

# Wetenschappelijke verantwoording Kleuter in beeld - Taal

Sanneke Schouwstra en Judith Vloedgraven

Taal





# **Wetenschappelijke verantwoording**

## **Kleuter in beeld - Taal**

Sanneke Schouwstra

Judith Vloedgraven

Met medewerking van Ron Engelen, Inge Groenen, Linda Horsels, Annet Mennink, Lonneke Nikkels

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	12
2.4	Doel	13
<b>3</b>	<b>Beschrijving meetinstrument</b>	<b>15</b>
3.1	Opbouw van het meetinstrument	15
3.2	Ontwikkel- en constructieproces in co-creatie	16
3.3	Operationalisatie van de inhoud	18
<b>4</b>	<b>Dataverzameling</b>	<b>29</b>
4.1	Proefonderzoek	29
4.2	Steekproef kwaliteitsonderzoek	30
4.3	Ontwerp van het kwaliteitsonderzoek	33
4.4	Verwerking en analyses	34
4.5	Samenstelling versies uitgave	36
<b>5</b>	<b>Kalibratie kindroute</b>	<b>45</b>
5.1	Bank met opdrachten en observatiepunten	45
5.2	Meetmodel	46
5.3	Kalibratie	49
5.4	Toetsing van het IRT-model	50
<b>6</b>	<b>Standaardbepaling voor de kindroute</b>	<b>57</b>
6.1	Doel van standaardbepaling	57
6.2	Methode en experts	57
6.3	Procedure	57
6.4	Materiaal	59
6.5	Definitieve expertoordelen	60
6.6	Conclusie	63
<b>7</b>	<b>Beschrijving algoritmes</b>	<b>67</b>
7.1	Algoritme voor het rapporteren van het functioneringsniveau op subdomeinen	67
7.2	Algoritme voor een signaal voor Taal	72
<b>8</b>	<b>Betrouwbaarheid en nauwkeurigheid</b>	<b>73</b>
8.1	Interbeoordelaarsbetrouwbaarheid activiteiten	73
8.2	Betrouwbaarheid	75
8.3	Nauwkeurigheid	77
<b>9</b>	<b>Validiteit</b>	<b>85</b>
9.1	Inhoudsvaliditeit	85
9.2	Begripsvaliditeit	86

**10 Afname en rapportage 99**

10.1 Afname 99

10.2 Rapportage 101

**Literatuur 105**

# 1 Inleiding

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

Kleuter in beeld - Taal is een instrument om de vaardigheid van kleuters op het gebied van taal 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 - Taal: 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)

## *Taal*

- Inhoudsverantwoording Kleuter in beeld - Taal

Om Kleuter in beeld - Taal 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 - Taal 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 - Taal is echt ontwikkeld in co-creatie met het veld. De wensen uit het veld hebben centraal gestaan bij de ontwikkeling. Er zijn onder andere focus-, screenings- en expertbijeenkomsten georganiseerd, zodat leerkrachten, intern begeleiders en andere experts feedback konden geven op de concepten en de inhoud. Er was een speciale online community, waarbij mensen uit het veld konden reageren op stellingen en zo hun voorkeuren konden aangeven. 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 tientallen 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 - Taal 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 onder andere SLO en het Expertisecentrum Nederlands;
- de mening van leerkrachten en taalspecialisten;
- 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 - Taal 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 - Taal. 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 indirecte observatie observeert de leerkracht de gevolgen of resultaten van het gedrag en niet het gedrag zelf. 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 Spreken wordt aan de kinderen gevraagd een verhaaltje te vertellen bij een strip en bij de activiteit Boek en verhaal leest de leerkracht een prentenboek voor.

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 op het gebied van taal van kleuters op een objectieve manier in beeld te brengen. Er is voor gekozen om Taal (hoofddomein) op te delen in Mondelinge taal (domein) en Beginnende geletterdheid (domein), vanwege de volgende redenen:

- Mondeling onderwijs en schriftelijk onderwijs wordt in de kerndoelen uit elkaar getrokken (Ministerie OCW, 2006).
- De keuze voor de verschillende inhoudsgebieden binnen taal is grotendeels gebaseerd op de inhoudskaart Taal (Nederlands) jonge kind (fase 1) van SLO (2018). De inhoudskaart van SLO hanteert ook de indeling in Mondelinge taal en Beginnende geletterdheid.
- Mondelinge taalvaardigheid vormt de basis voor schriftelijke taalvaardigheid; schriftelijke taalvaardigheid is secundair aan de mondelinge taalvaardigheid (Shanahan, MacArthur, Graham & Fitzgerald, 2006). Bij schriftelijke taalvaardigheid komt in tegenstelling tot mondelinge taalvaardigheid altijd expliciete instructie aan te pas.
- Sterke mondelinge taalvaardigheden gaan vaak samen met sterke lees- en schrijfvaardigheden (Shanahan et al., 2006). Sommige leerlingen komen wel tot spreken, maar niet tot schrijven (bijvoorbeeld leerlingen met een zml-indicatie). Relatief goede prestaties op het domein Mondelinge taalvaardigheid kunnen in theorie relatief slechte prestaties op het domein Beginnende geletterdheid compenseren. Ook het omgekeerde kan voorkomen. Wanneer beide domeinen samengevoegd worden en er maar één score voor Taal gerapporteerd zou worden, is er het risico van vals-positieven (leerlingen die niet gesignaleerd worden als zwak op het gebied van beginnende geletterdheid of mondelinge taal). Het is daarom belangrijk de vaardigheid op beide domeinen te meten en niet te kiezen voor een score op Taal algemeen.

Diverse onderzoeken laten zien dat vroege signalering van achterstanden op het gebied van taal van groot belang is (Carmiggelt, Uilenburg, Romeijn, Stam-van den Doel, & Pijpers, 2013; Zijlstra, 2015). Dan kan een kind ook tijdig de juiste ondersteuning krijgen, waardoor problemen met bijvoorbeeld communicatie, leren en gedrag worden voorkomen ([www.tosinbeeld.nl](http://www.tosinbeeld.nl)). Denk hierbij aan intensievere begeleiding vanuit school, maar ook aan doorverwijzing naar een taalspecialist, zoals een logopedist. Binnen de jeugdgezondheidszorg is het signaleren van een taalontwikkelingsstoornis (TOS) een van de aandachtspunten. Toch kan het gebeuren dat een kind niet in de voorschoolse periode gesignaleerd wordt. Daarom ontwikkelde Siméa een vragenlijst met handreikingen voor leerkrachten. Deze vragenlijst is bedoeld om in te vullen voor kinderen in de leeftijd van 4 tot en met 6 jaar. De leerkracht beantwoordt vragen op het gebied van spraak, taal en communicatieve redzaamheid. Bijvoorbeeld: Kunt u ongeveer de helft van de spraak van het kind verstaan? Kan het kind een mondeling aangeboden instructie uitvoeren? In de handleiding van Kleuter in beeld - Taal verwijzen we naar deze vragenlijst. Wanneer een kind (sterk) uitvalt op bepaalde subdomeinen is het raadzaam om deze vragenlijst in te vullen. Dit helpt bij het vroegtijdig opsporen van kinderen met een TOS.

Onderzoeken waarin de ontwikkeling van leerlingen langdurig gevolgd is, hebben aangetoond dat vooral een lage letterkennis en daarnaast een zwak fonologisch bewustzijn in groep 2 belangrijke voorspellers zijn van latere leesproblemen (Boets et al., 2010; Van der Leij, 2013). Gezien het cruciale belang van het preventief aanpakken van leesproblemen, stelt Van der Leij (2016) dat al in groep 2 onderscheiden moet worden welke kinderen risico lopen op het signaleren van leesproblemen. In een onderzoek van de Universiteit van Amsterdam (Zijlstra, 2015) werd de effectieve implementatie van een computerondersteund programma onderzocht dat vanaf groep 2 werd ingezet. Het onderzoek heeft veelbelovende resultaten opgeleverd als het gaat om het verminderen van ernstige leesproblemen en daaruit voortvloeiende secundaire problematiek. Daarom is het ook zo belangrijk dat kleuters die moeite hebben met bepaalde aspecten van taal tijdig gesignaleerd worden. Er is dan ook voor gekozen zowel Mondelinge taal als Beginnende geletterdheid verder op te delen in subdomeinen. In de rapportage staat niet het niveau op Mondelinge taal en Beginnende geletterdheid centraal, maar het niveau op elk van de subdomeinen. Dit geeft de leerkracht gerichtere informatie voor het handelen.

Zowel Mondelinge taal als Beginnende geletterdheid is onderverdeeld in vier subdomeinen. Er is gekozen voor de meest onderscheidende subdomeinen en onderdelen, gebaseerd op onder andere de inhoudskaart Taal van SLO. Alle doelen uit de inhoudskaart aan bod laten komen, is te omvangrijk en leidt mogelijk tot een gebruikersonvriendelijk instrument. De definitieve selectie van subdomeinen, onderdelen en observatiepunten is gebaseerd op literatuur. Ook is de selectie van de verschillende inhoudsvoorwerpen aan leerkrachten en intern begeleiders en hun feedback is ook meegenomen bij de selectie. In Tabel 2.1 staat een overzicht van de acht subdomeinen met hun bijbehorende onderdelen.

#### *Keuzes bij Mondelinge taal: subdomeinen*

Bij Mondelinge taal is de indeling van de inhoudskaart SLO Nederlandse taal – jonge kind fase 1 (2018) gevolgd. Hierin worden vier subdomeinen onderscheiden, namelijk 1) Woordenschat en woordgebruik, 2) Spreken (vloeiend en verstaanbaar), 3) Luisteren en 4) Gesprekken voeren.

Deze indeling wijkt enigszins af van de uitwerking van het referentiekader Nederlandse taal (Van der Beek, Hoogeveen, & Prenger, 2015). 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, 2008b). In de uitwerking onderscheiden Van der Beek et al. geen vier, maar drie subdomeinen, namelijk 1) Gesprekken, 2) Luisteren en 3) Spreken. Woordenschat is hier ondergebracht bij zowel het domein Gesprekken als bij Luisteren. 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.

#### *Keuzes bij Mondelinge taal: onderdelen*

Onder elk subdomein kunnen verschillende, uiteenlopende leerdoelen vallen. Naast de inhoudskaart van SLO is aanvullende literatuur geraadpleegd om de leerdoelen in kaart te brengen, zoals de Leerstoflijnen mondelinge taalvaardigheid beschreven (Van der Beek et al, 2015), Opbrengstgericht werken aan mondelinge taalvaardigheid (Gijsel & Van Druenen, 2011) en Praatjes peilen (SLO, 2014). Dit mondde uit in een lijst met 35 potentiële leerdoelen. Om de belastbaarheid van zowel de leerkracht als de kleuter te beperken, is voor het instrument Kleuter in beeld - Taal een selectie gemaakt. Hiervoor vroegen we een groepje van zes professionals, allen werkzaam in het kleuteronderwijs, om maximaal vijftien leerdoelen te selecteren. Welke aspecten vinden zij belangrijk om te monitoren bij hun kleuters? De punten die door minimaal vier van de zes aanwezigen zijn gekozen, bleken volgens de groep goed weer te geven waar de focus ligt in het taalonderwijs, zoals receptieve en productieve woordenschat, zinsbegrip, verhaalbegrip, verstaanbaarheid, zinsvorming, vrijuit spreken en vasthouden van het gespreksonderwerp. Doelen gericht op intonatie, non-verbale communicatie en vloeiendheid van het spreken werden bijvoorbeeld niet of nauwelijks gekozen.

In een volgende fase is aan twee experts deze selectie voorgelegd (een onderzoeker op het gebied van taalonderwijs en een medewerker van SLO). Op basis van onder meer de feedback van deze professionals is het observatiepunt 'vrijuit spreken' verwijderd. Een expert merkte bijvoorbeeld op: 'Let op het cruciale

verschil tussen 'kunnen' en 'willen'. Bij spreken manifesteert het verschil zich in 'praters' (extraverte kinderen) en 'denkers' (stille of introverte kinderen) en daar tussen in: de ambiverte kinderen (afhankelijk van de context/setting).'

Met behulp van deze input is bekeken welke doelen onder welk subdomein vallen (Woordenschat, Luisteren, Spreken of Gesprekken). Zo is bijvoorbeeld:

- zinsbegrip en verhaalbegrip samengenomen bij het subdomein (begrijpend) Luisteren;
- woorden en zinnen vormen samengenomen in het onderdeel Grammaticale beheersing bij het subdomein Spreken;
- het vasthouden van het gespreksonderwerp als één van de onderdelen bij het subdomein Gesprekken opgenomen.

Receptieve woordenschat en productieve woordenschat is ondergebracht in verschillende subdomeinen. Hierbij hebben we de vorm van de opdrachten zwaarder laten wegen. De receptieve woordenschat kan goed in kaart worden gebracht door een kind een woord te geven en te vragen op welk van de drie plaatjes hij dit woord ziet. Bij productieve woordenschat past deze vorm niet, en is voor observatiepunten bij een activiteit gekozen. Deze twee vormen laten zich niet goed op één schaal plaatsen. Daarom is de productieve woordenschat ondergebracht bij het subdomein Spreken. Dit is een goede aanvulling op de onderdelen verstaanbaarheid, grammaticale beheersing en verhalen vertellen, omdat met deze vier onderdelen niet alleen iets gezegd kan worden over hoe een kind iets zegt, maar ook wát een kind zegt.

#### *Keuzes bij Beginnende geletterdheid: subdomeinen*

De inhoudskaart van SLO (2018) onderscheidt vier subdomeinen binnen Beginnende geletterdheid: Oriëntatie op verhaal en tekst, Oriëntatie op geschreven taal, Fonemisch bewustzijn en alfabetisch principe en Leesplezier. Bij Kleuter in beeld - Taal is gekozen voor een vergelijkbare indeling. Omdat ook boekoriëntatie is opgenomen, een onderdeel van de tussendoelen Beginnende geletterdheid (De Blauw, Boland, Vernooij, & Van het Zandt, 2013), is gekozen voor de term Oriëntatie op boek en verhaal in plaats van Oriëntatie op verhaal en tekst. Verder is Oriëntatie op geschreven taal opgenomen. Het subdomein Fonemisch bewustzijn en alfabetisch principe is opgesplitst. Wij onderscheiden het subdomein Fonologisch bewustzijn, een bredere term dan fonemisch bewustzijn, omdat we ook het onderdeel Rijm hebben opgenomen. Ten slotte is Letterkennis opgenomen, wat valt onder het subdomein Alfabetisch principe. Ook in het Protocol preventie van leesproblemen – groep 1 en 2 (Van Druenen, Scheltinga, Wentink, & Verhoeven, 2017) en in de tussendoelen Beginnende geletterdheid (De Blauw et al., 2013) wordt het belang van deze vier subdomeinen onderschreven.

Onderzoek heeft aangetoond dat twee vaardigheden op het gebied van Beginnende geletterdheid, fonologisch bewustzijn en letterkennis, de belangrijkste factoren zijn die ten grondslag liggen aan de ontwikkeling van leesvaardigheid (Boets et al., 2010; Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004; Van der Leij, 2013). Deze twee vaardigheden komen ook voor in de Inhoudskaart Taal van SLO, maar dan onder één noemer. Vanwege hun belangrijke voorspellende waarde voor het latere lezen, is ervoor gekozen deze op te nemen als twee aparte subdomeinen.

