniveau. De andere dataverzamelmethode behoren tot het substantief onderzoek (Onderwijsinspectie, 2018a, 2021f).



Figuur 42. Het doorlichtingsdesign voor het gewoon basis- en secundair onderwijs (Figuur overgenomen uit Onderwijsinspectie, 2021h)

6.1.5.1 Dataverzamelmethode

Voordat de doorlichting plaatsvindt, wil de onderwijsinspectie a.d.h.v. een *deskanalyse* zicht krijgen op de specifieke context, input, resultaten en effecten van de school in kwestie. Het doorlichtingsteam bestudeert en analyseert informatie van de website van de school, de databankgegevens van het Departement Onderwijs en Vorming, het vorige doorlichtingsverslag en andere ontvangen informatie van de school. Op die manier kan de onderwijsinspectie relevante informatie verzamelen met het oog op het beleidsonderzoek, het onderzoek naar het kwaliteitsgebied en de onderwijsleerpraktijk (Onderwijsinspectie, 2018a, 2021f).

De data van het **beleidsonderzoek** verzamelt de onderwijsinspectie via verschillende methoden. Ten eerste organiseert ze een *gesprek met het beleidsteam* van de school over de kwaliteitsontwikkeling om na te gaan hoe de school in kwestie de interne kwaliteitszorg vormgeeft. Tijdens dat gesprek peilt de onderwijsinspectie naar de maatregelen die de school neemt met het oog op interne kwaliteitszorg.

Daarbij wordt aan de school gevraagd om de genomen maatregelen en afspraken te illustreren op basis van een gekozen thema.

Ten tweede worden vier *proceswandelingen*⁸⁰ georganiseerd. Die proceswandelingen hebben als doel om na te gaan in welke mate een school de schoolvisie en het schoolbeleid systematisch en effectief toepast in de dagelijkse schoolpraktijk. In elke doorlichting staan twee vaste thema's centraal bij de proceswandelingen: onderwijsleerpraktijk en woonbaarheid, veiligheid en hygiëne. Verder komt als derde het thema van het gekozen kwaliteitsgebied aan bod in een proceswandeling. De laatste proceswandeling focust op een thema dat de school zelf kiest.

Een derde methode die de onderwijsinspectie hanteert, is *de studie van relevante schooldocumenten*. Ten vierde maakt de onderwijsinspectie gebruik van *observaties*. Ze woont daarvoor lessen en andere activiteiten bij tijdens de doorlichting. Ten vijfde kunnen er eventuele *tussentijdse gesprekken* georganiseerd worden, waarbij het beleidsteam de eigen aanpak en standpunten over de kwaliteitsontwikkeling kan verduidelijken. Tot slot kunnen ook gegevens en vaststellingen uit het substantief onderzoek ingezet worden bij het beleidsonderzoek (Onderwijsinspectie, 2021a, 2021b).

Op het substantief niveau worden er drie aspecten gemeten (zie 6.1.2 Domein en construct). Voor het onderzoek naar de **onderwijsleerpraktijk**⁸¹ wordt er eerst een *startgesprek* georganiseerd met de leerkrachten. Het doel van dat gesprek is om zicht te krijgen op hoe die leerkrachten de onderwijsleerpraktijk voor een specifiek leergebied (basisonderwijs) of vak (secundair onderwijs) vormgeven rekening houdend met de context en visie van de school. Via dat gesprek gaat de onderwijsinspectie ook na of de schooleigen aanpak voor kwaliteitszorg (zoals beschreven op beleidsniveau) toegepast wordt in de klas.

Ten tweede verzamelt de onderwijsinspectie informatie via een *documentenstudie*. Een derde dataverzamelmethode omvat de *observaties* tijdens de lessen en andere activiteiten. Ten vierde kunnen er *tussentijdse gesprekken* georganiseerd worden om de inzichten te verfijnen na de lesobservaties. In die gesprekken kunnen leerkrachten bepaalde elementen uit de onderwijsleerpraktijk verduidelijken. Tot slot maakt de onderwijsinspectie ook gebruik van de informatie uit de gesprekken met ouders en leerlingen (Onderwijsinspectie, 2021a, 2021b).

Daarnaast wordt er één van de vier **kwaliteitsgebieden** (zie 6.1.2 Domein en construct) in kaart gebracht aan de hand van diverse onderzoeksmethoden. Ten eerste wordt de aanpak voor het geselecteerde kwaliteitsgebied toegelicht en geïllustreerd in een *gesprek* met leden van het schoolteam. Daarbij bevaart de onderwijsinspectie eventuele schoolafspraken die de school inzet om zelf het kwaliteitsgebied in kwestie te evalueren. Ten tweede verzamelt de onderwijsinspectie informatie via een *documentenstudie*. Een derde onderzoeksmethode omvat de *observatie* van enkele lessen en activiteiten indien dat relevant is voor het kwaliteitsgebied in de focus. Ten vierde maakt de onderwijsinspectie ook gebruik van de informatie uit de *gesprekken met ouders en leerlingen*. Met die

⁸⁰ Bij die proceswandelingen worden telkens de volgende vier elementen nagegaan binnen een bepaald thema:

1. Op welke manier de school resultaat- en doelgericht werkt en of de school voor zichzelf bepaalt wat de juiste dingen zijn.
2. Of de school die juiste dingen en schooleigen doelen met zorg implementeert.
3. Of de school op betrouwbare en systematische wijze nagaat of die doelen in de praktijk gebracht worden en of de juiste dingen op een kwaliteitsvolle manier gedaan worden.
4. Of de school haar eigen kwaliteit ontwikkelt en borgt (Onderwijsinspectie, 2018a).

⁸¹ De onderwijsinspectie vertrekt bij het onderzoek naar de onderwijsleerpraktijk vanuit de leerplannen waarmee de leerkrachten dagelijks aan de slag gaan. Bij de beoordeling van de onderwijsleerpraktijk wordt er enkel rekening gehouden met die leerplandoelstellingen die het bereiken of nastreven van de onderwijsdoelen beogen en niet met de leerplandoelstellingen die verder gaan dan de onderwijsdoelen (Onderwijsinspectie, 2018a).

gesprekken wil de onderwijsinspectie inzicht krijgen in het welbevinden van de leerlingen en hun mening over het schoolbeleid enerzijds en in de tevredenheid van ouders en hun mening over het schoolbeleid anderzijds. Tot slot kunnen gegevens en vaststellingen uit het onderzoek naar het schoolbeleid aangewend worden als dat relevant is voor het onderzoek naar het geselecteerde kwaliteitsgebied (Onderwijsinspectie, 2021a, 2021b).

Het laatste onderzoek richt zich op de **bewoonbaarheid, veiligheid en hygiëne**. Eerst wordt er een *gesprek* georganiseerd met het beleidsteam en de interne dienst voor preventie van de school. Via dat gesprek wil de onderwijsinspectie zicht krijgen op de schooleigen kwaliteitszorg voor de drie processen in de doorlichtingsfocus. Het schoolteam kan tijdens dat gesprek zijn aanpak toelichten aan de hand van ondersteunde schooldocumenten. Ook vraagt de onderwijsinspectie naar eventuele maatregelen op school rond de evaluatie van de bewoonbaarheid, veiligheid en hygiëne. Daarnaast worden relevante *documenten* bestudeerd. Verder doet de onderwijsinspectie een *bepaalde toetsing* van de kwaliteit aan de hand van observaties in de gebouwen en klaslokalen. Tot slot kunnen andere gegevens en vaststellingen gebruikt worden wanneer dat relevant is (Onderwijsinspectie, 2021a, 2021b). Tabel 65 geeft een overzicht van de dataverzamelmethode op het beleids- en substantief niveau.

Tabel 65

Overzicht dataverzamelmethode op beleids- en substantief niveau

Beleidsniveau		Substantief niveau	
Kwaliteitsontwikkeling	Onderwijsleerpraktijk	Kwaliteitsgebied	BVH
<ul style="list-style-type: none"> • Gesprekken met schoolteam • Proceswandelingen • Documentenstudie • Observaties • Vaststellingen uit andere onderzoeken 	<ul style="list-style-type: none"> • Gesprekken met schoolteam • Documentenstudie • Observaties • Gesprek met leerlingen • Gesprek met ouders 	<ul style="list-style-type: none"> • Gesprekken met schoolteam • Documentenstudie • Observaties • Gesprek met leerlingen • Gesprek met ouders • Vaststellingen uit andere onderzoeken 	<ul style="list-style-type: none"> • Gesprekken met schoolteam • Documentenstudie • Observaties • Vaststellingen uit andere onderzoeken

Noot. BVH = Bewoonbaarheid, veiligheid en hygiëne

6.1.5.2 Afnameprocedure bij de doorlichtingen

Eerst wordt de **doorlichting aangekondigd** aan de school en het schoolbestuur. Dat gebeurt ongeveer zes weken voor de feitelijke doorlichting. De onderwijsinspectie bepaalt in overleg met de school de doorlichtingsfocus en de weekplanning. Ook worden er praktische afspraken gemaakt. Ondertussen bereidt het doorlichtingsteam zich voor door de website van de school, de databankgegevens van het Departement Onderwijs en Vorming en de ontvangen informatie te bekijken. Op die manieren krijgt het team zicht op de context en de inputgegevens van de school (Onderwijsinspectie, 2021f).

Tijdens de feitelijke doorlichting volgt de **dataverzameling** via verschillende methoden (zie 6.1.5.1 Dataverzamelmethode). Na een kennismaking met het schoolteam en eventueel een rondleiding door leerlingen (enkel in het basisonderwijs) gaat de onderwijsinspectie aan de slag met observaties en documentenstudie. Op datzelfde moment starten ook de gesprekken in kader van het beleidsonderzoek en substantief onderzoek. Zoals eerder aangegeven, gaat het om de volgende gesprekken (Onderwijsinspectie, 2021f):

- startgesprek met het beleidsteam over de kwaliteitszorg;
- gesprek over de kwaliteit van hun onderwijsleerpraktijk;
- gesprek over een kwaliteitsgebied met de actoren die daarbij betrokken zijn;

- gesprek met het beleidsteam en de interne dienst voor preventie over de bewoonbaarheid, veiligheid en hygiëne;
- gesprek met leerlingen om inzicht te krijgen in de mening van de leerlingen over het schoolbeleid en hun welbevinden;
- gesprek met ouders om zicht te krijgen op de mening van de ouders over het schoolbeleid en hun tevredenheid.

Voor bovenstaande gesprekken zijn er uitgebreide instrumentenbundels en/of leidraden beschikbaar. De vragen die gesteld worden in de gesprekken zijn open en breed geformuleerd zodat de gesprekpartners kansen krijgen tot eigen inbreng. Aan het begin van elk gesprek schetst het doorlichtingsteam kort de context, doelen, rollen, aanpak en eventuele interactie (Onderwijsinspectie, 2018a).

Na die gesprekken organiseert de onderwijsinspectie een **teamoverleg** waarin de leden van het doorlichtingsteam de vaststellingen bespreken op basis van de instrumentenbundels. Bij dat overleg wordt er gestreefd naar een consensus bij de inschaling op de ontwikkelingsschalen. Daarnaast is er tijdens het teamoverleg ook ruimte om de reflectie- en synthesegesprekken voor te bereiden (Onderwijsinspectie, 2018a).

In een volgende stap worden er vier **reflectiegesprekken** georganiseerd. Eén gesprek over het beleidsniveau en drie gesprekken op substantief niveau over respectievelijk de onderwijsleerpraktijk, het kwaliteitsgebied en de bewoonbaarheid, veiligheid en hygiëne. Het doel van die gesprekken is de resultaten van de doorlichting terugkoppelen en samen met de school daarover reflecteren. Zowel sterke punten als ontwikkelingskansen worden besproken. Er kan ook ingegaan worden op mogelijke doelen en acties voor de toekomst. Tijdens die gesprekken verduidelijkt de onderwijsinspectie in welke mate de school voldoet aan de kwaliteitsverwachtingen uit het OK-kader (Onderwijsinspectie, 2021a, 2021b, 2021f).

Op het einde van de doorlichting vindt er een **synthesegesprek** plaats. Daarin worden de definitieve resultaten van de onderzoeken op beleids- en substantief niveau gedeeld met de school. Bovendien deelt het doorlichtingsteam het advies en de bijhorende motivatie mee aan de school. Tijdens dat gesprek kunnen de onderwijsinspectie en de school ook mogelijke doelen en acties bespreken die gericht zijn op verdere kwaliteitsontwikkeling (Onderwijsinspectie, 2021f).

Vervolgens worden de resultaten van de doorlichting en het advies over de verdere erkenning opgenomen in het **doorlichtingsverslag**. Om de kwaliteit van de feedback in het doorlichtingsverslag te garanderen, geeft elk lid van het doorlichtingsteam feedback op de deelverslagen van een ander teamlid. Die feedback wordt gebruikt om het doorlichtingsverslag te versterken. Uiterlijk enkele dagen na het synthesegesprek ontvangt de school het doorlichtingsverslag. Kort na ontvangst kan de school een bespreking van het verslag aanvragen of aanvullingen bezorgen aan de onderwijsinspectie die integraal worden toegevoegd aan het doorlichtingsverslag. De school moet de leerlingen en hun ouders informeren over de mogelijkheid om het doorlichtingsverslag in te kijken. Daarnaast dient het verslag integraal besproken te worden op de personeelsvergadering en de schoolraad. Na afronding van de volledige doorlichtingsprocedure ontvangt de school een online vragenlijst om zo feedback te bezorgen aan de onderwijsinspectie. Tot slot verschijnt het doorlichtingsverslag online⁸² (Onderwijsinspectie, 2021f).

⁸² Alle doorlichtingsverslagen van de onderwijsinspectie zijn openlijk toegankelijk en terug te vinden op de volgende website: www.onderwijs.vlaanderen.be/doorlichtingsverslagen/

6.2 Standaardbepaling

Voor elk proceskenmerk opgenomen in het conceptueel kader (zie 6.1.2 Domein en construct) is er een ontwikkelingsschaal opgesteld. Die ontwikkelingsschalen zijn steeds opgebouwd uit vier ontwikkelingsniveaus (Onderwijsinspectie, 2021f):

- Beneden de verwachting: dat niveau wijst erop dat meerdere fundamentele elementen vatbaar zijn voor verbetering.
- Benadert de verwachting: dat niveau betekent dat er zowel verbeterpunten als sterke punten zijn. Het geheel voldoet echter niet aan de verwachtingen uit het OK-kader.
- Volgens de verwachting: dat niveau houdt in dat er verschillende sterke punten zijn en geen fundamentele verbeterpunten. Het geheel voldoet aan de verwachtingen uit het OK-kader.
- Overstijgt de verwachting: dat niveau wijst op verschillende sterke punten, waaronder enkele voorbeelden van goede praktijk.

Aan elk niveau van een ontwikkelingsschaal wordt er een inhoudelijke omschrijving gekoppeld op basis van de kwaliteitsverwachtingen die opgenomen zijn in het OK-kader.⁸³ Die inhoudelijke beschrijvingen worden per ontwikkelingsschaal en per ontwikkelingsniveau geconcretiseerd aan de hand van kritische kenmerken. Die kenmerken zijn uitgedrukt in concreet en rechtstreeks waarneembaar gedrag en vormen het vertrekpunt voor de beoordeling (Onderwijsinspectie, 2018a). De ontwikkelingsschalen met bijhorende omschrijving zijn per onderwijsniveau beschikbaar op de website van de onderwijsinspectie. De kritische kenmerken⁸⁴ zijn daarin niet opgenomen (Onderwijsinspectie, 2021f).

Zoals eerder aangegeven, beoordeelt de onderwijsinspectie de onderwijskwaliteit van een school aan de hand van de kritische kenmerken bij de ontwikkelingsschalen. Door die concrete kritische kenmerken te vergelijken met de verzamelde data (zie 6.1.5.1 Dataverzamelmethode) bepaalt de onderwijsinspectie voor elke doorgelichte school het ontwikkelingsniveau voor een specifieke ontwikkelingsschaal. Normaal gezien bereikt een school een specifiek niveau, wanneer de school voldoet aan alle kritische kenmerken van dat ontwikkelingsniveau. Het is echter mogelijk dat enkele kritische kenmerken relevanter zijn dan andere gegeven de specifieke context van de school. In dat geval houdt de onderwijsinspectie daarmee rekening bij de inschaling op de ontwikkelingsschaal.

De onderwijsinspectie kiest voor het gebruik van ontwikkelingsschalen met bijhorende kritische kenmerken omwille van verschillende redenen. Ten eerste bevorderen ze een objectieve en betrouwbare beoordeling. Ten tweede ondersteunt het gebruik van ontwikkelingsschalen gelijkgerichtheid. Tot slot kunnen de resultaten van de doorlichting op een gestandaardiseerde manier geregistreerd worden (Onderwijsinspectie, 2018a).

6.3 Analytische methode

6.3.1 Van inschaling op de ontwikkelingsschalen naar het advies voor de school

Het onderzoek op beleidsniveau en de inschalingen op de gerelateerde ontwikkelingsschalen hebben geen impact op het advies dat een school ontvangt. Dat advies wordt enkel gebaseerd op de inschaling op de ontwikkelingsschalen van het substantief onderzoek (Onderwijsinspectie, 2018a). Met het oog op standaardisatie zijn er eenduidige criteria vastgelegd om te bepalen welke combinatie van inschalingen leidt tot een tekort. Zo leiden minstens drie inschalingen van 'benadert de verwachting' bij ontwikkelingsschalen L1 tot L4 bij leerlingenbegeleiding bijvoorbeeld automatisch tot een tekort voor leerlingenbegeleiding. Daarnaast zijn er ook criteria opgesteld voor de bepaling van

⁸³ Ter informatie, binnen de doorlichtingsaanpak van Inspectie 2.0 kan de onderwijsinspectie de inhoudelijke omschrijving van de ontwikkelingsschalen bijsturen en aanpassen.

⁸⁴ De onderwijsinspectie kiest ervoor om de kritische kenmerken niet vrij te geven.

ontwikkelkansen. Een ontwikkelkans bij leerlingenbegeleiding wordt bijvoorbeeld gerapporteerd wanneer een school bij minstens één van ontwikkelingsschalen L1 tot L4 'beneden de verwachting' of 'benadert de verwachting' scoort. Tot slot beschouwt de onderwijsinspectie alle onderzoeken die niet resulteren in ontwikkelkansen of tekorten als sterke punten (Onderwijsinspectie, 2018a).

De onderwijsinspectie onderscheidt twee varianten bij het gunstig advies: het gunstig advies zonder meer en het gunstig advies met verplichting om te werken aan tekorten. Bij dat laatste advies wordt van de school verwacht dat ze hun tekorten aan te pakken. Toch volgt de onderwijsinspectie de school niet op. Ze hoopt namelijk dat de school focust op de versterking van de interne kwaliteitszorg in zijn totaliteit in plaats van zich louter te richten op het wegwerken van tekorten. Wel kunnen de vastgestelde tekorten in de focus staan bij een volgende doorlichting.

Ook bij het ongunstig advies zijn er twee varianten. Ten eerste is er het ongunstig advies met mogelijkheid om te verzoeken de procedure tot intrekking van de erkenning niet op te starten mits externe begeleiding (Advies 2a)⁸⁵. Bij dat advies moet het schoolbestuur het engagement aangaan om te werken aan de tekorten met behulp van externe begeleiding. In dat geval volgt er opnieuw een doorlichting van de school, waarna de onderwijsinspectie een nieuw advies formuleert. Ten tweede is er het ongunstig advies zonder mogelijkheid om te verzoeken de procedure tot intrekking van de erkenning niet te starten (Advies 2b). Een school kan daartegen beroep indienen, waarna er een onderzoek start naar de argumentatie van de school om beroep in te dienen (Onderwijsinspectie, 2021f, 2021g).

Een advies 2a volgt wanneer (1) een tekort blijkt voor minstens 40 procent van de onderzoeken m.b.t. de erkenningsvoorwaarden⁸⁶, (2) over meerdere doorlichtingen heen herhaaldelijk een tekort blijkt voor hetzelfde onderzoek of (3) er acuut gevaar is voor de fysieke veiligheid van de leerlingen of het schoolteam. Bij dat advies worden in het doorlichtingsverslag alle tekorten, ontwikkelkansen en sterke punten vermeld. Een advies 2b reikt de onderwijsinspectie enkel uit in uitzonderlijke gevallen, waarbij er zeer ernstige tekorten zijn of er acuut gevaar is voor de veiligheid van de leerlingen of het schoolteam. Dat gebeurt steeds in overleg met de inspecteur-generaal en de coördinerend inspecteur. Bij dat advies worden enkel de tekorten m.b.t. de erkenningsvoorwaarden vermeld. De onderwijsinspectie vermeld geen ontwikkelkansen of sterke punten. In alle andere gevallen dan hierboven omschreven, ontvangt een school een gunstig advies waarbij eventuele tekorten, ontwikkelkansen en sterke punten vermeld worden (Onderwijsinspectie, 2018a).

6.3.2 Analyses in de Onderwijsspiegel

Elk jaar rapporteert de onderwijsinspectie in de Onderwijsspiegel over de onderwijsinstellingen die in het voorgaande schooljaar doorgelicht werden. Daarin wordt per onderwijsniveau gerapporteerd over de resultaten van het beleids- en substantieve onderzoek. Er worden daarbij zowel kwantitatieve als kwalitatieve analyses gepresenteerd. Concreet bestaat het kwantitatieve luik uit descriptieve analyses van de inschalingen op de ontwikkelingsschalen en de soorten uitgereikte adviezen. Het kwalitatieve luik daarentegen omvat een globale analyse van de doorlichtingsverslagen en brengt op die manier in kaart welke vaststellingen vaak geobserveerd werden door de onderwijsinspectie. Op basis van de resultaten van die analyses stelt de onderwijsinspectie een kwaliteitsprofiel op. Het kwaliteitsprofiel toont op systeemniveau per onderwijsniveau waar de sterktes en uitdagingen liggen voor de Vlaamse

⁸⁵ Rekening houdend met de specifieke context van een school kan het doorlichtingsteam waar nodig een ongunstig advies 2a delibereren naar een gunstig advies (Onderwijsinspectie, 2018a).

⁸⁶ Meer specifiek zijn dat de onderzoeken die focussen op de kwaliteitsontwikkeling, onderwijsleerpraktijk, leerlingenbegeleiding, BVH en regelgeving (Onderwijsinspectie, 2018a).

scholen (Onderwijsinspectie, 2021h). Figuur 43 toont een voorbeeld van een kwaliteitsprofiel voor het gewoon basisonderwijs.

	UITDAGING	KANS	KRACHT
Kwaliteitsontwikkeling	<ul style="list-style-type: none"> • Onderwijskundig beleid • Betrouwbare evaluatie van de kwaliteit 	<ul style="list-style-type: none"> • Systematische evaluatie van de kwaliteit • Borgen en bijsturen 	<ul style="list-style-type: none"> • Visie en strategisch beleid • Organisatiebeleid
Taalgericht onderwijs		<ul style="list-style-type: none"> • Taalgericht Onderwijs 	
Rapportering en oriëntering		<ul style="list-style-type: none"> • Rapportering 	<ul style="list-style-type: none"> • Studiebekrachtiging en oriëntering
Leerlingenbegeleiding		<ul style="list-style-type: none"> • Brede basiszorg • Ondersteuning van leraren 	<ul style="list-style-type: none"> • Passende begeleiding • Samenwerking met CLB
Onderwijsleerpraktijk		<ul style="list-style-type: none"> • Leer- en ontwikkelingsgericht aanbod • Feedback • Kleuter- Of leerlingen-evaluatie • Leereffecten 	<ul style="list-style-type: none"> • Afstemming van het aanbod op het gevalideerd doelenkader • Leer- en leefklimaat • Materiële leeromgeving

Figuur 43. Kwaliteitsprofiel van het gewoon basisonderwijs op basis van de doorlichtingen tijdens schooljaar 2019-2020 (Figuur overgenomen uit Onderwijsinspectie, 2021h)

6.4 Besluitvorming

De doorlichtingen hebben als het hoofddoel de onderwijskwaliteit van een onderwijsinstelling in kaart brengen om op basis daarvan een advies te formuleren over de erkenning van die onderwijsinstelling. Bij de doorlichtingen stelt de onderwijsinspectie een triangulatie van dataverzamelmethode en informatiebronnen voorop, wat bijdraagt aan de validiteit en betrouwbaarheid van de resultaten (Onderwijsinspectie, 2021a, 2021b). Met oog op standaardisatie wordt bij elke doorlichting dezelfde afnameprocedure gevolgd en gebruikt de onderwijsinspectie per onderwijsniveau vaste ontwikkelingsschalen met bijhorende kritische kenmerken (Onderwijsinspectie, 2021h). Daarnaast gaat de onderwijsinspectie in gesprek met verschillende actoren. Er is echter geen garantie dat de betrokken actoren representatief zijn voor alle leerlingen, cursisten, leerkrachten en ouders die verbonden zijn aan de onderwijsinstelling. Wel vormen die gesprekken – omwille van de triangulatie – slechts één informatiebron waarop de onderwijsinspectie beroep doet om tot een oordeel te komen (Onderwijsinspectie, 2021a, 2021b).

De doorlichtingen van de onderwijsinspectie zijn echter minder geschikt om valide en betrouwbare uitspraken te doen over de onderwijskwaliteit op systeemniveau. Doordat de jaarlijkse steekproef van scholen en centra niet noodzakelijk representatief is, is het niet aangewezen om de resultaten van de

doorlichtingen te veralgemenen naar de Vlaamse populatie. De gepresenteerde resultaten in de Onderwijsspiegels zijn daardoor louter illustratief.

Zoals eerder aangegeven, wil de onderwijsinspectie elke school constructieve feedback aanreiken op basis van de resultaten van de doorlichting en daarover in gesprek gaan met de school (zie 6.1.1 Doel). Concreet bespreekt de onderwijsinspectie met de school de sterke punten, ontwikkelkansen en eventuele tekorten die vastgesteld zijn tijdens de doorlichting. Ook kunnen samen mogelijke doelen en acties voor de toekomst besproken worden (Onderwijsinspectie, 2018c, 2021f, 2021h). Hoewel de feedback van de onderwijsinspectie waardevol kan zijn voor scholen om hun interne kwaliteitszorg te versterken, is er geen garantie dat scholen ook effectief aan de slag gaan met die feedback.

6.5 Doorlichtingen doorheen de tijd

Het decreet betreffende onderwijsinspectie en pedagogische begeleidingsdiensten uit 1991 vormde de wettelijke basis voor de doorlichtingen tot er in 2009 een nieuw decreet werd ingevoerd: het decreet betreffende de kwaliteit van het onderwijs (Rekenhof, 2011). Naar aanleiding van dat nieuwe decreet werd in 2009 de doorlichtingsmethode gewijzigd. Bij die nieuwe manier van doorlichting, de zogenaamde 'doorlichtingsronde 3', staat gedifferentieerd doorlichten centraal. Dat betekent concreet dat de doorlichting op basis van het specifieke profiel van een school gebeurt. Tijdens een vooronderzoek worden alle beschikbare gegevens verzameld om zo het profiel van de school in kaart te brengen. Dat profiel vormt op zijn beurt de basis voor de bepaling van de doorlichtingsfocus. De doorlichting zelf gebeurt aan de hand van het CIPO-referentiekader en resulteert in een verslag met advies aan de Vlaamse overheid (Onderwijsinspectie, 2010, 2011; Rekenhof, 2011).

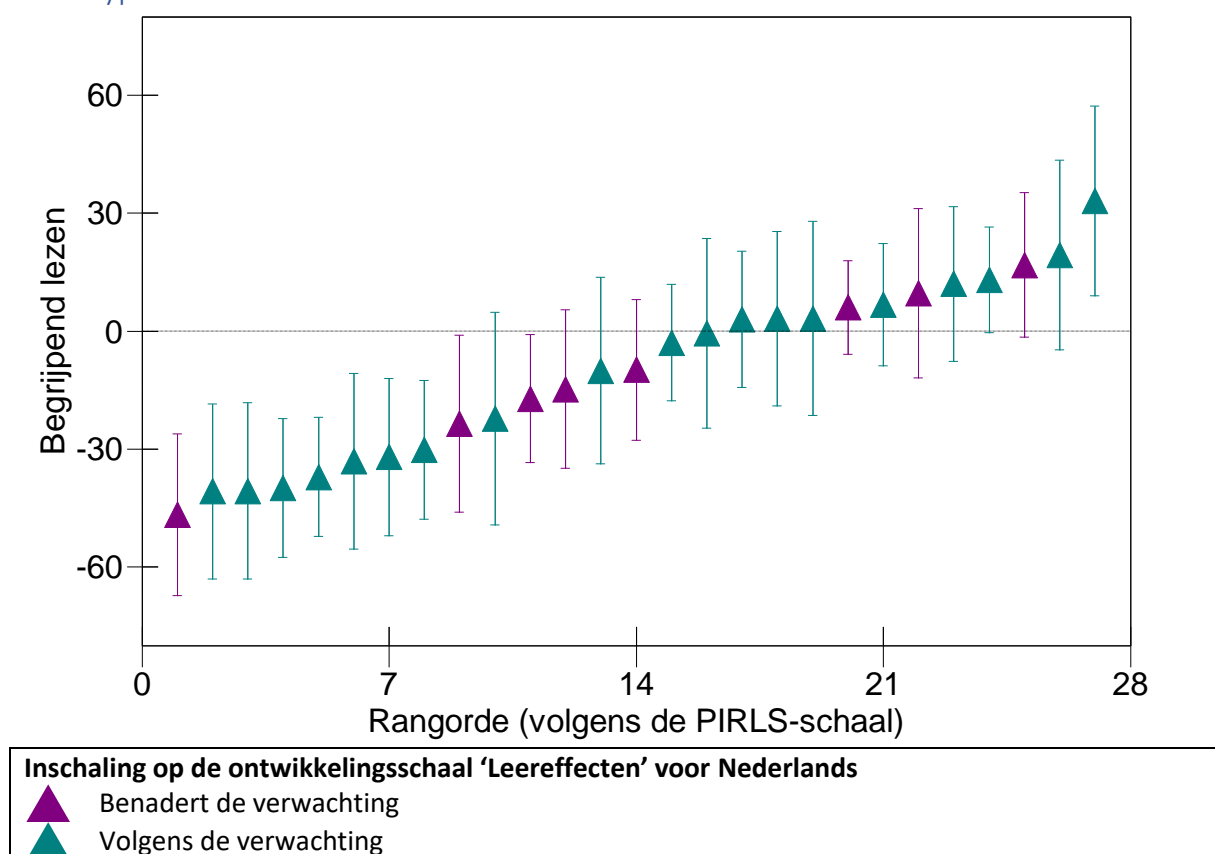
In 2018 kwam er een nieuwe doorlichtingsaanpak 'Inspectie 2.0' onder meer op basis van de feedback uit doorlichtingsronde 3. De introductie van Inspectie 2.0 ging samen met nieuwe onderzoeksvragen voor de doorlichtingen, een nieuw doorlichtingsdesign en een nieuw uitgangspunt, namelijk het OK-kader. De aanpak bij Inspectie 2.0 is degene die hierboven (zie 6 Doorlichtingen van de onderwijsinspectie) beschreven wordt (Onderwijsinspectie, 2018b, 2018c).

Bijlage 2 – Resultaten van studie 2

Vergelijking 1: Internationaal vergelijkend onderzoek en doorlichtingen

1. PIRLS 2016 en de doorlichting: 27 scholen

1.1 Type 0-schooleffect



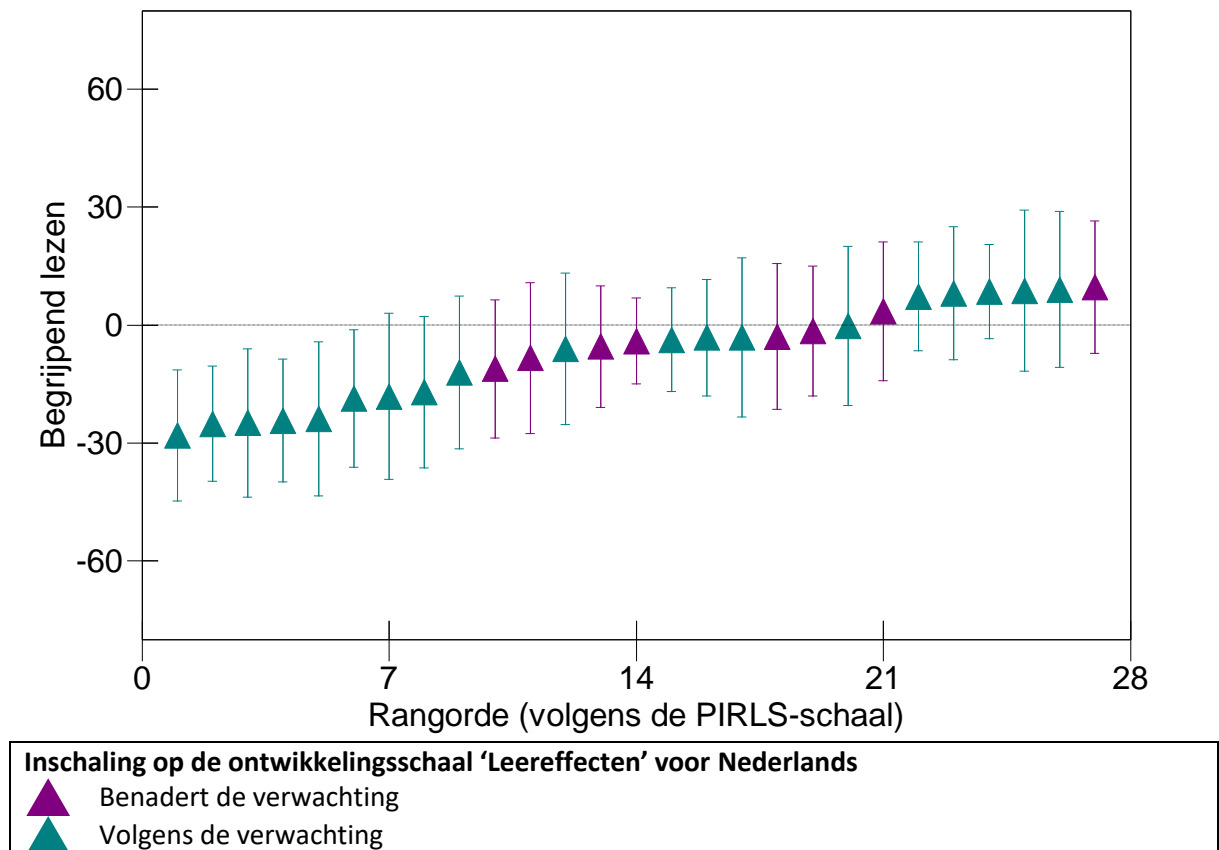
Figuur 44. Rangschikking van scholen volgens hun type 0-schooleffect op de PIRLS-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

Tabel 66

Kruistabel van scholen volgens hun type 0-schooleffect op de PIRLS-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

	Ontwikkelingsschaal leereffecten Nederlands (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
PIRLS					
Ondergemiddeld	-	3 (11,11%)	7 (25,93%)	-	10 (37,04%)
Gemiddeld	-	5 (18,52%)	11 (40,74%)	-	16 (59,26%)
Bovengemiddeld	-	-	1 (3,70%)	-	1 (3,70%)
Totaal	-	8 (29,63%)	19 (70,37%)	-	27 (100%)

1.2 Type A-schooleffect



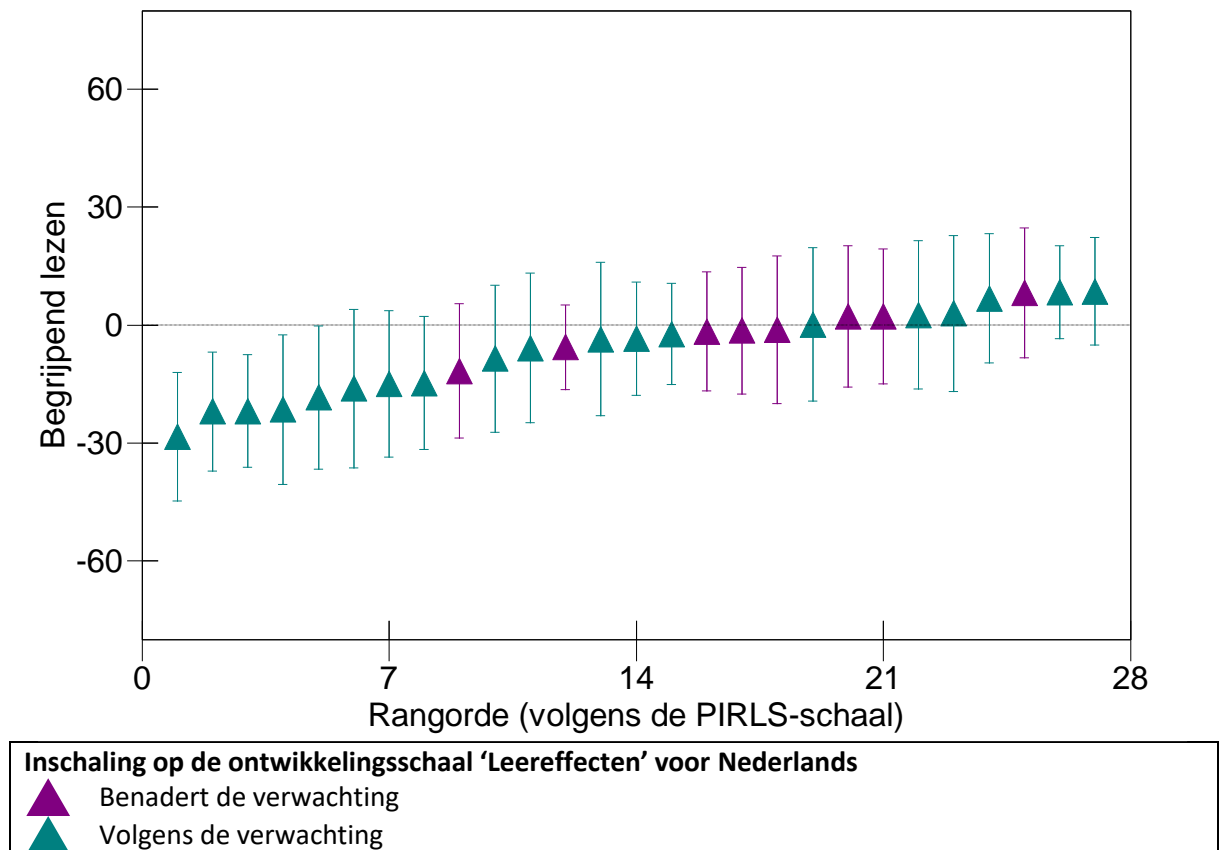
Figuur 45. Rangschikking van scholen volgens hun type A-schooleffect op de PIRLS-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

Tabel 67

Kruistabel van scholen volgens hun type A-schooleffect op de PIRLS-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

	Ontwikkelingsschaal leereffecten Nederlands (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
PIRLS					
Ondergemiddeld	-	-	6 (22,22%)	-	6 (22,22%)
Gemiddeld	-	8 (29,63%)	13 (48,15%)	-	21 (77,78%)
Bovengemiddeld	-	-	-	-	-
Totaal	-	8 (29,63%)	19 (70,37%)	-	27 (100%)

1.3 Type B-schooleffect



Figuur 46. Rangschikking van scholen volgens hun type B-schooleffect op de PIRLS-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

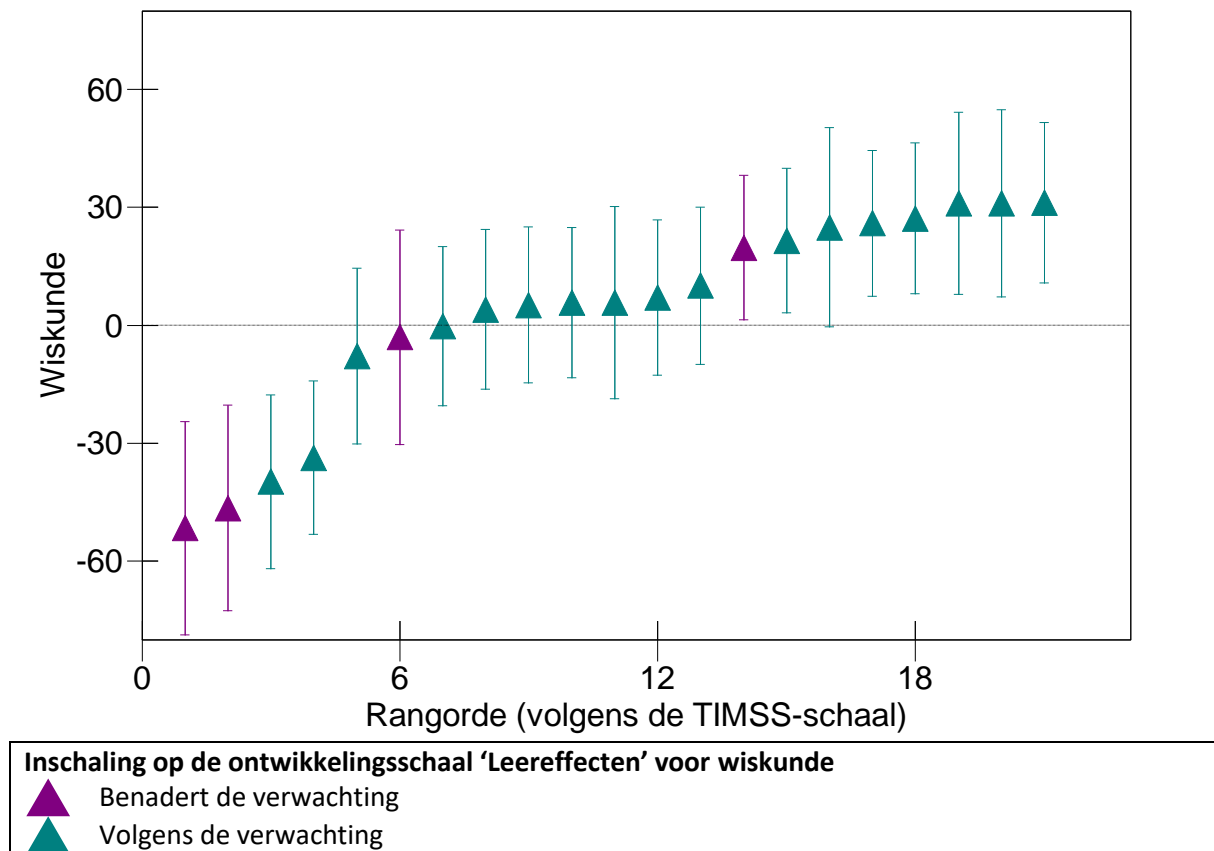
Tabel 68

Kruistabel van scholen volgens hun type B-schooleffect op de PIRLS-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

	Ontwikkelingsschaal leereffecten Nederlands (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
PIRLS					
Ondergemiddeld	-	-	5 (18,52%)	-	5 (18,52%)
Gemiddeld	-	8 (29,63%)	14 (51,85%)	-	22 (81,48%)
Bovengemiddeld	-	-	-	-	-
Totaal	-	8 (29,63%)	19 (70,37%)	-	27 (100%)

2. TIMSS 2019 en de doorlichting: 21 scholen

2.1 Type 0-schooleffect



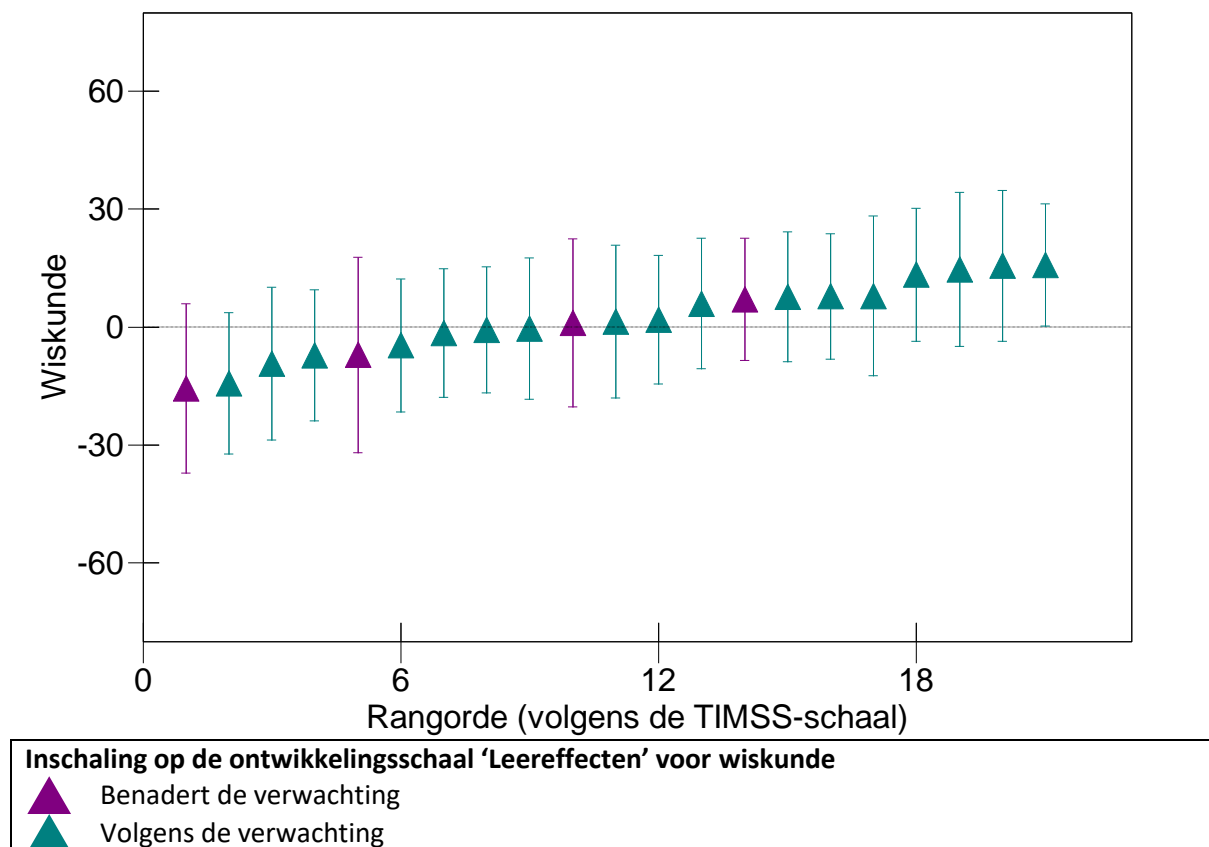
Figuur 47. Rangschikking van scholen volgens hun type 0-schooleffect op de TIMSS-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

Tabel 69

Kruistabel van scholen volgens hun type 0-schooleffect op de TIMSS-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
TIMSS					
Ondergemiddeld	-	2 (9,52%)	2 (9,52%)	-	4 (19,05%)
Gemiddeld	-	1 (4,76%)	10 (47,62%)	-	11 (52,38%)
Bovengemiddeld	-	1 (4,76%)	5 (23,81%)	-	6 (28,57%)
Totaal	-	4 (19,05%)	17 (80,95%)	-	21 (100%)

