

# Wetenschappelijke verantwoording Kleuter in beeld - Rekenen

Sanneke Schouwstra, Judith Vloedgraven, Agnes de Boer,  
Nienke Lansink en Lonneke Nikkels

Rekenen





## **Wetenschappelijke verantwoording Kleuter in beeld - Rekenen**

Sanneke Schouwstra

Judith Vloedgraven

Agnes de Boer

Nienke Lansink

Lonneke Nikkels

Met medewerking van Ron Engelen, Floor Scheltens, Julet Harms, Saskia van Berkel

Met dank aan alle leerkrachten, intern begeleiders en experts die in het proces van co-creatie met ons hebben meegedacht

© Cito B.V. Arnhem (2020)

Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van Cito B.V. worden openbaar gemaakt en/of verveelvoudigd door middel van druk, fotografie, scanning, computersoftware of andere elektronische verveelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook.

# Inhoud

<b>1</b>	<b>Inleiding</b>	<b>5</b>
<b>2</b>	<b>Uitgangspunten ontwikkeling instrument</b>	<b>7</b>
2.1	Algemene uitgangspunten	7
2.2	Meetpretentie	8
2.3	Doelgroep	11
2.4	Doel	12
<b>3</b>	<b>Beschrijving meetinstrument</b>	<b>13</b>
3.1	Opbouw van het meetinstrument	13
3.2	Ontwikkel- en constructieproces in co-creatie	14
3.3	Operationalisatie van de inhoud	16
<b>4</b>	<b>Dataverzameling</b>	<b>27</b>
4.1	Proefonderzoek	27
4.2	Steekproef kwaliteitsonderzoek	29
4.3	Ontwerp van het kwaliteitsonderzoek	31
4.4	Verwerking en analyses	33
4.5	Samenstelling versies uitgave	34
<b>5</b>	<b>Kalibratie kindroute</b>	<b>43</b>
5.1	Bank met opdrachten en observatiepunten	43
5.2	Meetmodel	44
5.3	Kalibratie	47
5.4	Toetsing van het IRT-model	48
<b>6</b>	<b>Standaardbepaling voor de kindroute</b>	<b>53</b>
6.1	Doel van standaardbepaling	53
6.2	Methode	53
6.3	Procedure en experts	53
6.4	Materiaal	55
6.5	Definitieve expertoordelen	56
6.6	Conclusie	59
<b>7</b>	<b>Beschrijving algoritmes</b>	<b>63</b>
7.1	Algoritme voor het rapporteren van het functioneringsniveau op subdomeinen	63
7.2	Algoritme voor een signaal voor Rekenen	67
<b>8</b>	<b>Betrouwbaarheid en nauwkeurigheid</b>	<b>69</b>
8.1	Interbeoordelaarsbetrouwbaarheid activiteit	69
8.2	Betrouwbaarheid	71
8.3	Nauwkeurigheid	72
<b>9</b>	<b>Validiteit</b>	<b>77</b>
9.1	Inhoudsvaliditeit	77
9.2	Begripsvaliditeit	78

<b>10</b>	<b>Afname en rapportage</b>	<b>87</b>
10.1	Afname	87
10.2	Rapportage	89

<b>Appendix</b>	<b>93</b>
-----------------	-----------

<b>Literatuur</b>	<b>97</b>
-------------------	-----------

# 1 Inleiding

Deze wetenschappelijke verantwoording heeft betrekking op Kleuter in beeld - Rekenen.

Kleuter in beeld - Rekenen is een instrument om de vaardigheid van kleuters op het gebied van rekenen in beeld te brengen, waarbij observaties centraal staan. Dit nieuwe observatie-instrument bestaat uit een leerkrachtmap met handleiding, instructies en materialen en daarnaast papieren opdrachtenboekjes. De opdrachten kunnen ook digitaal gemaakt worden via Basispoort. Bij dit instrument werkt de leerkracht vanuit een online omgeving: Kleuter in beeld online. Hierin kan de leerkracht observaties en opdrachten inplannen, observaties registreren en de resultaten van kinderen en groepen in handige overzichten bekijken.

Deze verantwoording biedt alle informatie die nodig is voor een beoordeling van de kwaliteit van het betreffende meetinstrument. Naast deze verantwoording horen de volgende producten in de beschouwingen te worden betrokken:

- Handleiding Kleuter in beeld - Rekenen: leerkrachtmap met handleiding en instructies leerkrachtroute, opdrachten en activiteiten
- Papieren opdrachtenboekjes
- Digitale opdrachten (via Basispoort)
- Kleuter in beeld online (via Basispoort)

Via het kopje 'Downloads' in Kleuter in beeld online zijn aanvullend nog de volgende documenten te vinden:

## **Algemeen**

- Handleiding Kleuter in beeld online
- Samen in gesprek (praatplaat voor scholen om in gesprek te gaan over het gebruik van Kleuter in beeld in de praktijk)

## **Rekenen**

- Inhoudsverantwoording Kleuter in beeld - Rekenen

Om Kleuter in beeld - Rekenen te kunnen beoordelen, beschrijven de hoofdstukken de volgende aspecten:

- Uitgangspunten ontwikkeling instrument (hoofdstuk 2)
- Beschrijving meetinstrument (hoofdstuk 3)
- Dataverzameling (hoofdstuk 4)
- Kalibratie (hoofdstuk 5)
- Standaardbepaling (hoofdstuk 6)
- Beschrijving algoritmes (hoofdstuk 7)
- Betrouwbaarheid (hoofdstuk 8)
- Validiteit (hoofdstuk 9)
- Afname en rapportage (hoofdstuk 10)

Het aspect validiteit betreft alleen begripsvaliditeit en géén criteriumvaliditeit. Kleuter in beeld - Rekenen is namelijk niet bedoeld voor 'voorspellend gebruik'.

De kwaliteit van het materiaal en de handleiding is te bepalen door kennis te nemen van de inhoud van de leerkrachtmap en van Kleuter in beeld online.

Voor de leesbaarheid van de tekst is gekozen voor de hij-vorm bij refereren aan de leerkracht en aan het kind. We benadrukken dat overal waar 'hij' staat, uiteraard ook 'zij' kan worden gelezen. Hetzelfde geldt voor 'zijn' of 'haar'.





## 2 Uitgangspunten ontwikkeling instrument

### 2.1 Algemene uitgangspunten

#### **Co-creatie**

Kleuter in beeld - Rekenen is ontwikkeld in co-creatie met het veld. De wensen uit het veld hebben centraal gestaan bij de ontwikkeling. Er zijn onder andere een focus-, screenings- en een expertbijeenkomst georganiseerd, zodat leerkrachten, intern begeleiders en andere experts mee konden denken over de concepten en de inhoud. Er was een speciale online community Kleuter in beeld - Taal, waarbij mensen uit het veld konden reageren op stellingen en zo hun voorkeuren konden aangeven. De input die op deze manier verkregen is, is ook voor Kleuter in beeld - Rekenen gebruikt. Daarnaast zijn meerdere toetsdeskundigen van Cito naar diverse scholen geweest om te praten met leerkrachten, intern begeleiders en kleuters zelf. Verder is er een proefonderzoek en grootschalig kwaliteitsonderzoek georganiseerd en hebben veel onderwijsprofessionals meegewerkt aan de standaardbepaling. Het resultaat is een instrument dat aansluit bij de wensen uit het veld en dat past bij de eigenheid van kleuters.

#### **Methodeonafhankelijk**

Kleuter in beeld - Rekenen is een methodeonafhankelijk volgsysteem. Dit betekent dat het te gebruiken is naast elke methode. Het belangrijkste doel van het instrument is om de vaardigheid van kleuters op een objectieve manier in beeld te brengen. Met deze informatie kan de leerkracht het onderwijsaanbod zo goed mogelijk afstemmen op de kinderen in de groep.

De inhoud van het instrument is gebaseerd op:

- de meest recente leerlijnen van SLO;
- de mening van leerkrachten en rekenspecialisten;
- de uitgebreide onderzoeken die zijn georganiseerd op scholen.

#### **Regie bij de leerkracht**

Op verschillende manieren is ervoor gezorgd dat de leerkracht weer de regie krijgt.

- De leerkracht ziet de kleuters dag in dag uit. Daardoor heeft hij al goed zicht op hun ontwikkeling. Dat is de reden dat de leerkrachtroute is ontwikkeld. Heeft de leerkracht twijfels bij een of meer kinderen over een of meer subdomeinen, dan kiest hij in die situaties voor de kindroute.
- Bij de opdrachten kan er gekozen worden voor papier of digitaal.
- Ook het moment waarop Kleuter in beeld - Rekenen wordt ingezet, is flexibel. De leerkracht plant de observaties, opdrachten en activiteiten op een moment waarop hij de informatie nodig heeft.

#### **Observaties centraal**

Observaties staan centraal bij Kleuter in beeld - Rekenen. Dit is een belangrijke wens vanuit het veld en doet recht aan de eigenheid van kleuters. Observaties zijn te onderscheiden in indirecte en directe observaties. Bij directe observatie gaat het om het observeren van gedrag op het moment dat het zich voordoet, zoals bijvoorbeeld bij de activiteiten uit de kindroute. Bij de leerkrachtroute daarentegen neemt de leerkracht het kind in gedachten en geef hij 'een oordeel' over de vaardigheid van het kind op basis van wat hij allemaal weet over het kind en wat het kind de afgelopen (wat langere) periode heeft laten zien in de groep. De leerkracht kijkt dus niet alleen op 'moment X' wat het kind laat zien, maar vormt zich een 'totaalbeeld' op basis van zijn dagelijkse observaties die hij op meerdere momenten heeft gedaan. In die zin is er sprake van een meer 'indirecte' observatie. Bij directe observatie gaat het om het observeren van gedrag op het moment dat het zich voordoet.

- De leerkrachtroute is volledig gebaseerd op observaties. Het is algemeen bekend dat kleuters zich sprongsgewijs ontwikkelen. Door het niveau in te schatten op basis van zijn eigen (indirecte) observaties over een langere periode, wordt voorkomen dat het bepalen van het niveau een momentopname is.

- De kindroute bestaat uit opdrachten en activiteiten. Bij de activiteiten staan directe observaties centraal. De leerkracht voert samen met het kind een activiteit uit en vult observatiepunten in op basis van wat hij op dat moment ziet.

Kwaliteit staat voorop, daarom zijn er diverse onderzoeken uitgevoerd om de observaties zo gestandaardiseerd en objectief mogelijk te maken.

### **Speels**

Kleuters komen het beste tot hun recht als de inhoud in een speelse context wordt aangeboden. Daarom zijn de opdrachten en activiteiten op een toegankelijke en eigentijdse manier opgebouwd en vormgegeven. Zo staan bij de opdrachten en activiteiten de vriendjes Anne en Mo en hun hondje Bink centraal. Verder is zoveel mogelijk de aansluiting gezocht bij de belevingswereld van kleuters. De opdrachten en activiteiten gaan over voor kleuters aansprekende onderwerpen, zoals tekenen, klimrek, zandbak, huisdieren, opa/oma en eten. Ook de activiteiten zelf zijn speels van opzet: bij bijvoorbeeld de activiteit Dierentuinspel (Verbanden) wordt aan de kinderen gevraagd plaatjes van dieren te sorteren en bij de activiteit Eten bij opa en oma (Bewerkingen) onder andere om broodjes eerlijk te verdelen.

De opdrachten en activiteiten die aan de kinderen worden voorgelegd, worden aangeboden in een dagelijkse schoolpraktijk, zonder tijdsdruk, waarbij er voor het kind geen merkbaar onderscheid is tussen leren, spelen en geobserveerd worden.

## **2.2 Meetpretentie**

Het doel van dit instrument is om de vaardigheid van kleuters op het gebied van rekenen op een objectieve manier in beeld te brengen. De keuze voor de verschillende inhoudsgebieden binnen Kleuter in beeld - Rekenen is grotendeels gebaseerd op de inhoudskaart Rekenen-Wiskunde kleuters (fase 1) van SLO (2018). De inhoudskaart van SLO hanteert een indeling in zes subdomeinen. Voor Kleuter in beeld - Rekenen is gekozen om te rapporteren op vijf subdomeinen, waarbij het subdomein Verhoudingen uit de SLO inhoudskaart is ondergebracht bij de subdomeinen Getalbegrip en Meten.

Er is bewust gekozen voor een rapportage waarin de groei en het niveau van het kind op elk van de vijf subdomeinen wordt weergegeven in plaats van alleen een resultaat op het totaal van Rekenen. De rapportage geeft zo handvaten om het onderwijs zo goed mogelijk op de kinderen af te stemmen. Daarmee is tijdige signalering mogelijk en is het instrument een hulpmiddel voor de leerkracht. Op welke subdomeinen gaat het boven verwachting en kan het kind extra uitdaging gebruiken? Op welke subdomeinen heeft het kind misschien extra hulp nodig?

Veel rekenproblemen zijn terug te voeren op onvoldoende ontwikkelde voorbereidende rekenvaardigheden (Geary, 1990; Van Luit, 2009). Voorbereidend rekenen vormt de basis van al het rekenen en richt zich op het verwerven van getalbegrip, classificeren, heen-en-terug tellen tot 20, verkennen van rekenkundige begrippen (veel, weinig, meer, minder, enzovoort), leren van basisvaardigheden voor meten (vormen, eigenschappen, vergelijken, ordenen), meten en tijd (tijdsverloop, synchroniteit, enzovoort) en operaties (erbij doen, samenvoegen, weghalen, telkens wegnemen) (Van Vugt & Wösten, 2013). Het is hierbij niet noodzakelijk dat een kind deze rekenactiviteiten direct met volledig begrip van hoeveelheid kan uitvoeren en kwantificeren. Voorbereidend rekenen speelt zich vooral af in groep 1 en 2 en gaat geleidelijk over in het aanvankelijk rekenonderwijs dat zich richt op optellen en aftrekken tot 20, overbruggen van het tiental, verkennen van het rekenkundig begrip 'is gelijk aan', tellen met grotere sprongen op de getallenlijn (2 en meer) en ontdekken van regelmatigigheden hierin, herhaald op- en aftellen, verkennen van de getallenrij tot 100, exploreren van eenheden en tientallen en verkennen van eenvoudige meet-, tijd- en geldbegrippen. Hiaten in het voorbereidend rekenen in de kleuterleeftijd zijn een belangrijke voorspeller voor rekenproblemen in de latere basisschoolcarrière. De koppeling tussen hoeveelheden en getalsymbolen en telwoorden is belangrijk voor het leren rekenen. Belangrijke voorwaarden voor het ontstaan van deze koppeling zijn vaardigheden als het kennen van de telrij en het herkennen van cijfersymbolen (Kolkman & Kroesbergen, 2014).

Bij kleuters spreek je nog niet van dyscalculie. Langberg, Van Zanten & Boswinkel (2014) geven aan dat in de literatuur verschillende omschrijvingen van de term dyscalculie te vinden zijn. Het protocol Ernstige RekenWiskunde-problemen en Dyscalculie (ERWD) (Van Groenestijn, Borghouts en Janssen, 2011) maakt onderscheid tussen ernstige rekenwiskunde-problemen en dyscalculie en hanteert de volgende werkdefinities:

‘Ernstige rekenwiskunde-problemen kunnen ontstaan wanneer het gedurende langere tijd niet lukt om de juiste afstemming te realiseren van het onderwijsaanbod op de onderwijsbehoeften van de leerling’ (p.50).

‘Wij spreken van dyscalculie als ernstige rekenwiskunde-problemen ontstaan ondanks tijdig ingrijpen, deskundige begeleiding en zorgvuldige pogingen tot afstemming. De problemen blijken hardnekkig te zijn. De rekenwiskundige ontwikkeling van de leerling wordt waarschijnlijk belemmerd door kindfactoren’ (p.50).

Jansen & Kroesbergen (2012) noemen een aantal specifieke signalen die bij kleuters erop (kunnen) duiden dat de rekenontwikkeling moeizaam verloopt of mogelijk vastloopt:

- mogelijk langzamer leren tellen, mogelijk in combinatie met meer fouten maken bij het tellen
- moeite met het ‘zien’ van structuur en met het zelf structureren van hoeveelheden
- moeite met het ordenen en vergelijken van hoeveelheden en met het schatten van aantallen
- moeite met betekenis verlenen aan rekentaal en symbolisering
- moeite met ruimtelijk voorstellen

Rekenproblemen vanaf groep 3 zijn vaak terug te voeren op een gebrekkige reken-wiskunde-ontwikkeling in de kleutertijd. Veel problemen kunnen voorkomen worden door in de kleuterperiode de ontwikkeling regelmatig in kaart te brengen. Als signalen tijdig worden opgemerkt, kan de leerkracht tijdig en gericht ingrijpen om erger te voorkomen. Kinderen krijgen zo de kans om optimaal van het onderwijsaanbod te profiteren (Van Groenestijn et al., 2011; Van Luit, 2009; Janssen & Kroesbergen, 2012).

Het protocol ERWD adviseert leerkrachten om een kind extra activiteiten aan te bieden binnen een rijke leeromgeving als er vermoedens zijn van een minder gunstige rekenontwikkeling (zie ook Treffers, Van den Heuvel-Panhuizen & Buys, 2009; Toll & Van Luit, 2014; Van den Heuvel-Panhuizen & Buys, 2009; Veltman & Van den Heuvel-Panhuizen, 2010). Dat kan bijvoorbeeld door het kind activiteiten aan te bieden waarbij het kan ervaren, verkennen en onderzoeken. Activiteiten waarbij het kind binnen een context kan doen alsof, experimenteren en concrete problemen oplossen, zijn ook geschikt. Door dit soort activiteiten in te passen in het reguliere aanbod, kunnen leerkrachten kinderen helpen om in allerlei alledaagse en speelse situaties hun vaardigheden verder te ontwikkelen en toe te passen. Een deel van de kinderen heeft meer nodig. Bij die kinderen kiest de leerkracht voor een intensievere aanpak, waarbij het kind ook meer aandacht en tijd krijgt. Daarbij is het van belang om de ontwikkeling die het kind doormaakt, goed te volgen.

### **Keuzes: subdomeinen**

Bij de keuzes voor de subdomeinen is de indeling van de inhoudskaart Rekenen-wiskunde kleuters (fase 1) van SLO (2018) gevolgd. Hierin worden vier domeinen onderscheiden, namelijk 1) Getallen (met de subdomeinen getalbegrip en bewerkingen), 2) Meten & Meetkunde (met de subdomeinen meten en meetkunde), 3) Verhoudingen en 4) Verbanden.

Deze indeling komt overeen met de uitwerking van het referentiekader taal en rekenen (Ministerie van OCW, 2009). In het referentiekader staat beschreven wat kinderen op verschillende momenten in hun schoolloopbaan op het gebied van taal en rekenen moeten kennen en kunnen (Expertgroep doorlopende leerlijnen Taal en Rekenen, 2008). In de nadere concretisering van het referentiekader wordt binnen Getallen geen expliciet onderscheid gemaakt tussen getalbegrip en bewerkingen, zoals dat wel in de inhoudskaart van SLO gebeurt. Er is gekozen voor de indeling van SLO, omdat deze inhoudskaart een belangrijke leidraad is voor het kleuteronderwijs, en leerkrachten deze indeling in subdomeinen in het instrument direct zullen herkennen. Het onderdeel Verhoudingen uit de inhoudskaart van SLO is inhoudelijk

te klein om apart te bevragen (en erover te rapporteren). Daarom is ervoor gekozen om dit onderdeel onder te brengen bij Getalbegrip en Meten.

### **Keuzes: onderdelen**

Onder elk subdomein vallen verschillende, uiteenlopende leerdoelen. Naast de inhoudskaart van SLO is aanvullende literatuur geraadpleegd. Literatuur waarin doelen en leerlijnen zijn uitgewerkt en waarin wordt besproken welke kerninzichten leerlingen op het gebied van rekenen-wiskunde tijdens de basisschooltijd moeten verwerven, zoals de publicaties:

- Jonge kinderen leren rekenen (Treffers, Van den Heuvel-Panhuizen & Buys, 2009) en Jonge kinderen leren meten en meetkunde (Van den Heuvel-Panhuizen & Buys, 2009) van het TAL-team
- TULE (Tussendoelen & leerlijnen) van SLO (2009)
- Tussendoelen Rekenen-Wiskunde voor het primair onderwijs (Noteboom, Aartsen & Lit, 2017)
- Rekenen met hele getallen op de basisschool: Tussendoelen annex leerlijnen (Veltman & Van den Heuvel-Panhuizen, 2010)
- Rekenen-wiskunde in de praktijk. Kerninzichten (Oonk, Keijzer, Lit, Barth, Den Engelsens, Lek & Van Waveren-Hogervorst, 2015)
- Rekenen een hele opgave. Deel 1 (Van Vugt & Wösten, 2015)

Ook hebben leerkrachten, ib'ers en experts input geleverd voor Kleuter in beeld - Rekenen, onder meer over welke doelen in het instrument aan bod moeten komen. Alle doelen uit de inhoudskaart van SLO aan bod laten komen, is te omvangrijk en leidt mogelijk tot een gebruikersonvriendelijk instrument. De betrokken leerkrachten, ib'ers en experts vonden veel doelen belangrijk. Bij de subdomeinen die in kaart worden gebracht met opdrachten (Getalbegrip en Meten) is een groot deel van de doelen meegenomen. Dat geldt ook voor de activiteiten Verbanden. De inhoud van de activiteit Bewerkingen is een uitwerking van de aanbodsdoelen uit de inhoudskaart van SLO. Hierbij is ook gekeken naar hoe het onderdeel Bewerkingen in kleutermethoden aan bod komt en beschreven staat in de literatuur. Voor de opdrachten is gekozen om te kijken naar vier onderdelen, namelijk optellen, aftrekken, splitsen en (ver)delen. Dit in tegenstelling tot hoe het subdomein Bewerkingen in de inhoudskaart van SLO is opgesplitst in de twee onderdelen 'optellen en aftrekken' (waar splitsen wel onder valt) en 'vermenigvuldigen en delen' (waaronder dan (ver)delen valt). De inhoudskaart van SLO richt zich op fase 1: de groepen 1, 2 en 3 van het basisonderwijs. Het aspect vermenigvuldigen past minder goed bij het aanbod dat leerkrachten in de groepen 1 en 2 verzorgen. De aspecten optellen, aftrekken, splitsen en (ver)delen sluiten wel goed aan bij het aanbod in de groepen 1 en 2. Daarom hebben we ervoor gekozen om ons in de activiteit Bewerkingen op deze aspecten te richten. Bij de activiteit Meetkunde zijn keuzes gemaakt (mede) op basis van wat praktisch goed uitvoerbaar is en op basis van feedback van leerkrachten, ib'ers en experts, bijvoorbeeld wat betreft de benodigde materialen en instructie die de leerkracht aan het kind moet geven.

In Tabel 2.1 staat een overzicht van de vijf subdomeinen met hun bijbehorende onderdelen die een plek hebben gekregen in Kleuter in beeld - Rekenen.

Tabel 2.1 Subdomeinen en onderdelen van *Kleuter in beeld – Rekenen*

Subdomein	Onderdelen
<b>Getalbegrip</b>	<ul style="list-style-type: none"> <li>- Telrij</li> <li>- Hoeveelheden</li> <li>- Getallen</li> <li>- Relaties tussen telrij, hoeveelheden en getallen incl. Verhoudingen</li> </ul>
<b>Bewerkingen</b>	<ul style="list-style-type: none"> <li>- Optellen en aftrekken met hele getallen (tot ten minste 20) (optellen, aftrekken, splitsen)</li> <li>- Vermenigvuldigen en delen met hele getallen (tot ten minste 20) ((ver)delen)</li> </ul>
<b>Meten</b>	<ul style="list-style-type: none"> <li>- Lengte en omtrek</li> <li>- Oppervlakte</li> <li>- Inhoud</li> <li>- Gewicht</li> <li>- Tijd</li> <li>- Geld</li> <li>incl. Verhoudingen</li> </ul>
<b>Meetkunde</b>	<ul style="list-style-type: none"> <li>- Oriënteren in de ruimte</li> <li>- Construeren</li> <li>- Opereren met vormen en figuren</li> </ul>
<b>Verbanden</b>	<ul style="list-style-type: none"> <li>- Verbanden</li> </ul>

## 2.3 Doelgroep

De doelgroep van dit instrument zijn kinderen in groep 1 en 2 van het primair en speciaal onderwijs. Het is daardoor ook in te zetten bij kinderen met een ontwikkelingsachterstand en/of extra onderwijsbehoeften. Bij kinderen voor wie Nederlands niet de moedertaal is, is het advies om te wachten tot het kind minimaal een half jaar op school zit.

Bij de ontwikkeling van het instrument is rekening gehouden met specifieke kenmerken van de doelgroep kleuters:

- Het niveau waar kleuters mee binnenkomen is zeer divers vanwege onder meer erfelijke factoren en het verschil in aanbod dat ze van thuis hebben meegekregen. De inhoud van het instrument past dan ook bij uiteenlopende vaardigheidsniveaus. Voor elk van de activiteiten is er één versie die geschikt is voor zowel kinderen in groep 1 als groep 2. Bij de opdrachten zijn er drie verschillende versies voor verschillende niveaus:
  - ★ = niveau groep 1
  - ★★ = niveau midden groep 2
  - ★★★ = niveau eind groep 2
 Dit betekent dat een kind uit groep 1 bijvoorbeeld ook de opdrachten van het niveau midden groep 2 kan maken, als een kind op een bepaald subdomein volgens de inschatting van de leerkracht enigszins voorloopt op andere kinderen.
- Kleuters ontwikkelen zich sprongsgewijs. Daarom is het bijvoorbeeld belangrijk kleuters op basis van de resultaten niet meteen in een hokje te plaatsen. Er is dan ook gekozen voor een absolute normering in plaats van een relatieve normering (zie voor meer informatie hoofdstuk 6 Standaardbepaling). En als een kind onder verwachting scoort, is dat wel een signaal om hier beter naar te kijken, maar nog geen reden tot directe ongerustheid.

- Bij kleuters speelt de leeftijd een belangrijke rol, nog meer dan bij kinderen uit groep 3 tot en met 8. Het maakt nogal uit of het gaat om een relatief jongere of oudere kleuter. Daarom benadrukken we bij het interpreteren van de resultaten ook de regie van de leerkracht. Het is belangrijk dat wanneer de leerkracht het niveau op de verschillende subdomeinen bekijkt, hij ook de leeftijd van het kind in gedachten houdt.
- Kleuters leren vooral spelenderwijs. Ze construeren kennis namelijk op een experimentele, interactieve, concrete en praktische wijze (Bredenkamp & Rosegrant, 1995) in plaats van door abstract redeneren, zoals oudere kinderen en volwassenen leren. Ter voorbereiding op groep 3 wordt bij kleuters ook aandacht besteed aan werken op het platte vlak. Bij Kleuter in beeld - Rekenen wordt zowel een beroep gedaan op het handelen (activiteiten) als op het platte vlak (opdrachten).

## 2.4 Doel

Kleuter in beeld - Rekenen is bedoeld om de rekenvaardigheid van kleuters op een objectieve manier in beeld te brengen. Dit kan voor vijf verschillende subdomeinen. Er worden drie verschillende doelen onderscheiden:

- Niveaubepaling: het bepalen van het niveau voor de vijf subdomeinen binnen Rekenen. In de online omgeving van Kleuter in beeld - Rekenen staat in de kindrapportage het behaalde niveau op elk van de vijf subdomeinen. In de groepsrapportage staat een overzicht van het niveau op de vijf subdomeinen voor een hele groep. De niveaus zijn gebaseerd op een absolute normering in plaats van een relatieve normering. Kinderen worden dus niet vergeleken met een landelijke referentiegroep (zoals A t/m E of I t/m V). Tijdens een standaardbepaling hebben leerkrachten en intern begeleiders op basis van leerlijnen, empirie van de onderzoeken en hun eigen ervaring bepaald wat doorgaans gezien wordt bij een kleuter midden groep 1, eind groep 1 en eind groep 2. De gekozen grenzen zijn daarmee inhoudelijk bepaald.
- Progressiebepaling: wanneer er sprake is van meerdere observatieperiodes, is de ontwikkeling van een kind in de loop van de kleuterperiode te zien. Het gaat dan om een vergelijking van de resultaten van hetzelfde kind met zijn eigen resultaten op eerdere of latere afnamemomenten van Kleuter in beeld. Het instrument geeft antwoord op vragen als: Is er sprake van vooruitgang, achteruitgang of van stabilisering? Om de progressie in het functioneringsniveau van de kinderen te kunnen bepalen is het domein opgevat als een unidimensionale vaardigheid, of 'latente trek'. Hierdoor ontstaat een vaardigheidsschaal. De ruwe scores bij de activiteiten en opdrachten (de ruwe score is het aantal opdrachten/observatiepunten goed) zijn daartoe getransformeerd in scores op één vaardigheidsschaal met behulp van het One Parameter Logistic Model (Verhelst, 1993; Verhelst & Glas, 1995; Verhelst, Glas, & Verstralen, 1995). De standaarden voor de grenzen tussen de functioneringsniveaus zijn ook uitgedrukt in scores op deze vaardigheidsschalen. Deze vaardigheidsgrenzen zijn vervolgens omgezet in een aantal correcte opdrachten voor elk van de drie niveauversies in de uitgave en een aantal correcte observatiepunten voor de activiteiten. Door het gehanteerde meetmodel en de standaardbepaling, kan het functioneringsniveau van de kinderen met elke niveauversie bepaald worden. Ook kunnen de functioneringsniveaus die in verschillende observatieperiodes behaald zijn met elkaar vergeleken worden.
- Afstemmen onderwijsaanbod: de rapportage geeft handvaten om het onderwijs zo goed mogelijk op de kinderen af te stemmen. Er is bewust gekozen voor een rapportage met niveaus op elk van de vijf subdomeinen, in plaats van alleen een resultaat op het totaal van Rekenen. Daarmee is tijdige signalering mogelijk en is het instrument een hulpmiddel voor de leerkracht. Op welke subdomeinen gaat het boven verwachting en kan het kind extra uitdaging gebruiken? Op welke subdomeinen heeft het kind misschien extra hulp nodig? Wanneer de digitale opdrachten gemaakt zijn, is het via de online kindrapportage mogelijk de gegeven antwoorden van een kind in te zien. Zo kun je zien welk soort opdrachten goed gaan en met welk soort opdrachten het kind nog wat meer moeite heeft. En is er een patroon te zien als je de resultaten van een hele groep bekijkt? In de handleiding in de leerkrachtmap wordt nader ingegaan op de stap naar het handelen.

## 3 Beschrijving meetinstrument

### 3.1 Opbouw van het meetinstrument

Voor de opzet van Kleuter in beeld - Rekenen is gekozen voor een observatie-instrument, dat bestaat uit een leerkrachtroute en een kindroute.

- De leerkrachtroute wordt ingezet als de leerkracht al een goed beeld heeft van een kind. Dit kan voor alle vijf subdomeinen, of een deel daarvan. De route is volledig gebaseerd op indirecte observaties over een langere periode (het is geen momentopname).
- De kindroute wordt ingezet als de leerkracht twijfelt over de vaardigheid van een kind op een bepaald subdomein, of als hij meer informatie wil. Bij deze route komen dezelfde vijf subdomeinen aan bod, maar nu met een actieve rol voor het kind. In de kindroute hebben we zowel voor opdrachten als activiteiten gekozen. In Figuur 3.1 staat de opbouw van Kleuter in beeld - Rekenen visueel weergegeven en staat ook bij welke subdomeinen voor opdrachten en bij welke subdomeinen voor een activiteit is gekozen.

Zoals al eerder beschreven, zijn de niveaus gebaseerd op een absolute normering in plaats van een relatieve normering. Kinderen worden niet vergeleken met een landelijke referentiegroep in een bepaalde periode. Dit betekent dat de observaties, opdrachten en activiteiten ook niet gebonden zijn aan een bepaalde periode. De leerkracht kan deze flexibel inplannen, op het moment waarop de informatie nodig is.

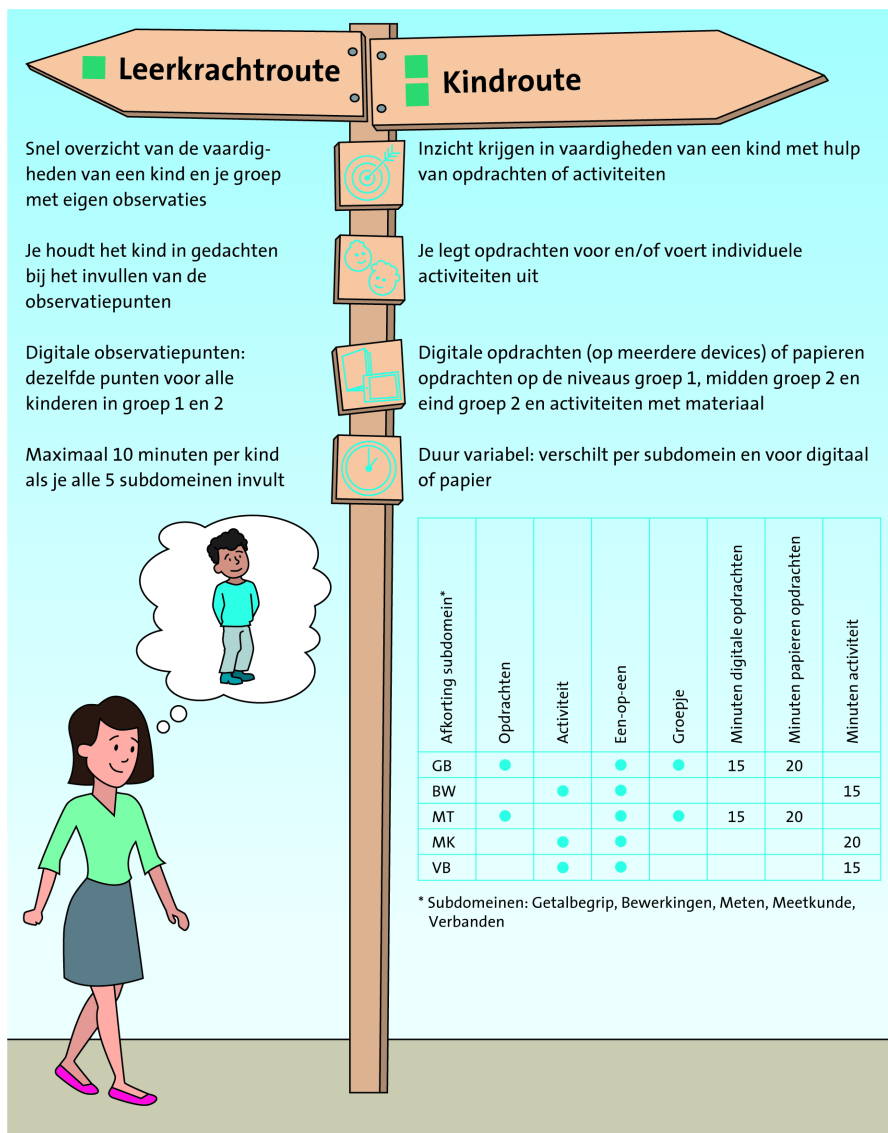
De leerkracht kan ervoor kiezen om de opdrachten op papier of digitaal voor te leggen. Bij de opdrachten zijn er voor elk van de twee subdomeinen drie verschillende niveauversies beschikbaar. De leerkracht stelt per subdomein vast welk niveau het beste bij het kind past.

- ★ = niveau groep 1
- ★★ = niveau midden groep 2
- ★★★ = niveau eind groep 2

Dit betekent dat een kind uit groep 1 bijvoorbeeld ook de opdrachten van het niveau midden groep 2 kan maken, als een kind op een bepaald subdomein enigszins voorloopt op andere kinderen.

Bij de drie activiteiten zijn er geen verschillende niveauversies. Voor de kinderen in groep 1 en 2 kunnen dezelfde activiteiten en observatieformulieren gebruikt worden. De leerkracht voert de activiteit samen met het kind uit en vult tegelijkertijd het bijbehorende observatieformulier in.

Figuur 3.1 Opbouw van Kleuter in beeld - Rekenen



### 3.2 Ontwikkel- en constructieproces in co-creatie

Dit instrument is in co-creatie met het veld tot stand gekomen. Allereerst is vastgesteld hoe het instrument er in grote lijnen uit gaat zien. In februari 2019 heeft een bijeenkomst van een focusgroep plaatsgevonden. Een groep van 15 kleuterleerkrachten, intern begeleiders en onderbouwcoördinatoren uit het hele land kwam naar Arnhem om mee te denken over de opbouw van het nieuwe instrument. In april 2019 is er een expertbijeenkomst bij Cito georganiseerd. Hierbij waren vijf experts (onder andere van SLO, Onderwijsadviseurs, Marnix Academie en MB-rekenadvies) aanwezig om de toetsdeskundigen van Cito van feedback te voorzien over de opzet van Kleuter in beeld - Rekenen. Ook is de online community die voor Kleuter in beeld - Taal was opgericht, voortgezet voor Kleuter in beeld - Rekenen: regelmatig zijn stellingen voorgelegd aan tientallen kleuterleerkrachten en intern begeleiders. Uit deze vormen van co-creatie kwamen tien duidelijke wensen vanuit het veld naar voren:

- Observatie centraal
- Kort, maar krachtig
- Flexibel in te zetten, hele jaar door
- Regie bij de leerkracht
- Inclusief concreet materiaal



- Geen normering/vergelijking met landelijke referentiegroep
- Passend bij belevingswereld kleuters
- Digitaal en op papier
- Objectief beeld
- Gebruiksvriendelijke rapportage

Bij de verdere ontwikkeling hebben deze wensen centraal gestaan en is met al deze wensen rekening gehouden. Door de verschillende bijeenkomsten en gesprekken met uiteenlopende onderwijsprofessionals kwam de opzet van de combinatie van de leerkrachtroute en de kindroute tot stand. Met de mogelijkheid van de leerkrachtroute wordt recht gedaan aan de kennis van de leerkracht, die in veel gevallen al een duidelijk beeld heeft van zijn kleuters. In de kindroute is zowel voor opdrachten als activiteiten gekozen. Dit was ook een belangrijke wens vanuit het veld. Met opdrachten kan op een zeer betrouwbare en relatief eenvoudige manier het niveau van kinderen bepaald worden. Bovendien, ter voorbereiding op groep 3, wordt er bij kleuters ook aandacht besteed aan werken op het platte vlak; hier sluiten de opdrachten goed bij aan. In de community is gevraagd of scholen een voorkeur hebben voor papieren of digitale opdrachten. Een deel van de scholen koos voor papieren opdrachten en een ander deel voor digitale opdrachten. Daarom is ervoor gekozen beide opties aan te bieden, zodat scholen hierin een keus hebben. Inhoudelijk is er geen verschil, alle drie niveauversies bevatten op papier en digitaal dezelfde opdrachten.