De inhoudskaart noemt verder nog Leesplezier. Er is bewust voor gekozen Leesplezier niet op te nemen in Kleuter in beeld - Taal. Bij Leesplezier is er geen sprake van een toenemende vaardigheid, een startend kind in groep 1 heeft niet altijd minder leesplezier dan een kind eind groep 2. Leesplezier geeft meer de motivatie aan dan de vaardigheid op het gebied van taal. Verder is het van belang dat de te meten vaardigheden zo objectief mogelijk zijn. Leesplezier is subjectiever van aard. Vanwege deze redenen is Leesplezier niet opgenomen in Kleuter in beeld - Taal.

#### *Keuzes bij Beginnende geletterdheid: onderdelen*

Het Expertisecentrum Nederlands publiceerde verschillende uitgaves over de inhoud en ontwikkeling van Beginnende geletterdheid: Tussendoelen beginnende geletterdheid (De Blauw et al., 2013), Protocol preventie van leesproblemen groep 1 en 2 (Van Druenen et al., 2017) en De Taallijn (Van Elsäcker, Van der Beek, Hillen, & Peters, 2013). Hierbij worden verschillende aspecten onderscheiden die vallen

onder het domein Beginnende geletterdheid: boekoriëntatie, verhaalbegrip, taalbewustzijn (waaronder fonologisch bewustzijn), alfabetisch principe (waaronder letterkennis), functioneel lezen en schrijven en technisch lezen en schrijven. Oriëntatie op boek en verhaal is daarom opgedeeld in de onderdelen Boekoriëntatie en Verhaalbegrip. Boekoriëntatie gaat om de vorm van een boek en begrippen die daarbij worden gebruikt. Dit onderdeel komt aan de orde voorafgaand aan het lezen van een verhaal. Verhaalbegrip gaat om de inhoud en begrip van het verhaal. Dit komt aan de orde tijdens en na het lezen van een verhaal. Deze tweedeling sluit zo goed aan bij het lezen van een prentenboek, zoals dat gaat in een kleutergroep.

Oriëntatie op geschreven taal is opgedeeld in de onderdelen Voorbereidend lezen en Voorbereidend schrijven, omdat het Expertisecentrum Nederlands de indeling Lezen en Schrijven hanteert. Omdat deze vaardigheden bij jonge kinderen uit groep 1 en 2 oriënterend en beginnend zijn, is gekozen voor de termen Voorbereidend lezen en Voorbereidend schrijven. Voorbereidend schrijven gaat niet over de vaardigheid om schriftelijk precies de juiste lettervormen te kunnen maken, maar om (delen van) een woord met letters (potlood, stempels) op papier te kunnen zetten. Dit is gelijk aan de uitwerking van de leerdoelen in de (aan het begin van deze paragraaf genoemde) literatuur. Zowel Voorbereidend lezen als Voorbereidend schrijven bevatten leerdoelen die door het Expertisecentrum Nederlands worden genoemd onder Alfabetisch principe en Technisch lezen en schrijven. Dit zijn doelen die in de literatuur gelden voor groep 1 tot en met 3. Door deze doelen op te nemen kunnen ook kinderen met een ontwikkelingsvoorsprong gevolgd worden.

De selectie van onderdelen en leerdoelen is ook voorgelegd aan leerkrachten en intern begeleiders en aan twee experts Jonge Kind (medewerker Expertisecentrum Nederlands en een onderzoeker/adviseur jonge kind). Daarmee zijn de leerdoelen en observatiepunten verder aangescherpt, bijvoorbeeld voor het observatiepunt voorbereidend lezen in de activiteit Oriëntatie op geschreven taal. Experts geven aan dat de aandacht bij kleuters eind groep 2 vooral ligt op het herkennen en benoemen van letters, en niet op het (globaal) lezen. Daarom gaan de meeste observatiepunten in de activiteit over het herkennen en benoemen van letters.

Binnen Fonologisch bewustzijn is gezocht naar opdrachten die goed de vaardigheid fonologisch bewustzijn weergeven en die latere leesvaardigheid kunnen voorspellen. Tevens is gezocht naar opdrachten die verschillen in moeilijkheid. Op basis van alle inzichten is in eerste instantie gekozen voor opdrachten rijm, beginfoneem, synthese en deletie (Van Druenen et al., 2017; Vloedgraven, 2008). Tijdens focus- en expertbijeenkomsten met onderwijsprofessionals gaven mensen uit het veld aan dat deletie niet of nauwelijks voorkomt in het aanbod aan kleuters. Bovendien bleek uit het proefonderzoek dat het onderdeel deletie te moeilijk en te weinig onderscheidend was (zie ook hoofdstuk 4 Dataverzameling). Daarom is deletie niet meer opgenomen in het kwaliteitsonderzoek en de uitgave.

Tabel 2.1 Subdomeinen en onderdelen van *Kleuter in beeld - Taal*

	Subdomein	Onderdelen
Mondelinge Taal	<b>Woordenschat</b>	- Receptieve woordenschat
	<b>Luisteren</b>	- Begrijpend luisteren
	<b>Spreken</b>	- Verstaanbaarheid - Productieve woordenschat - Grammaticale beheersing - Verhalen vertellen
	<b>Gesprekken</b>	- Vasthouden van gespreksonderwerp - Eenvoudige gespreksregels
Beginnende geletterdheid	<b>Fonologisch bewustzijn</b>	- Rijm - Beginfoneem - Synthese
	<b>Letterkennis</b>	- Receptieve letterkennis
	<b>Oriëntatie op boek en verhaal</b>	- Boekoriëntatie - Verhaalbegrip
	<b>Oriëntatie op geschreven taal</b>	- Voorbereidend lezen - Voorbereidend schrijven

## 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 - Taal wordt zowel een beroep gedaan op het handelen (activiteiten) als op het platte vlak (opdrachten).

## 2.4 Doel

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

- Niveaubepaling: het bepalen van het niveau voor de acht subdomeinen binnen Taal  
In de online omgeving van Kleuter in beeld - Taal staat in de kindrapportage het behaalde niveau op elk van de acht subdomeinen. In de groepsrapportage staat een overzicht van het niveau op de acht 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 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 kleuter in de loop van de kleuterperiode te zien. Het gaat dan om een vergelijking van de resultaten van dezelfde leerling 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 elk domein (Mondelinge taal en Beginnende geletterdheid) opgevat als een unidimensionale vaardigheid, of 'latente trek'. Hierdoor ontstaan twee vaardigheidsschalen. 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 Mondelinge taal en één vaardigheidsschaal Beginnende geletterdheid 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 niveaueversies 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 niveaueversie 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 acht subdomeinen, in plaats van alleen een resultaat op het totaal van Taal. 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 bijvoorbeeld bij Fonologisch bewustzijn in een oogopslag zien of een kind bijvoorbeeld meer moeite heeft met een of meer van de onderdelen. Bij Letterkennis kun je zien welke letters het kind al wel en niet kent of dat het

kind met een bepaalde categorie, zoals de lange klanken, 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 - Taal 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 acht 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 8 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 - Taal 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 vier subdomeinen drie verschillende niveauversies beschikbaar. De leerkracht stelt per subdomein vast welk niveau het beste bij het kind past.

★ = niveau groep 1

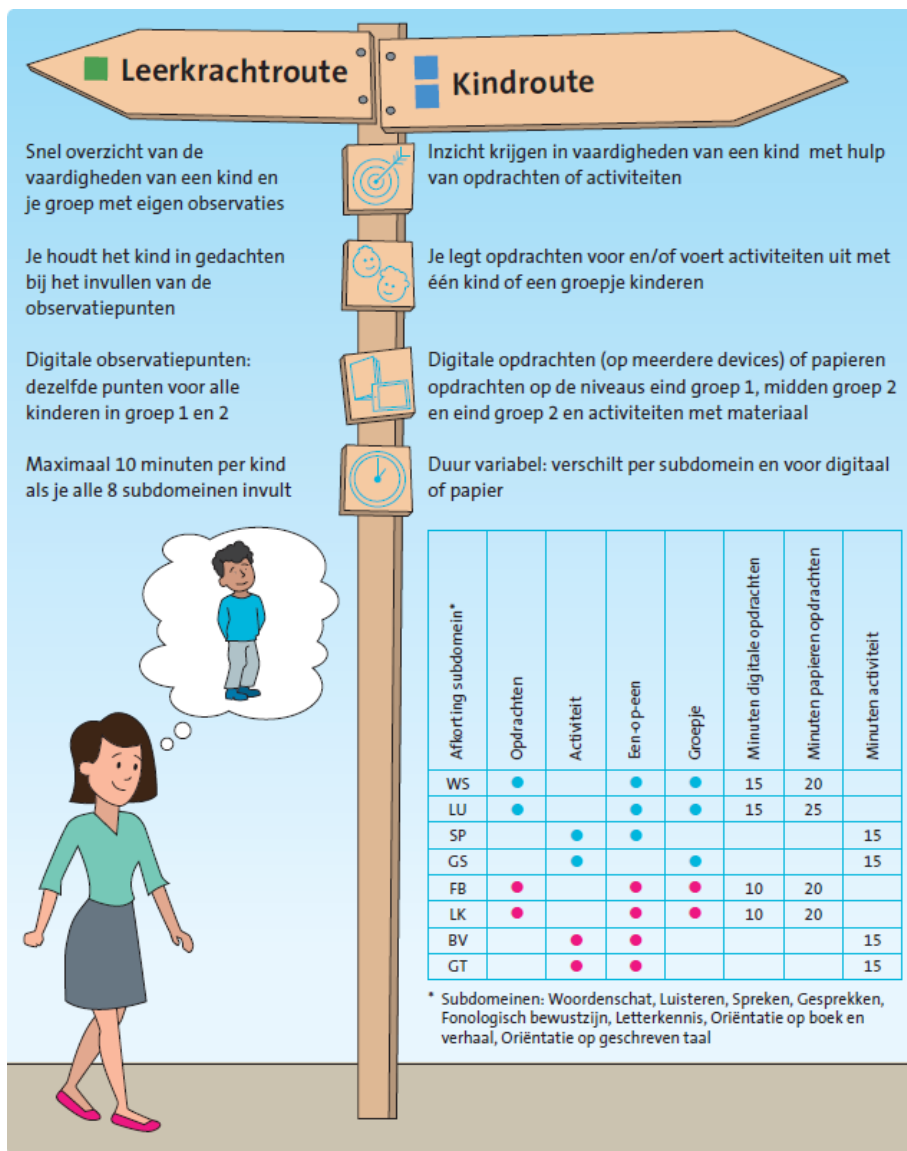
★★ = niveau midden groep 2

★★★ = 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 vier 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 - Taal



### 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 2018 heeft een bijeenkomst van een focusgroep plaatsgevonden. Een groep van 15 kleuterleerkrachten, intern begeleiders en onderbouwcoördinatoren uit het hele land kwamen naar Cito om mee te denken over de opbouw van het nieuwe instrument. In april 2018 is er een expertbijeenkomst bij Cito georganiseerd. Hierbij waren acht experts (onder andere van SLO en Expertisecentrum Nederlands) aanwezig om de toetsdeskundigen van Cito van feedback te voorzien over de opzet van Kleuter in beeld - Taal. Ook is in dezelfde periode een online community opgericht: 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:

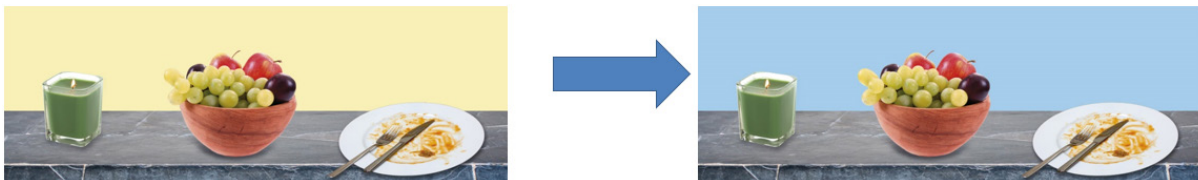
- Observatie centraal
- Kort, maar krachtig
- Flexibel in te zetten, hele jaar door
- Regie bij de leerkracht
- Inclusief concreet materiaal

- 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 items. Activiteiten kosten weliswaar wat meer tijd, maar passen goed bij deze leeftijdsgroep. Om te kunnen leren, moeten jonge kinderen objecten kunnen aanraken en manipuleren, moeten zij kunnen luisteren naar verhalen en rollen kunnen uitspelen, praten en zingen, en bewegen op verschillende manieren en plekken (Guddemi & Case, 2004). 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 Woordenschat, Luisteren, Fonologisch bewustzijn en Receptieve letterkennis waren opdrachten het meest passend, bij Spreken, Gesprekken, Oriëntatie op boek en verhaal en Oriëntatie op geschreven taal een activiteit.

In het voorjaar van 2018 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 - Taal 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/activiteiten gescreend op geschiktheid voor kinderen met speciale onderwijsbehoeften. Op basis daarvan zijn er opdrachten aangepast en ook afgefallen. Zo is bij onderstaande opdracht de achtergrondkleur aangepast, waardoor deze opdracht ook geschikt is voor kinderen die slechtziend zijn:



Voorbeeld Woordenschat

'Vers. Wat past het best bij vers?'

Vervolgens is in september 2018 een proefonderzoek georganiseerd en in februari 2019 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.

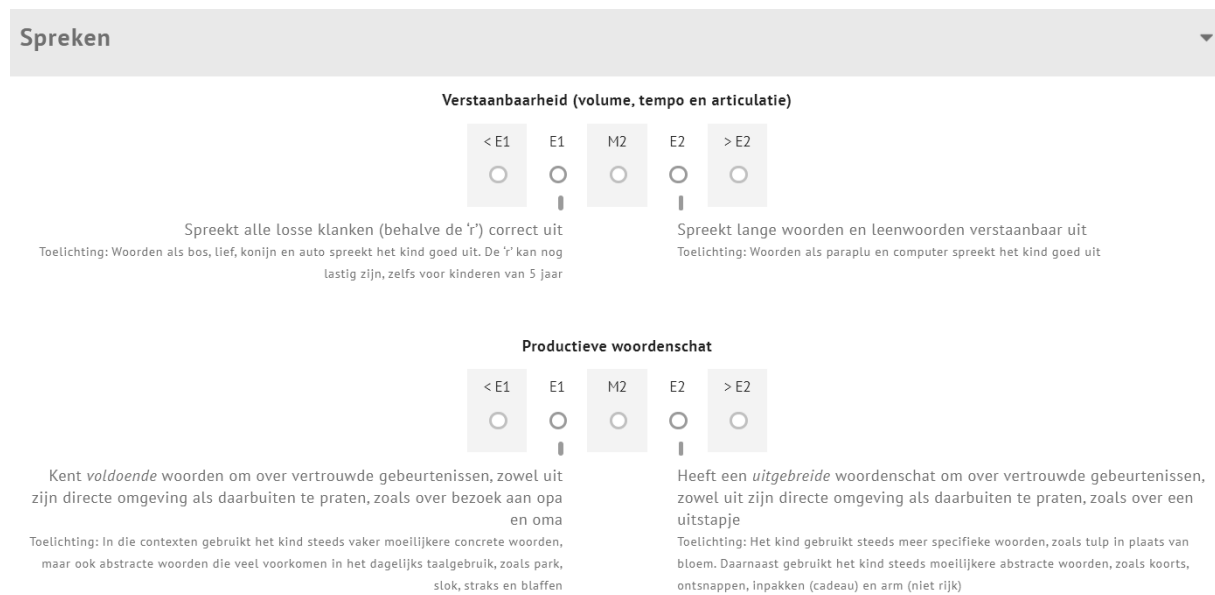
### **3.3 Operationalisatie van de inhoud**

In paragraaf 2.2 over de meetpretentie is al uiteengezet hoe de selectie van de acht 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 acht 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 vijf niveaus van vaardigheid onderscheiden: <E1, 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 (bronnen SLO en Expertisecentrum Nederlands). 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 constructiegroep van kleuterleerkrachten. 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. 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 Verstaanbaarheid en Productieve woordenschat bij Spreken.