2.2 Type A-schooleffect



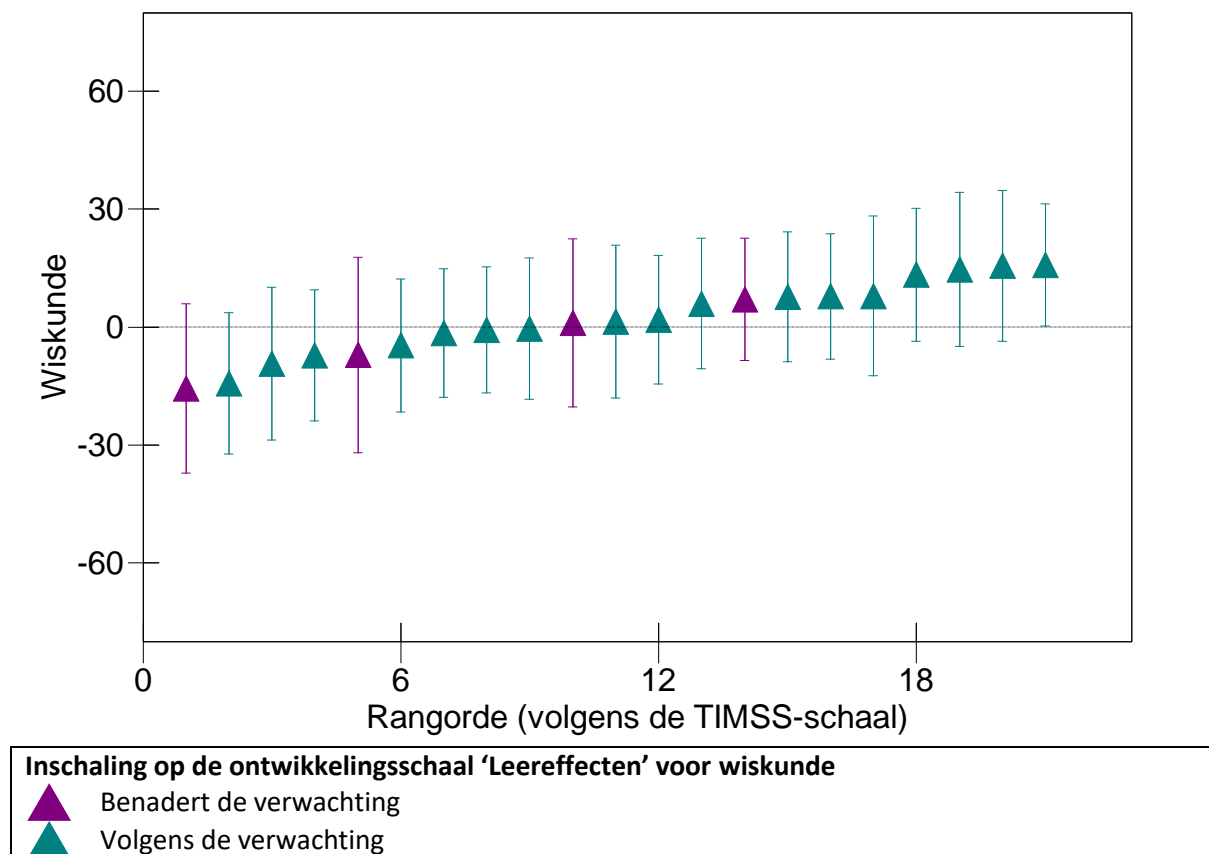
Figuur 48. Rangschikking van scholen volgens hun type A-schooleffect op de TIMSS-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

Tabel 70

Kruistabel van scholen volgens hun type A-schooleffect op de TIMSS-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
TIMSS					
Ondergemiddeld	-	-	-	-	-
Gemiddeld	-	4 (19,05%)	16 (76,19%)	-	20 (95,24%)
Bovengemiddeld	-	-	1 (4,76%)	-	1 (4,76%)
Totaal	-	4 (19,05%)	17 (80,95%)	-	21 (100%)

2.3 Type B-schooleffect



Figuur 49. Rangschikking van scholen volgens hun type B-schooleffect op de TIMSS-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

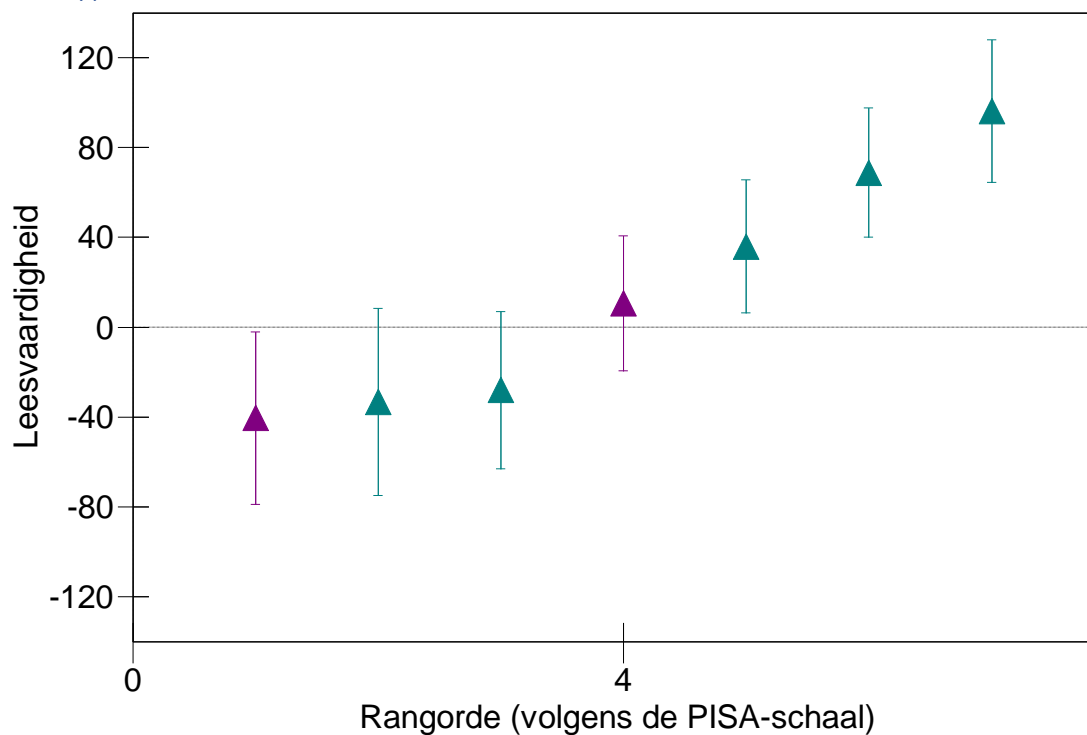
Tabel 71

Kruistabel van scholen volgens hun type B-schooleffect op de TIMSS-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
TIMSS					
Ondergemiddeld	-	-	-	-	-
Gemiddeld	-	4 (19,05%)	17 (80,95%)	-	21 (100%)
Bovengemiddeld	-	-	-	-	-
Totaal	-	4 (19,05%)	17 (80,95%)	-	21 (100%)

3. PISA 2018 leesvaardigheid en de doorlichting: 7 scholen

3.1 Type 0-schooleffect



Inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

- ▲ Benadert de verwachting
- ▲ Volgens de verwachting

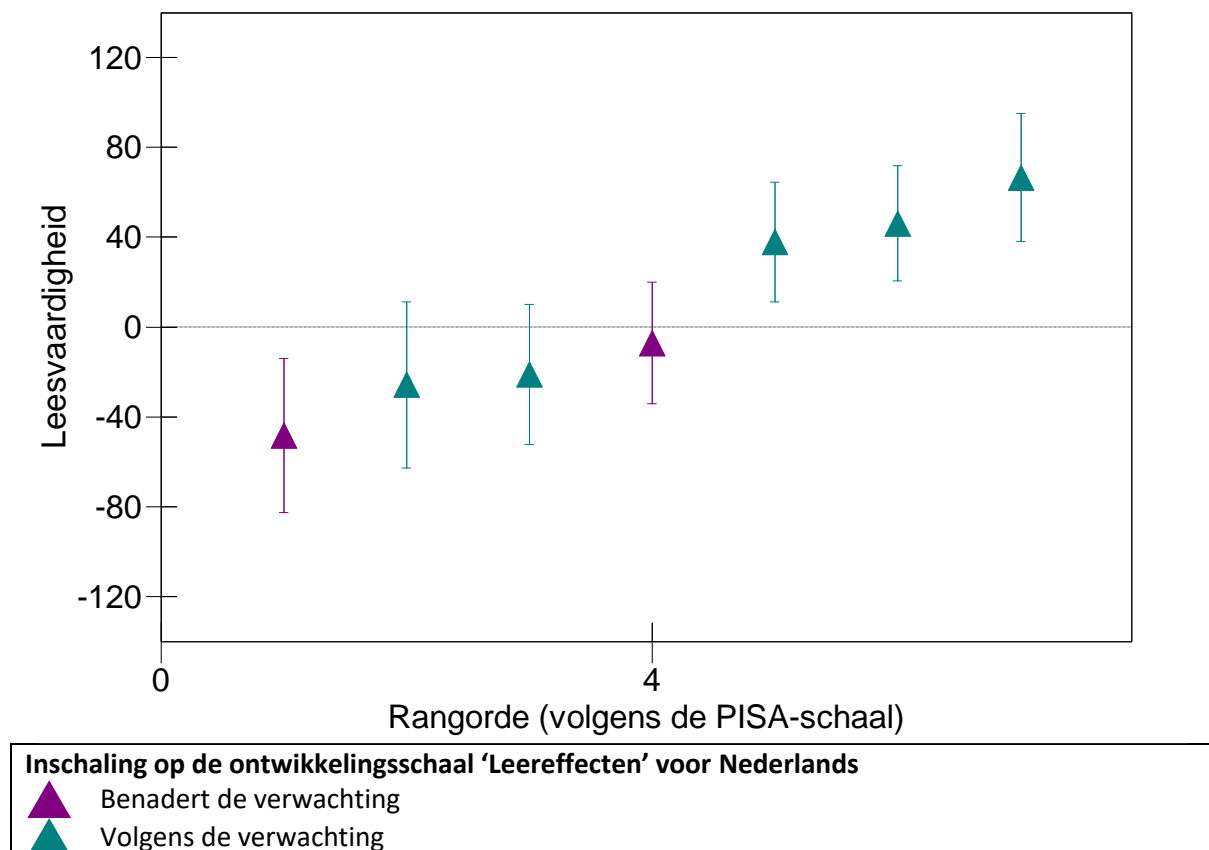
Figuur 50. Rangschikking van scholen volgens hun type 0-schooleffect op de PISA-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

Tabel 72

Kruistabel van scholen volgens hun type 0-schooleffect op de PISA-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

	Ontwikkelingsschaal leereffecten Nederlands (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
PISA					
Ondergemiddeld	-	1 (14,29%)	-	-	1 (14,29%)
Gemiddeld	-	1 (14,29%)	2 (28,57%)	-	3 (42,86%)
Bovengemiddeld	-	-	3 (42,86%)	-	3 (42,86%)
Totaal	-	2 (28,57%)	5 (71,43%)	-	7 (100%)

3.2 Type A-schooleffect



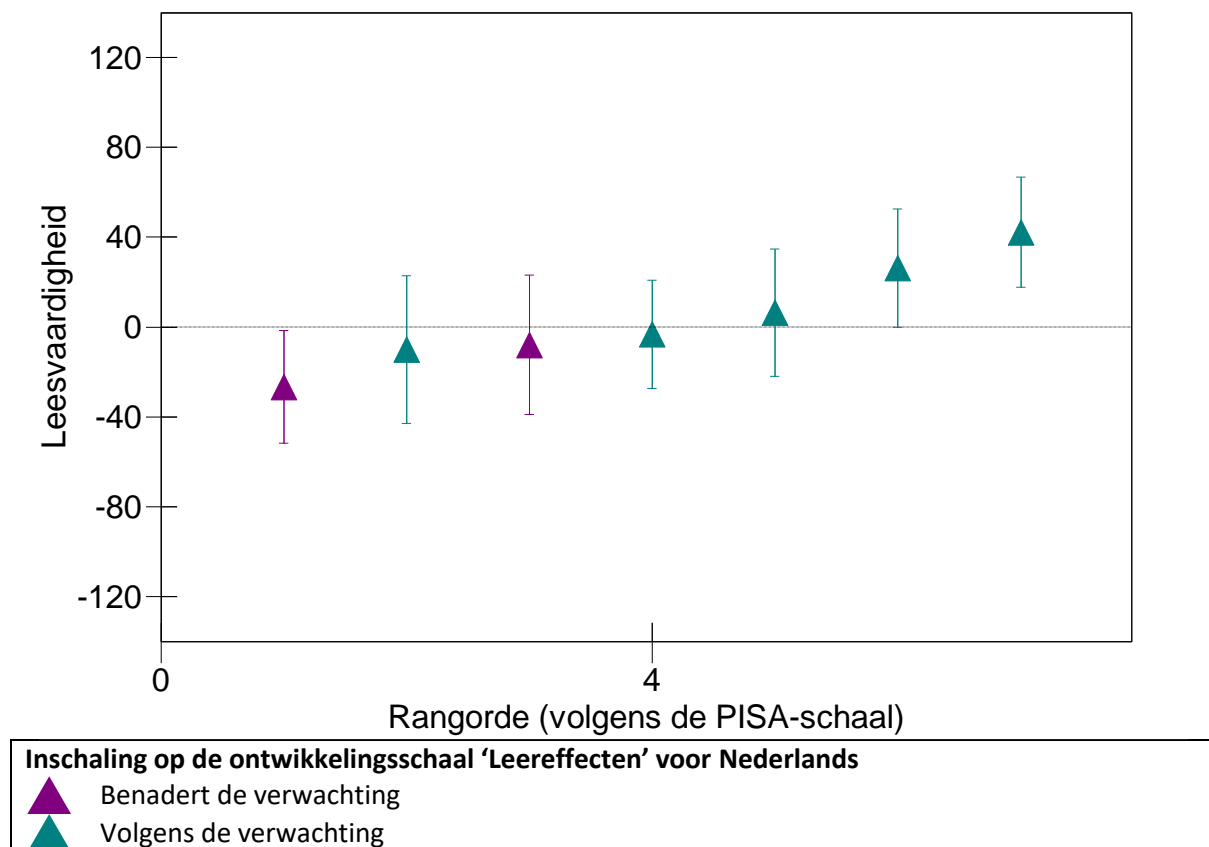
Figuur 51. Rangschikking van scholen volgens hun type A-schooleffect op de PISA-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

Tabel 73

Kruistabel van scholen volgens hun type A-schooleffect op de PISA-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

	Ontwikkelingsschaal leereffecten Nederlands (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
PISA					
Ondergemiddeld	-	1 (14,29%)	-	-	1 (14,29%)
Gemiddeld	-	1 (14,29%)	2 (28,57%)	-	3 (42,86%)
Bovengemiddeld	-	-	3 (42,86%)	-	3 (42,86%)
Totaal	-	2 (28,57%)	5 (71,43%)	-	7 (100%)

3.3 Type B-schooleffect



Figuur 52. Rangschikking van scholen volgens hun type B-schooleffect op de PISA-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

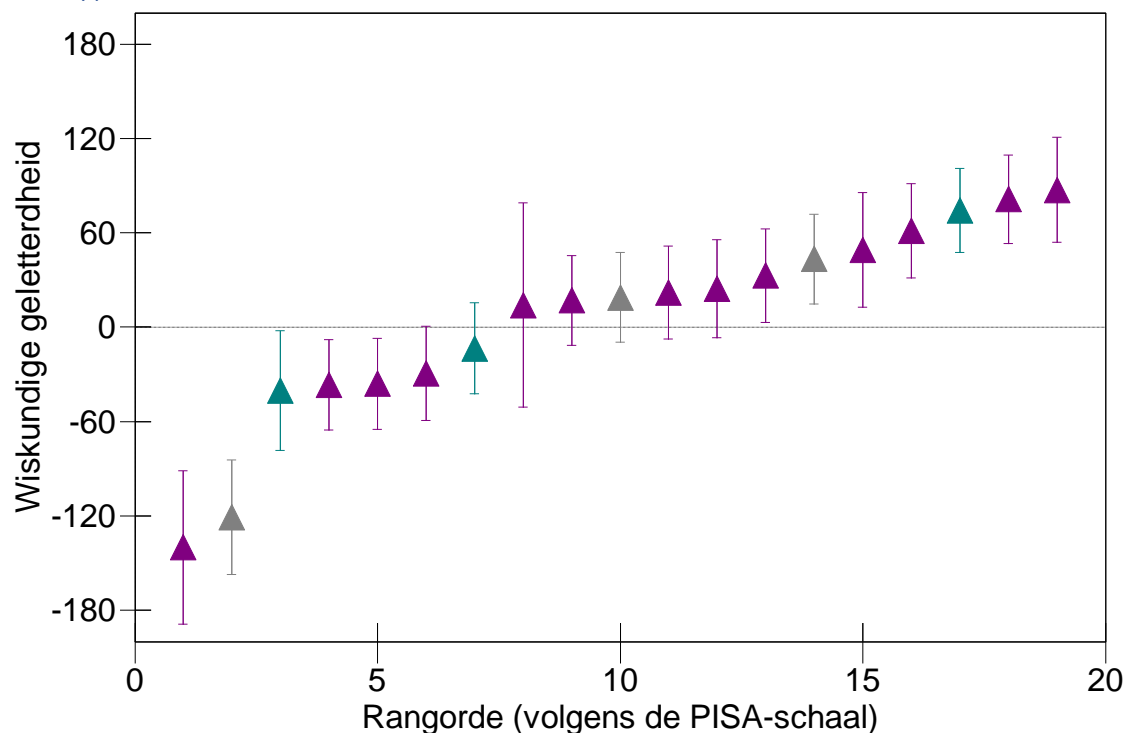
Tabel 74

Kruistabel van scholen volgens hun type B-schooleffect op de PISA-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands




	Ontwikkelingsschaal leereffecten Nederlands (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
PISA					
Ondergemiddeld	-	1 (14,29%)	-	-	1 (14,29%)
Gemiddeld	-	1 (14,29%)	4 (57,14%)	-	5 (71,43%)
Bovengemiddeld	-	-	1 (14,29%)	-	1 (14,29%)
Totaal	-	2 (28,57%)	5 (71,43%)	-	7 (100%)

4. PISA 2018 wiskundige geletterdheid en de doorlichting: 19 scholen

4.1 Type 0-schooleffect



Inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

-  Beneden de verwachting
-  Benadert de verwachting
-  Volgens de verwachting

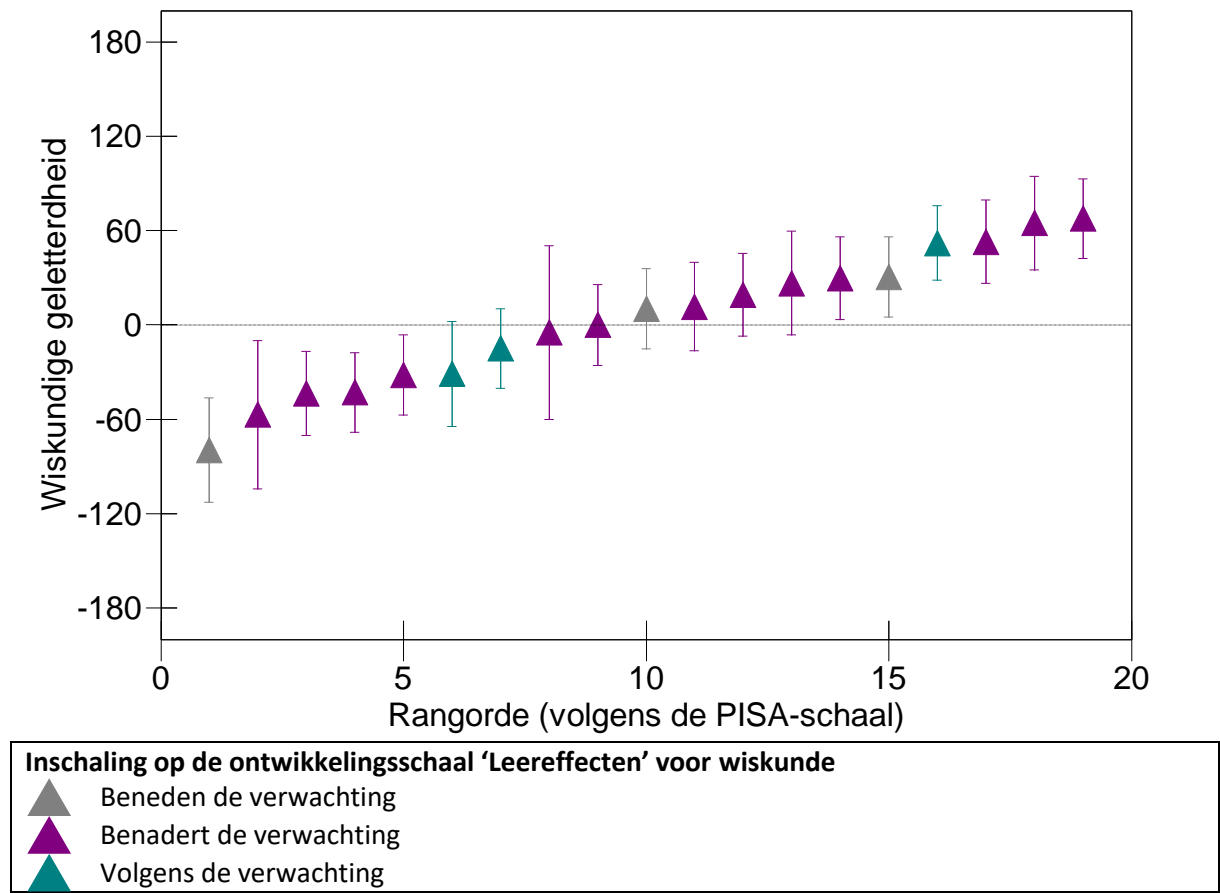
Figuur 53. Rangschikking van scholen volgens hun type 0-schooleffect op de PISA-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

Tabel 75

Kruistabel van scholen volgens hun type 0-schooleffect op de PISA-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
PISA					
Ondergemiddeld	1 (5,26%)	3 (15,79%)	1 (5,26%)	-	5 (26,32%)
Gemiddeld	1 (5,26%)	5 (26,32%)	1 (5,26%)	-	7 (36,84%)
Bovengemiddeld	1 (5,26%)	5 (26,32%)	1 (5,26%)	-	7 (36,84%)
Totaal	3 (15,79%)	13 (68,42%)	3 (15,79%)	-	19 (100%)

4.2 Type A-schooleffect



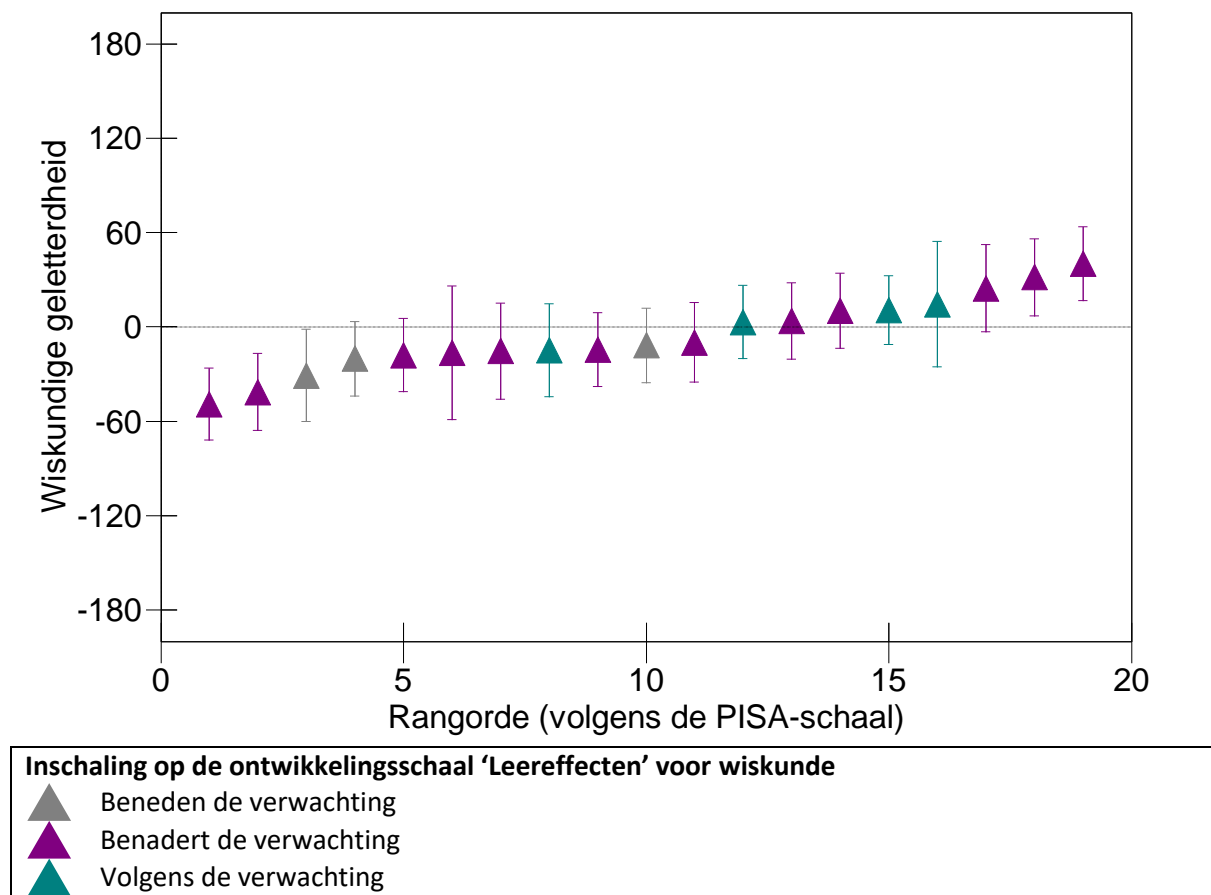
Figuur 54. Rangschikking van scholen volgens hun type A-schooleffect op de PISA-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

Tabel 76

Kruistabel van scholen volgens hun type A-schooleffect op de PISA-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
PISA					
Ondergemiddeld	1 (5,26%)	4 (21,05%)	-	-	5 (26,31%)
Gemiddeld	1 (5,26%)	5 (26,32%)	2 (10,53%)	-	8 (42,11%)
Bovengemiddeld	1 (5,26%)	4 (21,05%)	1 (5,26%)	-	6 (31,58%)
Totaal	3 (15,79%)	13 (68,42%)	3 (15,79%)	-	19 (100%)

4.3 Type B-schooleffect



Figuur 55. Rangschikking van scholen volgens hun type B-schooleffect op de PISA-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

Tabel 77

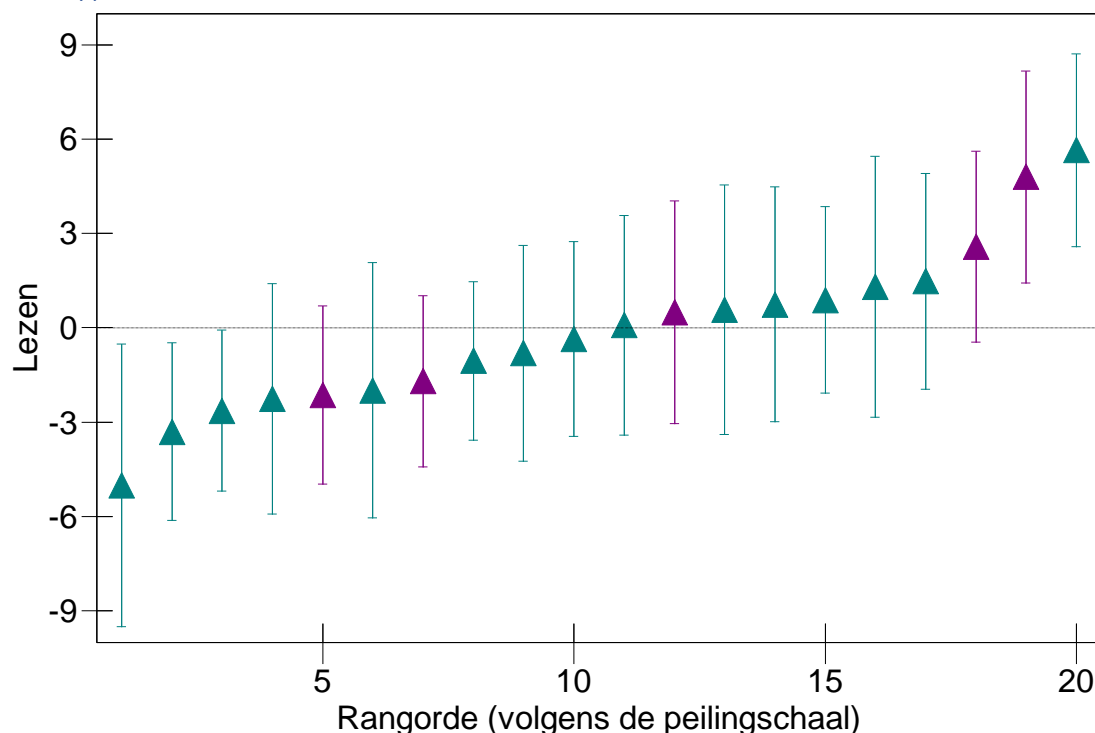
Kruistabel van scholen volgens hun type B-schooleffect op de PISA-schaal en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
PISA					
Ondergemiddeld	1 (5,26%)	2 (10,53%)	-	-	3 (15,79%)
Gemiddeld	2 (10,53%)	9 (47,37%)	3 (15,79%)	-	14 (73,68%)
Bovengemiddeld	-	2 (10,53%)	-	-	2 (10,53%)
Totaal	3 (15,79%)	13 (68,42%)	3 (15,79%)	-	19 (100%)

Vergelijking 2: Peilingsonderzoek en doorlichtingen

1. Peiling Nederlands 2018 en de doorlichting: 20 scholen

1.1 Type 0-schooleffect



Inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

▲ Benadert de verwachting

▲ Volgens de verwachting

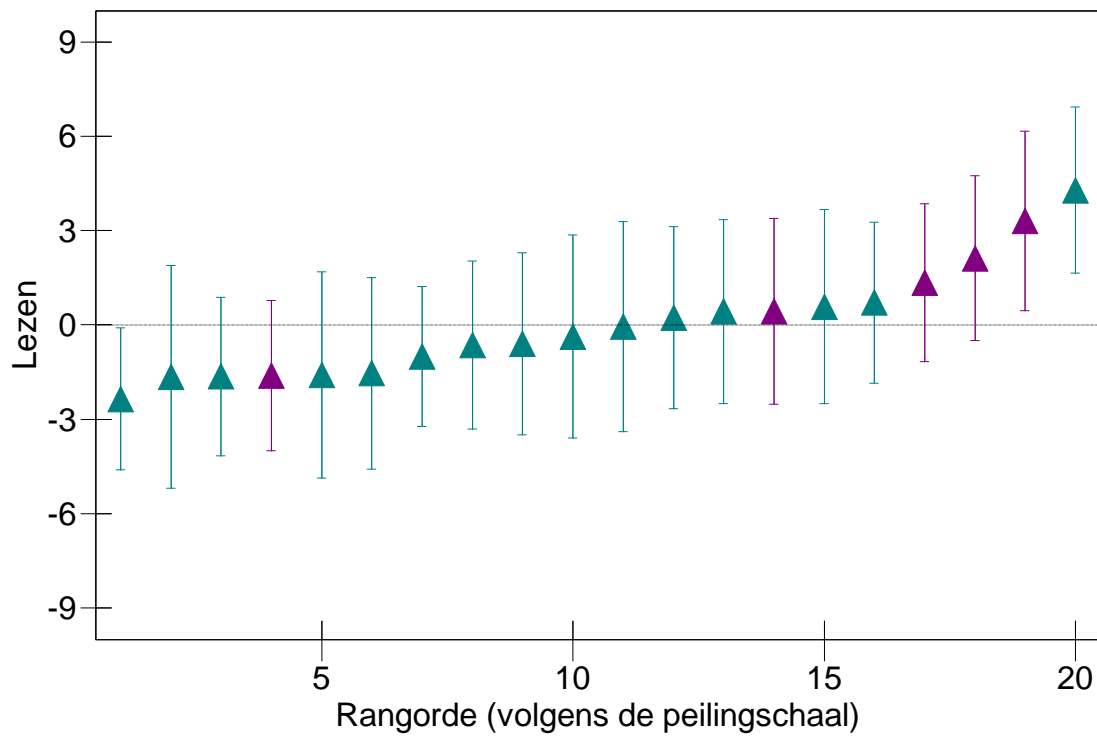
Figuur 56. Rangschikking van scholen volgens hun type 0-schooleffect op de peilingschaal lezen en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

Tabel 78

Kruistabel van scholen volgens hun type 0-schooleffect op de peilingschaal lezen en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

	Ontwikkelingsschaal leereffecten Nederlands (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Lezen (peiling)					
Ondergemiddeld	-	-	3 (15%)	-	3 (15%)
Gemiddeld	-	4 (20%)	11 (55%)	-	15 (75%)
Bovengemiddeld	-	1 (5%)	1 (5%)	-	2 (10%)
Totaal	-	5 (25%)	15 (75%)	-	20 (100%)

1.2 Type A-schooleffect



Inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

- ▲ Benadert de verwachting
- ▲ Volgens de verwachting

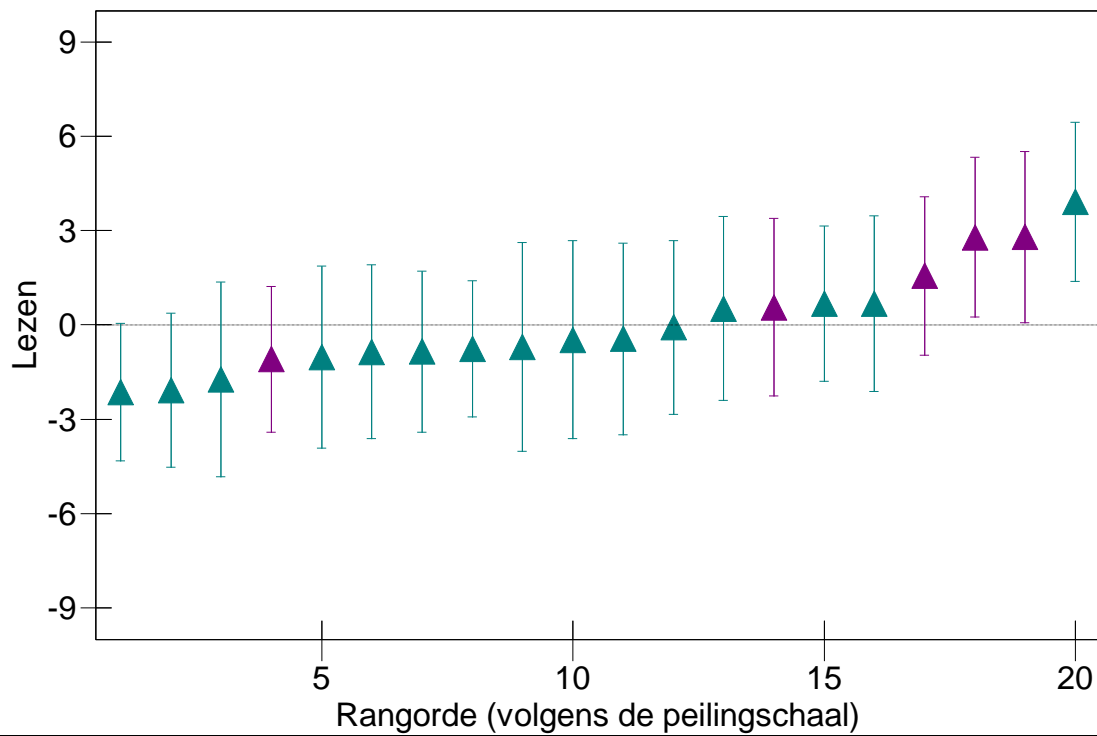
Figuur 57. Rangschikking van scholen volgens hun type A-schooleffect op de peilingschaal lezen en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

Tabel 79

Kruistabel van scholen volgens hun type A-schooleffect op de peilingschaal lezen en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

	Ontwikkelingsschaal leereffecten Nederlands (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Lezen (peiling)					
Ondergemiddeld	-	-	1 (5%)	-	1 (5%)
Gemiddeld	-	4 (20%)	13 (65%)	-	17 (85%)
Bovengemiddeld	-	1 (5%)	1 (5%)	-	2 (10%)
Totaal	-	5 (25%)	15 (75%)	-	20 (100%)

1.3 Type B-schooleffect



Inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

- ▲ Benadert de verwachting
- ▲ Volgens de verwachting

Figuur 58. Rangschikking van scholen volgens hun type B-schooleffect op de peilingschaal lezen en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

Tabel 80

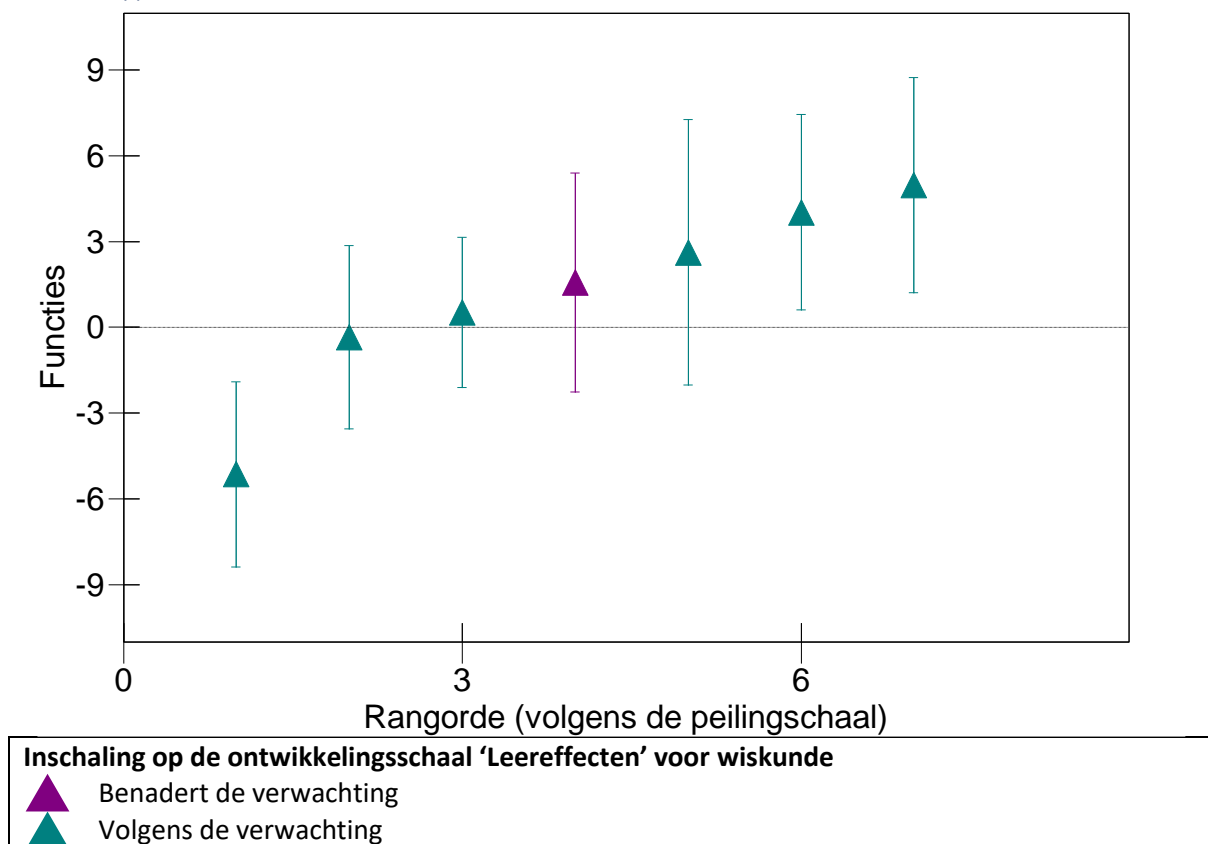
Kruistabel van scholen volgens hun type B-schooleffect op de peilingschaal lezen en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor Nederlands

	Ontwikkelingsschaal leereffecten Nederlands (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Lezen (peiling)					
Ondergemiddeld	-	-	-	-	-
Gemiddeld	-	3 (15%)	14 (70%)	-	17 (85%)
Bovengemiddeld	-	2 (10%)	1 (5%)	-	3 (15%)
Totaal	-	5 (25%)	15 (75%)	-	20 (100%)

2. Peiling wiskunde basisonderwijs 2021 en de doorlichting: 19 scholen

2.1 Functies (7 scholen)

2.1.1 Type 0-schooleffect



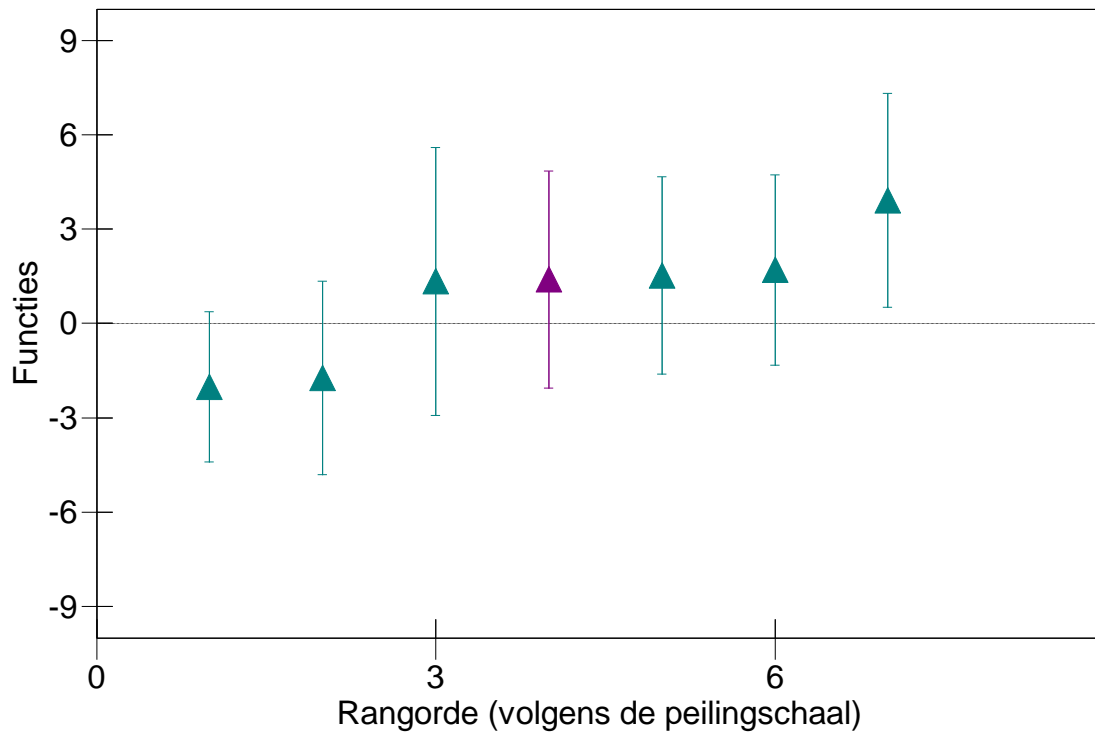
Figuur 59. Rangschikking van scholen volgens hun type 0-schooleffect op de peilingschaal functies en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

Tabel 81

Kruistabel van scholen volgens hun type 0-schooleffect op de peilingschaal functies en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Functies (peiling wiskunde)					
Ondergemiddeld	-	-	1 (14,29%)	-	1 (14,29%)
Gemiddeld	-	1 (14,29%)	3 (42,86%)	-	4 (57,14%)
Bovengemiddeld	-	-	2 (28,57%)	-	2 (28,57%)
Totaal	-	1 (14,29%)	6 (85,71%)	-	7 (100%)

2.1.2 Type A-schooleffect



Inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

- ▲ Benadert de verwachting
- ▲ Volgens de verwachting

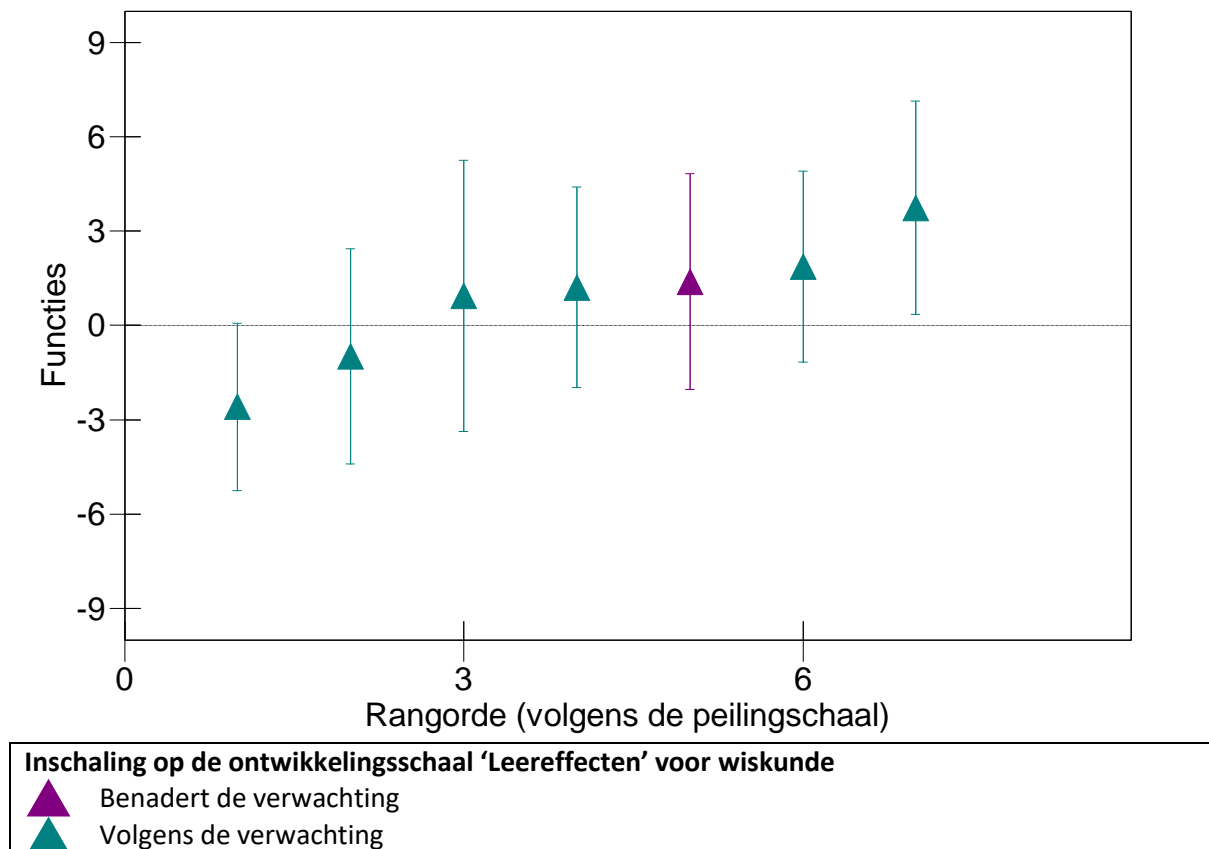
Figuur 60. Rangschikking van scholen volgens hun type A-schooleffect op de peilingschaal functies en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

Tabel 82

Kruistabel van scholen volgens hun type A-schooleffect op de peilingschaal functies en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Functies (peiling wiskunde)					
Ondergemiddeld	-	-	-	-	-
Gemiddeld	-	1 (14,29%)	5 (71,43%)	-	6 (85,71%)
Bovengemiddeld	-	-	1 (14,29%)	-	1 (14,29%)
Totaal	-	1 (14,29%)	6 (85,71%)	-	7 (100%)

2.1.3 Type B-schooleffect



Figuur 61. Rangschikking van scholen volgens hun type B-schooleffect op de peilingschaal functies en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