Voor Kleuter in beeld – Rekenen is niet enkel gekozen voor opdrachten, maar ook voor activiteiten. Dit was ook een uitdrukkelijke wens van het veld. De keuze voor activiteiten is ook vanuit de literatuur een goede keuze voor het in kaart brengen van de rekenvaardigheid bij kleuters. Kleuters construeren kennis namelijk op een experimentele, interactieve, concrete en praktische wijze (Bredekamp & Rosegrant, 2005) in plaats van door abstract redeneren, zoals oudere kinderen en volwassenen leren. Om te kunnen leren moeten jonge kinderen objecten kunnen aanraken en manipuleren, moeten zij kunnen bouwen en creëren met verschillende materialen, moeten zij kunnen luisteren naar en uitspelen van verhalen en rollen, maar ook praten, zingen, en bewegen op verschillende manieren en plekken (Guddemi & Case, 2004). Het is dan ook passend om de kennis en vaardigheden van kleuters op eenzelfde manier te meten. In plaats van enkel traditionele papier- en potloodassessments, kan waar mogelijk een activiteit met concreet materiaal ingezet worden, zodat kleuters goed kunnen laten zien wat zij kennen en kunnen (Guddemi & Case, 2004). Kinderen leren door spel. Behoeft aan spelen is hun eigen intensieve, actieve en spontane vorm van leren (Galema-Koolen, 2008). Een activiteit in de vorm van een spel past goed bij de ontwikkelingsstappen die kleuters kunnen zetten. Binnen spel, creatieve activiteiten en bouwactiviteiten doen kleuters spontane leerervaringen op, al dan niet gestimuleerd door een impuls van de leerkracht. Door gerichte rekenactiviteiten af te stemmen op wat de kleuters bezighoudt, sluiten deze aan bij de belevingswereld van de kleuters (Van Vugt & Wösten, 2015). Activiteiten kosten weliswaar wat meer tijd, maar passen goed bij deze leeftijdsgroep. Het geeft leerkrachten dan ook veel inzicht door kinderen handelend aan het werk te zien.

Per subdomein is gekeken naar wat het meest geschikt was: opdrachten of een activiteit. Bij Getallen en Meten waren opdrachten het meest passend, bij Bewerkingen, Meetkunde en Verbanden een activiteit. De betrokken leerkrachten en ib'ers hebben ook meegedacht over welke subdomeinen het beste via opdrachten en welke subdomeinen het beste via activiteiten te meten zijn. De groep vond dat Getalbegrip goed te meten is met opdrachten. Bij de andere subdomeinen had de groep een voorkeur voor activiteiten. Naar aanleiding van de ervaringen die inmiddels met Kleuter in beeld - Taal zijn opgedaan, is ervoor gekozen om nog een subdomein in kaart te brengen met opdrachten. De ervaring leert namelijk dat het voor scholen niet haalbaar is om voor alle subdomeinen activiteiten uit te voeren. Daarom is gekozen voor meer balans tussen opdrachten en activiteiten. Van de andere subdomeinen leent het subdomein Meten zich het beste voor opdrachten. Bij de subdomeinen Bewerkingen, Meetkunde en Verbanden ligt het meer voor de hand om de ontwikkeling van de kleuters te bekijken door ze te laten handelen met concrete materialen (activiteiten).

In het voorjaar van 2019 zijn de toetsdeskundigen begonnen met de constructie van de opdrachten en observatiepunten bij de activiteiten. Alle items (opdrachten en observatiepunten) die in Kleuter in beeld - Rekenen zijn opgenomen, zijn speciaal voor dit instrument ontwikkeld. Ook is in deze periode een try-out georganiseerd en hebben de toetsdeskundigen diverse scholen bezocht om de eerste prototypen uit te proberen bij kinderen. Op basis van de feedback zijn de prototypen verfijnd en aangepast. Vervolgens zijn per subdomein tientallen opdrachten geconstrueerd en de activiteiten ontwikkeld met daarbij de observatiepunten.

Leerkrachten werkzaam in het speciaal (basis)onderwijs hebben de opdrachten en activiteiten gescreend op geschiktheid voor kinderen met speciale onderwijsbehoeften. Op basis daarvan zijn er opdrachten aangepast en ook afgefallen. Zo was bij onderstaande opdracht zowel bij het eerste als het tweede kind het kleurcontrast tussen het boek en de kleding niet groot genoeg. Daarom zijn de kleuren aangepast, waardoor deze opdracht ook geschikt is voor kinderen die slechtziend zijn:



Voorbeeld Meten: 'Wie leest het kleinste boek?'

Vervolgens is in september en oktober 2019 een proefonderzoek georganiseerd en in januari en februari 2020 een grootschalig kwaliteitsonderzoek. Bij beide onderzoeken hebben toetsdeskundigen scholen bezocht om mee te kijken met de afnames en te praten met leerkrachten en kleuters over de wensen voor het nieuwe instrument. Daarnaast is bij beide onderzoeken een online evaluatieformulier gemaakt dat door tientallen scholen is ingevuld. Alle input is gebruikt om zo goed mogelijk aan te sluiten bij de wensen, behoeften en ideeën van de toekomstige gebruikers van het instrument. In hoofdstuk 4 wordt verder ingegaan op de dataverzameling van deze twee onderzoeken.

### 3.3 Operationalisatie van de inhoud

In paragraaf 2.2 over de meetpretentie is al uiteengezet hoe de selectie van de vijf subdomeinen en de 16 bijbehorende onderdelen tot stand is gekomen. Deze inhoud van subdomeinen en onderdelen vormt het uitgangspunt voor zowel de leerkrachtroute als de kindroute. In deze paragraaf 3.3 wordt verder ingegaan op de operationalisatie van de inhoud: op welke manier kunnen de vaardigheden op de verschillende subdomeinen gemeten worden? De keuzes hiervoor zijn enerzijds gebaseerd op de literatuur zoals genoemd in paragraaf 2.2. Anderzijds zijn de keuzes gebaseerd op de wensen vanuit het veld, zoals deze in de verschillende vormen van co-creatie naar voren zijn gekomen. Per subdomein wordt ook de toetsmatrijs bij de uitgave van de niveauversies van de opdrachten en van de observatiepunten van de activiteiten weergegeven, gebaseerd op de gewenste verdelingen van onderdelen, categorieën en bepaalde kenmerken.

#### **Leerkrachtroute**

De leerkrachtroute is dus gebaseerd op dezelfde vijf subdomeinen met bijbehorende 16 onderdelen als bij de kindroute. Het overzicht is terug te vinden in Tabel 2.1. Bij elk onderdeel hoort een observatiepunt; de leerkrachtroute bestaat dan ook uit 16 observatiepunten. Er worden zes niveaus van vaardigheid onderscheiden: <M1, M1, E1, M2, E2 en >E2. Bij elk observatiepunt is bij de niveaus E1 en E2 een beschrijving opgenomen. Die beschrijving geeft aan wat de meeste kinderen eind groep 1 en eind groep 2 beheersen. De keuze voor de beschrijvingen zijn gebaseerd op de genoemde literatuur in paragraaf 2.2. Belangrijk uitgangspunt was dat de observatiepunten te observeren zouden zijn in de dagelijkse klassensituatie én alle kleuterleerkrachten een observatiepunt op dezelfde manier zouden interpreteren.

Daarom is een concept-observatielijst voorgelegd aan een jonge kind rekenexpert van SLO. Dat heeft geholpen bij het aanscherpen van formuleringen. Daarnaast is ervoor gekozen om waar nodig een toelichting te geven op een observatiepunt, bijvoorbeeld in de vorm van een voorbeeld. Dit leidde tot de versie die we in het kwaliteitsonderzoek hebben meegenomen. Leerkrachten die aan dat onderzoek hebben meegedaan, konden feedback geven op de observatiepunten. Na het kwaliteitsonderzoek is de inhoud van de observatiepunten gevalideerd aan psychometrische gegevens (zie hiervoor hoofdstuk 9 Validiteit). Op basis daarvan zijn de definitieve observatiepunten vastgesteld. Zie Figuur 3.2 voor een voorbeeld van de onderdelen Telrij en Hoeveelheden bij Getalbegrip.

Figuur 3.2 Voorbeeld deel leerkrachtroute

### Getalbegrip

**Telrij (tot ten minste 20)**

<input type="radio"/> <M1	<input checked="" type="radio"/> M1	<input type="radio"/> E1	<input checked="" type="radio"/> M2	<input type="radio"/> E2	<input checked="" type="radio"/> >E2
---------------------------	-------------------------------------	--------------------------	-------------------------------------	--------------------------	--------------------------------------

Kan verder tellen vanaf elk getal onder 10.  
Voorbeeld: 'Ik tel hardop en als ik stop, tel jij verder:  
5, 6, 7, ...'

Kan rangtelwoorden tot en met 5 begrijpen en gebruiken.  
Voorbeeld: Zet 5 stoelen op een rij. Ga zelf op de eerste stoel zitten en vraag het kind om op de derde stoel te gaan zitten.

**Hoeveelheden (tot ten minste 20)**

<input type="radio"/> <M1	<input checked="" type="radio"/> M1	<input type="radio"/> E1	<input checked="" type="radio"/> M2	<input type="radio"/> E2	<input checked="" type="radio"/> >E2
---------------------------	-------------------------------------	--------------------------	-------------------------------------	--------------------------	--------------------------------------

Kan hoeveelheden vergelijken en ordenen op *meeste* en *minste*.  
Voorbeeld: Leg drie groepjes blokjes neer (een groepje met 4, 5 en 3 blokjes).  
'Waar liggen de meeste blokjes?'

Kan hoeveelheden tellen, vergelijken en ordenen en hierover redeneren ook aan de hand van begrippen als *evenveel* en *eerlijk verdelen*.  
Voorbeeld: Leg 3 groepjes (van bijv. 11, 12 en 13) knopen geordend op tafel. Hou in je hand 11 knopen.  
'In mijn hand heb ik 11 knopen.  
Waar op tafel liggen ook 11 knopen?'

## Opdrachten kindroute

- **Getalbegrip**

Op grond van de literatuur en aanbevelingen uit het veld, zijn de volgende uitgangspunten geformuleerd:

- verschillende vormen hotspotopdrachten: drie alternatieven in één afbeelding (zie voorbeeld 1), drie alternatieven in losse plaatjes (zie voorbeeld 2), hotspot met alternatieven (getallen) rechts onder elkaar (zie voorbeeld 3) en hotspot waarbij twee antwoorden gekozen moeten worden (zie voorbeeld 4);
- sleepopdrachten alleen voor ★★ en ★★★ (zie voorbeeld 5);
- uitvoerbaarheid opdrachten papier en digitaal;
- waar mogelijk passend voor leerlingen in het s(b)o.

Getalbegrip is iets wat jonge kinderen geleidelijk verwerven. Ze leren kleine hoeveelheden tellen, herkennen en schatten. In de dagelijkse werkelijkheid komen ze al in aanraking met de functies van getallen. Ze leren aantallen ordenen en vergelijken. Daarna leren kinderen eenvoudige erbij- en eraf-handelingen uit te voeren, eerst met concreet materiaal en later meer abstract rekenen met representaties (plaatjes of zelfs alleen getallen) (Koerhuis & Keuning, 2011). Getalbegrip bevat de onderdelen 'Telrij', 'Hoeveelheden', 'Getallen' en 'Relaties tussen telrij, hoeveelheden en getalen'. Binnen Getalbegrip zijn ook opdrachten opgenomen, waarbij kinderen verhoudingsgewijs vergelijken en ordenen.

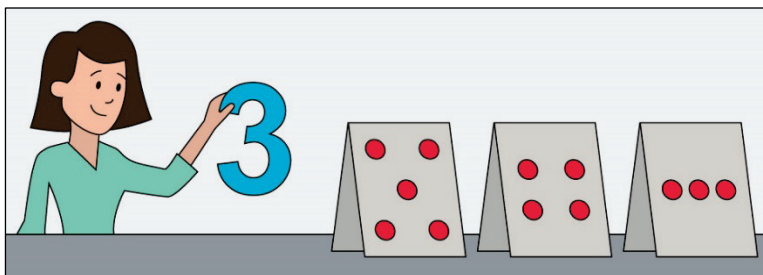
Telrij: het kind kan doortellen en teruggtellen vanaf willekeurige getallen en omgaan met rangtelwoorden. Verder kan het kind omgaan met begrippen rondom de telrij en tellen met sprongen van 2.

Hoeveelheden: het kind kan hoeveelheden tellen en omgaan met begrippen rond hoeveelheden zoals weinig/veel, minste/meeste en evenveel. Daarnaast kan het kind hoeveelheden vergelijken en ordenen en redeneren over hoeveelheden.

Getallen: het kind kan cijfers en getallen herkennen en omgaan met begrippen rond getallen zoals groter, kleiner, grootste en kleinste en oefenen met de vaste volgorde van getallen in de getallenrij. Ook kan het kind getallen in de getallenrij vergelijken en ordenen.

Relaties tussen telrij, hoeveelheden en getallen: het kind kan telwoorden, hoeveelheden en getallen koppelen en flexibel omgaan met de relatie tussen telrij, hoeveelheden en getallen. Ook kan het kind redeneren over hoeveelheden en hele getallen.

Bij een opdracht Getalbegrip hoort een kind een vraag, in sommige gevallen voorafgegaan door een zin met informatie. Het kind luistert, kijkt naar de afbeelding en kiest het juiste antwoord. Dit kan zowel een getal als een plaatje zijn.



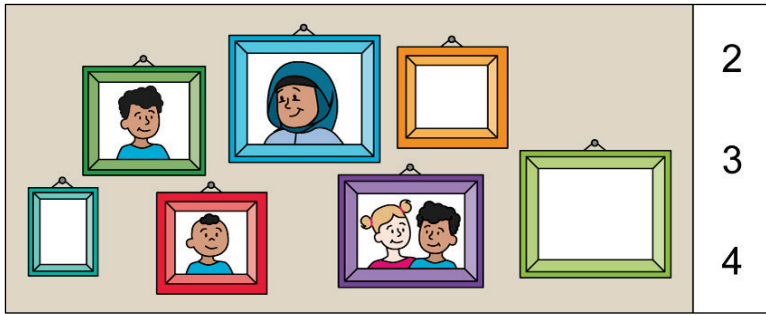
Voorbeeld 1) Hotspot met één grote afbeelding

Juf Ilse houdt een getal vast. Welke stippenkaart hoort daarbij?



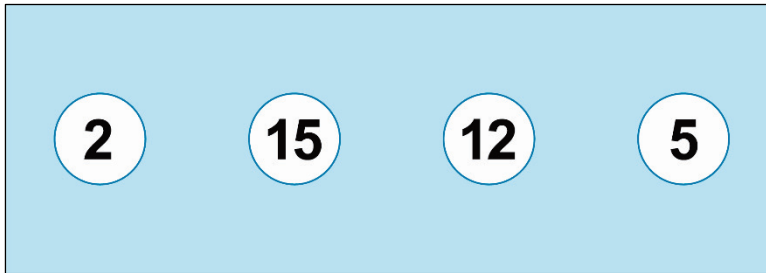
Voorbeeld 2) Hotspot met drie losse plaatjes

Op welk plaatje zie je 6 katten?



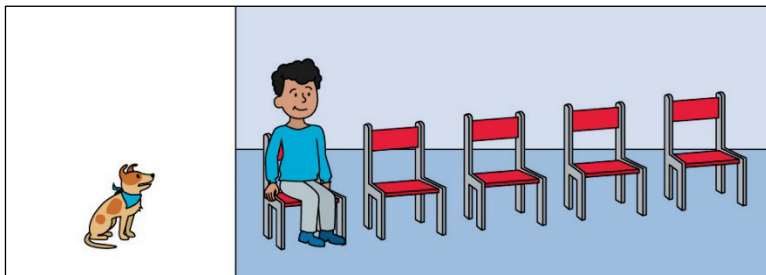
Voorbeeld 3) Hotspot met getallen rechts

'Hoeveel fotolijstjes zijn leeg?'



Voorbeeld 4) Hotspot met twee antwoorden

'Welke twee getallen zijn kleiner dan 11?'



Voorbeeld 5) Sleepopdracht

'Mo zit op de eerste stoel. Zet Bink op de derde stoel.'

In Tabel 3.1 staat de toetsmatrix voor de drie niveauversies van de uitgave.

Tabel 3.1 Toetsmatrix Uitgave Getalbegrip

Onderdeel	★	★★	★★★
Telrij	4	5	4
Hoeveelheden	5	3	4
Getallen	4	7	5
Relaties tussen telrij, hoeveelheden en getallen	5	3	5
<b>Totaal</b>	18	18	18

- **Meten**

Op grond van de literatuur en aanbevelingen uit het veld, zijn de volgende uitgangspunten geformuleerd:

- verschillende vormen hotspotopdrachten: drie alternatieven in één grote afbeelding (zie voorbeeld 6), drie alternatieven in losse plaatjes (zie voorbeeld 7), hotspot met alternatieven (getallen) rechts onder elkaar (zie voorbeeld 8), hotspot waarbij twee antwoorden gekozen moeten worden (zie voorbeeld 9) en hotspotopdrachten met volgorde in tijd (zie voorbeeld 10);
- uitvoerbaarheid opdrachten papier en digitaal;
- waar mogelijk passend voor leerlingen in het s(b)o.

Metten heeft betrekking op lengte en omtrek, oppervlakte, gewicht, inhoud, tijd en geld. Bij het meten worden eigenschappen vergeleken of afgestemd. Naast informele meetinstrumenten (schoen) worden ook formele meetinstrumenten gebruikt (weegschaal). Ook ordenen van lengte, gewicht, inhoud en tijd (wat doe je eerst, wat daarna) hoort bij Metten (Koerhuis & Keuning, 2011). Binnen Metten zijn ook opdrachten opgenomen, waarbij kinderen verhoudingsgewijs vergelijken en ordenen.

Lengte en omtrek: het kind kan vergelijken en ordenen op lengte en omtrek en redeneren over lengte.

Oppervlakte: het kind kan omgaan met begrippen rond oppervlakte en vergelijken naar oppervlakte.

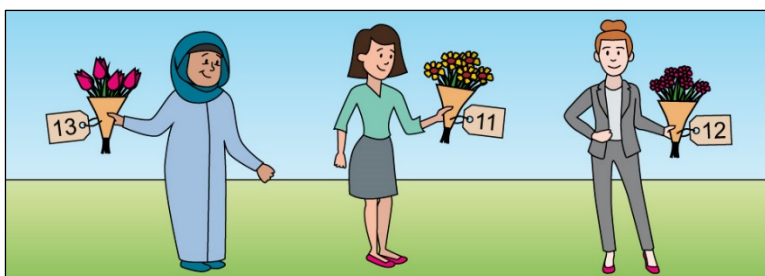
Inhoud: het kind kan omgaan met begrippen rond inhoud, vergelijken en ordenen op inhoud en redeneren over inhouden in eenvoudige probleem- en conflictsituaties.

Gewicht: het kind kan omgaan met begrippen rond gewicht en vergelijken en ordenen op gewicht. Daarnaast begrijpt het kind hoe informele 'weeginstrumenten' werken en kan het formele weeginstrumenten aflezen. Ook kan het kind redeneren over wegen en gewichten in eenvoudige probleem- en conflictsituaties.

Tijd: het kind kan omgaan met cyclische tijdsaanduidingen, gebeurtenissen plaatsen in tijdsvolgorde (leeftijd bepalen) en redeneren over tijd in eenvoudige probleem- en conflictsituaties.

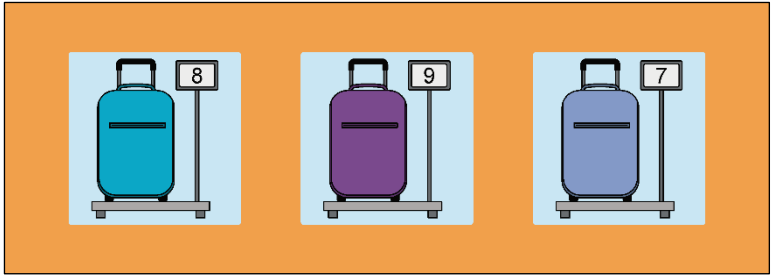
Geld: het kind kan omgaan met begrippen rond geld en begrijpt de rol van geld bij kopen, verkopen en betalen. Daarnaast kan het kind geldwaarden bepalen en eenvoudige geldbedragen samenstellen in hele euro's.

Bij een opdracht Metten hoort een kind een vraag, in sommige gevallen voorafgegaan door een zin met informatie. Het kind luistert, kijkt naar de afbeelding en kiest het juiste antwoord. Dit kan zowel een getal als een plaatje zijn.

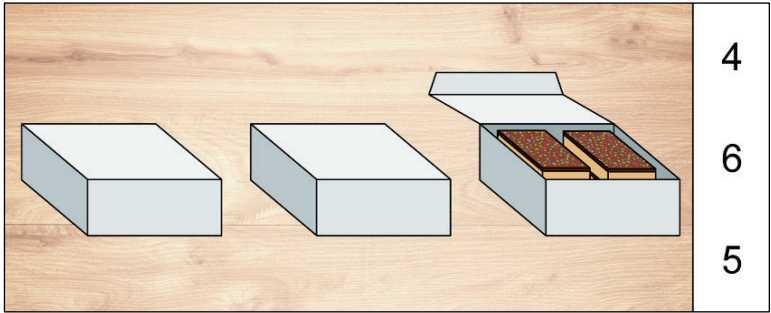


Voorbeeld 6) Hotspot met één grote afbeelding

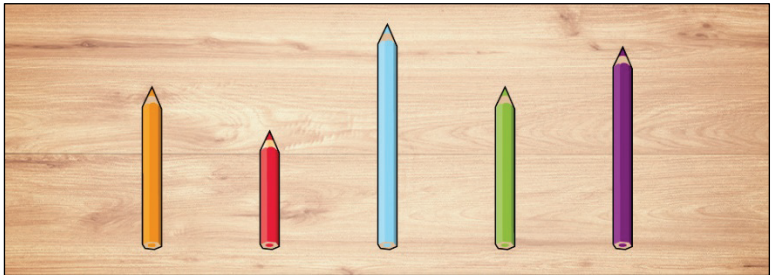
'Wie betaalt het minst voor de bos bloemen?'



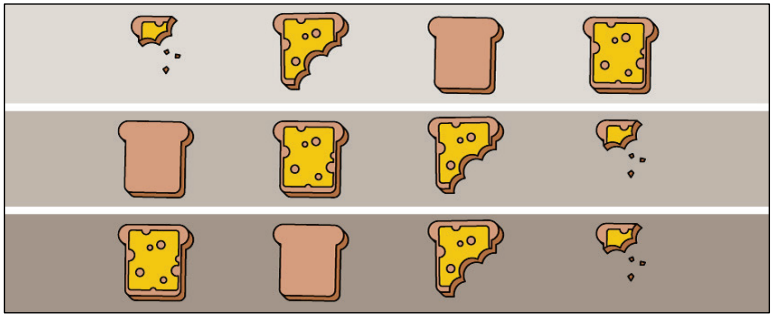
Voorbeeld 7) Hotspot met drie losse plaatjes  
 'Welke koffer weegt het meest?'



Voorbeeld 8) Hotspot met getallen rechts  
 'Op tafel staan 3 volle dozen met gebakjes. Hoeveel gebakjes zijn dat samen?'



Voorbeeld 9) Hotspot met twee antwoorden  
 'Welke twee potloden zijn even lang?'



Voorbeeld 10) Hotspot volgorde tijd  
 'Anne eet een boterham. In welke rij staan de plaatjes goed?'

In Tabel 3.2 staat de toetsmatrijs voor de drie niveaueversies van de uitgave.

Tabel 3.2 Toetsmatrijs Uitgave Meten

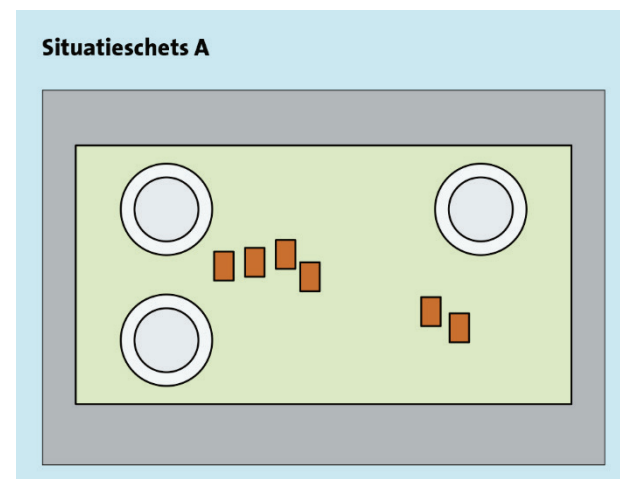
Onderdeel	★	★★	★★★
Lengte en omtrek	3	3	5
Oppervlakte	1	1	2
Inhoud	5	5	3
Gewicht	4	3	3
Tijd	1	4	3
Geld	4	2	2
Totaal	18	18	18

### Activiteiten kindroute

Elke activiteit bestaat uit een instructie en observatieformulier voor de leerkracht. Op het observatieformulier staan de concrete observatiepunten en de antwoordopties. De antwoordopties variëren: nee-ja (dichotome observatiepunten) en nee-met hulp-ja (polytome observatiepunten). In de leerkrachtinstructie en handleiding wordt duidelijk aangegeven wanneer je voor welke optie kiest. De leerkracht vult zoveel mogelijk tijdens de activiteit het bijbehorende observatieformulier in. De leerkracht beoordeelt in welke mate het kind het beschreven gedrag in het observatiepunt laat zien.

### Bewerkingen

In de groepen 1 en 2 doen kinderen ervaring op met het allereerste begin van het rekenen. In eerste instantie zijn kinderen nog bezig met puur tellen bij het 'uitrekenen' van een bewerking. Geleidelijk aan wordt de manier van tellen steeds handiger en gaat het over in rekenen. Dat zie je terug in hoe kinderen een opdracht aanpakken (Treffers et al., 2009; Veltman & Van den Heuvel-Panhuizen, 2010; Oonk et al., 2019). Bijvoorbeeld bij BW01: *Er liggen 4 broodjes in het mandje. (De leerkracht wijst het mandje aan.) Leg de broodjes maar op tafel. Dan kunnen we ze goed zien. Oma pakt er nog 2 broodjes bij. De leerkracht legt nog 2 broodjes op tafel (zie Situatieschets A). Hoeveel broodjes zijn dit samen?*



Een deel van de kinderen telt alle broodjes (al dan niet synchroon) om tot een antwoord te komen. Sommige kinderen zijn al in staat om vanuit het groepje van 4 broodjes meteen verder te tellen: 4 ... 5, 6 (een manier van verkort tellen). Er zijn ook kinderen die de broodjes per 2 tellen: 2, 4, 6 (een andere vorm van verkort tellen). Enkele kinderen hoeven misschien niet te tellen, maar weten dit zelfs al uit het hoofd (weten: 4 erbij 2 is 6). In de activiteit Bewerkingen zijn de opdrachten zo vormgegeven dat het kind op dit soort verschillende manieren tot 'de oplossing' kan komen. Dat gebeurt in een speelse context: de leerkracht speelt een doen-alsof-spel met het kind dat aansluit bij de belevingswereld van kleuters, namelijk



eten bij opa en oma. Tijdens het spel krijgt het kind allerlei opdrachten die gaan over optellen, aftrekken, splitsen en (ver)delen van broodjes en druiven.

Bij de deelactiviteiten beslist de leerkracht zelf of hij deze direct achter elkaar uitvoert of dat hij dat in delen doet (eerst optellen en aftrekken en op een ander moment splitsen en (ver)delen). Bij alle onderdelen mag het kind concreet materiaal gebruiken (onder andere blokjes die broodjes voorstellen en kralen die druiven voorstellen). Zo kan het kind de opdrachten die de leerkracht geeft al handelend met het materiaal uitvoeren. De instructie en de opdrachten die de leerkracht geeft, zijn letterlijk weergegeven in de activiteit, vanwege eenduidigheid en objectiviteit van de observatiepunten. De volgorde van observatiepunten is inhoudelijk opgebouwd. In het eerste deel van de activiteit krijgt het kind opdrachten over optellen en daarna over aftrekken. Het tweede deel van de activiteit begint met opdrachten over splitsen en eindigt met opdrachten over (ver)delen. De opdrachten zijn ook opgebouwd in moeilijkheid. Bij de opdrachten over optellen bijvoorbeeld, krijgt het kind eerst opdrachten met maximaal 6 broodjes, daarna opdrachten met maximaal 10 broodjes en ten slotte opdrachten met maximaal 15 druiven. Daarnaast ligt het materiaal de ene keer goed zichtbaar op tafel en de andere keer minder goed zichtbaar in een mandje. Naar aanleiding van het kwaliteitsonderzoek hebben we de activiteit *Bewerkingen* nog verder aangescherpt. Zo zijn onder andere duidelijke situatieschetsen opgenomen in de instructie, omdat leerkrachten het onduidelijk vonden waar zij het materiaal (de blokjes en de kralen) neer moesten leggen: op tafel of in het mandje. Ook is de context speelser gemaakt, waardoor het nog meer een 'doen-alsof-spel' is geworden. Daarnaast hebben we voor de leerkrachten nog een extra aanwijzing gegeven over wanneer je 'met hulp' invult. Bij het uitvoeren van de opdrachten maakten de kinderen soms telfouten. Voor de leerkrachten was het niet altijd duidelijk of zij 'met hulp' moesten invullen als zij het kind vroegen om opnieuw te tellen. Het maken van telfouten is onderdeel van de opeenvolgende fasen in het leren tellen. Daarom hebben we in de instructie opgenomen dat het kind opnieuw laten tellen als 'met hulp' moet worden gescoord. Het kind maakt dan namelijk nog fouten in het synchroon tellen.

In Tabel 3.3 staat de toetsmatrix voor de observatiepunten van deze activiteit.

Tabel 3.3 Toetsmatrix *Uitgave Bewerkingen*

Onderdeel	N observatiepunten
Optellen	5
Aftrekken	4
Splitsen	4
(Ver)delen	4
Totaal	17

### **Meetskunde**

Het meetkundeonderwijs op de basisschool richt zich op het ontwikkelen van het ruimtelijk voorstellings- en redeneervermogen van kinderen. In de groepen 1 en 2 doen kinderen ervaring op met het leren begrijpen van de ruimte, zowel de fysieke ruimte (3D) als het platte vlak (2D) (Van den Heuvel-Panhuizen, Veltman, Janssen & Hochstenbach, 2009). Kinderen verhouden zich tot de ruimte om hen heen (oriënteren: lokaliseren en het innemen van een standpunt). De oriëntatie in de ruimte ontwikkelt zich vanuit het eigen lichaam. Het lichaam is het referentiepunt van waaruit relaties worden gelegd (Ik sta voor de kast.). Langzamerhand neemt het kind afstand van zichzelf en leert dat er ook andere referentiepunten bestaan (De kast staat naast de tafel.). Daarnaast is er een ontwikkeling van concrete ruimtes naar verkleinde ruimtes (poppenhuis) naar abstracte afbeeldingen (plattegrond). Kinderen leren de eigenschappen van verschillende vormen kennen en construeren ruimtelijke en vlakke objecten, bijvoorbeeld bouwen met blokken, dozen of wc-rolletjes en vouwen met vouwblaadjes. Ze opereren ook met vormen en figuren,

bijvoorbeeld het samenstellen van patronen op een kralenplank, het rijgen van een ketting of het spiegelen van een bouwwerk in de bouwhoek (Koerhuis & Keuning, 2011).

De activiteit Meetkunde bestaat uit drie deelactiviteiten waarin de verschillende onderdelen van het subdomein Meetkunde terugkomen: 'Waar zijn Anne en Mo?' (oriënteren in de ruimte), 'Patronen en Vormen' (opereren met vormen en figuren) en 'Bouwen en Vouwen' (construeren). De deelactiviteiten bestaan uit verschillende speelse korte opdrachten met concreet materiaal die aansluiten bij de belevingswereld van kleuters, zoals: het nabouwen van een legohuisje, het navouwen van een hondje, het voortzetten van patronen en het bepalen van het standpunt van de fotograaf. Tijdens de deelactiviteiten krijgt het kind allerlei opdrachten die gaan over 'oriënteren in de ruimte', 'opereren met vormen en figuren' en 'construeren'. De leerkracht beslist zelf of hij deze deelactiviteiten direct achter elkaar uitvoert of dat hij dat in delen doet. Bij alle onderdelen gebruikt het kind concreet materiaal (onder andere de poppetjes Anne en Mo, fiches, lego en vouwblaadjes) om de opdrachten die de leerkracht geeft uit te voeren. Het kind voert de opdrachten die de leerkracht geeft handelend met het materiaal uit. De instructie en de opdrachten die de leerkracht geeft, zijn letterlijk weergegeven in de activiteit, vanwege eenduidigheid en objectiviteit van de observatiepunten. De inhoud van de opdrachten is een uitwerking van de aanbodsdoelen uit de inhoudskaart van SLO, waarbij ook gekeken is naar hoe het onderdeel 'Meetkunde' in kleutermethoden aan bod komt en beschreven staat in de literatuur. De volgorde van observatiepunten is inhoudelijk en qua moeilijkheid opgebouwd. Elke deelactiviteit begint met opdrachten die meer gericht zijn op groep 1, gevolgd door opdrachten die meer passen bij groep 2. Bij de deelactiviteit 'Bouwen en Vouwen' bijvoorbeeld, krijgt het kind eerst de opdracht om een legohuisje na te bouwen. Verderop in deze deelactiviteit vouwt het kind een hondje door de handelingen van de leerkracht na te doen. En bij de deelactiviteit 'Patronen en Vormen' wijst het kind eerst alle gele vierkanten aan, waarbij de vierkanten alleen verschillen in grootte en kleur. Vervolgens wijst het kind alle blauwe driehoeken aan, waarbij de driehoeken verschillen in grootte, kleur én tint van de kleur (donkerblauw en lichtblauw). Verderop in deze deelactiviteit legt het kind eerst een patroon met legosteentjes die alleen verschillen in kleur en vervolgens een patroon met legosteentjes die daarnaast ook verschillen in grootte (vorm). Naar aanleiding van het kwaliteitsonderzoek hebben we de activiteit Meetkunde nog verder aangescherpt. Zo is onder andere een aantal opdrachten verwijderd, bijvoorbeeld omdat de opdrachten te makkelijk waren (een aantal opdrachten over ruimtelijke begrippen; toren nabouwen van een 2D voorbeeld) of niet eenduidig te interpreteren waren (een aantal opdrachten over ruimtelijke begrippen) of lastiger praktisch uitvoerbaar waren vanwege de benodigde materialen en voorbereiding (opdrachten rondom een plattegrond).

In Tabel 3.4 staat de toetsmatrijs voor de observatiepunten van deze activiteit.

Tabel 3.4 Toetsmatrijs Uitgave Meetkunde

Onderdeel	N observatiepunten
Oriënteren in de ruimte	10
Construeren	6
Opereren met vormen en figuren	5
Totaal	21

### Verbanden

Bij verbanden gaat het over relaties tussen gegevens, bijvoorbeeld tussen buitentemperatuur en ijsverkoop. Het gaat ook over schematische weergaven van allerlei soorten informatie, bijvoorbeeld een lesrooster of een weersvoorspelling voor de komende dagen. Belangrijke vaardigheden bij verbanden zijn het kunnen aflezen en interpreteren, en selecteren en waarderen van informatie (Hutten, Van den Bergh, Van den Brom-Snijders & Van Zanten, 2014). In groep 1 en 2 doen kleuters al ervaring op met eenvoudige weergaven van grafieken en diagrammen. In een eenvoudige staafdiagram kan bijvoorbeeld weergegeven

worden hoe vaak een bepaalde (natuurlijke) maat wordt afgestemd. Het geeft inzicht in hoe op een bepaalde manier snel hoeveelheden in beeld gebracht kunnen worden (*Hoeveel olifanten zijn er in de dierentuin*). Door op een speelse manier kleuters te laten werken met een beelddiagram krijgen ze inzicht in ordenen van hoeveelheden en informatie. Voor kleuters is het werken met een beelddiagram en een staafdiagram een speelse manier om abstracte begrippen te koppelen aan betekenisvolle situaties. Begrippen als meer en minder (*Van welk dier is er 1 minder dan de ijsberen*) en meeste en minste (*Van welk dier zijn er de meeste*) krijgen op deze manier betekenis (Borghouts & Veltman, 2012/2013).