Figuur 3.2 Voorbeeld deel leerkrachtroute



## Opdrachten kindroute

- **Woordenschat**

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

- waar mogelijk een hotspot (drie alternatieven in één plaatje, zie voorbeeld 1), en een traditionele meerkeuzeopdracht (drie alternatieven in losse kaders, zie voorbeeld 2) waar nodig;
- waar mogelijk foto's en tekeningen waar nodig;
- dat elke opdracht drie alternatieven heeft;
- waar mogelijk ook passend voor kinderen in het speciaal (basis)onderwijs;
- zowel opdrachten Betekenis (zie voorbeeld 3) als opdrachten Betekenisrelatie (zie voorbeeld 4).

**Betekenis:** het kind begrijpt verschillende zelfstandig naamwoorden, werkwoorden en bijvoeglijke naamwoorden waarbij het woord direct zichtbaar is in een van de drie keuzemogelijkheden van een opdracht.

**Betekenisrelatie:** het kind begrijpt verschillende zelfstandig naamwoorden, werkwoorden en bijvoeglijke naamwoorden waarbij hij dit woord aan een persoon, voorwerp, handeling of situatie moet koppelen die er qua betekenis aan gerelateerd is. Dat betekent dat het woord niet direct zichtbaar is in een van de keuzemogelijkheden van een opdracht.

Bij een opdracht Woordenschat hoort het kind een woord. Hij hoort dit woord twee keer: een keer geïsoleerd en een keer in zinsverband. Het kind kiest het plaatje waarop het woord staat afgebeeld (betekenis) of een plaatje dat het best bij het woord past (betekenisrelatie).



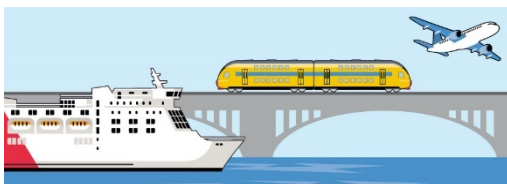
Voorbeeld 1) Hotspot met foto's  
'Buggy. Waar zie je een buggy?'



Voorbeeld 2) Traditionele meerkeuzeopdracht met tekeningen  
'Angstig. Op welk plaatje kijkt Mo angstig?'



Voorbeeld 3) Woordenschat betekenis  
'Thermometer. Op welk plaatje zie je een thermometer?'



Voorbeeld 4) Woordenschat betekenisrelatie  
'Piloot. Wat past het best bij een piloot?'

Bij het maken van de opdrachten Woordenschat is gekozen voor de **woordsoorten** zelfstandige naamwoorden (Z), werkwoorden (W) en bijvoeglijke naamwoorden (B). In onze gesproken taal gebruiken wij meer zelfstandige naamwoorden dan bijvoeglijke naamwoorden en werkwoorden. Daarom komen er relatief veel zelfstandige naamwoorden voor in de opdrachten en wat minder werkwoorden en bijvoeglijke naamwoorden. In Tabel 3.1 staat de toetsmatrix voor de drie niveaueversies van de uitgave.

Tabel 3.1 Toetsmatrix Uitgave Woordenschat

Categorie	★	★★	★★★
Betekenis	Z = 9 W = 3 B = 4	Z = 10 W = 2 B = 3	Z = 7 W = 3 B = 2
Betekenisrelaties			Z = 3 B = 1
<b>Totaal</b>	16	15	16

De gekozen woorden hadden verschillende **kenmerken**:

- van concreet (bijvoorbeeld *schort*) naar abstract (bijvoorbeeld *vers*)
- van dicht bij de belevingswereld (bijvoorbeeld *lantaarnpaal*) naar verder van de belevingswereld af (bijvoorbeeld *piloot*)

Hoe zijn de woorden bij Woordenschat gekozen?

De woorden zijn geselecteerd uit verschillende woordenlijsten:

- Basiswoordenlijst Amsterdamse Kleuters (BAK) (Mulder, Timman, & Verhallen, 2009)
- BasiLex<sup>1</sup> (Tellings, 2015)
- Digitale Woordenlijst Alle Kinderen (Digiwak<sup>2</sup>) (Stichting Digiwak, UvA)

Er is gestart met het selecteren van woorden uit de BAK. De BAK is bedoeld om tot een minimale woordenschat van 2000 woorden te komen. Gemiddeld begrijpen kinderen van 5 en 6 jaar echter tussen de 6000 tot 8000 woorden (Goorhuis & Schaerlaekens, 2000). Uit de onderzoeken bleek dat veel kinderen al veel woorden kenden. Daarom is ook Digiwak en BasiLex geraadpleegd. Voor alle woorden geldt dat de feedback van kleuterleerkrachten is meegenomen bij de selectie van de woorden. Verder was het uitgangspunt dat de woorden in een foto of tekening 'te vangen' moesten zijn, en dat de woorden aansloten bij verschillende thema's, zoals eten en drinken, verkeer en dieren.

- **(Begrijpend) Luisteren**

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

- sleepopdrachten (zie voorbeeld 5) en hotspots waar mogelijk, traditionele meerkeuzeopdracht waar nodig;
- foto's waar mogelijk, tekeningen waar nodig;
- drie alternatieven waar mogelijk, en vier waar nodig;
- dat de setting van de tekst aansluit bij de belevingswereld van de kleuter;
- waar mogelijk passend ook voor kinderen in het speciaal (basis)onderwijs;
- een combinatie van opdrachten zinsbegrip en verhaalbegrip.

Voor het subdomein (Begrijpend) Luisteren is de definitie van Gijsel en Van Druenen (2011) gehanteerd: het achterhalen van de betekenis van een gesproken tekst. In het instrument Kleuter in beeld - Taal wordt daarbij uit gegaan van teksten op zinsniveau (opdrachten Zinsbegrip) en op verhaalniveau (opdrachten Verhaalbegrip).

Zinsbegrip: het kind begrijpt de instructies (zie voorbeeld 5) en niet-instructies van een zin (zie voorbeeld 6), waarbij het combineren of selecteren van inhoudselementen, verschillende grammaticale constructies en enkele zinnen met figuurlijk taalgebruik centraal staan.

Verhaalbegrip: Het kind begrijpt informatieve en verhalende teksten met vier of meer uitingen achter elkaar waarbij het begrijpen van de strekking van de tekst centraal staat (zie voorbeeld 7).

Bij een opdracht Luisteren hoort het kind een zin of kort verhaaltje, gevolgd door een vraag. Elk luisterfragment kan een keer herhaald worden. Het kind kiest het juiste plaatje of sleept een plaatje naar de juiste plek (bij papieren opdrachten trekt het kind een lijn).

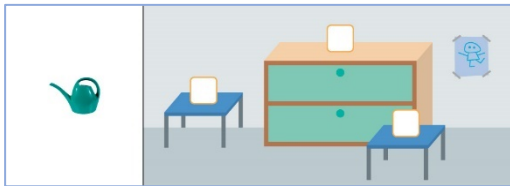
Bij het construeren van de opdrachten was het uitgangspunt dat deze aansluiten bij de belevingswereld van het kind. Het kunnen bijvoorbeeld instructies zijn die ze in de klas krijgen. Bij niet-instructies hoort het kind een zin die afkomstig zou kunnen zijn uit een verhalende tekst. De woorden die gebruikt zijn in de zinnen en verhalen zijn relatief gemakkelijk. Dit is om te voorkomen dat de woordenschat gemeten wordt, tenzij dit het te bevragen woord is (bijvoorbeeld een signaalwoord of figuurlijk taalgebruik).

---

<sup>1</sup> BasiLex is een verzameling van geschreven teksten die Nederlandse kinderen in de basisschoollleeftijd (vanaf groep 3) aangeboden krijgen. Het corpus telt zo'n 11,5 miljoen woorden.

<sup>2</sup> Digiwak is een online overzicht van de woorden die leerlingen van het primair onderwijs moeten beheersen. De lijst bevat 9052 woorden voor leerlingen uit groep 3 tot en met 8.

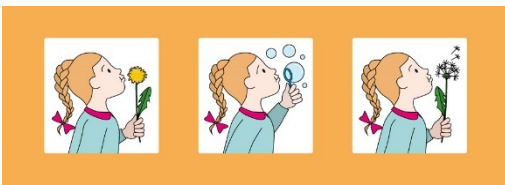
De meeste opdrachten zijn meerkeuzeopdrachten, maar er zijn ook een aantal sleepopdrachten ontwikkeld (zowel bij Zinsbegrip als Verhaalbegrip). In alle taken (niveaueversies) komt een mix voor van verschillende soorten opdrachten met verschillende kenmerken. De opdrachten figuurlijk taalgebruik worden alleen op niveau midden groep 2 en eind groep 2 aangeboden.



Voorbeeld 5) Zinsbegrip instructie in de vorm van een sleepopdracht  
'Zet de gieter op de tafel naast de kast.'



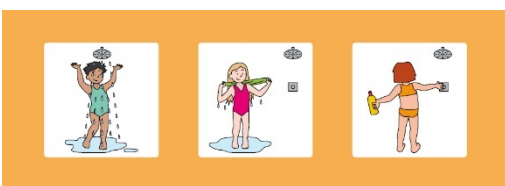
Voorbeeld 6) Zinsbegrip niet-instructie met vier keuzemogelijkheden in de vorm van een hotspot  
'Lola heeft een pet op haar hoofd en een vogeltje op haar schouder. Waar zie je Lola?'



Voorbeeld 7) Verhaalbegrip met drie keuzemogelijkheden in de vorm van een traditionele meerkeuzeopdracht  
'In de wei staat een paardenbloem. Eerst waren de bloemblaadjes geel. Nu zit er een grijze pluizenbol aan de steel. Blaas maar eens ... Dan vliegen de pluizjes door de lucht! Welk plaatje hoort bij dit verhaal?'

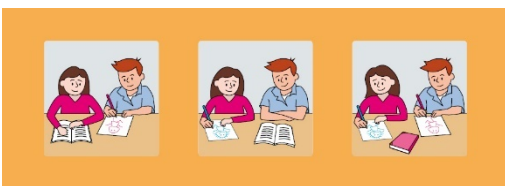
Bij het construeren van de zinnen is gekozen voor verschillende **categorieën**:

- Begrijpen van verschillende grammaticale constructies (zie voorbeeld 8)



Voorbeeld 8) Grammaticale constructie  
'Yara heeft gedoucht. Op welk plaatje zie je dat?'

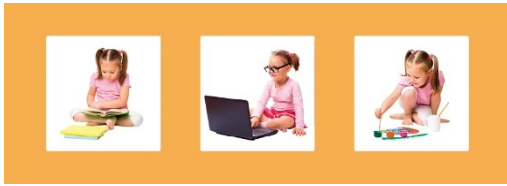
- Combineren van twee of drie inhoudselementen (zie voorbeeld 9)



Voorbeeld 9) Combineren van twee inhoudselementen  
'Jim maakt een tekening en Nadia leest een boek. Op welk plaatje zie je dat?'



- Selecteren van inhoudselementen (zie voorbeeld 10)



Voorbeeld 10) Selecteren van inhoudselementen

'Gisteren speelde Liva op de computer, vandaag zit ze te lezen en morgen wil ze gaan schilderen. Wat deed Liva gisteren?'

- Opdrachten over figuurlijk taalgebruik (geselecteerd uit Basiswoordenlijst Amsterdamse Kleuters; (Mulder et al., 2009)

Bij het construeren van de verhaaltjes is gekozen voor verhalen van vier tot zes zinnen. Daarnaast zijn zowel verhalende als informatieve teksten opgenomen. In Tabel 3.2 staat de toetsmatrijs voor de drie niveaueversies van de uitgave.

Tabel 3.2 Toetsmatrijs Uitgave Luisteren

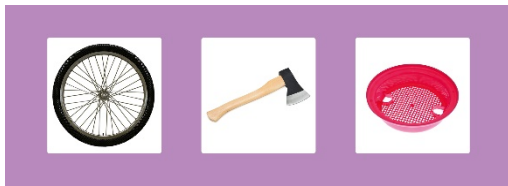
		★	★★	★★★
<b>Zinsbegrip</b>		16 zinnen, waarvan:	15 zinnen, waarvan:	14 zinnen, waarvan:
(Niet-) Instructie	Instructie	6	4	3
	Niet-instructie	10	11	11
<b>Categorie</b>	<b>Grammaticale constructies</b>	7	7	5
	Combineren van inhoudselementen	7	4	3
	Selecteren van inhoudselementen	2	3	4
	Figuurlijk taalgebruik	0	1	2
<b>Verhaalbegrip</b>		4 verhalen, waarvan:	5 verhalen, waarvan:	5 verhalen, waarvan:
	Informatief	3	3	4
	Verhalend	1	2	1
<b>Totaal</b>		20	20	19

- **Fonologisch bewustzijn**

De uitgangspunten bij het maken van de opdrachten waren:

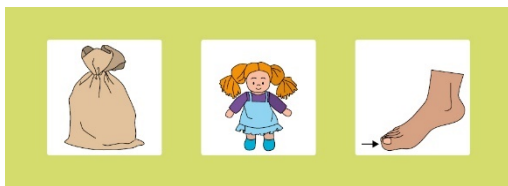
- foto's waar mogelijk, tekeningen waar nodig;
- de afbeeldingen zijn eenduidig en uniek, hebben geen samenhang,
- zo kan het kind zich goed richten op de klank van de woorden;
- elke opdracht heeft drie alternatieven;
- bij de papieren opdrachten benoemt de leerkracht de woorden van de drie alternatieven, bij de digitale opdrachten benoemt de computerstem deze woorden;
- waar mogelijk ook passend voor kinderen in het s(b)o.

Op basis van literatuur en wensen uit het veld is ervoor gekozen de onderdelen rijm (zie voorbeeld 11), beginfoneem (zie voorbeeld 12) en synthese (zie voorbeeld 13) op te nemen. Bij een opdracht Fonologisch bewustzijn hoort het kind een vraag. De drie afbeeldingen worden benoemd. Het kind kiest de juiste afbeelding.