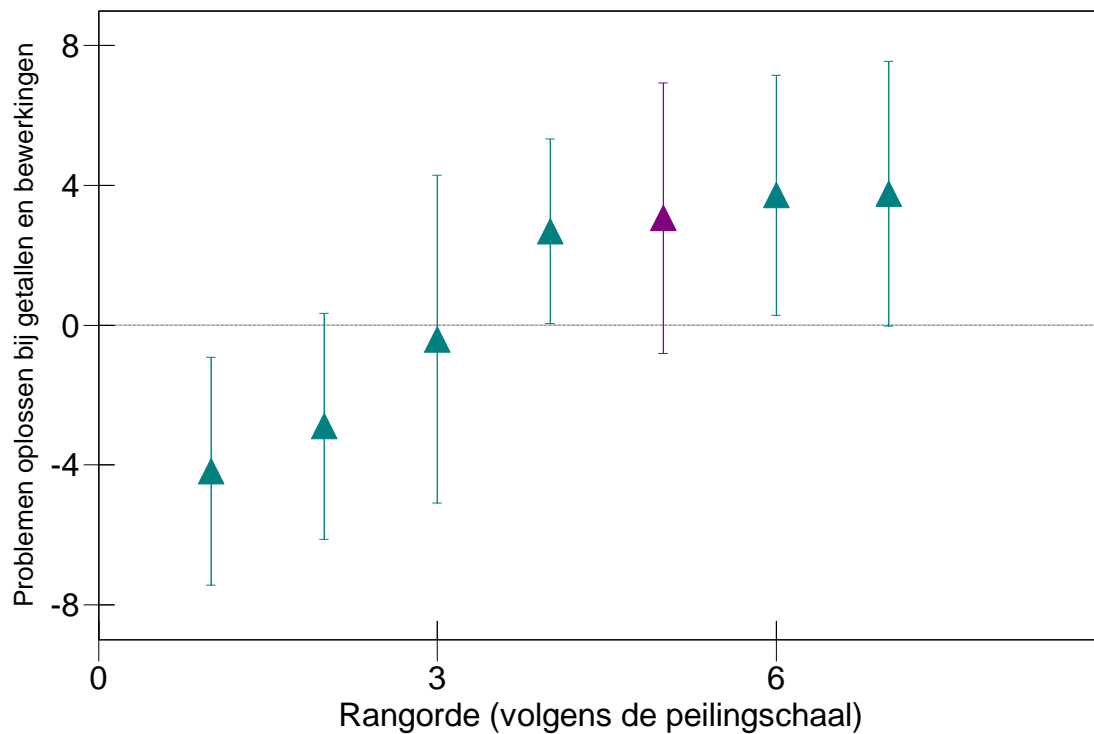
Tabel 83

Kruistabel van scholen volgens hun type B-schooleffect op de peilingschaal functies en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Functies (peiling wiskunde)					
Ondergemiddeld	-	-	-	-	-
Gemiddeld	-	1 (14,29%)	5 (71,43%)	-	6 (85,71%)
Bovengemiddeld	-	-	1 (14,29%)	-	1 (14,29%)
Totaal	-	1 (14,29%)	6 (85,71%)	-	7 (100%)

2.2 Problemen oplossen bij getallen en bewerkingen (7 scholen)

2.2.1 Type 0-schooleffect



Inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

- ▲ Benadert de verwachting
- ▲ Volgens de verwachting

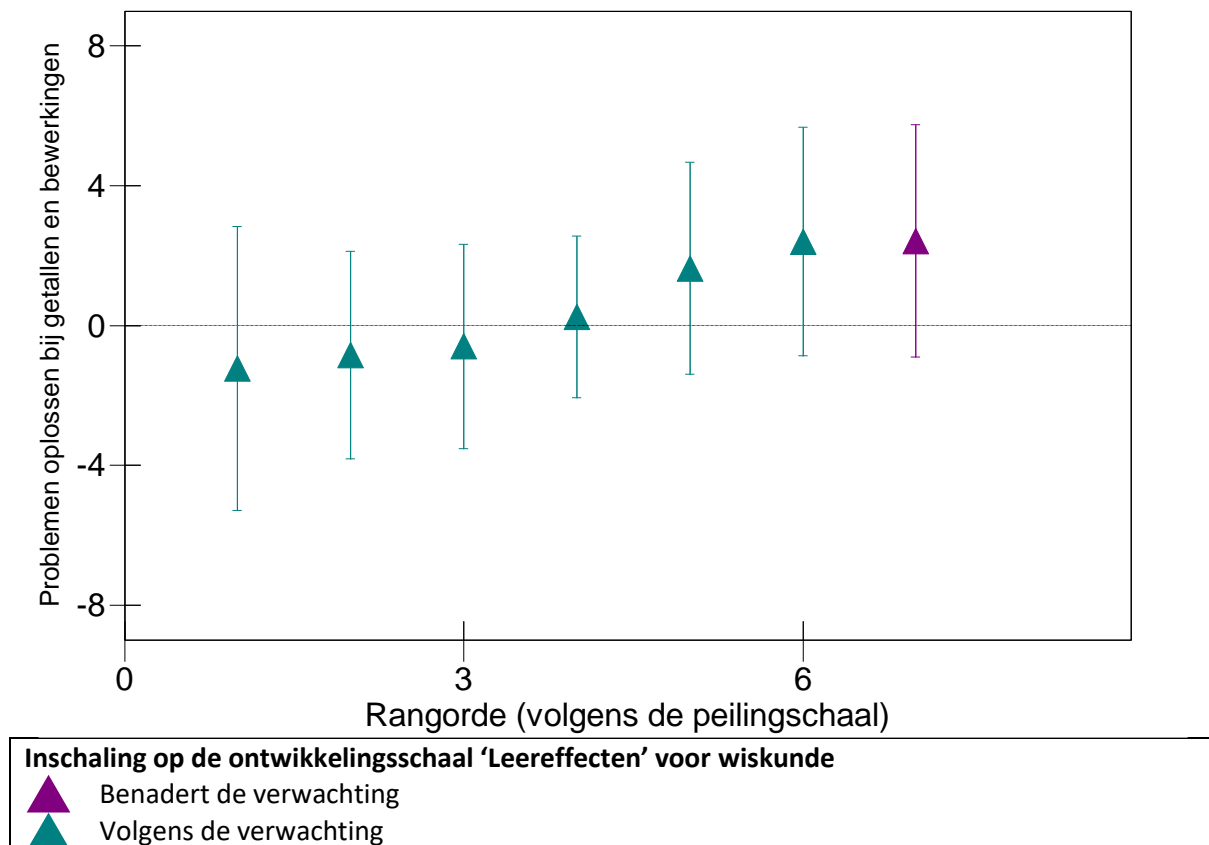
Figuur 62. Rangschikking van scholen volgens hun type 0-schooleffect op de peilingschaal problemen oplossen bij getallen en bewerkingen en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

Tabel 84

Kruistabel van scholen volgens hun type 0-schooleffect op de peilingschaal problemen oplossen bij getallen en bewerkingen en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Problemen oplossen bij getallen en bewerkingen (peiling wiskunde)					
Ondergemiddeld	-	-	1 (14,29%)	-	1 (14,29%)
Gemiddeld	-	1 (14,29%)	3 (42,86%)	-	4 (57,14%)
Bovengemiddeld	-	-	2 (28,57%)	-	2 (28,57%)
Totaal	-	1 (14,29%)	6 (85,71%)	-	7 (100%)

2.2.2 Type A-schooleffect



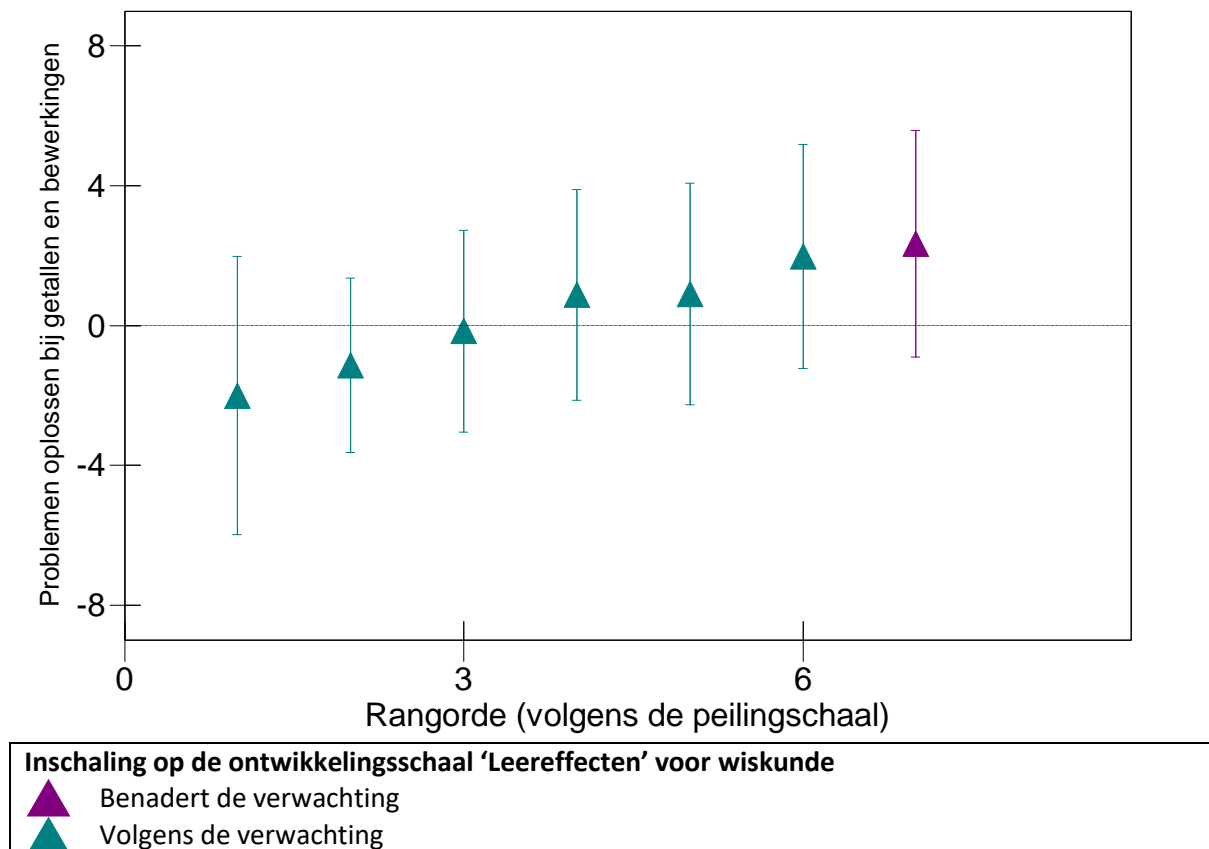
Figuur 63. Rangschikking van scholen volgens hun type A-schooleffect op de peilingschaal problemen oplossen bij getallen en bewerkingen en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

Tabel 85

Kruistabel van scholen volgens hun type A-schooleffect op de peilingschaal problemen oplossen bij getallen en bewerkingen en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Problemen oplossen bij getallen en bewerkingen (peiling wiskunde)					
Ondergemiddeld	-	-	-	-	-
Gemiddeld	-	1 (14,29%)	6 (85,71%)	-	7 (100%)
Bovengemiddeld	-	-	-	-	-
Totaal	-	1 (14,29%)	6 (85,71%)	-	7 (100%)

2.2.3 Type B-schooleffect



Figuur 64. Rangschikking van scholen volgens hun type B-schooleffect op de peilingschaal problemen oplossen bij getallen en bewerkingen en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

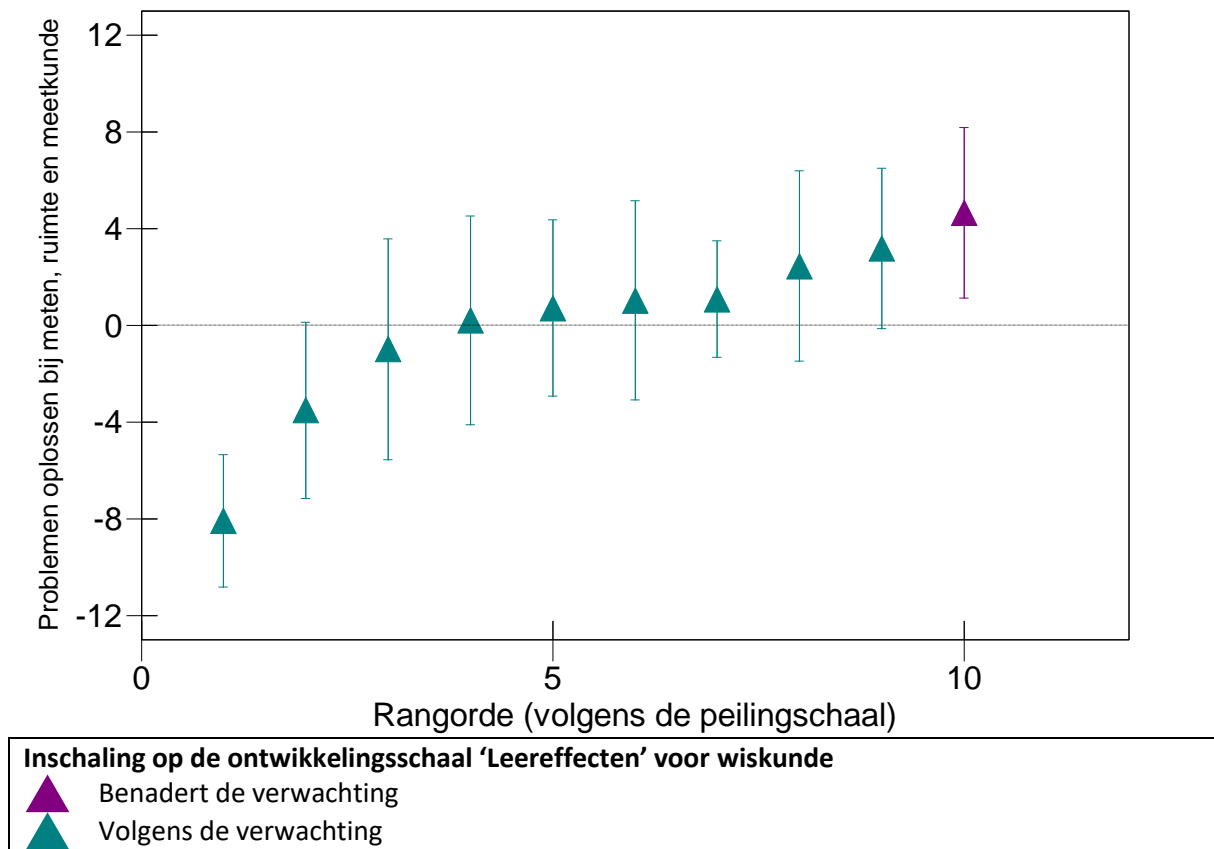
Tabel 86

Kruistabel van scholen volgens hun type B-schooleffect op de peilingschaal problemen oplossen bij getallen en bewerkingen en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Problemen oplossen bij getallen en bewerkingen (peiling wiskunde)					
Ondergemiddeld	-	-	-	-	-
Gemiddeld	-	1 (14,29%)	6 (85,71%)	-	7 (100%)
Bovengemiddeld	-	-	-	-	-
Totaal	-	1 (14,29%)	6 (85,71%)	-	7 (100%)

2.3 Problemen oplossen bij meten, ruimte en meetkunde (10 scholen)

2.3.1 Type 0-schooleffect



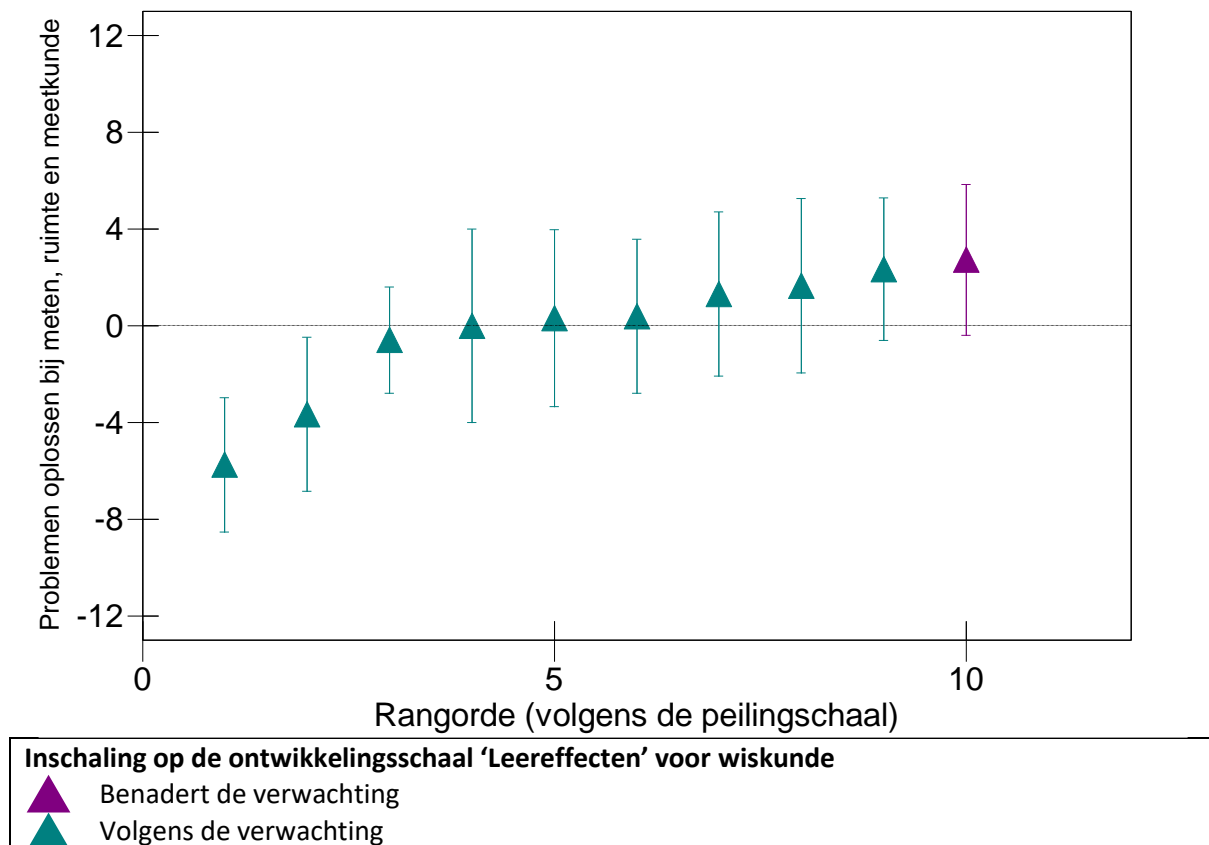
Figuur 65. Rangschikking van scholen volgens hun type 0-schooleffect op de peilingschaal problemen oplossen bij meten, ruimte en meetkunde en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

Tabel 87

Kruistabel van scholen volgens hun type 0-schooleffect op de peilingschaal problemen oplossen bij meten, ruimte en meetkunde en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Problemen oplossen bij meten, ruimte en meetkunde (peiling wiskunde)					
Ondergemiddeld	-	-	1 (10%)	-	1 (10%)
Gemiddeld	-	-	8 (80%)	-	8 (80%)
Bovengemiddeld	-	1 (10%)	-	-	1 (10%)
Totaal	-	1 (10%)	9 (90%)	-	10 (100%)

2.3.2 Type A-schooleffect



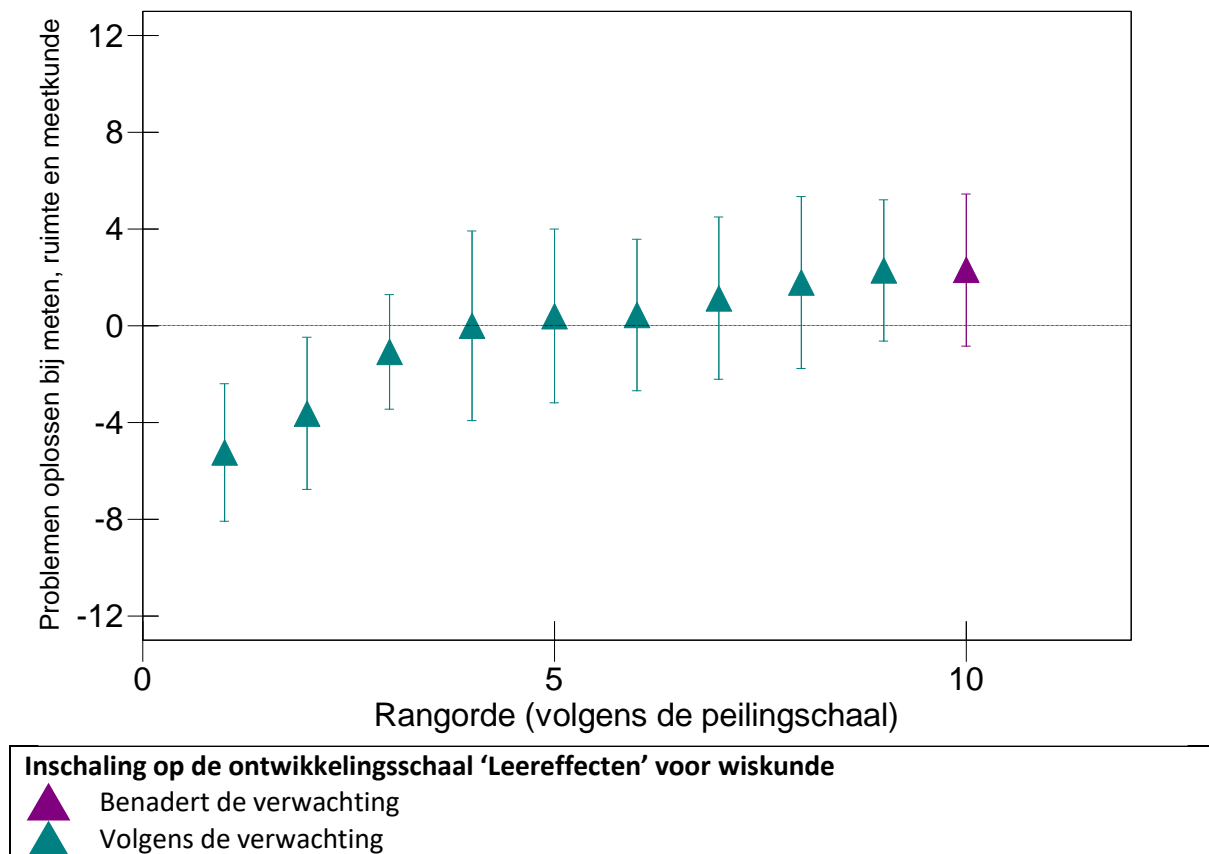
Figuur 66. Rangschikking van scholen volgens hun type A-schooleffect op de peilingschaal problemen oplossen bij meten, ruimte en meetkunde en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

Tabel 88

Kruistabel van scholen volgens hun type A-schooleffect op de peilingschaal problemen oplossen bij meten, ruimte en meetkunde en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Problemen oplossen bij meten, ruimte en meetkunde (peiling wiskunde)					
Ondergemiddeld	-	-	2 (20%)	-	2 (20%)
Gemiddeld	-	1 (10%)	7 (70%)	-	8 (80%)
Bovengemiddeld	-	-	-	-	-
Totaal	-	1 (10%)	9 (90%)	-	10 (100%)

2.3.3 Type B-schooleffect



Figuur 67. Rangschikking van scholen volgens hun type B-schooleffect op de peilingschaal problemen oplossen bij meten, ruimte en meetkunde en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

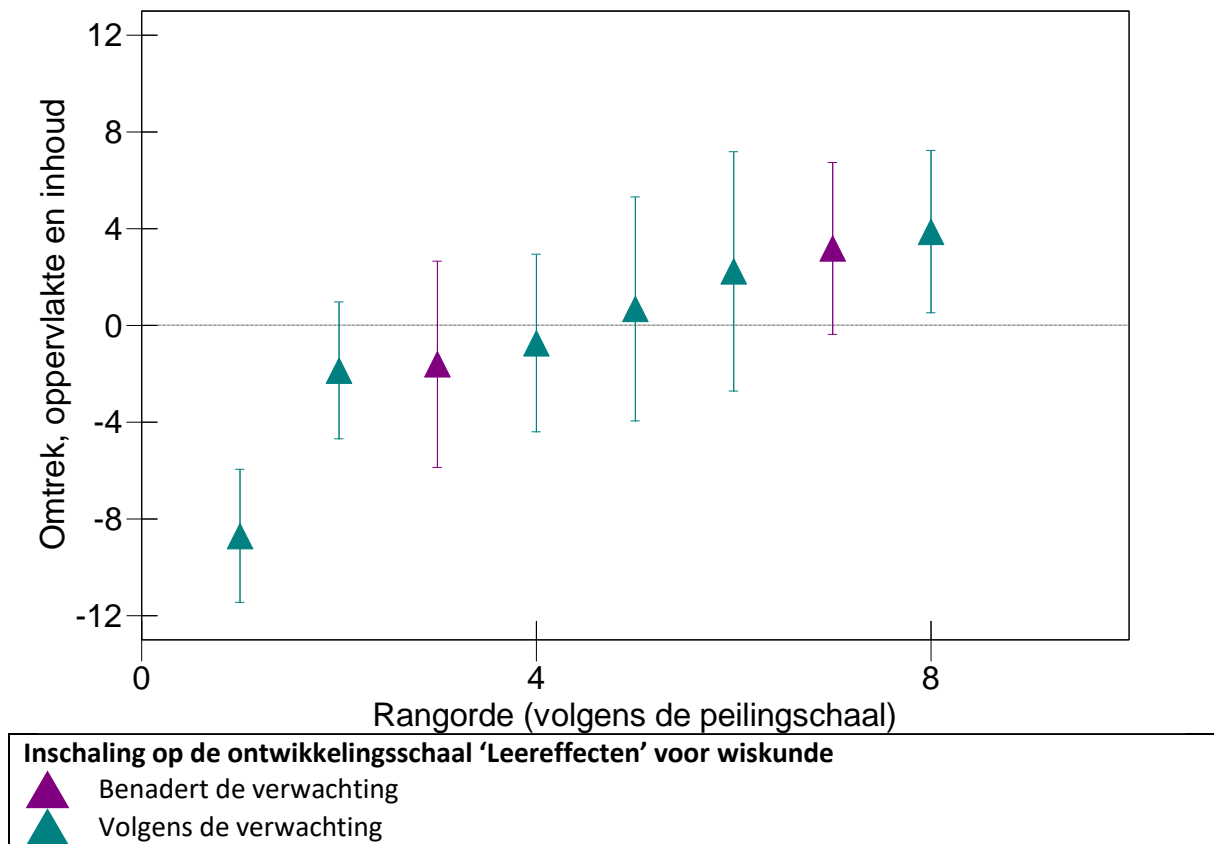
Tabel 89

Kruistabel van scholen volgens hun type B-schooleffect op de peilingschaal problemen oplossen bij meten, ruimte en meetkunde en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Problemen oplossen bij meten, ruimte en meetkunde (peiling wiskunde)					
Ondergemiddeld	-	-	2 (20%)	-	2 (20%)
Gemiddeld	-	1 (10%)	7 (70%)	-	8 (80%)
Bovengemiddeld	-	-	-	-	-
Totaal	-	1 (10%)	9 (90%)	-	10 (100%)

2.4 Omtrek, oppervlakte en inhoud (8 scholen)

2.4.1 Type 0-schooleffect



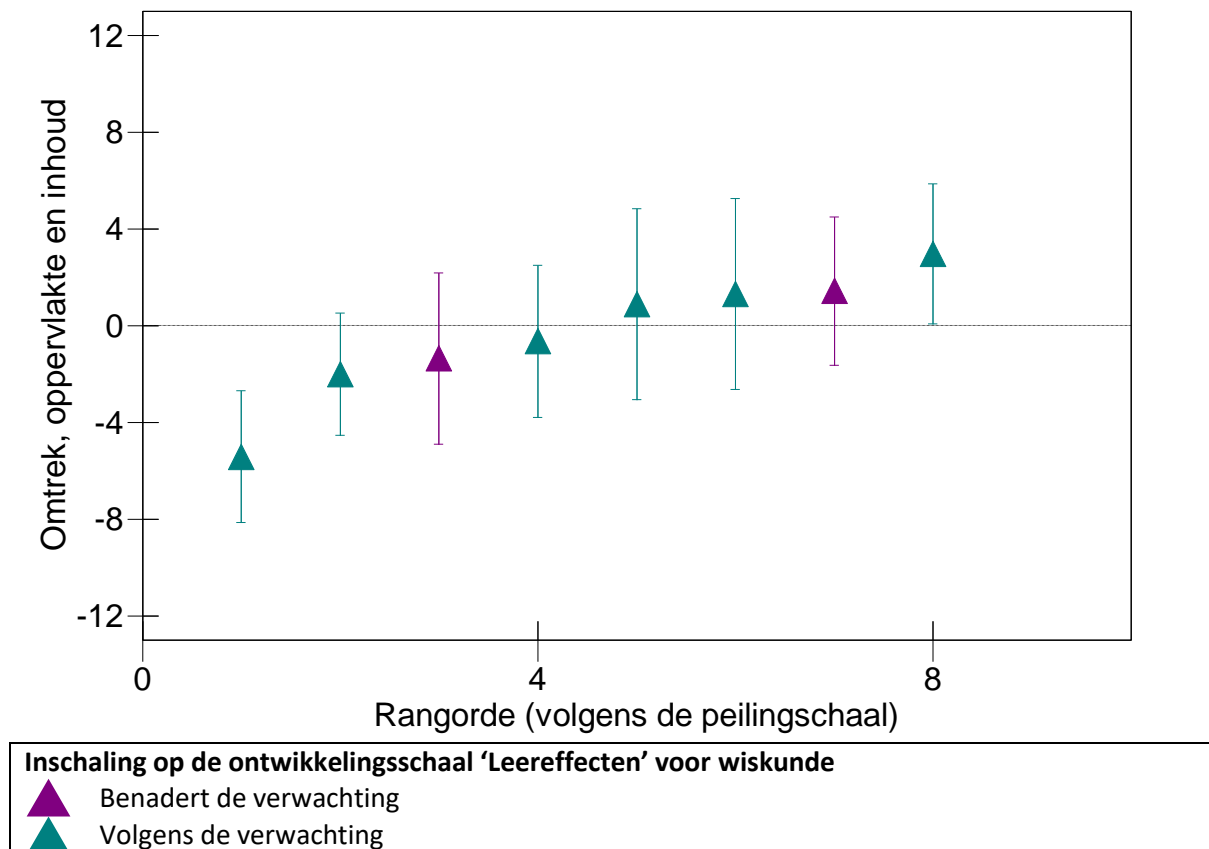
Figuur 68. Rangschikking van scholen volgens hun type 0-schooleffect op de peilingschaal omtrek, oppervlakte en inhoud en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

Tabel 90

Kruistabel van scholen volgens hun type 0-schooleffect op de peilingschaal omtrek, oppervlakte en inhoud en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Omtrek, oppervlakte en inhoud (peiling wiskunde)					
Ondergemiddeld	-	-	1 (12,5%)	-	1 (12,5%)
Gemiddeld	-	2 (25,0%)	4 (50,0%)	-	6 (75,0%)
Bovengemiddeld	-	-	1 (12,5%)	-	1 (12,5%)
Totaal	-	2 (25,0%)	6 (75,0%)	-	8 (100%)

2.4.2 Type A-schooleffect



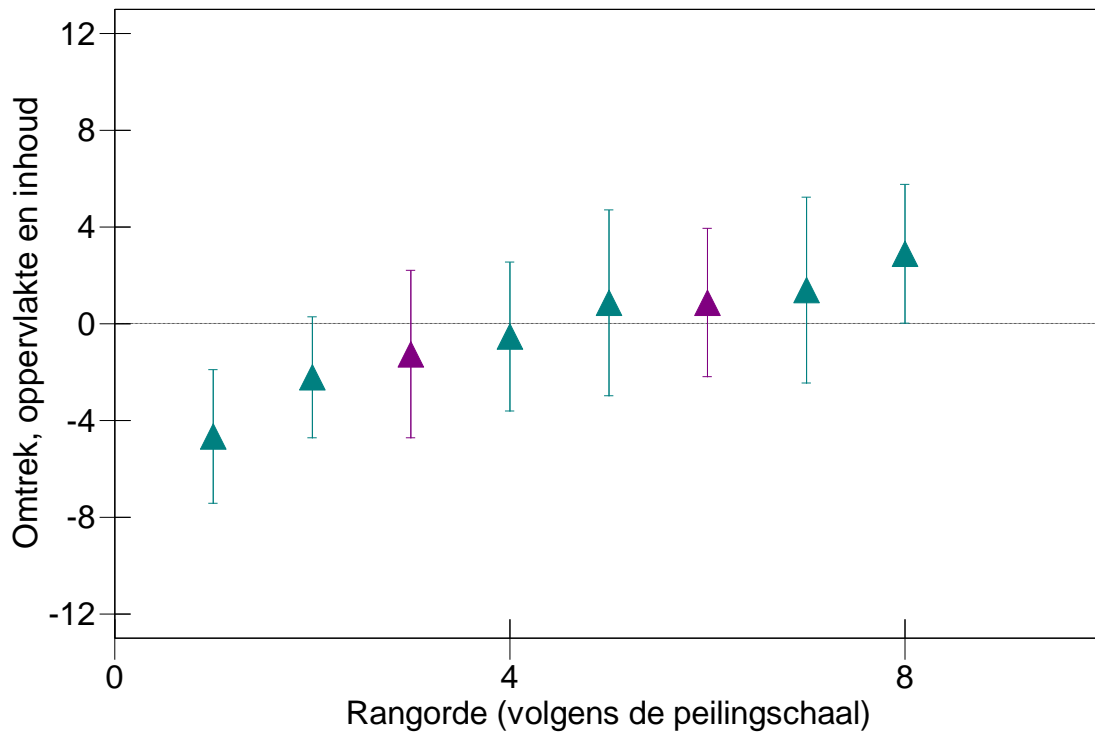
Figuur 69. Rangschikking van scholen volgens hun type A-schooleffect op de peilingschaal omtrek, oppervlakte en inhoud en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

Tabel 91

Kruistabel van scholen volgens hun type A-schooleffect op de peilingschaal omtrek, oppervlakte en inhoud en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Omtrek, oppervlakte en inhoud (peiling wiskunde)					
Ondergemiddeld	-	-	1 (12,5%)	-	1 (12,5%)
Gemiddeld	-	2 (25,0%)	4 (50,0%)	-	6 (75,0%)
Bovengemiddeld	-	-	1 (12,5%)	-	1 (12,5%)
Totaal	-	2 (25,0%)	6 (75,0%)	-	8 (100%)

2.4.3 Type B-schooleffect



Inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

- ▲ Benadert de verwachting
- ▲ Volgens de verwachting

Figuur 70. Rangschikking van scholen volgens hun type B-schooleffect op de peilingschaal omtrek, oppervlakte en inhoud en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

Tabel 92

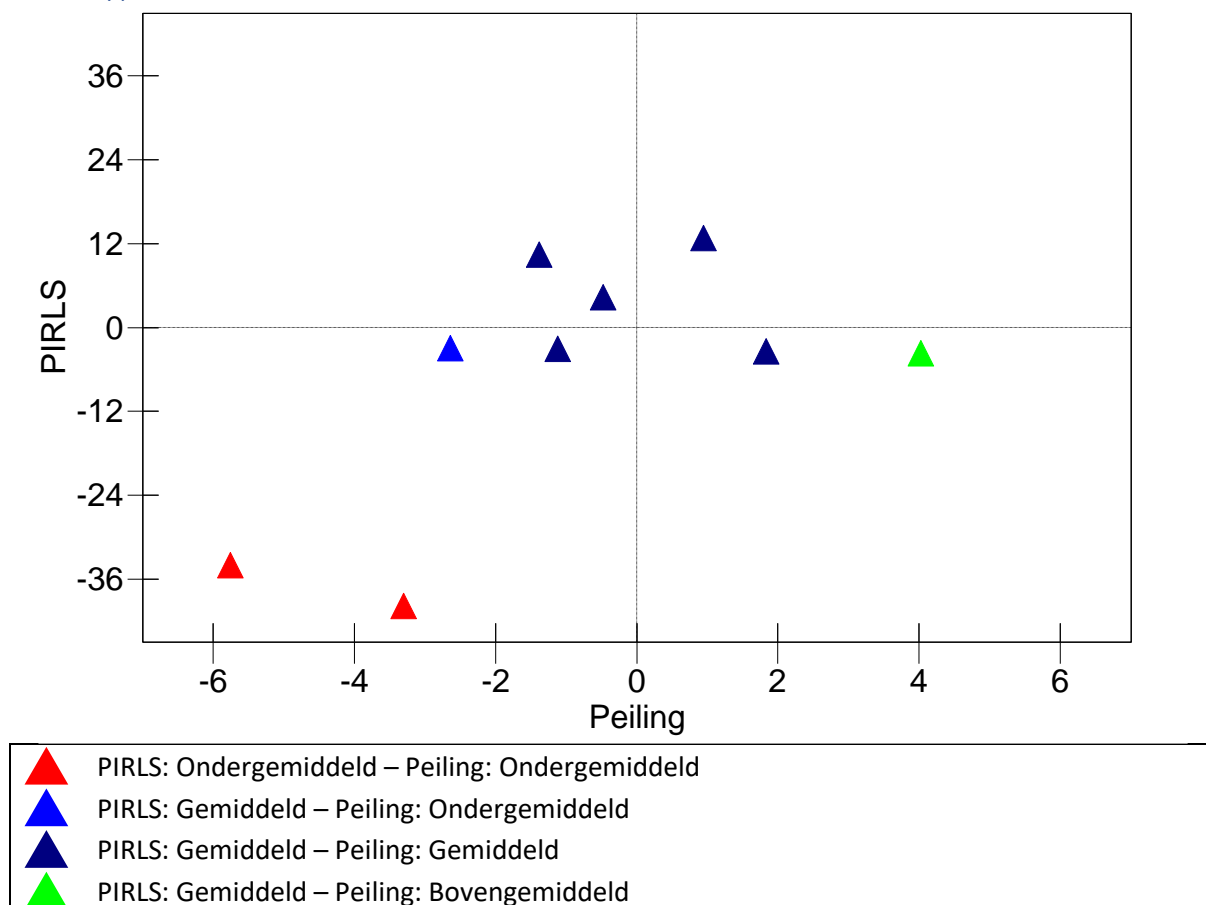
Kruistabel van scholen volgens hun type B-schooleffect op de peilingschaal omtrek oppervlakte en inhoud en volgens hun inschaling op de ontwikkelingsschaal 'Leereffecten' voor wiskunde

	Ontwikkelingsschaal leereffecten wiskunde (doorlichtingen)				Totaal
	Beneden de verwachting	Benadert de verwachting	Volgens de verwachting	Overstijgt de verwachting	
Omtrek, oppervlakte en inhoud (peiling wiskunde)					
Ondergemiddeld	-	-	1 (12,5%)	-	1 (12,5%)
Gemiddeld	-	2 (25,0%)	4 (50,0%)	-	6 (75,0%)
Bovengemiddeld	-	-	1 (12,5%)	-	1 (12,5%)
Totaal	-	2 (25,0%)	6 (75,0%)	-	8 (100%)

Vergelijking 3: Internationaal vergelijkend onderzoek en peilingsonderzoek

1. Peiling Nederlands basisonderwijs 2018 & PIRLS 2016: 9 scholen

1.1 Type 0-schooleffect



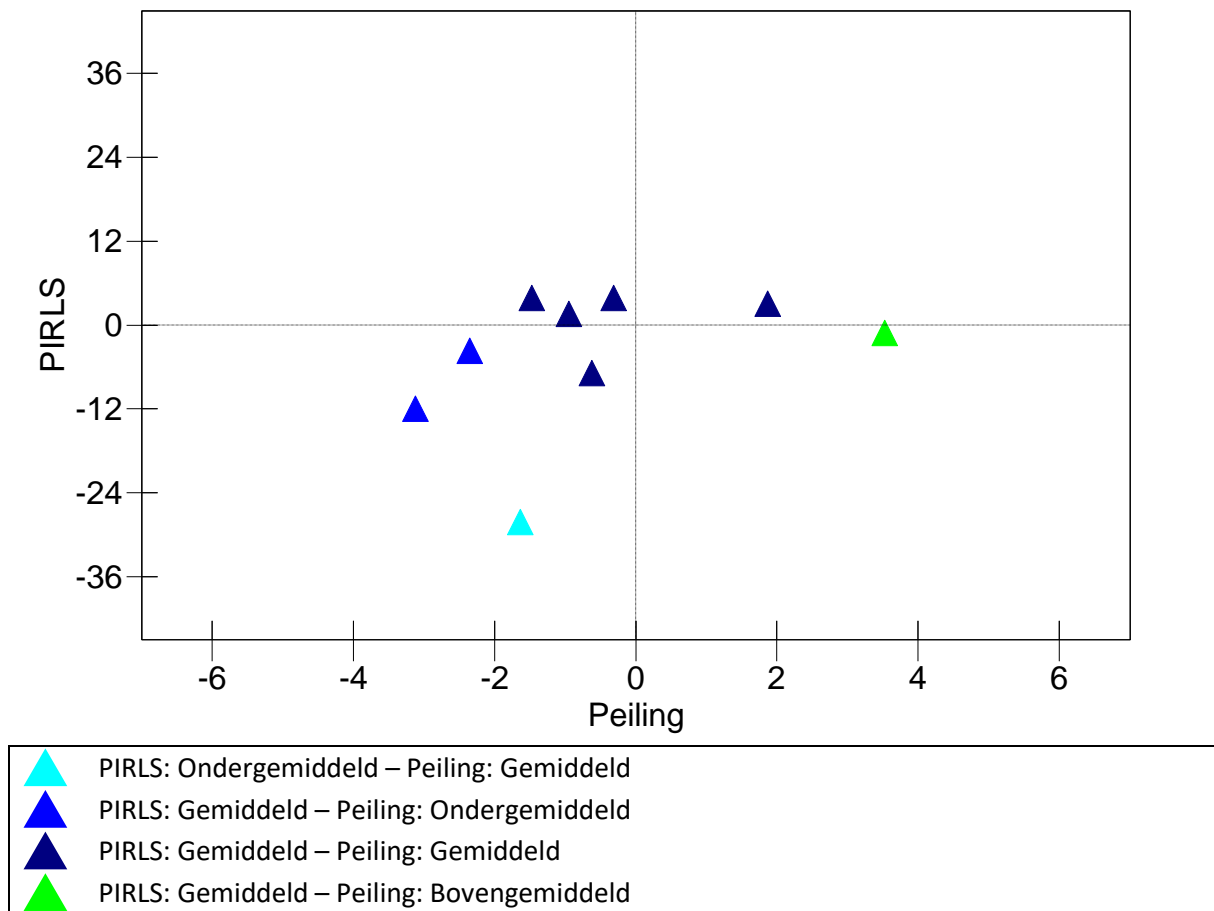
Figuur 71. Puntgrafiek van scholen volgens hun type 0-schooleffect op PIRLS en op lezen bij de peiling Nederlands in het basisonderwijs

Tabel 93

Classificatie van scholen volgens hun type 0-schooleffect op PIRLS en op lezen bij de peiling Nederlands in het basisonderwijs

	Peiling			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PIRLS				
Ondergemiddeld	2 (22,22%)	-	-	2 (22,22%)
Gemiddeld	1 (11,11%)	5 (55,56%)	1 (11,11%)	7 (77,78%)
Bovengemiddeld	-	-	-	-
Totaal	3 (33,33%)	5 (55,56%)	1 (11,11%)	9 (100%)

1.2 Type A-schooleffect



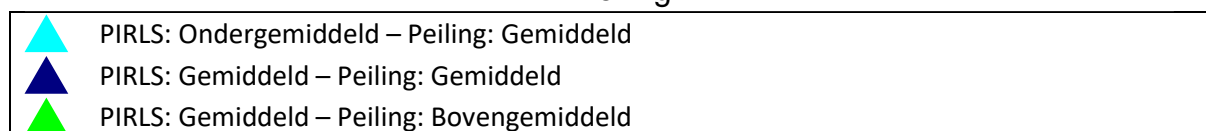
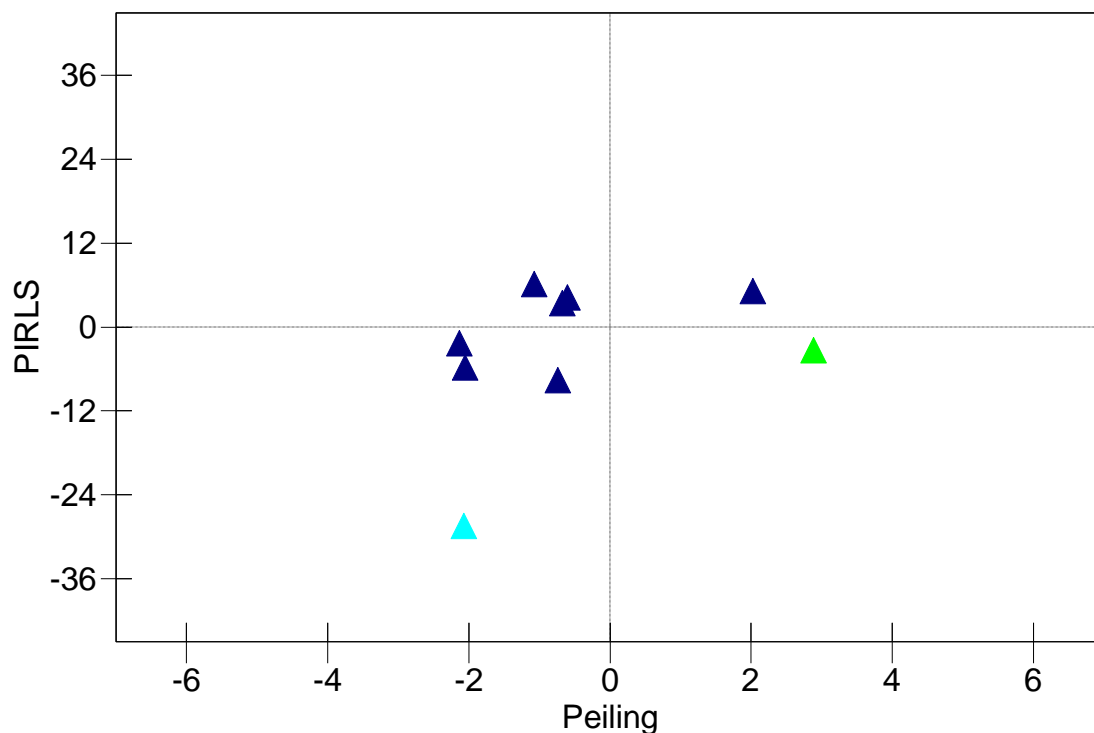
Figuur 72. Puntgrafiek van scholen volgens hun type A-schooleffect op PIRLS en op lezen bij de peiling Nederlands in het basisonderwijs

Tabel 94

Classificatie van scholen volgens hun type A-schooleffect op PIRLS en op lezen bij de peiling Nederlands in het basisonderwijs

	Peiling			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PIRLS				
Ondergemiddeld	-	1 (11,11%)	-	1 (11,11%)
Gemiddeld	2 (22,22%)	5 (55,56%)	1 (11,11%)	8 (88,89%)
Bovengemiddeld	-	-	-	-
Totaal	2 (22,22%)	6 (66,67%)	1 (11,11%)	9 (100%)

1.3 Type B-schooleffect



Figuur 73. Puntgrafiek van scholen volgens hun type B-schooleffect op PIRLS en op lezen de peiling Nederlands in het basisonderwijs

Tabel 95

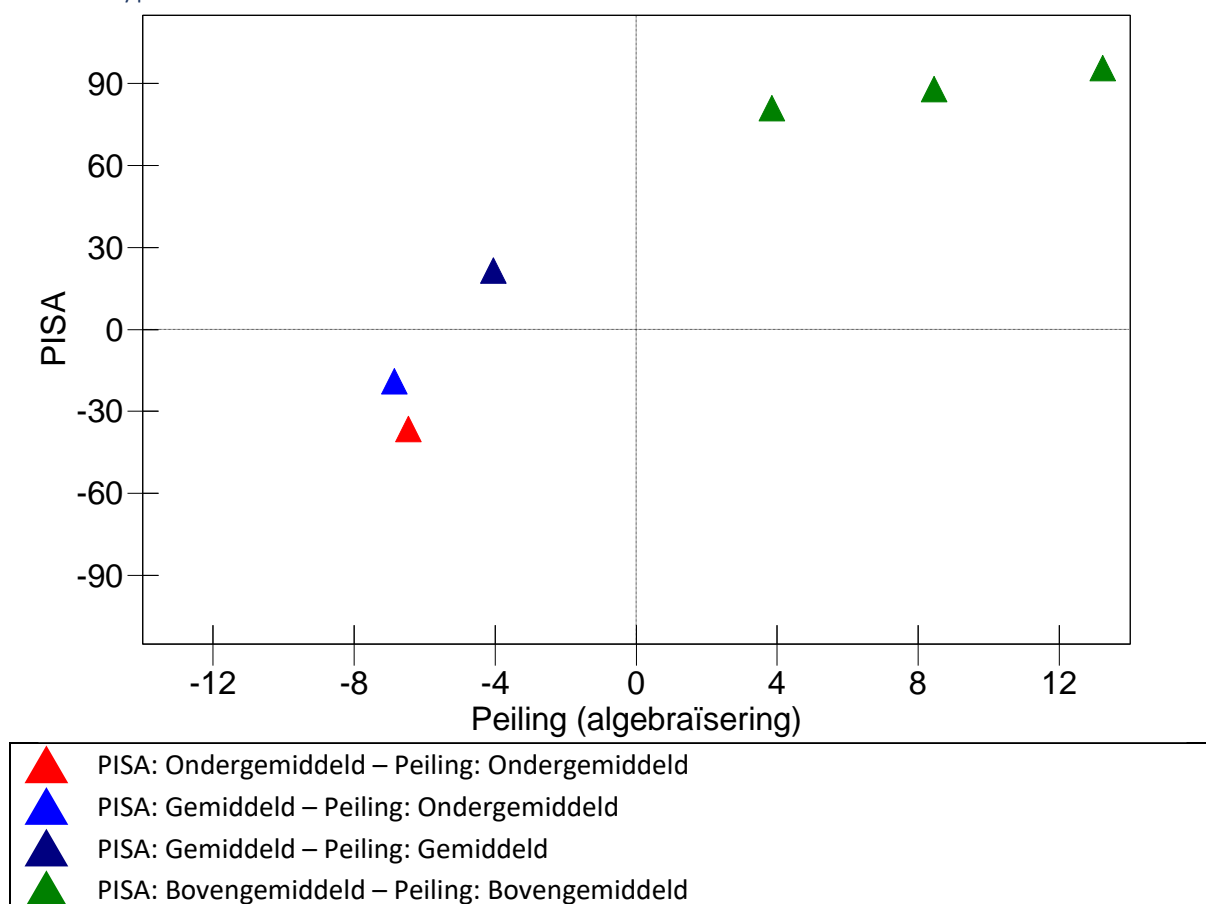
Classificatie van scholen volgens hun type B-schooleffect op PIRLS en op lezen bij de peiling Nederlands in het basisonderwijs