In de activiteit *Verbanden* speelt de leerkracht samen met het kind een dierspel. Met behulp van kaartjes van dieren in de dierentuin en blanco kaartjes gaat het kind zelf een beelddiagram van de dieren en een staafdiagram van het voer van de dieren construeren. Tijdens het spel krijgt het kind opdrachten die gaan over het construeren van een beelddiagram, het aflezen van informatie uit een grafische voorstelling (beelddiagram, staafdiagram) en een beelddiagram of een staafdiagram gebruiken om hoeveelheden en informatie te ordenen en te vergelijken. De instructie en de opdrachten die de leerkracht geeft, zijn letterlijk weergegeven in de activiteit, vanwege eenduidigheid en objectiviteit van de observatiepunten. De volgorde van observatiepunten is inhoudelijk opgebouwd. Eerst construeert het kind met concreet materiaal een beelddiagram, daarna krijgt het kind opdrachten om informatie af te lezen uit het diagram en vervolgens om het diagram te gebruiken om hoeveelheden en informatie te vergelijken en te ordenen.

Naast inzicht in een beeld- en staafdiagram heeft het kind ook begrip van getallen en de telrij nodig. In de activiteit *Verbanden* koppelt het kind de betekenis van een cijfer aan een hoeveelheid door af te lezen of te tellen. Zo beantwoordt het kind bij het beelddiagram bijvoorbeeld de vraag 'Hoeveel olifanten zijn er in de dierentuin?' door af te lezen of te tellen hoeveel olifanten er zijn. De olifanten zijn hier gerepresenteerd door plaatjes van olifanten. Op een hoger niveau hebben we een meer abstractere uitwerking in de vorm van een staafdiagram, waarin het kind afleest hoeveel voer een olifant eet. In het staafdiagram wordt het voer weergegeven door blanco kaartjes.

In Tabel 3.5 staat de toetsmatrix voor de observatiepunten van deze activiteit.

Tabel 3.5 Toetsmatrix Uitgave Verbanden

Onderdeel	N observatiepunten
Construeren beelddiagram	1
Aflezen informatie uit grafische voorstelling	4
Gebruiken van beelddiagram of staafdiagram om hoeveelheden en informatie te ordenen en te vergelijken	12
Totaal	17



## 4 Dataverzameling

### 4.1 Proefonderzoek

In september 2019 is een proefonderzoek georganiseerd. Een belangrijk doel van het onderzoek was om de kwaliteit en de moeilijkheidsgraad van de opdrachten en activiteiten te bepalen. Een ander doel was om op basis van de eerste ervaringen en feedback van leerkrachten aanpassingen te doen, zodat het uiteindelijke instrument zo goed mogelijk aansluit bij de wensen uit het veld. Zo twijfelden we voorafgaand aan het proefonderzoek welke vorm voor Getalbegrip het meest passend is. Daarom zijn in dit onderzoek zowel opdrachten Getalbegrip afgenomen als een activiteit Getalbegrip uitgevoerd.

In het steekproefkader voor het proefonderzoek zaten ruim 6000 scholen. Al deze scholen hebben per post een uitnodiging ontvangen om mee te doen met het onderzoek. 42 scholen meldden zich aan voor het onderzoek. Uiteindelijk hebben 38 scholen en 792 kinderen uit groep 1 en 2 daadwerkelijk meegedaan.

Alle kinderen maakten een papieren boekje met opdrachten Getalbegrip en een papieren boekje met opdrachten Meten. Daarnaast voerden kleuterleerkrachten activiteiten uit met de kleuters. Bij de ene helft van de kleuters werd de activiteit Getalbegrip afgenomen, bij de andere helft de activiteit Meetkunde. Om zoveel mogelijk verschillende opdrachten uit te proberen, zijn er verschillende boekjes gemaakt en hebben niet alle kinderen dezelfde opdrachten gemaakt. In totaal zijn 60 verschillende opdrachten voor Getalbegrip voorgelegd en 60 opdrachten voor Meten.

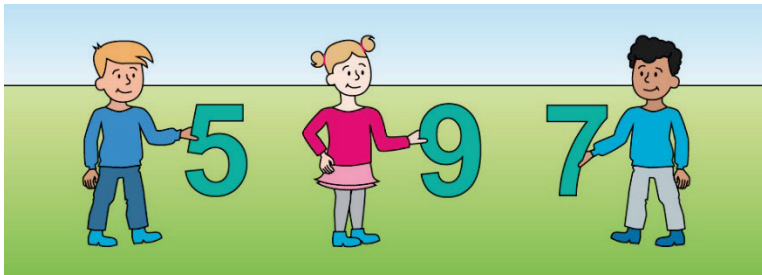
Nadat de scholen klaar waren met het onderzoek, hebben ze alle materialen teruggestuurd naar Cito. Daar zijn alle opdrachtenboekjes handmatig nagekeken. De gegeven antwoorden zijn ingevoerd in een verwerkingssysteem. Dit systeem bevat alle sleutels van de opdrachten, zodat alle opdrachten gescoord konden worden (goed/fout en welk alternatief is gekozen). De observatieformulieren van de activiteiten waren voorafgaand aan het onderzoek optisch leesbaar gemaakt. Dit betekent dat na het onderzoek de ingevulde observatieformulieren ingelezen konden worden. Op deze manier kwamen alle observaties automatisch in het systeem en was handmatige verwerking van de observatieformulieren niet nodig. Verder hebben de deelnemers na afloop een online evaluatieformulier ingevuld, waarbij ze feedback konden geven op onder meer de inhoud en vorm van de opdrachten.

Op basis van de resultaten van het evaluatieformulier, zijn een aantal aanpassingen doorgevoerd en beslissingen genomen, zoals:

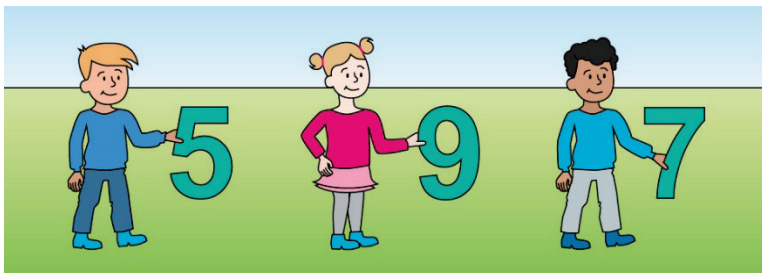
- Er is besloten om voor Getalbegrip te kiezen voor opdrachten en niet voor een activiteit. Beide vormen bleken geschikt, maar het veld gaf aan een voorkeur te hebben voor opdrachten, omdat de activiteiten individueel afgenomen moeten en bij teveel activiteiten in een instrument is dit lastig te organiseren in de klas.
- De opdrachten waarbij kinderen drie lijnen moeten trekken, bleken ongeschikt om op papier te doen. Deze opdrachten zijn daarom niet meegenomen in het kwaliteitsonderzoek en in de uitgave.
- De individuele activiteiten mogen niet te lang duren. Het speelse element is positief, maar niet als het daardoor meer tijd kost en gekunsteld overkomt. Deze feedback namen we mee bij de verdere ontwikkeling van de activiteiten.

Na het onderzoek zijn toets- en itemanalyses uitgevoerd. Op basis van psychometrische gegevens (p- en rit-waarden) en feedback van de deelnemers zijn de beste opdrachten geselecteerd voor het kwaliteitsonderzoek. Bij sommige opdrachten is de afbeelding aangepast naar aanleiding van de psychometrische gegevens en de evaluatie van de leerkrachten. Zo is bij onderstaand voorbeeld het getal dat Mo vasthoudt aan de andere kant naast hem geplaatst, zodat duidelijker te zien is welk getal bij welk kind hoort.

Voorbeeld Getalbegrip | Getallen  
Wie houdt het grootste getal vast?



Oorspronkelijke afbeelding

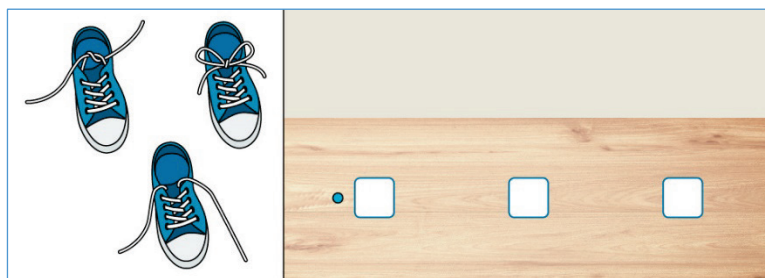


Aangepaste afbeelding

De oude versie van de opdracht had een p-waarde van ,26 in groep 2 (proefonderzoek september), terwijl in de nieuwe versie (kwaliteitsonderzoek februari) de p-waarde in groep 2 ,61 was. De nieuwe versie van deze opdracht is opgenomen in de uitgave.

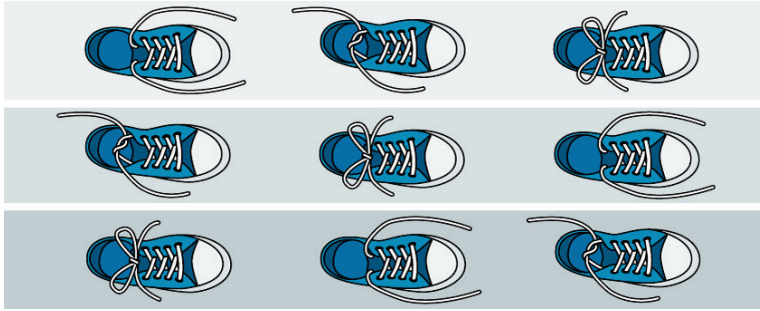
Ook bleek bij het proefonderzoek dat sommige opdrachttypen erg lastig waren voor de kinderen, zoals bij het trekken van meerdere lijnen in één opdracht. De motorische vaardigheid speelde een grote rol bij het maken van deze opdrachten, waardoor de rekenvaardigheid niet zuiver meer kon worden gemeten. Daarom is ervoor gekozen deze opdrachten aan te passen en opnieuw in te zetten bij het kwaliteitsonderzoek. Dat is bijvoorbeeld gebeurd bij onderstaande opdracht.

Voorbeeld Meten | Tijd



Mo wil zijn veters strikken. Zet zijn schoenen in de goede volgorde. Begin bij het hokje met de stip.

Oorspronkelijke opdracht:



Aangepaste opdracht:

Mo strikt zijn veters. In welke rij staan de plaatjes goed?

Ook waren er een paar opdrachten die te moeilijk bleken voor groep 1, maar verder wel geschikt waren; deze hebben we bij het kwaliteitsonderzoek verplaatst naar groep 2. Ook het omgekeerde kwam voor. Zowel voor Getalbegrip als voor Meten bleven er onvoldoende opdrachten over na het proefonderzoek. Er zijn voor Getalbegrip en Meten respectievelijk 15 en 10 opdrachten in meer of mindere mate aangepast (bijvoorbeeld aanpassingen aan de kleuren die in een afbeelding zijn gebruikt). Daarnaast zijn er voor het kwaliteitsonderzoek twee nieuwe opdrachten geconstrueerd voor Getalbegrip en zeven voor Meten.

## 4.2 Steekproef kwaliteitsonderzoek

In februari 2020 is een grootschalig kwaliteitsonderzoek georganiseerd. In dit onderzoek zijn de leerkrachtroute, de activiteiten uit de kindroute en papieren en digitale opdrachten afgenomen om de kwaliteit van het instrument te onderzoeken. In het steekproefkader voor het kwaliteitsonderzoek zaten ruim 6000 scholen. In eerste instantie is een aselechte steekproef van 910 scholen getrokken voor de afname (uitgaande van een response rate van 0,04%) in groep 1 en groep 2. Hierbij is rekening gehouden met het percentage gewichtenleerlingen en regio. Omdat het aantal scholen dat zich aanmeldde te gering was, zijn vervolgens alle scholen aangeschreven. Om de deelnamebereidheid te vergroten is besloten een attentie (VVV-bon van 25 euro) te geven aan deelnemende scholen.

Uiteindelijk hebben 49 scholen meegedaan. Scholen konden zelf kiezen of ze met de papieren of digitale afname van de opdrachten mee wilden doen. Ze hadden geen duidelijke voorkeur voor de papieren of digitale opdrachten (zie Tabel 4.1). Vier scholen deden met zowel de papieren als de digitale versie mee. Twee van deze scholen namen de papieren versie af in andere groepen dan de digitale versie. De helft van de scholen (24) deed met twee leerjaren mee (zie Tabel 4.2).

Tabel 4.1 Aantal leerlingen dat heeft deelgenomen per leerjaar en wijze van afnemen (papier of digitaal)

	Papier	Digitaal	Totaal
<b>Groep 1</b>	181	105	286
<b>Groep 2</b>	352	399	751
<b>Totaal</b>	533	504	1037

Tabel 4.2 Aantal scholen dat heeft deelgenomen met de papieren en digitale afnames per leerjaar

Digitaal			Papier			Totaal	
Groep	Aantal scholen		Groep	Aantal scholen	Groep	Aantal scholen	
Groep 1&2	7				Groep 1&2	7	
Groep 1	1	&	Groep 2	1	Groep 1&2**	1	
Groep 2	1	&	Groep 1	1	Groep 1&2**	1	
Groep 2	2	&	Groep 2	2	Groep 2*	2	
Groep 2	14				Groep 2	14	
			Groep 1&2	15	Groep 1&2	15	
			Groep 2	9	Groep 2	9	
<b>Totaal</b>	<b>25</b>			<b>28</b>		<b>49</b>	

\* Scholen die zowel de papieren als de digitale versie hebben afgenomen in dezelfde leerjaren

\*\* Scholen die de papieren versie in andere groepen hebben afgenomen dan de digitale versie

### Representativiteit van de steekproef

Het kwaliteitsonderzoek vond plaats bij 49 scholen; 1037 kinderen uit groep 1 en 2 deden mee.

De representativiteit van de steekproef uit het kwaliteitsonderzoek<sup>1</sup> is onderzocht door te kijken naar vier variabelen: regio, urbanisatiegraad, percentage achterstandsleerlingen en sekse.

- **Regio.** Bij de definitie van de variabele regio is uitgegaan van de CBS-indeling naar landsdeel. Dit betekent dat er vier regio's te onderscheiden zijn. Regio Noord omvat de provincies Groningen, Friesland en Drenthe; regio Oost de provincies Overijssel, Gelderland en Flevoland; regio West de provincies Utrecht, Noord-Holland, Zuid-Holland en Zeeland en regio Zuid de provincies Noord-Brabant en Limburg.
- **Urbanisatiegraad.** Bij de definitie van de variabele urbanisatiegraad is ervoor gekozen om de indeling naar vijf niveaus die gebruikelijk is bij het CBS te reduceren tot een tweedeling in enerzijds niet tot matig verstedelijkt (platteland) en anderzijds sterk tot zeer sterk verstedelijkt (stad). Een dergelijke tweedeling blijkt in de praktijk goed te volstaan (cf. Van Boxtel & Hemker, 2009).
- **Percentage achterstandsleerlingen.** Bij de definitie van de variabele 'schooltype' is gebruikgemaakt van de formatiegewichten van de leerlingen binnen een school volgens de meest recente regeling van OCW. Daarin worden drie niveaus onderscheiden die gebaseerd zijn op het opleidingsniveau van de ouders:
  - 0.0: één van de ouders of beide ouders heeft of hebben een opleiding gehad uit categorie 3
  - 0.3: beide ouders of de ouder die belast is met de dagelijkse verzorging hebben of heeft een opleiding uit categorie 2 gehad
  - 1.2: één van de ouders heeft een opleiding gehad uit categorie 1 en de andere ouder een opleiding uit categorie 1 óf 2

In deze indeling wordt verwezen naar de volgende categorieën in het opleidingsniveau van de ouders: 1 = maximaal basisonderwijs of (V)SO-ZMLK, 2 = maximaal LBO/VBO, praktijkonderwijs of VMBO basis- of kaderberoepsgerichte leerweg, en 3 = overig VO en hoger. Leerlingen met een formatiegewicht van 0.3 of 1.2 zijn te definiëren als achterstandsleerlingen.

Scholen zijn ingedeeld naar het percentage achterstandsleerlingen volgens een indeling in twee typen: (1) percentage achterstandsleerlingen [0 tot 0,10], (2) percentage achterstandsleerlingen [0,10 tot 1].

- **Sekse.** Bij de variabele sekse is een tweedeling naar jongens en meisjes gehanteerd.

<sup>1</sup> Drie scholen konden niet meegenomen worden in de analyse van de representativiteit, omdat voor deze scholen de regio, de urbanisatiegraad en het percentage achterstandskinderen onbekend waren.



Tabel 4.3 Representativiteit van de steekproef uit het kwaliteitsonderzoek

Variabele	Categorie	Populatie		Steekproef Kleuter in beeld - Rekenen	
		%	N	N	%
<b>Regio</b>	Noord	15,0	8	8	17,4
	Oost	24,6	13	13	28,3
	West	41,9	17	17	37,0
	Zuid	18,5	8	8	17,4
<b>Urbanisatie</b>	Platteland	65,0	32	32	69,6
	Stad	35,0	14	14	30,4
<b>Percentage achterstandskinderen</b>	0 – 10%	68,1	30	30	65,2
	>10%	31,9	16	16	34,8
<b>Sekse</b>	Jongen	50,5	65	65	51,6
	Meisje	49,5	61	61	48,4

We zien dat het percentage scholen met meer dan 10% achterstandskinderen niet significant afwijkt van de hele populatie scholen in Nederland. Ook de verdeling over regio's en de urbanisatiegraad in de steekproef wijkt niet af van de verdeling in heel Nederland (zie Tabel 4.4). De verhouding van jongens en meisjes in de steekproef is ook niet significant verschillend van de verhouding in de populatie (zie Tabel 4.4). Hierbij dient wel opgemerkt te worden dat voor het merendeel van de kinderen het geslacht niet ingevuld is (N = 911, 88%). Omdat in Kleuter in beeld geen relatieve normering wordt gebruikt, maar gewerkt wordt met standaarden die verbonden zijn aan de inhoud heeft de representativiteit van de steekproef geen impact op de standaarden.

Tabel 4.4 Toetsing representativiteit

Variabele	$\chi^2$	df	N	P
<b>Regio</b>	0,73	3	46	0,867
<b>Urbanisatie</b>	0,42	1	46	0,518
<b>Percentage achterstandskinderen</b>	0,17	1	46	0,678
<b>Sekse</b>	0,06	1	126	0,800

### 4.3 Ontwerp van het kwaliteitsonderzoek

Voor de twee subdomeinen Getalbegrip en Meten zijn zestig opdrachten ontwikkeld om uit te testen in het kwaliteitsonderzoek. Voor elk van deze twee subdomeinen zijn boekjes gemaakt van 15 opdrachten om voor te leggen aan de kinderen. Voor elk subdomein zijn de zestig opdrachten verdeeld over vier boekjes voor leerjaar 2. Uit elk boekje zijn de drie tot vier makkelijkste opdrachten ook opgenomen in het boekje voor leerjaar 1 (zie Tabel 4.5). De opdrachten die de link zijn tussen de boekjes komen in meer boekjes voor. Deze boekjes konden digitaal gemaakt worden of op papier.

Tabel 4.5 Ontwerp voor boekjes voor Getalbegrip en Meten en aantal observaties per boekje

Groep	Boekje	Makkelijk				Gemiddeld tot moeilijk				GB	MT
		4	4	4	3	11	11	11	12		
1	1	1	1	1	1					254	252
2	2	1				1				208	155
2	3		1				1			140	172
2	4			1				1		199	164
2	5				1				1	163	166
										964	909

In het kwaliteitsonderzoek vulde de leerkracht het (indirecte) observatieformulier van de leerkrachtroute in. Hierdoor zijn de gebruikte observatoren uit de steekproef van 49 scholen een goede afspiegeling van de observatoren die het observatie-instrument daadwerkelijk in de praktijk gaan gebruiken. De leerkrachten in dit onderzoek hebben dezelfde voorbereiding als leerkrachten die het definitieve instrument gebruiken. Daarnaast zijn de randvoorwaarden waaronder de observaties worden gedaan hetzelfde als in de praktijk in de klas. De leerkrachtroute is bedoeld voor de leerkracht en is ontwikkeld, omdat de leerkracht al een goed zicht heeft op de ontwikkeling van de kleuters. De leerkracht ziet de kleuters immers dag in dag uit. Voor de kindroute is de leerkracht ook de beoogde observator (zie hoofdstuk 2). Er heeft dus geen selectie van observatoren plaatsgevonden, maar uitsluitend een selectie van scholen (zie paragraaf 4.2).

Alle kinderen maakten twee subdomeinen uit de kindroute: de twee subdomeinen met opdrachten (Getalbegrip en Meten). Daarnaast deden minstens acht kinderen één activiteit (Bewerkingen, Meetkunde of Verbanden). Voor deze acht kinderen vulde de leerkracht ook het observatieformulier van de leerkrachtroute in. Het ontwerp voor de combinaties van de verschillende subdomeinen staat in Tabel 4.6. In Tabel 4.7 staat het aantal observaties per versie. De (indirecte) observaties in de leerkrachtroute en de directe observaties van de activiteiten (Bewerkingen, Meetkunde of Verbanden) werden op een papieren observatieformulier ingevuld tijdens het kwaliteitsonderzoek.

Tabel 4.6 Ontwerp voor de combinaties van de verschillende subdomeinen

LKroute	Getalbegrip (GB)	Meten (MT)	Bewerkingen (BW)	Meetkunde (MK)	Verbanden (VB)
1	1	1	1		
1	1	1		1	
1	1	1			1
	1	1			

Tabel 4.7 Aantal observaties per versie

LKroute	Getalbegrip (GB)	Meten (MT)	Bewerkingen (BW)	Meetkunde (MK)	Verbanden (VB)
155	154	152	166	0	0
166	169	157	0	181	0
232	231	215	0	0	252
122	150	103	0	0	0
	260	282	0	0	0
675	964	909	166	181	252

## 4.4 Verwerking en analyses

Nadat de scholen klaar waren met het onderzoek, hebben ze alle materialen teruggestuurd naar Cito. Daar zijn alle papieren opdrachtenboekjes handmatig nagekeken en de gegeven antwoorden ingevoerd in het verwerkingssysteem. Bij de digitale opdrachten werden de resultaten in een database opgeslagen. Vanuit de database is een export gemaakt. De geëxporteerde bestanden zijn vervolgens geïmporteerd in het verwerkingssysteem, samen met de data van de kinderen die de papieren opdrachten hebben gemaakt. De observatieformulieren van de leerkrachtroute en van de activiteiten van de kindroute waren voorafgaand aan het onderzoek optisch leesbaar gemaakt. Dit betekent dat na het onderzoek de ingevulde observatieformulieren ingelezen konden worden. Op deze manier kwamen alle observaties automatisch in het systeem en was handmatige verwerking van de observatieformulieren niet nodig. Nadat alle data van de leerkrachtroute en kindroute (digitale en papieren opdrachten en activiteiten) opgenomen waren in het verwerkingssysteem, zijn er analysebestanden gegenereerd. Verder hebben de deelnemers na afloop een online evaluatieformulier ingevuld, waarbij ze feedback konden geven op onder meer de inhoud en vorm van de opdrachten.

Met behulp van itemanalyses heeft eerst een sleutelcontrole plaatsgevonden. Er zijn geen foutieve sleutels geconstateerd. Hierna zijn analyses verricht om aan scholen een rapportage te verstrekken. Per leerling werd het aantal correcte opdrachten per subdomein uitgerekend en het aantal correcte observatiepunten per activiteit (waarbij observatiepunten die 'met hulp' of 'een beetje' correct waren als 0,5 werden geteld). Indien een leerling meer dan 40% ontbrekende waarnemingen had, werd dit met een ster aangegeven. Ook werd een overzicht van de indirecte observaties uit de leerkrachtroute verstrekt. Zie Figuur 4.1 voor een voorbeeld van de rapportage zoals deze aan scholen is gestuurd.

Figuur 4.1 Rapportage aan de scholen die hebben deelgenomen aan het kwaliteitsonderzoek: leerkrachtroute en kindroute

Jaargroep: 1 Groep: Groep 1																											
		Getalbegrip - Tellij		Getalbegrip - Hoofdesheiden		Getalbegrip - Getallen		Getalbegrip - Relaties tussen tellij, hoofdesheiden en getallen		Bewerkingen - Optellen en aftrekken met hele getallen en delen met hele getallen		Meten - Lenge en omtrek		Meten - Oppervlakte		Meten - Inhoud		Meten - Gewicht		Meten - Tijd		Meetkunde - Meeten in de ruimte		Meetkunde - Construeren		Verbanden	
Naam		E1	<E1	E1	<E1	<E1	<E1	E1	E1	E1	E1	E1	E2	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	<E1	<E1	E1	E1
K B		E1	<E1	E1	<E1	<E1	<E1	E1	E1	E1	E1	E1	E2	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	<E1	<E1	E1	E1
S. vd B		M2	M2	M2	E2	E1	E1	E1	E1	E1	E1	E1	E2	<E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	<E1	E1	E1	E1
L. v D		E1	E1	E1	E1	<E1	E1	E1	E1	E1	E1	E1	E2	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1
P. G		E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E2	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1
M. N		M2	M2	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	<E1	E1	E1	E1
N. S		M2	M2	E1	E1	E1	E1	E1	E1	E1	E1	E1	E2	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1

Jaargroep: 1 Groep: Groep 1													
		Opdrachten (papier)						Activiteiten					
		Totaal		Getalbegrip		Meten		Bewerkingen		Meetkunde		Verbanden	
Naam		#	goed	#	goed	#	goed	#	goed	#	goed	#	goed
B. B		30	15	15	5*	15	10						
K B		60	24	15	5*	15	6			30	13		
S. vd B		60	43,5	15	10	15	8			30	25,5		
L. v D		60	31,5	15	8	15	5*			30	18,5		
P. G		60	49	15	13	15	12			30	24		
M. N		60	37,5	15	12	15	7			30	18,5		
E. S		30	16	15	11	15	5*						
N. S		45	34,5	15	12					30	22,5		
V. v W		30	19	15	7	15	12						

De scholen kregen een korte uitleg over de rapportages:

“#: het aantal opdrachten/observatiepunten dat een kind in totaal aangeboden heeft gekregen. In de kolom ‘goed’ bij de activiteiten staat het aantal beheerste observatiepunten. Bij de observatiepunten kon soms ‘nee, beetje of ja’ ingevuld worden. ‘Nee’ telt dan als 0 punten, ‘beetje’ als 0,5 punt en ‘ja’ als 1 punt. Op dit moment analyseren we de resultaten en kunnen we helaas nog niet zeggen welk niveau (<E1, E1, M2, E2 en >E2) het beste past bij elk ingevuld observatieformulier. Dit hangt namelijk af van de uitkomsten van de standaardbepaling en is pas mogelijk bij de uitgave van Kleuter in beeld - Rekenen.”

#### 4.5 Samenstelling versies uitgave

Bij de opdrachten zijn voor elk subdomein drie verschillende niveauseries (taken) samengesteld. Omdat met elke niveauserie het mogelijk moest zijn om te constateren dat een kind een hoger of lager functioneringsniveau had, is gekozen om bij de versie voor groep 1 ook een aantal moeilijker opdrachten op te nemen en bij de versie voor midden groep 2 en eind groep 2 ook een aantal makkelijker opdrachten. In de versie voor groep 1 was twee derde van de opdrachten makkelijk tot enigszins makkelijk en een derde enigszins moeilijk tot moeilijk. Voor de versie voor eind groep 2 was dit precies andersom (zie Figuur 4.2). Om de opdrachten in de komende jaren goed te kunnen evalueren en eventueel nieuwe opdrachten te kalibreren en op te nemen is besloten om een kleine overlap tussen de opeenvolgende versies te creëren.

Figuur 4.2 Ontwerp voor de niveauseries van de uitgave

	Versie groep 1 ★	Versie midden groep 2 ★★	Versie eind groep 2 ★★★
Makkelijk	2/3	1/6	1/3
Enigszins makkelijk		2/3	
Enigszins moeilijk	1/3		2/3
Moeilijk		1/6	

Bij de activiteiten zijn de definitieve observatiepunten geselecteerd. Hierbij is rekening gehouden met inhoudelijke indelingen (zie ook paragraaf 3.3 met de toetsmatrices voor elk van de subdomeinen) en waren de opmerkingen uit het veld van belang. Naast deze inhoudelijke overwegingen speelden psychometrische gegevens uit het kwaliteitsonderzoek een rol (zoals moeilijkheid, onderscheidend vermogen en de betrouwbaarheid) en efficiëntie. De afname van de activiteiten en opdrachten mag niet te veel tijd in beslag nemen. Per subdomein worden er één of meer voorbeelden gegeven van opdrachten/observatiepunten die met een bepaalde reden juist wel of niet zijn opgenomen in de uitgave/opgenomen zijn in een bepaalde niveauserie.

## **Opdrachten**

Scholen hebben behoefte aan korte taken, waarmee ze op een snelle manier de vaardigheid van kinderen in beeld kunnen brengen. Daarom bevat elke taak 18 opdrachten. Niveauversie ★ bevat vooral de relatief makkelijke opdrachten, maar ook enkele opdrachten die wat moeilijker zijn. Niveauversie ★★ bevat de moeilijkere opdrachten. Zo was het mogelijk om met elke niveauversie verschillende niveaus in beeld te brengen.

### **Getalbegrip**

In de taken (niveauversies) Getalbegrip komen verschillende onderdelen voor, namelijk: telrij, hoeveelheden, getallen en relaties tussen telrij, hoeveelheden en getallen. Voor elk onderdeel zijn de opdrachten zoveel mogelijk eerlijk verdeeld over de drie niveaus.

Bij het maken van de opdrachten Getalbegrip is niveauverschil aangebracht door rekening te houden met:

- de positie van de aspecten/doelen op de leerlijn
- hoogte van de getallen
- grootte van de hoeveelheden
- complexiteit van de afbeeldingen
- mate van abstractie
- opdrachttype

Bijvoorbeeld de aspecten 'tellen met sprongen' en 'splitsen van getallen' zijn niet in de ★ opdrachten opgenomen vanwege de positie van deze doelen op de leerlijn.

Voorafgaand aan het kwaliteitsonderzoek was het bij sommige opdrachten moeilijk in te schatten bij welk niveau deze het beste zouden aansluiten. Daarom zijn deze opdrachten zowel in groep 1 als in groep 2 ingezet. Op basis van de psychometrische gegevens van het kwaliteitsonderzoek zijn de betreffende opdrachten vervolgens toegewezen aan het niveau waar ze het beste bij pasten.

### **Meten**

In de taken (niveauversies) Meten komen verschillende onderdelen voor, namelijk: lengte en omtrek, oppervlakte, inhoud, gewicht, tijd en geld. Voor elk onderdeel zijn de opdrachten zoveel mogelijk eerlijk verdeeld over de drie niveaus.

Bij het maken van de opdrachten Meten is niveauverschil aangebracht door rekening te houden met:

- de positie van de aspecten/doelen op de leerlijn
- hoogte van de getallen
- grootte van de hoeveelheden
- complexiteit van de afbeeldingen
- mate van abstractie
- opdrachttype

De begrippen 'breed' en 'smal' zijn bijvoorbeeld niet opgenomen in de ★ opdrachten vanwege hun positie op de leerlijn, evenals de opdrachten 'plaatsen van gebeurtenissen in tijdsvolgorde'.

Voorafgaand aan het kwaliteitsonderzoek was het bij sommige opdrachten moeilijk in te schatten bij welk niveau deze het beste zouden aansluiten. Daarom zijn deze opdrachten zowel in groep 1 als in groep 2 ingezet. Op basis van de psychometrische gegevens van het kwaliteitsonderzoek zijn de betreffende opdrachten vervolgens toegewezen aan het niveau waar ze het beste bij pasten.

## **Activiteiten**

### **Bewerkingen**

Voor het kwaliteitsonderzoek zijn 49 observatiepunten geformuleerd bij de activiteit Bewerkingen. Daarnaast zijn zes afbreekpunten ingebouwd, zodat een deel van de activiteit overgeslagen kon worden als het kind een onderdeel echt niet kon of gefrustreerd raakte tijdens dat onderdeel van de activiteit. Zo kon de leerkracht bijvoorbeeld na de eerste twee observatiepunten van het onderdeel 'optellen' besluiten om de volgende vier observatiepunten van dat onderdeel niet aan te bieden en gelijk door te gaan naar het

onderdeel 'aftrekken'. Ook in het onderdeel 'aftrekken' en het onderdeel 'splitsen met materiaal' zijn dit soort afbreekpunten ingebouwd, waarbij de leerkracht wordt doorverwezen naar het volgende onderdeel in de activiteit. Uit het kwaliteitsonderzoek is gebleken dat de afbreekpunten nauwelijks zijn gebruikt (maximaal drie keer per afbreekpunt op 166 leerlingen). Het gros van de kinderen kan de activiteit goed uitvoeren, zonder frustratie. Daarom is besloten om de afbreekpunten niet op te nemen in het observatieformulier voor de uitgave. (Uiteraard kan een leerkracht alsnog besluiten om onderdelen van de activiteit over te slaan of de activiteit af te breken bij kinderen voor wie dat van belang is.) Bij de onderdelen over 'optellen' en 'aftrekken' kon de leerkracht bij elke opdracht die hij het kind gaf, aangeven welke aanpak het kind koos bij het uitvoeren van de opdracht. Uit een Principale Componenten Analyse bleek dat de verschillende soorten aanpak niet samenhangen met de vaardigheid. Daarom hebben we besloten om alle observatiepunten (12 punten) die 'de aanpak' betreffen, te laten vervallen.

Na het kwaliteitsonderzoek hebben we ook geëvalueerd welke opdrachten qua moeilijkheidsgraad het meest geschikt zijn voor de uitgave. Uit de data blijkt dat de opdrachten tot en met 6 vaak te gemakkelijk zijn. Veel van deze opdrachten zijn daarom komen te vervallen. Opdrachten tot en met 10 en 15 blijken beter aan te sluiten bij het niveau van de kinderen. Van de 22 opdrachten met bijbehorende observatiepunten die een bewerking (optellen, aftrekken, splitsen, (ver)delen) betreffen, zijn er uiteindelijk acht vervallen.

Van de zes bewerkingen over optellen zijn er twee vervallen: ' $2+3=5$ ' en ' $2+5=7$ '. De eerste had een p-waarde groter dan ,90 in zowel groep 1 als groep 2. De tweede was voor groep 2 te makkelijk (p-waarde groter dan ,90). Bij de bewerkingen over aftrekken hebben we drie van de zes observatiepunten laten vervallen, namelijk beide opdrachten over aftrekken tot en met 6 (' $5-3=2$ ' en ' $6-5=1$ '). Deze hadden een p-waarde groter dan of gelijk aan ,90 in zowel groep 1 als groep 2. Daarnaast hebben we ook één van de opdrachten over aftrekken tot en met 10 (' $9-6=3$ ') laten vervallen. Deze was voor groep 2 te makkelijk (p-waarde groter dan ,90) en het onderscheidend vermogen in groep 2 was niet zo groot bij deze opdracht. De andere opdracht over aftrekken tot en met 10 (' $10-4=6$ ') is voor groep 2 ook te makkelijk (p-waarde groter dan ,90), maar voor groep 1 wel geschikt qua moeilijkheidsgraad. Het onderscheidend vermogen van deze opdracht is in beide groepen goed. Omwille van een goede inhoudelijke opbouw en opbouw in moeilijkheid in de activiteit hebben we besloten deze opgave wel op te nemen in de uitgave.

Het tweede deel van de activiteit Bewerkingen gaat over splitsen met materiaal en (ver)delen met materiaal. Van de zes bewerkingen over splitsen zijn er twee vervallen, namelijk splitsen tot en met 6: ' $5$  is 2 en ...' en ' $5$  is 1 en ...'. Deze hadden een p-waarde groter dan of gelijk aan ,90 in zowel groep 1 als groep 2. Van de bewerkingen over splitsen tot en met 10 was één opdracht (' $9$  is 6 en ...') makkelijker voor groep 1 ( $p = ,91$ ) dan voor groep 2 ( $p = ,89$ ). De andere opdracht die over splitsen tot en met 10 gaat (het kind kan zelf een andere verdeling bedenken van 9) heeft in groep 1 een minder goed onderscheidend vermogen. Deze opdracht hebben we ook behouden vanwege de opbouw qua inhoud en moeilijkheid én omdat deze opdracht in groep 2 wel een goed onderscheidend vermogen heeft. Het laatste deel van de activiteit betreft bewerkingen over (ver)delen met materiaal. Van de vier bewerkingen over (ver)delen is er één vervallen, namelijk 'kan een bepaalde hoeveelheid materiaal uitdelen'. Deze opdracht was te makkelijk voor zowel groep 1 als groep 2 (p-waarde groter dan ,90).

Bij de opdrachten over optellen, aftrekken en (ver)delen konden de leerkrachten ook aangeven of de kinderen de 'begrippen' over respectievelijk optellen (samen en erbij), aftrekken (over) en (ver)delen (delen, verdelen, gelijk, eerlijk, oneerlijk) kennen. Uit het kwaliteitsonderzoek bleek dat het niet zinvol is om dit bij iedere bewerking te laten scoren. Eén keer per onderdeel volstaat. Daarom hebben we de meeste observatiepunten over de 'begrippen' laten vervallen en vragen we daar per onderdeel (optellen, aftrekken en (ver)delen) nog maar één keer naar in de uitgave.

Na de selectie bleven in de activiteit Bewerkingen uiteindelijk 17 observatiepunten over, 3 dichotome en 14 polytome observatiepunten.

### **Meetskunde**

Voor het kwaliteitsonderzoek zijn op basis van literatuur en een gesprek met een rekenexpert jonge kind dertig opdrachten met bijbehorende observatiepunten geformuleerd bij de activiteit Meetskunde. Na dit onderzoek is geëvalueerd welke opdrachten qua moeilijkheidsgraad het meest geschikt zijn voor de uitgave. Daarbij hebben we ook de praktische uitvoerbaarheid laten meewegen.