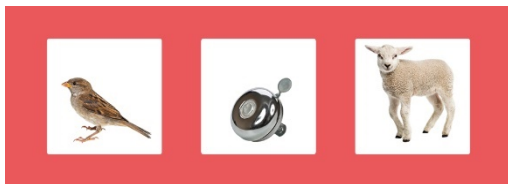
Voorbeeld 11) Rijm

'Wat rijmt op neef? Wiel, bijl, zeef'



Voorbeeld 12) Beginfoneem

'De z van zon. Welk woord begint ook met de z? zak, pop, teen'



Voorbeeld 13) Synthese

'Op welk plaatje zie je -l-a-m-? mus, bel, lam'

Voor de opdrachten Fonologisch bewustzijn is gekozen voor korte, klankzuivere woorden die uitgedrukt kunnen worden in een afbeelding. Ook zijn woordenlijsten gebruikt om aan te sluiten bij het niveau en de belevingswereld van kinderen in groep 1 en 2:

- Basiswoordenlijst Amsterdamse Kleuters (Mulder et al., 2009)
- Digitale Woordenlijst Alle Kinderen (Digiwak) (Stichting Digiwak, UvA)
- Woordenlijst Schrooten & Vermeer (1994)

De woordenlijsten geven aan in welke groep het kind een woord beheerst. Vervolgens is aan leerkrachten gevraagd of de woorden en afbeeldingen aansluiten bij de belevingswereld en het niveau van kinderen van groep 1 en 2. In Tabel 3.3 staat de toetsmatrijs voor de drie niveauversies van de uitgave.

Tabel 3.3 Toetsmatrijs Uitgave Fonologisch bewustzijn

Onderdeel	★	★★	★★★
Rijm	5	4	3
Beginfoneem	6	6	6
Synthese	4	5	6
<b>Totaal</b>	15	15	15

- **(Receptieve) Letterkennis**

Bij de opdrachten (Receptieve) Letterkennis worden alle klinkers en medeklinkers van het alfabet aangeboden, met uitzondering van de c, q, x en y. Deze letters komen niet of nauwelijks voor in woorden die kleuters beheersen, daarom zijn ze niet opgenomen. Hierbij wordt het onderwijs gevolgd, waar letters worden aangeboden in een betekenisvolle context en deze letters ook niet of nauwelijks voorkomen. Ook in de literatuur (Van Druenen et al., 2017) zijn deze letters niet opgenomen als lesstof voor kleuters. Bij een opdracht (Receptieve) Letterkennis hoort het kind een letter en wordt die letter uitgesproken in een kort, klankzuiver en betekenisvol woord. Zo hoort een kind duidelijk de uitspraak en de klank van de letter. Het kind kiest de juiste letter uit drie verschillende afbeeldingen (zie voorbeeld 14).



Voorbeeld 14) (Receptieve) Letterkennis  
 'Waar zie je de m van maan?'

De drie letters staan naast elkaar in wolkjes. Zo zijn ze afzonderlijk te 'lezen' door het kind en is het duidelijk dat het om drie verschillende letters gaat. Er is gekozen voor wolkjes, omdat dit een eenvoudig en speels aanbod is van de letters. Door het veld werd dit goed ontvangen.

In Tabel 3.4 staat de toetsmatrijs voor de drie niveauversies van de uitgave.

Tabel 3.4 Toetsmatrijs Uitgave Letterkennis

Categorie	★	★★	★★★
Medeklinker	17	17	17
Klinker	5	5	5
Lange klank			4
<b>Totaal</b>	22	22	26

**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: variëren: nee-ja (dichotome observatiepunten), nee-beetje-ja en nee-met hulp-ja (polytome observatiepunten). 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.

- **Spreken**

Het spreken van kinderen kan uitstekend geobserveerd worden in de dagelijkse klassenpraktijk of tijdens een activiteit. Kleuters gebruiken vaak concrete hulpmiddelen, zoals plaatjes of een voorwerp om iets te vertellen (Van der Beek et al., 2015). Om daarbij aan te sluiten, is een activiteit ontwikkeld waarbij een kind iets vertelt bij een stripje van Anne en Mo. Er zijn drie stripjes met vier of vijf plaatjes ontwikkeld.

De onderwerpen sluiten aan bij de belevingswereld van kleuters (zoals het bouwen van een tent), zodat er ook doorgevraagd kan worden naar de eigen ervaringen van het kind met het onderwerp. Voor het formuleren van observatiepunten voor verschillende niveaus is literatuur gebruikt. Zo bevat bijvoorbeeld Praatjes peilen (SLO, 2014) een bijlage met niveaugrids. Deze geven houvast bij het maken van een inschatting van het taalniveau van kinderen. Op elk onderdeel van de taalontwikkeling staan op vijf niveaus indicatoren benoemd die overeenkomen met belangrijke mijlpalen in de taalontwikkeling: niveau 2 is

aanvang groep 1 en niveau 4 is eind groep 2. In Tabel 3.5 staat de toetsmatrijs voor de observatiepunten van deze activiteit.

*Tabel 3.5 Toetsmatrijs Uitgave Spreken*

Onderdeel	N observatiepunten
Verstaanbaarheid	6
Productieve woordenschat	3
Grammaticale beheersing	8
Verhalen vertellen	4
<b>Totaal</b>	<b>21</b>

- **Gesprekken**

Gesprekken kunnen uitstekend geobserveerd worden in de dagelijkse klassenpraktijk of tijdens een activiteit. Een gespreksstaak voor groep 1/2 die in de Leerstoflijnen mondelinge taalvaardigheid wordt genoemd (Van der Beek et al., 2015), is het deelnemen aan discussie. Er is een activiteit bedacht waarbij een groepje kinderen met de leerkracht in gesprek gaat over een waarschijnlijk onbekend voorwerp (een plumeau) met verschillende kenmerken wat betreft bijvoorbeeld kleur en materiaal. Wat zou het zijn? Wat zou je er (nog meer) mee kunnen doen? Naar aanleiding van het kwaliteitsonderzoek is er bij de uitgave voor gekozen om geen standaardvoorwerp te leveren bij de activiteit, maar de keus voor een geschikt voorwerp bij de leerkracht zelf te laten. Bijvoorbeeld een voorwerp dat aansluit bij het thema of de interesse van een groepje kinderen. Voor het formuleren van observatiepunten voor verschillende niveaus gebruikten wij net als bij Spreken de niveaugrids van Praatjes peilen. In Tabel 3.6 staat de toetsmatrijs voor de observatiepunten van deze activiteit.

*Tabel 3.6 Toetsmatrijs Uitgave Gesprekken*

Onderdeel	N observatiepunten
Vasthouden van gespreksonderwerp	3
Eenvoudige gespreksregels	5
<b>Totaal</b>	<b>8</b>

- **Oriëntatie op boek en verhaal**

In de activiteit Oriëntatie op boek en verhaal leest de leerkracht een prentenboek voor aan het kind. Er zijn drie fasen: voor het lezen, tijdens het lezen en na het lezen. In de eerste fase komt het onderdeel Boekoriëntatie aan bod, in de tweede en derde fase het onderdeel Verhaalbegrip. In elke leesfase worden vier vragen aan het kind gesteld. De inhoud van de vragen zijn een uitwerking van de doelen boekoriëntatie en verhaalbegrip in de literatuur. De vragen zijn concreet en gaan over de inhoud van het prentenboek. De vragen zijn letterlijk weergegeven in de activiteit, vanwege eenduidigheid en objectiviteit van de observatiepunten. De volgorde van observatiepunten is inhoudelijk opgebouwd. Eerst kijkt het kind naar de afbeeldingen op de omslag en in het prentenboek en dan pas naar de tekst in het boek. In Tabel 3.7 staat de toetsmatrijs voor de observatiepunten van deze activiteit.

Tabel 3.7 Toetsmatrijs Uitgave Oriëntatie op boek en verhaal

Onderdeel	Fase	N observatiepunten
Boekoriëntatie	voor het lezen	4
Verhaalbegrip	tijdens het lezen	4
	na het lezen	4
<b>Totaal</b>		12

• **Oriëntatie op geschreven taal**

Dit subdomein bestaat uit de onderdelen Voorbereidend lezen en Voorbereidend schrijven. De activiteit is een uitwerking van de doelen oriëntatie op geschreven taal, alfabetisch principe en functioneel lezen, gebaseerd op de eerdergenoemde literatuur in paragraaf 2.2. De leerkracht doet met het kind spelletjes met concrete materialen, zoals kaartjes met pictogrammen en plaatjes en naamkaartjes van kinderen uit de groep. De activiteiten zijn opgebouwd van dichtbij (eigen naam) naar veraf (korte algemene woordjes). Omdat we motorische vaardigheid willen uitsluiten, mogen kinderen ook letterstempels gebruiken in plaats van te schrijven met potlood. Wat de leerkracht zegt tegen en vraagt aan het kind is letterlijk weergegeven in de activiteit, vanwege eenduidigheid en objectiviteit van de observatiepunten. Bij het onderdeel Voorbereidend lezen is bij het aspect letters benoemen gekozen voor het actief benoemen van letters. Dit is een andere vaardigheid dan het herkennen van letters, zoals bij de opdrachten (Receptieve) Letterkennis. Dezelfde klinkers, medeklinkers en lange klanken komen aan bod. De volgorde van observatiepunten in de activiteiten is inhoudelijk opgebouwd. Bijvoorbeeld bij het spel met de naamkaartjes bij Voorbereidend lezen is ervoor gekozen om eerst het eigen naamkaartje te herkennen (GT02) en dan pas letters te benoemen in de naam (GT03). In Tabel 3.8 staat de toetsmatrijs voor de observatiepunten van deze activiteit.

Tabel 3.8 Toetsmatrijs Uitgave Oriëntatie op geschreven taal

Onderdeel	N observatiepunten
Vorbereidend lezen	5
Vorbereidend schrijven	6
<b>Totaal</b>	11



## 4 Dataverzameling

### 4.1 Proefonderzoek

In september 2018 is een proefonderzoek georganiseerd. Een belangrijk doel van het onderzoek was om de kwaliteit en de moeilijkheidsgraad van de opdrachten 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.

In het steekproefkader voor het proefonderzoek zaten ruim 6000 scholen. Hieruit is een aselecte steekproef van 840 scholen getrokken. Hierbij is rekening gehouden met het percentage gewichtenleerlingen en regio. Deze scholen hebben per post een uitnodiging ontvangen om mee te doen met het onderzoek. 41 scholen meldden zich aan voor het onderzoek. Uiteindelijk hebben 37 scholen en 922 kinderen uit groep 1 en 2 daadwerkelijk meegedaan.

Alle kinderen maakten de opdrachten Woordenschat en Luisteren uit een papieren opdrachtenboekje Mondelinge taal. Daarnaast maakten alle kinderen opdrachten Fonologisch bewustzijn (rijm, beginfoneem, synthese en deletie) uit een papieren opdrachtenboekje Beginnende geletterdheid. Bij de kinderen in groep 2 waren aan het boekje Beginnende geletterdheid ook een aantal opdrachten Receptieve letterkennis toegevoegd. Om zoveel mogelijk verschillende opdrachten uit te proberen, zijn er verschillende boekjes gemaakt en hebben niet alle kinderen dezelfde opdrachten gemaakt. In totaal zijn 73 verschillende opdrachten voor Woordenschat voorgelegd, voor Luisteren 44, voor Fonologisch bewustzijn 75 en voor Receptieve letterkennis 34.

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). 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, zoals:

- Opdrachtenboekje op A5-formaat  
Opmerkingen uit het evaluatieformulier en eigen observaties bij de schoolbezoeken gaven de doorslag: het papieren opdrachtenboekje is verkleind van A4 naar A5. Kleuters krijgen dan telkens één opdracht in beeld in plaats van vier, worden niet meer afgeleid door andere opdrachten en hebben geen afdekblad meer nodig.
- Lange vraagvorm bij letterkennis  
'De m van maan' of 'Waar zie je de m van maan?'. Het was duidelijk dat de laatste optie de voorkeur had, dus die vraagstelling is gekozen voor het kwaliteitsonderzoek en de uitgave.
- Bij een aantal opdrachten zijn de foto's/tekeningen bij de alternatieven aangepast.
- Onderdeel Deletie weglaten  
Leerkrachten gaven aan dat het onderdeel Deletie nauwelijks wordt aangeboden aan kleuters en dat ze dit onderdeel liever willen laten vervallen in de uiteindelijke taken voor Fonologisch bewustzijn.

Na het onderzoek zijn toets- en itemanalyses uitgevoerd. Het oordeel van de leerkrachten over het onderdeel Deletie werd bevestigd door de data: de opdrachten Deletie bleken te moeilijk (alle deletie-opdrachten hadden in groep 2 p-waarden lager dan 0,5) en maakten weinig onderscheid tussen vaardige en minder vaardige kleuters. Daarom zijn de opdrachten Deletie niet meer opgenomen in het kwaliteitsonderzoek en de uitgave. Op basis van psychometrische gegevens (p- en rit-waarden) en feedback van de deelnemers zijn de beste opdrachten geselecteerd voor het kwaliteitsonderzoek. Er waren opdrachten die

voor groep 1 nog te makkelijk bleken en andere opdrachten die voor groep 2 nog te moeilijk bleken. Deze opdrachten zijn afgevallen.



Voorbeeld Woordenschat

'Timmeren. Wat past het best bij timmeren?'

De p-waarde van deze opdracht was ,92 in groep 1, dit is erg gemakkelijk, daarom is deze opdracht afgevallen.

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.

Bij Woordenschat en Fonologisch bewustzijn bleven er na het proefonderzoek voldoende geschikte opdrachten over voor het kwaliteitsonderzoek. Voor Luisteren waren 44 opdrachten geconstrueerd voor het proefonderzoek en een aantal opdrachten vielen na het proefonderzoek af. Voor Luisteren zijn er daarom voor het kwaliteitsonderzoek nog 15 extra opdrachten geconstrueerd.

## 4.2 Steekproef kwaliteitsonderzoek

In februari 2019 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 361 scholen getrokken voor de digitale afname van de opdrachten en 344 scholen voor de papieren afname van de opdrachten (uitgaande van een response rate van 0,04%) in groep 1 en groep 2. Hierbij is rekening gehouden met het percentage gewichtenleerlingen en regio. De tweede steekproef voor de papieren afname bestond uit 856 scholen en digitaal uit 571 scholen. De getrokken steekproeven zijn tweemaal aangeschreven. Omdat het aantal scholen dat zich aanmeldde te gering was, zijn uiteindelijk alle scholen aangeschreven. Daarnaast is ervoor gekozen om extra te werven door middel van een advertentie op Facebook en te flyeren op beurzen. Om de deelnamebereidheid te vergroten is besloten een beloning (VVV-bon van 25 euro) te geven aan deelnemende scholen.

Uiteindelijk hebben 82 scholen meegedaan. Scholen hadden de voorkeur voor de papieren afname van de opdrachten (zie Tabel 4.1). Vijf scholen deden met zowel de papieren als de digitale versie mee. Drie van deze scholen namen de papieren versie af in andere groepen dan de digitale versie. Meer dan de helft van de scholen (53) deden met twee of zelfs drie leerjaren mee (Tabel 4.2). Negen scholen deden daarnaast een afname bij kinderen uit groep 3. Deze gegevens zijn alleen gebruikt om te controleren of de opdrachten/observatiepunten voldoende onderscheid maakten tussen leerjaar 2 en leerjaar 3.