	Peiling			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PIRLS				
Ondergemiddeld	-	1 (11,11%)	-	1 (11,11%)
Gemiddeld	-	7 (77,78%)	1 (11,11%)	8 (88,89%)
Bovengemiddeld	-	-	-	-
Totaal	-	8 (88,89%)	1 (11,11%)	9 (100%)

2. Peiling wiskunde A-stroom 2018 & PISA 2018: 11 scholen

2.1 Algebraïsering (6 scholen)

2.1.1 Type 0-schooleffect



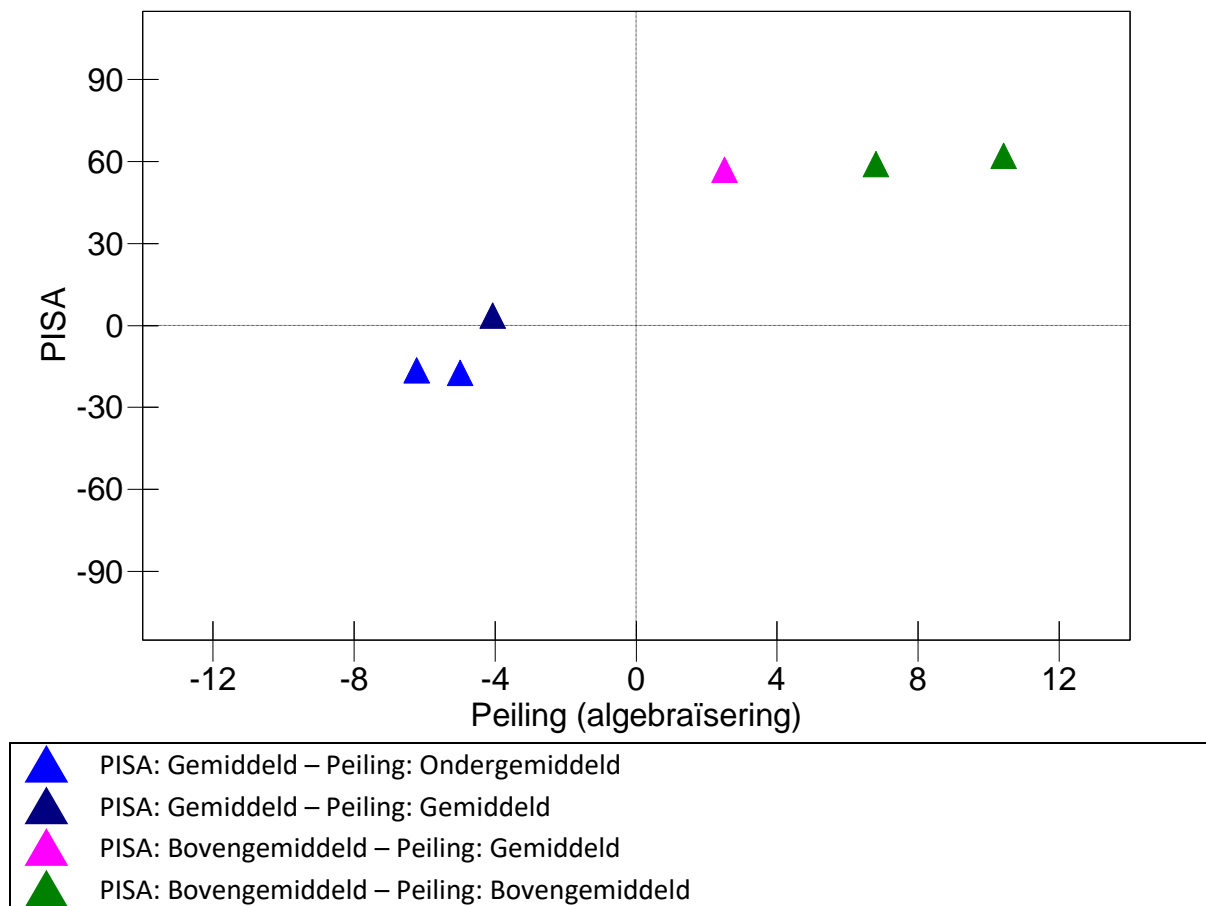
Figuur 74. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op algebraïsering bij de peiling wiskunde in de A-stroom

Tabel 96

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op algebraïsering bij de peiling wiskunde in de A-stroom

	Algebraïsering (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	1 (16,67%)	-	-	1 (16,67%)
Gemiddeld	1 (16,67%)	1 (16,67%)	-	2 (33,33%)
Bovengemiddeld	-	-	3 (50,00%)	3 (50,00%)
Totaal	2 (33,33%)	1 (16,67%)	3 (50,00%)	6 (100%)

2.1.2 Type A-schooleffect



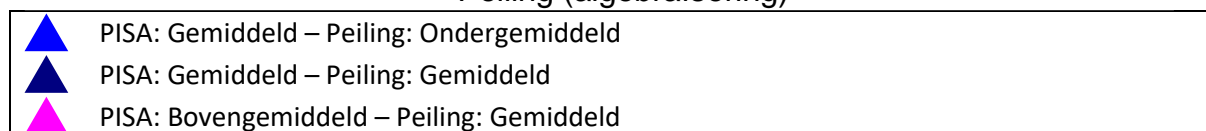
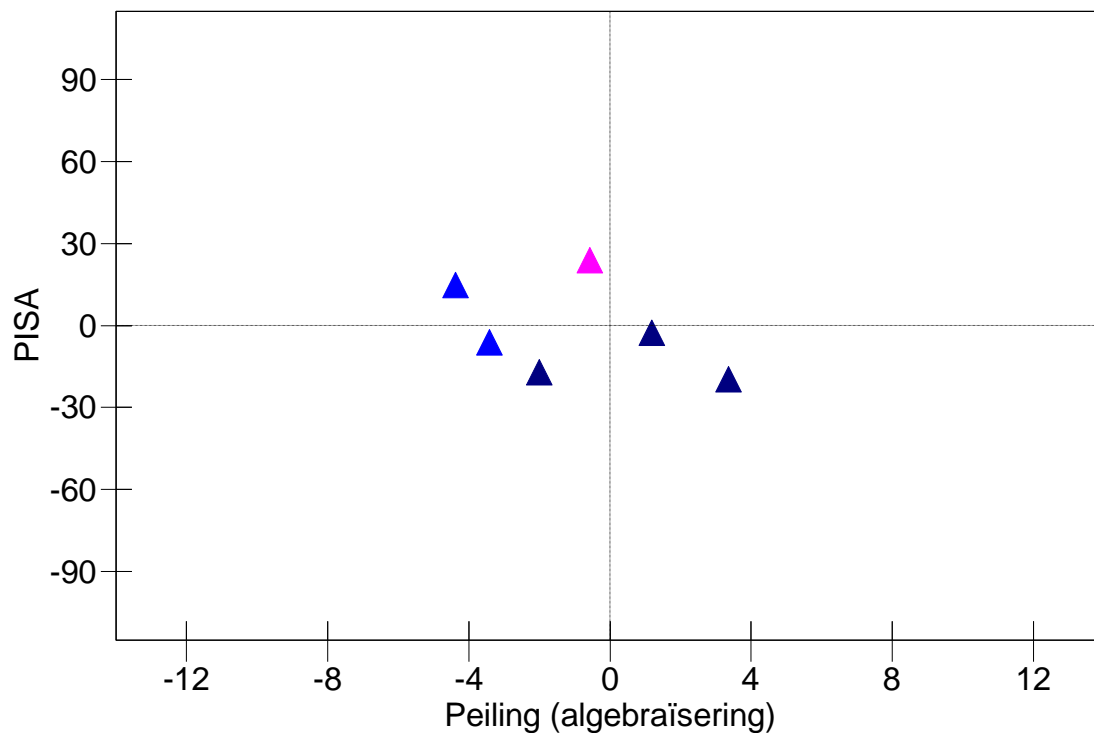
Figuur 75. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op algebraïsering bij de peiling wiskunde in de A-stroom

Tabel 97

Classificatie van scholen volgens hun type A-schooleffect op PISA en op algebraïsering bij de peiling wiskunde in de A-stroom

	Algebraïsering (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	-	-	-
Gemiddeld	2 (33,33%)	1 (16,67%)	-	3 (50,00%)
Bovengemiddeld	-	1 (16,67%)	2 (33,33%)	3 (50,00%)
Totaal	2 (33,33%)	2 (33,33%)	2 (33,33%)	6 (100%)

2.1.3 Type B-schooleffect



Figuur 76. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op algebraïsering bij de peiling wiskunde in de A-stroom

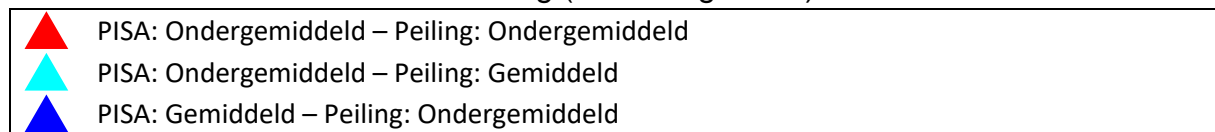
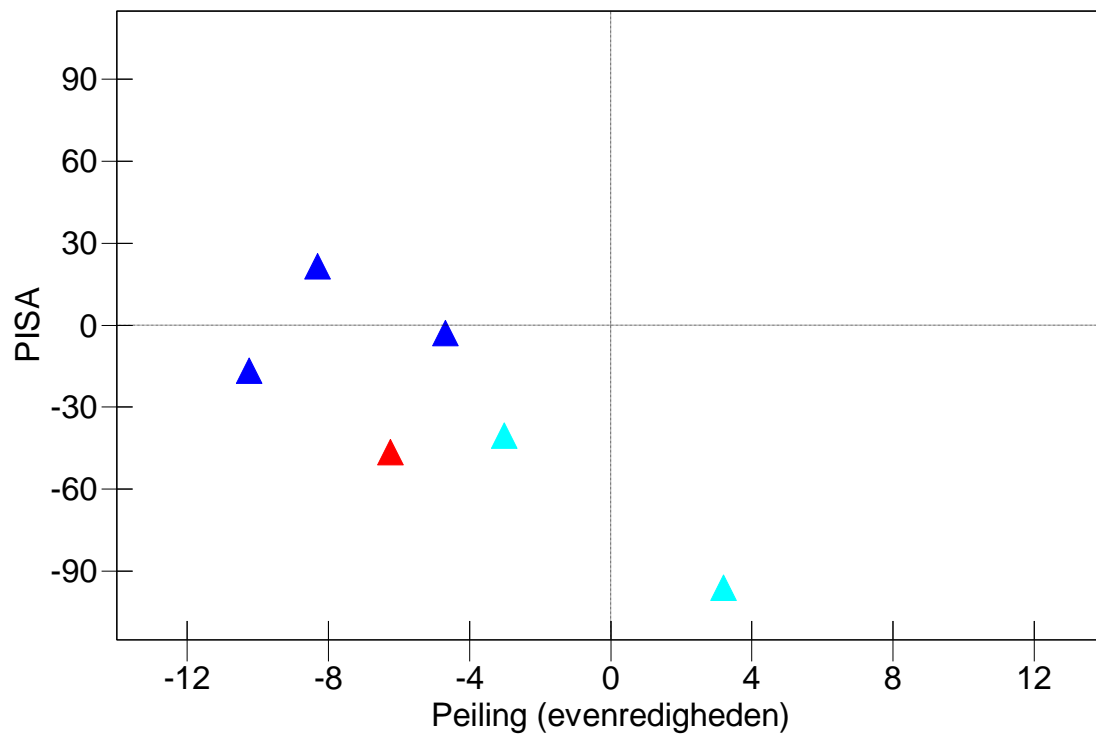
Tabel 98

Classificatie van scholen volgens hun type B-schooleffect op PISA en op algebraïsering bij de peiling wiskunde in de A-stroom

	Algebraïsering (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	-	-	-
Gemiddeld	2 (33,33%)	3 (50,00%)	-	5 (83,33%)
Bovengemiddeld	-	1 (16,67%)	-	1 (16,67%)
Totaal	2 (33,33%)	4 (66,67%)	-	6 (100%)

2.2 Evenredigheden (6 scholen)

2.2.1 Type 0-schooleffect



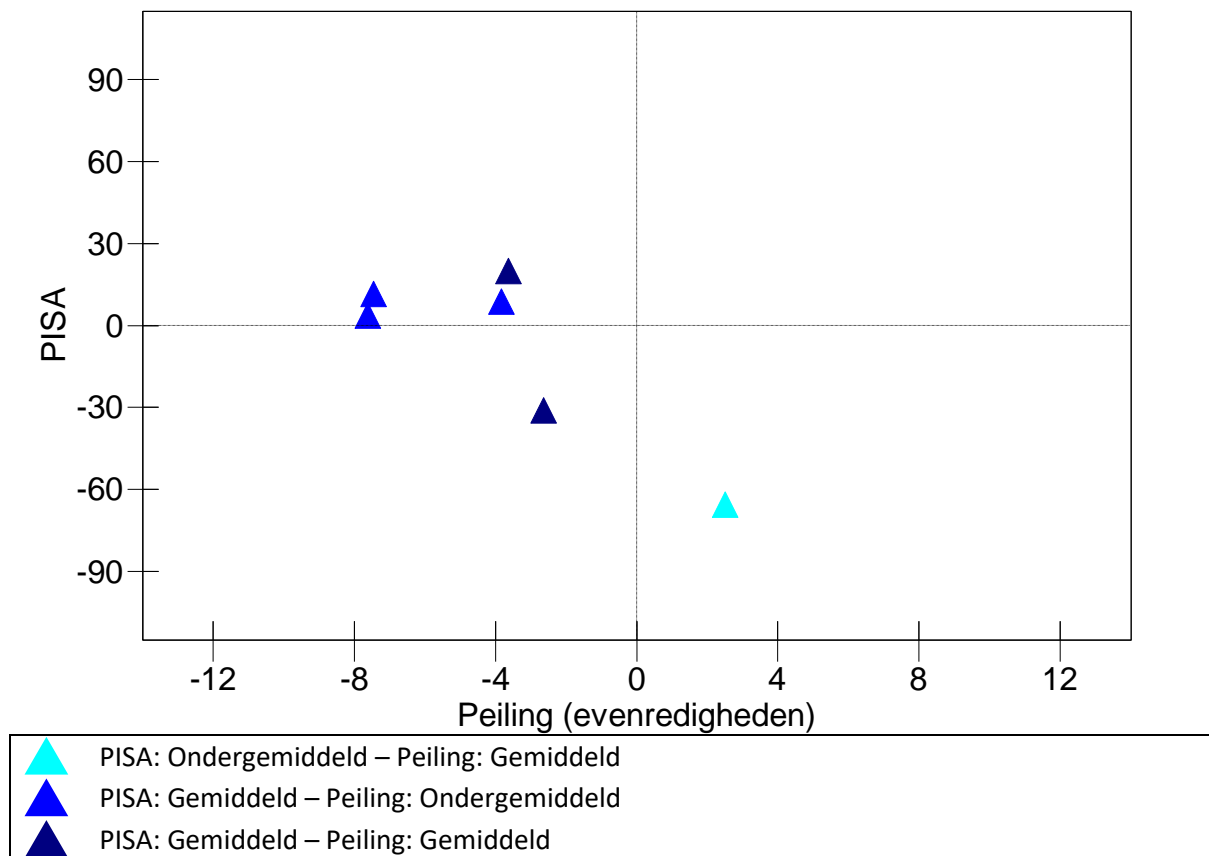
Figuur 77. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op evenredigheden bij de peiling wiskunde in de A-stroom

Tabel 99

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op evenredigheden bij de peiling wiskunde in de A-stroom

	Evenredigheden (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	1 (16,67%)	2 (33,33%)	-	3 (50,00%)
Gemiddeld	3 (50,00%)	-	-	3 (50,00%)
Bovengemiddeld	-	-	-	-
Totaal	4 (66,67%)	2 (33,33%)	-	6 (100%)

2.2.2 Type A-schooleffect



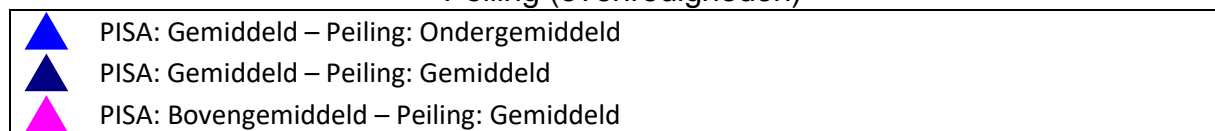
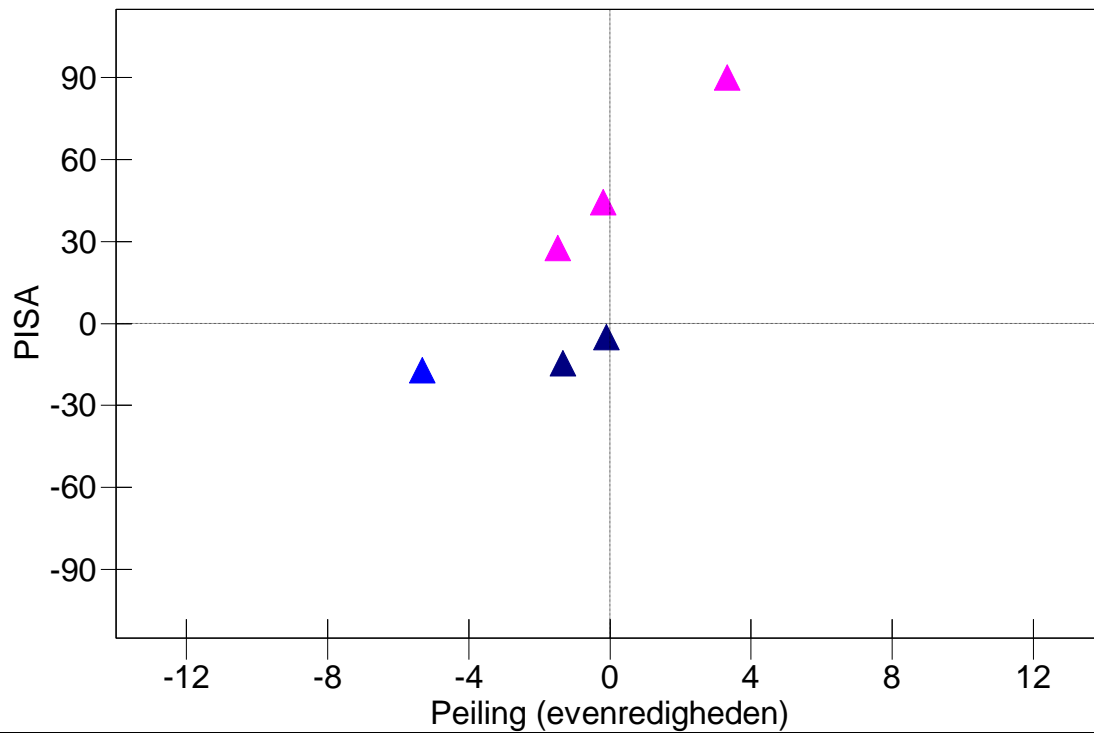
Figuur 78. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op evenredigheden bij de peiling wiskunde in de A-stroom

Tabel 100

Classificatie van scholen volgens hun type A-schooleffect op PISA en op evenredigheden bij de peiling wiskunde in de A-stroom

	Evenredigheden (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (16,67%)	-	1 (16,67%)
Gemiddeld	3 (50,00%)	2 (33,33%)	-	5 (83,33%)
Bovengemiddeld	-	-	-	-
Totaal	3 (50,00%)	3 (50,00%)	-	6 (100%)

2.2.3 Type B-schooleffect



Figuur 79. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op evenredigheden bij de peiling wiskunde in de A-stroom

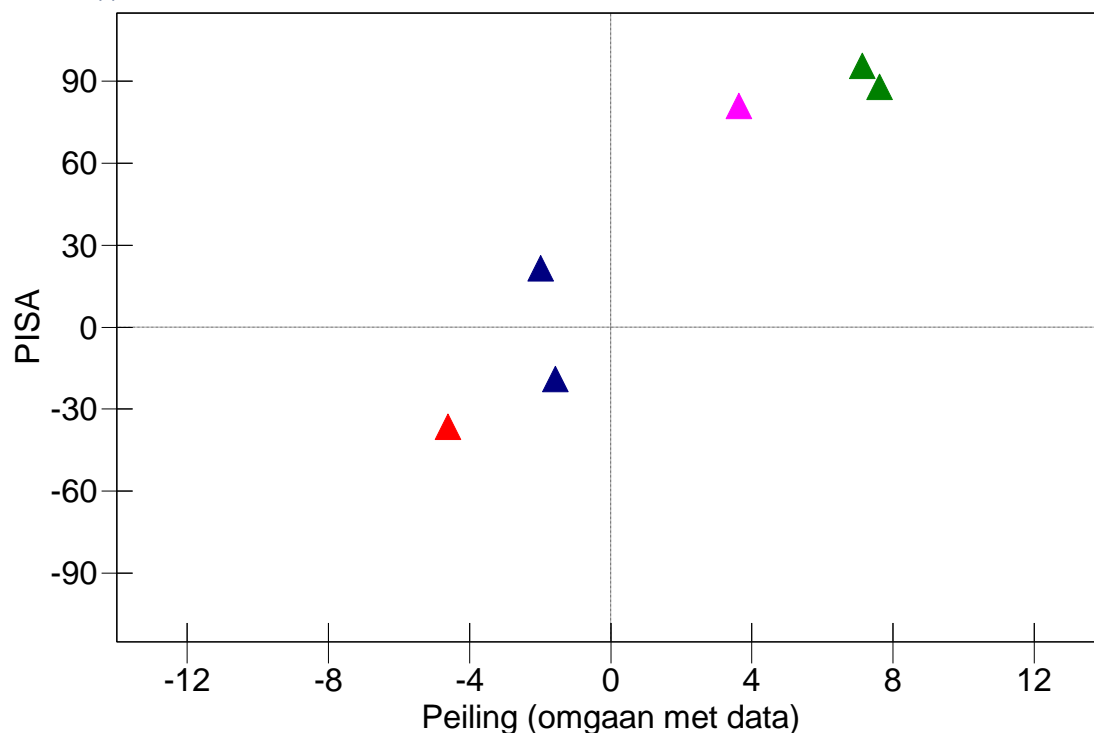
Tabel 101

Classificatie van scholen volgens hun type B-schooleffect op PISA en op evenredigheden bij de peiling wiskunde in de A-stroom

	Evenredigheden (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	-	-	-
Gemiddeld	1 (16,67%)	2 (33,33%)	-	3 (50,00%)
Bovengemiddeld	-	3 (50,00%)	-	3 (50,00%)
Totaal	1 (16,67%)	5 (83,33%)	-	6 (100%)

2.3 Omgaan met data (6 scholen)

2.3.1 Type 0-schooleffect



	PISA: Ondergemiddeld – Peiling: Ondergemiddeld
	PISA: Gemiddeld – Peiling: Gemiddeld
	PISA: Bovengemiddeld – Peiling: Gemiddeld
	PISA: Bovengemiddeld – Peiling: Bovengemiddeld

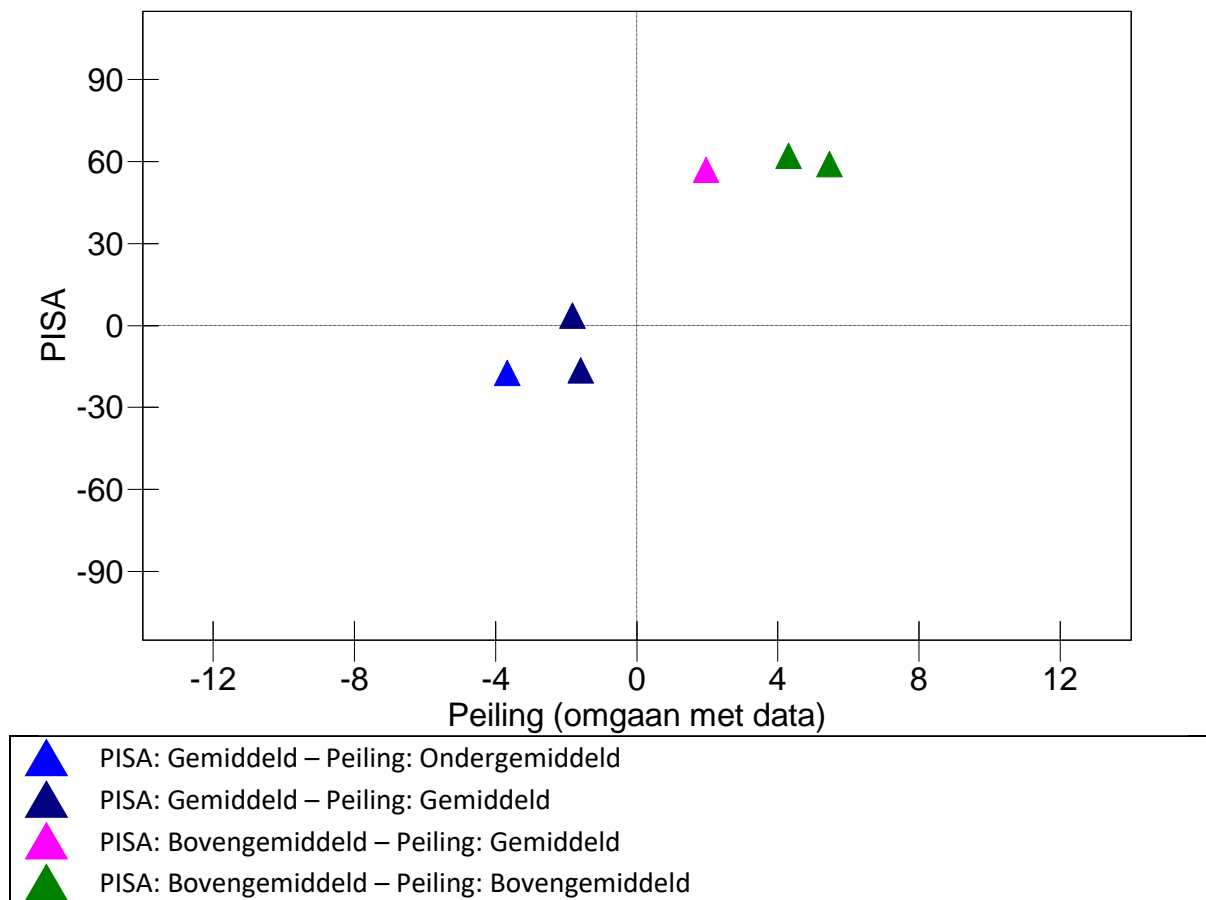
Figuur 80. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op omgaan met data bij de peiling wiskunde in de A-stroom

Tabel 102

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op omgaan met data bij de peiling wiskunde in de A-stroom

	Omgaan met data (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	1 (16,67%)	-	-	1 (16,67%)
Gemiddeld	-	2 (33,33%)	-	2 (33,33%)
Bovengemiddeld	-	1 (16,67%)	2 (33,33%)	3 (50,00%)
Totaal	1 (16,67%)	3 (50,00%)	2 (33,33%)	6 (100%)

2.3.2 Type A-schooleffect



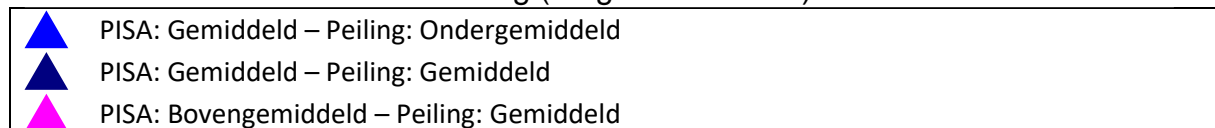
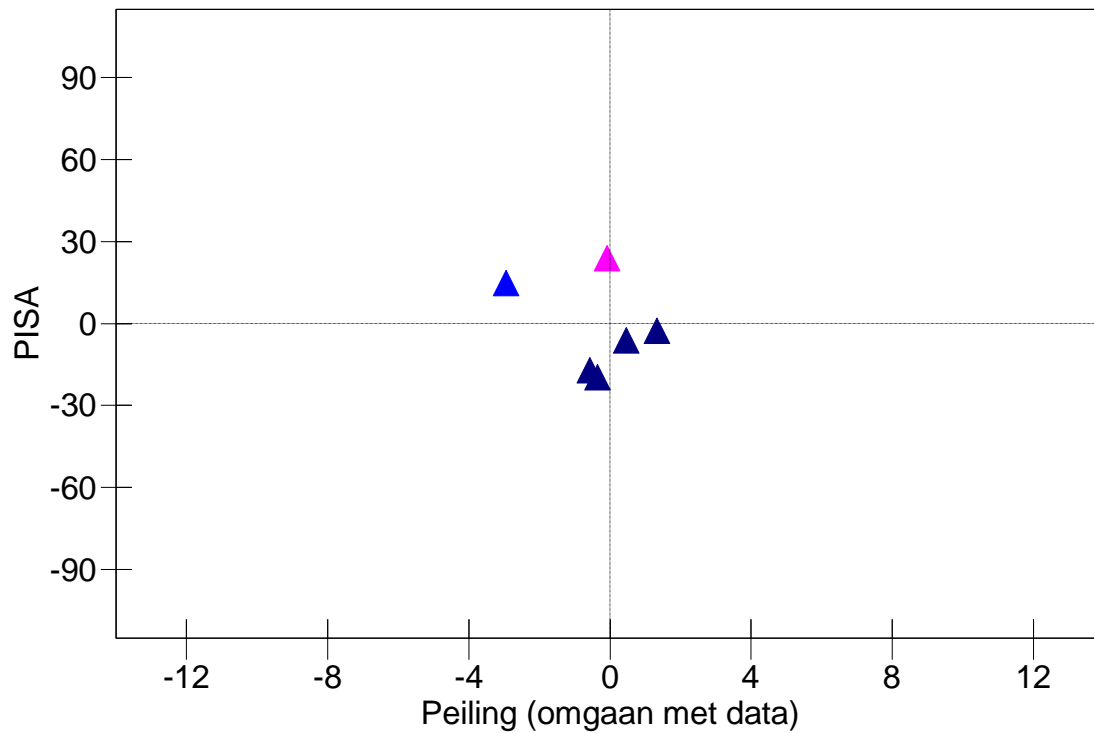
Figuur 81. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op omgaan met data bij de peiling wiskunde in de A-stroom

Tabel 103

Classificatie van scholen volgens hun type A-schooleffect op PISA en op omgaan met data bij de peiling wiskunde in de A-stroom

	Omgaan met data (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	-	-	-
Gemiddeld	1 (16,67%)	2 (33,33%)	-	3 (50,00%)
Bovengemiddeld	-	1 (16,67%)	2 (33,33%)	3 (50,00%)
Totaal	1 (16,67%)	3 (50,00%)	2 (33,33%)	6 (100%)

2.3.3 Type B-schooleffect



Figuur 82. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op omgaan met data bij de peiling wiskunde in de A-stroom

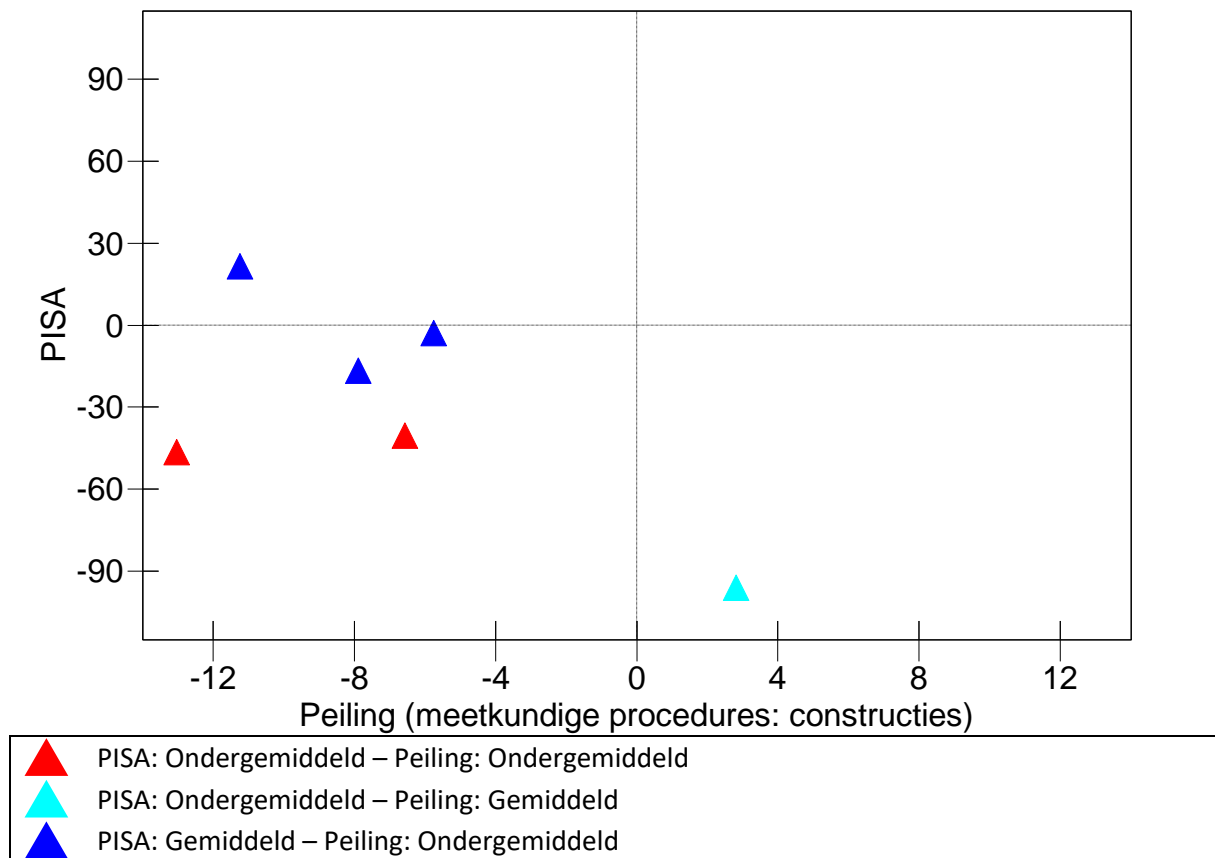
Tabel 104

Classificatie van scholen volgens hun type B-schooleffect op PISA en op omgaan met data bij de peiling wiskunde in de A-stroom

	Omgaan met data (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	-	-	-
Gemiddeld	1 (16,67%)	4 (66,67%)	-	5 (83,33%)
Bovengemiddeld	-	1 (16,67%)	-	1 (16,67%)
Totaal	1 (16,67%)	5 (83,33%)	-	6 (100%)

2.4 Meetkundige procedures: constructies (6 scholen)

2.4.1 Type 0-schooleffect



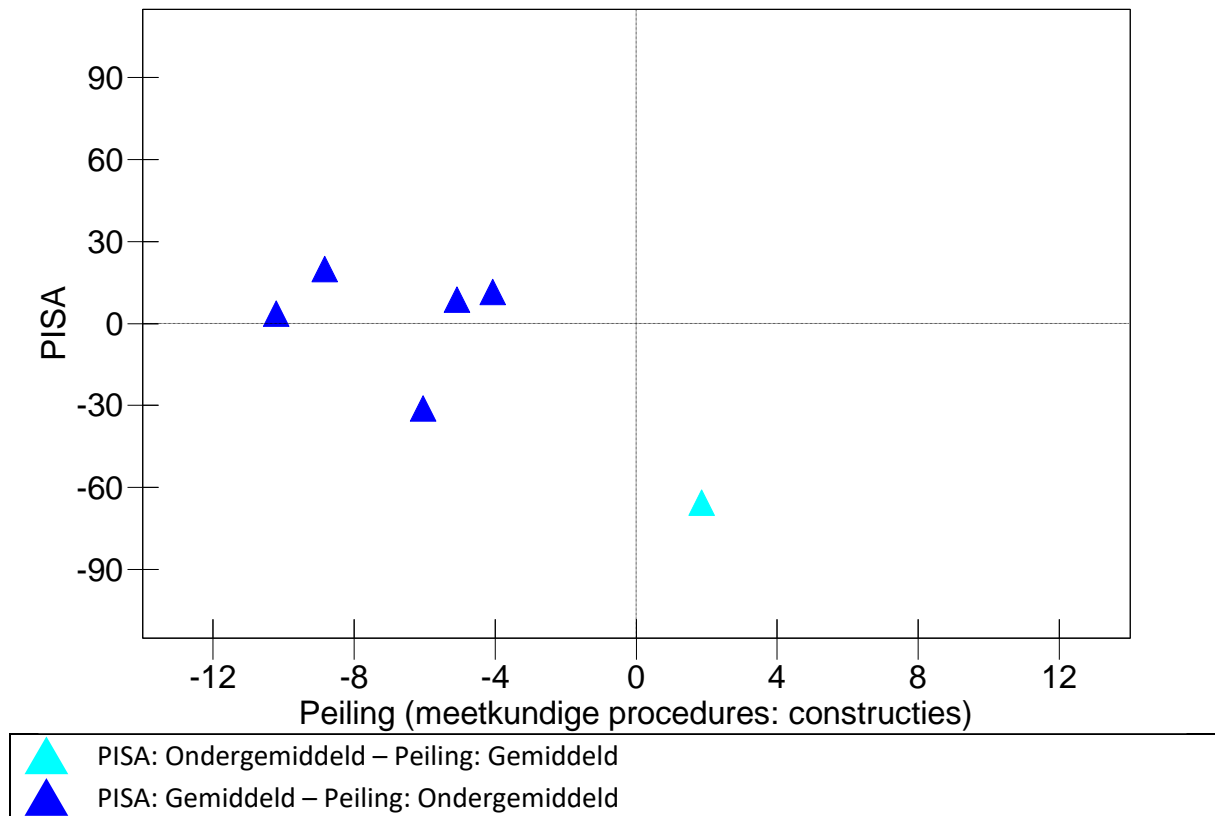
Figuur 83. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op meetkundige procedures: constructies bij de peiling wiskunde in de A-stroom

Tabel 105

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op meetkundige procedures: constructies bij de peiling wiskunde in de A-stroom

	Meetkundige procedures: constructies (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	2 (33,33%)	1 (16,67%)	-	3 (50,00%)
Gemiddeld	3 (50,00%)	-	-	3 (50,00%)
Bovengemiddeld	-	-	-	-
Totaal	5 (83,33%)	1 (16,67%)	-	6 (100%)

2.4.2 Type A-schooleffect



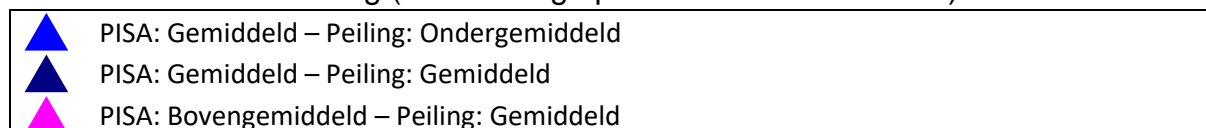
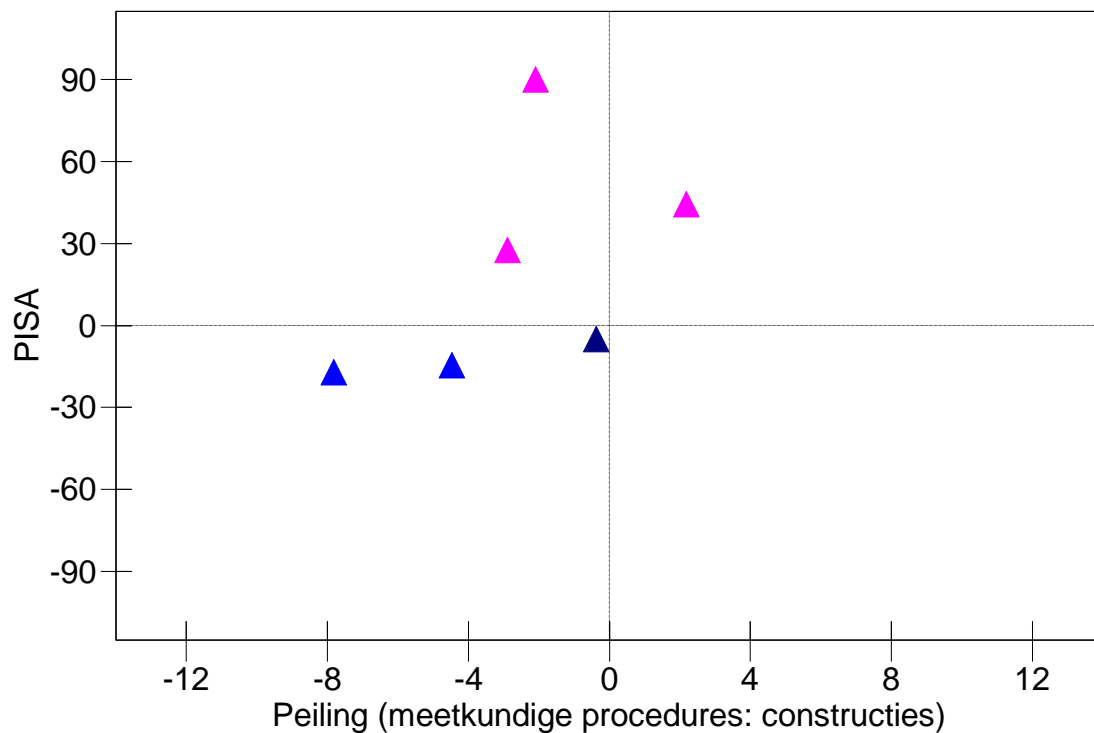
Figuur 84. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op meetkundige procedures: constructies bij de peiling wiskunde in de A-stroom

Tabel 106

Classificatie van scholen volgens hun type A-schooleffect op PISA en op meetkundige procedures: constructies bij de peiling wiskunde in de A-stroom

	Meetkundige procedures: constructies (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (16,67%)	-	1 (16,67%)
Gemiddeld	5 (83,33%)	-	-	5 (83,33%)
Bovengemiddeld	-	-	-	-
Totaal	5 (83,33%)	1 (16,67%)	-	6 (100%)

2.4.3 Type B-schooleffect



Figuur 85. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op meetkundige procedures: constructies bij de peiling wiskunde in de A-stroom

Tabel 107

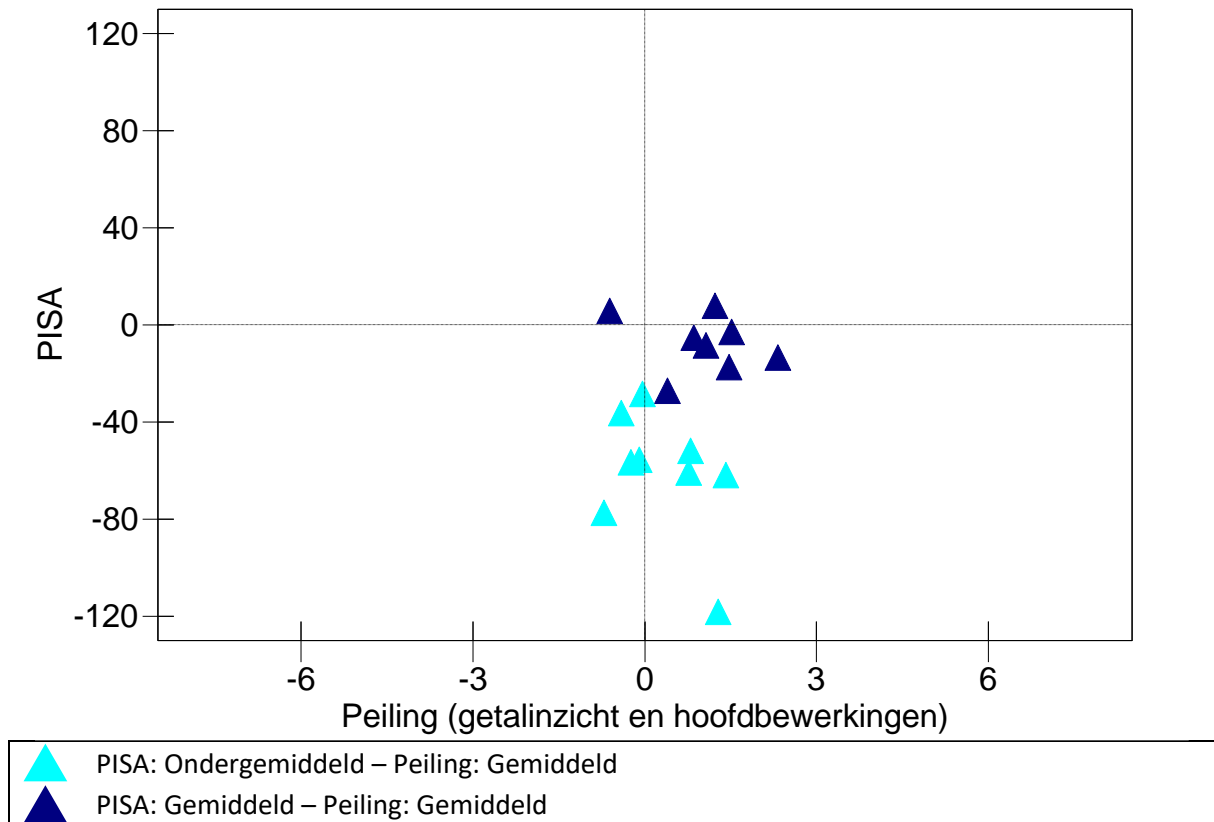
Classificatie van scholen volgens hun type B-schooleffect op PISA en op meetkundige procedures: constructies bij de peiling wiskunde in de A-stroom

	Meetkundige procedures: constructies (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	-	-	-
Gemiddeld	2 (33,33%)	1 (16,67%)	-	3 (50,00%)
Bovengemiddeld	-	-	3 (50,00%)	3 (50,00%)
Totaal	2 (33,33%)	1 (16,67%)	3 (50,00%)	6 (100%)