Uit de analyses na het kwaliteitsonderzoek bleek dat acht observatiepunten zowel in groep 1 als in groep 2 een p-waarde groter dan ,90 hadden. Zes van deze acht observatiepunten hebben we niet opgenomen in de uitgave. Het betreft de observatiepunten: 'gebruikt begrip onder', 'begrijpt begrip op', 'gebruikt begrip in', 'begrijpt begrip dichtbij', 'begrijpt begrip veraf' en 'torentje nabouwen'. De andere twee observatiepunten ('begrijpt begrip achter' en 'begrijpt begrip naast') hebben we behouden om ervoor te zorgen dat we binnen de activiteit een goede opbouw zowel inhoudelijk als qua moeilijkheid hebben. Daarnaast hadden drie observatiepunten alleen in groep 2 een p-waarde groter dan ,90. Eén van deze observatiepunten hebben we niet opgenomen in de uitgave. Het gaat om het observatiepunt 'begrijpt het begrip voor'. Tijdens het kwaliteitsonderzoek bleek dat deze opdracht verwarrend was en daardoor niet eenduidig werd geïnterpreteerd. De andere twee observatiepunten, 'bepalen wat vanuit het standpunt ander niet te zien is' en 'aanpassingen aanbrengen in gebouwd bouwwerk', hebben we behouden om ervoor te zorgen dat we binnen de activiteit een goede inhoudelijke opbouw en opbouw in moeilijkheid hebben. De betreffende observatiepunten hebben een goed onderscheidend vermogen. De twee opdrachten over het werken met een plattegrond waren qua moeilijkheidsgraad en onderscheidend vermogen heel geschikt om in de uitgave op te nemen. Tijdens het kwaliteitsonderzoek bleek echter dat deze opdrachten om praktische redenen moeilijk uitvoerbaar waren. Daarom hebben we besloten om deze opdrachten niet op te nemen in de uitgave. Na de selectie bleven 21 observatiepunten over, 12 dichotome en 9 polytome observatiepunten.

### **Verbanden**

Voor het kwaliteitsonderzoek zijn 23 observatiepunten geformuleerd bij de activiteit die inhoudelijk goed passen bij de vaardigheid Verbanden. Daarnaast was het mogelijk de activiteit af te breken na start 1 als het kind start 1 echt niet kon of gefrustreerd raakte tijdens de activiteit. Uit het kwaliteitsonderzoek is gebleken dat kinderen de gehele activiteit goed kunnen uitvoeren, zonder frustratie. Daarop is besloten om het afbreekpunt niet op te nemen in het observatieformulier voor de uitgave. Uit de analyses na dit onderzoek bleek dat bij start 2 zes observatiepunten 'Kan een staafdiagram construeren (olifant)', 'Kan een staafdiagram construeren (leeuw)', 'Kan een staafdiagram construeren (aap)', 'Kan een staafdiagram construeren (papegaai)', 'Kan een staafdiagram construeren (zebra)' en 'Kan een staafdiagram construeren (beer)' een p-waarde hadden van respectievelijk ,89 ,91 ,89 ,88 ,88 en ,97 voor groep 1 en ,93, 97 ,99 ,99 ,98 en ,99 voor groep 2. Deze observatiepunten bleken te makkelijk voor groep 1 en 2 en maakten niet genoeg onderscheid tussen de vaardige en minder vaardige kinderen. Deze zes observatiepunten zijn daarom afgevallen voor de uitgave. Bij start 1 zijn vier observatiepunten en bij start 2 zijn twee observatiepunten die een p-waarde hebben hoger dan ,90 voor groep 2. Deze zes observatiepunten 'Kan een beelddiagram construeren', 'Kan informatie over de olifanten aflezen uit het beelddiagram', 'Kan informatie uit het beelddiagram met elkaar vergelijken en bepalen van welk dier er de meeste zijn', 'Kan informatie uit het beelddiagram met elkaar vergelijken en bepalen van welke dieren er meer en minder zijn', 'Kan informatie uit het staafdiagram met elkaar vergelijken en bepalen van welk voer er het minste is' en 'Kan informatie uit het staafdiagram met elkaar vergelijken en bepalen van welk voer er meer is' hadden voor groep 1 een p-waarde van respectievelijk ,83 ,80 ,66 ,73 ,88 en ,88. Voor groep 1 blijken deze zes observatiepunten moeilijker te zijn dan voor groep 2 en maken deze observatiepunten wel een goed onderscheid tussen de vaardige en minder vaardige kinderen in groep 1. Daarom zijn deze zes observatiepunten wel opgenomen in de uitgave. Na de selectie bleven 17 observatiepunten over, alle 17 polytome observatiepunten.

### Leerkrachtroute

Zoals beschreven in hoofdstuk 3, bevat de leerkrachtroute 16 observatiepunten (alle 16 onderdelen behorend bij de vijf subdomeinen). Per observatiepunt kan de leerkracht kiezen uit zes niveaus: <M1, M1, E1, M2, E2 of >E2. Daarnaast is ter ondersteuning bij elk observatiepunt een beschrijving en voorbeeld opgenomen bij E1 en E2 gebaseerd op de literatuur (zie overzicht in paragraaf 2.2) en gegevens vanuit de toets Rekenen voor kleuters (Koerhuis, 2010). Het eerste concept van de leerkrachtroute is meegenomen in het kwaliteitsonderzoek. De adequaatheid van de beschrijvingen en voorbeelden uit de leerkrachtroute zijn vervolgens gevalideerd aan de observaties uit de kindroute.

Bij de observatiepunten voor Getalbegrip en Meten worden in de beschrijving voorbeelden genoemd, die zijn ontleend aan de opdrachten in de kindroute. Bijvoorbeeld, het observatiepunt over Relaties tussen telrij, hoeveelheden en getallen bij E1 (subdomein Getalbegrip) is als volgt beschreven:

“Kan telwoorden, hoeveelheden en voor het kind bekende getsymbolen tot 10 aan elkaar koppelen. Voorbeeld: De juf heeft 3 blokjes, het kind heeft 2 blokjes. Als je vraagt ‘Hoeveel blokjes hebben we samen?’ kan het kind het getalkaartje met het symbool 5 aanwijzen.”

In de kindroute is gezocht naar geschikte voorbeelden die genoemd kunnen worden in de leerkrachtroute. Als voorbeeld van E1 zijn opdrachten gezocht waarbij ongeveer 75% in groep 1 de opdracht goed maakte en meer dan 90% in groep 2. Als voorbeeld van E2 zijn opdrachten gezocht waarbij ongeveer 75% in groep 2 de opdracht goed maakte en minder dan 60% in groep 1. In het bovengenoemde voorbeeld kon 77% van de kinderen uit groep 1 de relatie tussen het totaal aantal blokjes en het symbool 5 leggen en 91% van de kinderen uit groep 2.

Voor de observatiepunten in de leerkrachtroute van Bewerkingen, Meetkunde en Verbanden is bij de activiteiten gekeken naar geschikte beschrijvingen voor E1 en E2.

- Een observatiepunt uit de kindroute werd geschikt gevonden voor de beschrijving van E1 in de leerkrachtroute als ongeveer 75% in groep 1 het punt goed doet en meer dan 90% in groep 2.
- Een observatiepunt uit de kindroute werd geschikt gevonden voor de beschrijving van E2 in de leerkrachtroute als ongeveer 75% in groep 2 het punt goed doet en minder dan 60% in groep 1.

Bij de uiteindelijke samenstelling van de leerkrachtroute in de uitgave hebben we gezocht naar consensus tussen de inhoud van het observatiepunt en de moeilijkheid van het voorbeeld en/of toelichting die we daarbij gaven. Een deel van de observatiepunten uit het kwaliteitsonderzoek hebben we voor de uitgave in beperkte of grotere mate aangepast (zie E2 van Meten - Lengte en omtrek). Een ander deel bleef hetzelfde of nagenoeg hetzelfde (zie E1 van Meten - Lengte en omtrek).

Meten - Lengte en omtrek in het Kwaliteitsonderzoek:

Meten - Lengte en omtrek				
<E1	E1	M2	E2	>E2
<input type="checkbox"/>	<input type="checkbox"/> Kan voorwerpen en lengtes vergelijken en ordenen. Begrijpt en gebruikt daarbij begrippen als ‘groot’ en ‘klein’, ‘lang’ en ‘kort’. Voorbeeld: Het kind kiest 3 kinderen uit de klas uit. ‘Wie is het kleinste? Wie is het grootste? Zet de kinderen op een rij van klein naar groot.’	<input type="checkbox"/>	<input type="checkbox"/> Kan redeneren over lengte en omtrek en gebruikt hierbij informele instrumenten. Voorbeeld: Leg enkele stroken met een verschillende lengte neer (bijv. lengte en breedte van een A4). ‘De tafel is 5 stroken lang. Welke stroken heb ik gebruikt?’	<input type="checkbox"/>



Meten - Lengte en omtrek in de Uitgave:

Meten	
<b>Lengte en omtrek</b>	
<M1   M1   E1   M2   E2   >E2 ○   ○   ○   ○   ○   ○	
Kan voorwerpen en lengtes vergelijken en ordenen. Begrijpt en gebruikt daarbij begrippen als <i>groot</i> en <i>klein</i> , <i>lang</i> en <i>kort</i> . Voorbeeld: Het kind kiest 3 kinderen uit de klas uit. 'Wie is het kleinste?' 'Wie is het grootste?' 'Zet de kinderen van klein naar groot.'	Kan voorwerpen en lengtes vergelijken en ordenen. Begrijpt en gebruikt daarbij begrippen als <i>dik</i> en <i>dun</i> , <i>breed</i> en <i>smal</i> . Voorbeeld: Geef het kind 4 boeken van verschillende dikte. 'Zet de boeken op een rij van dik naar dun.'

Deze paragraaf wordt afgesloten met een statistische beschrijving met toets- en itemkenmerken van de versies voor de uitgave.

**Statistische beschrijving van de uitgave**

In de hoofdstukken 5 en 6 zullen de kalibratie en de standaardbepaling van de kindroute uitgebreid worden beschreven. Voorafgaand aan deze hoofdstukken geven we hier een globaal overzicht van de beschrijvende gegevens van de kind- en leerkrachtroute. In Tabel 4.8 en Tabel 4.9 worden de beschrijvende gegevens van de subdomeinen gegeven op de somscore en de vaardigheidsschaal.

Bij de niveauversies met opdrachten (zie Tabel 4.8) zijn de taken beperkt gehouden tot 18 opdrachten. Aan de item-restcorrelatie (R<sub>ir</sub>) en de item-testcorrelatie (R<sub>it</sub>) is te zien dat de opdrachten gemiddeld goed onderscheid maken tussen kinderen met een lage en hoge vaardigheid. De gemiddelde vaardigheid loopt op met de niveauversies, zoals bedoeld. Het verschil in de geschatte moeilijkheid tussen de niveauversies is vrij klein. Het maximale verschil tussen niveauversies in de gemiddelde p-waarde is respectievelijk 0,09 (GB) en 0,11 (MT). Bij beide subdomeinen is de niveauversie voor midden groep 2 (★★) iets makkelijker (geschatte gemiddelde p-waarde = 0,75) dan de niveauversie voor eind groep 1. Voor leerlingen van groep 2 zijn, zoals beoogd, de niveauversies midden groep 2 (★★) makkelijker dan de versies eind groep 2 (★★★).

Tabel 4.8 Beschrijvende gegevens van de niveauversies met opdrachten

Sub-domein	Versie	Groep	Som			Theta		Gem. p.	Gem. Rit	Gem. Rir
			Max.	Gem.	St.dev.	Gem.	St.dev.			
GB	★	Groep 1	18	11,9	3,82	0,05	0,39	0,66	0,47	0,37
	★★	Groep 2	18	13,6	3,42	0,44	0,43	0,75	0,46	0,36
	★★★	Groep 2	18	12,0	3,63	0,49	0,43	0,67	0,45	0,35
MT	★	Groep 1	18	11,4	3,95	0,06	0,40	0,64	0,48	0,38
	★★	Groep 2	18	13,5	3,71	0,41	0,46	0,75	0,49	0,39
	★★★	Groep 2	18	12,7	3,93	0,45	0,46	0,71	0,50	0,42

Tabel 4.9 Beschrijvende gegevens van de activiteiten

Sub-domein	Groep	No	Max.	Som		Theta		Gem. p	Gem. Rit	Gem. Rir
				Gem.	St.dev.	Gem.	St.dev.			
<b>BW</b>	Groep 1	17	31	21,9	6,87	0,29	0,37	0,71	0,56	0,48
	Groep 2	17	31	26,5	4,92	0,45	0,41	0,86	0,53	0,44
<b>MK</b>	Groep 1	21	30	18,7	5,98	0,19	0,32	0,66	0,56	0,50
	Groep 2	21	30	23,9	5,06	0,44	0,37	0,86	0,44	0,36
<b>VB</b>	Groep 1	17	34	20,5	8,91	0,21	0,44	0,60	0,64	0,58
	Groep 2	17	34	28,6	5,54	0,53	0,38	0,84	0,54	0,46

Bij de activiteiten wordt hetzelfde observatieformulier gebruikt voor groep 1 en groep 2. We zien aan de item-restcorrelatie (Rir) en de item-testcorrelatie (Rit) dat de observatiepunten heel goed onderscheid maken tussen kinderen met een lage en hoge vaardigheid (zie Tabel 4.9). De gemiddelde vaardigheid loopt op met de groep, zoals bedoeld. De maximale somscore is hoger dan het aantal observatiepunten, omdat bij veel observatiepunten een onderscheid wordt gemaakt of een kind een activiteit zelfstandig kan uitvoeren of met hulp. Doordat dezelfde observatiepunten gehanteerd worden bij groep 1 en groep 2 is de gemiddelde p-waarde bij kinderen in groep 1 duidelijk lager dan bij kinderen in groep 2.

De (indirecte) observaties uit de leerkrachtroute zijn op een ordinaal meetniveau. Voor elk observatiepunt is daarom het percentage leerlingen weergegeven per functioneringsniveau in het kwaliteitsonderzoek, zie Tabel 4.10. In groep 1 is de modus voor 14 observatiepunten E1. Alleen voor twee onderdelen (Bewerkingen - Optellen en aftrekken met hele getallen en Verbanden) is de modus <E1. Bij groep 2 variëren alle onderdelen tussen <E1 en >E2. Bij 13 van de observatiepunten ligt de modus bij M2 en bij E2 voor de andere drie observatiepunten (drie onderdelen van Getalbegrip: Telrij, Hoeveelheden en Relaties tussen telrij, hoeveelheden en getallen).

Tabel 4.10 Beschrijvende gegevens van de leerkrachtroute

	Groep 1					Groep 2				
	<E1	E1	M2	E2	>E2	<E1	E1	M2	E2	>E2
Getalbegrip - telrij	23	51	18	7	1	4	8	33	44	11
Getalbegrip - hoeveelheden	21	59	16	5	0	5	11	38	39	7
Getalbegrip - getallen	22	58	13	7	0	6	10	41	35	9
Getalbegrip - relaties tussen telrij, hoeveelheden en getallen	28	52	14	6	1	6	10	37	41	7
Bewerkingen - optellen en aftrekken met hele getallen	55	29	11	3	1	12	17	42	24	5
Bewerkingen - vermenigvuldigen en delen met hele getallen	40	43	9	6	1	11	14	44	25	6
Meten - lengte en omtrek	25	57	15	3	0	4	13	55	25	3
Meten - oppervlakte	42	45	10	3	0	5	15	53	24	2
Meten - inhoud	36	51	10	3	0	7	14	51	25	4
Meten - gewicht	31	49	10	9	1	6	10	43	38	4
Meten - tijd	37	48	9	5	1	7	12	44	34	4
Meten - geld	23	63	12	3	0	5	10	56	25	5
Meetkunde - oriënteren in de ruimte	21	60	11	8	1	7	9	50	30	3
Meetkunde - construeren	32	47	11	10	1	8	9	43	37	4
Meetkunde - opereren met vormen en figuren	31	47	13	8	1	7	10	41	38	4
Verbanden	50	35	8	6	1	8	15	41	30	5



## 5 Kalibratie kindroute

Voor de constructie van de kindroute van Kleuter in beeld - Rekenen is een bank met opdrachten en observatiepunten ontwikkeld (zie hoofdstuk 3 en 4). Een dergelijke bank is nadrukkelijk niet eenvoudigweg een verzameling opdrachten en observatiepunten waaruit een toetsdeskundige min of meer naar willekeur een aantal opdrachten of observatiepunten selecteert om het nieuwe meetinstrument samen te stellen. In deze paragraaf wordt beschreven wat de vereisten zijn om van een deugdelijke en psychometrisch goed gefundeerde bank te kunnen spreken om de vaardigheid van een kind op een bepaald subdomein in beeld te brengen.

### 5.1 Bank met opdrachten en observatiepunten

#### Veronderstellingen

##### *Unidimensionaal continuüm*

Het algemene uitgangspunt is dat elk van de subdomeinen van rekenvaardigheid kunnen worden opgevat als een unidimensionaal continuüm (de reële lijn), en dat elke leerling voorgesteld kan worden als een punt op die lijn, met andere woorden: als een getal. Het getal drukt de mate van rekenvaardigheid uit op dat subdomein, waarbij een groter getal wijst op een grotere rekenvaardigheid. Het doel van de meetprocedure – het observeren van kinderen tijdens activiteiten en opdrachten – is de plaats van de leerling op dit continuüm zo nauwkeurig mogelijk te bepalen. De uitkomst van de meetprocedure bestaat strikt genomen uit twee grootheden: de eerste is de schatting van de plaats van de leerling op het vaardigheidscontinuüm. De tweede grootheid geeft aan hoe nauwkeurig die schatting is, en heeft dus de status van een standaardfout, te vergelijken met de standaardmeetfout uit de klassieke testtheorie.

##### *Latente vaardigheid*

De wijze waarop een kind de opdrachten en/of activiteiten uitvoert, wordt beschouwd als indicator van de vaardigheid, hetgeen ruwweg betekent dat men verwacht dat alle opdrachten en observatiepunten in de bank een subdomein van rekenvaardigheid meten. De vaardigheid zelf wordt als niet-observeerbaar beschouwd, en daarom gewoonlijk omschreven als een latente vaardigheid.

##### *'Moeilijkheid' in de Item Respons Theorie*

Hoewel verschillende opdrachten dezelfde vaardigheid meten, kunnen ze toch systematisch van elkaar verschillen. Het belangrijkste verschil tussen de opdrachten is hun moeilijkheidsgraad. Ditzelfde geldt voor de observatiepunten. In de klassieke testtheorie wordt moeilijkheidsgraad uitgedrukt met een zogeheten p-waarde, de proportie correct uitgevoerde opdrachten of observatiepunten. In de Item Respons Theorie (IRT) die voor het construeren van de banken met opdrachten en observatiepunten wordt gebruikt, hanteert men echter een andere definitie van moeilijkheid: ruwweg gesproken is het de mate van vaardigheid die nodig is om de opdracht of het observatiepunt goed te kunnen uitvoeren. Dit verschil in definitie van de moeilijkheidsgraad tussen klassieke theorie en IRT is uitermate belangrijk: men kan verwachten dat de p-waarde van een opdracht in groep 2 groter zal zijn dan in groep 1, waardoor duidelijk wordt dat de p-waarde een relatief begrip is: ze geeft de moeilijkheid aan van een opdracht of observatiepunt in een bepaalde populatie (groep in ons geval). Binnen de IRT is de moeilijkheid van een opdracht/observatiepunt gedefinieerd in termen van de onderliggende vaardigheid, zonder enige referentie aan een bepaalde populatie van leerlingen. Zo kan men ook de uitspraak begrijpen dat in de IRT de vaardigheid en moeilijkheid op eenzelfde schaal liggen.

### **Kansmodel**

De ruwe omschrijving van de moeilijkheidsgraad die in de vorige alinea werd gehanteerd (de mate van vaardigheid die nodig is om de opdracht of het observatiepunt goed te kunnen uitvoeren) behoeft enige verdere uitwerking. Men zou deze omschrijving kunnen opvatten als een drempel: heeft een leerling die mate van vaardigheid niet, dan kan hij de opdracht of activiteit niet juist uitvoeren; heeft hij die drempel wel gehaald, dan voert het kind de opdracht of activiteit juist uit. Deze interpretatie weerspiegelt een deterministische kijk op het gedrag van het kind, die echter in de praktijk geen standhoudt, omdat eruit volgt dat een kind dat een moeilijke opdracht correct heeft uitgevoerd geen fout kan maken op een meer gemakkelijke opdracht. Daarom wordt in de IRT een kansmodel gebruikt: hoe groter de vaardigheid, des te groter de kans dat een opdracht/observatiepunt juist wordt uitgevoerd. De moeilijkheidsgraad van een opdracht/observatiepunt wordt dan gedefinieerd als de mate van vaardigheid die nodig is om met een kans van precies een half de opdracht correct uit te voeren.

### **Aantonen veronderstellingen**

In de voorgaande tekst zijn nogal wat veronderstellingen gedaan (unidimensionaliteit, alle opdrachten of observatiepunten zijn indicatoren voor dezelfde vaardigheid, kansmodel) die niet zonder meer voor waar kunnen worden aangenomen; er moet aangetoond worden dat al die veronderstellingen deugdelijk zijn. Dit 'aantonen' gebeurt met statistische gereedschappen waar in de volgende paragraaf dieper op wordt ingegaan. Maar voor de opdrachten en observatiepunten in een meetinstrument gebruikt kunnen worden, moet ook geprobeerd worden de waarden van de moeilijkheidsgraden te achterhalen. Dit gebeurt met een statistische schattingsmethode die wordt toegepast op de responsen op de opdrachten en observatiepunten die bij een steekproef van leerlingen zijn verzameld. Het hele proces van moeilijkheidsgraden schatten en verifiëren of de modelveronderstellingen houdbaar zijn, wordt kalibratie of ijking genoemd.

Meestal bevat een bank meer opdrachten en observatiepunten dan een doorsnee meetinstrument, zodat het praktisch niet haalbaar is om alle opdrachten en observatiepunten aan alle kinderen voor te leggen. Elk kind in de steekproef van het kwaliteitsonderzoek krijgt derhalve slechts een (klein) gedeelte van de opdrachten en observatiepunten uit de bank voorgelegd. Dit gedeeltelijk voorleggen gebeurt aan de hand van een zogeheten 'onvolledig design'. Dit moet met de nodige voorzichtigheid gebeuren. In het vorige hoofdstuk is ingegaan op het afnamedesign dat voor het kwaliteitsonderzoek is gebruikt, de geïnteresseerde lezer wordt verwezen naar Eggen (1993).

Als de kalibratie met succes uitgevoerd is, is het resultaat een zogenoemde gekalibreerde bank. In dat proces worden de opdrachten/observatiepunten die niet passen bij de verzameling uit de collectie verwijderd (zie ook hoofdstuk 4). De bank bevat voor elke opdracht en elk observatiepunt niet alleen de feitelijke inhoud, maar ook de psychometrische eigenschappen, en de statistische zekerheid dat alle opdrachten en/of observatiepunten dezelfde vaardigheid aanspreken.

## **5.2 Meetmodel**

In het kwaliteitsonderzoek is gebruikgemaakt van zowel de klassieke testtheorie als een op de item-responstheorie (IRT) gebaseerd meetmodel, zoals dat bij Cito gebruikelijk is. IRT-modellen verschillen in een aantal opzichten van de klassieke testtheorie (Verhelst, 1993; Verhelst & Kleintjes, 1993; Verhelst & Glas, 1995). Bij de klassieke testtheorie staan het meetinstrument en de score centraal. Het theoretisch belangrijkste begrip in deze theorie is de zogeheten ware score, de gemiddelde score die de persoon zou behalen indien de meting een oneindig aantal keren onder dezelfde condities zou worden afgenomen. Die notie geeft een van de belangrijkste (praktische) obstakels van deze theorie voor ons onderzoek weer: het is problematisch om scores te vergelijken die verkregen zijn in een onvolledig design, zoals het geval is bij twee subdomeinen in de kindroute van Kleuter in beeld - Rekenen (Getalbegrip en Meten).

In de IRT staat het te meten begrip of de te meten eigenschap centraal. De IRT beschouwt het correct uitvoeren van de opdracht of het observatiepunt (vanaf hier de itemrespons genoemd) als een indicator voor de mate waarin die eigenschap aanwezig is. Het verband tussen eigenschap en itemrespons is van probabilistische aard en wordt weergegeven in de zogenoemde itemresponsfunctie. Die geeft aan hoe groot de kans is op een correcte respons op een opdracht of observatiepunt (vanaf hier item genoemd) als functie van de onderliggende eigenschap of vaardigheid. Formeler:  $X_i$  is de toevalsvariabele die de respons op item  $i$  voorstelt.  $X_i$  neemt de waarde 1 aan in geval van een correcte respons en 0 in geval van een foute respons. Als symbool voor de vaardigheid wordt  $\theta$  (theta) gekozen. De vaardigheid  $\theta$  is niet rechtstreeks observeerbaar. Dat zijn alleen de responsen op de opdrachten en observatiepunten. Dat is de reden waarom  $\theta$  een 'latente' variabele wordt genoemd<sup>2</sup>. De itemresponsfunctie  $f_i(\theta)$  is gedefinieerd als een conditionele kans:

$$f_i(\theta) = P(X_i = 1|\theta) \quad (5.1)$$

Een IRT-model is een speciale toepassing van (5.1) waarbij aan de functie  $f_i(\theta)$  een meer of minder specifieke functionele vorm wordt toegekend.

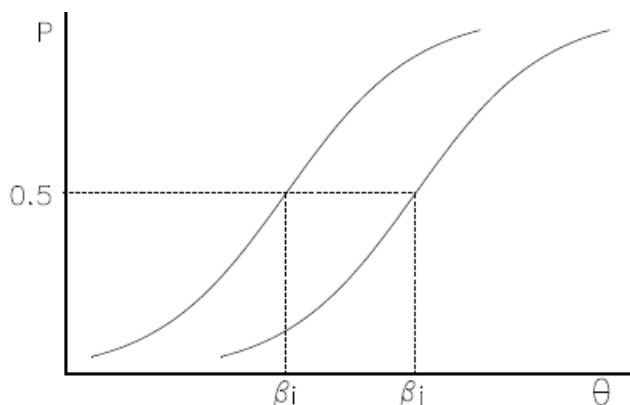
Een eenvoudig en zeer populair voorbeeld is het Raschmodel (Rasch, 1960) waarin  $f_i(\theta)$  gegeven is door

$$f_i(\theta) = \frac{\exp(\theta - \beta_i)}{1 + \exp(\theta - \beta_i)} \quad (5.2)$$

waarin  $\beta_i$  de moeilijkheidsparameter van item  $i$  is. Dat is een onbekende grootte die geschat wordt uit de observaties. De grafiek van (5.2) is weergegeven in Figuur 5.1 voor twee items,  $i$  en  $j$ , die in moeilijkheid verschillen. Deze figuur illustreert dat de itemresponsfunctie een stijgende functie is van  $\theta$ : hoe groter de vaardigheid, des te groter de kans op een juiste respons. Indien de latente vaardigheid precies gelijk is aan de moeilijkheidsparameter  $\beta_i$ , volgt

$$f_i(\theta) = \frac{\exp(\beta_i - \beta_i)}{1 + \exp(\beta_i - \beta_i)} = \frac{1}{1 + 1} = \frac{1}{2} \quad (5.3)$$

Figuur 5.1 Twee itemresponscurven in het Raschmodel



<sup>2</sup> Dit maakt duidelijk waarom men de modellen die vallen onder de IRT, ook wel aanduidt met 'latente trek'-modellen.

Formule (5.2) is geen beschrijving van de werkelijkheid, het is een hypothese over de werkelijkheid die getoetst kan worden op haar houdbaarheid. Hoe zo'n toetsing grofweg verloopt, is te verduidelijken aan de hand van Figuur 5.1. Daaruit blijkt dat, voor welk vaardigheidsniveau dan ook, de kans om item  $j$  juist te uit te voeren steeds kleiner is dan de kans op een juiste respons op item  $i$ . Hieruit volgt de statistisch te toetsen voorspelling dat de verwachte proportie juiste responsen op item  $j$  kleiner is dan op item  $i$  in een willekeurige steekproef van personen. Splitst men nu een grote steekproef in twee deelsteekproeven, een 'laaggroep', met de vijftig procent laagste scores, en een 'hooggroep', met de vijftig procent hoogste scores, dan kan men nagaan of de geobserveerde  $p$ -waarden van de opdrachten/observatiepunten in beide deelsteekproeven op dezelfde wijze geordend zijn. Daarvan kan strikt genomen alleen sprake zijn als, in termen van de klassieke testtheorie uitgedrukt, alle opdrachten/observatiepunten eenzelfde discriminatie-index hebben. Dat blijkt echter lang niet altijd zo te zijn, ook in dit geval niet. Veel van de items blijken dan ook niet beschreven te kunnen worden met het Raschmodel. Daarom is bij dit instrument gekozen voor een ander IRT-model.

Alvorens het hier gebruikte model te introduceren, is eerst een kanttekening nodig bij het schatten van de moeilijkheidsparameters in het Raschmodel. Een vaak toegepaste schattingsmethode is de 'conditionele grootste aannemelijkheidsmethode' (in het Engels: *Conditional Maximum Likelihood*, verder aangeduid als CML). Die maakt gebruik van het feit dat in het Raschmodel een afdoende steekproefgrootte (*sufficient statistic*) bestaat voor de latente variabele  $\theta$ , namelijk de ruwe score of het aantal correct uitgevoerde items. Dat betekent grofweg dat, indien de itemparameters bekend zijn, alle informatie die het responspatroon over de vaardigheid bevat, kan worden samengevat in de ruwe score; het doet er dan verder niet meer toe welke opdrachten/observatiepunten goed en welke fout zijn gemaakt. Hieruit vloeit voort dat de conditionele kans op een juiste respons op item  $i$ , gegeven de ruwe score, een functie is die alleen afhankelijk is van de itemparameters en onafhankelijk van de waarde van  $\theta$ . Een gedetailleerde uiteenzetting hierover kan men vinden in Verhelst, 1992. De CML-schattingsmethode maakt gebruik van deze functie. Deze methode maakt geen enkele veronderstelling over de verdeling van de vaardigheid in de populatie, en is ook onafhankelijk van de wijze waarop de steekproef is getrokken.

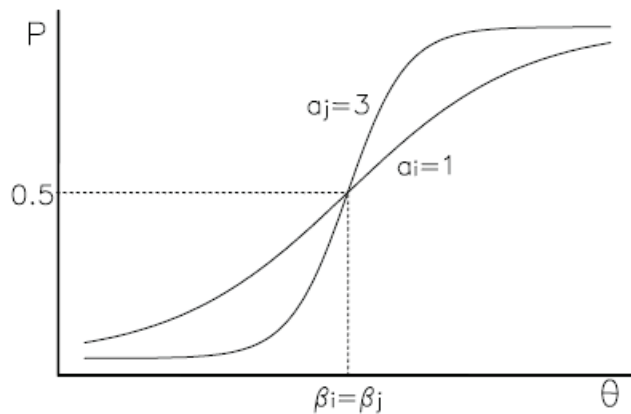
De CML-schattingsmethode is echter niet bij elk meetmodel toepasbaar. In het zogeheten éénparameter logistisch model (One Parameter Logistic Model, afgekort: OPLM) is CML mogelijk. Dit model is, anders dan het Raschmodel, wel bestand tegen 'omwisseling' van 'proporties juist' in verschillende steekproeven (Glas & Verhelst, 1993; Eggen, 1993; Verhelst & Kleintjes, 1993). De itemresponsfunctie van het OPLM is gegeven door

$$f_i(\theta) = \frac{\exp[\alpha_i(\theta - \beta_i)]}{1 + \exp[\alpha_i(\theta - \beta_i)]} \quad (5.4)$$

waarin  $\alpha_i$  de zogenoemde discriminatie-index van het item is. Door deze indices te beperken tot (positieve) gehele getallen, en door ze a priori als constanten in te voeren, is het mogelijk CML-schattingen van de itemparameters  $\beta_i$  te maken. In Figuur 5.2 is de itemresponscurve weergegeven van twee items  $i$  en  $j$ , die even moeilijk zijn maar verschillend discrimineren.



Figuur 5.2 Twee itemresponscurven in het OPLM: zelfde moeilijkheid, verschillende discriminatie



De schattingen worden berekend met het computerprogramma OPLM (Verhelst, Glas en Verstralen, 1995). Dit programma voert ook statistische toetsen uit op grond waarvan kan worden bepaald of het model de gegevens adequaat beschrijft. Omdat een aantal van deze toetsen bijzonder gevoelig is voor een verkeerde specificatie van de discriminatie-indices, zijn de uitkomsten van deze toetsen bruikbaar als modificatie-indices: ze geven een aanwijzing in welke richting deze discriminatie-indices moeten worden aangepast om een betere overeenkomst tussen model en gegevens te verkrijgen. Kalibratie van items volgens het OPLM is dan ook een iteratief proces waarin alternerend de modelfit van items wordt onderzocht door middel van statistische toetsing en de waarden van de discriminatie-indices worden aangepast op grond van de resultaten van deze toetsen. Deze aanpassingen geschieden in de praktijk op basis van een en hetzelfde gegevensbestand. Er kan dus kanskapitalisatie optreden. Indien een steekproef een voldoende grootte heeft, is het effect van deze kanskapitalisatie echter gering (Verhelst, Verstralen en Eggen, 1991).

### 5.3 Kalibratie

Met kalibratie wordt bedoeld dat we kengetallen zoeken bij de opdrachten/observatiepunten die de responsen van de leerlingen goed representeren. Hoe de kengetallen gezocht worden, ligt deels vast door het gekozen model (zie vorige paragraaf). Hoe succesvol deze operatie is, kan statistisch getoetst worden. In OPLM schatten we met de CML-methode de itemparameters en we controleren of deze parameters de data goed voorspellen. Voor een exacte beschrijving van de statistische toetsen die in OPLM gebruikt worden, hun eigenschappen en feitelijke implementatie in OPLM, verwijzen we naar Verhelst (1993). Hier beperken we ons tot een korte beschrijving van de principes van de statistische toetsen die gebruikt zijn in de kalibratieprocedure.

De statistische toetsen in OPLM hebben goede statistische en asymptotische eigenschappen, omdat OPLM behoort tot de exponentiële familie, met de gewogen somscore als een 'afdoende statistiek' (*sufficient statistic*) voor de vaardigheid:

$$S = \sum_{i=1}^k a_i x_i \tag{5.5}$$

Dit betekent dat alle informatie in de data met betrekking tot de vaardigheid in deze maat aanwezig is. Hiervan wordt gebruikgemaakt bij de statistische toetsen in OPLM. Het basisprincipe van de statistische toetsen in OPLM is dat op grond van de afdoende statistiek  $S$  de personen in de data kunnen worden gegroepeerd. En binnen deze groepen kan de verwachte proportie goede responsen op een item onder het model,  $p(+|s)$ , vergeleken worden met de feitelijk geobserveerde proportie goede responsen,  $prop(+|s)$ . Via de basisvergelijking van OPLM kunnen we eenvoudig de conditionele kans op het goed uitvoeren van de items afleiden en daarmee kunnen we  $p(+|s)$  evalueren,  $prop(+|s)$  volgt uit de data. Discrepancies

tussen  $p(+|s)$  en  $prop(+|s)$  duiden op schendingen van het model. Deze discrepanties vormen de basis voor de diverse statistische toetsen in OPLM. De toetsingsgrootheid voor de veronderstelde discriminatie-indices is gegeven door:

$$M = f_{s \in H}(p(+|s) - prop(+|s)) + f_{s \in L}(p(+|s) - prop(+|s)) \quad (5.6)$$

Deze zogeheten  $M$ -toetsen verdelen de scoregroepen in een laag deel ( $L$ ) en een hoog deel ( $H$ ) en  $f$  is een monotone functie.  $M$ -toetsen hebben een duidelijke interpretatie: is  $M$  significant positief dan is de veronderstelde steilheid van de ICC (in het Engels: *Item Characteristic Curve*, verder aangeduid als ICC) overschat in het model, is  $M$  daarentegen erg laag dan is de index te klein. In OPLM zijn drie verschillende  $M$ -toetsen geïmplementeerd die verschillen in de definitie van de hoge en lage scoregroepen.

Naast deze  $M$ -toetsen is er een algemene itemtoets die de volgende vorm heeft:

$$S = f(p(+|s) - prop(+|s)) \quad (5.7)$$

Deze zogeheten  $S$ -toets heeft een  $\chi^2$ -verdeling onder het model. Als globale modeltoets is de R1c-toets (Glas, 1988) geschikt. Ook de distributie van de rechteroverschrijdingskansen van alle afzonderlijke  $S$ -toetsen komt hiervoor in aanmerking. Als we deze  $S$ -toetsen opvatten als onafhankelijk, wat ze strikt genomen niet zijn, dan zouden de overschrijdingskansen uniform verdeeld moeten zijn over het  $(0,1)$  interval.

Kortom, als we afzien van de formeel-statistische achtergrond van de gehanteerde toetsen, kan de kalibratieprocedure als volgt worden samengevat:

- (1) Met behulp van het programma OPCAT worden de discriminatie-indices in OPLM ingesteld.
- (2) Vervolgens worden de itemparameters geschat met behulp van de CML-methode.
- (3) Met behulp van de  $M$ -toetsen wordt gecontroleerd of de discriminatie-indices goed zijn ingesteld.
- (4) Een volgende controle betreft de overschrijdingskansen van de  $S$ -toetsen en een grafische modelcontrole door middel van het programma WOPPLOT (grafische inspectie van de ICC's).
- (5) Vervolgens vindt een globale modelcontrole plaats in de vorm van een R1c-toets en de verdeling van de overschrijdingskansen van de  $S$ -toetsen.