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

	Papier	Digitaal	Totaal
Groep 1	501	206	707
Groep 2	957	438	1395
Groep 3	83	67	150
	1541	711	2252

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

Papier Groep	Aantal scholen	Digitaal Groep	Aantal scholen	Groep	Totaal Aantal scholen
Groep 1	3	-	-	Groep 1	3
Groep 1	1	Groep 2	1	Groep 1 & 2**	1
Groep 2	1	Groep 2	1	Groep 2*	1
-	-	Groep 2	9	Groep 2	9
Groep 2	16	-	-	Groep 2	16
Groep 2	1	Groep 1	1	Groep 1 & 2**	1
Groep 2	1	Groep 1 & 2	1	Groep 1 & 2**	1
Groep 1 & 2	1	Groep 1 & 2	1	Groep 1 & 2*	1
-	-	Groep 1 & 2	13	Groep 1 & 2	13
Groep 1 & 2	27	-	-	Groep 1 & 2	27
Groep 1, 2 & 3	4	-	-	Groep 1, 2 & 3	4
-	-	Groep 1, 2 & 3	2	Groep 1, 2 & 3	2
Groep 2 & 3	2	-	-	Groep 2 & 3	2
-	-	Groep 2 & 3	1	Groep 2 & 3	1
	57		30		82

\* 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 82 scholen; 2252 kinderen uit groep 1, 2 en 3 deden mee.

De representativiteit van de steekproef uit het kwaliteitsonderzoek<sup>3</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 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).

<sup>3</sup> Vier 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.

- *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.

Tabel 4.3 *Representativiteit van de steekproef uit het kwaliteitsonderzoek*

Variabele	Populatie		Steekproef Kleuter in beeld - Taal	
	Categorie	%	N	%
<b>Regio</b>	Noord	10,0	13	16,7
	Oost	22,4	26	33,3
	West	47,9	23	29,5
	Zuid	19,7	16	20,5
<b>Urbanisatie</b>	Platteland	54,8	55	70,5
	Stad	45,2	23	29,5
<b>Percentage achterstandskinderen</b>	0 – 10%	67,3	60	76,9
	>10%	32,7	18	23,1
<b>Sekse</b>	Jongen	50,5	981	51,9
	Meisje	49,5	908	48,1
	Onbekend		363	

We zien dat het percentage scholen met meer dan 10% achterstandskinderen niet significant afwijkt van de hele populatie scholen in Nederland. De verhouding van jongens en meisjes in de steekproef is ook niet significant verschillend van de verhouding in de populatie (zie Tabel 4.4).

Wel zien we dat de verdeling over regio's en de urbanisatiegraad in de steekproef afwijkt van de verdeling in heel Nederland, zie Tabel 4.4. We zien dat minder scholen uit het westen (29,5% in de steekproef versus 47,9% van de hele populatie) hebben deelgenomen en meer uit het noorden en oosten. Het aantal scholen uit het zuiden is wel representatief. We zien ook dat naar verhouding meer scholen uit niet tot matig verstedelijkte gebieden (platteland) hebben deelgenomen (70,5% in de steekproef, tegenover 54,8% in heel Nederland, zie Tabel 4.3). 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
Regio	13,18	3	78	0,004
Urbanisatie	7,77	1	78	0,005
Percentage achterstandskinderen	3,28	1	78	0,070
Sekse	1,64	1	1889	0,201

### 4.3 Ontwerp van het kwaliteitsonderzoek

Voor de drie subdomeinen, Woordenschat, Luisteren en Fonologisch bewustzijn, zijn 56 tot 60 opdrachten ontwikkeld om uit te testen in het kwaliteitsonderzoek. Voor elk van deze drie subdomeinen zijn boekjes gemaakt van 15 opdrachten om voor te leggen aan de kinderen. Voor elk subdomein zijn 12 opdrachten-blokjes gemaakt van 5 opdrachten, die als makkelijk, gemiddeld of moeilijk ingeschat werden. Vervolgens zijn daarmee 11 verschillende boekjes gemaakt met overlap volgens het onderstaande ontwerp (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 Woordenschat, Luisteren en Fonologisch bewustzijn en aantal observaties per boekje

Groep	Boekje	Makkelijk			Gemiddeld			Moeilijk			LU	WS	FB		
		5	5	5	5	5	5	5	5	5					
1	1	1	1		1						164	167	152		
1	2		1	1		1					120	125	123		
1	3			1	1	1					148	151	136		
1	4	1			1	1					170	172	174		
2	5				1		1	1			207	210	206		
2	6					1		1	1		201	233	202		
2	7				1				1	1	213	217	210		
2	8					1				1	1	219	226	212	
2	9				1						1	1	194	200	195
2	10					1	1					1	220	224	201
3	11									1	1	1	103	143	104
											1959	2068	1915		

In het kwaliteitsonderzoek vulde de leerkracht voor elke leerling het (indirecte) observatieformulier van de leerkrachtroute in. Hierdoor zijn de gebruikte observatoren uit de steekproef van 83 scholen een goede afspiegeling van de observatoren die het observatie-instrument daadwerkelijk in de praktijk gaan gebruiken. 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).

Daarnaast deden de kinderen vier subdomeinen uit de kindroute: drie subdomeinen met opdrachten (Luisteren, Woordenschat en Fonologisch bewustzijn) en één activiteit (Spreken, Gesprekken, Oriëntatie op

boek en verhaal, Oriëntatie op geschreven taal) of de drie subdomeinen met opdrachten, aangevuld met opdrachten Receptieve letterkennis. Het ontwerp voor de combinaties van de verschillende subdomeinen staat in Tabel 4.6. De (indirecte) observaties in de leerkrachtroute en de directe observaties van de activiteiten (Spreken, Gesprekken, Oriëntatie op boek en verhaal en Oriëntatie op geschreven taal) werden op een papieren observatieformulier ingevuld tijdens het kwaliteitsonderzoek.

Tabel 4.6 Ontwerp voor de combinaties van de verschillende subdomeinen

LKroute	Woordenschat (WS)	Luisteren (LU)	Fonologisch bewustzijn (FB)	Receptieve letterkennis (LK)	Spreken (SP)	Gesprekken (GS)	Oriëntatie op boek en verhaal (BV)	Oriëntatie op geschreven taal (GT)
1	1	1	1	1				
1	1	1	1		1			
1	1	1	1			1		
1	1	1	1				1	
1	1	1	1					1

Tabel 4.7 Aantal observaties per versie

LKroute	Woordenschat (WS)	Luisteren (LU)	Fonologisch bewustzijn (FB)	Receptieve letterkennis (LK)	Spreken (SP)	Gesprekken (GS)	Oriëntatie op boek en verhaal (BV)	Oriëntatie op geschreven taal (GT)
404	518	521	524	540	0	0	0	0
175	169	157	164	0	184	0	0	0
480	502	500	486	0	0	510	0	0
220	208	201	194	0	0	0	242	0
215	244	230	242	0	0	0	0	284
188	427	350	305	0	0	0	0	0
<b>1682</b>	<b>2068</b>	<b>1959</b>	<b>1915</b>	<b>540</b>	<b>184</b>	<b>510</b>	<b>242</b>	<b>284</b>

Het valt op dat bij het subdomein Gesprekken de meeste observaties bij een activiteit zijn gedaan (zie ook Tabel 4.7). Dit is te verklaren doordat deze activiteit bij twee of drie kinderen tegelijk moest worden uitgevoerd. De overige activiteiten werden individueel afgenomen.

#### 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: 2		Groep: Groep 2																																	
		Receptieve woordenschat		Begrijpend luisteren		Spreken - Voorzetsbaarheid		Spreken - Productieve woordenschat		Spreken - Grammaticale beheersing		Spreken - Verhalen vertellen		Gepraakt - Voorhouden		Gepraakt - Vergelegenheidsverp		Gepraakt - Eenvoudige problemen		Fonologisch bewustzijn - rijn		Fonologisch bewustzijn - riefloemen		Fonologisch bewustzijn - systeme		Letterkennis		Oriëntatie op boek en verhaal - Besluitvorming		Oriëntatie op boek en verhaal - Verhaal		Oriëntatie op geschreven taal - Voorbereidingsfase		Oriëntatie op geschreven taal - Voorbereidingsfase	
Naam																																			
Leerling_001		M2	M2	M2	M2	E1	M2	M2	M2	M2	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	M2	E1	E1	E1	M2	M2						
Leerling_002		M2	M2	E1	M2	E1	M2	E1	M2	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	E1	M2	E1	E1	E1	M2	M2						
Leerling_003		E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	M2	E2	E2	E2	E2	E2	E2	E2	E2	>E2	>E2	E2	E2	>E2	E2	E2					
Leerling_004		>E2	>E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	E2	>E2	E2	E2	E2	E2	E2	E2	E2	E2	>E2	E2	E2	E2	>E2	E2	E2					

Papieren/digitale opdrachtjes						Activiteiten											
Totaal		Receptieve woordenschat		Begrijpend luisteren		Fonologisch bewustzijn		Receptieve Letterkennis		Spreken		Gesprekken		Oriëntatie op boek en verhaal		Oriëntatie op geschreven taal	
#	goed	#	goed	#	goed	#	goed	#	goed	#	goed	#	goed	#	goed	#	goed
45	41	15	14	15	12	15	15					10	9				
45	30	15	11	15	12	15	7					10	9				
45	33	15	13	15	10	15	10					10	6,5				
45	34	15	10	15	11	15	13					10	7				
45	41	15	14	15	14	15	13					10	6				
45	40	15	13	15	14	15	13					10	9,5				
45	34	15	12	15	11	15	11					10	6,5				
45	30	15	9	15	9	15	12					10	4,5				
45	33	15	12	15	11	15	10										
45	27	15	10	15	9	15	8					10	7,5				
45	38	15	12	15	13	15	13					10	10				
45	28	15	11	15	10	15	7					10	3,5				
45	25	15	10	15	8	15	7					10	9				
45	37	15	14	15	12	15	11					10	8				

De scholen kregen een korte uitleg over de rapportages:

"#: het aantal opdrachten 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 - Taal."

## 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.

### Opdrachten

Scholen hebben behoefte aan korte taken, waarmee ze op een snelle manier de vaardigheid van kinderen in beeld kunnen brengen. Het uitgangspunt was een taak van 15 tot 20 opdrachten. Indien de betrouwbaarheid het toeliet, werd de lengte zover mogelijk teruggebracht tot 15 opdrachten, anders werden 20 opdrachten opgenomen in de taak.

### Woordenschat

Bij de opdrachten Woordenschat is een onderscheid gemaakt in Betekenis en Betekenisrelatie. Opdrachten Betekenisrelatie bleken qua moeilijkheid het beste bij de niveauserie ★★★ te passen. Daarom zijn deze opdrachten alleen in de moeilijkste versie opgenomen. Verder is ervoor gezorgd dat in elke versie de verschillende woordsoorten (zelfstandige naamwoorden, bijvoeglijke naamwoorden en werkwoorden) voorkomen. Zelfstandige naamwoorden komen in elke versie het meeste voor, omdat we deze woordsoort in onze gesproken taal ook het meest gebruiken. Verder is ervoor gezorgd dat in alle taken een mix

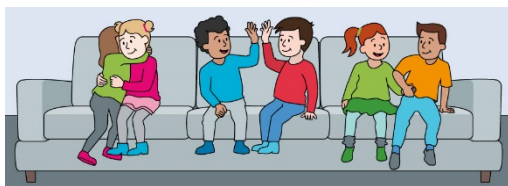
voorkomt van woorden met verschillende kenmerken (concreet versus abstract en dicht bij de belevingswereld versus verder van de belevingswereld af).

Er zijn opdrachten afgevallen, omdat deze te makkelijk bleken en geen onderscheid maakten tussen vaardige en minder vaardige kinderen. Een voorbeeld hiervan is de opdracht 'munt' met een p-waarde van ,96 in groep 1. De moeilijkste opdracht bleek 'omhelzen' en had een p-waarde van ,48 in groep 2. Om ook de betere kinderen goed te kunnen onderscheiden, is deze opdracht opgenomen in de versie ★★★.



Voorbeeld

'Munt. Op welk plaatje zie je een munt?'

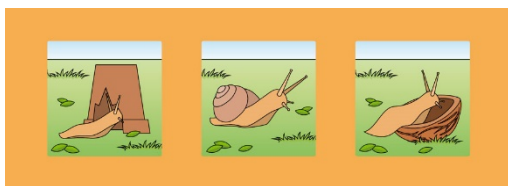


Voorbeeld

'Omhelzen. Waar zie je twee kinderen elkaar omhelzen?'

### **(Begrijpend) Luisteren**

Bij Luisteren zijn opdrachten Zinsbegrip en Verhaalbegrip ontwikkeld. Bij de opdrachten Zinsbegrip is een onderscheid gemaakt tussen Instructie en Niet-instructie en is gezorgd voor opdrachten met verschillende grammaticale constructies. Daarnaast zijn een aantal sleepopdrachten geconstrueerd (zowel bij Zinsbegrip als Verhaalbegrip). In alle taken (niveaueversies) komt een mix voor van verschillende soorten opdrachten met verschillende kenmerken. Opdrachten over figuurlijk taalgebruik komen vanwege hun relatieve moeilijkheid alleen voor in niveaueversie ★★ (1 opdracht) en ★★★ (2 opdrachten). Na het kwaliteitsonderzoek vielen ook wat opdrachten af. Op basis van de opmerkingen in het evaluatieformulier bleek de opdracht over de slak bijvoorbeeld inhoudelijk niet goed genoeg. De sleutel is B, maar daar kan over getwist worden. Deze opdracht is daarom niet opgenomen in de uitgave.

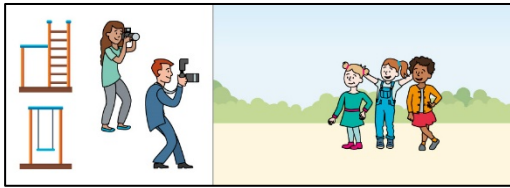


Voorbeeld

De slak heeft een huisje op zijn rug. Dat zit aan hem vast. Als de slak bang is, kruipt hij erin. Daar is hij veilig. Welk plaatje hoort bij dit verhaal?

Over de opdracht met twee fotografen kwamen veel opmerkingen, het bijbehorende verhaaltje zou te ingewikkeld zijn en de vraag niet duidelijk voor het kind. Dit werd ondersteund door de psychometrie: de Rar bleek hoger dan de Rir.

Dit betekent dat het verkeerde antwoord een hogere Item-Restcorrelatie heeft dan het juiste antwoord. Deze opdracht is daarom ook komen te vervallen.



### **Fonologisch bewustzijn**

In alle taken (niveaueversies) Fonologisch bewustzijn komen de drie verschillende onderdelen Rijm, Beginfoneem en Synthese voor. Bij het subdomein Fonologisch bewustzijn is sprake van een gradueel toenemende vaardigheid, een kind beheerst de vaardigheid steeds beter. Niveaueversie ★ bevat vooral de relatief makkelijke opdrachten, maar ook enkele opdrachten die wat moeilijker zijn. Niveaueversie ★★★ bevat de moeilijker opdrachten. Zo was het mogelijk om met elke niveaueversie verschillende niveaus in beeld te brengen. Omdat rijm relatief makkelijker bleek te zijn en synthese moeilijker, zijn in taak ★ wat meer opdrachten rijm opgenomen en in taak ★★★ wat meer opdrachten synthese. De synthese-opdracht over 'vos' is in het kwaliteitsonderzoek zowel in groep 1 als in groep 2 afgenomen. In groep 1 was de p-waarde ,73 en ,93 in groep 2. Voor groep 2 is deze opdracht te makkelijk. Daarom is deze opdracht opgenomen in niveaueversie ★.