3. Peiling wiskunde B-stroom 2019 & PISA 2018: 20 scholen

3.1 Getalinzicht en hoofdbewerkingen (17 scholen)

3.1.1 Type 0-schooleffect



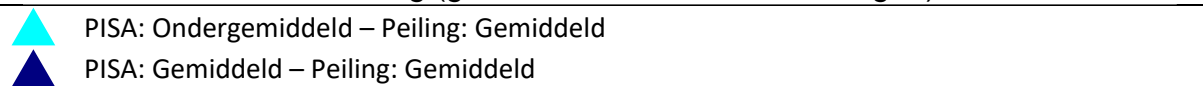
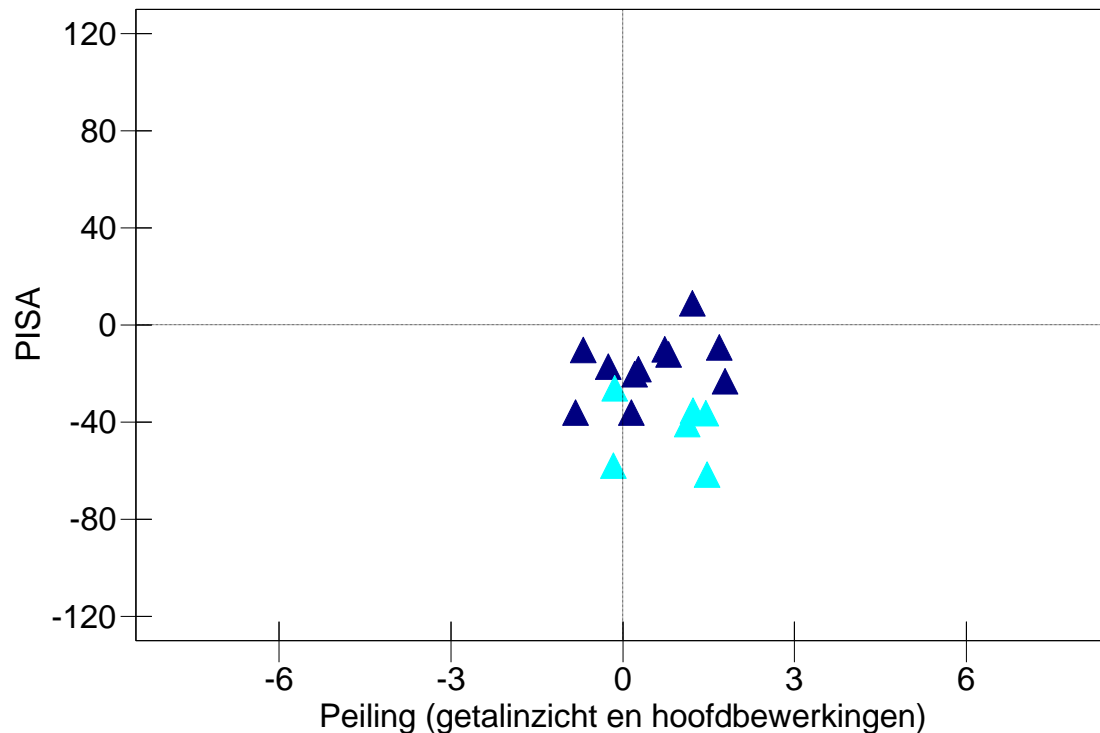
Figuur 86. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op getalinzicht en hoofdbewerkingen bij de peiling wiskunde in de B-stroom

Tabel 108

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op getalinzicht en hoofdbewerkingen bij de peiling wiskunde in de B-stroom

	Getalinzicht en hoofdbewerkingen (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	9 (52,94%)	-	9 (52,94%)
Gemiddeld	-	8 (47,06%)	-	8 (47,06%)
Bovengemiddeld	-	-	-	-
Totaal	-	17 (100%)	-	17 (100%)

3.1.2 Type A-schooleffect



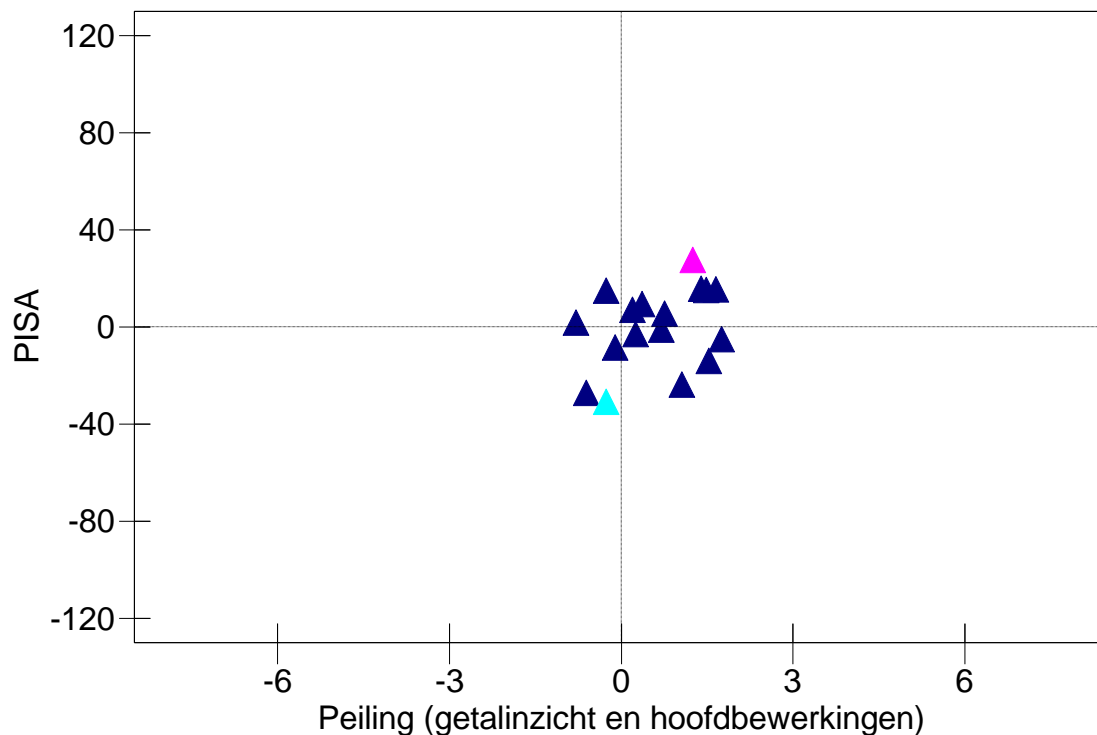
Figuur 87. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op getalinzicht en hoofdbewerkingen bij de peiling wiskunde in de B-stroom

Tabel 109

Classificatie van scholen volgens hun type A-schooleffect op PISA en op getalinzicht en hoofdbewerkingen bij de peiling wiskunde in de B-stroom

	Getalinzicht en hoofdbewerkingen (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	6 (35,29%)	-	6 (35,29%)
Gemiddeld	-	11 (64,71%)	-	11 (64,71%)
Bovengemiddeld	-	-	-	-
Totaal	-	17 (100%)	-	17 (100%)

3.1.3 Type B-schooleffect



	PISA: Ondergemiddeld – Peiling: Gemiddeld
	PISA: Gemiddeld – Peiling: Gemiddeld
	PISA: Bovengemiddeld – Peiling: Gemiddeld

Figuur 88. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op getalinzicht en hoofdbewerkingen bij de peiling wiskunde in de B-stroom

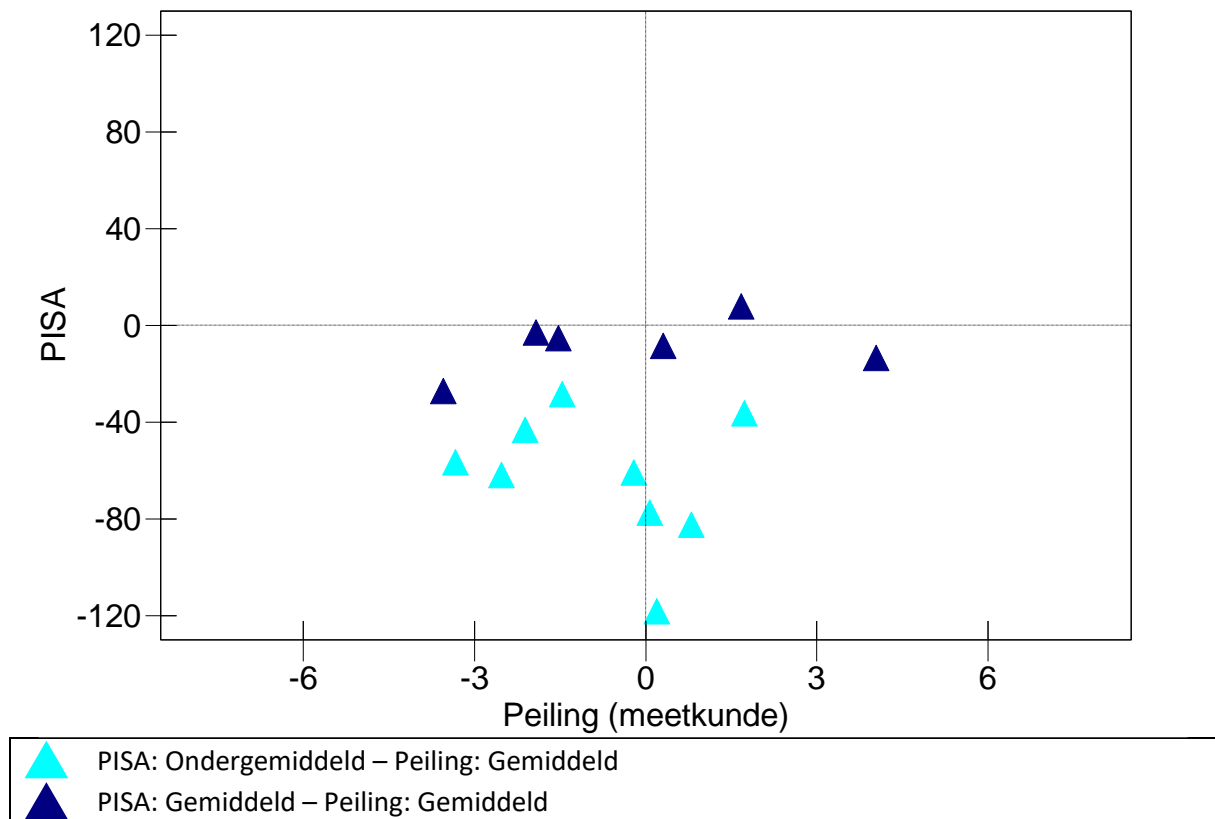
Tabel 110

Classificatie van scholen volgens hun type B-schooleffect op PISA en op getalinzicht en hoofdbewerkingen bij de peiling wiskunde in de B-stroom

	Getalinzicht en hoofdbewerkingen (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (5,88%)	-	1 (5,88%)
Gemiddeld	-	15 (88,24%)	-	15 (88,24%)
Bovengemiddeld	-	1 (5,88%)	-	1 (5,88%)
Totaal	-	17 (100%)	-	17 (100%)

3.2 Meetkunde (15 scholen)

3.2.1 Type 0-schooleffect



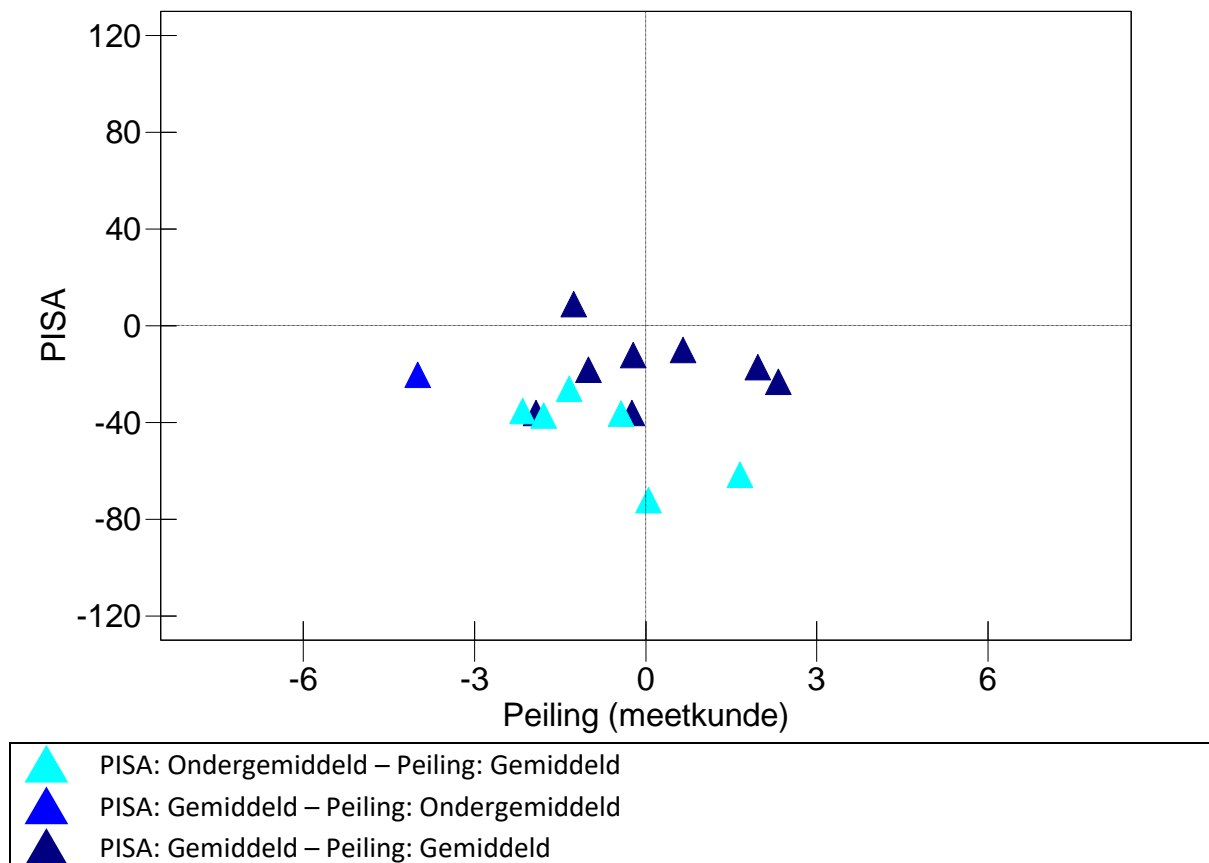
Figuur 89. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op meetkunde bij de peiling wiskunde in de B-stroom

Tabel 111

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op meetkunde bij de peiling wiskunde in de B-stroom

	Meetkunde (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	9 (60,00%)	-	9 (60,00%)
Gemiddeld	-	6 (40,00%)	-	6 (40,00%)
Bovengemiddeld	-	-	-	-
Totaal	-	15 (100%)	-	15 (100%)

3.2.2 Type A-schooleffect



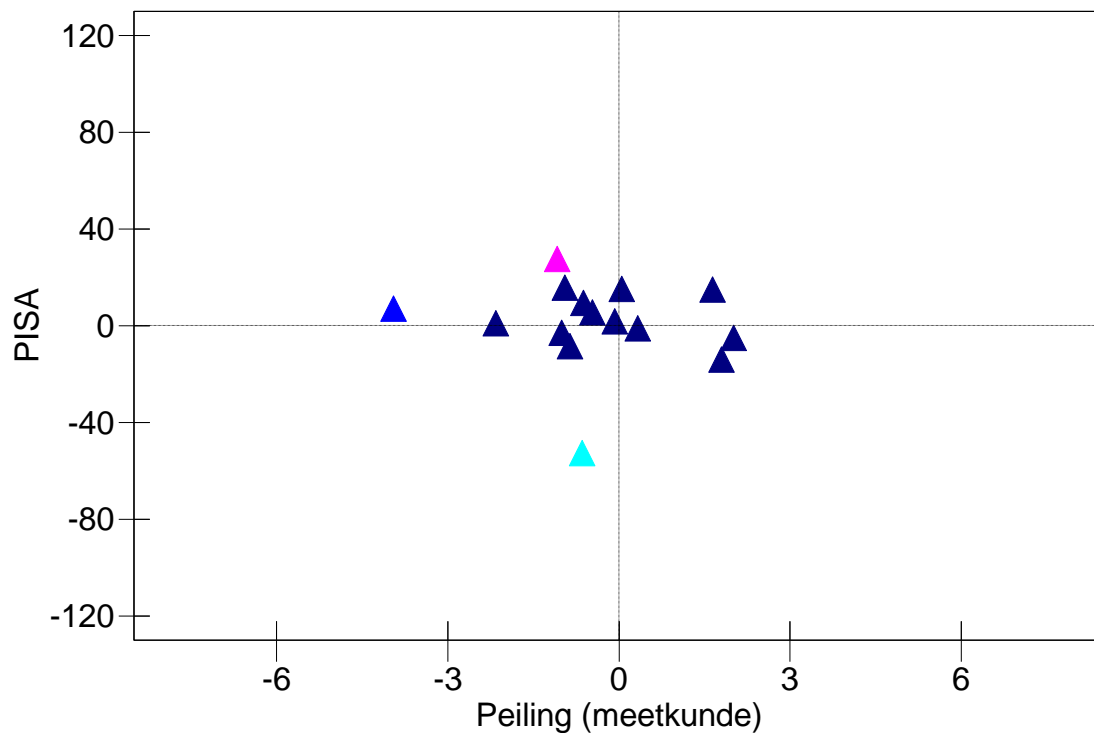
Figuur 90. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op meetkunde bij de peiling wiskunde in de B-stroom

Tabel 112

Classificatie van scholen volgens hun type A-schooleffect op PISA en op meetkunde bij de peiling wiskunde in de B-stroom

	Meetkunde (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	6 (40,00%)	-	6 (40,00%)
Gemiddeld	1 (6,67%)	8 (53,33%)	-	9 (60,00%)
Bovengemiddeld	-	-	-	-
Totaal	1 (6,67%)	14 (93,33%)	-	15 (100%)

3.2.3 Type B-schooleffect



Figuur 91. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op meetskunde bij de peiling wiskunde in de B-stroom

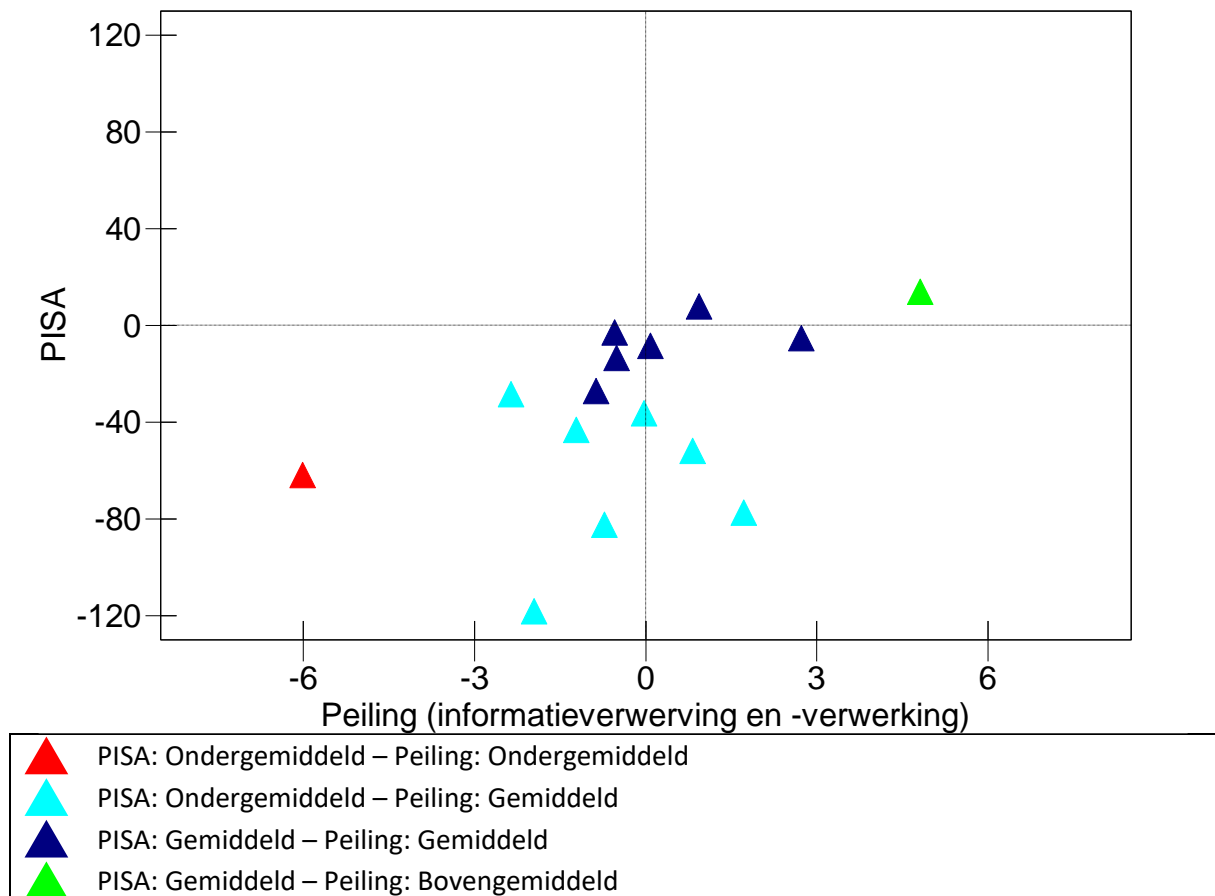
Tabel 113

Classificatie van scholen volgens hun type B-schooleffect op PISA en op meetskunde bij de peiling wiskunde in de B-stroom

	Meetskunde (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (6,67%)	-	1 (6,67%)
Gemiddeld	1 (6,67%)	12 (80,00%)	-	13 (86,67%)
Bovengemiddeld	-	1 (6,67%)	-	1 (6,67%)
Totaal	1 (6,67%)	14 (93,33%)	-	15 (100%)

3.3 Informatieverwerking en -verwerking (15 scholen)

3.3.1 Type 0-schooleffect



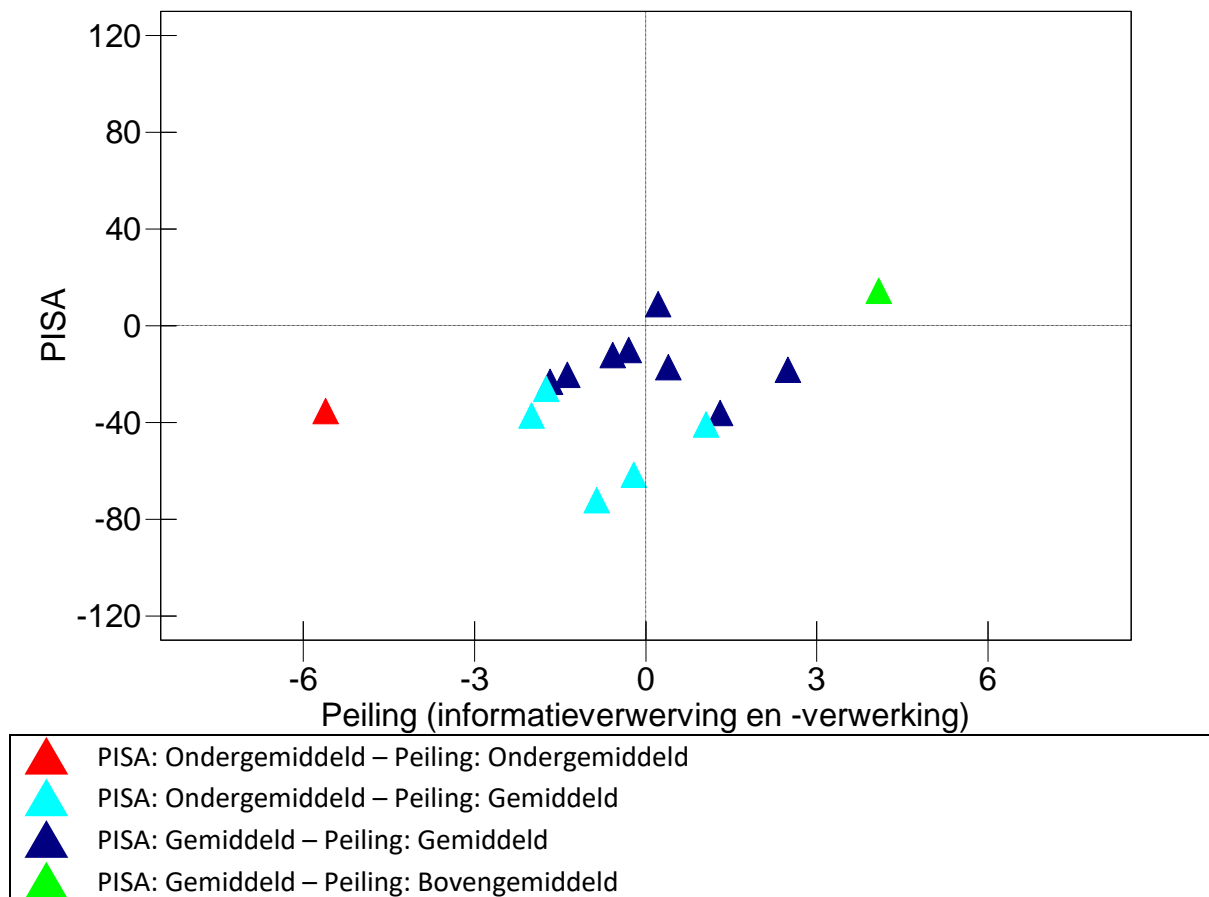
Figuur 92. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op informatieverwerking en -verwerking bij de peiling wiskunde in de B-stroom

Tabel 114

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op informatieverwerking en -verwerking bij de peiling wiskunde in de B-stroom

	Informatieverwerking en -verwerking (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	1 (6,67%)	7 (46,67%)	-	8 (53,33%)
Gemiddeld	-	6 (40,00%)	1 (6,67%)	7 (46,67%)
Bovengemiddeld	-	-	-	-
Totaal	1 (6,67%)	13 (86,67%)	1 (6,67%)	15 (100%)

3.3.2 Type A-schooleffect



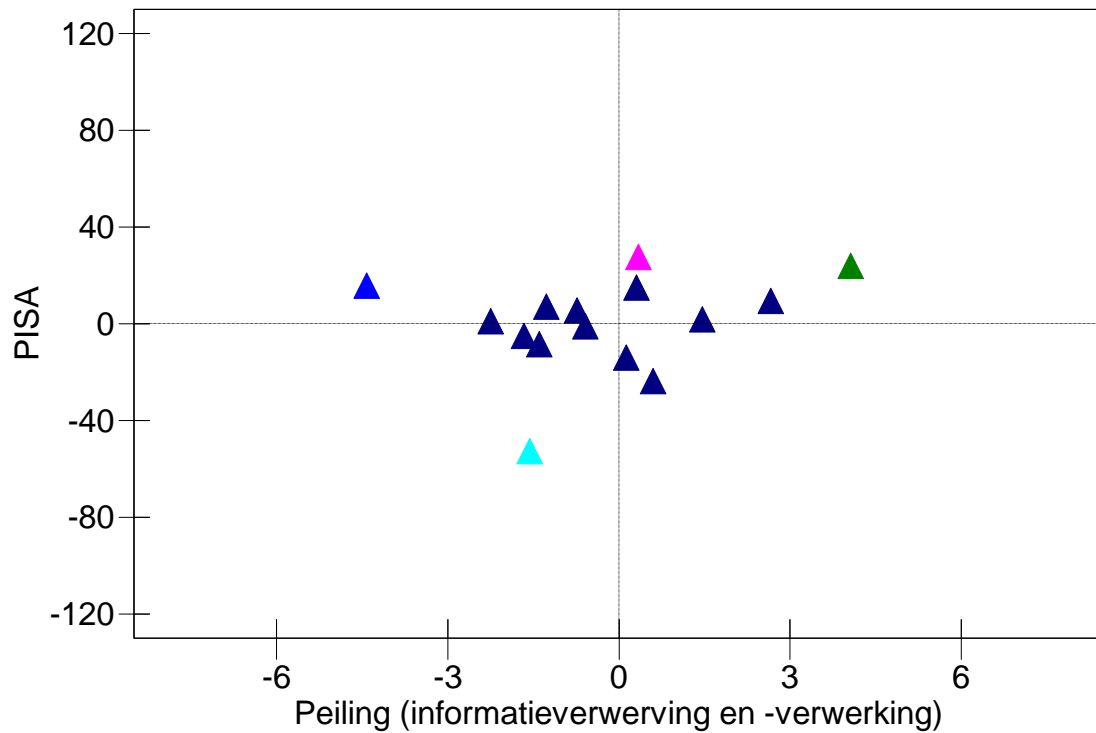
Figuur 93. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op informatieverwerking en -verwerking bij de peiling wiskunde in de B-stroom

Tabel 115

Classificatie van scholen volgens hun type A-schooleffect op PISA en op informatieverwerking en -verwerking bij de peiling wiskunde in de B-stroom

	Informatieverwerking en -verwerking (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	1 (6,67%)	5 (33,33%)	-	6 (40,00%)
Gemiddeld	-	8 (53,33%)	1 (6,67%)	9 (60,00%)
Bovengemiddeld	-	-	-	-
Totaal	1 (6,67%)	13 (86,67%)	1 (6,67%)	15 (100%)

3.3.3 Type B-schooleffect



	PISA: Ondergemiddeld – Peiling: Gemiddeld
	PISA: Gemiddeld – Peiling: Ondergemiddeld
	PISA: Gemiddeld – Peiling: Gemiddeld
	PISA: Bovengemiddeld – Peiling: Gemiddeld
	PISA: Bovengemiddeld – Peiling: Bovengemiddeld

Figuur 94. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op informatieverwerking en -verwerking bij de peiling wiskunde in de B-stroom

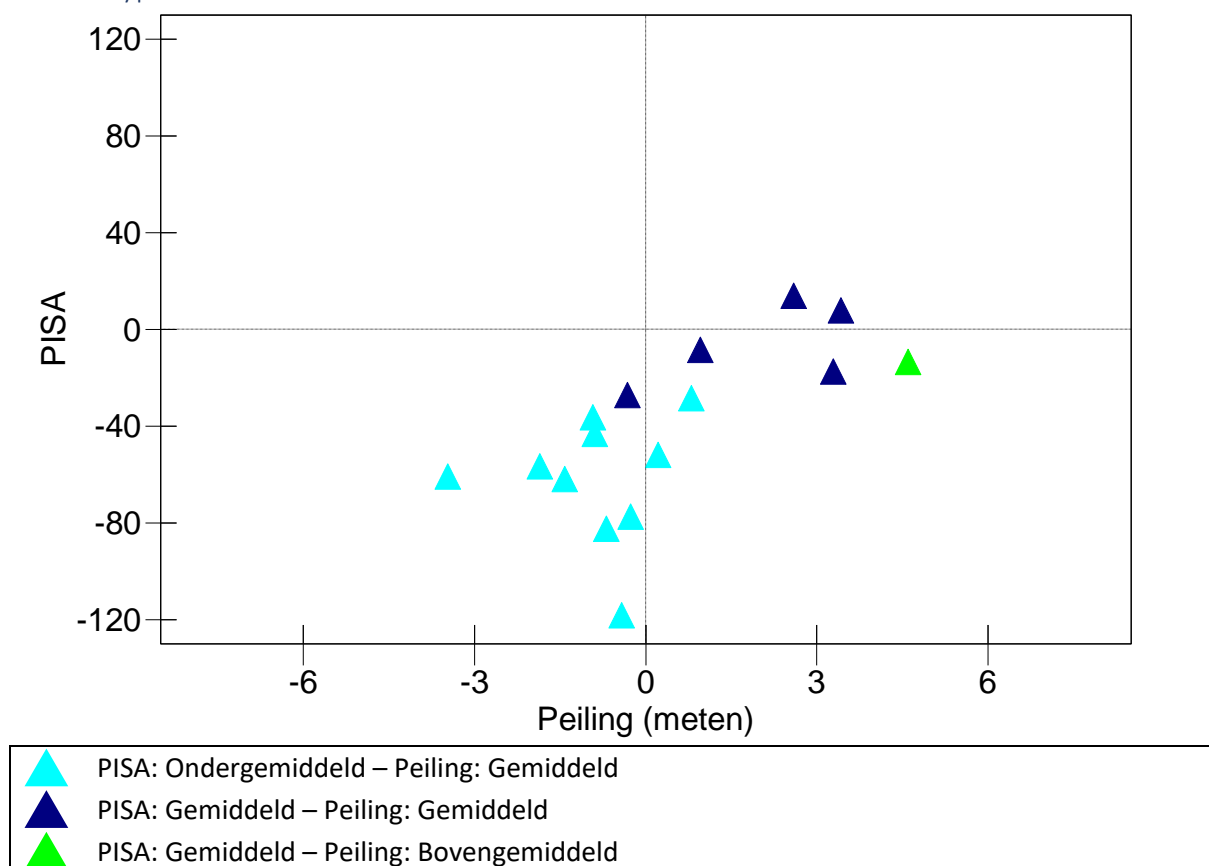
Tabel 116

Classificatie van scholen volgens hun type B-schooleffect op PISA en op informatieverwerking en -verwerking bij de peiling wiskunde in de B-stroom

	Informatieverwerking en -verwerking (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (6,67%)	-	1 (6,67%)
Gemiddeld	1 (6,67%)	11 (76,66%)	-	12 (80,00%)
Bovengemiddeld	-	1 (6,67%)	1 (6,67%)	2 (13,33%)
Totaal	1 (6,67%)	13 (86,67%)	1 (6,67%)	15 (100%)

3.4 Meten (16 scholen)

3.4.1 Type 0-schooleffect



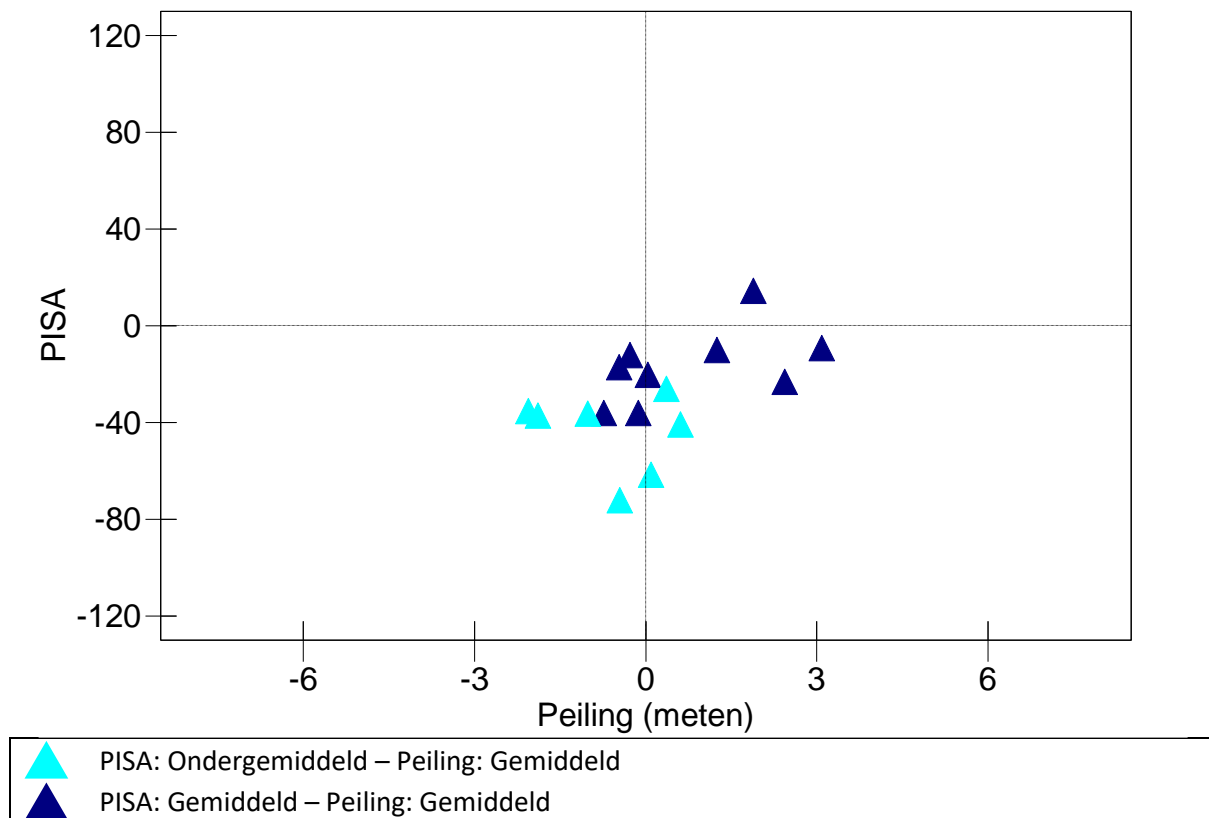
Figuur 95. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op meten bij de peiling wiskunde in de B-stroom

Tabel 117

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op meten bij de peiling wiskunde in de B-stroom

	Meten (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	10 (62,50%)	-	10 (62,50%)
Gemiddeld	-	5 (31,25%)	1 (6,25%)	6 (37,50%)
Bovengemiddeld	-	-	-	-
Totaal	-	15 (93,75%)	1 (6,25%)	16 (100%)

3.4.2 Type A-schooleffect



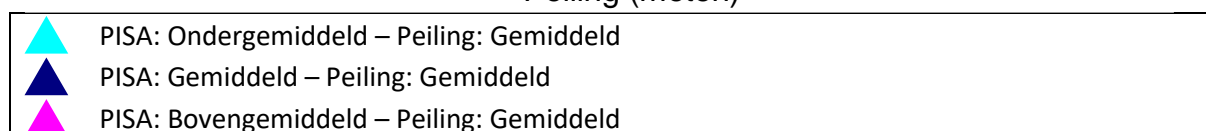
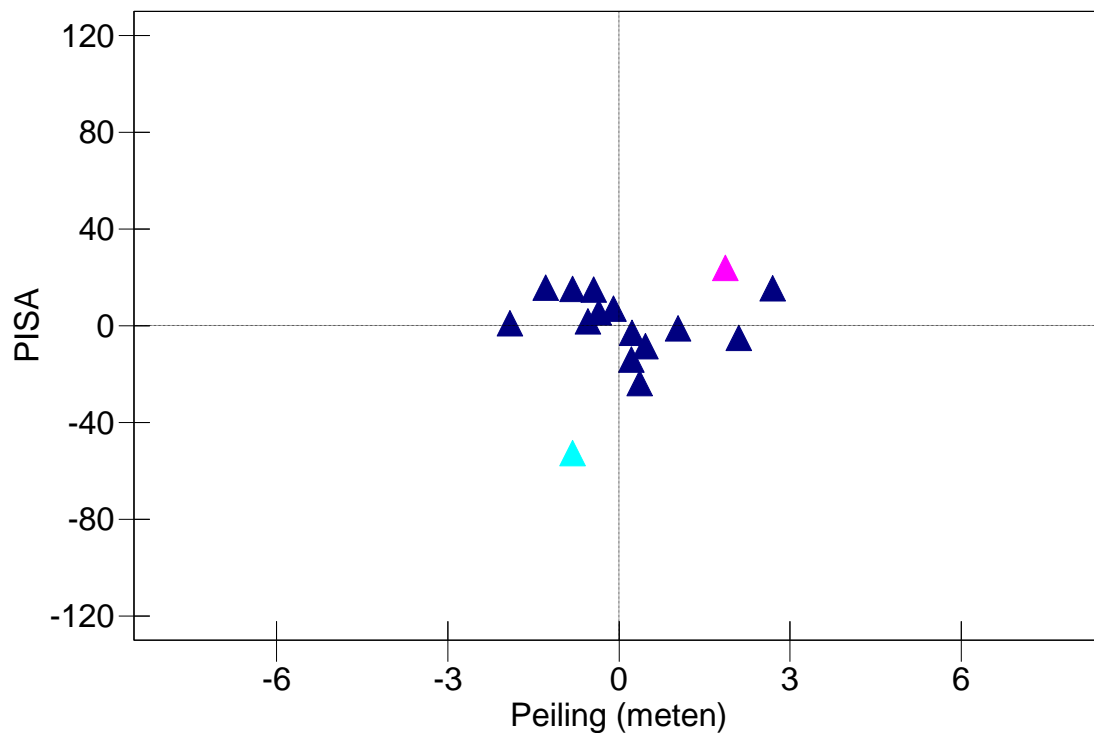
Figuur 96. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op meten bij de peiling wiskunde in de B-stroom

Tabel 118

Classificatie van scholen volgens hun type A-schooleffect op PISA en op meten bij de peiling wiskunde in de B-stroom

	Meten (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	7 (43,75%)	-	7 (43,75%)
Gemiddeld	-	9 (56,25%)	-	9 (56,25%)
Bovengemiddeld	-	-	-	-
Totaal	-	16 (100%)	-	16 (100%)

3.4.3 Type B-schooleffect



Figuur 97. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op meten bij de peiling wiskunde in de B-stroom

Tabel 119

Classificatie van scholen volgens hun type B-schooleffect op PISA en op meten bij de peiling wiskunde in de B-stroom

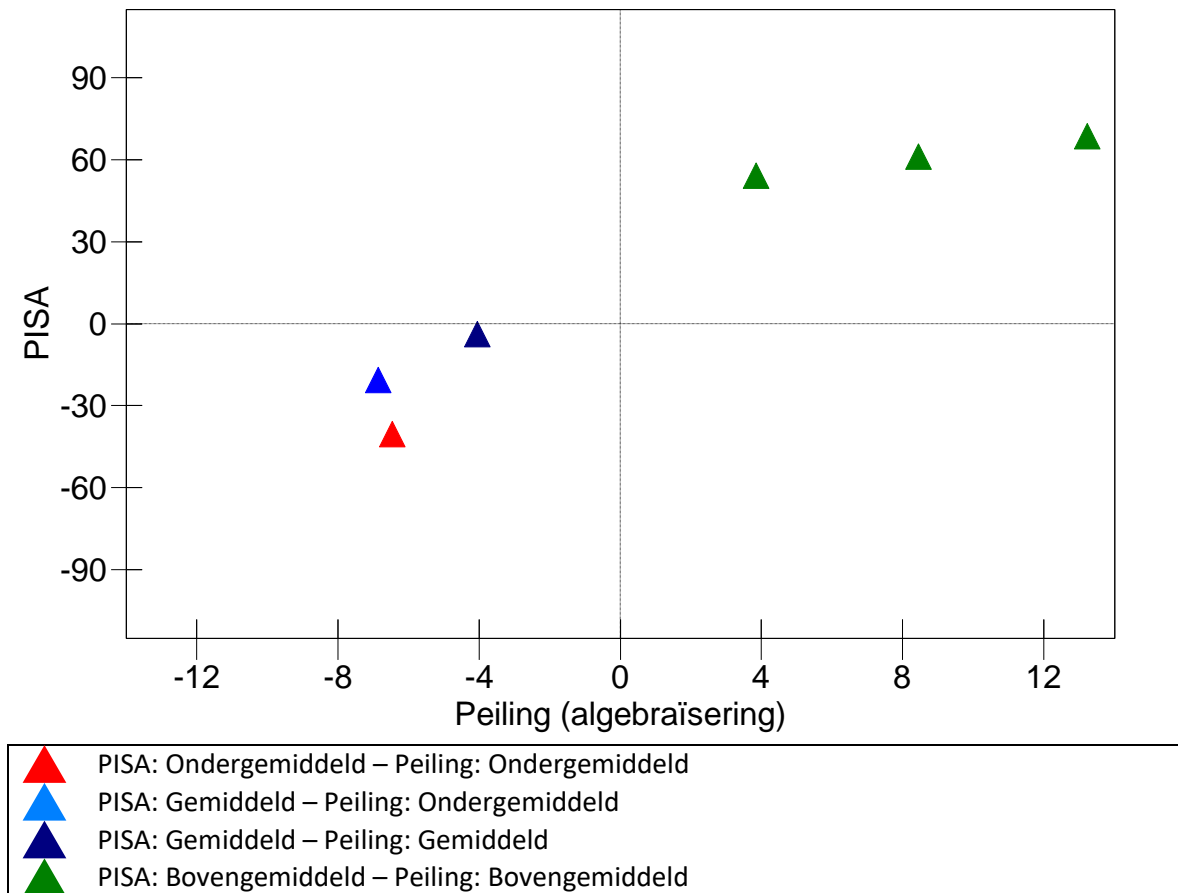
	Meten (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (6,25%)	-	1 (6,25%)
Gemiddeld	-	14 (87,50%)	-	14 (87,50%)
Bovengemiddeld	-	1 (6,25%)	-	1 (6,25%)
Totaal	-	16 (100%)	-	16 (100%)

Bijkomende analyses: Peilingen eerste graad secundair onderwijs & PISA

1. Peiling wiskunde A-stroom 2018 & PISA 2018: 11 scholen

1.1 Algebraïsering (6 scholen)

1.1.1 Type 0-schooleffect



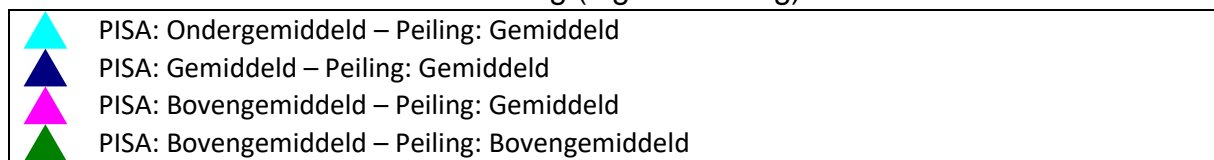
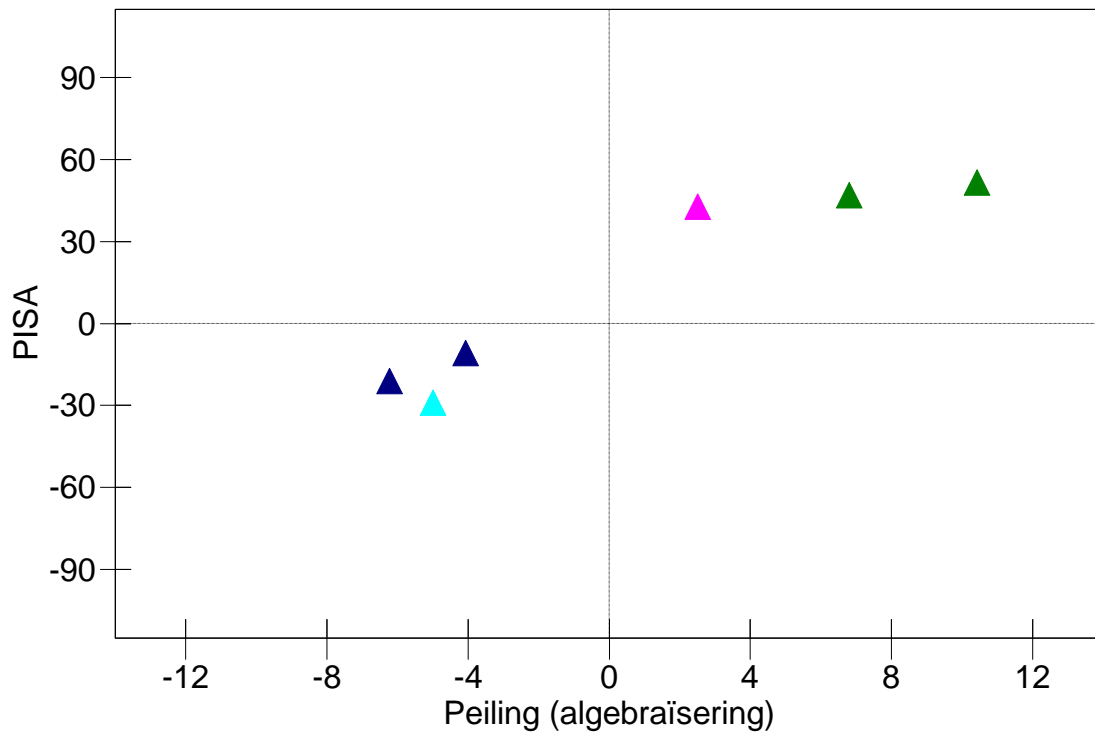
Figuur 98. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op algebraïsering bij de peiling wiskunde in de A-stroom

Tabel 120

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op algebraïsering bij de peiling wiskunde in de A-stroom

	Algebraïsering (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	1 (16,67%)	-	-	1 (16,67%)
Gemiddeld	1 (16,67%)	1 (16,67%)	-	2 (33,33%)
Bovengemiddeld	-	-	3 (50,00%)	3 (50,00%)
Totaal	2 (33,33%)	1 (16,67%)	3 (50,00%)	6 (100%)