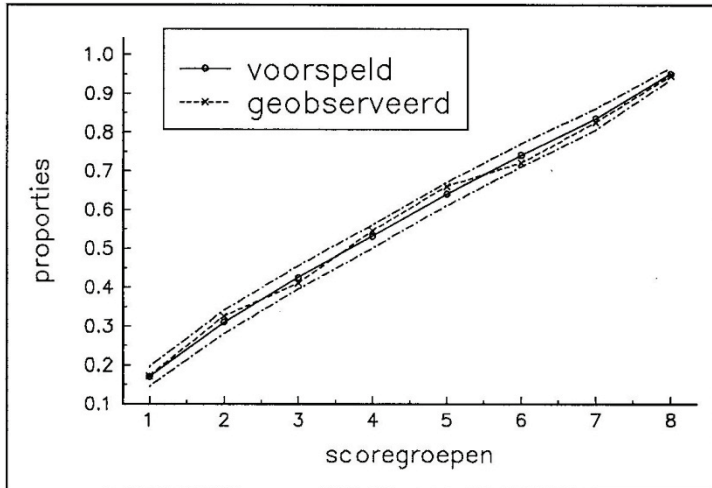
De stappen 1 tot en met 5 zijn een aantal malen doorlopen. Allereerst zijn de stappen doorlopen met alle observaties uit leerjaar 1 en leerjaar 2. Vervolgens zijn afhankelijk van inhoudelijke overwegingen en de uitkomsten opdrachten/observatiepunten verwijderd (zie ook hoofdstuk 4).

## 5.4 Toetsing van het IRT-model

Het is niet eenvoudig om de kwaliteit van de kalibratie aan te tonen. De belangrijkste statistische instrumenten om de passing van een item in het IRT-model te bewerkstelligen en uiteindelijk te documenteren betreffen de hierboven al besproken  $S$ -toetsen. Het lastige daarvan is, dat de toetsing voor een groot deel visueel gebeurt. Dit kunnen we illustreren aan de hand van Figuur 5.3 (zie Staphorsius, 1994, blz. 239). Figuur 5.3 beeldt voor een item de gegevens af waarop de betreffende  $S$ -toetsen gebaseerd zijn (zie handleiding OPLM: Verhelst; 1992). Ten behoeve van deze toetsing wordt de totale groep van leerlingen die een verzameling items gemaakt heeft, ingedeeld in een aantal (meestal acht) scoregroepen. Elke groep bestaat uit leerlingen met een ongeveer even hoge score. De geobserveerde proporties juiste responsen van deze groepen (telkens gesymboliseerd door een x) zijn door de middelste stippellijn verbonden. De volle lijn daarentegen verbindt de proporties die op grond van de parameterschattingen voorspeld kunnen worden. De twee buitenste lijnen geven het 95%-betrouwbaarheidsinterval aan. De breedte van dit interval is in belangrijke mate afhankelijk van het aantal leerlingen dat de opdracht/het observatiepunt heeft uitgevoerd.

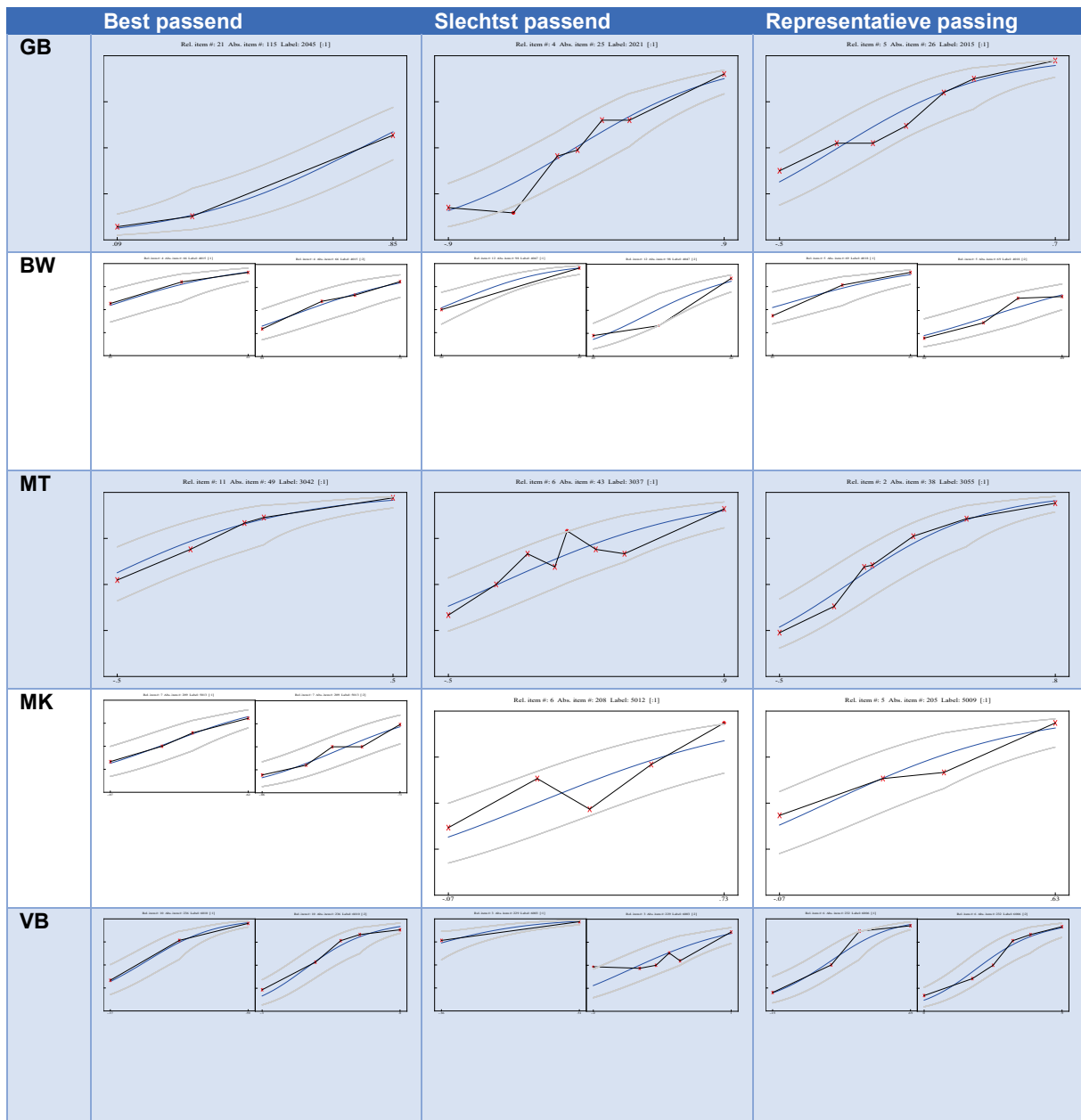
Uit de figuur blijkt heel duidelijk dat de geobserveerde proporties, zoals bedoeld, binnen het 95%-betrouwbaarheidsinterval van de (geschatte) voorspelde proporties liggen, en dit komt in grote lijnen overeen met een niet-significante S-toetsingsgrootheid (Verhelst, et al., 1995).

Figuur 5.3 Grafische voorstelling van een Si-toets



Het is ondoenlijk om voor alle opdrachten/observatiepunten dergelijke grafische voorstellingen in deze verantwoording op te nemen. Daarom beperken we ons steeds per subdomein tot de opdracht/het observatiepunt met de slechtste en de beste S-passing, aangevuld met een qua S-toetsingsresultaat representatieve passing. De voorbeelden in Figuur 5.4 illustreren dat zelfs bij de slechtst passende opdracht (of observatiepunt) sprake is van een redelijk aanvaardbaar beeld. Bij alle subdomeinen voldoen de meeste opdrachten/observatiepunten aan de eis dat de geobserveerde proportie binnen het 95%-betrouwbaarheidsinterval van de geschatte proporties ligt voor alle scoregroepen. De afbeeldingen voor de representatieve en best passende items illustreren dit. Dit leidt tot de conclusie dat bij vrijwel alle opdrachten en observatiepunten in Kleuter in beeld - Rekenen een grafische voorstelling van de S-toetsing hoort die in grote lijnen met Figuur 5.3 overeenkomt. Dit is een zeer sterke aanwijzing dat het meetinstrument en het meetmodel dat ontwikkeld is, respectievelijk gebruikt is, adequaat zijn om het gedrag van de leerlingen te beschrijven. Bovendien blijkt, en dat is vanuit theoretisch oogpunt nog belangrijker, dat gemeten verschillen in gedrag tussen de leerlingen te verklaren zijn door één unidimensioneel concept.

**Figuur 5.4** Voorbeelden van S-toetsen voor de subdomeinen in de kindroute met de best passende, de slechtst passende en een qua passing representatieve opdracht/observatiepunt



In feite kan men bij de kalibratie beter varen op deze grafische weergaven dan op toetsingsresultaten in termen van exacte getallen en de significantie daarvan. Niettemin zijn er bij de kalibratie S-toetsen uitgevoerd die een indicatie geven van de kwaliteit van de kalibratie. Daarbij zijn we vooral geïnteresseerd in de distributie van de overschrijdingskansen van deze verzameling toetsingsresultaten. Tabel 5.1 waarin het (0,1) interval is opgedeeld in tien gelijke stukken, geeft een beeld van de uitkomsten bij een kalibratie van alle opdrachten en observatiepunten van de subdomeinen van Kleuter in beeld - Rekenen. Daarnaast is aangegeven in hoeveel gevallen de overschrijdingskans kleiner was dan ,01, respectievelijk ,05. Het is duidelijk dat bij elk van de subdomeinen de verdeling redelijk gelijkmatig is over het gehele interval van overschrijdingskansen. Dit resultaat geeft een bevestiging van het eerder geschetste beeld, dat er met uitzondering van enkele opdrachten en/of observatiepunten, sprake is van niet-significante S-toetsen. Zij vormen een kwantitatieve ondersteuning van de conclusie dat de opdrachten en observatiepunten een unidimensioneel construct representeren van een subdomein.

Tabel 5.1 Verdeling van overschrijdingskansen bij S-toetsen voor de subdomeinen uit de Kindroute

GB	0.--/ 0/	---/ 0/	---.1 1	---.2 1	---.3 4	---.4 4	---.5 6	---.6 6	---.7 3	---.8 5	---.9 4	----1 12
BW	0.--/ 0/	---/ 0/	---.1 0	---.2 1	---.3 0	---.4 1	---.5 2	---.6 0	---.7 2	---.8 6	---.9 7	----1 12
MT	0.--/ 0/	---/ 1/	---.1 1	---.2 0	---.3 1	---.4 7	---.5 5	---.6 4	---.7 3	---.8 0	---.9 7	----1 12
MK	0.--/ 0/	---/ 0/	---.1 0	---.2 1	---.3 0	---.4 1	---.5 1	---.6 4	---.7 2	---.8 4	---.9 4	----1 13
VB	0.--/ 0/	---/ 4/	---.1 3	---.2 2	---.3 0	---.4 2	---.5 1	---.6 4	---.7 3	---.8 3	---.9 3	----1 9

In Tabel 5.2 zijn de R1c-waarden weergegeven. Voor de activiteiten Meetkunde en Bewerkingen waren onvoldoende waarnemingen (respectievelijk N = 169 en N = 159) om de R1C te berekenen. R1c is een statistiek die zicht geeft op de modelpassing van de toets als geheel. Voor een acceptabele modelfit geldt als vuistregel dat R1c niet groter dan ongeveer anderhalf maal het aantal vrijheidsgraden (*df*) zou moeten zijn en bij voorkeur niet significant. Voor de subdomeinen waarvoor de R1c uitgerekend kon worden, geldt dat R1c niet groter is dan ongeveer anderhalf maal het aantal vrijheidsgraden. De subdomeinen Getalbegrip en Meten zijn daarnaast niet significant. De modelpassing is dus acceptabel.

Tabel 5.2 R1c-waarden voor de subdomeinen

Subdomein	R1c		df		P	
<b>GB</b>	R1c* =	17,63	df =	24	p =	0,82
<b>MT</b>	R1c* =	19,94	df =	22	p =	0,59
<b>VB</b>	R1c* =	54,32	df =	33	p =	0,01

Ten slotte bespreken we nog een methode om de kwaliteit van de kalibratie te verantwoorden die wordt besproken in het COTAN Beoordelingssysteem (Evers, Lucassen, Meijer & Sijtsma, 2010, p. 40). Het betreft hier een poging om de nauwkeurigheid van de itemparameterschattingen te beoordelen op basis van een constante (in het COTAN Beoordelingssysteem met 'c' aangeduid) die weergeeft hoe de relatie is tussen de standaardfout van de moeilijkheidsparameter van een item en de standaarddeviatie van de vaardigheidsverdeling van de kalibratiepopulatie. Het beoordelingssysteem geeft ook richtlijnen voor het beoordelen van de grootte van deze 'c'. Deze dient te worden beoordeeld als goed als de waarde lager is dan of gelijk aan ,20. Waarden tussen ,30 en ,40 kunnen nog als voldoende worden beschouwd.

In Tabel 5.3 zijn het gemiddelde en de frequentieverdeling van deze waarden voor alle opdrachten en observatiepunten in de niveaueversies weergegeven. De gemiddelde waarde van de constante is voldoende tot goed in alle niveaueversies. Bij de grote meerderheid van alle itemparameters (86%) is de nauwkeurigheid voldoende tot goed (een  $c \leq ,4$ ).

Tabel 5.3 Nauwkeurigheid van de itemparameterschattingen (constante 'c')

	Groep	Versie	Gemiddelde	Goed		Voldoende		Onvoldoende
				$c \leq 0,2$	0,2-0,3	$0,3 \leq c \leq 0,4$	0,4-0,5	$c \geq 0,5$
<b>GB</b>	1	★	0,20	9	7	2	0	0
	2	★★	0,23	6	11	0	1	0
	2	★★★	0,23	6	11	0	0	1
<b>ME</b>	1	★	0,24	7	4	6	0	1
	2	★★	0,20	9	9	0	0	0
	2	★★★	0,18	11	7	0	0	0
<b>BW</b>	1	–	0,31	5	10	9	5	2
	2	–	0,28	6	13	9	1	2
<b>MK</b>	1	–	0,30	2	15	7	6	0
	2	–	0,25	8	14	8	0	0
<b>VB</b>	1	–	0,23	21	9	2	0	2
	2	–	0,27	14	14	3	1	2

Op basis van de hierboven beschreven resultaten kan geconcludeerd worden dat de kalibratie voor de kindroute van Kleuter in beeld - Rekenen geslaagd is. Hiermee is het laatste woord nog niet gezegd over de validiteit, maar het kalibratieonderzoek brengt in ieder geval een essentieel aspect van het validiteitsvraagstuk naar voren: de rechtvaardiging van wat in de meeste toepassingen gebruikelijk is, namelijk het reduceren van alles wat de leerling heeft uitgevoerd tot een enkele somscore op grond waarvan zijn of haar functioneringsniveau wordt geschat. De kalibratie-analyse, als puur formeel proces, kan geen uitspraken doen over de inhoudsvaliditeit of over de constructvaliditeit. In hoofdstuk 9 over validiteit zal worden nagegaan of de gemeten concepten inderdaad overeenkomen met het begrip zoals bedoeld. Een geslaagde kalibratie op een unidimensioneel construct beschouwen we als een noodzakelijke voorwaarde voor deze validiteit.

## 6 Standaardbepaling voor de kindroute

### 6.1 Doel van standaardbepaling

Met het observatie-instrument Kleuter in beeld kan de leerkracht nagaan hoe de kleuters ervoor staan, zodat de leerkracht het onderwijsaanbod daarop kan afstemmen. Het doel van de standaardbepaling is om met leerkrachten en andere onderwijsprofessionals na te gaan wat qua rekenen doorgaans gezien wordt bij kleuters midden groep 1 (M1), eind groep 1 (E1) en wat doorgaans gezien wordt bij kleuters eind groep 2 (E2). De standaard is bedoeld om vroegtijdig te kunnen signaleren of een kind mogelijk behoefte heeft aan extra aandacht of juist aan extra uitdaging.

Tijdens de standaardbepaling dienen voor Rekenen de grenzen bepaald te worden tussen zes functioneringsniveaus (<M1, M1, E1, M2, E2 en >E2). Tijdens de standaardbepaling worden dus uiteindelijk vijf standaarden bepaald.

### 6.2 Methode

Voor de standaardbepaling is de 3DC-methode gebruikt (Feskens, Keuning, Van Til, & Verheyen, 2014; Keuning, Straat & Feskens, 2017). Deze methode is al succesvol toegepast voor onder meer het bepalen van de standaarden voor het Europees Referentie Kader (Feskens, Keuning, Van Til, & Verheyen, 2014) en voor de Diagnostische Tussentijdse Toets (Schouwstra, 2018). Bij de 3DC-standaardbepaling wordt gewerkt met een team van experts op het vakgebied. Om tot goed gefundeerde standaarden te komen, worden de experts voorzien van verschillende informatiebronnen.

Op grond van de leerlijnen en de inhoud van het observatie-instrument bepalen onderwijsprofessionals samen de standaarden. Daarnaast wordt empirische informatie uit het kwaliteitsonderzoek gebruikt. De experts bekijken clusters van opdrachten en zetten per cluster een grens. De methode veronderstelt dus dat een instrument uit meerdere opdrachten bestaat die in te delen zijn in een aantal clusters. Kleuter in beeld - Rekenen bestaat uit vijf clusters, namelijk elk subdomein vormt een cluster: Getalbegrip, Bewerkingen, Meten, Meetkunde en Verbanden.

Kenmerkend aan de 3DC-methode is dat de standaarden worden gezet door te kijken naar de leerlijnen en de inhoud van het observatie-instrument en niet door leerlingen met elkaar te vergelijken.

### 6.3 Procedure en experts

Op vier verschillende dagen, twee dagen in juni en twee dagen in september, is een standaardbepaling georganiseerd. In maart startte de werving voor deelnemers aan de standaardbepaling. Per mail is een uitnodiging gestuurd aan de contactpersonen van de deelnemende scholen aan het proef- en kwaliteitsonderzoek, aan de mensen die in focusgroepen mee hebben gedacht over de concepten voor Kleuter in beeld en aan degenen die screeningswerkzaamheden hebben uitgevoerd voor dit instrument. Dit waren allemaal leerkrachten en intern begeleiders. Daarnaast is een mail gestuurd aan experts die hebben deelgenomen aan onze expertbijeenkomsten voor Kleuter in beeld en aan experts die bekend zijn uit de netwerken van Cito. Tot slot is ook een oproep op *social media* geplaatst met de vraag of onderwijsprofessionals wilden meewerken aan de standaardbepaling.

In totaal hebben 31 onderwijsprofessionals zich aangemeld, waarvan er uiteindelijk 30 ook mee hebben gedaan aan de vier standaardbepalingsdagen: 15 kleuterleerkrachten, 7 intern begeleiders, 1 kleuterleerkrachten die tevens intern begeleider is en 7 experts op het gebied van rekenen/het jonge kind (onderwijsadviseurs/ontwikkelaars/rekenexperts).

Wegens de coronacrisis heeft de standaardbepaling online plaatsgevonden. Voorafgaand aan elke standaardbepalingsdag kregen de deelnemers informatie over leerlijnen. In verband met de online standaardbepaling kregen de deelnemers ook de opdrachten en observatiepunten toegestuurd en het verzoek een oordeel te vormen hoe een leerling eind groep 1 de activiteiten en opdrachten zou doen. De standaardbepaling verliep via Teams en begon met een uitleg over het ontwikkelingsproces van Kleuter in beeld - Rekenen en een beschrijving van de inhoud van het instrument. Vervolgens werd het doel van de standaardbepaling en de methode uitgelegd. Daarna gingen de experts in Teams in drie groepjes uiteen om de eerste grens te bepalen. In elk groepje konden de experts onderling overleggen. Er was ook een toetsdeskundige van Cito aanwezig bij elk groepje om vragen te kunnen beantwoorden.

De eerste dag in juni zetten de deelnemers standaarden voor de clusters Meten, Meetkunde en Verbanden (7 experts) en de tweede dag voor de clusters Getalbegrip, Bewerkingen en Verbanden (7 experts). Na de zomer werd deze opzet herhaald: de eerste dag in september werden de standaarden gezet voor de clusters Meten, Meetkunde en Verbanden (7 experts) en de tweede dag voor de clusters Getalbegrip, Bewerkingen en Verbanden (9 experts).

Tijdens de sessie bekeken de experts voor drie subdomeinen een cluster met opdrachten of observatiepunten en vormden zij zich een oordeel over hoeveel opdrachten/observatiepunten het kind naar verwachting correct zou hebben op dat cluster als zijn/haar vaardigheid zich precies Onder E1 van Rekenen bevindt. De onderwijsprofessionals werd gevraagd de grensleerling voor ogen te houden. Bij de ondergrens is de grensleerling een leerling van wie je de ontwikkeling extra in de gaten wilt houden, omdat de leerling mogelijk extra aandacht nodig heeft.

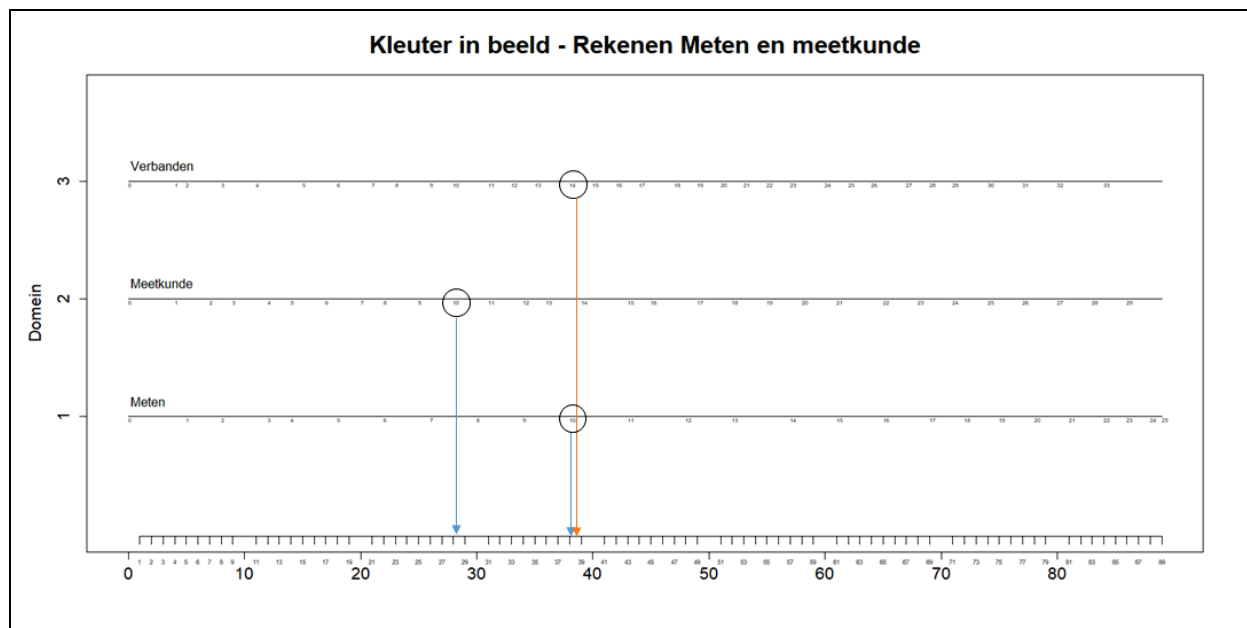
### ***Het standaardbepalingsformulier***

De experts gaven hun oordeel aan op een formulier (zie een fictief voorbeeld in Figuur 6.1). Op de horizontale lijnen van het formulier is het aantal correcte opdrachten/observatiepunten voor elk cluster te zien. Met behulp van de gegevens uit het kwaliteitsonderzoek is de moeilijkheid bepaald van de opdrachten/observatiepunten die gebruikt worden in de standaardbepaling. De moeilijkheid is bepaald door alle opdrachten en observatiepunten van Rekenen op één vaardigheidsschaal te leggen met behulp van het One Parameter Logistic Model (Verhelst, Glas, & Verstralen, 1995). Met behulp van simulaties is voor elk mogelijk aantal correcte opdrachten/observatiepunten in totaal uitgerekend wat het verwachte aantal correcte opdrachten/observatiepunten is op een subdomein (cluster). Deze uitkomsten zijn weergegeven op het standaardbepalingsformulier (Figuur 6.1). In de standaardbepaling voor Kleuter in beeld wordt het One Parameter Logistic Model alleen gebruikt voor de weergave op dit formulier.

Het aantal correcte opdrachten per cluster is gerelateerd aan het totaal aantal correcte opdrachten voor een dag (bijvoorbeeld de eerste dag met de drie clusters Meten, Meetkunde en Verbanden). Bijvoorbeeld, op het formulier in Figuur 6.1 is te zien dat tien correcte observatiepunten bij *Meten* te verwachten zouden zijn bij een leerling die in totaal ongeveer 38 correcte opdrachten/punten heeft en veertien correcte opdrachten bij *Verbanden*. Op dit formulier zien de experts dus de relatieve moeilijkheid van het cluster in verhouding tot de andere clusters. De visuele relatie tussen de drie subdomeinen helpt de experts een realistisch oordeel te geven.



Figuur 6.1 Een fictief standaardbepalingsformulier



### Oordelen van de experts

Nadat de experts hun oordelen hadden gegeven over het aantal correcte observatiepunten/opdrachten van de grensleerling, werden enkele experts uitgenodigd hun oordeel toe te lichten en volgde een discussie. Bij de eerste discussie werden experts uitgenodigd met oordelen aan de uiteinden van het beoordelings-spectrum en juist een expert die zich meer in het midden bevond. De discussie was niet gericht op het bewerkstelligen van consensus, maar diende te waarborgen dat de inhoud waarop het oordeel gevormd was op dezelfde manier geïnterpreteerd werd door alle experts en dat de experts dezelfde grensleerling voor ogen hadden. Gedurende de dag werd er voor gezorgd dat telkens andere experts als eerste werden uitgenodigd om er voor te zorgen dat alle meningen gehoord werden en de discussie niet gedomineerd kon worden door één expert. Soms kregen de experts ook feedback over de moeilijkheid van bepaalde opdrachten (indien daar een discussie over ontstond) of over de impact van hun oordelen op het percentage leerlingen dat in een bepaald functioneringsniveau valt. Tijdens de eerste discussie werd ruim de tijd genomen om te zorgen dat de experts eenzelfde beeld hadden van de grensleerling. Na de discussie kregen de experts de gelegenheid om hun definitieve oordeel op het formulier aan te geven.

Nadat de experts op alle clusters een definitief oordeel hebben gegeven over de grens Onder E1/E1, gingen ze door naar de tweede grens (E1/boven E1), de derde grens (onder E2/E2), de vierde grens (E2/boven E2) en ten slotte de laatste grens (onder M1/M1). Op elke dag gaven de onderwijsprofessionals dus voor drie subdomeinen een oordeel over de onder- en bovengrens eind leerjaar 1 (E1), een onder- en bovengrens voor eind leerjaar 2 (E2) en de ondergrens midden leerjaar 1 (M1).

## 6.4 Materiaal

Omdat het totaal aantal opdrachten en observatiepunten uit alle niveauversies van de uitgave te omvangrijk is, is een selectie gemaakt voor de standaardbepaling. Voor elke dag zijn drie subdomeinen gekozen. De eerste en derde dag zijn de subdomeinen Meten, Meetkunde en Verbanden behandeld en de tweede en vierde dag de subdomeinen Getalbegrip, Bewerkingen en Verbanden.

Voor de activiteiten (Bewerkingen, Meetkunde en Verbanden) zijn alle observatiepunten uit de uitgave voorgelegd. Bij Getalbegrip en Meten zijn 25 opdrachten geselecteerd binnen de inhoudelijke categorieën van de niveaueversies van de uitgave, zie Tabel 6.1.

Tabel 6.1 Aantal opdrachten per subdomein en categorie die geselecteerd zijn voor de standaardbepaling

Subdomein	Subcategorie	Niveaueversies uitgave		In standaardbepaling	
		Aantal	Percentage	Aantal	Percentage
<b>Getalbegrip</b>	Getallen	15	32,6%	8	32,0%
	Hoeveelheden	10	21,7%	5	20,0%
	Relaties	12	26,1%	7	28,0%
	Telrij	9	19,6%	5	20,0%
		46	100%	25	100%
<b>Meten</b>	Geld	7	17,1%	4	16,0%
	Gewicht	7	17,1%	4	16,0%
	Inhoud	9	22,0%	5	20,0%
	Lengte en omtrek	8	19,5%	5	20,0%
	Oppervlakte	4	9,8%	3	12,0%
	Tijd	6	14,6%	4	16,0%
		41	100%	25	100%

## 6.5 Definitieve expertoordelen

### **Meten, Meetkunde en Verbanden (dag 1 en 3)**

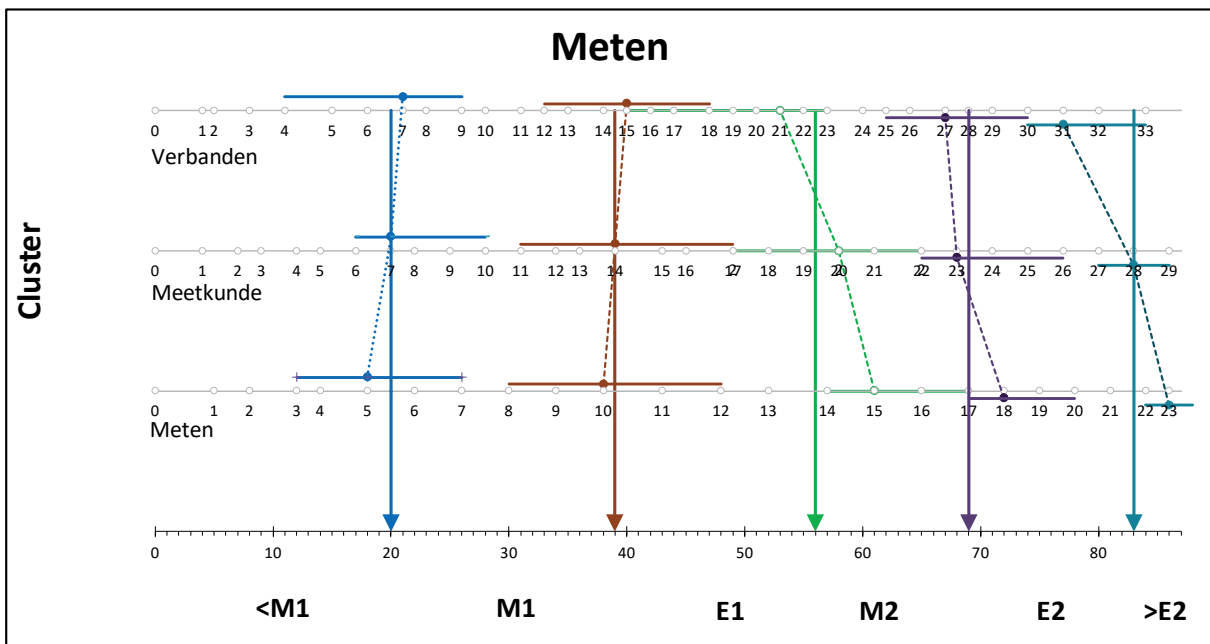
De oordelen van de experts per cluster (subdomein) op dag 1 en 3 (Meten, Meetkunde en Verbanden) zijn te zien in Tabel 6.3. De grens voor het subdomein ( $G_T$ ) per expert is gelijk aan de som van de expertoordelen per cluster ( $C_k$ ) en wordt dus niet direct door de experts zelf vastgelegd (zie de kolom "totaal" in Tabel 6.3):  $G_T = \sum C_k$ . In Figuur 6.2 is per subdomein het bereik van de oordelen per grens weergegeven en de gemiddelde grens over experts heen. We zien dat de oordelen van de experts niet ver uit elkaar lagen. Alleen bij het oordeel over Boven E1 bij Verbanden lagen de oordelen iets verder uit elkaar.

Bij het analyseren van de resultaten van de standaardbepaling is naar de impact van elke individuele expert op de grens van het hele domein gekeken en naar de samenhang tussen oordelen van de individuele expert en de oordelen van de overige experts. De impact van een individueel oordeel op de grens van het hele domein is bepaald door het absolute verschil te berekenen tussen de grens zoals bepaald ( $G_T$ ) en de grens zoals die zou zijn als de betreffende expert buiten beschouwing is gelaten ( $G'_T$ ):  $|G_T - G'_T|$ . Daarnaast is per expert de Ranking Similarity Index berekend (RSI, de gemiddelde correlatie tussen de oordelen van een expert met de oordelen van de overige experts). De overeenstemming tussen de expert-oordelen, ten slotte, is geëvalueerd met behulp van de gemiddelde interbeoordelaarscorrelatie ( $GIR$ ) en de Finn-coëfficiënt (Finn, 1970). In tegenstelling tot de gemiddelde interbeoordelaarscorrelatie is de Finn-coëfficiënt niet op de covariantie gebaseerd, maar op de ratio tussen de geobserveerde variantie en de variantie die je zou verwachten als de beoordelaars een random oordeel geven. Als beoordelaars systematisch strenger of milder zijn heeft dat invloed op de Finn coëfficiënt, maar niet op de interbeoordelaarscorrelatie. De Finn-coëfficiënt wordt vooral geschikt geacht als de variantie tussen beoordelaars gering is, dus als de overeenstemming hoog is (Tinsley & Weiss, 1975; Gamer, Lemon, & Singh, 2019). Omdat de clusters niet evenveel opdrachten/observatiepunten omvatten, zijn voor de analyse de grenzen op de clusters omgezet naar het corresponderende totaal. Bijvoorbeeld, als een expert de grens voor Meetkunde bij 7 zette is dat

voor de analyse omgezet in het corresponderende totaal van 20 (zie Figuur 6.2) en de grens van een andere expert bij 8 in het corresponderende totaal van 22.

Aan de Ranking Similarity Index is te zien dat voor elke expert de oordelen zeer goed samenhangen met de oordelen van de andere experts ( $RSI = 0,96 - 0,98$ ). De impact van de oordelen van de individuele experts op de domeingrenzen is gering. Als een individuele expert buiten beschouwing wordt gelaten, verandert een domeingrens in alle gevallen met minder dan een punt (op een schaal van 0 tot 90). De gemiddelde interbeoordelaarscorrelatie ( $GIR = 0,98$ ) en de Finn-coëfficiënt ( $Finn = 0,98$ ) laten zien dat de overeenstemming tussen de experts zeer goed is.

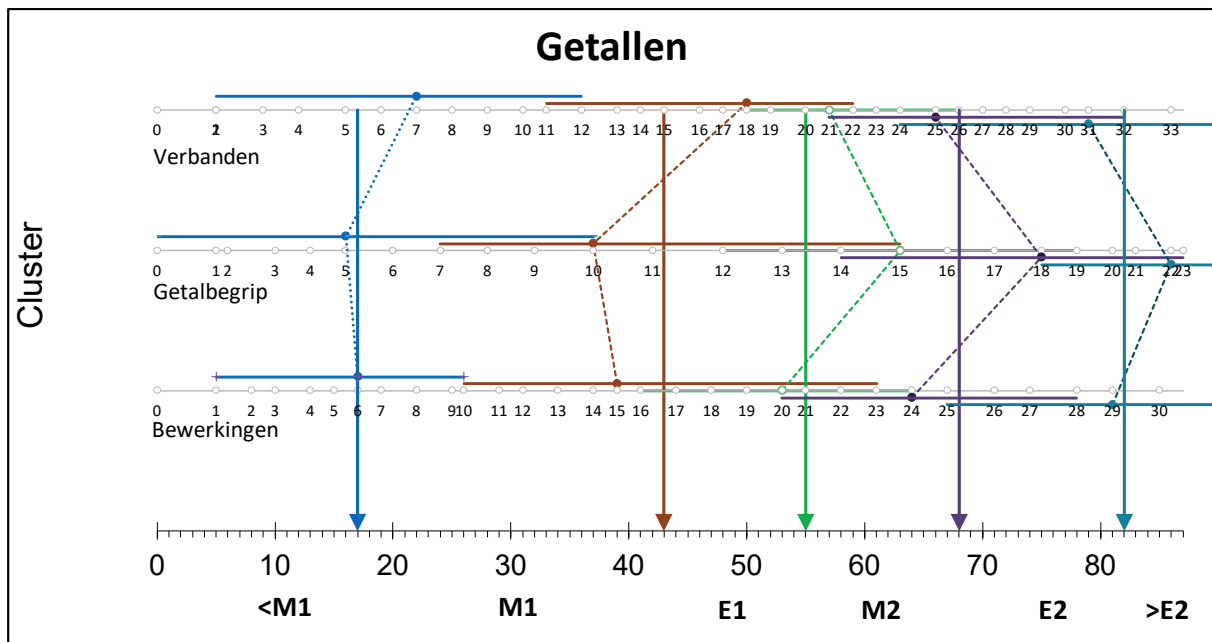
**Figuur 6.2** *Meten, Meetkunde en Verbanden: Plot van de oordelen per subdomein en gemiddelde oordeel op het totaal van dag 1 en 3 (Meten, Meetkunde en Verbanden)*



**Getalbegrip, Bewerkingen en Verbanden (dag 2 en 4)**

De oordelen van de experts per cluster op dag 2 en 4 (Getalbegrip, Bewerkingen en Verbanden) zijn te zien in Tabel 6.4. Te zien is dat de oordelen van de experts meer uiteenlopen dan bij Meten, Meetkunde en Verbanden het geval was. De oordelen lopen het meeste uiteen bij de Onder M1-grens van het subdomein Getalbegrip (zie Figuur 6.3). Aan de RSI, echter, is te zien dat voor elke expert de oordelen goed samenhangen met de oordelen van de andere experts ( $RSI = 0,87 - 0,95$ ). De impact van de oordelen van de individuele experts op de domeingrenzen is gering. Als een individuele expert buiten beschouwing wordt gelaten, verandert een domeingrens slechts in drie gevallen met 1 schaalpunt, in alle andere gevallen met minder dan 1 punt (op een schaal van 0 tot 89). De gemiddelde interbeoordelaarscorrelatie ( $GIR = 0,93$ ) en de Finn-coëfficiënt ( $Finn = 0,91$ ) laten zien dat de overeenstemming tussen de experts goed is.