Voorbeeld Synthese

'Op welk plaatje zie je v-o-s?'

### **(Receptieve) Letterkennis**

Bij het subdomein Receptieve letterkennis is sprake van een gradueel toenemende kennis, een kind kent steeds meer letters. In de niveaueversie ★ en ★★ worden alle veelvoorkomende klinkers en medeklinkers aangeboden in de opdrachten. Elk kind kent weer andere letters, afhankelijk van de omgeving en het aanbod. Om elk kind een eerlijke kans te geven, zijn alle veelvoorkomende klinkers en medeklinkers opgenomen in de opdrachten en is geen selectie gemaakt op basis van moeilijkheid. In niveaueversie ★★★ worden naast klinkers en medeklinkers ook alle lange klanken (aa, oo, ee, uu) aangeboden. Het kind moet dan goed onderscheid kunnen maken tussen een korte /e/ en een lange klank /ee/. Daarom zijn de lange klanken alleen in de moeilijkste versie opgenomen. Daarnaast gaven experts het belang aan van het aanbod van tweetekenklanken (zoals ui, oe en ei). Daarom zijn deze klanken in eerste instantie ook opgenomen in het proef- en kwaliteitsonderzoek. Uit het kwaliteitsonderzoek bleek dat alle tweetekenklanken te moeilijk waren, ook in groep 2. Zie het voorbeeld van de /oe/ in poes; in groep 2 wist slechts 26 procent het goede antwoord op deze opdracht. Daarom zijn tweeklanken uiteindelijk niet opgenomen in de opdrachten Letterkennis.



Voorbeeld (Receptieve) Letterkennis

'Waar zie je de oe in poes?'



## **Activiteiten**

### ***Spreken***

Voor het kwaliteitsonderzoek zijn op basis van literatuur en diverse gesprekken met het veld 24 observatiepunten geformuleerd bij de activiteit Spreken. Na dit onderzoek is geëvalueerd welke observatiepunten qua moeilijkheid het meest geschikt zijn voor de uitgave. De drie observatiepunten 'Spreekt verstaanbaar in een een-op-een gesprek met de leerkracht', 'Kent voldoende woorden om iets te vertellen over zijn directe omgeving', 'Vormt goede korte, enkelvoudige zinnen', hadden elk een p-waarde van rond de ,95. Deze observatiepunten bleken te makkelijk en maakten niet genoeg onderscheid tussen de vaardige en minder vaardige kinderen en vielen daarom af voor de uitgave. Na de selectie bleven 21 observatiepunten over. Dit betrof 12 dichotome observatiepunten en 9 polytome observatiepunten.

### ***Gesprekken***

Voor het kwaliteitsonderzoek zijn 10 observatiepunten geformuleerd bij de activiteit die inhoudelijk goed passen bij de vaardigheid Gesprekken. Uit de analyses na dit onderzoek bleek dat de twee observatiepunten 'Kan korte tijd (enkele minuten) bij het gespreksonderwerp blijven in een een-op-een gesprek' en 'Laat de leerkracht uitpraten in een een-op-een gesprek' een p-waarde hadden van respectievelijk ,96 en ,91. Deze observatiepunten bleken te makkelijk en maakten niet genoeg onderscheid tussen de vaardige en minder vaardige kinderen en vielen daarom af voor de uitgave. Na de selectie bleven acht observatiepunten over, alle acht polytome observatiepunten.

### ***Oriëntatie op boek en verhaal***

Bij het voorlezen van een prentenboek zijn er drie fasen te onderscheiden: voor het lezen, tijdens het lezen en na het lezen. Uit data blijkt dat de eerste fase het makkelijkst is, de tweede fase wat moeilijker en de derde fase het moeilijkst. In de laatste fase komen vaardigheden aan bod als samenvatten, verbanden leggen en het begrijpen van boekentaal. De vaardigheid van kinderen in boekoriëntatie en verhaalbegrip neemt toe gedurende de kleuterperiode. Omdat observatiepunten van uiteenlopende moeilijkheid zijn opgenomen in deze activiteit, is die ontwikkeling goed te volgen.

Voor het kwaliteitsonderzoek zijn op basis van literatuur en diverse gesprekken met het veld 16 observatiepunten geformuleerd bij deze activiteit. Na dit onderzoek is geëvalueerd welke observatiepunten qua moeilijkheid het meest geschikt zijn voor de uitgave. De drie observatiepunten 'Weet wat de voor- en achterkant van een boek is', 'Kan op juiste wijze een bladzijde omslaan' en 'Kan aan de hand van een boek laten zien dat een boek wordt gelezen van voor naar achter', hadden elk een p-waarde van rond de ,95. Deze observatiepunten bleken te makkelijk en maakten niet genoeg onderscheid tussen de vaardige en minder vaardige kinderen en vielen daarom af voor de uitgave. Het observatiepunt 'Kan zelf kenmerken van een boek benoemen (zoals titel, binnenkant, rug, schrijver en illustrator)' bleek te moeilijk (p-waarde van ,32 in groep 1) en viel daarom af voor de uitgave. Na de selectie bleven 12 observatiepunten over. Dit betrof twee dichotome observatiepunten en tien polytome observatiepunten.

### ***Oriëntatie op geschreven taal***

Bij de activiteit Oriëntatie op geschreven taal voert het kind voorbereidende lees- en schrijfactiviteiten uit. Uit de data blijkt dat de observatiepunten van beide onderdelen toenemen in moeilijkheid, de eerste punten zijn makkelijker dan de laatste punten. De vaardigheid van kinderen in voorbereidend lezen en schrijven neemt toe gedurende de kleuterperiode. Omdat observatiepunten van uiteenlopende moeilijkheid zijn opgenomen in deze activiteit is die ontwikkeling goed te volgen.

Voor het kwaliteitsonderzoek zijn 12 observatiepunten geformuleerd bij de activiteit die inhoudelijk goed passen bij Oriëntatie op geschreven taal. Na het onderzoek bleek het observatiepunt 'Herkent dezelfde woordkaartjes' voor alle kinderen veel te makkelijk, zowel in groep 1 als 2 was de p-waarde ,99. Dit punt is daarom afgevallen voor de uitgave.

Leg 4 paar woordkaartjes (woordparen) door elkaar gehusseld neer voor het kind (op afstand van elkaar, geordend). Wijs een woordkaartje aan. Kun jij hetzelfde woordje zoeken als dit woordje? Leg hetzelfde kaartje er maar naast. Doe dit bij alle woordparen.

P-waarde	
Groep 1	.99
Groep 2	.99

#### Voorbeeld observatiepunt Herkent dezelfde woordkaartjes

Bij deze activiteit wordt bij een aantal observatiepunten gevraagd naar het aantal handelingen dat een kind correct uitvoert. Bijvoorbeeld, het eerste observatiepunt gaat over het kunnen benoemen van pictogrammen. Bij dit observatiepunt worden zes kaartjes met pictogrammen gebruikt. In het kwaliteitsonderzoek is leerkrachten gevraagd hoeveel pictogrammen elk kind goed kon benoemen (0-6). Na het onderzoek zijn de gegevens geanalyseerd en gehercodeerd tot een driepuntsschaal (nee, beetje, ja), zoals bij de andere observatiepunten. In de uitgave is de driepuntsschaal opgenomen (nee, beetje, ja) waarbij de toegepaste regel is beschreven: 1-2 goed = nee, 3-4 goed = beetje, 5-6 goed = ja. Bij alle observatiepunten waarbij het gaat om aantal goed zijn de observaties op een vergelijkbare manier gehercodeerd en is in de uitgave de regel beschreven. Na de selectie bleven 11 observatiepunten over. Dit betrof na hercodering vier dichotome observatiepunten en zeven polytome observatiepunten.

#### Leerkrachtroute

Zoals beschreven in hoofdstuk 3, bevat de leerkrachtroute 16 observatiepunten (alle 16 onderdelen behorend bij de acht subdomeinen). Per observatiepunt kan de leerkracht kiezen uit vijf niveaus: <E1, E1, M2, E2 of >E2. Daarnaast is ter ondersteuning bij elk observatiepunt een beschrijving opgenomen bij E1 en E2 gebaseerd op de literatuur (bronnen SLO en Expertisecentrum Nederlands). De adequaatheid van deze beschrijvingen en voorbeelden zijn gevalideerd aan de observaties uit de kindroute.

Bij de observatiepunten voor Woordenschat, Luisteren, Fonologisch bewustzijn en Letterkennis worden in de beschrijving voorbeelden genoemd, die zijn ontleend aan de opdrachten in de kindroute. Bijvoorbeeld, het observatiepunt voor Woordenschat is als volgt beschreven:

“Kent voldoende woorden om de inhoud van gesprekken en verhalen over zijn directe omgeving te volgen: thuis, de mensen die hij kent (familie, vriendjes, leerkracht) en school.

Toelichting: zijn receptieve woordenschat bestaat vooral uit concrete woorden die veel voorkomen in het dagelijkse taalgebruik, zoals bestek en thermometer.”

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 rond de 50% in groep 1. In het bovengenoemde voorbeeld kende 76% van de kinderen uit groep 1 het woord “bestek” en 90% van de kinderen uit groep 2. Het woord “zwanger” dat opgenomen is als voorbeeld in de toelichting bij E2 kende 75% van de kinderen uit groep 2 en 55% van de kinderen uit groep 1.

Voor de observatiepunten in de leerkrachtroute van Gesprekken, Spreken, Oriëntatie op boek en verhaal en Oriëntatie op geschreven taal 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 50% in groep 1.

In een aantal gevallen kon de omschrijving verbeterd worden. Bijvoorbeeld, bij het observatiepunt Voorbereidend lezen (van het subdomein Oriëntatie op geschreven taal) staat in de uitgave als beschrijving onder E1 “Herkennt zijn eigen naam” en onder E2 “Kan 12 of meer letters benoemen” of “Kan goed pictogrammen benoemen”. Uit de observaties van de kindroute bleek dat 78% van de kinderen uit groep 1 de eigen naam kon herkennen en 96% van de kinderen uit groep 2. Uit groep 2 kon 72% 12 of meer letters benoemen (tegenover 41% uit groep 1). Pictogrammen benoemen stond oorspronkelijk als voorbeeld bij E1, maar de gemiddelde p-waarde in groep 1 was 0,48 en in groep 2 ,79. “Kan goed pictogrammen benoemen” is daarom in de uitgave als voorbeeld bij E2 geplaatst en niet onder E1. 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 15 tot 20 opdrachten. Alleen bij letterkennis zijn 22 tot 26 letters aangeboden, zodat elk kind kan aangeven welke letters hij of zij al kent. Aan de item-restcorrelatie (Rir) en de item-testcorrelatie (Rit) 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. Bij Woordenschat en Luisteren is het verschil in de geschatte moeilijkheid tussen de niveauversies vrij klein. Het maximale verschil tussen niveauversies in de gemiddelde p-waarde is respectievelijk 0,11 (WS) en 0,04 (LU). Bij Fonologisch bewustzijn en Letterkennis is het maximale verschil wat groter: 0,13 (FB) en 0,19 (LK). Bij Fonologisch bewustzijn is de niveauversie voor midden groep 2 (★★) vrij gemakkelijk (geschatte gemiddelde p-waarde = 0,83) en bij Letterkennis is de niveauversie voor groep 1 (★) vrij moeilijk (geschatte gemiddelde p-waarde is 0,60). Voor groep 2 leerlingen zijn, zoals beoogd, de midden groep 2 niveauversies (★★) makkelijker dan de eind groep 2 versies (★★★).

Tabel 4.8 Beschrijvende gegevens van de niveauversies met opdrachten

Sub-domein	Versie	Groep	Max.	Som			Theta		Gem. p.	Gem. Rit	Gem. Rir
				Gem.	St.dev.	Gem.	St.dev.				
WS	★	Groep 1	16	11,30	2,99	0,22	0,35	0,70	0,43	0,30	
	★★	Groep 2	15	12,10	2,59	0,41	0,43	0,81	0,46	0,33	
	★★★	Groep 2	16	11,90	2,86	0,59	0,46	0,75	0,43	0,31	
LU	★	Groep 1	20	14,40	3,41	0,24	0,31	0,72	0,39	0,27	
	★★	Groep 2	20	14,60	3,36	0,39	0,33	0,73	0,39	0,28	
	★★★	Groep 2	19	13,20	3,29	0,40	0,33	0,69	0,40	0,28	
FB	★	Groep 1	15	9,80	3,36	0,09	0,34	0,65	0,48	0,36	
	★★	Groep 2	15	12,40	3,06	0,24	0,44	0,83	0,55	0,45	
	★★★	Groep 2	15	11,70	3,13	0,45	0,50	0,78	0,54	0,44	
LK	★	Groep 1	22	13,20	5,18	0,13	0,37	0,60	0,50	0,42	
	★★	Groep 2	22	17,50	4,34	0,36	0,42	0,79	0,51	0,44	
	★★★	Groep 2	26	20,50	5,10	0,41	0,42	0,79	0,50	0,44	

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>SP</b>	Groep 1	21	30	18,10	9,48	0,12	0,78	0,58	0,73	0,70
	Groep 2	21	30	22,90	6,75	0,49	0,61	0,74	0,63	0,58
<b>GS</b>	Groep 1	8	16	10,90	3,75	0,37	0,53	0,68	0,67	0,56
	Groep 2	8	16	12,50	3,57	0,42	0,55	0,78	0,69	0,59
<b>BV</b>	Groep 1	12	22	12,20	5,12	0,10	0,50	0,57	0,65	0,57
	Groep 2	12	22	16,40	4,38	0,44	0,50	0,76	0,63	0,55
<b>GT</b>	Groep 1	11	18	8,50	4,28	-0,06	0,59	0,48	0,62	0,53
	Groep 2	11	18	13,40	3,84	0,43	0,58	0,76	0,62	0,54

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. 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, Tabel 4.10. In groep 1 is de modus voor 13 observatiepunten E1. Alleen voor twee onderdelen van Fonologisch bewustzijn en een onderdeel van Oriëntatie op geschreven taal is de modus <E1. Slechts bij twee onderdelen komt >E2 voor, bij alle overige 14 onderdelen varieert het geobserveerde functioneringsniveau tussen <E1 en E2.

Bij groep 2 variëren alle onderdelen tussen <E1 en >E2. Bij de helft van de observatiepunten ligt de modus bij M2 en bij de andere helft bij E2.