1.1.2 Type A-schooleffect



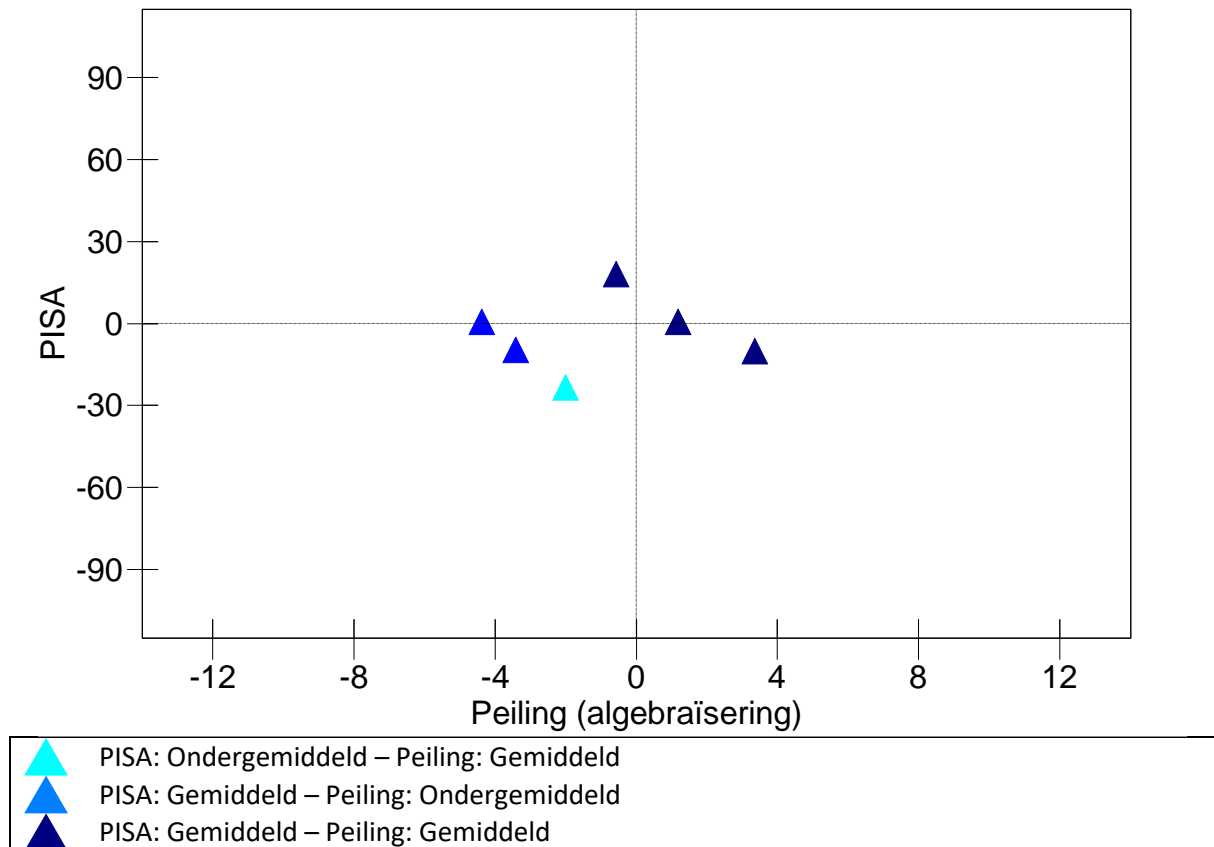
Figuur 99. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op algebraïsering bij de peiling wiskunde in de A-stroom

Tabel 121

Classificatie van scholen volgens hun type A-schooleffect op PISA en op algebraïsering bij de peiling wiskunde in de A-stroom

	Algebraïsering (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	-	-	-
Gemiddeld	2 (33,33%)	1 (16,67%)	-	3 (50,00%)
Bovengemiddeld	-	1 (16,67%)	2 (33,33%)	3 (50,00%)
Totaal	2 (33,33%)	2 (33,33%)	2 (33,33%)	6 (100%)

1.1.3 Type B-schooleffect



Figuur 100. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op algebraïsering bij de peiling wiskunde in de A-stroom

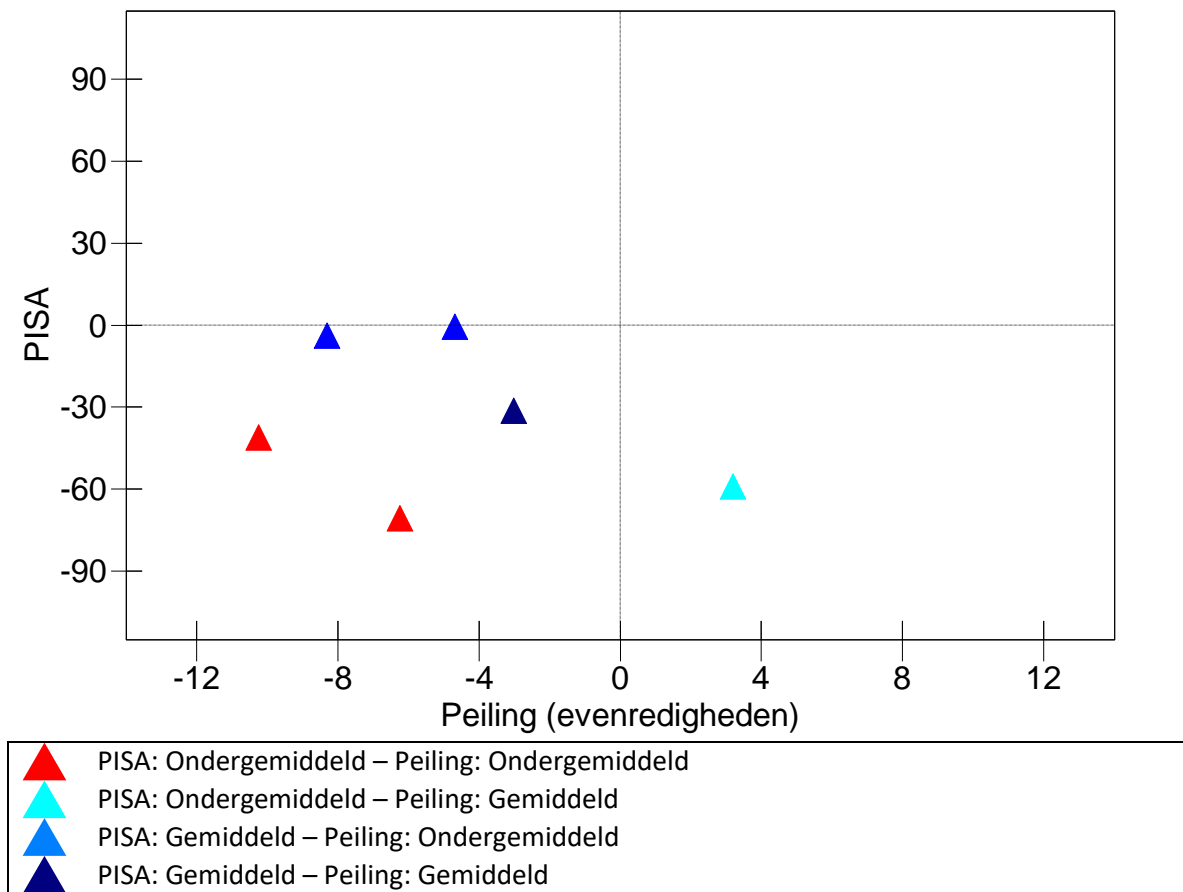
Tabel 122

Classificatie van scholen volgens hun type B-schooleffect op PISA en op algebraïsering bij de peiling wiskunde in de A-stroom

	Algebraïsering (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (16,67%)	-	1 (16,67%)
Gemiddeld	2 (33,33%)	3 (50,00%)	-	5 (83,33%)
Bovengemiddeld	-	-	-	-
Totaal	2 (33,33%)	4 (66,67%)	-	6 (100%)

1.2 Evenredigheden (6 scholen)

1.2.1 Type 0-schooleffect



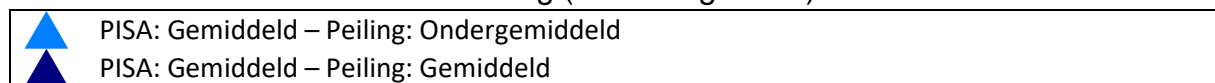
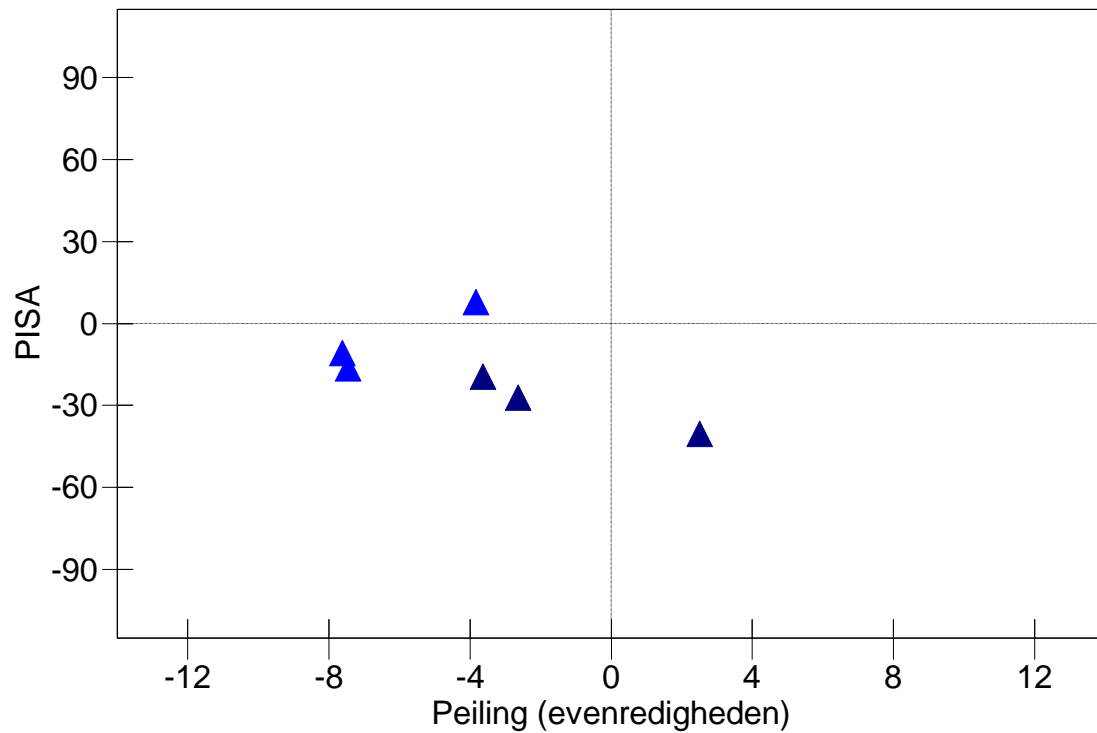
Figuur 101. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op evenredigheden bij de peiling wiskunde in de A-stroom

Tabel 123

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op evenredigheden bij de peiling wiskunde in de A-stroom

	Evenredigheden (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	2 (33,33%)	1 (16,67%)	-	3 (50,00%)
Gemiddeld	2 (33,33%)	1 (16,67%)	-	3 (50,00%)
Bovengemiddeld	-	-	-	-
Totaal	4 (66,67%)	2 (33,33%)	-	6 (100%)

1.2.2 Type A-schooleffect



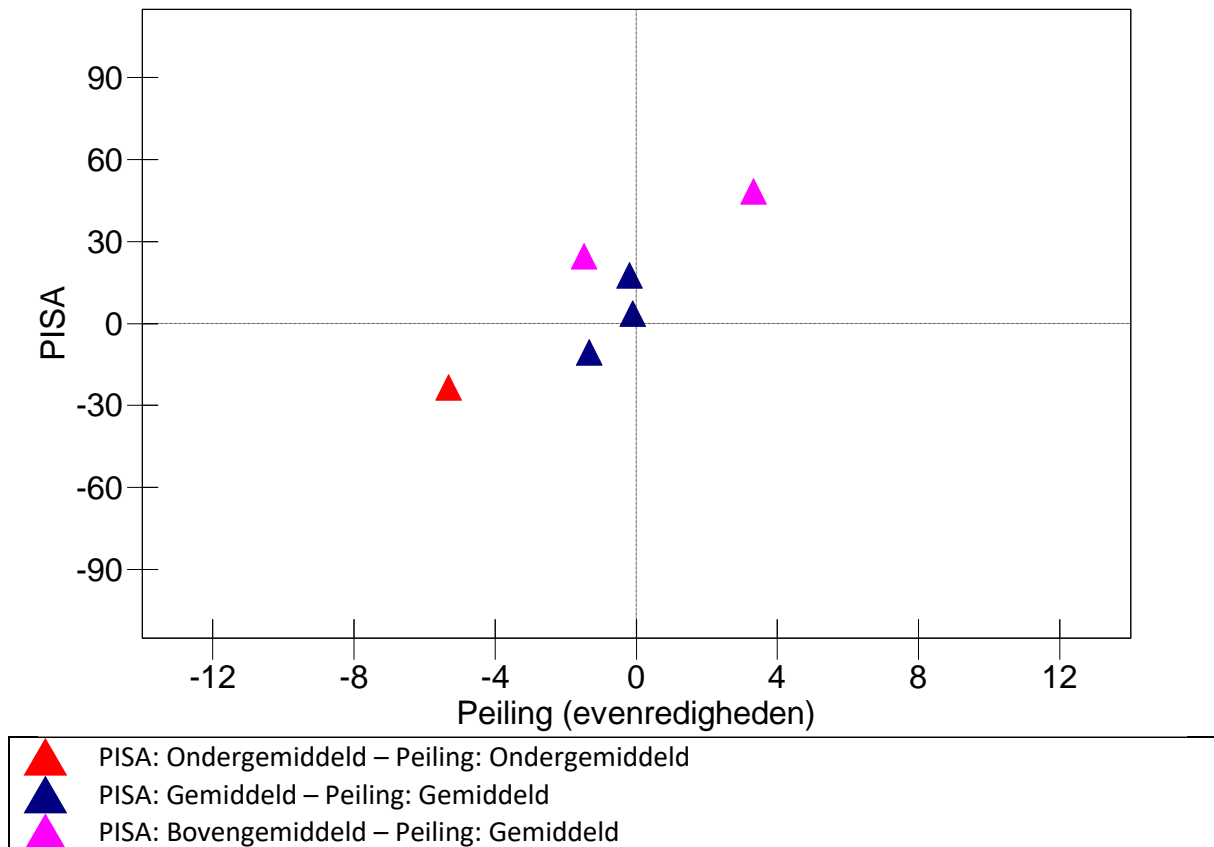
Figuur 102. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op evenredigheden bij de peiling wiskunde in de A-stroom

Tabel 124

Classificatie van scholen volgens hun type A-schooleffect op PISA en op evenredigheden bij de peiling wiskunde in de A-stroom

	Evenredigheden (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	-	-	-
Gemiddeld	3 (50,00%)	3 (50,00%)	-	6 (100%)
Bovengemiddeld	-	-	-	-
Totaal	3 (50,00%)	3 (50,00%)	-	6 (100%)

1.2.3 Type B-schooleffect



Figuur 103. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op evenredigheden bij de peiling wiskunde in de A-stroom

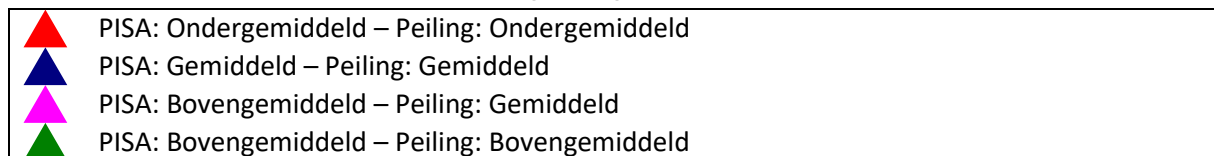
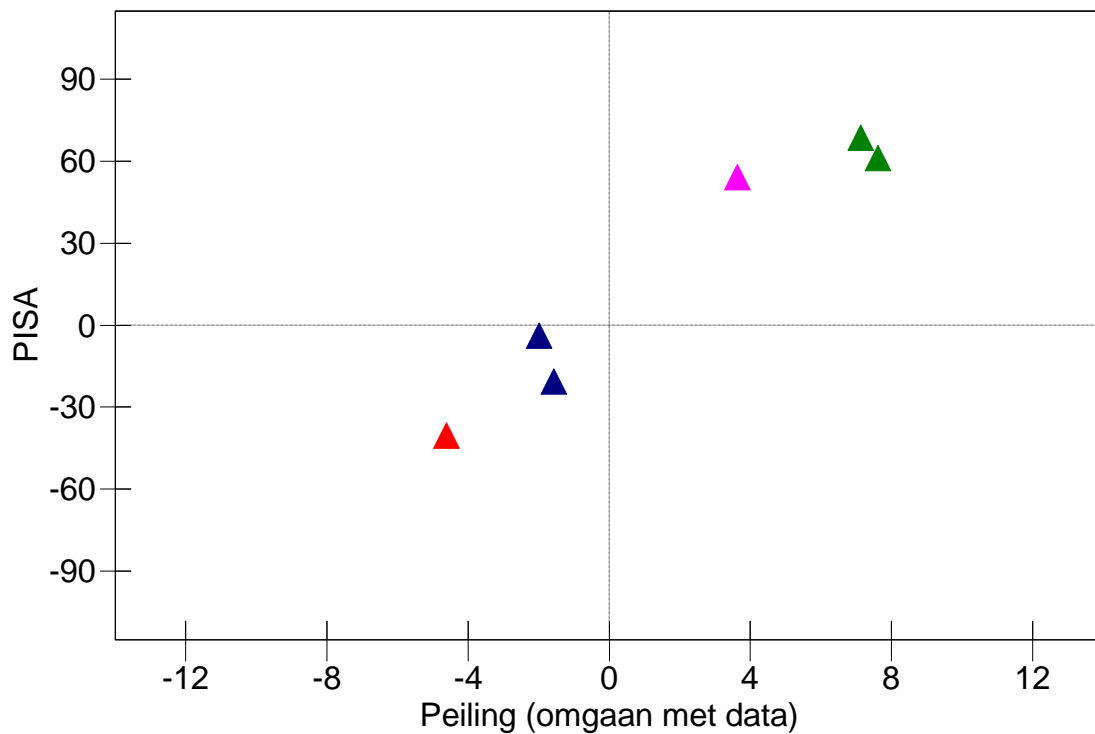
Tabel 125

Classificatie van scholen volgens hun type B-schooleffect op PISA en op evenredigheden bij de peiling wiskunde in de A-stroom

	Evenredigheden (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	1 (16,67%)	-	-	1 (16,67%)
Gemiddeld	-	3 (50,00%)	-	3 (50,00%)
Bovengemiddeld	-	2 (33,33%)	-	2 (33,33%)
Totaal	1 (16,67%)	5 (83,33%)	-	6 (100%)

1.3 Omgaan met data (6 scholen)

1.3.1 Type 0-schooleffect



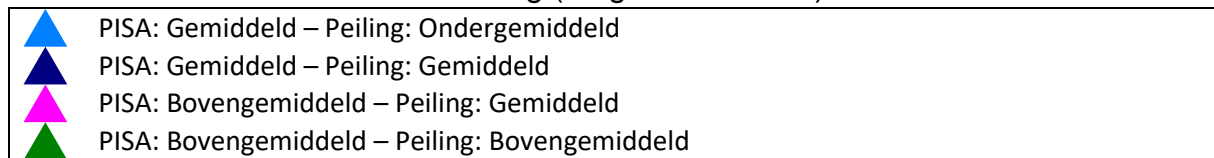
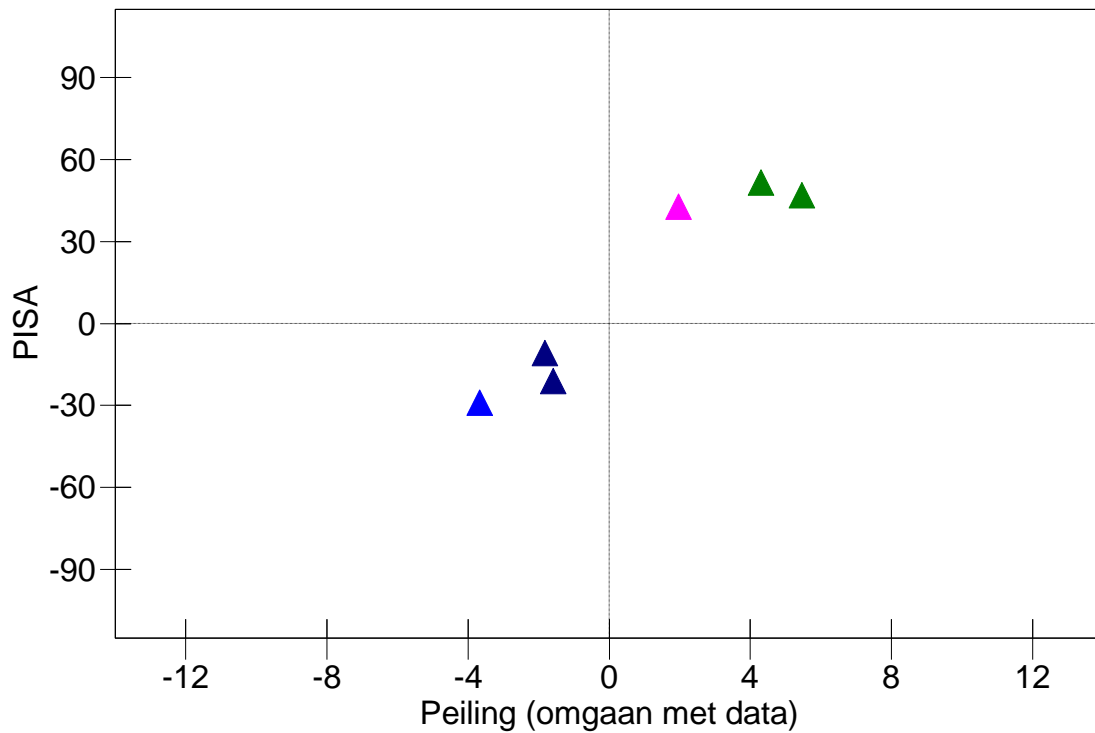
Figuur 104. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op omgaan met data bij de peiling wiskunde in de A-stroom

Tabel 126

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op omgaan met data bij de peiling wiskunde in de A-stroom

	Omgaan met data (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	1 (16,67%)	-	-	1 (16,67%)
Gemiddeld	-	2 (33,33%)	-	2 (33,33%)
Bovengemiddeld	-	1 (16,67%)	2 (33,33%)	3 (50,00%)
Totaal	1 (16,67%)	3 (50,00%)	2 (33,33%)	6 (100%)

1.3.2 Type A-schooleffect



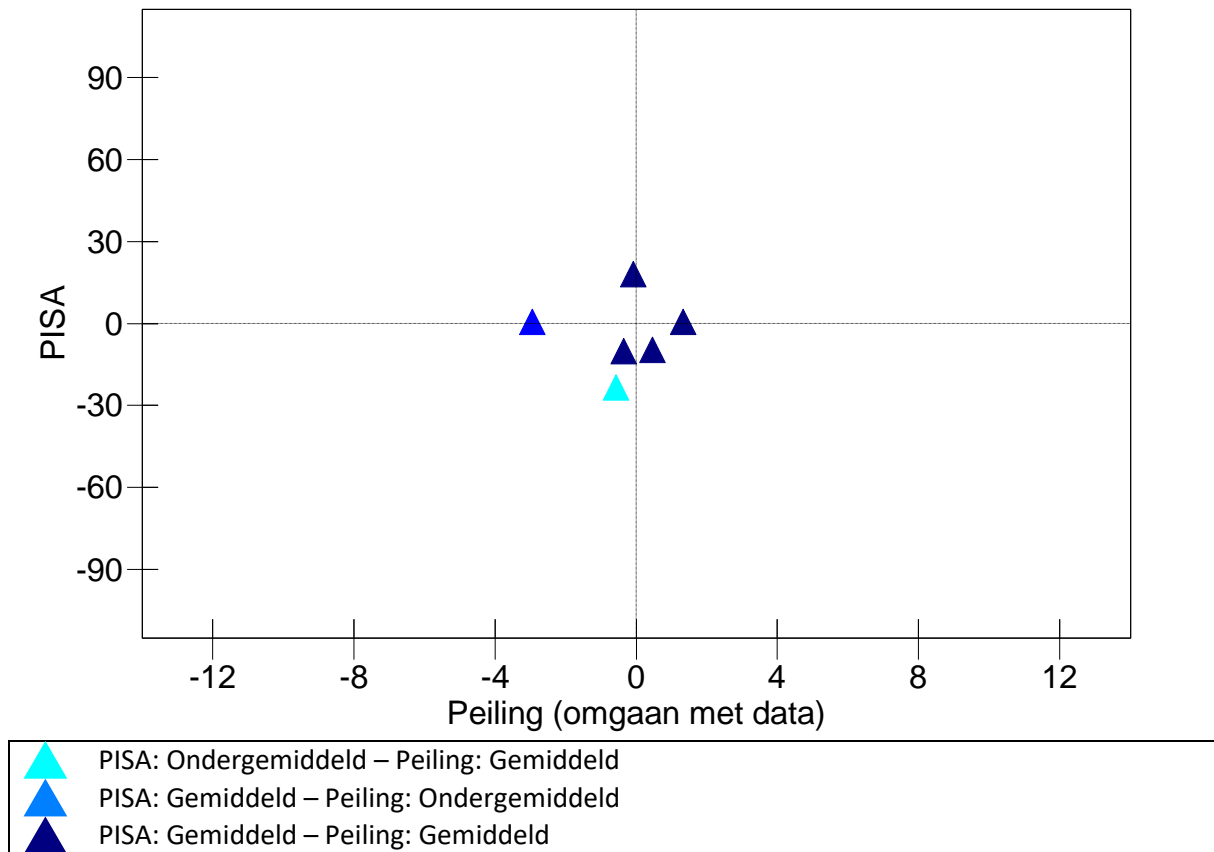
Figuur 105. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op omgaan met data bij de peiling wiskunde in de A-stroom

Tabel 127

Classificatie van scholen volgens hun type A-schooleffect op PISA en op omgaan met data bij de peiling wiskunde in de A-stroom

	Omgaan met data (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	-	-	-
Gemiddeld	1 (16,67%)	2 (33,33%)	-	3 (50,00%)
Bovengemiddeld	-	1 (16,67%)	2 (33,33%)	3 (50,00%)
Totaal	1 (16,67%)	3 (50,00%)	2 (33,33%)	6 (100%)

1.3.3 Type B-schooleffect



Figuur 106. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op omgaan met data bij de peiling wiskunde in de A-stroom

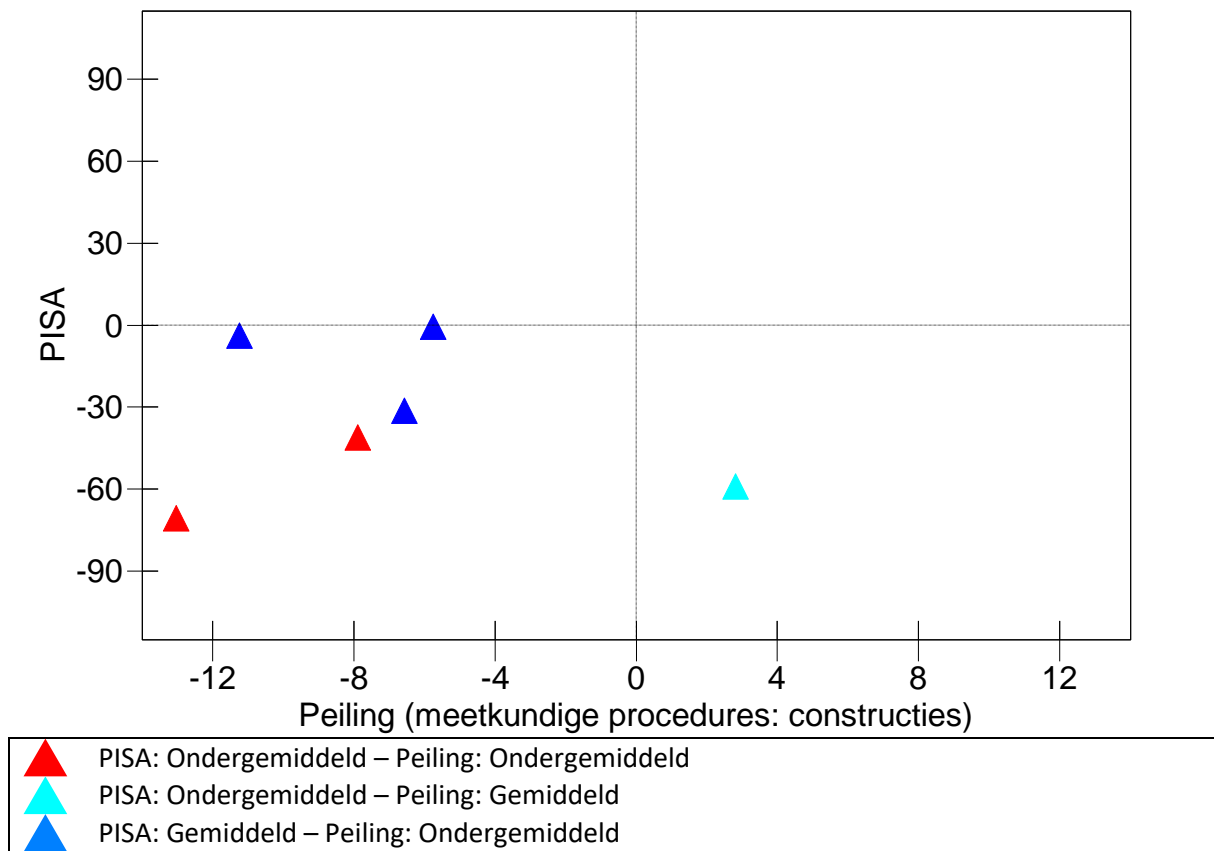
Tabel 128

Classificatie van scholen volgens hun type B-schooleffect op PISA en op omgaan met data bij de peiling wiskunde in de A-stroom

	Omgaan met data (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (16,67%)	-	1 (16,67%)
Gemiddeld	1 (16,67%)	4 (66,67%)	-	5 (83,33%)
Bovengemiddeld	-	-	-	-
Totaal	1 (16,67%)	5 (83,33%)	-	6 (100%)

1.4 Meetkundige procedures: constructies (6 scholen)

1.4.1 Type 0-schooleffect



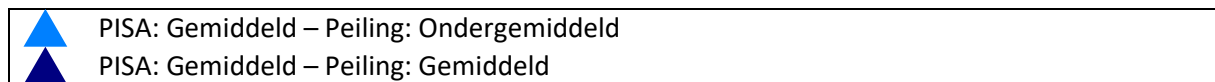
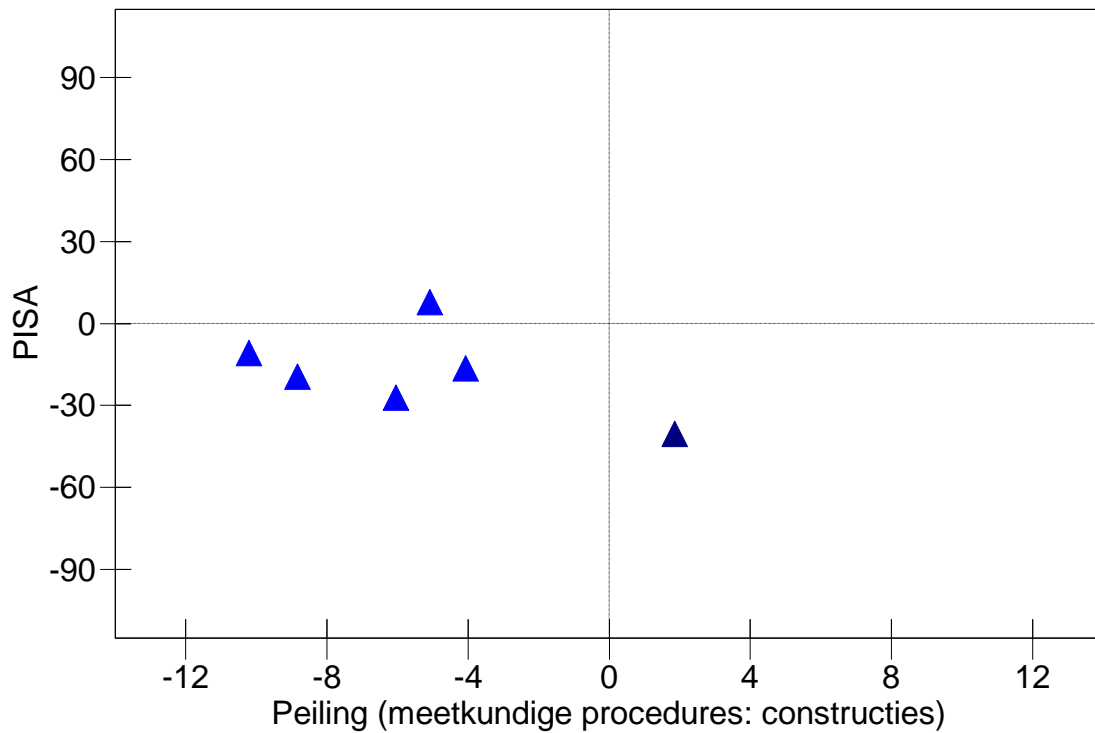
Figuur 107. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op meetkundige procedures: constructies bij de peiling wiskunde in de A-stroom

Tabel 129

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op meetkundige procedures: constructies bij de peiling wiskunde in de A-stroom

	Meetkundige procedures: constructies (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	2 (33,33%)	1 (16,67%)	-	3 (50,00%)
Gemiddeld	3 (50,00%)	-	-	3 (50,00%)
Bovengemiddeld	-	-	-	-
Totaal	5 (83,33%)	1 (16,67%)	-	6 (100%)

1.4.2 Type A-schooleffect



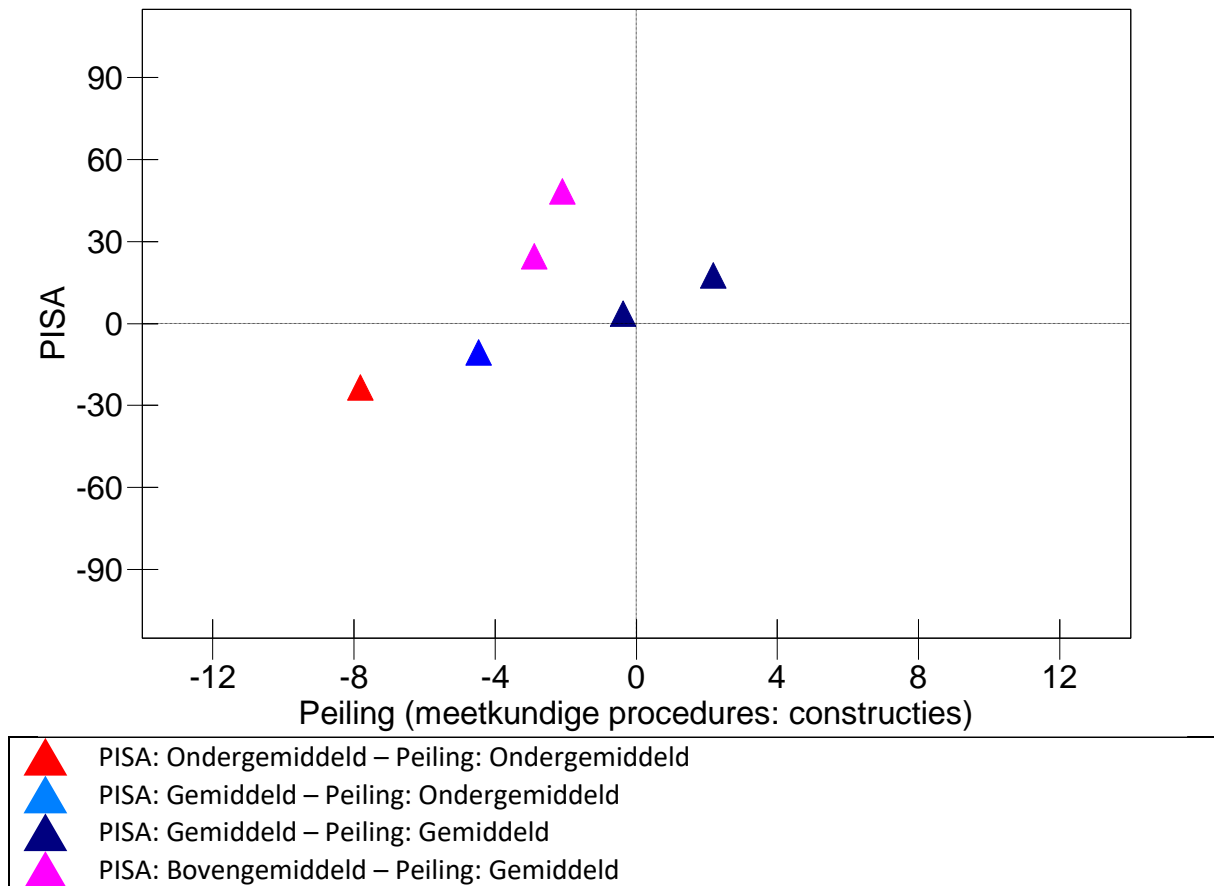
Figuur 108. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op meetkundige procedures: constructies bij de peiling wiskunde in de A-stroom

Tabel 130

Classificatie van scholen volgens hun type A-schooleffect op PISA en op meetkundige procedures: constructies bij de peiling wiskunde in de A-stroom

	Meetkundige procedures: constructies (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	-	-	-
Gemiddeld	5 (83,33%)	1 (16,67%)	-	6 (100%)
Bovengemiddeld	-	-	-	-
Totaal	5 (83,33%)	1 (16,67%)	-	6 (100%)

1.4.3 Type B-schooleffect



Figuur 109. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op meetkundige procedures: constructies bij de peiling wiskunde in de A-stroom

Tabel 131

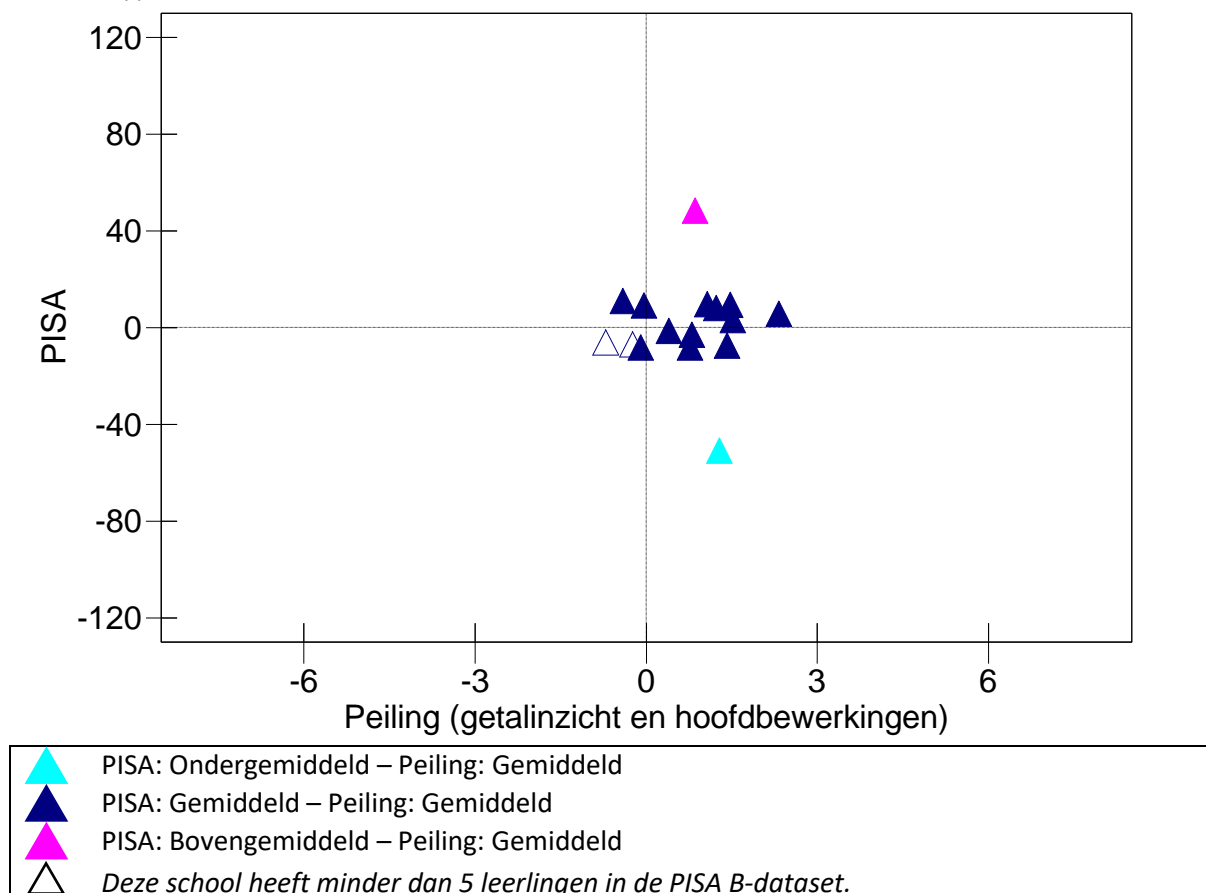
Classificatie van scholen volgens hun type B-schooleffect op PISA en op meetkundige procedures: constructies bij de peiling wiskunde in de A-stroom

	Meetkundige procedures: constructies (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	1 (16,67%)	-	-	1 (16,67%)
Gemiddeld	1 (16,67%)	2 (33,33%)	-	3 (50,00%)
Bovengemiddeld	-	2 (33,33%)	-	2 (33,33%)
Totaal	2 (33,33%)	4 (66,67%)	-	6 (100%)

2. Peiling wiskunde B-stroom 2019 & PISA 2018: 16 scholen⁸⁷

2.1 Getalinzicht en hoofdbewerkingen (16 scholen)

2.1.1 Type 0-schooleffect



Figuur 110. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op getalinzicht en hoofdbewerkingen bij de peiling wiskunde in de B-stroom

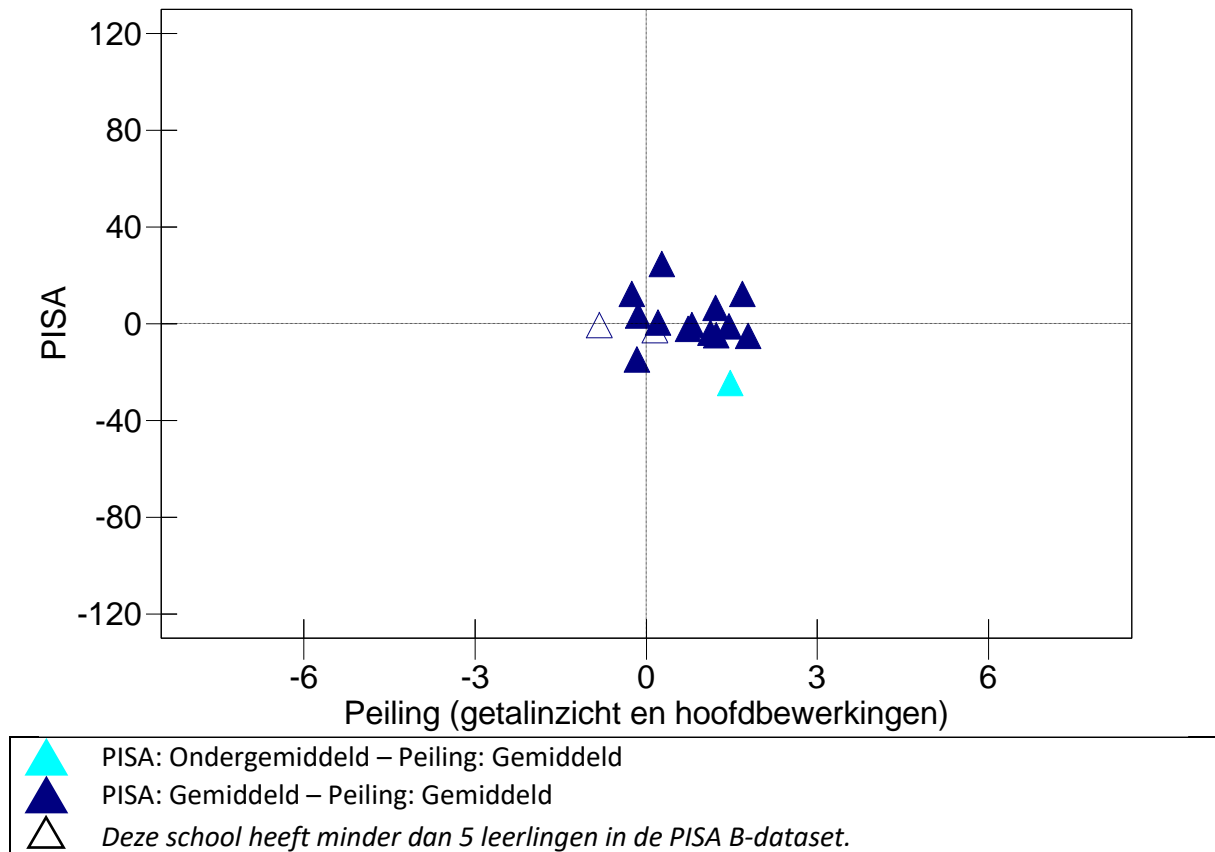
Tabel 132

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op getalinzicht en hoofdbewerkingen bij de peiling wiskunde in de B-stroom

	Getalinzicht en hoofdbewerkingen (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (6,25%)	-	1 (6,25%)
Gemiddeld	-	14 (87,50%)	-	14 (87,50%)
Bovengemiddeld	-	1 (6,25%)	-	1 (6,25%)
Totaal	-	16 (100%)	-	16 (100%)

⁸⁷ Twee scholen werden wel opgenomen bij de resultaten van cross-instrumenteel onderzoek (vergelijking 3) en niet in de bijkomende analyses. Dat wijst erop dat die scholen wel een B-stroom en/of BSO-studierichtingen aanbieden, maar dat die studierichtingen te klein zijn in verhouding met de andere studierichtingen en het geselecteerde aantal leerlingen voor PISA (35 leerlingen). Daardoor zijn er in die scholen voor PISA geen leerlingen geselecteerd die de B-stroom of het BSO volgen.

2.1.2 Type A-schooleffect



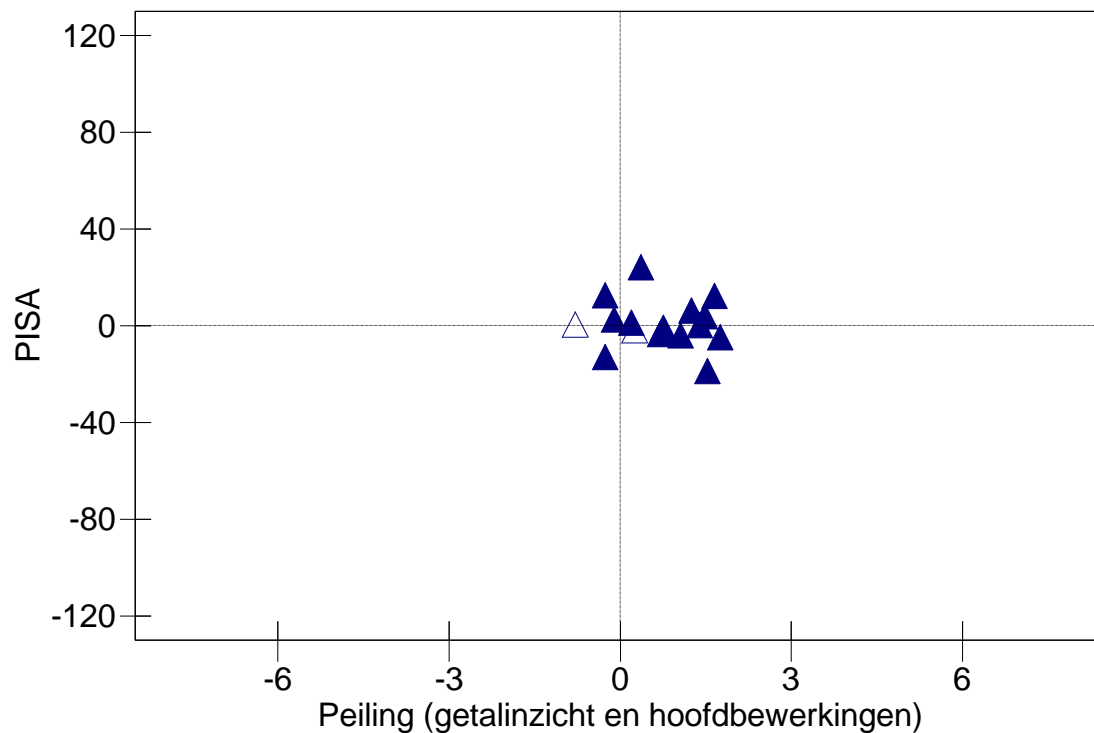
Figuur 111. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op getalinzicht en hoofdbewerkingen bij de peiling wiskunde in de B-stroom

Tabel 133

Classificatie van scholen volgens hun type A-schooleffect op PISA en op getalinzicht en hoofdbewerkingen bij de peiling wiskunde in de B-stroom

	Getalinzicht en hoofdbewerkingen (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (6,25%)	-	1 (6,25%)
Gemiddeld	-	15 (93,75%)	-	15 (93,75%)
Bovengemiddeld	-	-	-	-
Totaal	-	16 (100%)	-	16 (100%)

2.1.3 Type B-schooleffect



▲ PISA: Gemiddeld – Peiling: Gemiddeld

△ Deze school heeft minder dan 5 leerlingen in de PISA B-dataset.