Figuur 6.3 Getalbegrip, Bewerkingen en Verbanden: Plot van de oordelen per subdomein en gemiddelde oordeel op het totaal van dag 2 en 4 (Getalbegrip, Bewerkingen en Verbanden)



### Standaarden

De vijf definitieve standaarden voor Rekenen zijn berekend door de vijf grenzen van elke expert uit te drukken in een theta-schatting op de onderliggende vaardigheidsschaal. Vervolgens is per domeingrens de gemiddelde vaardigheid uitgerekend over alle experts als standaard. We zien dat de standaarddeviatie bij de onderste en bovenste standaard groter is dan de middelste drie standaarden. Nadat de standaarden zijn berekend, is voor leerjaar 1 en leerjaar 2 geschat welk percentage leerlingen in elk functioneringsniveau zou vallen (zie Tabel 6.2). Te zien is dat de bovengrens van leerjaar 1 vrij laag ligt, omdat 48% van de leerlingen midden leerjaar 1 in het functioneringsniveau Boven E1 zou vallen.

Tabel 6.2 Standaarden Rekenen en de verwachte impact (geschat percentage leerlingen dat onder M1, op M1, op E1, op M2, op E2 of boven E2 niveau valt, gegeven de standaarden)

Standaarden	<M1	M1	E1	M2	E2	>E2
Grens		-0,40	-0,05	0,12	0,30	0,73
Stdev		0,19	0,06	0,05	0,09	0,30
Leerjaar 1	4%	25%	23%	23%	23%	2%
Leerjaar 2	0%	4%	9%	18%	53%	16%
	<M1	M1	E1	>E1		
Leerjaar 1	4%	25%	23%	48%		
				<E2	E2	>E2
Leerjaar 2				31%	53%	16%

## 6.6 Conclusie

Onderwijsprofessionals hebben met de 3DC-methode de standaarden bepaald voor de grenzen tussen de zes functioneringsniveaus bij Rekenen. Deze standaarden worden gebruikt om het functioneringsniveau per subdomein te bepalen en te rapporteren op basis van de opdrachten van de niveaoversies en observatiepunten van de verschillende activiteiten van de kindroute. In hoofdstuk 7 wordt dit uitgebreid beschreven.

Gebruikmakend van hun ervaring hebben onderwijsprofessionals de standaarden gezet door te kijken naar de leerlijnen en de inhoud van Kleuter in beeld en niet door leerlingen met elkaar te vergelijken. Het proces is ondersteund door de empirische informatie. De onderwijsprofessionals bereikten een goede overeenstemming.

Standaarden en normen zijn aan slijtage onderhevig (Evers, Lucassen, Meijer & Sijsma, 2010). De standaarden zijn vastgesteld voor een termijn van tien jaar, mits de inhoud van het instrument en de impact van de standaarden niet aanmerkelijk veranderd zijn. Tussentijds (na drie en zeven jaar) zal derhalve een controle plaatsvinden van de inhoud en de impact.

Tabel 6.3 *Expertoordelen per subdomein voor Meten, Meetkunde, Verbanden*

Expert	Onder M1				Onder E1				Boven E1				Onder E2				Boven E2				Impact					RSI
	Meten	Meetkunde	Verbanden	Totaal	Meten	Meetkunde	Verbanden	Totaal	Meten	Meetkunde	Verbanden	Totaal	Meten	Meetkunde	Verbanden	Totaal	Meten	Meetkunde	Verbanden	Totaal	Onder M1	Onder E1	Boven E1	Onder E2	Boven E2	
1	5	7	8	20	9	12	13	34	16	21	21	58	20	26	27	73	24	29	31	84	0.0	0.4	0.1	0.3	0.1	0.97
2	6	8	9	23	8	12	12	32	16	22	21	59	17	23	27	67	23	28	32	83	0.2	0.5	0.2	0.1	0.0	0.97
3	7	8	9	24	11	13	15	39	15	19	22	56	18	22	25	65	24	27	30	81	0.3	0.0	0.0	0.3	0.1	0.98
4	6	8	9	23	11	13	17	41	14	19	22	55	18	22	26	66	23	27	30	80	0.2	0.2	0.1	0.2	0.2	0.98
5	6	8	8	22	10	13	14	37	14	20	21	55	19	22	27	68	23	27	32	82	0.2	0.2	0.1	0.0	0.0	0.98
6	5	10	8	23	8	11	15	34	15	21	23	59	18	23	27	68	22	28	32	82	0.2	0.4	0.2	0.0	0.0	0.97
7	5	8	8	21	10	13	15	38	16	17	23	56	20	22	27	69	24	28	31	83	0.1	0.1	0.0	0.0	0.0	0.97
8	6	7	6	19	12	12	13	37	17	21	15	53	19	22	26	67	24	29	33	86	0.1	0.2	0.2	0.1	0.3	0.96
9	5	6	4	15	11	15	12	38	16	20	23	59	18	26	27	71	23	29	32	84	0.4	0.1	0.2	0.2	0.1	0.97
10	5	6	7	18	11	16	18	45	14	21	22	57	18	25	27	70	22	29	31	82	0.1	0.5	0.1	0.1	0.0	0.97
11	4	7	8	19	12	17	17	46	14	19	20	53	18	22	30	70	23	28	33	84	0.1	0.5	0.2	0.1	0.1	0.96
12	5	8	7	20	10	15	15	40	14	21	21	56	17	24	26	67	23	28	30	81	0.0	0.1	0.0	0.1	0.1	0.98
13	4	7	7	18	11	13	18	42	14	20	22	56	17	23	28	68	22	28	32	82	0.1	0.2	0.0	0.0	0.0	0.98
14	3	6	5	14	11	15	17	43	14	20	20	54	18	26	27	71	22	29	30	81	0.5	0.3	0.2	0.2	0.1	0.97
Minimum	3	6	4	14	8	11	12	32	14	17	15	53	17	22	25	65	22	27	30	80						
Maximum	7	10	9	24	12	17	18	46	17	22	23	59	20	26	30	73	24	29	33	86						
Range	5	7	7	20	10	14	15	39	15	20	21	56	18	23	27	69	23	28	31	83						
Gemiddelde	5	8	8	23	11	13	15	34	14	21	21	56	18	22	27	67	23	28	32	82						
Maximum cluster	25	30	34	89	25	30	34	89	25	30	34	89	25	30	34	89	25	30	34	89						

Tabel 6.4 *Expertoordelen per subdomein voor Getalbegrip, Bewerkingen, Verbanden*

Expert	Onder M1				Onder E1				Boven E1				Onder E2				Boven E2				Impact					RSI
	Bewerkingen	Getalbegrip	Verbanden	Totaal	Bewerkingen	Getalbegrip	Verbanden	Totaal	Bewerkingen	Getalbegrip	Verbanden	Totaal	Bewerkingen	Getalbegrip	Verbanden	Totaal	Bewerkingen	Getalbegrip	Verbanden	Totaal	Onder M1	Onder E1	Boven E1	Onder E2	Boven E2	
1	9	8	9	26	18	12	17	47	21	14	21	56	23	16	25	64	28	20	30	78	0.6	0.3	0.0	0.2	0.3	0.95
2	6	4	9	19	15	7	17	39	16	12	19	47	22	16	24	62	30	21	31	82	0.1	0.3	0.6	0.4	0.0	0.93
3	3	5	3	11	15	11	15	41	21	12	22	55	23	14	23	60	29	21	33	83	0.4	0.1	0.0	0.5	0.1	0.93
4	5	2	7	14	10	8	16	34	18	14	18	50	23	16	22	61	29	22	31	82	0.2	0.6	0.4	0.4	0.0	0.94
5	5	2	6	13	22	10	18	50	24	15	20	59	28	17	28	73	30	21	33	84	0.3	0.5	0.2	0.4	0.1	0.92
6	10	6	12	28	20	11	20	51	17	12	20	49	25	15	22	62	27	19	27	73	0.7	0.5	0.4	0.4	0.6	0.91
7	6	3	4	13	16	10	17	43	22	14	20	56	24	17	25	66	27	23	30	80	0.3	0.0	0.0	0.1	0.1	0.95
8	10	10	12	32	14	15	20	49	17	17	20	54	28	23	24	75	30	23	32	85	1.0	0.4	0.1	0.5	0.2	0.88
9	1	0	1	2	11	8	17	36	21	19	23	63	24	20	31	75	27	24	33	84	1.0	0.5	0.5	0.5	0.1	0.93
10	5	5	7	17	13	14	11	38	22	16	20	58	23	19	22	64	29	23	29	81	0.0	0.3	0.2	0.2	0.1	0.91
11	4	4	7	15	23	8	22	53	21	14	19	54	23	18	21	62	25	18	24	67	0.1	0.7	0.1	0.4	1.0	0.87
12	6	4	4	14	18	9	18	45	23	12	21	56	25	18	23	66	29	24	30	83	0.2	0.1	0.0	0.1	0.1	0.93
13	5	5	8	18	10	10	18	38	21	17	23	61	20	23	28	71	31	25	34	90	0.1	0.3	0.4	0.2	0.5	0.93
14	7	7	7	21	10	10	18	38	17	16	26	59	27	22	32	81	31	25	34	90	0.3	0.3	0.2	0.9	0.5	0.92
15	4	4	6	14	13	9	18	40	22	15	20	57	25	18	24	67	31	25	33	89	0.2	0.2	0.1	0.0	0.5	0.95
16	5	4	7	16	14	14	18	46	19	14	20	53	25	22	27	74	26	24	33	83	0.1	0.2	0.2	0.4	0.1	0.92
Minimum	1	0	1	2	10	7	11	34	16	12	18	47	20	14	21	60	25	18	24	67						
Maximum	10	10	12	32	23	15	22	53	24	19	26	63	28	23	32	81	31	25	34	90						
Range	9	10	11	30	13	8	11	19	8	7	8	16	8	9	11	21	6	7	10	23						
Gemiddelde	6	5	7	17	15	10	18	43	20	15	21	55	24	18	25	68	29	22	31	82						
Maximum cluster	31	25	34	90	31	25	34	90	31	25	34	90	31	25	34	90	31	25	34	90						





## 7 Beschrijving algoritmes

Voor Kleuter in beeld - Rekenen wordt op grond van de observaties het functioneringsniveau van kleuters op subdomeinen gerapporteerd en een signaal gegeven over het hoofddomein Rekenen. Dit hoofdstuk beschrijft de algoritmes waarmee het functioneringsniveau op subdomeinen wordt gerapporteerd. Het beschrijft ook het algoritme waarmee het signaal op het hoofddomein wordt gegeven.

Een algoritme is een beperkte set regels die van een begintoestand (in ons geval observaties) naar een beoogd doel leidt (de rapportage van functioneringsniveaus en signalering). De implementatie van die regels gebeurt in de online omgeving van Kleuter in beeld.

Tabel 7.1 Aantal observaties per subdomein in de leerkrachtroute en aantal observatiepunten/-opdrachten in de kindroute

Subdomein	Leerkracht-route		Kindroute			
		aantal (indirecte) observaties	aantal observatiepunten	aantal opdrachten per versie		
				Groep 1	Midden Groep 2	Eind Groep 2
<b>Getalbegrip</b>	(GB)	4		18	18	18
<b>Bewerkingen</b>	(BW)	2	17			
<b>Meten</b>	(MT)	6		18	18	18
<b>Meetkunde</b>	(MK)	3	21			
<b>Verbanden</b>	(VB)	1	17			

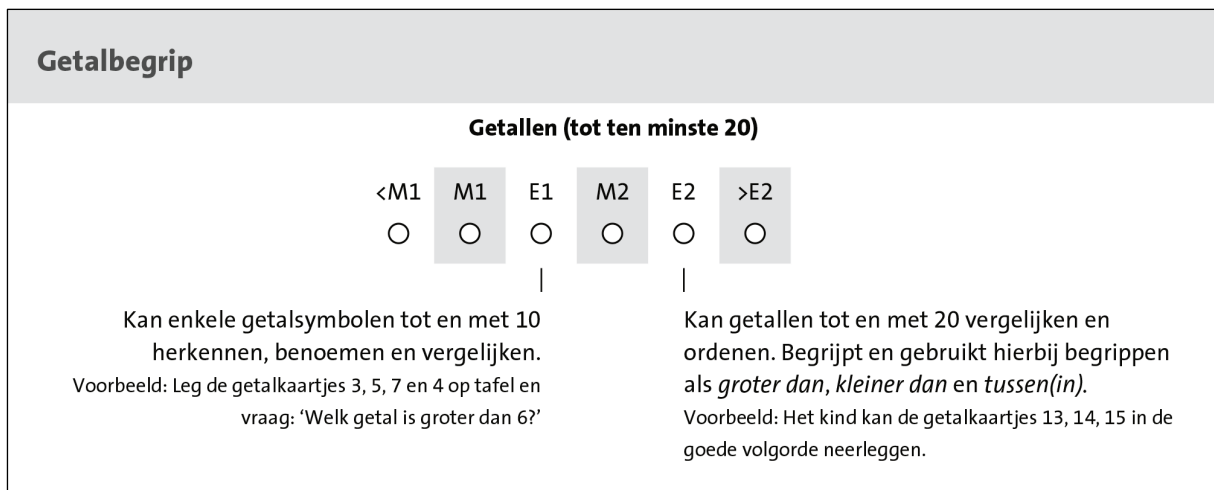
### 7.1 Algoritme voor het rapporteren van het functioneringsniveau op subdomeinen

#### Leerkrachtroute

In de leerkrachtroute voert de leerkracht per subdomein een of meerdere (indirecte) observaties uit en vult per observatie een oordeel in op een zespuntsschaal (zie Figuur 7.1). De oordelen worden gecodeerd van 0 tot 5. In Tabel 7.1 staat het aantal observaties dat de leerkracht per subdomein uitvoert in de leerkracht-route.

- Bij een oneven aantal observaties is het gerapporteerde functioneringsniveau gelijk aan het gemiddelde van de leerkrachtoordelen afgerond op een geheel getal. Als sprake is van één observatie voor een subdomein, zoals bij Verbanden, is het gerapporteerde functioneringsniveau dus gelijk aan dat ene leerkrachtoordeel.
- Bij een even aantal oordelen kan een kleuter exact tussen twee functioneringsniveaus in zitten. Indien dit het geval is, wordt het lagere functioneringsniveau gerapporteerd, zie het voorbeeld in Tabel 7.2. Toelichting: Het subdomein Bewerkingen bestaat uit twee onderdelen: Optellen en aftrekken met hele getallen en Vermenigvuldigen met hele getallen. Stel, in observatieperiode E1 schat de leerkracht het kind bij Vermenigvuldigen met hele getallen in op E1 en bij Optellen en aftrekken met hele getallen op M2. Als deze scores naar boven afgerond zouden worden, zou de algehele score op dit subdomein M2 worden. Dat zou onterecht de indruk kunnen wekken dat het kind voorloopt: het kind heeft immers nog niet alle doelen van dit subdomein voor M2 behaald. Daarom is gekozen voor afronding naar beneden.

Figuur 7.1 Voorbeeld van de zespuntsschaal in de leerkrachtroute



**Algoritme 1 Bepalen van het functioneringsniveau per subdomein op basis van de leerkrachtoordelen van de observaties in de leerkrachtroute**

Het functioneringsniveau ( $f$ ) is het gemiddelde van de oordelen ( $\bar{o}$ ) afgerond op een geheel getal. Het cijfer wordt naar beneden afgerond ( $\lfloor \bar{o} \rfloor$ ) of naar boven afgerond ( $\lceil \bar{o} \rceil$ ).

$$\lfloor \bar{o} \rfloor \in \mathbb{Z} \text{ en } \bar{o} - 1 < \lfloor \bar{o} \rfloor \leq \bar{o}$$

$$\lceil \bar{o} \rceil \in \mathbb{Z} \text{ en } \bar{o} < \lceil \bar{o} \rceil \leq \bar{o} + 1$$

Als het verschil tussen het gemiddelde en het gemiddelde afgerond naar beneden groter dan 0,5 is, dan is het functioneringsniveau het gemiddelde afgerond naar boven, anders is het functioneringsniveau het gemiddelde afgerond naar beneden.

$$\text{Als } \bar{o} - \lfloor \bar{o} \rfloor > 0,5 \text{ dan } f = \lceil \bar{o} \rceil \text{ anders } f = \lfloor \bar{o} \rfloor$$

Tabel 7.2 Voorbeeld van twee observaties en het gerapporteerde functioneringsniveau

Observatie 1		Observatie 2		Gerapporteerd functioneringsniveau op subdomein	
<M1	(0)	<M1	(0)	<M1	(0)
<M1	(0)	M1	(1)	<M1	(0)
<M1	(0)	E1	(2)	M1	(1)
<M1	(0)	M2	(3)	M1	(1)
<M1	(0)	E2	(4)	E1	(2)
<M1	(0)	>E2	(5)	E1	(2)
M1	(1)	<M1	(0)	<M1	(0)
M1	(1)	M1	(1)	M1	(1)
M1	(1)	E1	(2)	M1	(1)
M1	(1)	M2	(3)	E1	(2)
M1	(1)	E2	(4)	E1	(2)
M1	(1)	>E2	(5)	M2	(3)
E1	(2)	<M1	(0)	M1	(1)
E1	(2)	M1	(1)	M1	(1)
E1	(2)	E1	(2)	E1	(2)
E1	(2)	M2	(3)	E1	(2)
E1	(2)	E2	(4)	M2	(3)
E1	(2)	>E2	(5)	M2	(3)
M2	(3)	<M1	(0)	M1	(1)
M2	(3)	M1	(1)	E1	(2)
M2	(3)	E1	(2)	E1	(2)
M2	(3)	M2	(3)	M2	(3)
M2	(3)	E2	(4)	M2	(3)
M2	(3)	>E2	(5)	E2	(4)
E2	(4)	<M1	(0)	E1	(2)
E2	(4)	M1	(1)	E1	(2)
E2	(4)	E1	(2)	M2	(3)
E2	(4)	M2	(3)	M2	(3)
E2	(4)	E2	(4)	E2	(4)
E2	(4)	>E2	(5)	E2	(4)
>E2	(5)	<M1	(0)	E1	(2)
>E2	(5)	M1	(1)	M2	(3)
>E2	(5)	E1	(2)	M2	(3)
>E2	(5)	M2	(3)	E2	(4)
>E2	(5)	E2	(4)	E2	(4)
>E2	(5)	>E2	(5)	>E2	(5)

### Kindroute

De Kindroute bestaat uit vijf subdomeinen. Bij drie subdomeinen wordt het functioneren van een individuele kleuter geobserveerd met behulp van activiteiten en bij twee subdomeinen met behulp van opdrachten. Tijdens de standaardbepaling zijn door experts de grenzen bepaald tussen de zes functioneringsniveaus voor Rekenen. Deze grenzen worden gebruikt om per subdomein de functioneringsniveaus te bepalen.

Elke grens die de experts hebben gezet, wordt uitgedrukt in het aantal punten (opdrachten/observatiepunten) dat de kleuter beheerst als zijn/haar vaardigheid op het domein precies op de grens ligt (zie ook hoofdstuk 6). Deze punten kunnen voor elke expert ook uitgedrukt worden in bijbehorende vaardigheidsscores. Vervolgens is per domeingrens de gemiddelde vaardigheid uitgerekend over alle experts als standaard (gerapporteerd in Tabel 6.2). Voor de activiteiten en voor de opdrachten in de niveauseries (uit de uitgave) is een omzettingstabel gemaakt van het aantal correcte opdrachten/observatiepunten (de ongewogen scores) naar de vaardigheidsscore. Vervolgens wordt bij de vaardigheidsscores die bij de gezette grenzen horen, het bijbehorende correcte aantal opdrachten voor elk subdomein en elke niveauserie in de uitgave gezocht, zie Tabel 7.3.

Tabel 7.3 Grenzen uit de standaardbepaling en op de subdomeinen en niveauseries in de uitgave

Grens	Versie	Max	<M1 – M1	M1 - E1	E1 - M2	M2 - E2	E2 - >E2
<b>Vaardigheidsscore</b>			-0,40	-0,05	0,12	0,30	0,73
<b>Punten subdomeinen:</b>							
<b>Getalbegrip (GB)</b>	★	18	5,5	10,7	12,9	14,9	17,2
	★★	18	3,5	7,5	9,9	12,3	16,1
	★★★	18	3,1	6,4	8,4	10,6	15,2
<b>Bewerkingen (BW)</b>		31	6,2	16,7	22,3	26,4	30,3
<b>Meten (MT)</b>	★	18	4,8	9,9	12,5	14,6	17,3
	★★	18	2,7	6,9	9,6	12,3	16,5
	★★★	18	2,2	5,9	8,3	10,9	15,5
<b>Meetkunde (MK)</b>		30	5,7	14,1	18,9	23,1	28,4
<b>Verbanden (VB)</b>		34	5,4	14,8	21,6	27,5	33,2

**Algoritme 2 Bepalen van het functioneringsniveau per subdomein op basis van de observatiepunten uit de verschillende activiteiten en de opdrachten uit de niveauseries van de kindroute**

Er zijn zes mogelijke gerapporteerde functioneringsniveaus: <M1 ( $f_{1ij}$ ), M1 ( $f_{2ij}$ ), E1 ( $f_{3ij}$ ), M2 ( $f_{4ij}$ ), E2 ( $f_{5ij}$ ), en >E2 ( $f_{6ij}$ ). Op grond van de domeingrenzen die door de experts zijn bepaald tijdens de standaardbepaling, zijn per uitgave  $j$  van elk subdomein  $i$  vijf subdomeingrenzen ( $g_{1ij}$ ,  $g_{2ij}$ ,  $g_{3ij}$ ,  $g_{4ij}$ ,  $g_{5ij}$ ) uitgerekend, waarbij geldt dat  $g_{1ij} < g_{2ij} < g_{3ij} < g_{4ij} < g_{5ij}$ .

- Als de kleuter de observatiepunten/opdrachten ( $\sum o_i$ ) in mindere mate beheerst dan de eerste grens ( $g_{1i}$ ) dan is het gerapporteerde functioneringsniveau <M1 ( $f_{1i}$ ).  
IF ( $\sum o_i < g_{1i}$ ) THEN  $f_{1i}$
- Als de kleuter de observatiepunten/opdrachten ( $\sum o_i$ ) in mindere mate beheerst dan de tweede grens en evenveel of in meerdere mate dan de eerste grens dan is het gerapporteerde functioneringsniveau M1.  
IF ( $\sum o_i \geq g_{1i} \wedge \sum o_i < g_{2i}$ ) THEN  $f_{2i}$
- Als de kleuter de observatiepunten/opdrachten ( $\sum o_i$ ) in mindere mate beheerst dan de derde grens en evenveel of meer dan de tweede grens dan is het gerapporteerde functioneringsniveau E1.  
IF ( $\sum o_i \geq g_{2i} \wedge \sum o_i < g_{3i}$ ) THEN  $f_{3i}$
- Als de kleuter de observatiepunten/opdrachten ( $\sum o_i$ ) in mindere mate beheerst dan de vierde grens en evenveel of meer dan de derde grens dan is het gerapporteerde functioneringsniveau M2.  
IF ( $\sum o_{ij} \geq g_{3i} \wedge \sum o_i < g_{4i}$ ) THEN  $f_{4i}$

- Als de kleuter de observatiepunten/opdrachten ( $\sum o_i$ ) in mindere mate beheerst dan de vijfde grens en evenveel of meer dan de vierde grens dan is het gerapporteerde functioneringsniveau E2.  
IF ( $\sum o_{ij} \geq g_{4i} \wedge \sum o_i < g_{5i}$ ) THEN  $f_{5i}$
- Als de kleuter de observatiepunten/opdrachten ( $\sum o_i$ ) in meerdere mate beheerst dan de vijfde grens dan is het gerapporteerde functioneringsniveau >E2.  
IF ( $\sum o_i \geq g_{5i}$ ) THEN  $f_{6i}$

In Tabel 7.4 staan de grenzen weergegeven die zijn toegepast bij de uitgave van 2020.

Tabel 7.4 *Bereik aantal correcte observatiepunten/opdrachten en bijbehorend gerapporteerd functioneringsniveau*

Subdomein	Versie	<M1 ( $f_{1i}$ )	M1 ( $f_{2i}$ )	E1 ( $f_{3i}$ )	M2 ( $f_{4i}$ )	E2 ( $f_{5i}$ )	>E2 ( $f_{6i}$ )
<b>Getalbegrip</b>	★	0-5	6-10	11-12	13-14	15-17	18
	★★	0-3	4-7	8-9	10-12	13-16	17-18
	★★★	0-3	4-6	7-8	9-10	11-15	16-18
<b>Bewerkingen</b>		0-6	7-16	17-22	23-26	27-30	31
<b>Meten</b>	★	0-4	5-9	10-12	13-14	15-17	18
	★★	0-2	3-6	7-9	10-12	13-16	17-18
	★★★	0-2	3-5	6-8	9-10	11-15	16-18
<b>Meetkunde</b>		0-5	6-14	15-18	19-23	24-28	29-30
<b>Verbanden</b>		0-5	6-14	15-21	22-27	28-33	34

## 7.2 Algoritme voor een signaal voor Rekenen

Het onderwijsveld gaf aan dat het, naast het niveau voor elk van de subdomeinen, ook een signaal voor het hoofddomein Rekenen wil om zicht te krijgen op hoe elke kleuter het doet op Rekenen totaal. Er kunnen drie signalen gegeven worden: een signaal “wees alert”, een “ga zo door” of een “uitstekend”. Het signaal is afhankelijk van het aantal subdomeinen dat boven of onder het verwachte niveau ligt, gegeven de periode dat de kleuter is geobserveerd.

Het signaal “uitstekend” is bedoeld voor de rekensterke kinderen. Dit signaal wordt gegeven wanneer vier of meer subdomeinen boven verwachting zijn. Het signaal “wees alert” is bedoeld voor het signaleren van de (relatief) rekenzwakke kinderen: kinderen met mogelijke rekenproblemen en/of risico op latere rekenproblemen. Het is belangrijk een signaal te geven aan de leerkracht als een kind uitvalt op minimaal drie subdomeinen. Tijdig signaleren van deze kinderen is belangrijk, zodat nader bekeken kan worden of deze kinderen extra hulp nodig hebben en ernstige problemen op het gebied van rekenen zoveel mogelijk voorkomen kunnen worden.

### Algoritme 3 Bepalen van een signaal voor het hoofddomein Rekenen op basis van de functioneringsniveaus van alle subdomeinen

Er zijn in totaal vijf subdomeinen (A, B, C, D, E). Per subdomein is bepaald welk functioneringsniveau ( $f$ ) het kind heeft: 0 (<M1); 1 (M1); 2 (E1); 3 (M2); 4 (E2); 5 (>E2).

Er kunnen drie signalen ( $s$ ) gegeven worden: een “wees alert” (-1), een “ga zo door” (0) en “uitstekend” signaal (1). Het signaal is afhankelijk van het aantal subdomeinen met een bepaald functioneringsniveau ( $f$ ) dat boven of onder het verwachte niveau ( $v$ ) ligt gegeven de periode waarin het kind geobserveerd is ( $p$ ).

Per observatieperiode is een bepaald verwacht functioneringsniveau ( $v$ ):

- Begin leerjaar 1 ( $p = 1$ ) is het verwachte niveau <M1 ( $v = 0$ )
- Medio leerjaar 1 ( $p = 2$ ) is het verwachte niveau M1 ( $v = 1$ )
- Eind leerjaar 1 ( $p = 3$ ) en Begin leerjaar 2 ( $p=4$ ) is het verwachte niveau E1 ( $v = 2$ )
- Medio leerjaar 2 ( $p = 5$ ) is het verwachte niveau M2 ( $v = 3$ )
- Eind leerjaar 2 ( $p = 6$ ) is het verwachte niveau E2 ( $v = 4$ )

**Regel:**

Als in de observatieperiode  $p$  met een verwacht niveau  $v_0$

- vier of meer subdomeinen (van de vijf) een functioneringsniveau boven verwacht niveau hebben, dan krijgt het kind het signaal "uitstekend".
- drie of meer subdomeinen (van de vijf) een functioneringsniveau onder verwacht niveau hebben, dan krijgt het kind het signaal "wees alert".
- het kind geen signaal "uitstekend" en geen signaal "wees alert" krijgt, dan krijgt het kind het signaal "ga zo door".

## 8 Betrouwbaarheid en nauwkeurigheid

### 8.1 Interbeoordelaarsbetrouwbaarheid activiteiten

Om de betrouwbaarheid van de directe observaties bij de activiteiten in de kindroute te onderzoeken is de interbeoordelaarsbetrouwbaarheid bepaald. Scholen is gevraagd of een van de vijf Cito-experts met ervaring voor de klas, mocht langskomen terwijl de leerkracht een activiteit met een kind uitvoerde. Zowel de leerkracht als de Cito-expert vulden steeds hetzelfde observatieformulier in. In totaal hebben 8 scholen meegedaan aan dit betrouwbaarheidsonderzoek.

Tabel 8.1 Aantal kinderen dat is geobserveerd door een leerkracht en een Cito-expert

Activiteit	Groep 1	Groep 2	Totaal
<b>Bewerkingen</b>	13	28	41
<b>Meetskunde</b>	19	23	42
<b>Verbanden</b>	16	23	39
	48	74	122

Om de interbeoordelaarsbetrouwbaarheid te berekenen, is de correlatie tussen het leerkrachtoordeel en het oordeel van de Cito-expert berekend. Hiervoor is per activiteit de som van de oordelen berekend van de observatiepunten uit de uitgave. Daarnaast hebben we de absolute G-coëfficiënt uitgerekend in een generaliseerbaarheidsstudie (gstudy in R). De absolute G-coëfficiënt laat zien of er naast relatieve overeenstemming sprake is van systematische verschillen. Bij alle drie subdomeinen is de interbeoordelaarsbetrouwbaarheid en -overeenstemming goed (zie Tabel 8.2), bij Meetskunde en Verbanden zelfs uitstekend.

Tabel 8.2 Correlatie tussen de observaties van de leerkracht en de observaties van een Cito-expert

	Correlatie	g	N items	Maximum schaal	N leerlingen
<b>Bewerkingen</b>	0,85	0,84	17	31	41
<b>Meetskunde</b>	0,99	0,98	21	30	42
<b>Verbanden</b>	0,98	0,97	17	34	39

Vervolgens is de standaard toegepast op de leerkrachtoordelen en de expertoordelen om het functioneringsniveau te bepalen. In Tabel 8.3, Tabel 8.4 en Tabel 8.5 is de overeenstemming tussen het functioneringsniveau volgens het leerkrachtoordeel en volgens het Cito-expertoordeel te zien.

De absolute overeenstemming (percentage leerlingen dat volgens de oordelen van de leerkracht en de expert in hetzelfde functioneringsniveau valt) is met waarden tussen de 0,74 en 0,86 goed te noemen (zie Tabel 8.6). In de overeenstemmingstabellen 8.3, 8.4 en 8.5 wordt de conditionele verdeling gegeven (elke rij sommeert tot 100%). De absolute overeenstemming wordt berekend op de simultane verdeling (alle cellen samen sommeren tot 100%). Daarnaast is de gewogen Cohen's kappa uitgerekend met lineaire gewichten. Bij de gewogen Cohen's kappa wordt rekening gehouden met toevallige overeenstemming en met partiële overeenstemming. Volgens de vuistregel van Landis en Koch (1977) kunnen waarden van 0 – 0,2 als onvoldoende worden beschouwd; 0,21 – 0,40 als matig; 0,41 – 0,60 als redelijk; 0,61 – 0,80 als

goed en 0,81 – 1,00 als uitstekend. Volgens die vuistregel kan de overeenstemming dus als goed tot uitstekend worden beschouwd.

Tabel 8.3 Overeenstemmingstabel voor Bewerkingen

Functioneringsniveau op grond van leerkrachtoordeel						
Cito-expertoordeel	1 (<M1)	2 (M1)	3 (E1)	4 (M2)	5 (E2)	6 (>E2)
1 (<M1)	100.0	0.0	0.0	0.0	0.0	0.0
2 (M1)	0.0	100.0	0.0	0.0	0.0	0.0
3 (E1)	0.0	0.0	66.7	33.3	0.0	0.0
4 (M2)	0.0	0.0	15.4	84.6	0.0	0.0
5 (E2)	0.0	0.0	4.3	17.4	78.3	0.0
6 (>E2)	0.0	0.0	0.0	0.0	0.0	0.0

Tabel 8.4 Overeenstemmingstabel voor Meetkunde

Functioneringsniveau op grond van leerkrachtoordeel						
Cito-expertoordeel	1 (<M1)	2 (M1)	3 (E1)	4 (M2)	5 (E2)	6 (>E2)
1 (<M1)	0.0	0.0	0.0	0.0	0.0	0.0
2 (M1)	0.0	100.0	0.0	0.0	0.0	0.0
3 (E1)	0.0	0.0	100.0	0.0	0.0	0.0
4 (M2)	0.0	0.0	7.1	71.4	21.4	0.0
5 (E2)	0.0	0.0	0.0	8.3	83.3	8.3
6 (>E2)	0.0	0.0	0.0	0.0	0.0	100.0

Tabel 8.5 Overeenstemmingstabel voor Verbanden

Functioneringsniveau op grond van leerkrachtoordeel						
Cito-expertoordeel	1 (<M1)	2 (M1)	3 (E1)	4 (M2)	5 (E2)	6 (>E2)
1 (<M1)	100.0	0.0	0.0	0.0	0.0	0.0
2 (M1)	0.0	100.0	0.0	0.0	0.0	0.0
3 (E1)	0.0	0.0	60.0	30.0	0.0	10.0
4 (M2)	0.0	0.0	40.0	40.0	20.0	0.0
5 (E2)	0.0	0.0	0.0	8.3	75.0	16.7
6 (>E2)	0.0	0.0	0.0	0.0	0.0	100.0

Tabel 8.6 Overeenstemming tussen functioneringsniveaus volgens het leerkrachtoordeel en volgens het Cito-expertoordeel

	Absolute overeenstemming	Gewogen kappa
Bewerkingen	0.80	0.76
Meetkunde	0.86	0.88
Verbanden	0.74	0.82



## 8.2 Betrouwbaarheid

Door het onvolledige design kan bij de subdomeinen Getalbegrip en Meten geen traditionele maat als Cronbach's alpha uitgerekend worden als maat voor betrouwbaarheid. Het is mogelijk om de betrouwbaarheid van deze subdomeinen te schatten door gebruik te maken van het feit dat alle opdrachten en observatiepunten die zijn opgenomen in de kindroute OPLM-geschaald zijn. Ook andere beschrijvende gegevens, zoals het gemiddelde en de standaardmeetfout, zijn te schatten op grond van het feit dat de kindroute volledig bestaat uit OPLM-gekalibreerde opdrachten en observatiepunten. Om relevante beschrijvende gegevens bij de verschillende subdomeinen te genereren, is gebruikgemaakt van het programma OPLAT (Verhelst, Glas en Verstralen, 1995). In OPLAT wordt een door Verhelst et al. (1995, pp. 99-100) ontwikkelde coëfficiënt berekend die qua interpretatie een grote overeenkomst vertoont met betrouwbaarheidscoëfficiënten uit de klassieke testtheorie: de MAcc 'Accuracy of Measurement' (zie Box 1).

### Box 1 'Accuracy of Measurement'

Het begrip ware score is wat meer geëxpliciteerd, namelijk als de verwachte score op een (vaste) toets, maar dan gezien als functie van de latente variabele  $\theta$ . Deze verwachte waarde wordt aangeduid met  $\tau(\theta)$ . Als bovendien bekend is hoe  $\theta$  in de populatie verdeeld is, kunnen ook het gemiddelde en de variantie van de ware scores in de populatie bepaald worden. De variantie van de ware scores in de populatie wordt aangegeven met het symbool  $Var(\tau)$ . Tussen  $\theta$  en  $\tau(\theta)$  bestaat een een-op-een-relatie, immers de een kan uit de andere berekend worden. Het is echter niet zo dat een persoon met vaardigheid  $\theta$  per se de toetsscore  $\tau(\theta)$  moet behalen (dat is alleen zo als de toets oneindig lang wordt). De geobserveerde score bij een eenmalige afname zal dan ook een afwijking vertonen van de verwachte score, waardoor met een eenmalige toetsafname niet meer zonder fout de waarde van  $\theta$  bepaald kan worden. De variantie van de geobserveerde toetsscore wordt aangegeven met  $Var(t|\tau(\theta))$ , en door weer gebruik te maken van de distributie van  $\theta$  in de populatie kan ook de gemiddelde variantie van de geobserveerde toetsscores berekend worden.

$$Var(t) = E[Var(t|\tau(\theta))]$$

Deze variantie kan opgevat worden als de (gemiddelde) meetfoutvariantie in de metriek van de geobserveerde scores ( $t$ ). In analogie met de theorie over de betrouwbaarheid volgt dan

$$MAcc = \frac{Var(\tau)}{Var(\tau) + Var(t)}$$

waarin MAcc staat voor 'Accuracy of Measurement'.

Tabel 8.7 bevat een beschrijving van de meeteigenschappen van de niveaueversies van de subdomeinen. Voor elk subdomein van Rekenen worden de eigenschappen beschreven in de groep waarvoor de niveaueversie is bedoeld. Bij de drie activiteiten zijn er geen verschillende niveaueversies, maar worden de beschrijvende gegevens wel per jaargroep gegeven. In de vierde kolom staat het aantal observatiepunten/opdrachten in elke niveaueversie van een subdomein. Vervolgens worden het gemiddelde, de standaarddeviatie en het maximum van het subdomein in de niveaueversie getoond. In de laatste kolom staat de geschatte betrouwbaarheidscoëfficiënt (MAcc).