Tabel 4.10 Beschrijvende gegevens van de leerkrachtroute

	Groep 1					Groep 2				
	<E1	E1	M2	E2	>E2	<E1	E1	M2	E2	>E2
Woordenschat	16	56	18	9	1	3	12	40	39	7
Begrijpend luisteren	12	60	19	8	0	2	10	39	43	6
Spreken: verstaanbaarheid	26	45	18	11	0	4	11	39	41	5
Spreken: productieve woordenschat	17	55	19	9	0	2	14	43	35	5
Spreken: grammatica	27	50	18	6	0	4	15	40	38	3
Spreken: verhalen vertellen	27	49	17	7	0	4	16	43	33	3
Gesprekken: vasthouden onderwerp	18	57	18	7	0	2	11	40	42	4
Gesprekken: gespreksregels	21	53	19	7	0	3	14	42	37	4
Fonologisch bewustzijn: rijm	31	47	13	9	0	5	11	32	45	6
Fonologisch bewustzijn: beginfoneem	50	33	10	7	0	6	14	31	42	7
Fonologisch bewustzijn: synthese	52	33	8	7	0	5	15	34	39	7
Receptieve letterkennis	33	53	10	4	0	4	19	37	30	10
Oriëntatie op boek en verhaal: boekoriëntatie	22	56	17	5	1	2	9	37	48	4
Oriëntatie op boek en verhaal: verhaalbegrip	33	49	14	4	0	4	13	46	34	3
Oriëntatie op geschreven taal: lezen	25	60	12	3	0	3	18	41	29	10
Oriëntatie op geschreven taal: schrijven	43	42	10	4	0	5	13	37	39	6



## 5 Kalibratie kindroute

Voor de constructie van de kindroute van Kleuter in beeld - Taal 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 taalvaardigheid 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 taalvaardigheid uit op dat subdomein, waarbij een groter getal wijst op een grotere taalvaardigheid. 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 taalvaardigheid 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) heeft 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 die 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 drie subdomeinen in de kindroute van Kleuter in beeld - Taal (Luisteren, Woordenschat en Fonologisch bewustzijn).



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>4</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.

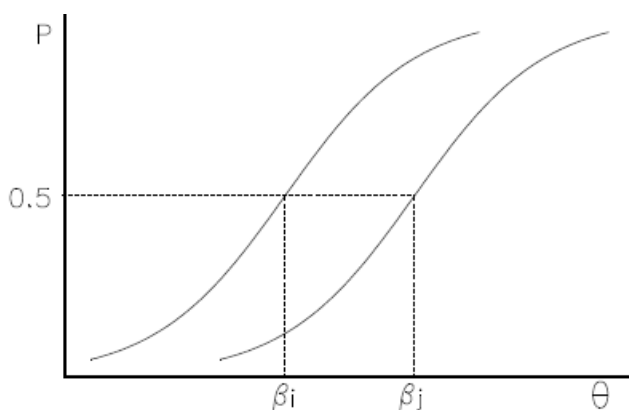
Een eenvoudig en zeer populair voorbeeld is het Raschmodel (Rasch, 1960) waarin  $f_i(\theta)$  gegeven is door

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

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



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

<sup>4</sup> Dit maakt duidelijk waarom men de modellen die vallen onder de IRT, ook wel aanduidt met 'latente trek'-modellen.

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.

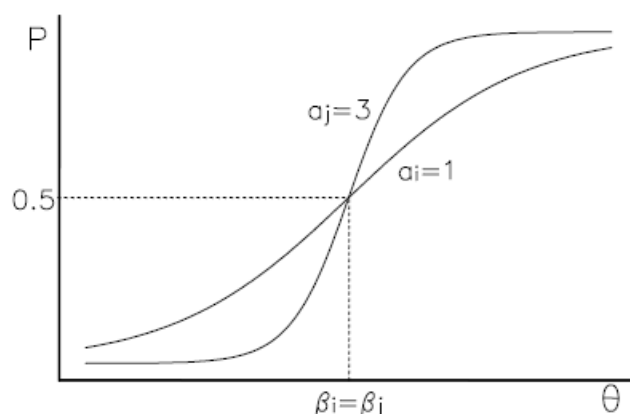
Alvorens het hier gebruikte model te introduceren, is eerst een kanttekening nodig bij het schatten van de moeilijkheidsparameters in het Raschmodel. Een vaak toegepaste schattingsmethode is de 'conditionele grootste aannemelijkheidsmethode' (in het Engels: *Conditional Maximum Likelihood*, verder aangeduid 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:

$$S = \sum_{i=1}^k a_i x_i \quad (5.5)$$

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 toetsingsgrootte 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:

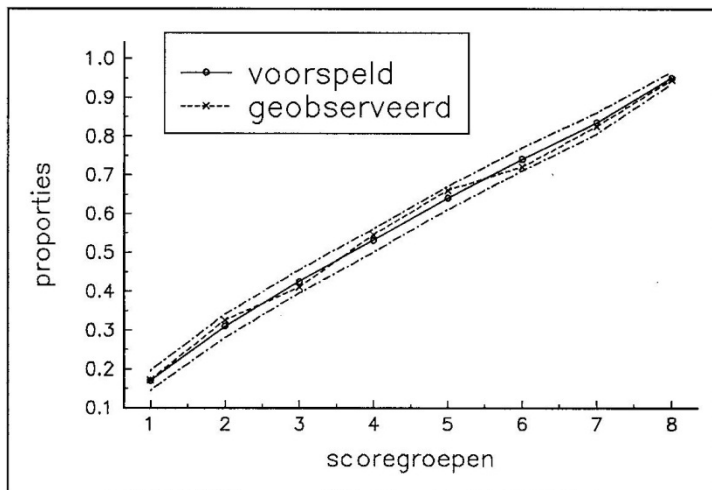
- (1) Met behulp van het programma OPCAT worden de discriminatie-indices in OPLM ingesteld.
- (2) Vervolgens worden de itemparameters geschat met behulp van de CML-methode.
- (3) Met behulp van de M-toetsen wordt gecontroleerd of de discriminatie-indices goed zijn ingesteld.
- (4) Een volgende controle 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, leerjaar 2 en leerjaar 3. Hierna zijn de stappen doorlopen met uitsluitend de observaties uit de doelgroep: 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**

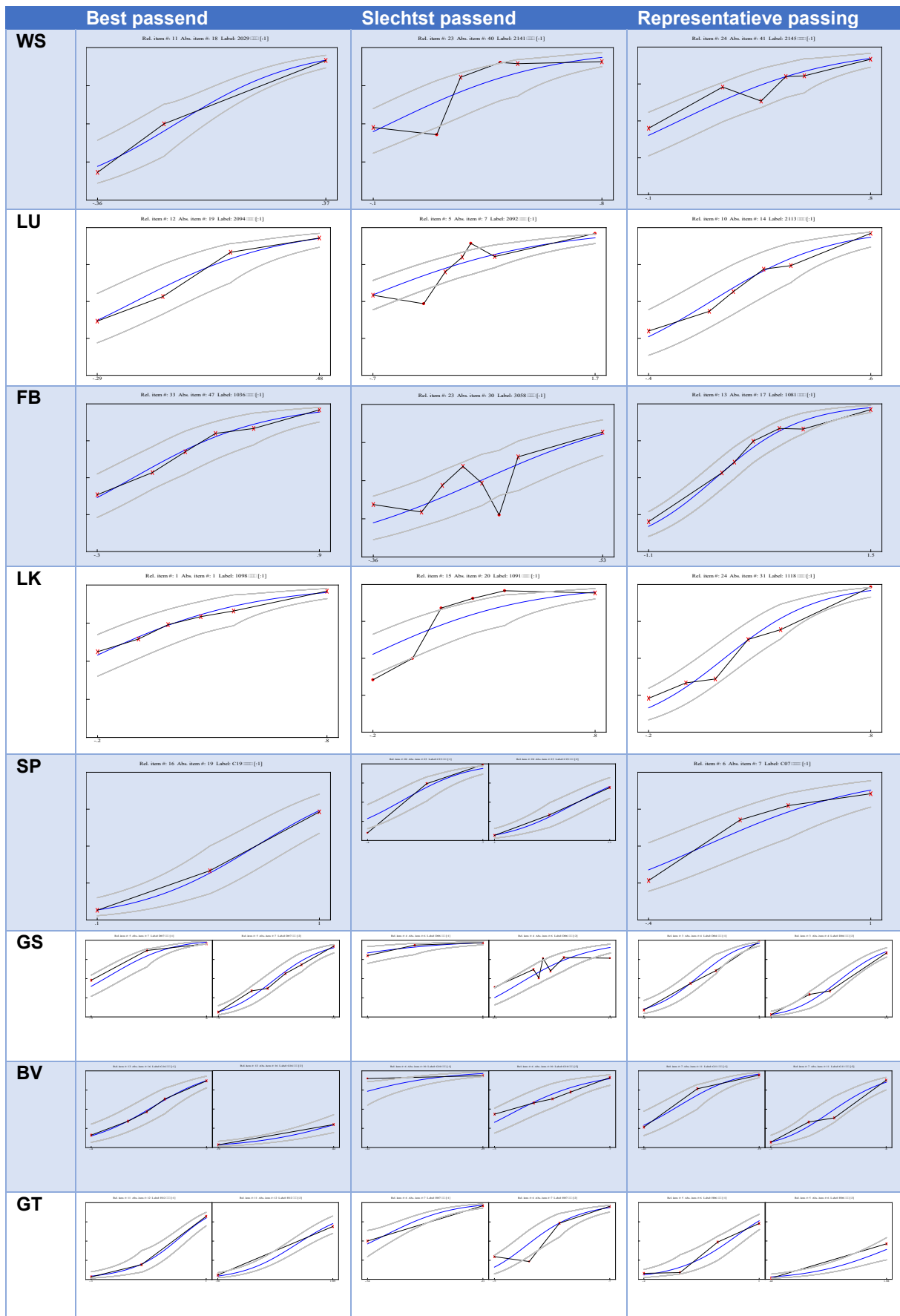
Het is niet eenvoudig om de kwaliteit van de kalibratie aan te tonen. De belangrijkste statistische instrumenten om de passing van een item in het IRT-model te bewerkstelligen en uiteindelijk te documenteren betreffen de hierboven al besproken S-toetsen. Het lastige daarvan is, dat de toetsing voor een 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-toetsingsgrootte (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. Er wordt in dit geval voor een deel (van de onderscheiden scoregroepen) niet beantwoord aan de eis dat de geobserveerde proportie binnen het 95%- betrouwbaarheidsinterval van de geschatte proporties ligt. Alleen bij gesprekken zien we dat bij de meeste observatiepunten een deel van de onderscheiden scoregroepen niet beantwoordt aan die eis. Bij de andere subdomeinen en de leerkrachtroute voldoen de meeste opdrachten/observatiepunten wel aan die eis 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 - Taal 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 - Taal. 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

WS	0	1	2	3	4	5	6	7	8	9	1	
	0/	2/	1	3	4	4	2	5	2	4	10	1
LU	0	1	2	3	4	5	6	7	8	9	1	
	2/	0/	3	5	3	2	2	3	8	2	3	6
FB	0	1	2	3	4	5	6	7	8	9	1	
	0/	3/	0	7	3	2	2	5	3	2	6	6
LK	0	1	2	3	4	5	6	7	8	9	1	
	2/	3/	1	2	4	1	2	3	2	3	2	1
SP	0	1	2	3	4	5	6	7	8	9	1	
	0/	1/	2	1	1	3	1	0	2	2	3	12
GS	0	1	2	3	4	5	6	7	8	9	1	
	3/	3/	1	1	0	1	1	0	3	1	1	1
GT	0	1	2	3	4	5	6	7	8	9	1	
	1/	0/	1	4	1	1	1	2	0	1	1	3
BV	0	1	2	3	4	5	6	7	8	9	1	
	0/	3/	3	0	0	1	2	1	4	2	0	6

In Tabel 5.2 zijn de R1c-waarden weergegeven voor dezelfde afnames waarvoor in Tabel 5.1 de resultaten van de S-toetsen zijn weergegeven. 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. Met uitzondering van het subdomein Gesprekken geldt voor de subdomeinen dat R1c niet groter is dan ongeveer anderhalf maal het aantal vrijheidsgraden. De subdomeinen Spreken en (Receptieve) Letterkennis zijn daarnaast niet significant. De modelpassing is dus acceptabel.

Tabel 5.2 R1c-waarden voor de subdomeinen

Subdomein	R1c	df	P
<b>WS</b>	R1c* = 414,083	df = 314	p = 0,000
<b>LU</b>	R1c* = 504,075	df = 345	p = 0,000
<b>FB</b>	R1c* = 361,513	df = 292	p = 0,003
<b>LK</b>	R1c* = 202,846	df = 175	p = 0,072
<b>GS</b>	R1c* = 264,866	df = 120	p = 0,000
<b>SP</b>	R1c* = 80,425	df = 87	p = 0,678
<b>BV</b>	R1c* = 137,843	df = 105	p = 0,016
<b>GT</b>	R1c* = 154,023	df = 102	p = 0,001

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 niveauseries weergegeven. De gemiddelde waarde van de constante is goed in alle niveauseries. Bij de grote meerderheid van alle itemparameters (86%) is de nauwkeurigheid goed (een  $c \leq ,2$ ).

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

Groep	Versie	Gemiddelde	Goed $c \leq 0,2$	Voldoende $0,3 \leq c \leq 0,4$	Onvoldoende $c \geq 0,5$			
<b>WS</b>	1	★	0,18	14	1	0	0	1
	2	★★	0,14	12	3	0	0	0
	2	★★★	0,16	13	0	3	0	0
<b>LU</b>	1	★	0,19	12	8	0	0	0
	2	★★	0,17	13	7	0	0	0
	2	★★★	0,17	15	4	0	0	0
<b>FB</b>	1	★	0,15	13	2	0	0	0
	2	★★	0,09	15	0	0	0	0
	2	★★★	0,12	14	0	1	0	0
<b>LK</b>	1	★	0,11	21	0	1	0	0
	2	★★	0,09	21	1	0	0	0
	2	★★★	0,09	25	1	0	0	0
<b>SP</b>	1	–	0,12	16	1	0	0	0
	2	–	0,17	16	0	1	0	0
<b>GS</b>	1	–	0,14	13	2	0	1	0
	2	–	0,11	15	0	1	0	0
<b>BV</b>	1	–	0,15	18	3	1	0	0
	2	–	0,15	19	2	1	0	0
<b>GT</b>	1	–	0,18	14	1	2	1	0
	2	–	0,16	15	1	2	0	0



Op basis van de hierboven beschreven resultaten kan geconcludeerd worden dat de kalibratie voor de kindroute van Kleuter in beeld - Taal 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 stemmen. Het doel van de standaardbepaling is om met leerkrachten en andere onderwijsprofessionals na te gaan wat qua taal doorgaans gezien wordt bij kleuters 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 de twee domeinen, Mondelinge taal en Beginnende geletterdheid, de grenzen bepaald te worden tussen vijf functioneringsniveaus (<E1, E1, M2, E2 en >E2). Tijdens de standaardbepaling worden per domein dus uiteindelijk vier standaarden bepaald.

### 6.2 Methode en experts

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 - Taal bestaat uit acht clusters, namelijk elk subdomein vormt een cluster:

- Mondelinge taal bestaat uit vier clusters: Luisteren, Woordenschat, Spreken en Gesprekken.
- Beginnende geletterdheid bestaat ook uit vier clusters: Fonologisch bewustzijn, Receptieve letterkennis, Oriëntatie op boek en verhaal, Oriëntatie op geschreven taal.

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

Op twee verschillende dagen, in juni en 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.