Figuur 112. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op getalinzicht en hoofdbewerkingen bij de peiling wiskunde in de B-stroom

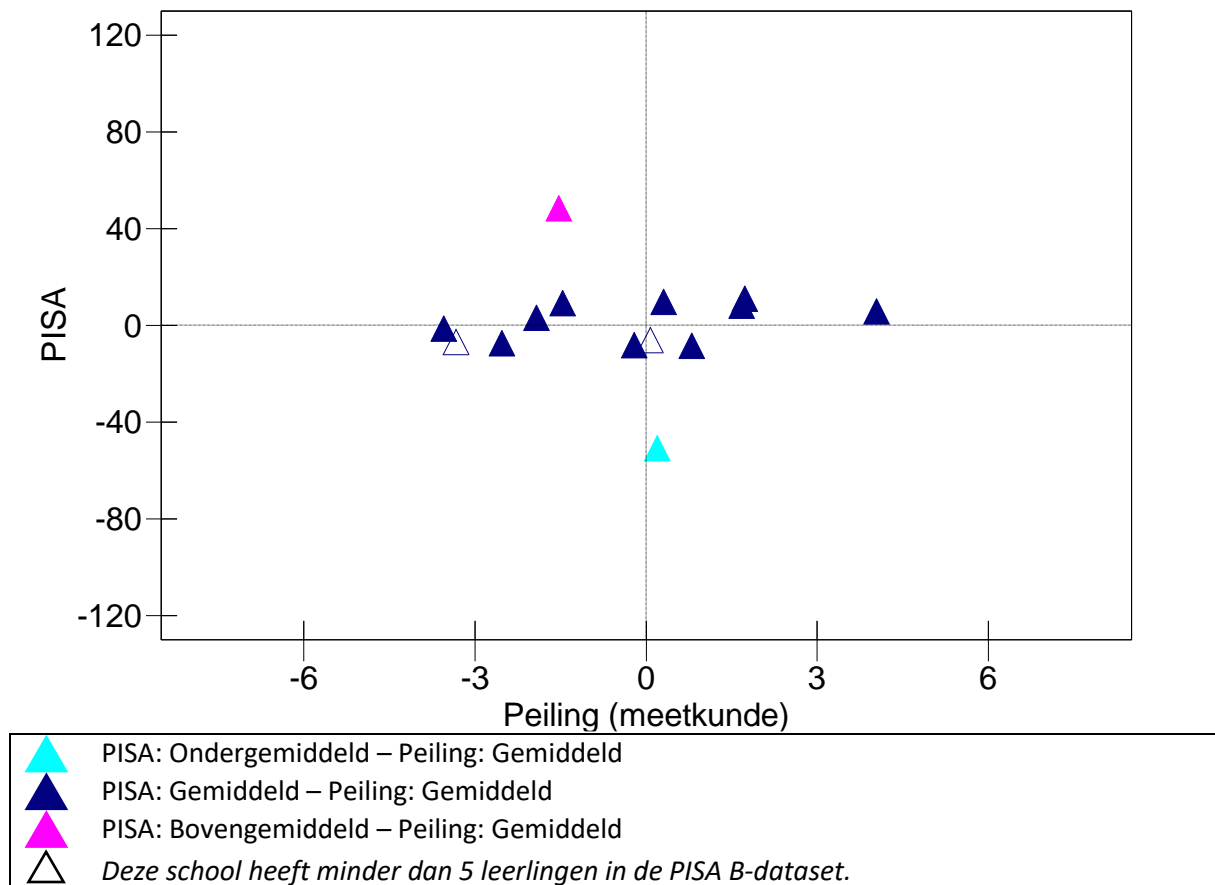
Tabel 134

Classificatie van scholen volgens hun type B-schooleffect op PISA en op getalinzicht en hoofdbewerkingen bij de peiling wiskunde in de B-stroom

	Getalinzicht en hoofdbewerkingen (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	-	-	-
Gemiddeld	-	16 (100%)	-	16 (100%)
Bovengemiddeld	-	-	-	-
Totaal	-	16 (100%)	-	16 (100%)

2.2 Meetkunde (15 scholen)

2.2.1 Type 0-schooleffect



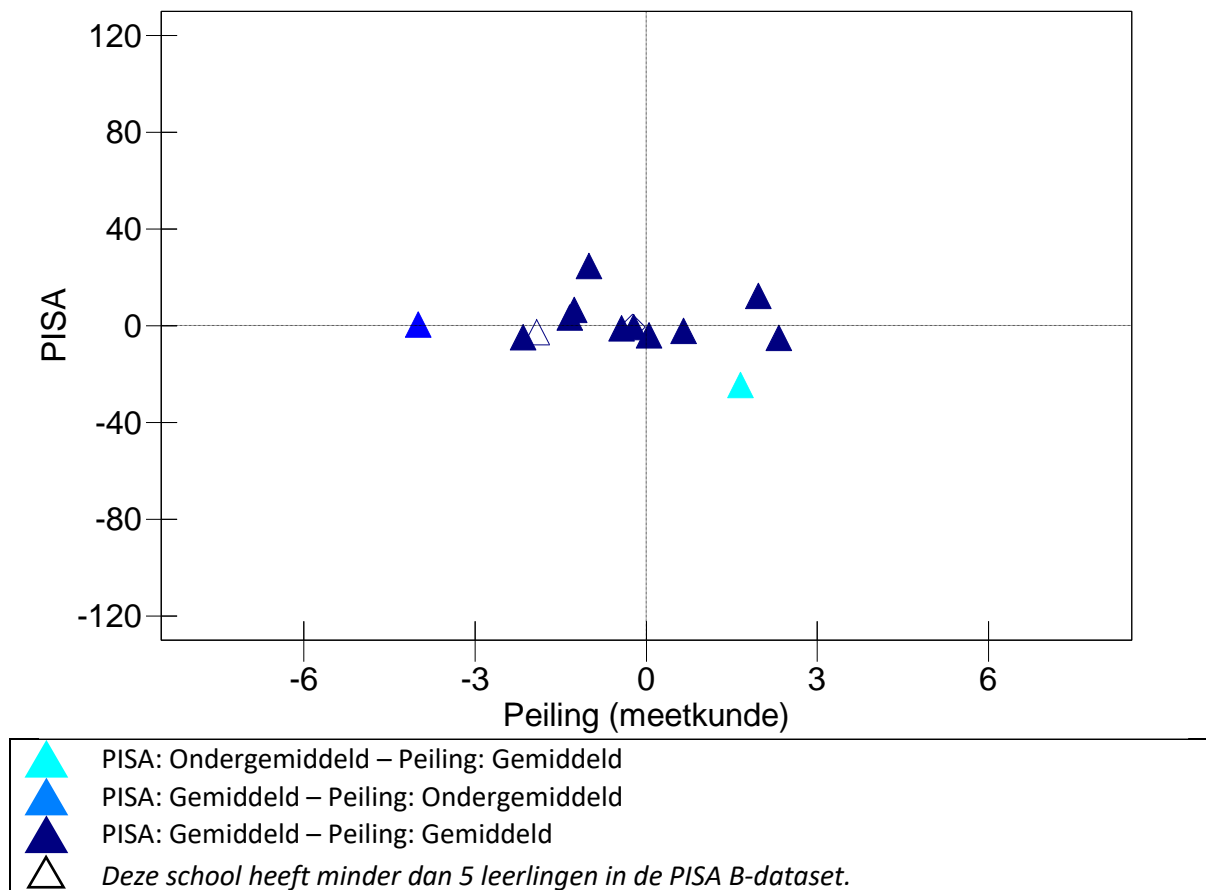
Figuur 113. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op meetkunde bij de peiling wiskunde in de B-stroom

Tabel 135

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op meetkunde bij de peiling wiskunde in de B-stroom

	Meetkunde (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (6,67%)	-	1 (6,67%)
Gemiddeld	-	13 (86,67%)	-	13 (86,67%)
Bovengemiddeld	-	1 (6,67%)	-	1 (6,67%)
Totaal	-	15 (100%)	-	15 (100%)

2.2.2 Type A-schooleffect



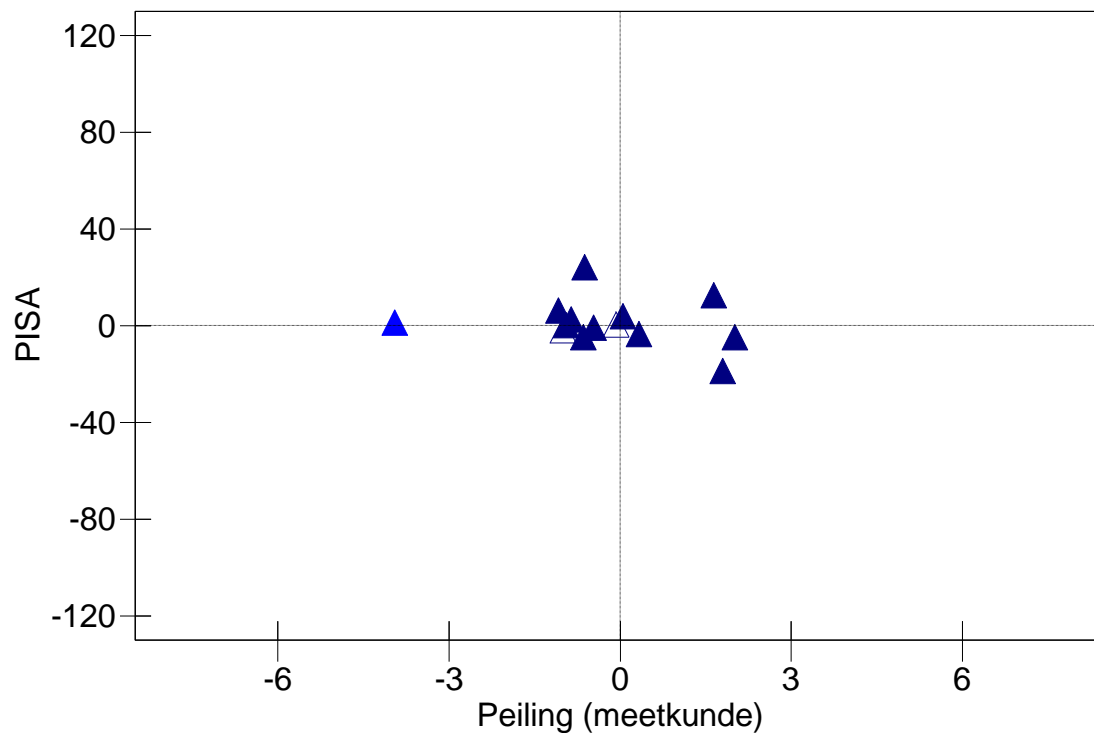
Figuur 114. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op meetkunde bij de peiling wiskunde in de B-stroom




Tabel 136

Classificatie van scholen volgens hun type A-schooleffect op PISA en op meetkunde bij de peiling wiskunde in de B-stroom

	Meetkunde (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (6,67%)	-	1 (6,67%)
Gemiddeld	1 (6,67%)	13 (86,67%)	-	14 (93,33%)
Bovengemiddeld	-	-	-	-
Totaal	1 (6,67%)	14 (93,33%)	-	15 (100%)

2.2.3 Type B-schooleffect



	PISA: Gemiddeld – Peiling: Ondergemiddeld
	PISA: Gemiddeld – Peiling: Gemiddeld
	Deze school heeft minder dan 5 leerlingen in de PISA B-dataset.

Figuur 115. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op meetkunde bij de peiling wiskunde in de B-stroom

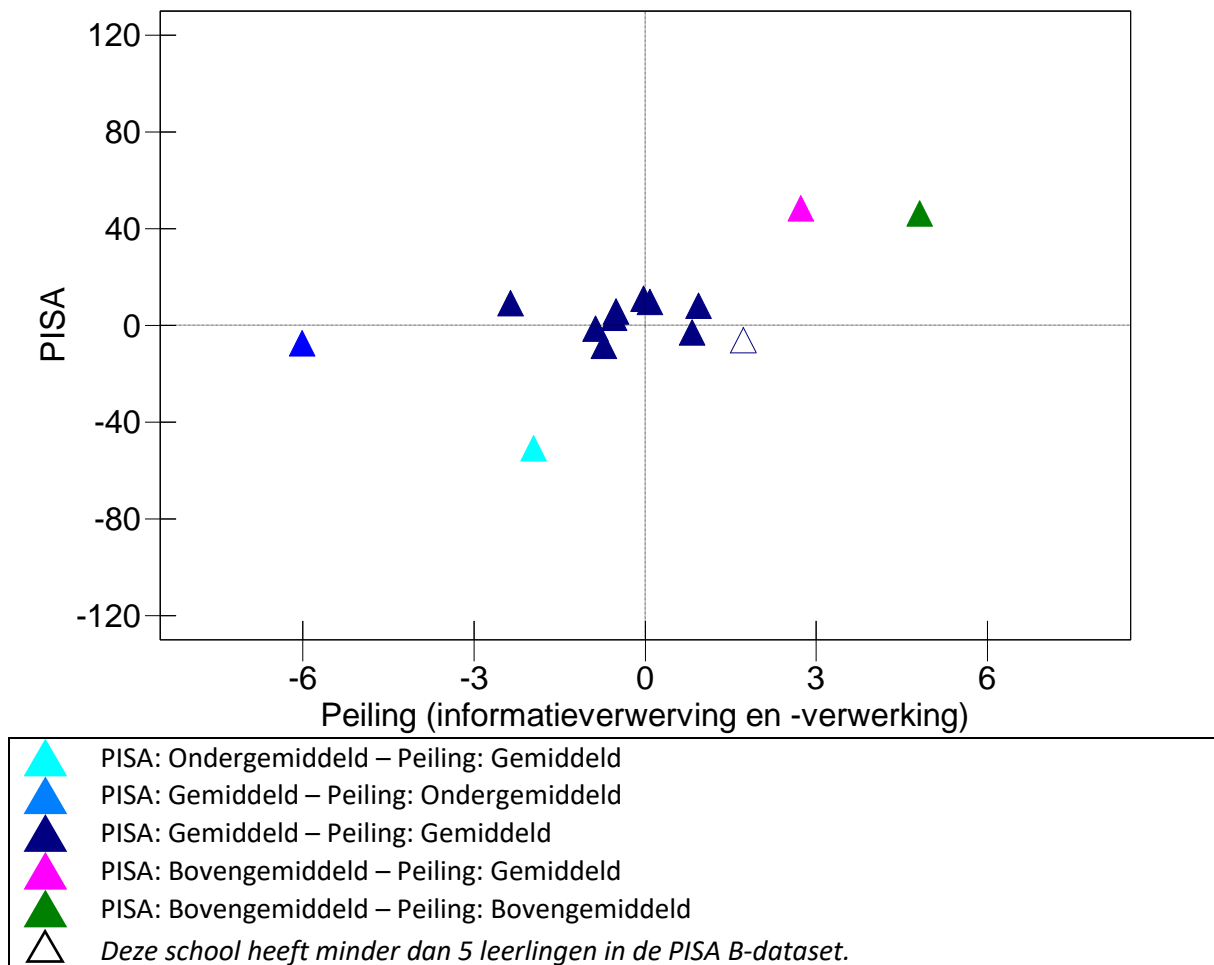
Tabel 137

Classificatie van scholen volgens hun type B-schooleffect op PISA en op meetkunde bij de peiling wiskunde in de B-stroom

	Meetkunde (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	-	-	-
Gemiddeld	1 (6,67%)	14 (93,33%)	-	15 (100%)
Bovengemiddeld	-	-	-	-
Totaal	1 (6,67%)	14 (93,33%)	-	15 (100%)

2.3 Informatieverwerking en -verwerking (14 scholen)

2.3.1 Type 0-schooleffect



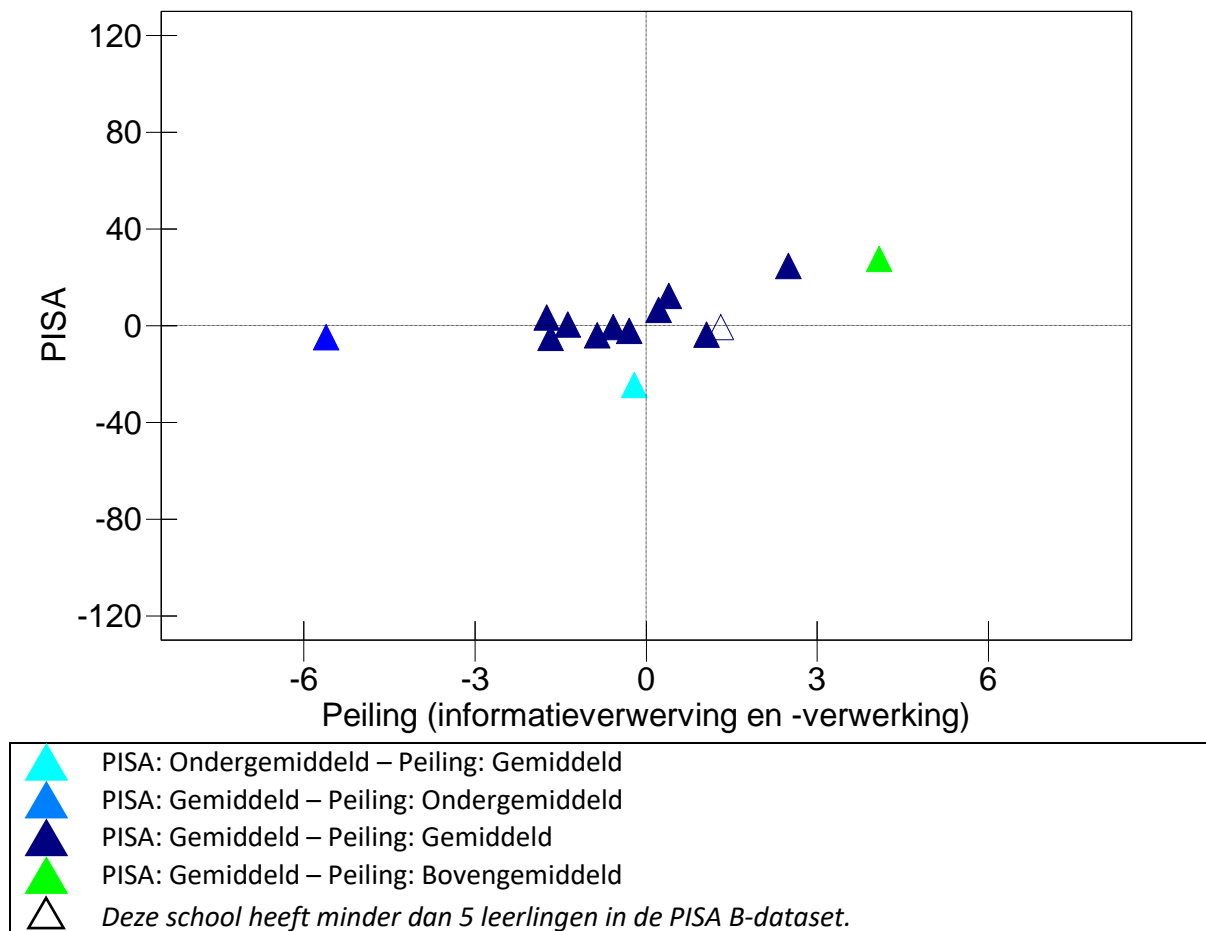
Figuur 116. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op informatieverwerking en -verwerking bij de peiling wiskunde in de B-stroom

Tabel 138

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op informatieverwerking en -verwerking bij de peiling wiskunde in de B-stroom

	Informatieverwerking en -verwerking (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (7,14%)	-	1 (7,14%)
Gemiddeld	1 (7,14%)	10 (71,43%)	-	11 (78,57%)
Bovengemiddeld	-	1 (7,14%)	1 (7,14%)	2 (14,29%)
Totaal	1 (7,14%)	12 (85,71%)	1 (7,14%)	14 (100%)

2.3.2 Type A-schooleffect



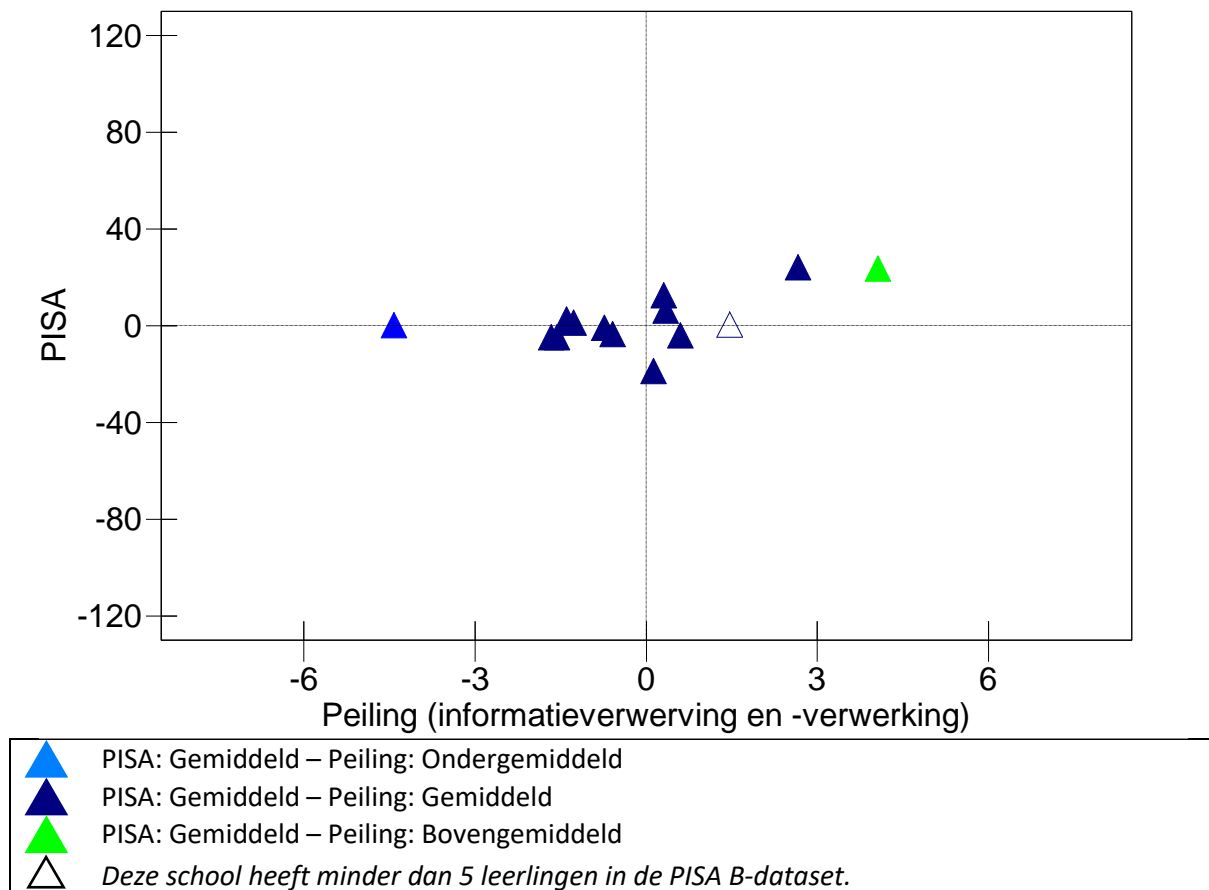
Figuur 117. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op informatieverwerking en -verwerking bij de peiling wiskunde in de B-stroom

Tabel 139

Classificatie van scholen volgens hun type A-schooleffect op PISA en op informatieverwerking en -verwerking bij de peiling wiskunde in de B-stroom

	Informatieverwerking en -verwerking (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (7,14%)	-	1 (7,14%)
Gemiddeld	1 (7,14%)	11 (71,43%)	1 (7,14%)	13 (92,86%)
Bovengemiddeld	-	-	-	-
Totaal	1 (7,14%)	12 (85,71%)	1 (7,14%)	14 (100%)

2.3.3 Type B-schooleffect



Figuur 118. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op informatieverwerking en -verwerking bij de peiling wiskunde in de B-stroom

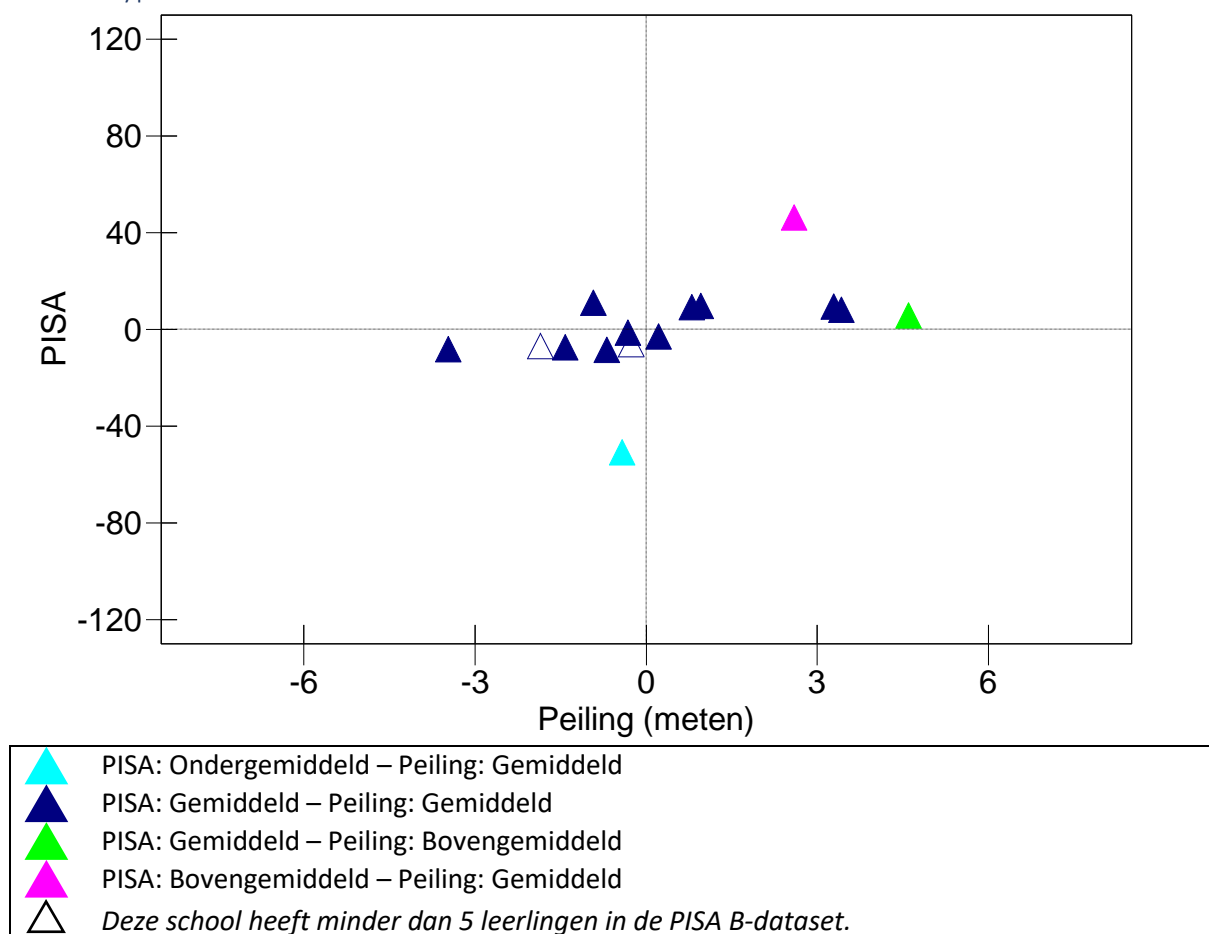
Tabel 140

Classificatie van scholen volgens hun type B-schooleffect op PISA en op informatieverwerking en -verwerking bij de peiling wiskunde in de B-stroom

	Informatieverwerking en -verwerking (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	-	-	-
Gemiddeld	1 (7,14%)	12 (71,43%)	1 (7,14%)	14 (100%)
Bovengemiddeld	-	-	-	-
Totaal	1 (7,14%)	12 (71,43%)	1 (7,14%)	14 (100%)

2.4 Meten (15 scholen)

2.4.1 Type 0-schooleffect



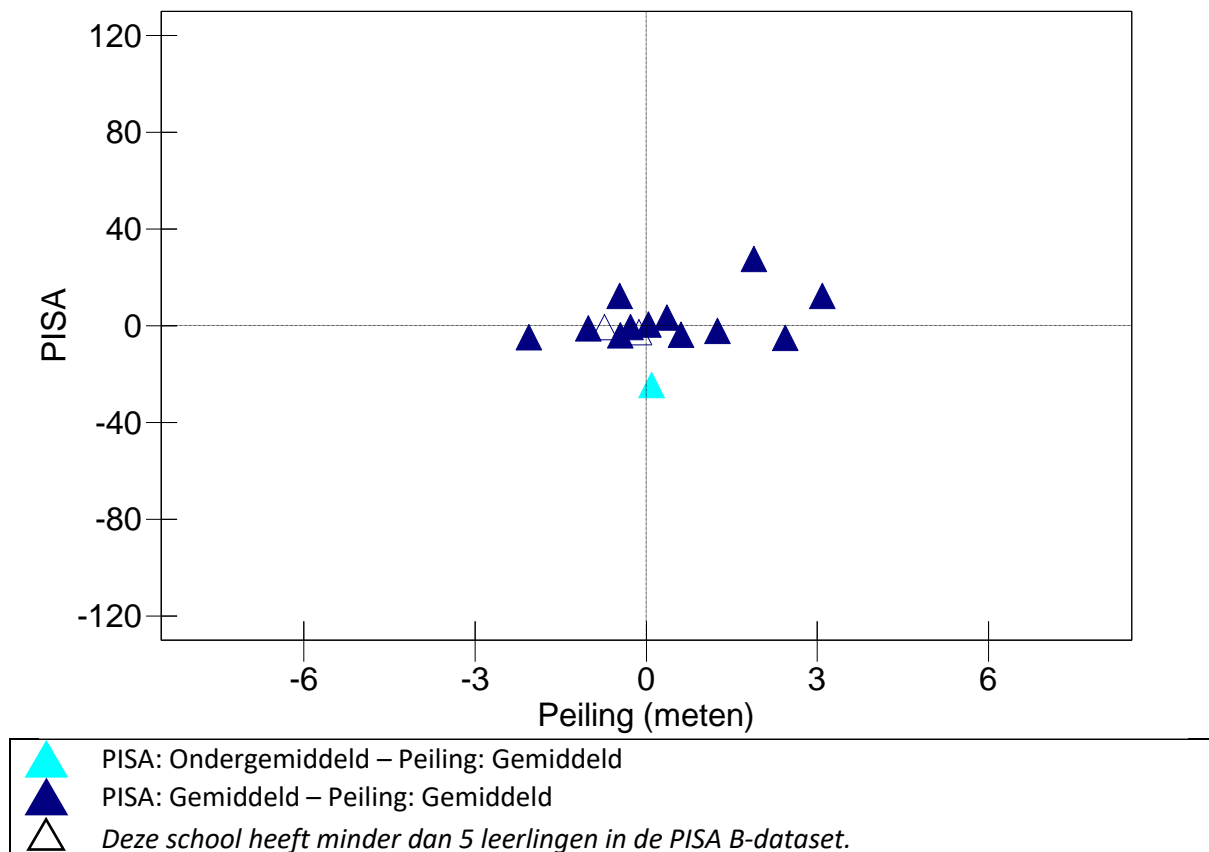
Figuur 119. Puntgrafiek van scholen volgens hun type 0-schooleffect op PISA en op meten bij de peiling wiskunde in de B-stroom

Tabel 141

Classificatie van scholen volgens hun type 0-schooleffect op PISA en op meten bij de peiling wiskunde in de B-stroom

	Meten (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (6,67%)	-	1 (6,67%)
Gemiddeld	-	12 (86,67%)	1 (6,67%)	13 (86,67%)
Bovengemiddeld	-	1 (6,67%)	-	1 (6,67%)
Totaal	-	14 (93,33%)	1 (6,67%)	15 (100%)

2.4.2 Type A-schooleffect



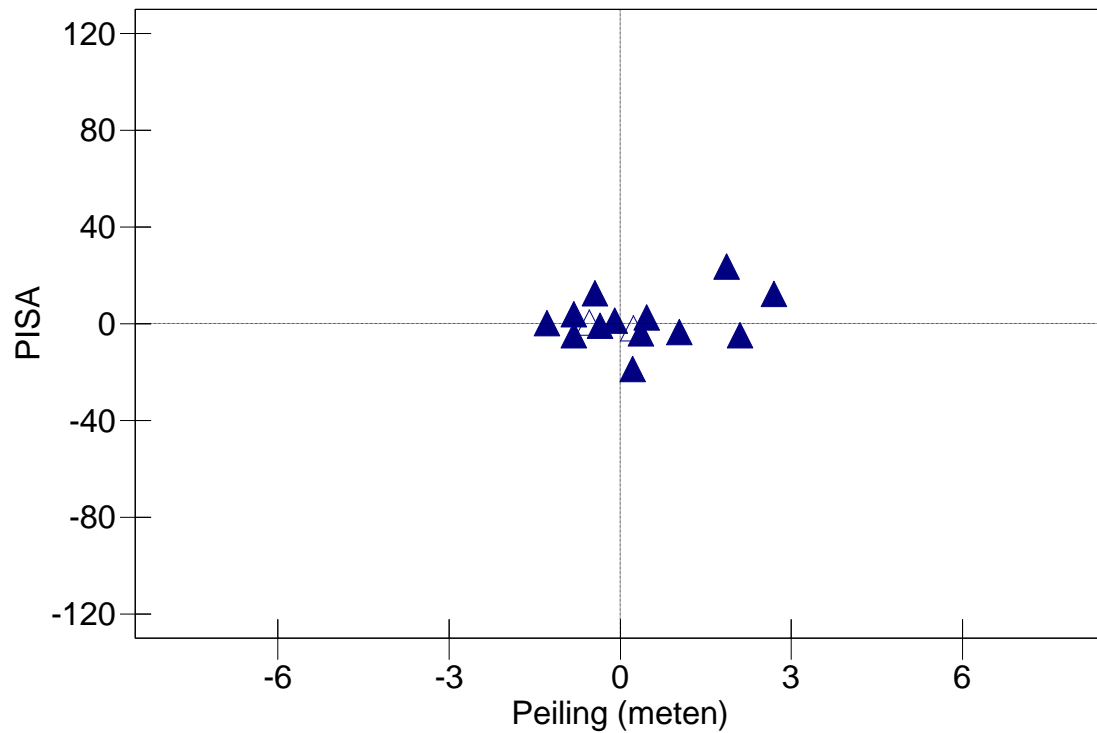
Figuur 120. Puntgrafiek van scholen volgens hun type A-schooleffect op PISA en op meten bij de peiling wiskunde in de B-stroom

Tabel 142

Classificatie van scholen volgens hun type A-schooleffect op PISA en op meten bij de peiling wiskunde in de B-stroom

	Meten (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	1 (6,67%)	-	1 (6,67%)
Gemiddeld	-	14 (93,33%)	-	14 (93,33%)
Bovengemiddeld	-	-	-	-
Totaal	-	15 (100%)	-	15 (100%)

2.4.3 Type B-schooleffect



▲ PISA: Gemiddeld – Peiling: Gemiddeld

△ Deze school heeft minder dan 5 leerlingen in de PISA B-dataset.

Figuur 121. Puntgrafiek van scholen volgens hun type B-schooleffect op PISA en op meten bij de peiling wiskunde in de B-stroom

Tabel 143

Classificatie van scholen volgens hun type B-schooleffect op PISA en op meten bij de peiling wiskunde in de B-stroom

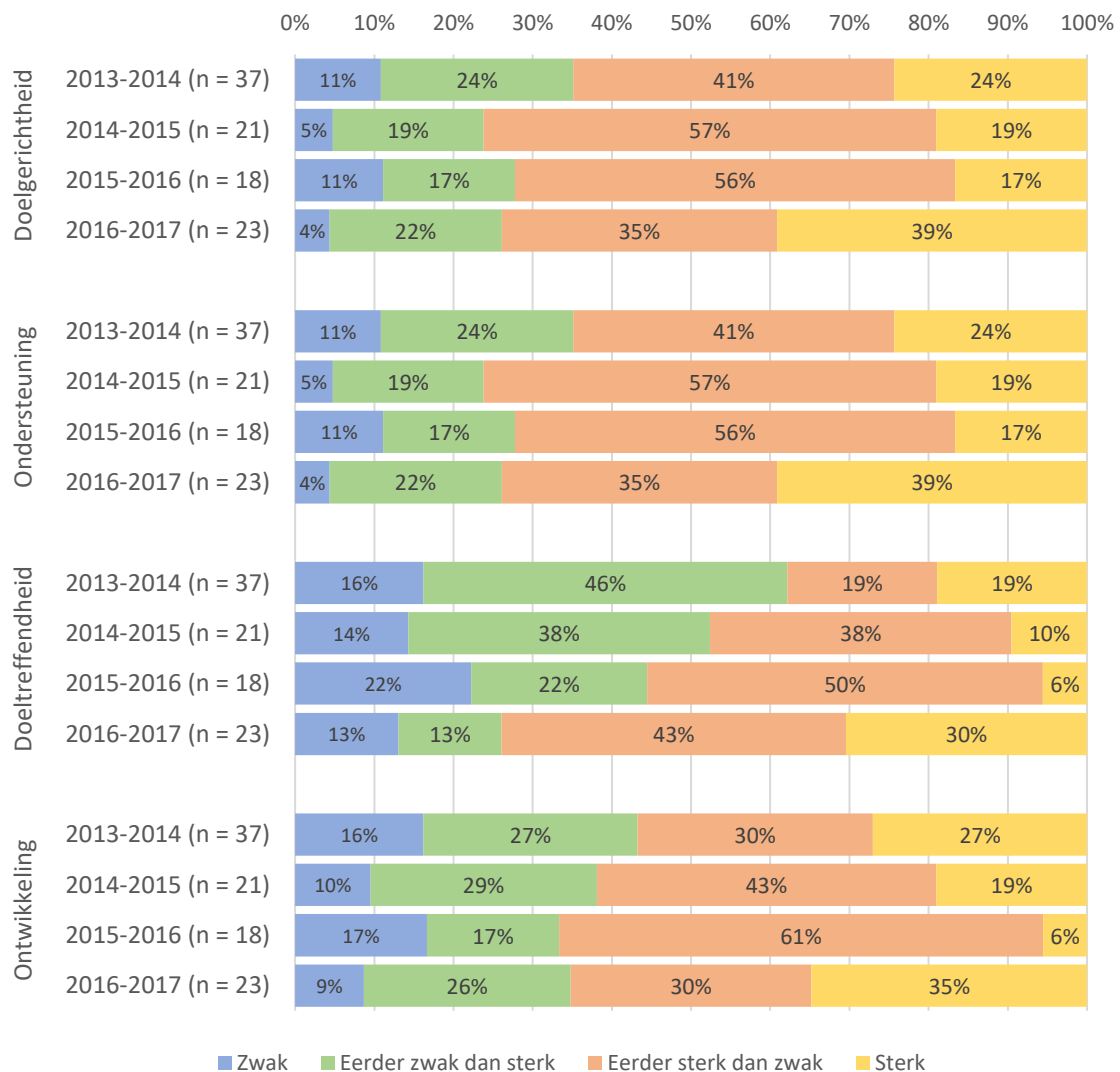
	Meten (peiling wiskunde)			Totaal
	Ondergemiddeld	Gemiddeld	Bovengemiddeld	
PISA				
Ondergemiddeld	-	-	-	-
Gemiddeld	-	15 (100%)	-	15 (100%)
Bovengemiddeld	-	-	-	-
Totaal	-	15 (100%)	-	15 (100%)

Bijlage 3 – Resultaten van studie 3

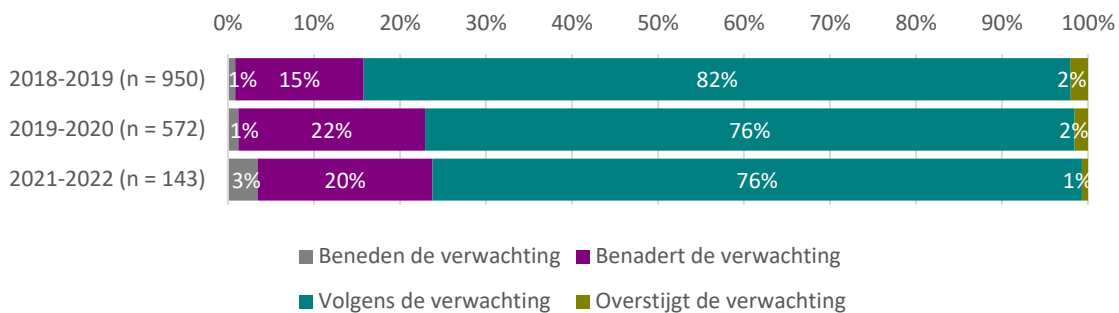
1. Doorlichtingen doorheen de tijd

1.1 Onderwijsaanbod (doorlichtingsronde 3) & Afstemming van het aanbod op een gevalideerd doelenkader (Inspectie 2.0)

1.1.1 Basisonderwijs (gewoon en buitengewoon)

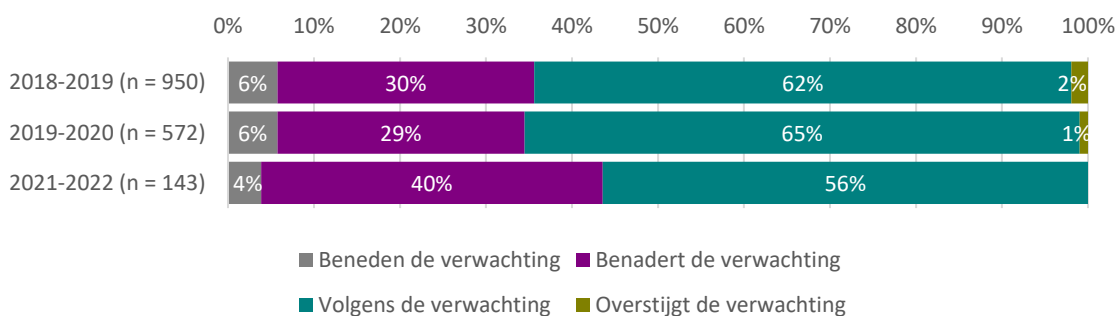


Figuur 122. Overzicht van de inschaling op het proceskenmerk 'onderwijsaanbod' in het basisonderwijs per schooljaar (doorlichtingsronde 3)



Figuur 123. Overzicht van de inschaling op de U1- en O1-schaal 'Afstemming van aanbod op gevalideerd doelenkader' in het basisonderwijs per schooljaar (Inspectie 2.0)

1.1.2 Secundair onderwijs (deeltijds, gewoon en buitengewoon)

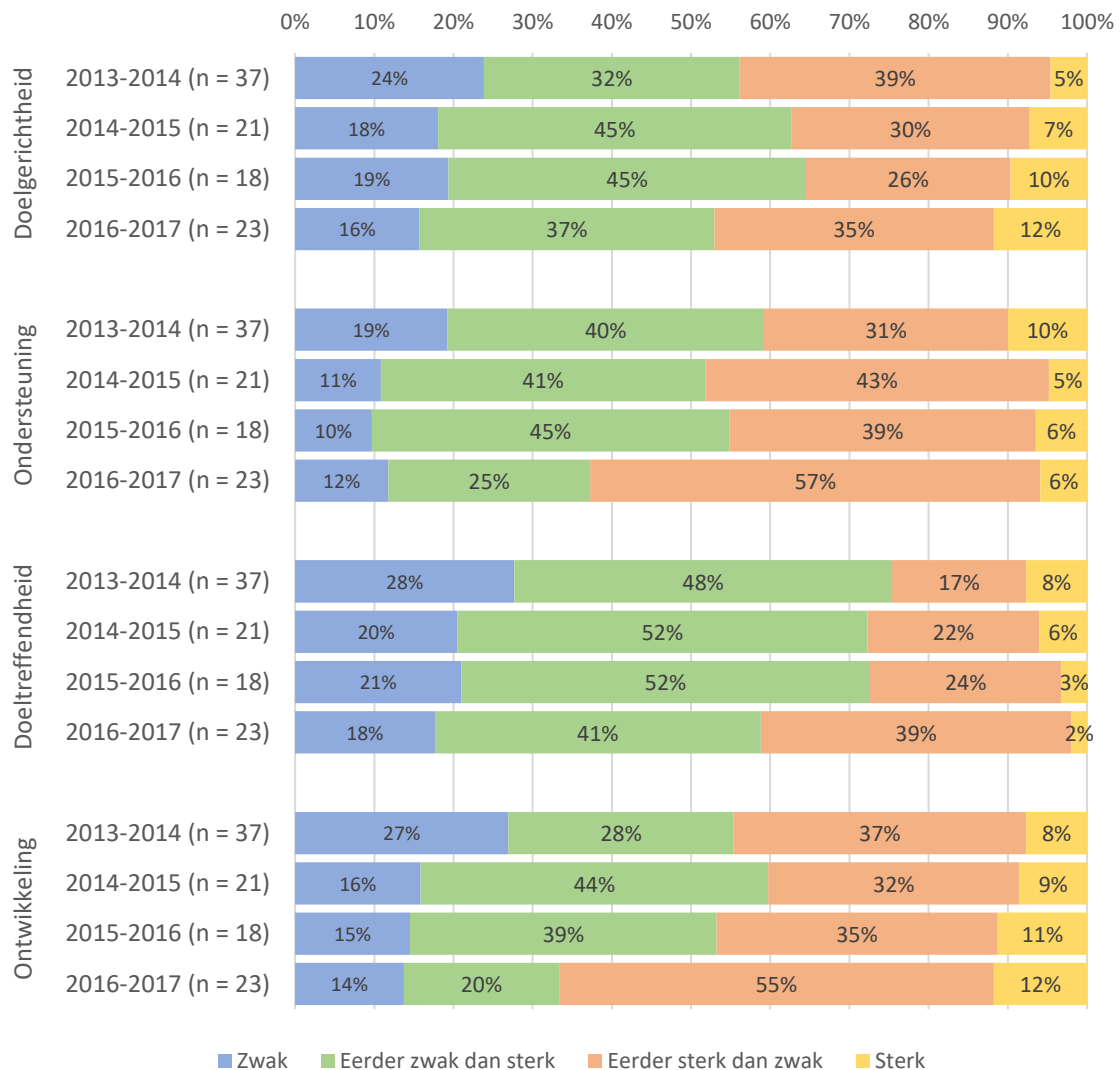


Figuur 124. Overzicht van de inschaling op de U1- en O1-schaal 'Afstemming van aanbod op gevalideerd doelenkader' in het secundair onderwijs per schooljaar (Inspectie 2.0)

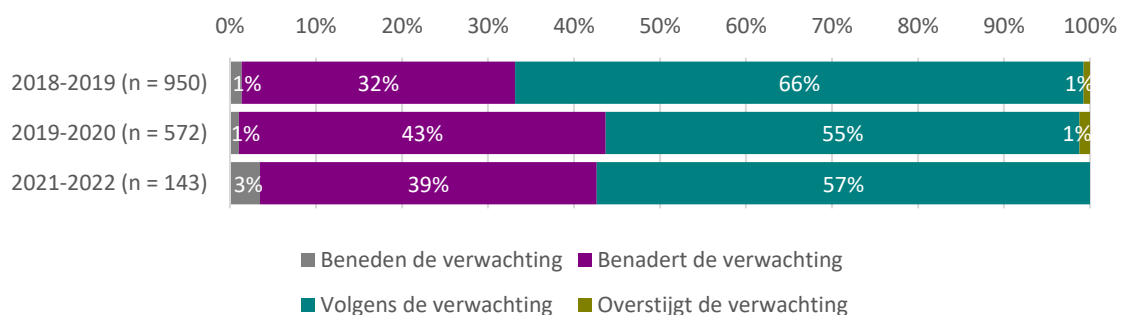
Van schooljaar 2013-2014 tot schooljaar 2016-2017 is er slechts een school voor secundair onderwijs beoordeeld voor het proceskenmerk 'onderwijsaanbod' in doorlichtingsronde 3. Bijgevolg konden we geen gelijkaardige figuur opstellen voor het proceskenmerk 'onderwijsaanbod'.