Voor meetinstrumenten waaraan geen zware consequenties voor leerlingen verbonden zijn, zoals bij Kleuter in beeld, geeft de COTAN (Commissie Testaangelegenheden Nederland van het Nederlands Instituut van Psychologen) aan dat een betrouwbaarheidscoëfficiënt lager dan 0,70 onvoldoende is, een betrouwbaarheidscoëfficiënt tussen 0,70 en 0,80 voldoende, en een betrouwbaarheidscoëfficiënt hoger dan 0,80 goed (Evers, Lucassen, Meijer en Sijtsma, 2010, p. 33). Op grond van dit criterium is de betrouwbaarheid van de niveaueversies van Kleuter in beeld - Rekenen voldoende tot goed te noemen.

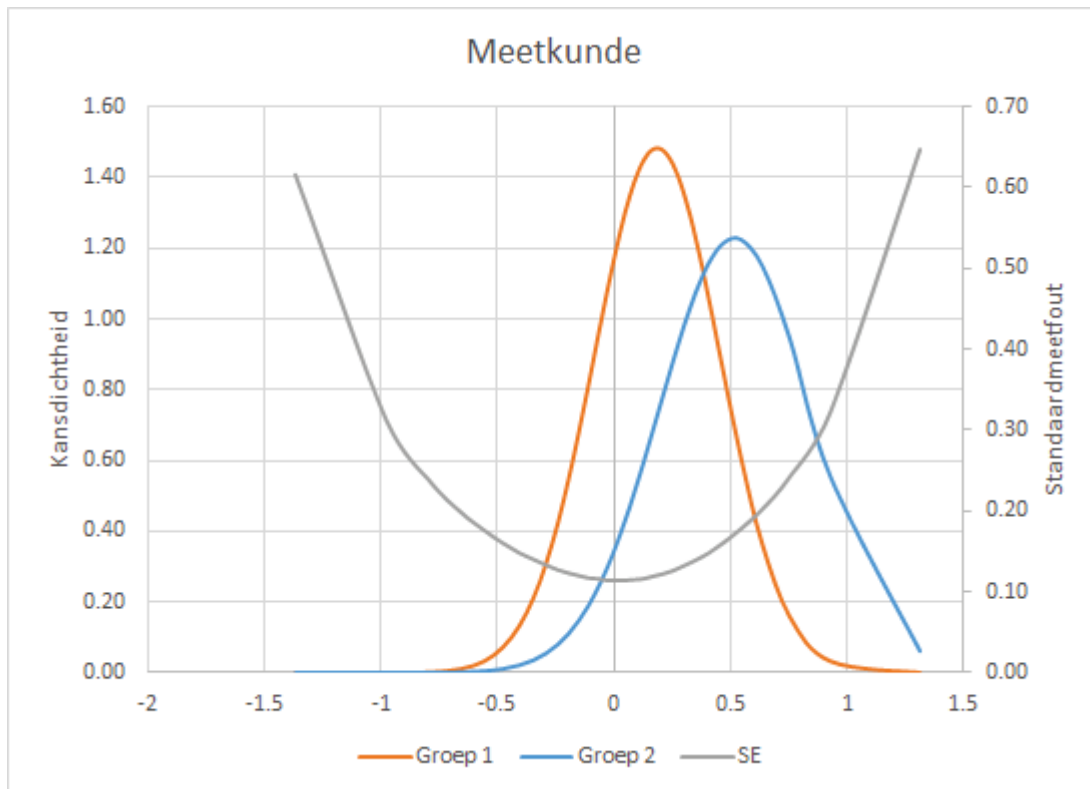
Tabel 8.7 Beschrijvende gegevens van de niveaueversies van de subdomeinen van Rekenen in de kindroute

Subdomein	Versie	Groep	Aantal	Gemiddelde	Stdev	Maximum	MAcc
Getalbegrip	★	Groep 1	18	11,9	3,82	18	0,79
	★★	Groep 2	18	13,6	3,42	18	0,78
	★★★	Groep 2	18	12,0	3,63	18	0,78
Meten	★	Groep 1	18	11,4	3,95	18	0,80
	★★	Groep 2	18	13,5	3,71	18	0,81
	★★★	Groep 2	18	12,7	3,93	18	0,83
Meetkunde	-	Groep 1	21	18,70	5,98	30	0,81
	-	Groep 2	21	23,90	5,06	30	0,82
Verbanden	-	Groep 1	17	20,5	8,91	34	0,92
	-	Groep 2	17	28,6	5,54	34	0,86
Bewerkingen	-	Groep 1	17	21,9	6,87	31	0,86
	-	Groep 2	17	26,5	4,92	31	0,84

### 8.3 Nauwkeurigheid

De hiervoor vermelde betrouwbaarheidscoëfficiënten hebben alleen betrekking op de globale meetnauwkeurigheid van de subdomeinen en geven geen beeld van de lokale meetnauwkeurigheid van de subdomeinen. In Figuur 8.1 is de grootte van de meetfout op de vaardigheidsschaal afgebeeld van het subdomein Meetkunde. Ook is de verdeling van de vaardigheid van kleuters in groep 1 en in groep 2 weergegeven in de figuur. We zien dat de meetfout kleiner is in het gemiddelde vaardigheidsgebied. Voor de andere subdomeinen kan op dezelfde wijze de lokale meetnauwkeurigheid afgebeeld worden en hieruit blijkt ook dat de meetfout kleiner is in het gemiddelde vaardigheidsgebied.

Figuur 8.1 Grootte van de meetfouten voor het subdomein Meetkunde en de bijbehorende kansdichtheidsfuncties



De betekenis van de (lokale) meetnauwkeurigheid voor de functieniveaus die met het meetinstrument gerapporteerd worden, is af te leiden uit betrouwbaarheidstabellen (ook wel verwarringsmatrices genoemd). Met behulp van simulaties is voor elk subdomein en niveauserie nagegaan hoe vaak het functieniveau dat volgt uit de resultaten gelijk is aan het ware (gesimuleerde) functieniveau op het domein.

Tabel 8.8 en Tabel 8.9 laten voor alle subdomeinen en niveauseries zien hoe vaak het werkelijke functieniveau overeenkomt met het waargenomen functieniveau. Zo is in Tabel 8.8 te zien dat 71% van de kleuters die volgens de resultaten in functieniveau 1 (<M1) vallen op het subdomein Meetkunde, ook werkelijk in functieniveau 1 zitten. Aan de rechterkant van de tabel zien we dat 72% van de kleuters waarbij het waargenomen functieniveau 5 (E2) is, ook werkelijk het functieniveau E2 hebben voor Meetkunde.

Tabel 8.8 Betrouwbaarheidstabellen voor de subdomeinen van Rekenen met activiteiten

Subdomein			Ware functieniveau					
			1 (<M1)	2 (M1)	3 (E1)	4 (M2)	5 (E2)	6 (>E2)
Meetkunde	Waargenomen functieniveau	<b>1 (&lt;M1)</b>	<b>0.71</b>	<b>0.29</b>	0.00	0.00	0.00	0.00
		<b>2 (M1)</b>	<b>0.06</b>	<b>0.73</b>	<b>0.19</b>	0.01	0.00	0.00
		<b>3 (E1)</b>	0.00	<b>0.20</b>	<b>0.56</b>	<b>0.23</b>	0.02	0.00
		<b>4 (M2)</b>	0.00	0.01	<b>0.22</b>	<b>0.53</b>	<b>0.24</b>	0.00
		<b>5 (E2)</b>	0.00	0.00	0.01	<b>0.15</b>	<b>0.72</b>	<b>0.13</b>
		<b>6 (&gt;E2)</b>	0.00	0.00	0.00	0.00	<b>0.36</b>	<b>0.64</b>
Bewerkingen	Waargenomen functieniveau	<b>1 (&lt;M1)</b>	<b>0.72</b>	<b>0.28</b>	0.00	0.00	0.00	0.00
		<b>2 (M1)</b>	<b>0.06</b>	<b>0.77</b>	<b>0.17</b>	0.01	0.00	0.00
		<b>3 (E1)</b>	0.00	<b>0.18</b>	<b>0.61</b>	<b>0.19</b>	0.02	0.00
		<b>4 (M2)</b>	0.00	0.01	<b>0.24</b>	<b>0.53</b>	<b>0.22</b>	0.00
		<b>5 (E2)</b>	0.00	0.00	0.01	<b>0.17</b>	<b>0.71</b>	<b>0.11</b>
		<b>6 (&gt;E2)</b>	0.00	0.00	0.00	0.00	<b>0.51</b>	<b>0.49</b>
Verbanden		<b>1 (&lt;M1)</b>	<b>0.66</b>	<b>0.34</b>	0.00	0.00	0.00	0.00
		<b>2 (M1)</b>	<b>0.06</b>	<b>0.78</b>	<b>0.16</b>	0.00	0.00	0.00
		<b>3 (E1)</b>	0.00	<b>0.16</b>	<b>0.66</b>	<b>0.18</b>	0.00	0.00
		<b>4 (M2)</b>	0.00	0.00	<b>0.18</b>	<b>0.63</b>	<b>0.18</b>	0.00
		<b>5 (E2)</b>	0.00	0.00	0.00	<b>0.12</b>	<b>0.78</b>	<b>0.10</b>
		<b>6 (&gt;E2)</b>	0.00	0.00	0.00	0.00	<b>0.44</b>	<b>0.56</b>

Tabel 8.9 Betrouwbaarheidstabellen voor de subdomeinen en niveaueversies van Rekenen met opdrachten

Subdomein			Ware functieniveau						
			1 (<M1)	2 (M1)	3 (E1)	4 (M2)	5 (E2)	6 (>E2)	
Getalbegrip	★	Waargenomen functieniveau	1 (<M1)	0.76	0.24	0.00	0.00	0.00	0.00
			2 (M1)	0.11	0.69	0.17	0.03	0.00	0.00
			3 (E1)	0.00	0.27	0.45	0.23	0.05	0.00
			4 (M2)	0.00	0.04	0.27	0.41	0.27	0.01
			5 (E2)	0.00	0.00	0.04	0.22	0.64	0.10
			6 (>E2)	0.00	0.00	0.00	0.01	0.56	0.43
			6 (>E2)	0.00	0.00	0.00	0.01	0.56	0.43
Getalbegrip	★★	Waargenomen functieniveau	1 (<M1)	0.63	0.37	0.00	0.00	0.00	0.00
			2 (M1)	0.10	0.65	0.21	0.03	0.00	0.00
			3 (E1)	0.00	0.26	0.44	0.25	0.04	0.00
			4 (M2)	0.00	0.04	0.24	0.46	0.26	0.01
			5 (E2)	0.00	0.00	0.02	0.16	0.64	0.18
			6 (>E2)	0.00	0.00	0.00	0.00	0.30	0.70
			6 (>E2)	0.00	0.00	0.00	0.00	0.30	0.70
Getalbegrip	★★★	Waargenomen functieniveau	1 (<M1)	0.64	0.34	0.01	0.00	0.00	0.00
			2 (M1)	0.13	0.57	0.25	0.05	0.00	0.00
			3 (E1)	0.01	0.26	0.41	0.26	0.06	0.00
			4 (M2)	0.00	0.05	0.24	0.40	0.31	0.00
			5 (E2)	0.00	0.00	0.02	0.12	0.71	0.14
			6 (>E2)	0.00	0.00	0.00	0.00	0.26	0.74
			6 (>E2)	0.00	0.00	0.00	0.00	0.26	0.74
Meten	★	Waargenomen functieniveau	1 (<M1)	0.72	0.28	0.00	0.00	0.00	0.00
			2 (M1)	0.10	0.70	0.18	0.02	0.00	0.00
			3 (E1)	0.00	0.26	0.48	0.22	0.05	0.00
			4 (M2)	0.00	0.03	0.27	0.42	0.27	0.00
			5 (E2)	0.00	0.00	0.03	0.20	0.67	0.10
			6 (>E2)	0.00	0.00	0.00	0.01	0.51	0.48
			6 (>E2)	0.00	0.00	0.00	0.01	0.51	0.48
Meten	★★	Waargenomen functieniveau	1 (<M1)	0.61	0.39	0.00	0.00	0.00	0.00
			2 (M1)	0.09	0.67	0.21	0.03	0.00	0.00
			3 (E1)	0.00	0.25	0.46	0.26	0.04	0.00
			4 (M2)	0.00	0.03	0.22	0.47	0.27	0.00
			5 (E2)	0.00	0.00	0.01	0.13	0.72	0.14
			6 (>E2)	0.00	0.00	0.00	0.00	0.34	0.66
			6 (>E2)	0.00	0.00	0.00	0.00	0.34	0.66
Meten	★★★	Waargenomen functieniveau	1 (<M1)	0.63	0.37	0.01	0.00	0.00	0.00
			2 (M1)	0.11	0.62	0.23	0.04	0.00	0.00
			3 (E1)	0.01	0.25	0.45	0.26	0.05	0.00
			4 (M2)	0.00	0.04	0.22	0.44	0.30	0.00
			5 (E2)	0.00	0.00	0.01	0.12	0.73	0.14
			6 (>E2)	0.00	0.00	0.00	0.00	0.29	0.70
			6 (>E2)	0.00	0.00	0.00	0.00	0.29	0.70

Er zijn verschillende indices die de nauwkeurigheid van een betrouwbaarheidstabel samenvatten (zie onder andere Lee, Hanson, & Brennan, 2002). In navolging van Pillner (1969) hebben verschillende onderzoekers geconcludeerd dat leerlingen in een bepaald niveau een waar niveau hebben dat 1 hoger of 1 lager is, tenzij de meting perfect betrouwbaar is (Wheadon en Stockford, 2010). Daarom stelt Pillner als ambitieniveau dat 95 procent van de leerlingen dat in een functieniveau valt in werkelijkheid ook in dat functieniveau moet vallen, of één functieniveau daarboven of één functieniveau daaronder. In de tabellen zijn dit de gearceerde cellen. Dit ambitieniveau is dus gebaseerd op de veronderstelling dat geen enkel meetinstrument perfect meet en dat er dus altijd sprake is van foutieve classificaties. In dat licht is de maximale accuraatheid die op het individuele niveau bereikt kan worden plus of minus één functieniveau. In Tabel 8.10 zien we dat alle subdomeinen en niveaoversies boven dit ambitieniveau liggen.

Een studie van Wheadon en Stockford (2010) laat zien dat, onder andere, kortere toetsen en makkelijke toetsen een lagere nauwkeurigheid hadden dan lange toetsen en toetsen met een grote spreiding in itemmoeilijkheid. Ook bij ons is zichtbaar dat de makkelijke, korte niveaoversies voor groep 1 (★) een iets lagere nauwkeurigheid hebben dan bijvoorbeeld de activiteiten met een grotere spreiding in moeilijkheid en meer schaalpunten.

Nauwkeurigheid wordt ook vaak geëvalueerd door te kijken naar de vals-positief en vals-negatief fouten. Het conditionele vals-positief foutpercentage (P+) is de kans dat een kleuter in een hoger functioneringsniveau valt dan het ware functioneringsniveau van de kleuter. Het conditionele vals-negatief foutpercentage (P-) is de kans dat een kleuter in een lager functioneringsniveau valt dan het ware functioneringsniveau van de kleuter. Aan deze percentages zien we dat de misclassificaties iets vaker in positieve zin zijn: als sprake is van een misclassificatie vallen kleuters iets vaker in een hoger functieniveau dan hun werkelijke functieniveau en niet in een lager functieniveau.

Tabel 8.10 Samenvattende indices voor de nauwkeurigheid van de subdomeinen en niveaoversies

Subdomein	Niveau- versie	+/- index	Foutpercentage		Aantal	
			Vals negatief (P-)	Vals positief (P+)	schaalpunten	Aantal items
<b>Bewerkingen</b>	–	0.99	0.14	0.22	31	17
<b>Meetkunde</b>	–	0.99	0.16	0.19	30	21
<b>Verbanden</b>	–	1.00	0.12	0.18	34	17
<b>Getalbegrip</b>	★	0.96	0.21	0.22	18	18
	★★	0.98	0.16	0.23	18	18
	★★★	0.97	0.16	0.20	18	18
<b>Meten</b>	★	0.97	0.20	0.22	18	18
	★★	0.98	0.15	0.21	18	18
	★★★	0.98	0.15	0.20	18	18

## 9 Validiteit

Validiteit is de mate waarin een test aan zijn doel beantwoordt: kan men uit de testcores de conclusies trekken die men op het oog heeft? Kleuter in beeld - Rekenen heeft geen voorspellende pretentie, daarom is de criteriumvaliditeit niet van belang. Bij een meetinstrument zoals Kleuter in beeld - Rekenen, dat tot doel heeft om de vaardigheid op een objectieve manier in beeld te brengen en problemen tijdig te signaleren, speelt de inhoudsvaliditeit een belangrijke rol. De argumenten over de inhoudsvaliditeit zijn uitgebreid besproken in hoofdstuk 2 en 3, maar zullen in de eerste paragraaf samengevat worden.

Naast de inhoudsvaliditeit is de begripsvaliditeit van belang. De begripsvaliditeit van een meetinstrument gaat over de vraag in hoeverre de uitkomsten toe te schrijven zijn aan verklarende concepten en constructen die deel uitmaken van het theoretische kader dat aan de ontwikkeling van het meetinstrument ten grondslag ligt. In de tweede paragraaf worden verschillende aspecten besproken die informatie geven over de begripsvaliditeit.

### 9.1 Inhoudsvaliditeit

Validering begint bij het specificeren van de concepten en de ontwikkeling van de inhoud van het meetinstrument (Anastasi, 1986; Schouwstra, 2000). De inhoudelijke verantwoording van Kleuter in beeld - Rekenen is uitgebreid beschreven in hoofdstuk 2 en 3. Bij Kleuter in beeld - Rekenen hebben vier informatiebronnen de ontwikkeling gestuurd: de input van onderwijsprofessionals, wetenschappelijke kennis over rekenen, de expertise van toetsdeskundigen en (psychometrische) informatie verkregen uit empirisch onderzoek. Zoals in hoofdstuk 3 beschreven, zijn eerst met behulp van een focusgroep, een expertbijeenkomst en een online community, de wensen vanuit het veld geïnventariseerd en de globale opzet van het instrument bepaald.

De inhoudskaart Rekenen-Wiskunde kleuters (fase 1) van SLO (2018) heeft een belangrijke rol gespeeld bij het bepalen van de subdomeinen en onderdelen. Aanvullende literatuur is geraadpleegd om de potentiële leerdoelen die het meetinstrument zou moeten meten in kaart te brengen. Na input en feedback van onderwijsprofessionals zijn de uiteindelijk te meten leerdoelen geselecteerd en geplaatst onder de subdomeinen. Dit proces heeft geleid tot een toetsmatrijs bestaande uit vijf subdomeinen en 16 onderdelen (zie hoofdstuk 2 en Tabel 9.1).

Hierna zijn de toetsdeskundigen begonnen met de constructie van opdrachten en observatiepunten bij activiteiten om de concepten en constructen uit de toetsmatrijs in beeld te brengen. Onderwijsprofessionals konden feedback geven op een prototype dat uitgetoetst is in een try-out. Tijdens een proefonderzoek is de eerste keer empirische en psychometrische informatie verzameld en vervolgens in het kwaliteitsonderzoek. Tijdens beide onderzoeken is de empirische en psychometrische informatie aangevuld met observaties van de toetsdeskundigen tijdens de afnames en feedback van de leerkrachten dat verzameld is via een online evaluatieformulier.

Na het kwaliteitsonderzoek zijn alle opdrachten en observatiepunten geëvalueerd. De suggesties van de leerkrachten en de observaties van de toetsdeskundigen hebben hierbij een belangrijke rol gespeeld. Daarnaast zijn alle opdrachten en observatiepunten langs de psychometrische meetlat gelegd. Hierbij is onder andere gekeken naar de moeilijkheid en het onderscheidend vermogen van de opdrachten en observatiepunten, maar ook het instrument en de niveaueversies als geheel zijn bekeken, zoals de structuur (zie hoofdstuk 5) en de betrouwbaarheid (zie hoofdstuk 8).

Tabel 9.1 Subdomeinen en onderdelen van *Kleuter in beeld – Rekenen*

Subdomein	Onderdelen
<b>Getalbegrip</b>	Telrij Hoeveelheden Getallen Relaties tussen telrij, hoeveelheden en getallen incl. Verhoudingen
<b>Bewerkingen</b>	Optellen en aftrekken met hele getallen (tot ten minste 20) (optellen, aftrekken, splitsen) Vermenigvuldigen en delen met hele getallen (tot ten minste 20) ((ver)delen)
<b>Meten</b>	Lengte en omtrek Oppervlakte Inhoud Gewicht Tijd Geld incl. Verhoudingen
<b>Meetkunde</b>	Oriënteren in de ruimte Construeren Opereren met vormen en figuren
<b>Verbanden</b>	Verbanden

## 9.2 Begripsvaliditeit

Bij begripsvaliditeit gaat het erom te toetsen of het meetinstrument inderdaad de eigenschap meet die wordt verondersteld. We presenteren verschillende onderzoeksgegevens die relevant zijn, namelijk gegevens over de structuur (dimensionaliteit), de psychometrische kwaliteit van de opdrachten en observatiepunten, soortgenoten validiteit en gegevens over verschillen tussen relevante groepen.

### **Structuur**

Het algemene uitgangspunt van het meetmodel is dat elk van de subdomeinen van rekenvaardigheid unidimensionaal is. Naast de inhoudelijk relevante indeling van de opdrachten per subdomein zijn unidimensionale schalen gerealiseerd. Dit betekent dat met elke willekeurige subset van opdrachten uit één subdomein dezelfde onderliggende vaardigheid kan worden vastgesteld. Zoals in hoofdstuk 5 al aangegeven zijn bij de kalibratie voor alle opdrachten *S*-toetsen uitgevoerd die een indicatie geven van de kwaliteit van de kalibratie. Daarbij is duidelijk geworden dat de verdeling van overschrijdingskansen bij deze statistische toetsen redelijk gelijkmatig is over het gehele interval waarin de overschrijdingskansen kunnen liggen (i.e. tussen 0 en 1). Dit resultaat geeft een bevestiging van het eerder geschetste beeld, dat er met uitzondering van enkele opdrachten/observatiepunten, sprake is van niet-significante *S*-toetsen. Zij vormen een kwantitatieve ondersteuning van de conclusie dat elke set van opdrachten/observatiepunten behorend tot één subdomein, een unidimensioneel construct representeren (zie Tabel 5.1). In hoofdstuk 5 zijn ook de *R1c*-waarden gepresenteerd als maat voor de modelpassing. Omdat deze waarden eveneens ondersteuning bieden voor de validiteit refereren we daar nogmaals aan. *R1c* is een statistiek die zicht geeft op de modelpassing van het meetinstrument als geheel. Voor een acceptabele modelpassing geldt als vuistregel dat *R1c* niet groter is dan ongeveer anderhalf maal het aantal vrijheidsgraden (*df*) zou moeten zijn en bij voorkeur niet significant. Voor de subdomeinen waarvoor de *R1c* uitgerekend kon worden, geldt dat *R1c* niet groter is dan ongeveer anderhalf maal het aantal vrijheidsgraden. De subdomeinen Getalbegrip en Meten zijn daarnaast niet significant. De modelpassing is dus acceptabel.



De leerkrachtroute vormt ook een eendimensionale schaal, zoals blijkt uit een principale component-analyse. De principale componentenanalyse liet zien dat er één component met een eigenwaarde groter dan 1 is en slechts één component boven de knip in een screeplot. Alle observatiepunten hadden een hoge componentenlading, zie Tabel 9.2.

Tabel 9.2 *Uitkomst principale componentenanalyse van de leerkrachtroute*

Observatiepunt	Componentlading
Getalbegrip – telrij	0,90
Getalbegrip – hoeveelheden	0,91
Getalbegrip – getallen	0,90
Getalbegrip – relaties tussen telrij, hoeveelheden en getallen	0,90
Bewerkingen – optellen en aftrekken met hele getallen	0,91
Bewerkingen – vermenigvuldigen en delen met hele getallen	0,91
Metten – lengte en omtrek	0,90
Metten – oppervlakte	0,90
Metten – inhoud	0,91
Metten – gewicht	0,90
Metten – tijd	0,90
Metten – geld	0,88
Meetkunde – oriënteren in de ruimte	0,86
Meetkunde – construeren	0,87
Meetkunde – opereren met vormen en figuren	0,88
Verbanden	0,91

### **Itemkwaliteit**

In hoofdstuk 4 is een globaal overzicht gegeven van de beschrijvende gegevens van de opdrachten en observatiepunten en zijn de resultaten besproken. In Tabel 9.3 zijn het bereik en de gemiddelden weergegeven voor de p-waarden en de  $R_{it}$ -waarden van de opdrachten en observatiepunten van de subdomeinen in de kindroute. Alle  $R_{it}$ -waarden liggen boven de ,23 en zijn daarmee alle voldoende (tussen de ,20 en ,29) tot goed (boven de ,30) volgens het COTAN Beoordelingssysteem (p. 40).

Tabel 9.3 Het bereik en de gemiddelden voor de p-waarden en de Rit-waarden van de opdrachten en observatiepunten in de kindroute

Sub-domein	Niveau-versie	Groep	Ni	Max	Min. P	Gem. P	Max. P	Min. Rit	Gem. Rit	Max. Rit
GB	★	Groep 1	18	18	0,39	0,66	0,88	0,35	0,47	0,59
	★★	Groep 2	18	18	0,47	0,75	0,89	0,39	0,46	0,61
	★★★	Groep 2	18	18	0,35	0,67	0,87	0,31	0,45	0,58
BW	–	Groep 1	17	31	0,57	0,71	0,85	0,43	0,56	0,71
		Groep 2	17	31	0,74	0,86	0,94	0,39	0,53	0,66
MT	★	Groep 1	18	18	0,39	0,64	0,85	0,39	0,48	0,55
	★★	Groep 2	18	18	0,48	0,75	0,91	0,37	0,49	0,60
	★★★	Groep 2	18	18	0,36	0,71	0,91	0,36	0,50	0,62
MK	–	Groep 1	21	30	0,40	0,64	0,90	0,26	0,44	0,64
		Groep 2	21	30	0,64	0,81	0,96	0,24	0,45	0,65
VB	–	Groep 1	17	34	0,32	0,60	0,84	0,39	0,64	0,74
		Groep 2	17	34	0,65	0,84	0,92	0,30	0,54	0,68

Voor de subdomeinen met opdrachten zien we dat de p-waarden liggen tussen 0,35 (GB ★★★) en 0,91 (MT ★★ en MT ★★★). Omdat met alle niveaiversies gesignaleerd moet kunnen worden of een kleuter op een hoger of lager functioneringsniveau zit, is gekozen om bij de versie voor groep 1 ook een klein aantal moeilijkere opdrachten op te nemen en bij de versie voor midden groep 2 en eind groep 2 ook een klein aantal makkelijkere opdrachten. De gemiddelde moeilijkheid past goed bij de groepen (tussen de ,64 en ,75).

Voor de activiteiten is gekozen om dezelfde activiteit en observatiepunten te gebruiken voor groep 1 en groep 2. Hierdoor is het bereik van de p-waarden iets groter voor de observatiepunten (tussen de ,32 en ,96). Veel van de observatiepunten zijn polytoom, waarbij een onderscheid wordt gemaakt tussen een activiteit niet goed uitvoeren (0), goed uitvoeren met hulp (1) en een activiteit zelfstandig goed uitvoeren (2). In groep 1 zal vaker een activiteit goed uitgevoerd worden met hulp en in groep 2 vaker zelfstandig. Deze opzet voorkomt dat in groep 1 kinderen gefrustreerd raken, omdat ze (met hulp van de leerkracht) toch een succeservaring hebben.

In hoofdstuk 5 is ten slotte informatie gegeven over de nauwkeurigheid van de itemparameters (constante 'c'). De gemiddelde waarde van de constante is goed in alle niveaiversies. Bij de grote meerderheid van alle itemparameters (86%) is de nauwkeurigheid voldoende tot goed (een  $c \leq ,4$ ).

### Itembias

Er kon geen onderzoek uitgevoerd worden naar differentieel itemfunctioneren (*Differential Item Functioning*, DIF) met betrekking tot sekse. De leerkrachten hebben bij slechts 126 kleuters het geslacht aangegeven, hierdoor kan itembias met betrekking tot sekse niet geanalyseerd worden. Ook onderzoek naar differentieel itemfunctioneren (*Differential Item Functioning*, DIF) met betrekking tot de wijze van afname is niet mogelijk door bestaande verschillen tussen scholen. Het gehele onderzoek is uitgevoerd op een vrijwillige basis. Scholen konden zelf kiezen tussen een digitale afname van de opdrachten of een papieren afname. Alle activiteiten waren voor alle scholen identiek, omdat deze op observaties gebaseerd zijn. Desondanks zagen we per observatiepunt van de activiteiten verschillen tussen "digitale afname"-scholen en "papieren afname"-scholen. Deze verschillen illustreren dat de scholen die voor een digitale afname van opdrachten hebben gekozen en de scholen die voor een papieren afname van opdrachten hebben gekozen duidelijk

verschillend zijn. Door deze bestaande verschillen is een onderzoek naar DIF met betrekking tot wijze van afname niet mogelijk.

### **Convergente validiteit**

In het kader van de convergente validiteit, is de samenhang bekeken tussen de (indirecte) observaties uit de leerkrachtroute en de vaardigheden zoals gemeten in de kindroute. In de leerkrachtroute vult de leerkracht per onderdeel een oordeel in op basis van indirecte observaties. Voor de subdomeinen met slechts één indirecte observatie (Verbanden) is het functioneringsniveau gelijk aan de observatie. Voor de subdomeinen met meerdere observatiepunten, is het functioneringsniveau bepaald zoals beschreven in hoofdstuk 7.

Eerst is per subdomein voor elk functioneringsniveau uit de leerkrachtroute gekeken wat de gemiddelde vaardigheid is zoals gemeten in de kindroute, zie Tabel 9.4 en Tabel 9.5. Zoals verwacht is bij elk subdomein een significant verschil in vaardigheid (kindroute) tussen de functioneringsniveaus uit de leerkrachtroute. De vaardigheid loopt op per functioneringsniveau conform de verwachting. Bijvoorbeeld, kleuters die volgens de leerkrachtroute in het functioneringsniveau <E1 vallen bij Getalbegrip hebben een gemiddelde vaardigheid in Getalbegrip van -0,08 zoals gemeten in de kindroute. Kleuters die volgens de leerkrachtroute in het functioneringsniveau E1 vallen bij Getalbegrip hebben een hogere gemiddelde vaardigheid in Getalbegrip, namelijk 0,15. Posthoc-toetsen laten zien dat het verschil tussen opeenvolgende functioneringsniveaus niet altijd significant is. Bij Verbanden, Bewerkingen en Meten is het verschil tussen boven E2 en respectievelijk E2 en M2 niet significant.

*Tabel 9.4 De gemiddelde vaardigheid Rekenen per subdomein met opdrachten binnen elk functioneringsniveau van de leerkrachtroute*

Functioneringsniveau leerkrachtroute	Vaardigheid uit kindroute				
	Gemiddelde	N	Standaard-deviatie	F	Sig.
<b>Getalbegrip</b>					
0 (<E1)	-0,08	54	0,40	67,83	0,000
1 (E1)	0,15	149	0,34		
2 (M2)	0,40	200	0,38		
3 (E2)	0,64	169	0,41		
4 (>E2)	0,88	34	0,39		
Totaal	0,39	606	0,46		
<b>Meten</b>					
0 (<E1)	-0,08	58	0,36	47,83	0,000
1 (E1)	0,15	145	0,37		
2 (M2)	0,50	262	0,41		
3 (E2)	0,61	97	0,45		
4 (>E2)	0,78	12	0,40		
Totaal	0,38	574	0,46		

Tabel 9.5 De gemiddelde vaardigheid Rekenen per subdomein met activiteiten binnen elk functioneringsniveau van de leerkrachtroute

Functioneringsniveau leerkrachtroute	Vaardigheid uit kindroute				
	Gemiddelde	N	Standaard-deviatie	F	Sig.
<b>Bewerkingen</b>					
0 (<E1)	0,00	21	0,44	24,31	0,000
1 (E1)	0,40	46	0,28		
2 (M2)	0,64	56	0,35		
3 (E2)	0,93	20	0,30		
4 (>E2)	0,89	5	0,36		
Totaal	0,52	148	0,44		
<b>Meetkunde</b>					
0 (<E1)	-0,14	6	0,25	25,02	0,000
1 (E1)	0,13	38	0,27		
2 (M2)	0,44	61	0,29		
3 (E2)	0,70	53	0,37		
4 (>E2)	1,29	1	0,00		
Totaal	0,43	159	0,40		
<b>Verbanden</b>					
0 (<E1)	0,00	42	0,50	31,93	0,000
1 (E1)	0,38	37	0,24		
2 (M2)	0,67	76	0,35		
3 (E2)	0,72	49	0,33		
4 (>E2)	0,82	11	0,24		
Totaal	0,51	215	0,45		

Vervolgens is ook de samenhang uitgerekend. Omdat de observatieschaal in de leerkrachtroute een ordinaal meetniveau heeft en de samenhang met de vaardigheid niet per definitie lineair is, maar wel monotoon stijgend, is Spearman's rho uitgerekend. We vinden een positieve samenhang tussen de (indirecte) observaties uit de leerkrachtroute en de vaardigheid zoals gemeten in de kindroute (Tabel 9.6). We zien dat de samenhang bij Meten iets zwakker is dan bij de overige vier subdomeinen.

Tabel 9.6 Samenhang per subdomein tussen de (indirecte) observaties uit de leerkrachtroute en de vaardigheid in kindroute

	Spearman's rho	p	N
<b>Getalbegrip</b>	0,56	0,000	606
<b>Bewerkingen</b>	0,64	0,000	148
<b>Meten</b>	0,49	0,000	574
<b>Meetkunde</b>	0,64	0,000	159
<b>Verbanden</b>	0,59	0,000	215

In de leerkrachtroute vult de leerkracht voor 16 onderdelen een observatiepunt in (zie hoofdstuk 3.3). Ter validatie is de leerkrachten ook gevraagd een niveau-inschatting voor elk subdomein als geheel te geven

(zie Figuur 9.1), naast de observaties per onderdeel uit de leerkrachtroute. We hebben per subdomein de samenhang bekeken tussen de subdomein niveau-inschatting en het functioneringsniveau dat bepaald is op grond van de observatiepunten (zoals beschreven in hoofdstuk 7). We zien dat de samenhang tussen beide (indirecte) observaties goed is, zie Tabel 9.7.

*Figuur 9.1 Niveau-inschatting voor elk subdomein door de leerkracht*

<b>Niveau-inschatting domeinen</b>				
<b>Getalbegrip</b>				
<b>&lt;E1</b>	<b>E1</b>	<b>M2</b>	<b>E2</b>	<b>&gt;E2</b>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Bewerkingen</b>				
<b>&lt;E1</b>	<b>E1</b>	<b>M2</b>	<b>E2</b>	<b>&gt;E2</b>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Meten</b>				
<b>&lt;E1</b>	<b>E1</b>	<b>M2</b>	<b>E2</b>	<b>&gt;E2</b>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Meetkunde</b>				
<b>&lt;E1</b>	<b>E1</b>	<b>M2</b>	<b>E2</b>	<b>&gt;E2</b>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
<b>Verbanden</b>				
<b>&lt;E1</b>	<b>E1</b>	<b>M2</b>	<b>E2</b>	<b>&gt;E2</b>
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

De divergente validiteit is niet onderzocht, omdat het kwaliteitsonderzoek met de leerkrachtroute, vijf subdomeinen uit de kindroute en het interbeoordelaarsonderzoek tijdsintensief is voor scholen en voor kinderen en een omvangrijke steekproef vereist. Om voor scholen en kinderen additionele belasting te vermijden, is ervoor gekozen om het kwaliteitsonderzoek te beperken tot het te onderzoeken instrument Kleuter in beeld - Rekenen.

Tabel 9.7 *Samenhang subdomein niveau-inschatting van de leerkracht en de (indirecte) observaties per subdomein uit de leerkrachtroute*

	Spearman's rho	p	N
<b>Getalbegrip</b>	0,92	0,000	632
<b>Bewerkingen</b>	0,91	0,000	608
<b>Meten</b>	0,90	0,000	631
<b>Meetkunde</b>	0,86	0,000	617
<b>Verbanden</b>	0,89	0,000	616

### **Verschillen tussen relevante subgroepen**

In het kwaliteitsonderzoek zijn geboortedatum, geslacht en leerlinggewicht van de leerlingen opgevraagd. Het leerlinggewicht is echter bij geen enkele leerling ingevuld, om die reden kunnen de verschillen tussen de gewichtscategorieën niet getoetst worden.

Voor leeftijd en geslacht zullen de verschillen tussen subgroepen worden besproken. Voor de kindroute wordt de gemiddelde vaardigheid van de subgroepen vergeleken. Met behulp van een chikwadraattoets wordt gekeken of de functioneringsniveaus die gerapporteerd zijn in de leerkrachtroute significant verschillen tussen subgroepen. Voorts wordt bekeken of sprake is van een samenhang tussen functioneringsniveau enerzijds en geslacht en leeftijdscategorie anderzijds.