1.2 Evaluatiepraktijk (doorlichtingsronde 3) & Leerlingenevaluatie (Inspectie 2.0)

1.2.1 Basisonderwijs (gewoon en buitengewoon)

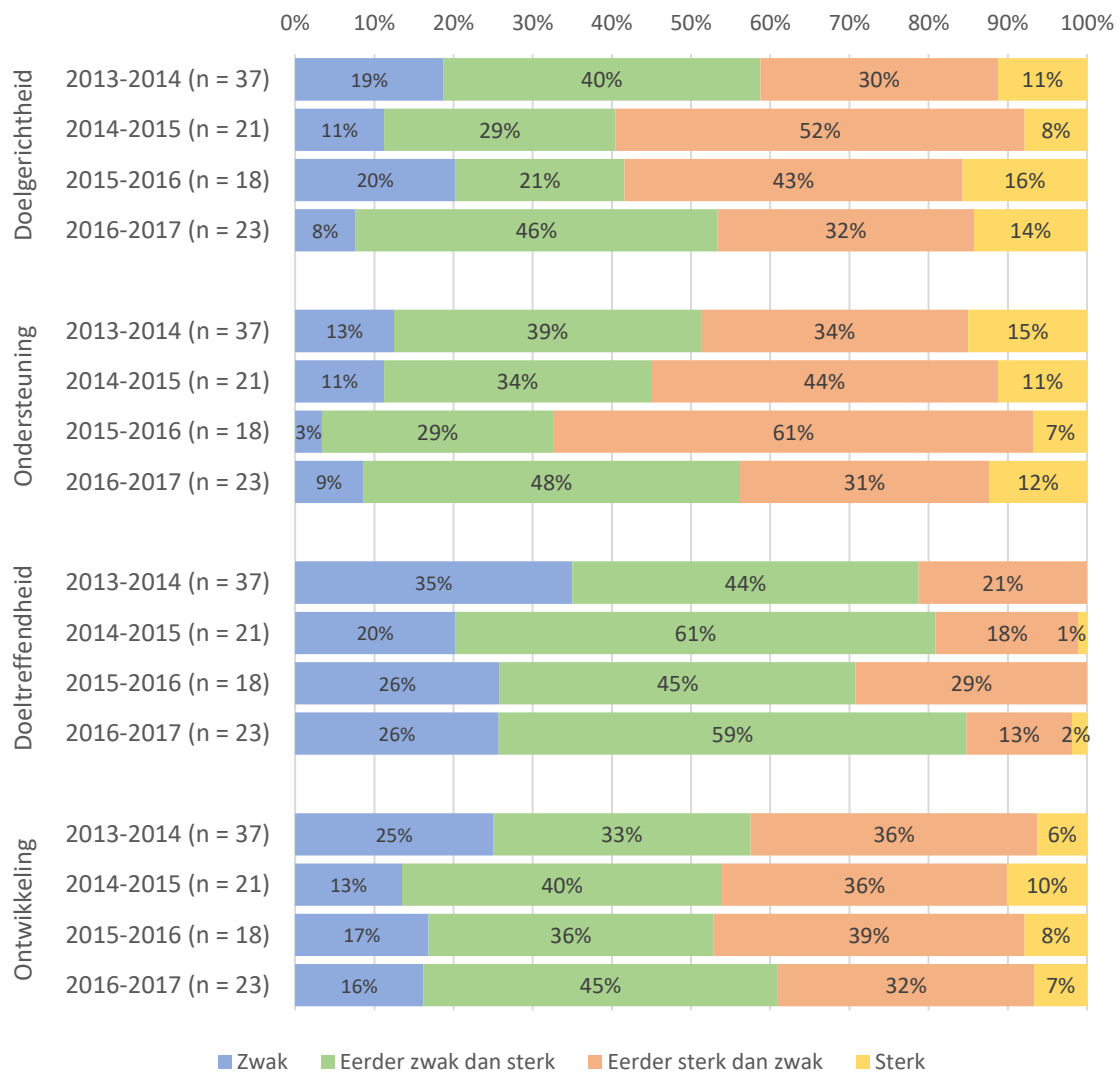


Figuur 125. Overzicht van de inschaling op het proceskenmerk 'evaluatiepraktijk' in het basisonderwijs per schooljaar (doorlichtingsronde 3)

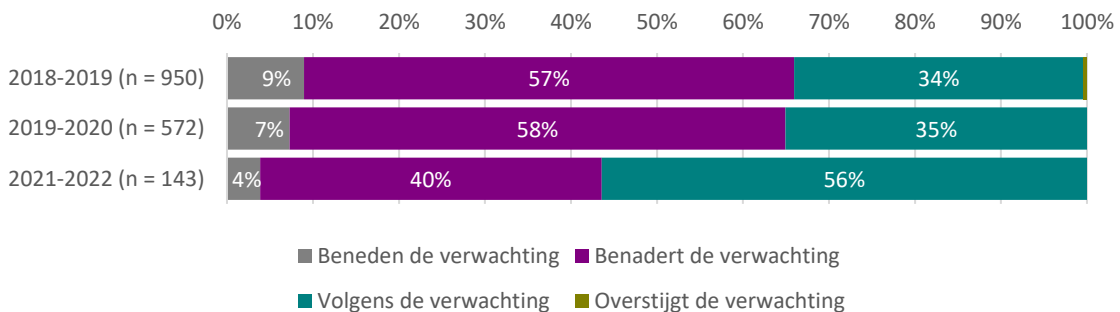


Figuur 126. Overzicht van de inschaling op de U6- en O7-schaal "Leerlingenevaluatie" in het basisonderwijs per schooljaar (Inspectie 2.0)

1.2.2 Secundair onderwijs (deeltijds, gewoon en buitengewoon)



Figuur 127. Overzicht van de inschaling op het proceskenmerk 'evaluatiepraktijk' in het secundair onderwijs per schooljaar (doorlichtingsronde 3)



Figuur 128. Overzicht van de inschaling op de U6- en O7-schaal 'Leerlingenevaluatie' in het secundair onderwijs per schooljaar (Inspectie 2.0)

2. Vergelijking van trends bij de peilingen en het internationaal vergelijkend onderzoek

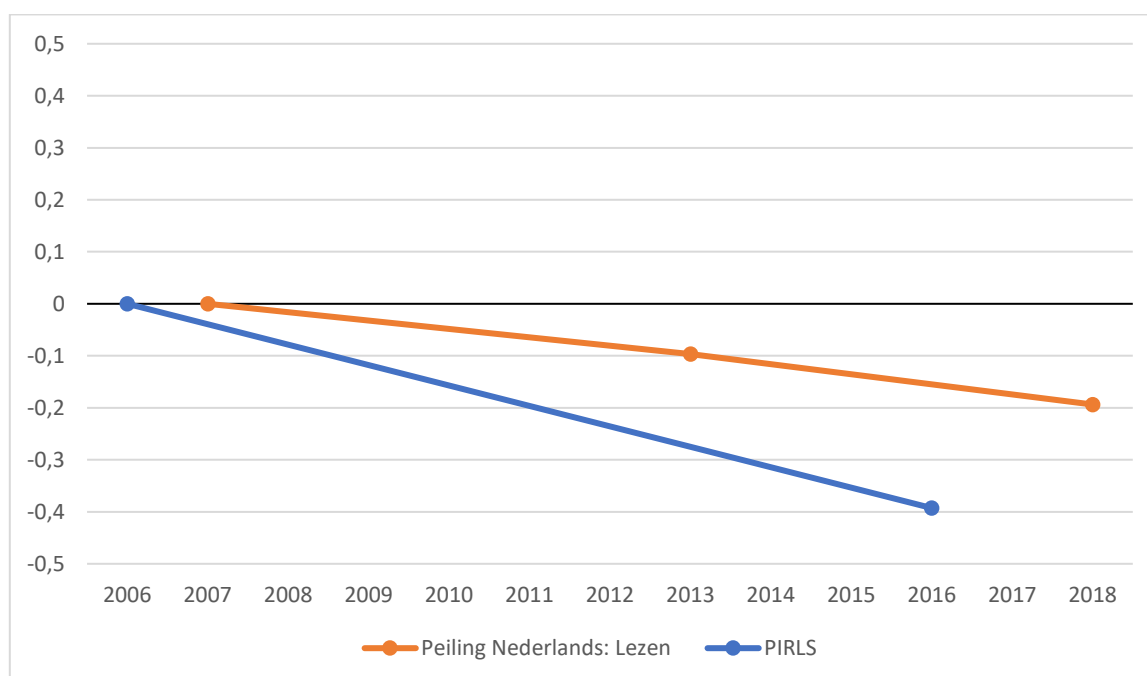
2.1 Peiling Nederlands en PIRLS (basisonderwijs)

Tabel 144

Trends in (begrijpend) lezen volgens de peiling Nederlands basisonderwijs en PIRLS

	Herschaalde trend		Jaarlijkse trend
	Helling	SE	Helling
PIRLS			
Begrijpend lezen	-0,3929	***	0,0493
Peiling Nederlands			
Lezen	-0,0969	0,0804	-0,0176

*Noot. *Het effect is statistisch significant op een significantieniveau van .05; ** op een significantieniveau van .01; *** op een significantieniveau van .001*



Figuur 129. Trends in (begrijpend) lezen volgens de peiling Nederlands basisonderwijs en PIRLS

2.2 Peiling wiskunde en TIMSS (basisonderwijs)

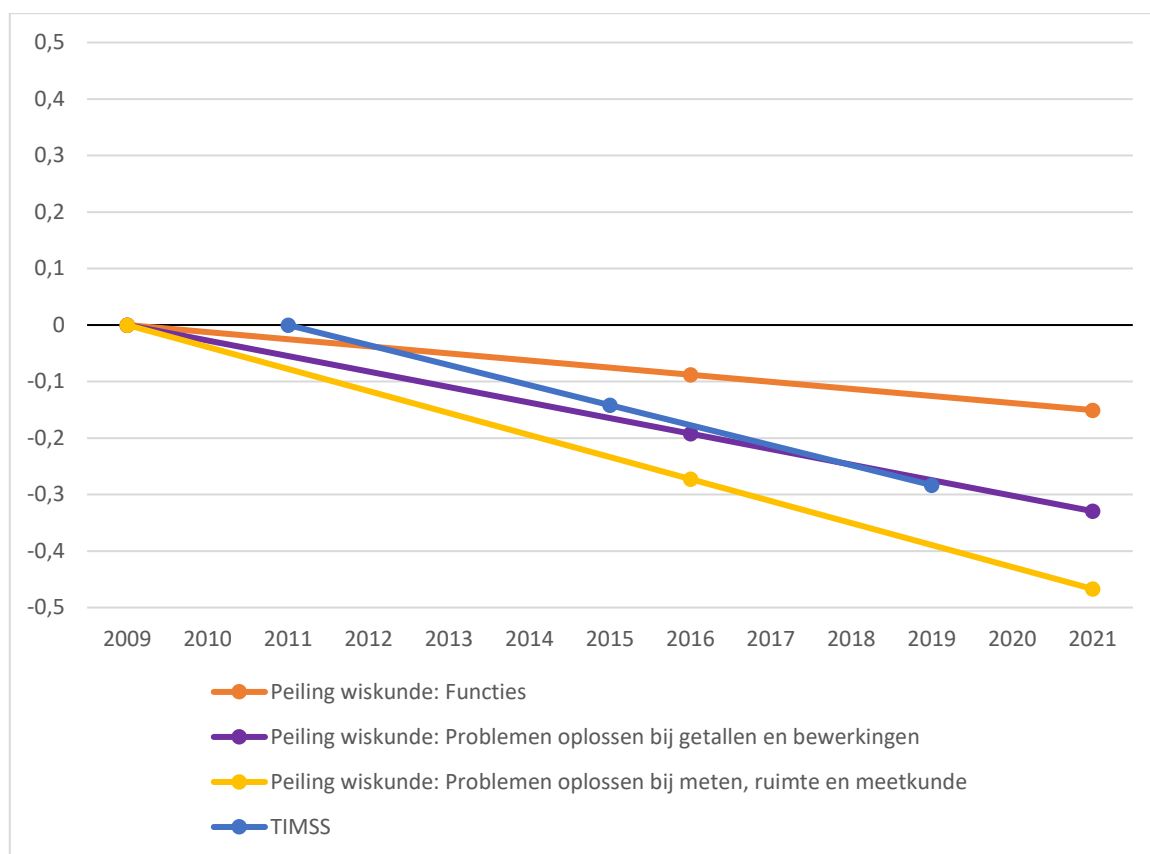
Tabel 145

Trends in wiskunde volgens de peiling wiskunde basisonderwijs en TIMSS

	Herschaalde trend		Jaarlijkse trend
	Helling	SE	Helling
TIMSS			
Wiskunde	-0,1417	**	0,0529
Peiling wiskunde^a			
Functies	-0,0761	***	0,0017
Problemen oplossen bij getallen en bewerkingen	-0,1708	***	0,0466
Problemen oplossen bij meten, ruimte en meetkunde	-0,2381	***	0,0249

*Noot. *Het effect is statistisch significant op een significantieniveau van .05; ** op een significantieniveau van .01; *** op een significantieniveau van .001*

^a Bij de peiling wiskunde basisonderwijs is de tijdsperiode tussen het eerste en tweede afnamemoment – 2009 en 2016 – groter dan de tijdsperiode tussen het tweede en derde afnamemoment – 2016 en 2021. Omwille van dat verschil hebben we de herschaalde en de jaarlijkse trend apart berekend. De herschaalde trend weerspiegelt de gemiddelde trend tussen twee afnamemomenten, terwijl de jaarlijkse trend de gemiddelde trend per jaar weergeeft.



Figuur 130. Trends in wiskunde volgens de peiling wiskunde basisonderwijs en TIMSS

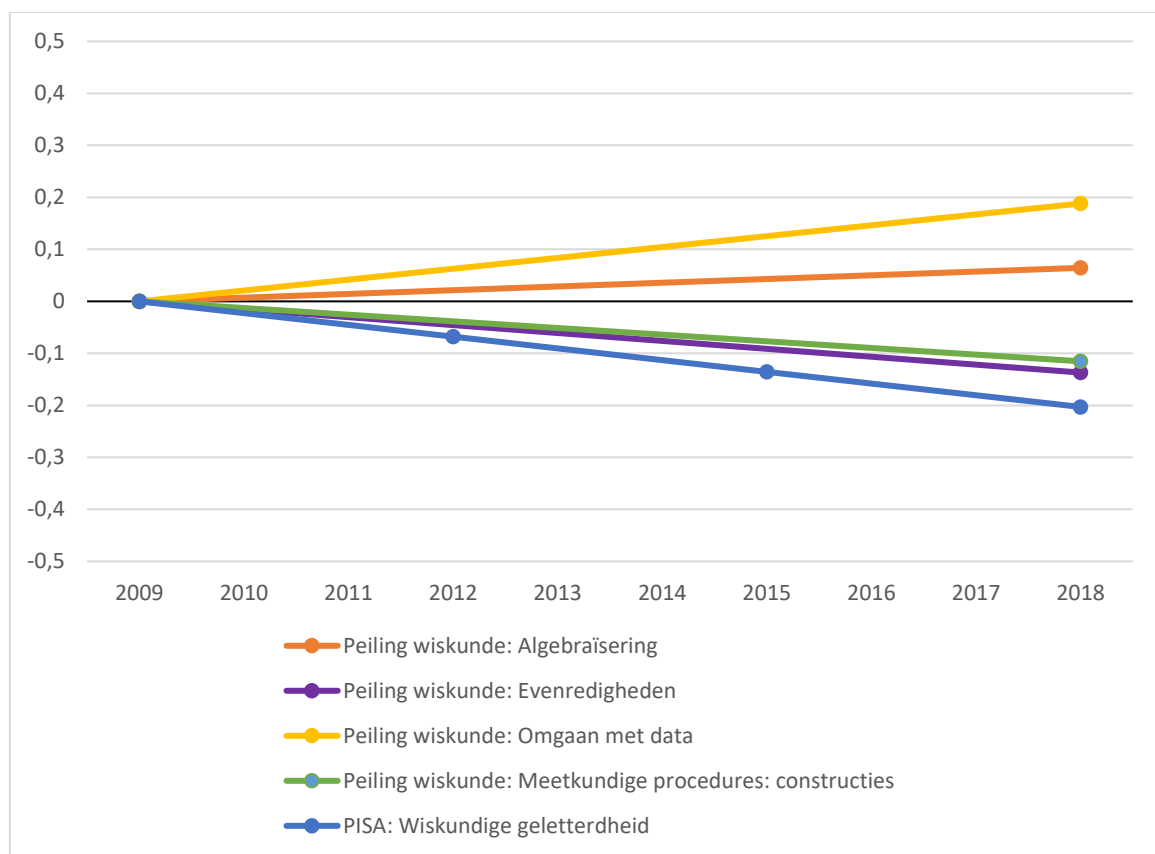
2.3 Peiling wiskunde in de A- en B-stroom en PISA (secundair onderwijs)

Tabel 146

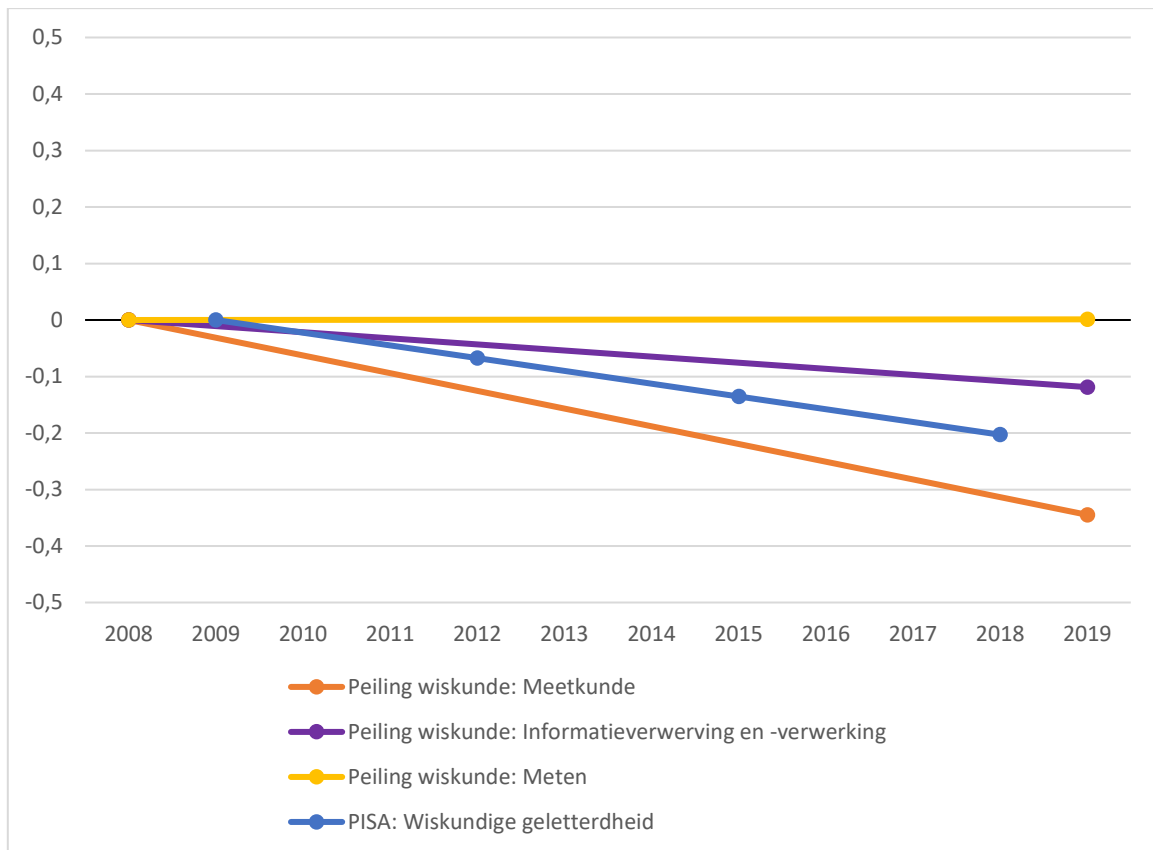
Trends in wiskunde volgens de peiling wiskunde A- en B-stroom en PISA

	Herschaalde trend		Jaarlijkse trend
	Helling	SE	Helling
PISA			
Wiskundige geletterdheid	-0,0677	***	0,0147
Peiling wiskunde A-stroom			
Algebraïsering	0,0644		0,0359
Evenredigheden	-0,1369	***	0,0358
Omgaan met data	0,1881	***	0,0406
Meetkundige procedures: constructies	-0,1152	**	0,0355
Peiling wiskunde B-stroom			
Meetkunde	-0,3449	***	0,0296
Informatieverwerking en -verwerking	-0,1188	**	0,0385
Meten	0,0012		0,0371

*Noot. *Het effect is statistisch significant op een significantieniveau van .05; ** op een significantieniveau van .01; *** op een significantieniveau van .001*



Figuur 131. Trends in wiskunde volgens de peiling wiskunde A-stroom en PISA



Figuur 132. Trends in wiskunde volgens de peiling wiskunde B-stroom en PISA

2.4 Bijkomende analyses

2.4.1 Inleiding

Zoals eerder aangehaald bij Studie 3 wordt de *equating error* waarschijnlijk onderschat bij de internationaal vergelijkende onderzoeken. PIRLS, PISA en TIMSS houden immers geen rekening met de instabiliteit van de parameterschattingen tussen landen en over de afnamecycli binnen landen (Glas & Jehangir, 2013; Robitzsch & Lüdtke, 2019; Sachse & Haag, 2017). De vergelijking van geschatte trends in verschillende onderwijssystemen heeft daarom een hogere onzekerheid dan traditioneel gedacht wordt. Als het doel is om de trend van slechts één onderwijssysteem te schatten, dan kan die onzekerheid wel verminderd worden door aparte itemparameters te schatten voor dat specifieke onderwijssysteem (Gebhardt & Adams, 2007; Sachse et al., 2016). Zo vonden Robitzsch en Lüdtke (2019) dat de trend van een individueel land bij PISA beter geschat kan worden door eerst een meetschaal specifiek voor dat land schatten.

Tegelijkertijd waren de peilingsonderzoeken voornamelijk gericht op de rapportering van het aandeel leerlingen dat de vooropgestelde cesuur – en dus de eindtermen – bereikt. Wegens de aard van de cesuurbepaling, waarin de volgorde van de itemmoeilijkheden een hoofdrol heeft, was de *equating* procedure van de peilingen vooral gericht op het behouden van de itemvolgorde. Bijgevolg werd een lineaire *equating* procedure toegepast die goed is voor dit doel, maar minder geschikt is om de herhalingspeiling op de meetschaal van de oorspronkelijke peiling te plaatsen. Daarvoor is een *true-score equating method* – zoals de internationaal vergelijkende onderzoeken gebruiken – beter geschikt.

In deze bijkomende analyses onderzoeken we daarom de trends van het internationaal vergelijkend onderzoek en peilingsonderzoek, waarbij we voor beiden eerst een nieuwe meetschaal schatten die

geoptimaliseerd is voor het schatten van een trend voor Vlaanderen. Zo voorkomen we dat instabiliteit van de itemparameters tussen landen bij internationaal vergelijkend onderzoek de trendschatting vertekent. Bovendien hanteren we een *true-score equating method* voor de peilingsonderzoeken. Voor deze analyses gebruiken we dezelfde gegevens van de peilingsonderzoeken en internationaal vergelijkende studies als de gegevens die in de voorgaande secties gebruikt werden.

2.4.2 Aanpak in de bijkomende analyses

Voor de schatting van nieuwe vaardigheidsscores gebruikten we de gescoorde itemantwoorden van de leerlingen. De meest voorkomende itemtype waren enkelvoudige meerkeuzevragen, samengestelde meerkeuzevragen en open vragen (Martin et al., 2020). Hieronder rapporteren we voor de verschillende studies het totaal aantal unieke items:

- TIMSS 2011, 2015 en 2019 meten de vaardigheid in wiskunde (319 items).
- PIRLS 2006 en 2016 meten de vaardigheid in begrijpend lezen (249 items)
- PISA 2009, 2012, 2015 en 2018 meten de vaardigheid in wiskundige geletterdheid (85 items)
- De peilingsonderzoeken Nederlands basisonderwijs in 2007, 2013 en 2018 meten de vaardigheid in lezen (137 items).
- De peilingsonderzoeken wiskunde basisonderwijs 2009, 2016 en 2021 meten verschillende deelvaardigheden die tot het domein wiskunde behoren (460 items).
- De peilingsonderzoeken wiskunde secundair onderwijs A-stroom in 2009 en 2018 meten verschillende deelvaardigheden die tot het domein wiskunde behoren (257 items).
- De peilingsonderzoeken wiskunde secundair onderwijs B-stroom 2009 en 2018 meten verschillende deelvaardigheden die tot het domein wiskunde behoren (251 items).

Om de gescoorde itemantwoorden van iedere leerling over de verschillende afnamecycli op één meetschaal te plaatsen (bijvoorbeeld PIRLS in 2006 en 2016) werden er IRT-modellen gebruikt. Een IRT-model beschrijft hoe de score op een item gerelateerd is aan de meetschaal van een vaardigheid. Op basis van die IRT-modellen en de leerlingen hun gescoorde itemantwoorden werden er vervolgens *plausible values* geschat.

De uitkomstanalyse van elke studie bestonden steeds uit een *multiple group* regressiemodel, waarbij we voor iedere afnamecyclus het gemiddelde en de variantie schatten. Bij elke analyse werd ook het steekproefontwerp, waaronder de stratificatie op schoolniveau in rekening gebracht. Zo gebruikten we bij TIMSS, PIRLS en PISA de *replicate weights* om de standaardfouten te schatten. Bij de peilingsonderzoeken gebruikten we de empirische *bootstrap* methode. Elke set van *plausible values* werd apart geanalyseerd, en de resultaten werden gecombineerd via Rubin's methode. Tenslotte gebruikten we ook de gewichten van TIMSS, PIRLS en PISA. Om de trend te beschrijven werd het verschil geschat tussen de gemiddeldes van de eerste en laatste afname.

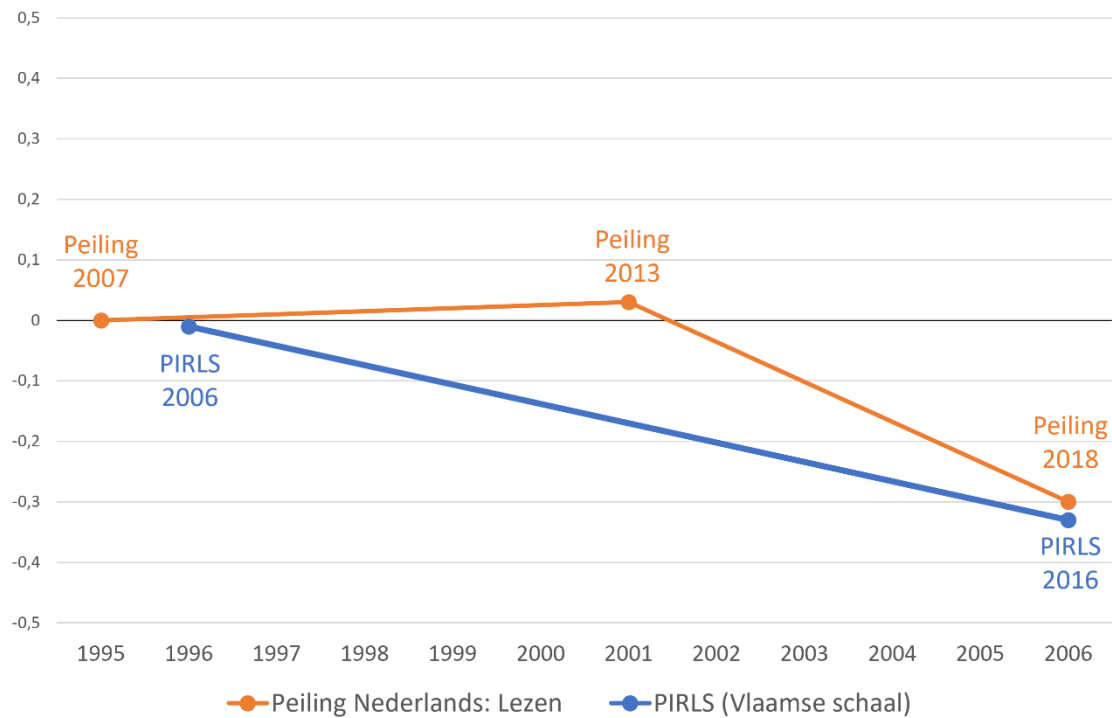
2.4.3 Resultaten

De trend volgens het peilingsonderzoek en de trend volgens de Vlaamse schaal van het internationaal vergelijkend onderzoek worden hieronder in dezelfde figuur weergegeven. De x-as geeft telkens het geboortjaar weer van de normaalvorderende leerlingen. Op die manier kunnen we visueel weergeven dat sommige bronnen eenzelfde geboortecohort meten. Dat is bijvoorbeeld het geval bij de peiling Nederlands lezen in 2018 en PIRLS in 2016 die beiden gericht waren op de geboortecohorte van 2006.

Ter informatie, de herkalibratie van PISA op basis van enkel de Vlaamse gegevens gebeurde voor alle onderwijsvormen samen. De trends voor de Vlaamse schaal van PISA worden in Figuur 135 en Figuur 136 wel apart weergegeven voor respectievelijk de A-stroom, ASO, TSO en KSO enerzijds en voor de B-stroom en BSO anderzijds.

2.4.3.1 Peiling Nederlands en PIRLS (basisonderwijs)

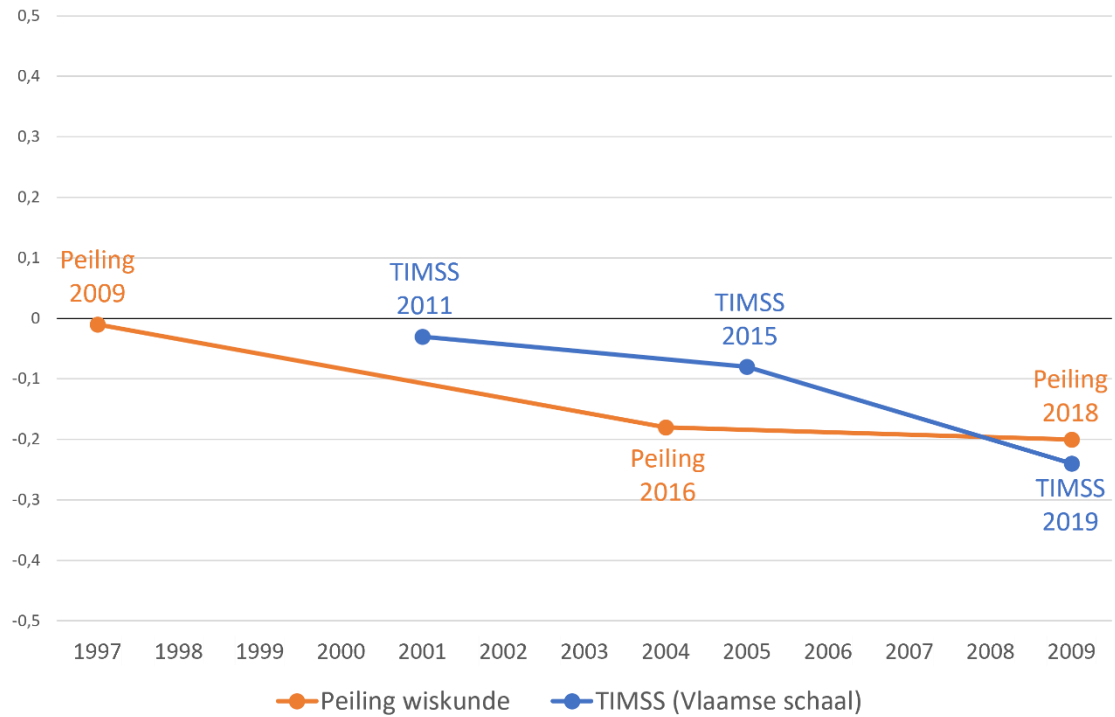
Bij het peilingsonderzoek is er tussen 2007 en 2018 een significant negatieve trend van -0.30, terwijl er bij PIRLS tussen 2006 en 2016 een significant negatieve trend is van -0.32. Voor het peilingsonderzoek lezen betreft het de trend volgens de geboortecohortes van 1995 en 2006, en voor PIRLS volgens 1996 en 2006. Er is geen significant verschil in trend tussen het peilingsonderzoek en de Vlaamse schaal van PIRLS (delta= 0.01, 95% BI [-0.10, 0.13]).



Figuur 133. Trends in (begrijpend) lezen volgens de peiling Nederlands basisonderwijs en PIRLS

2.4.3.2 Peiling wiskunde en TIMSS (basisonderwijs)

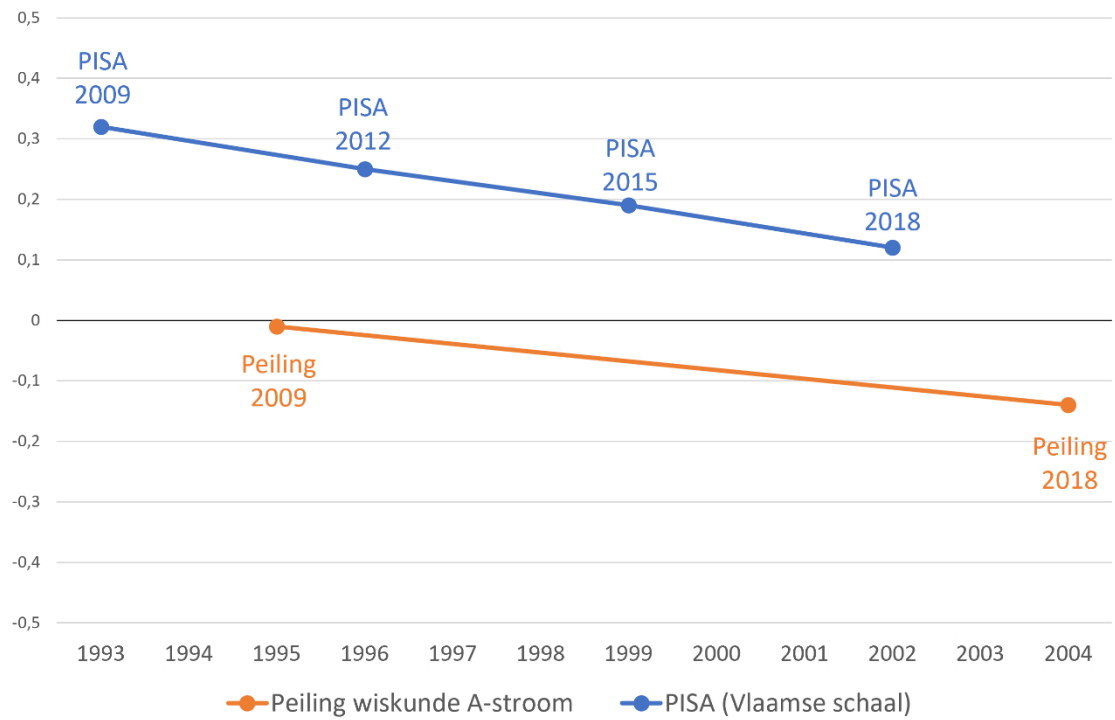
Voor de algemene schaal van wiskunde in het zesde leerjaar lager onderwijs tussen 2009 en 2021 is er in het peilingsonderzoek een significant negatieve trend van -0.19. Voor wiskunde in het vierde leerjaar is er in TIMSS tussen 2011 en 2019 een significant negatieve trend van -0.21. Voor het peilingsonderzoek wiskunde betreft het de trend volgens de geboortecohortes van 1997 en 2009, en voor TIMSS volgens 2001 en 2009. Er is geen significant verschil in trend tussen het peilingsonderzoek en de Vlaamse schaal van TIMSS (delta= 0.01, 95% BI [-0.10, 0.13]).



Figuur 134. Trends in wiskunde volgens de peiling wiskunde basisonderwijs en TIMSS

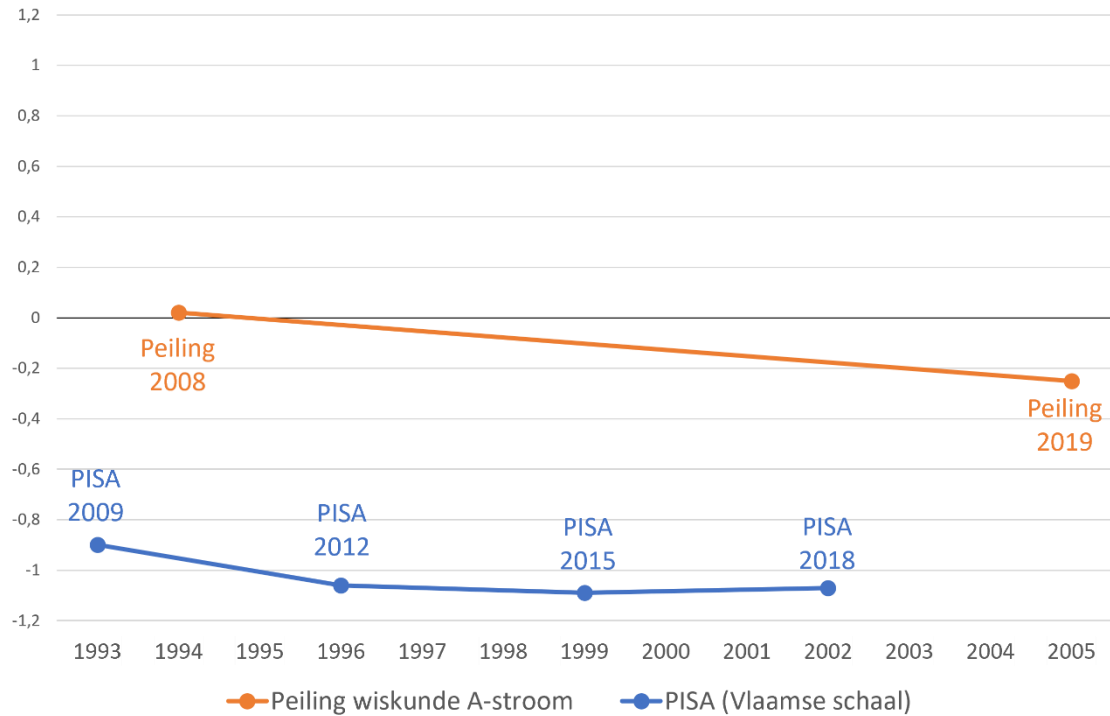
2.4.3.3 Peiling wiskunde in de A- en B-stroom en PISA (secundair onderwijs)

Bij het peilingsonderzoek vinden we voor de algemene schaal van wiskunde in het tweede leerjaar A tussen 2009 en 2018 een trend van -0,13, maar die trend is niet significant verschillend van 0. Voor wiskundige geletterdheid bij de 15-jarigen in de A-stroom is er in PISA tussen 2009 en 2018 een significant negatieve trend van -0.20. Er is geen significant verschil in trends tussen het peilingsonderzoek en de Vlaamse schaal van PISA (delta= -0.12, 95% BI [-0.26, 0.02]).



Figuur 135. Trends in wiskunde volgens de peiling wiskunde A-stroom en PISA

Voor de algemene schaal van de peiling wiskunde in de B-stroom is er tussen 2008 en 2019 in het peilingsonderzoek een significant negatieve trend van -0.26. Voor wiskundige geletterdheid bij de 15-jarigen in de B-stroom is er in PISA tussen 2009 en 2018 een significant negatieve trend van -0.18. Er is geen significant verschil in trends tussen het peilingsonderzoek en de Vlaamse schaal van PISA (delta= -0.01, 95% BI [-0.24, 0.21]).



Figuur 136. Trends in wiskunde volgens de peiling wiskunde B-stroom en PISA

Bijlage 4 – Leidraad focusgroepen

In **blauw** staan er concrete scenario's die voorgelegd kunnen worden aan de focusgroep om meer concrete en gedetailleerde antwoorden te ontrafelen.

<p>Openingsfase (5 min.)</p>	<p>Algemene informatie</p> <ul style="list-style-type: none"> • Bedankt voor deelname aan focusgroep. • Korte voorstelling van mezelf en (de doelstellingen van) het onderzoek. <p>Opbouw en doel van de focusgroep</p> <p>Het focusgroepgesprek verloopt als volgt:</p> <ul style="list-style-type: none"> • Eerst presenteer ik jullie de resultaten van het eerste deel van het onderzoek. Ik ga het hebben over: <ul style="list-style-type: none"> ○ De gelijkenissen en verschillen tussen de doorlichtingen, de peilingen en het internationaal vergelijkend onderzoek; ○ De mate waarin de resultaten van die verschillende bronnen overeenkomen of niet. • Daarna wil ik met jullie in gesprek gaan, aangezien jullie allemaal ervaring hebben met een of meerdere van die bronnen. We gaan het onder meer hebben over: <ul style="list-style-type: none"> ○ Mogelijke verklaringen voor de mate waarin verschillende bronnen al dan niet tot dezelfde resultaten komen; ○ Mogelijke manieren om de verschillende bronnen met elkaar te combineren. <p>Het gesprek duurt in totaal ongeveer twee tot drie uur.</p> <p>Administratieve zaken</p> <p>Vooraleer we starten met de presentatie, wil ik nog graag enkele zaken in orde brengen:</p> <ul style="list-style-type: none"> • Opname: Graag zouden wij deze focusgroepen digitaal willen opnemen. Zo moeten we niet alles noteren tijdens het gesprek zelf. Enkel ik zal toegang hebben tot de opname. De overheid heeft als opdrachtgever geen toegang tot de opname. • Anonimisering: Alle input uit de focusgroepen zal vertrouwelijk behandeld worden.
<p>Presentatie (50 min.)</p>	<p>Uitgebreide presentatie van de resultaten uit Module 2.1 en 2.2</p> <p>Eerst neem ik jullie mee door de resultaten van het eerste onderzoeksdeel. Stel gerust vragen als iets om meer verduidelijking vraagt.</p> <p>Zijn er nog zaken die om verduidelijking vragen bij het eerste deel?</p>
<p>Pauze (10 min.)</p>	

<p>(15 min.)</p>	<p>Het systeemniveau</p> <ul style="list-style-type: none"> • Welke verklaringen spelen er mogelijk op het systeemniveau? <u>Mogelijkheden tot doorvragen:</u> Heeft iemand iets anders opgeschreven? Hoe gaat dat juist in zijn werk? • Hoe hangen die verklaringen samen met de gelijkenissen en verschillen die er tussen de bronnen zijn? <u>Eventueel in herinnering brengen:</u> Design Doel Domein en construct Populatie Steekproef Modus Standaardbepaling Analytische methode <u>Mogelijkheden tot doorvragen:</u> In hoeverre zijn de verklaringen op systeemniveau hetzelfde als die op schoolniveau? Op welke vlakken verschillen ze?
<p>(5 min.)</p>	<p>We komen dus tot de volgende verklaringen: ... <i>(korte samenvatting van gesprek)</i>. Wil iemand daar nog iets aan toevoegen?</p>
<p>Pauze (10 min.)</p>	
<p>Vervolg (20 min.)</p>	<p>Manieren waarop bronnen gecombineerd kunnen worden</p> <p>Het schoolniveau</p> <ul style="list-style-type: none"> • Op welke manier kunnen we de doorlichtingsverslagen, de schoolfeedback van de peilingen en de schoolfeedback van de internationaal vergelijkende onderzoeken combineren op schoolniveau? <u>Mogelijkheden tot doorvragen:</u> Is dat wenselijk? Hoe kunnen die bronnen elkaar aanvullen? <u>Concreet scenario:</u> Als jouw school resultaten ontvangt van twee bronnen, hoe combineren jullie die resultaten? Hoe interpreteren jullie die resultaten? <u>Concreet scenario:</u> Als jouw school een ander resultaat behaalt op bron A dan op bron B, hoe ga je daarmee om? Kan je daarvan een voorbeeld geven?

<p>(20 min.)</p>	<p>Het systeemniveau</p> <ul style="list-style-type: none"> • Terug naar het systeemniveau, hoe kunnen we die bronnen combineren om tot meer algemene uitspraken over de Vlaamse onderwijskwaliteit te komen? <p><u>Concreet scenario:</u> Bijvoorbeeld, PIRLS 2016 wees op een sterke daling in presentatie voor begrijpend lezen. De peiling lezen uit 2013 gaf daarentegen aan dat 92% van de leerlingen de eindtermen voor lezen bereikte. De onderwijsspiegels tonen dan weer aan dat de overgrote meerderheid van de scholen volgens de verwachting scoort voor het leergebied Nederlands. Hoe moet het onderwijsveld dat dan interpreteren?</p> <p><u>Mogelijkheden tot doorvragen:</u> Hoe gaat dat concreet in zijn werk? Hoeveel gewicht geven we aan elke bron? Hoeveel waarde hechten we aan de resultaten van elke bron?</p> <ul style="list-style-type: none"> • Wat is zeker geen goede manier om bronnen te combineren? • Wat zijn randvoorwaarden om bronnen te kunnen combineren?
<p>(15 min.)</p>	<p>In de toekomst worden de peilingen vervangen door de centrale toetsen.</p> <ul style="list-style-type: none"> • Wat kunnen we uit dit onderzoek meenemen met oog op de centrale toetsen? <p><u>Mogelijkheden tot doorvragen:</u> Waar moet er rekening mee gehouden worden? Wat zijn mogelijke aandachtspunten uit dit onderzoek?</p> <ul style="list-style-type: none"> • De onderwijsinspectie zal inzage krijgen in de toetsresultaten van scholen op de centrale toetsen. Waar moet er rekening mee gehouden worden? <p>Dus als ik het goed begrijp, ... <i>(korte samenvatting geven)</i>. Klopt dat?</p>
<p>Afsluiting (5 min.)</p>	<p>Afronden van de focusgroep</p> <ul style="list-style-type: none"> • Zijn er nog zaken die jullie graag willen meegeven of kwijt willen? • Bedankt om deel te nemen.

Bijlage 5 – Codeboom

- Verklaringen
 - Doel (& Matrix)
 - Domein en construct
 - Populatie en steekproef
 - Modus
 - Standaardbepaling
 - Analytische methode
 - Andere
 - Meerderheid gemiddeld
 - Verandering/Momentopname
 - Algemene tendensen
 - Strategisch gedrag
 - Toetsen meten meer
 - Context
- Combinatie
 - Puzzelstukje (meerwaarde)
 - Duiding
 - Meerdere bronnen
 - Vertaling onderwijspraktijk
 - Schoolniveau
 - Interpretatie feedback
 - Toegankelijkheid
 - Tijd en ruimte gebrek
 - Divergerende resultaten
 - Nood begeleiding en datageletterdheid
 - Eigenaarschap
 - Databad
 - Systemniveau
 - Divergerende resultaten
 - Ontbrekende puzzelstukjes
 - Nood datageletterdheid en duiding
- Vlaamse toetsen
 - Puzzelstukje (meerwaarde)
 - Duiding
 - Meerdere bronnen
 - Vertaling onderwijspraktijk
 - Kansen en uitdagingen