## **Kindroute**

### **Leeftijd**

Voor alle leerlingen is de leeftijd berekend op 1 januari 2020. Vervolgens zijn de kleuters in vier leeftijds-categorieën ingedeeld: (0)  $\leq$  5 jaar; (1) 5 – 5,5 jaar; (2) 5,5 – 6 jaar; (3)  $>$  6 jaar. Conform verwachting heeft de leeftijdscategorie een significant effect op de gemiddelde vaardigheid op de subdomeinen (zie Tabel 9.8). In Tabel 0.1 (Appendix) is voor elk subdomein de gemiddelde vaardigheid per leeftijdscategorie te vinden. We zien dat wanneer sprake is van een significant verschil deze conform verwachting is, dat wil zeggen dat hogere leeftijdscategorieën gemiddeld een hogere vaardigheid hebben. Bij Bewerkingen is ook een significant effect van leeftijdscategorie, maar de verschillen tussen opeenvolgende leeftijdscategorieën zijn niet significant.

Tabel 9.8 *Effect van leeftijdscategorie op de vaardigheid van de subdomeinen*

Subdomein	F	df1	df2	Sig.	Eta
<b>Getalbegrip</b>	6,61	3	109	0,000	0,39
<b>Bewerkingen</b>	48,88	3	719	0,000	0,41
<b>Meten</b>	46,55	3	765	0,000	0,39
<b>Meetkunde</b>	12,05	3	149	0,000	0,44
<b>Verbanden</b>	13,30	3	202	0,000	0,41

### **Geslacht**

We vinden bij alle subdomeinen geen significante verschillen tussen jongens en meisjes (zie Tabel 9.9). In Tabel 0.2 is voor elk subdomein de gemiddelde vaardigheid per leeftijdscategorie te vinden.

Tabel 9.9 Effect van geslacht op de vaardigheid van de subdomeinen

Subdomein	F	df1	df2	Sig.	Eta
<b>Getalbegrip</b>	0,48	1	35	0,495	0,013
<b>Bewerkingen</b>	0,58	1	123	0,446	0,005
<b>Meten</b>	1,17	1	123	0,283	0,009
<b>Meetkunde</b>	0,22	1	23	0,645	0,009
<b>Verbanden</b>	1,29	1	24	0,268	0,051

## Leerkrachtroute

### Leeftijd

Het functioneringsniveau is significant verschillend tussen de leeftijdscategorieën, net zoals het geval was in de kindroute (zie Tabel 9.10). Conform verwachting is er een positieve samenhang tussen functioneringsniveau en leeftijdscategorie. In Appendix is de tabel te vinden met het percentage kinderen in elke leeftijdscategorie per functioneringsniveau (Tabel 0.3). Het patroon in de tabel is naar verwachting. Hoe ouder het kind, hoe hoger het functioneringsniveau.

Tabel 9.10 Chikwadraattoets voor de verschillen in functioneringsniveau tussen leeftijdscategorieën en de samenhang tussen functioneringsniveau en leeftijdscategorie

Subdomein	Chi-Square Tests	df	Sig.	Eta	Gamma
<b>Getalbegrip</b>	221,56	12	0,000	0,56	0,58
<b>Bewerkingen</b>	244,16	12	0,000	0,60	0,62
<b>Meten</b>	211,96	12	0,000	0,56	0,58
<b>Meetkunde</b>	235,98	12	0,000	0,58	0,58
<b>Verbanden</b>	240,66	12	0,000	0,59	0,56

### Geslacht

Net zoals bij de Kindroute, vinden we in de leerkrachtroute geen significante verschillen tussen jongens en meisjes bij alle subdomeinen, zie Tabel 9.11. In Tabel 0.4 (Appendix) is het percentages jongens en meisjes in elk functioneringsniveau beschreven.

Tabel 9.11 Chikwadraattoets voor de verschillen in functioneringsniveau tussen meisjes en jongens en de samenhang tussen functioneringsniveau en geslacht

Subdomein	Chi-kwadraat	df	Sig.	Eta	Gamma
<b>Getalbegrip</b>	8,75	4	0,068	0,03	-0,01
<b>Bewerkingen</b>	5,38	4	0,251	0,00	0,03
<b>Meten</b>	5,41	4	0,248	0,05	-0,01
<b>Meetkunde</b>	4,13	3	0,247	0,04	0,12
<b>Verbanden</b>	3,42	4	0,490	0,00	0,05





## 10 Afname en rapportage

De leerkrachtmap bevat naast alle instructies, opdrachtenboekjes en observatieformulieren een beknopte handleiding. Deze is gebruikersvriendelijk en visueel van opzet en geeft door middel van schema's in korte tijd een duidelijk beeld van de mogelijkheden en het gebruik van Kleuter in beeld - Rekenen. De handleiding bevat bijvoorbeeld een stappenplan voor het gebruik van Kleuter in beeld - Rekenen. Een van de hoofdstukken (H.3) gaat over het interpreteren en analyseren en gaat kort in op de rapportagevormen. In de digitale handleiding van Kleuter in beeld online staat stap voor stap uitgelegd hoe ingelogd kan worden in Basispoort (één inlogprocedure voor digitale materialen in het primair onderwijs) en hoe Kleuter in beeld online via een webbrowser geopend kan worden. Ook staat hierin onder meer beschreven hoe de digitale opdrachten opgestart kunnen worden. Onder het kopje 'Downloads' in Kleuter in beeld online zijn aanvullende documenten te vinden, zoals bijvoorbeeld de praatplaat Samen in gesprek. Deze is bedoeld voor scholen om vóór het gebruik van Kleuter in beeld - Rekenen eerst samen in gesprek te gaan op welke wijze dit instrument het beste ingezet kan worden.

### 10.1 Afname

De observaties bij de leerkrachtroute en de opdrachten en activiteiten uit de kindroute zijn niet gebonden aan een bepaalde periode, dit kan het hele schooljaar door. Om de voortgang en het overzicht van de groep te kunnen bewaken, is vanwege praktische overwegingen gekozen voor drie observatieperiodes. Dit sluit ook aan op de praktijk in het onderwijs en het leerlingvolgsysteem van Cito. De online-omgeving werkt met de volgende drie observatieperiodes in een schooljaar:

- Begin: van augustus tot en met november
- Medio: van december tot en met maart
- Eind: van april tot en met juli

De leerkracht kiest per kind en per subdomein voor de leerkrachtroute of de kindroute. De leerkracht vult de leerkrachtroute in op basis van eigen observaties en/of legt de opdrachten en activiteiten aan de kinderen voor. We gaan daarbij uit van de professionaliteit van de leerkracht. Deze wordt geacht in staat te zijn om – aan de hand van de aanwijzingen in de handleiding en instructies – een gestandaardiseerde afname te realiseren.

#### ***Leerkrachtroute***

De leerkrachtroute kan eenvoudig via Kleuter in beeld online digitaal ingevuld worden. Als de leerkracht voor alle vijf subdomeinen voor de leerkrachtroute kiest, vult hij alle 16 observatiepunten in. Op basis van beschrijvingen bij E1 en E2 kiest de leerkracht welk van de zes niveaus voor elk observatiepunt het beste bij het kind past. Dit kost maximaal tien minuten per kind. Zie ook paragraaf 3.3 en Figuur 3.2.

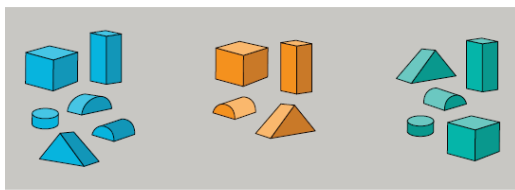
#### ***Kindroute opdrachten***

Het kind kan de opdrachten op papier of digitaal maken. Elke taak begint met een voorbeeldopdracht, zodat het kind weet wat de bedoeling is. Als een taak verschillende soorten opdrachten bevat, dan zijn er meer voorbeeldopdrachten. In Figuur 3.1 staat per subdomein hoe lang een taak ongeveer duurt.

Bij de digitale opdrachten is een korte instructie voor de leerkracht beschikbaar (in leerkrachtmap). Daarin staat bijvoorbeeld hoe hij de kinderen helpt met het opstarten van de opdrachten. De digitale opdrachten kunnen op verschillende devices gemaakt worden: iPad, Chromebook, Windows computer of Macbook of iMac. De opdrachten worden individueel gemaakt. Afhankelijk van de faciliteiten van de school kunnen meerdere kinderen tegelijk de digitale opdrachten maken. Een computerstem leest de instructie en de opdrachten voor. De instructie is ingesproken door een vrouw. De opdrachten zijn ingesproken door een jongen van 12 jaar. Een meerderheid uit het veld gaf namelijk aan een voorkeur te hebben voor een

kinderstem bij de opdrachten, omdat die beter aansluit bij de belevingswereld van kleuters. De kinderen kunnen de opdrachten in principe zelfstandig doorlopen. De kinderen beantwoorden de opdrachten door op een afbeelding te klikken met de muis of met de vinger op het scherm te tikken (bijvoorbeeld op de iPad). Vanwege de jonge leeftijd van kinderen is het een 'geleide afname', dat wil zeggen dat het kind niet zelf door de opdrachten heen kan navigeren. Na het kiezen van een antwoordalternatief wordt de button 'verder' actief en kan het kind door naar de volgende opdracht. Het is mogelijk het antwoord aan te passen zolang het kind in het betreffende scherm blijft. Het kind kan niet terug naar voorgaande opdrachten. Als een afname om wat voor reden dan ook wordt afgebroken, kan het kind een volgende keer verder met de opdracht waar hij gebleven was. De antwoorden worden automatisch nagekeken. Na het afronden van een taak wordt het niveau automatisch berekend en zichtbaar in de rapportage.

Bij de papieren opdrachten werken kinderen individueel in een opdrachtenboekje op A5-formaat. Elke opdracht staat op een aparte bladzijde, zodat kinderen niet afgeleid worden door andere opdrachten. Per subdomein en per niveauversie is er een gebruikersvriendelijke instructie voor de leerkrachten beschikbaar. De leerkracht ziet alle opdrachten van dat boekje in beeld. Bij elke opdracht hoort een symbool in plaats van een paginanummer (in het voorbeeld een appel). Hiermee kan de leerkracht de kinderen naar de juiste bladzijde verwijzen. Per opdracht staat de voorleeszin vermeld die de leerkracht voorleest.



Waar zie je de meeste blokken?

[Voorbeeld uit Instructie leerkracht](#)

De leerkracht kiest zelf of hij de opdrachten individueel voorlegt of met een groepje aan de slag gaat.

Het kind geeft antwoord door een hokje of twee hokjes onder of naast het juiste plaatje of getal aan te kruisen of in te kleuren. Bij Getalbegrip zijn er ook enkele opdrachten waarbij het kind een lijn moet trekken. Na afloop kan de leerkracht de opdrachten nakijken, aan de hand van de nakijkaarten uit de leerkrachtmap. Daarna kan hij het aantal 'goed' invoeren in Kleuter in beeld online. Vervolgens wordt het niveau automatisch berekend en wordt dit meteen zichtbaar in de rapportage.

### **Kindroute activiteiten**

De leerkracht neemt de activiteiten af. Bij het ontwikkelen van de activiteiten moest er gekozen worden voor een individuele afname of in een groepje. Een activiteit afnemen in een groepje heeft het voordeel dat dit tijds winst oplevert en makkelijker uitvoerbaar is in de klassensituatie. In de try-out van Kleuter in Beeld - Taal is dit uitgetoet, maar het bleek te veel nadelen te hebben. Het bleek dat kinderen dan op elkaar reageren en elkaar napraten, zodat de vaardigheid van elk kind niet onafhankelijk en minder goed in kaart gebracht kan worden. Bij Kleuter in beeld - Rekenen hebben we er daarom meteen voor gekozen om alleen met individuele afnames van de activiteiten te werken.

Alle drie de activiteiten hebben een speels karakter en sluiten aan bij de belevingswereld van kleuters. Bij Bewerkingen speelt het kind een doen-alsof-spel, waarbij Anne gaat eten bij opa en oma. Bij de drie deelactiviteiten van Meetkunde gaat het kind onder andere een legohuisje nabouwen en een hondje navouwen. Bij Verbanden gaat het kind een dierentuinspel spelen met Anne en Mo.

Bij elk van de drie activiteiten hoort een instructie voor de leerkracht. In de instructie staat onder meer precies vermeld welke materialen nodig zijn voor de activiteit. Dit betreffen bijvoorbeeld poppetjes van Anne en Mo, fiches, blanco kaartjes en dierenkaartjes. Deze zitten handig opgeborgen in de leerkrachtmap. Het is niet mogelijk om alle materialen die nodig zijn voor de activiteiten te leveren in de leerkrachtmap. Bij de activiteit Bewerkingen en Meetkunde vragen we de leerkrachten om een deel van de benodigde materialen te verzamelen, zoals kralen, blokjes en legostenen, maar bijvoorbeeld ook een tafeltje en een stoeltje. Het gaat om materialen die beschikbaar zijn in scholen/kleutergroepen.

In de instructie staat de activiteit stap voor stap uitgelegd en wordt aangegeven wat de leerkracht moet zeggen. Daarnaast bevat elke activiteit een observatieformulier met de bijhorende observatiepunten en antwoordmogelijkheden. De leerkracht beoordeelt in welke mate het kind het beschreven gedrag in het observatiepunt laat zien met bijvoorbeeld nee, met hulp en ja. Zie hieronder voor een voorbeeld.

Start 2	Staafdiagram	nee	met hulp	ja
VB11	Kan informatie over de bananen aflezen uit het staafdiagram. Noemt het juiste aantal bananen (7) door te tellen of af te lezen a.d.h.v. de kolom met dobbelstenen of getallen.			

#### Voorbeeld observatiepunt uit Verbanden

De leerkracht vult tijdens de activiteit al zoveel mogelijk het bijbehorend observatieformulier in. Dit kan direct in Kleuter in beeld online, bijvoorbeeld met een iPad. De leerkracht kan er ook voor kiezen eerst het papieren observatieformulier in te vullen en de observaties op een later moment digitaal in te vullen. De observatiepunten zijn van verschillende moeilijkheid en bij het bepalen van de observatiepunten is er rekening mee gehouden dat de activiteit ingezet wordt bij zowel groep 1 als groep 2.

#### **Privacyverklaring en technische ondersteuning (KIB-rekenen)**

Cito houdt zich in alle gevallen aan de eisen die de Algemene verordening gegevensbescherming (AVG) aangaande de verwerking van persoonsgegevens stelt. Op de [website](#) van Cito staat de Verwerkersovereenkomst met bijbehorende Privacybijsluiters PO-VSO en de Beveiligingsbijlage. Deze gelden voor de uitgevoerde onderzoeken en het uiteindelijke product Kleuter in beeld - Rekenen.

Bij technische en inhoudelijke vragen rondom het product kan telefonisch contact opgenomen worden met Cito (026 – 352 11 11) of een mail gestuurd worden naar [kleuters@cito.nl](mailto:kleuters@cito.nl). De klant wordt dan zo snel als mogelijk geholpen door een medewerker van Cito.

## 10.2 Rapportage

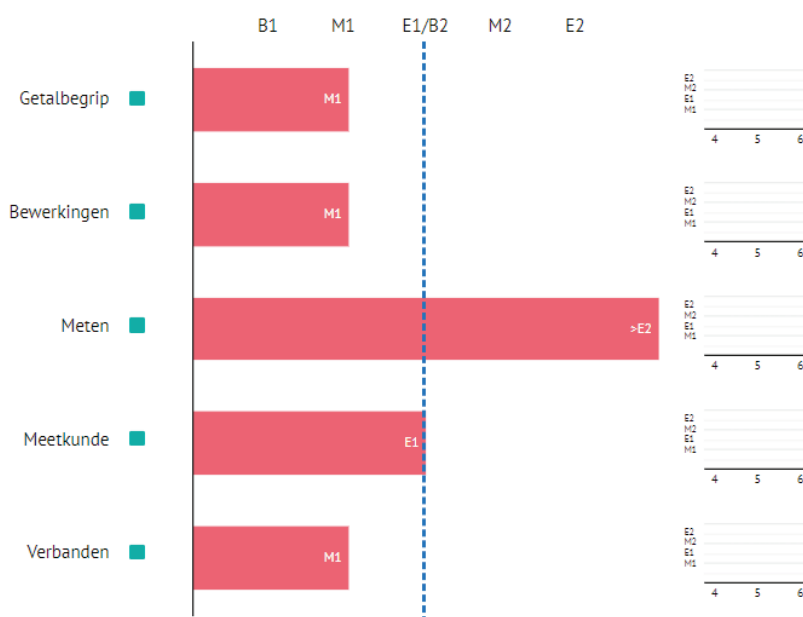
Een belangrijke wens vanuit het veld was een gebruikersvriendelijke, beknopte rapportage. Specifieke wensen daarbij waren:

- Niet alleen een resultaat op Rekenen als totaal, maar informatie over subdomeinen. Dit levert belangrijke informatie voor de leerkracht om het onderwijs op de kinderen af te stemmen.
- Signaal als het kind het boven verwachting doet en een signaal als een kind het onder verwachting doet.
- Inzicht in groei, hoe het kind zich ontwikkelt in de loop van de tijd.
- Rapportage per kind op het hoofddomein (zoals Rekenen), rapportage per groep en een overzicht van alle kinderen over meerdere hoofddomeinen (zoals Taal en Rekenen).

Bij de ontwikkeling van de rapportage is rekening gehouden met al deze wensen. Op basis van alle wensen is er gekozen voor drie rapportagevormen: kindrapportage, groepsrapportage en groepsoverzicht. Vervolgens zijn prototypes gemaakt en voorgelegd aan het veld. De feedback op de prototypes is meegenomen bij de ontwikkeling van de definitieve rapportagevormen. Het resultaat is een gebruikersvriendelijke, intuïtieve rapportage. Alle rapportagevormen kunnen opgevraagd worden via Kleuter in beeld online. Hierna volgt een voorbeeld en korte uitleg bij elk van de drie rapportagevormen.

### Kindrapportage





De kindrapportage geeft een beeld van alle vijf subdomeinen van Rekenen van een kind, zie het voorbeeld Kindrapportage. De stippelijijn geeft de observatieperiode aan. De balkjes geven het niveau per subdomein aan. Zo is in een oogopslag te zien op welke subdomeinen het kind boven, op of onder niveau zit. Ook is de groei per subdomein zichtbaar als er meerdere observatieperiodes zijn. Wanneer de digitale opdrachten gemaakt zijn, is het via de online kindrapportage mogelijk de gegeven antwoorden van een kind in te zien. Zo kun je zien welk soort opdrachten goed gaan en met welk soort opdrachten het kind nog wat meer moeite heeft.



### Voorbeeld Kindrapportage

Er zijn drie signalen voor het totaal van Rekenen als alle vijf subdomeinen zijn gemaakt, zie ook Tabel 10.1. De algoritmes voor deze signalen staan uitgewerkt in paragraaf 7.2. Het signaal met de hardloper betekent dat het kind het boven verwachting doet (uitstekend) en mogelijk extra uitdaging kan gebruiken. Het signaal met het plantje geeft aan dat het kind het volgens verwachting doet (ga zo door). En het oogje betekent dat het kind het onder verwachting doet (wees alert) en mogelijk wat extra hulp kan gebruiken.

Tabel 10.1 Betekenis signalen

-  boven verwachting
-  volgens verwachting
-  onder verwachting
-  geen signaal: niet alle 5 subdomeinen zijn geobserveerd

De leerkracht kan de kindrapportage opslaan als PDF en printen per kind of tegelijk voor meerdere kinderen. Om leerkrachten de mogelijkheid te geven de kindrapportage ook met ouders te delen, is ervoor gekozen een leerkrachtversie en een ouderversie beschikbaar te stellen. Er zijn drie verschillen tussen de leerkracht- en ouderversie. Bij de ouderversie zijn het totaalsignaal, de gevolgde route en persoonlijke notities niet zichtbaar, bij de leerkrachtversie wel. Leerkrachten gaven aan dat het totaalsignaal mogelijk gevoelig kan liggen bij ouders en waarschijnlijk meer vragen oproept dan beantwoordt. De gevolgde route (leerkracht- of kindroute) vergt de nodige uitleg en is voor ouders niet relevant. Het is niet wenselijk dat persoonlijke notities van leerkrachten voor ouders zichtbaar zijn. In de ouderrapportage staat dan heel duidelijk welk niveau een kind voor elk subdomein heeft en hoe het met de groei per subdomein staat. Dit rapport is niet bedoeld om naar ouders te sturen, maar om te bespreken tijdens bijvoorbeeld een tienminutengesprek. De leerkracht kan de rapportage dan toelichten en de gegevens bespreken in samenhang met bijvoorbeeld observaties in de klas en opbrengsten uit de methode.

### **Groepsrapportage**

In de groepsrapportage staan de resultaten van de observaties voor de kinderen uit een groep voor elk van de afgeronde subdomeinen, zie het voorbeeld Groepsrapportage. Ook is te zien of het resultaat gebaseerd is op de leerkrachtroute (1 blokje) of op de kindroute (2 blokjes). Het is mogelijk te ordenen op voornaam, achternaam en leeftijd. Bij het enveloppe is het mogelijk notities te maken over een bepaald subdomein of over één kind. De verschillende kleuren paars geven een indicatie van het niveau. Hoe lichter de kleur, hoe hoger het behaalde niveau. Op deze manier kan een leerkracht snel zien of een groepje kleuters bijvoorbeeld moeite heeft met een bepaald subdomein en juist op dat gebied wat extra uitleg en oefening nodig heeft.

Dashboard > Groepsrapportage

Rekenen GROEPSRAPPORTAGE GROEPSVOORTGANG PDF BEGIN 2020-2021

Resultaat	Getalbegrip	Bewerkingen	Meren	Meeikunde	Verbanden	Legenda
4% 1 van 25 56% 14 van 25 24% 6 van 25 16% 4 van 25	GB	BW	MT	MK	VB	Leerkracht-route Kindroute
Bekijk niet afgeronde observaties en opdrachten bij <a href="#">Groepsvoortgang</a> .						
Sorteer: ↑ Voornaam z-a						notities
Wandena Ammar 7 jaar 3 maanden	M1	M1	>E2	E1	M1	
Tiofilo Ortela 7 jaar 1 maand	M1	E2	<M1	M2	M1	
Sjafie Ahmed 7 jaar 0 maanden	M2	M2	<M1	E1	E1	

Voorbeeld Groepsrapportage

### Groepsoverzicht

Kleuter in beeld online bevat naast Rekenen ook andere domeinen. Het groepsoverzicht laat voor alle kinderen in een groep zien welke signalen ze hebben voor alle domeinen die zijn ingevuld.

BEGIN 2020-2021

Sorteer: ↓ Voornaam a-z	Sociaal gedrag	Welbevinden	Speelwerk-houding	Taal	Rekenen	Grove motoriek	Fijne motoriek
Wandena Ammar 7 jaar 3 maanden							
Tiofilo Ortela 7 jaar 1 maand							
Sjafie Ahmed 7 jaar 0 maanden							
Shakila Kamau 6 jaar 11 maand							
Sehri Azzai 6 jaar 3 maanden							

Voorbeeld Groepsoverzicht

## Appendix

Tabel 0.1 Gemiddelde vaardigheid op de subdomeinen in de kindroute per leeftijdscategorie en de significante verschillen tussen leeftijdscategorieën (sig)

Leeftijds-categorie	(0) ≤ 5 jaar				(1) 5 - 5,5				(2) 5,5 - 6				(3) > 6			
	Gem.	Sig.	N	Std. Dev.	Gem.	Sig.	N	Std. Dev.	Gem.	Sig.	N	Std. Dev.	Gem.	Sig.	N	Std. Dev.
<b>Getalbegrip</b>	0,09	1,2,3	177	0,43	0,37	0,2,3	232	0,41	0,56	0,1	256	0,45	0,53	0,1	104	0,38
<b>Bewerkingen</b>	0,35	2	33	0,41	0,54		28	0,36	0,76	0	34	0,38	0,64		18	0,38
<b>Meten</b>	0,07	1,2,3	171	0,41	0,40	0,2,3	211	0,44	0,56	0,1	239	0,48	0,61	0,1	102	0,47
<b>Meetkunde</b>	0,15	1,2,3	38	0,30	0,38	0,3	48	0,24	0,53	0	42	0,46	0,65	0,1	25	0,41
<b>Verbanden</b>	0,26	1,2,3	47	0,44	0,51	0,2	72	0,41	0,71	0,1	68	0,35	0,69	0	19	0,39

Tabel 0.2 Gemiddelde vaardigheid op de subdomeinen in de kindroute bij jongens en bij meisjes

Geslacht Subdomein	Jongens			Meisjes		
	Gem.	N	Stdev.	Gem.	N	Stdev.
<b>Getalbegrip</b>	0,45	64	0,47	0,36	61	0,47
<b>Bewerkingen</b>	0,65	19	0,41	0,56	18	0,42
<b>Meten</b>	0,43	64	0,45	0,37	61	0,50
<b>Meetkunde</b>	0,50	15	0,21	0,47	10	0,16
<b>Verbanden</b>	0,50	17	0,30	0,65	9	0,37

Tabel 0.3 Percentage kinderen van elke leeftijdscategorie in elk functioneringsniveau van Rekenen volgens de leerkrachtroute

Subdomein	Leeftijdscategorie	0 (<E1)	1 (E1)	2 (M2)	3 (E2)	4 (>E2)	Totaal	Aantal	
<b>Getalbegrip</b>	0	≤ 5 jaar	26,7%	55,0%	12,5%	5,0%	0,8%	100%	120
	1	5 – 5,5 jaar	4,8%	21,4%	44,6%	28,0%	1,2%	100%	168
	2	5,5 – 6 jaar	1,8%	11,0%	35,0%	40,5%	11,7%	100%	163
	3	> 6 jaar	1,4%	8,3%	43,1%	40,3%	6,9%	100%	72
<b>Bewerkingen</b>	0	≤ 5 jaar	53,8%	34,2%	6,0%	4,3%	1,7%	100%	117
	1	5 – 5,5 jaar	9,0%	25,7%	52,1%	12,0%	1,2%	100%	167
	2	5,5 – 6 jaar	2,5%	17,3%	48,1%	25,3%	6,8%	100%	162
	3	> 6 jaar	4,2%	12,5%	55,6%	23,6%	4,2%	100%	72
<b>Meten</b>	0	≤ 5 jaar	35,3%	52,1%	9,2%	3,4%	0,0%	100%	119
	1	5 – 5,5 jaar	3,0%	23,1%	60,4%	13,6%	0,0%	100%	169
	2	5,5 – 6 jaar	0,6%	12,3%	56,8%	26,5%	3,7%	100%	162
	3	> 6 jaar	1,4%	12,5%	63,9%	19,4%	2,8%	100%	72
<b>Meetkunde</b>	0	≤ 5 jaar	23,9%	54,7%	12,8%	8,5%	0,0%	100%	117
	1	5 – 5,5 jaar	4,2%	17,9%	50,0%	28,0%	0,0%	100%	168
	2	5,5 – 6 jaar	0,0%	5,6%	46,3%	44,4%	3,7%	100%	162
	3	> 6 jaar	1,4%	2,8%	59,7%	31,9%	4,2%	100%	72
<b>Verbanden</b>	0	≤ 5 jaar	47,9%	41,0%	4,3%	6,0%	0,9%	100%	117
	1	5 – 5,5 jaar	7,1%	19,6%	48,8%	22,6%	1,8%	100%	168
	2	5,5 – 6 jaar	0,6%	13,7%	43,5%	36,6%	5,6%	100%	161
	3	> 6 jaar	1,4%	15,3%	51,4%	26,4%	5,6%	100%	72



Tabel 0.4 Percentage jongens en meisjes in elk functioneringsniveau van de leerkrachtroute

Subdomein	Geslacht	0 (<E1)	1 (E1)	2 (M2)	3 (E2)	4 (>E2)	Totaal	Aantal
<b>Getalbegrip</b>	Jongens	3,9%	29,4%	45,1%	13,7%	7,8%	100%	51
	Meisjes	4,5%	20,5%	61,4%	13,6%	0,0%	100%	44
<b>Bewerkingen</b>	Jongens	16,0%	28,0%	36,0%	14,0%	6,0%	100%	50
	Meisjes	7,0%	30,2%	58,1%	4,7%	0,0%	100%	43
<b>Meten</b>	Jongens	9,8%	29,4%	43,1%	15,7%	2,0%	100%	51
	Meisjes	6,7%	22,2%	64,4%	6,7%	0,0%	100%	45
<b>Meetkunde</b>	Jongens	6,0%	24,0%	56,0%	14,0%	*	100%	50
	Meisjes	11,4%	9,1%	63,6%	15,9%	*	100%	44
<b>Verbanden</b>	Jongens	10,0%	32,0%	40,0%	14,0%	4,0%	100%	50
	Meisjes	13,6%	18,2%	54,5%	11,4%	2,3%	100%	44

\*Bij Meetkunde had slechts één leerling Boven E2 en het geslacht van die leerling is onbekend.



## Literatuur

- Anastasi, A. (1986). Evolving concepts of test validation. *Annual Review of Psychology*, 37, 1-15.
- Borghouts, C. & Veltman, A. (2012/2013). *De Vertaalcirkel bij kleuters. Werken aan begrip en inzicht bij (zwakke) rekenaars*. Volgens Bartjens, jaargang 33, nr. 3, 22-25.
- Boxtel, H. van & Hemker, B.T. (2009). *Wetenschappelijke verantwoording van de Intelligentietest Eindtoets Basisonderwijs*. Arnhem: Cito.
- Bredenkamp, S. & Rosegrant, T. (1995). *Reaching potentials: Transforming early childhood curriculum and assessment. Volume 2*. Washington, DC: National Association for the Education of Young Children.
- Eggen, T.J.H.M. (1993). Itemresponstheorie en onvolledige gegevens. In: T.J.H.M. Eggen & P.F. Sanders (red.). *Psychometrie in de praktijk* (pp. 239-284). Arnhem: Cito.
- Evers, A., Lucassen, W., Meijer, R. & Sijtsma, K. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP/COTAN.
- Expertgroep Doorlopende Leerlijnen Taal en Rekenen (2008). *Over de drempels met rekenen. Consolideren, onderhouden, gebruiken en verdiepen*. Enschede: SLO.
- Feskens, R., Keuning, J., Til, A. van & Verheyen, R. (2014). *Prestatiestandaarden voor ERK in het examenjaar: Een internationaal ijkingsonderzoek*. Arnhem: Cito.
- Finn, R.H. (1970). A note on estimating the reliability of categorical data. *Educational and Psychological Measurement*, 30, 71-76. doi:10.1177/001316447003000106
- Geary, D.C. (1990). A componential analysis of an early learning deficit in mathematics. *Journal of Experimental Child Psychology*, 49, 363-383.
- Glas, C.A.W. & Verhelst, N.D. (1993). Een overzicht van itemresponsmodellen. In: T.J.H.M. Eggen & P.F. Sanders (red.). *Psychometrie in de praktijk* (pp. 179-238). Arnhem: Cito.
- Glas, C.A.W. (1988). The Rasch model and multi-stage testing. *Journal of Educational Statistics*, 13, 45-52.
- Groenestijn, M. van, Borghouts, C. & Janssen, C. (2011). *Protocol Ernstige RekenWiskunde-problemen en Dyscalculie. BAO, SBO, SO*. Assen: Van Gorcum.
- Guddemi, M. & Case, B.J. (2004). *Assessing young children*. Assessment report, Pearson Education.
- Heuvel-Panhuizen, M. van den & Buys, K. (2009). *TAL - Jonge kinderen leren meten en meetkunde*. Groningen: Wolters-Noordhoff.
- Heuvel-Panhuizen, M. van den, Veltman, A., Janssen, C. & Hochstenbach, J. (2009). Meetkunde in groep 1 en 2. In: M. van den Heuvel-Panhuizen & K. Buys (2009). *TAL - Jonge kinderen leren meten en meetkunde*. Groningen: Wolters-Noordhoff.
- Hutten, O., Bergh, J. van den, Brom-Snijders, P. van den & Zanten, M. van (2014). *Metten en meetkunde. Reken-wiskundendidactiek*. Amersfoort: ThiemeMeulenhoff.

- Janssen, C. & Kroesbergen, E. (2012). Profielschets 1: Leerlingen met dyscalculie. In: N. Boswinkel (red.). *Passende perspectieven rekenen, profielschetsen* (pp. 7-10). Enschede: SLO.
- Keuning, J., Straat, J.H. & Feskens, R.C.W. (2017). The Data-Driven Direct Consensus (3DC) procedure: A new approach to standard setting. In: S. Blömeke & J. Gustafsson (Eds.). *Standard setting in education: the Nordic countries in an international perspective* (pp. 263-278). Cham, Switzerland: Springer International Publishing.
- Koerhuis, I. (2010). *LOVS Rekenen voor kleuters groep 1 en 2*. Arnhem: Cito.
- Koerhuis, I. & Keuning, J. (2011). *Wetenschappelijke verantwoording van de toetsen Rekenen voor kleuters*. Arnhem: Cito.
- Kolkman, M.E. & Kroesbergen, E.H. (2014). Rekenen met Kleuters. Voorwaarden voor een goede rekenstart in groep 3. *Het Jonge Kind*, vol. 41, januari 2014, 4-7.
- Landis, J.R. & Koch, G.G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33, 159-174.
- Langberg, M., Zanten, M. van & Boswinkel, N. (2014). *Dyslexie en dyscalculie: een kwestie van aanpakken. Verkenning naar aanleiding van motie 17*. Enschede: SLO.
- Lee, W., Hanson, B.A. & Brennan, R.L. (2002). Estimating Consistency and Accuracy Indices for Multiple Classifications. *Applied Psychological Measurement*, 26, 412-432.
- Luit, J.E.H. van (2009). *De ontwikkeling van tellen en getalbegrip bij kleuters*. Utrecht: Projectbureau Kwaliteit PO Raad.
- Ministerie van Onderwijs, Cultuur en Wetenschap (2009). *Referentiekader taal en rekenen. De referentieniveaus*. Enschede: SLO.
- Noteboom, A., Aartsen, A. & Lit, S. (2017). *Tussendoelen Rekenen-Wiskunde voor het primair onderwijs. Uitwerking van rekendoelen voor groep 2 tot en met 8 op weg naar streefniveau 1S*. Enschede: SLO.
- Oonk, W., Keijzer, R., Lit, S.A., Barth, F., Engelsens, J.F.M. den, Lek, A. & Waveren Hogervorst, C. van (2015). *Rekenen-wiskunde in de praktijk. Kerninzichten*. Groningen/Houten: Noordhoff Uitgevers.
- Pillner, A. (1969). *Estimation of number of grades to be awarded in an examination by consideration of its reliability coefficient*. Edinburgh: The Godfrey Thomson Unit for Educational Research.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Nielsen & Lydiche.
- Schouwstra, S. (2000). *On testing plausible threats to construct validity*. Proefschrift. Amsterdam: Universiteit van Amsterdam.
- Schouwstra, S. (2018). *De ontwikkeling van de diagnostische tussentijdse toets (2012-2017)*. Arnhem: Cito.
- SLO (2018). *Inhoudskaart Rekenen-Wiskunde jonge kind (fase 1). Leerplankader jonge kind*. Geraadpleegd van [www.slo.nl](http://www.slo.nl) <https://www.slo.nl/thema/meer/jonge-kind/doelen-jonge-kind/>
- Staphorsius, G. (1994). *Leesbaarheid en leesvaardigheid: de ontwikkeling van een domeingericht meetinstrument*. Enschede: Universiteit Twente.

- Toll, S.W.M. & Luit, J.E.H. van (2014). Structurele ondersteuning aan kleuters met een achterstand in getalbegrip. *Pedagogische Studiën*, 91, 82-96.
- Treffers, A., Heuvel-Panhuizen, M. van den & Buys, K. (2009). *TAL - Jonge kinderen leren rekenen*. Groningen: Wolters-Noordhoff.
- Veltman, A. & Heuvel-Panhuizen, M. van den (2010). *Rekenen met hele getallen op de basisschool: tussendoelen annex leerlijnen*. Groningen: Noordhoff Uitgevers.
- Verhelst, N.D. & Glas, C.A.W. (1995). The one parameter logistic model. In: G.H. Fischer & I.W. Molenaar (Eds.). *Rasch models: foundations, recent developments and applications* (pp. 215-239). New York: Springer.
- Verhelst, N.D. & Kleintjes, F.G.M. (1993). Toepassingen van itemresponsetheorie. In: T.J.H.M. Eggen en P.F. Sanders (red.). *Psychometrie in de praktijk*. (pp. 285-310). Arnhem: Cito.
- Verhelst, N.D. (1992). *Het één parameter model (OPLM): een theoretische inleiding en een handleiding bij het computerprogramma*. Arnhem: Cito.
- Verhelst, N.D. (1993). Itemresponstheorie. In: T.J.H.M. Eggen & P.F. Sanders (red.). *Psychometrie in de praktijk* (pp. 83-178). Arnhem: Cito.
- Verhelst, N.D., Glas, C.A.W. & Verstralen, H.H.F.M. (1995). *OPLM: One Parameter Logistic Model. Computer program and manual*. Arnhem: Cito.
- Verhelst, N.D., Verstralen, H.H.F.M. & Eggen, T.H.J.M. (1991). Finding starting values for the itemparameters and suitable discrimination indices in the one-parameter logistic model. *Measurement and Research Department Reports 91-10*. Arnhem: Cito.
- Vugt, J. van & Wösten, A. (2015). *Rekenen: een hele opgave. Deel 1*. Amersfoort: ThiemeMeulenhoff.
- Wheadon, C. & Stockford, I. (2010). *Classification accuracy and consistency in GCSE and A level examinations offered by the Assessment and Qualifications Alliance (AQA)*. Ofqual/11/4823. <https://pdfs.semanticscholar.org/d725/366c7c074ddff4faa9f8bf115d825b900968.pdf>

**Cito**

Amsterdamseweg 13  
6814 CM Arnhem  
Postbus 1034  
6801 MG Arnhem  
T (026) 352 11 11