Voorafgaand aan de eerste standaardbepalingsdag is met behulp van 18 onderwijsprofessionals de procedure uitgeprobeerd en verbeterd. Het was onduidelijk of het mogelijk zou zijn om alle grenzen op één dag te zetten en bij welke grens het beste begonnen kon worden. Tijdens deze eerste dag bleek het mogelijk te zijn om alle grenzen op dezelfde dag te zetten. Wel werd duidelijk dat het beter was om met de onderste grens te beginnen (onder E1/E1).

In totaal hebben 31 onderwijsprofessionals zich aangemeld, waarvan er uiteindelijk 20 ook mee hebben gedaan aan de twee standaardbepalingsdagen: 10 kleuterleerkrachten, 2 intern begeleiders, 3 kleuterleerkrachten die tevens intern begeleider zijn en 5 experts op het gebied van taal (onderwijsadviseurs/-ontwikkelaars/logopedist).

Voorafgaand aan de standaardbepalingsdag kregen de deelnemers informatie over leerlijnen. Elke standaardbepalingsdag begon met een uitleg over het ontwikkelingsproces van Kleuter in beeld - Taal en een beschrijving van de inhoud van het instrument. Vervolgens werd het doel van de standaardbepaling en de methode uitgelegd. Daarna is de groep in tweeën gesplitst. De ene helft van de deelnemers zette standaarden voor Mondelinge taal en de andere helft voor Beginnende geletterdheid.

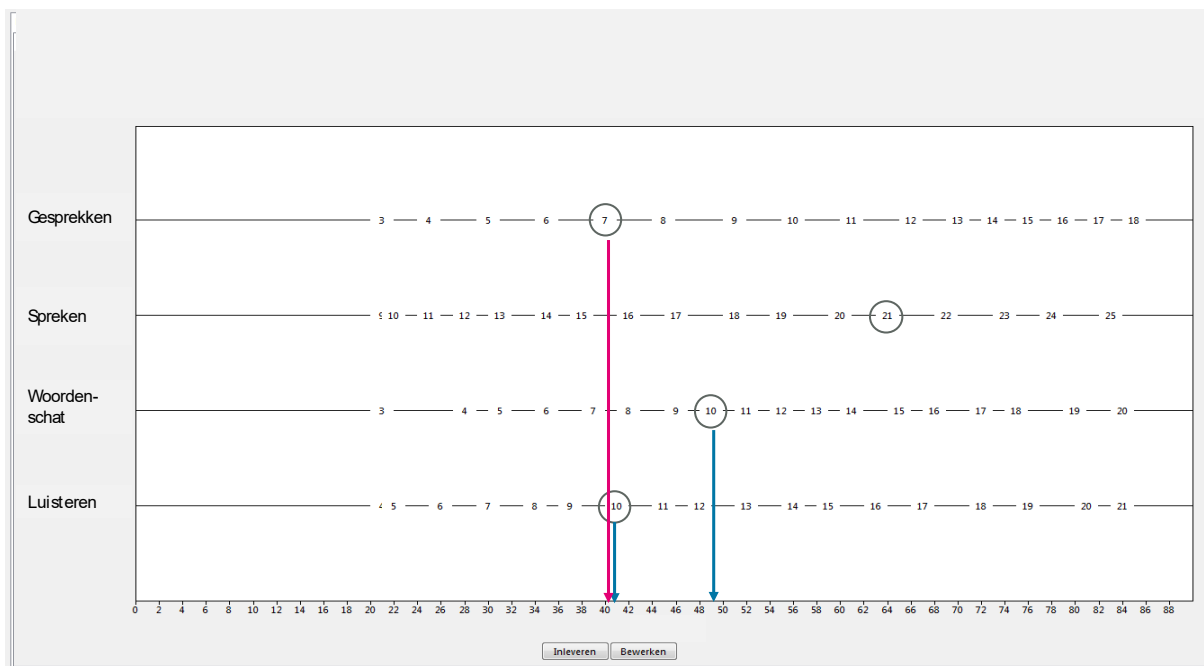
Tijdens de sessie kregen de experts voor elk subdomein een cluster met opdrachten of observatiepunten te zien en vormden zij zich een oordeel over hoeveel opdrachten/observatiepunten het kind naar verwachting correct zou hebben op dat cluster als zijn/haar vaardigheid zich precies Onder E1 van Mondelinge taal/Beginnende geletterdheid 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 Mondelinge taal op één vaardigheidsschaal te leggen en alle opdrachten en observatiepunten van Beginnende geletterdheid op één vaardigheidsschaal te leggen met behulp van het One Parameter Logistic Model (Verhelst, Glas, & Verstralen, 1995). Voor beide domeinen is met behulp van simulaties voor elk mogelijk aantal correcte opdrachten/observatiepunten op het volledige domein 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 model is niet gebruikt voor het uitdrukken van de standaard.

Het aantal correcte opdrachten per cluster is gerelateerd aan het totaal aantal correcte opdrachten voor het domein (bijvoorbeeld Mondelinge taal). Bijvoorbeeld, op het formulier in Figuur 6.1 is te zien dat zeven correcte observatiepunten bij *Gesprekken* te verwachten zouden zijn bij een leerling die in totaal ongeveer 40 correcte opdrachten/punten heeft en tien correcte opdrachten bij *Luisteren*. Op dit formulier zien de experts dus de relatieve moeilijkheid van het cluster in verhouding tot de andere clusters. De visuele relatie tussen de vier subdomeinen helpt de experts een realistisch oordeel te geven.

Figuur 6.1 Een fictief standaardbepalingsformulier



### Oordelen van de experts

Nadat de experts hun oordelen hebben 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 beoordelingsspectrum 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) en ten slotte de vierde grens (E2/boven E2). Op elke dag gaven de onderwijsprofessionals dus voor elk subdomein een oordeel over de onder- en bovengrens eind leerjaar 1 (E1) en een onder- en bovengrens voor eind leerjaar 2 (E2).

## 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 elk subdomein is een cluster gemaakt.

- Mondelinge taal bestaat uit vier clusters: twee clusters met opdrachten (Woordenschat en Luisteren) en twee clusters met observatiepunten van activiteiten (Spreken en Gesprekken).
- Beginnende geletterdheid bestaat in totaal ook uit vier clusters: twee clusters met opdrachten (Fonologisch bewustzijn en Receptieve letterkennis) en twee clusters met observatiepunten van activiteiten (Oriëntatie op boek en verhaal en Oriëntatie op geschreven taal).

Voor de activiteiten (Spreken, Gesprekken, Oriëntatie op boek en verhaal, Oriëntatie op geschreven taal) zijn alle observatiepunten van de activiteiten voorgelegd. Ook voor Receptieve letterkennis zijn alle opdrachten van de E1- en M2-versies voorgelegd. Bij Woordenschat, Luisteren en Fonologisch bewustzijn 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
Woordenschat	Bijvoeglijk naamwoorden	5	13%	3	12%
	Werkwoorden	10	26%	7	28%
	Zelfstandig naamwoorden <sup>5</sup>	23	61%	15	60%
		38		25	
Luisteren	Verhaalbegrip	8	21%	5	20%
	Zinsbegrip	30	79%	20	80%
		38		25	
Fonologisch bewustzijn	Beginfoneem	16	41%	10	40%
	Rijm	10	26%	7	28%
	Synthese	13	33%	8	32%
		39		25	

## 6.5 Definitieve expertoordelen

### Mondelinge taal

De oordelen van de experts per cluster (subdomein) van Mondelinge taal zijn te zien in Tabel 6.4. Te zien is dat bij de eerste grens die is gezet, de ondergrens voor leerjaar 1 bij het subdomein Luisteren, de oordelen van de experts nog vrij veel uiteenliepen. Bij de latere grenzen waren de experts het meer eens (zie Figuur 6.2).

De grens voor het hele domein ( $G_T$ ) per expert, Mondelinge taal of Beginnende geletterdheid, is gelijk aan de som van de expert-oordelen per cluster ( $C_k$ ) en wordt dus niet direct door de experts zelf vastgelegd (zie de kolom "totaal" in Tabel 6.4):  $G_T = \sum C_k$ .

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 te

<sup>5</sup> Alle vier opdrachten over basisregels zijn opgenomen.

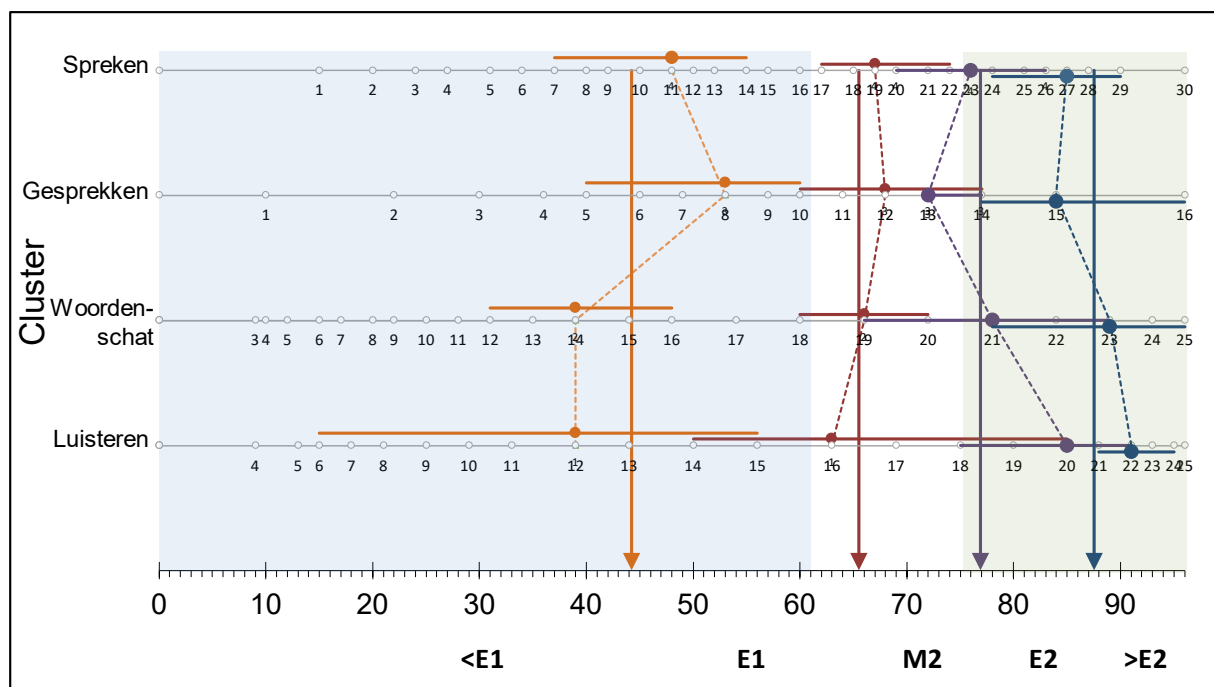
verwachten zou zijn als de beoordelaars een random oordeel geven. Als beoordelaars systematisch strenger of milder zijn heeft dat invloed op de Finn-coëfficiënt, maar niet op de interbeoordelaarscorrelatie. De Finn-coëfficiënt wordt vooral geschikt geacht als de variantie tussen beoordelaars gering is, 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 de schaal van het hele domein.

Aan de Ranking Similarity Index is te zien dat voor elke expert de oordelen redelijk tot goed samenhangen met de oordelen van de andere experts ( $RSI = 0,79 - 0,93$ ). 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 één geval met 1 schaalpunt, in alle andere gevallen met minder dan een punt (op een schaal van 0 tot 96). De gemiddelde interbeoordelaarscorrelatie ( $GIR = 0,94$ ) en de Finn-coëfficiënt (Finn = 0,95) laten zien dat de overeenstemming tussen de experts goed is.

### Standaarden

De vier definitieve standaarden voor Mondelinge taal zijn berekend door het gemiddelde van de expert domeingrenzen te nemen ( $\bar{G}_T$ ). De grenzen staan weergegeven in Figuur 6.2 en in Tabel 6.2. 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 42% van de leerlingen midden leerjaar 1 in het functioneringsniveau Boven E1 zou vallen. De ondergrens van leerjaar 2 is daarentegen vrij hoog: meer dan de helft van de leerlingen midden leerjaar 2 zou in het functioneringsniveau Onder E2 vallen.

Figuur 6.2 Mondelinge taal: Plot van de oordelen per subdomein en de standaarden gebaseerd op de oordelen



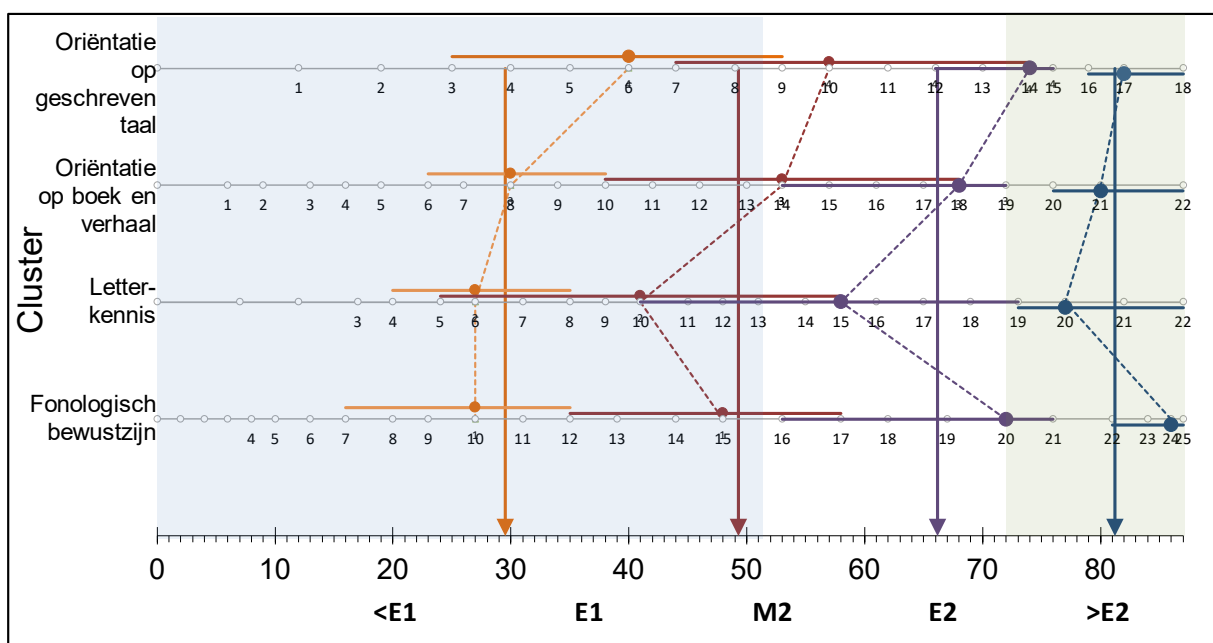
Tabel 6 2 Standaarden Mondelinge taal en de verwachte impact (geschat percentage leerlingen dat onder E1, op E1, op M2, op E2 of boven E2 niveau valt, gegeven de standaarden)

Standaarden MT	Onder E1	Boven E1	Onder E2	Boven E2	
	44,2	65,5	76,9	87,5	
	5,0	1,8	5,3	4,5	
	<E1	E1	M2	E2	>E2
<b>Bereik</b>	0-44	45-65	66-76	77-87	88-96
<b>Leerjaar 1</b>	23%	35%	21%	18%	3%
<b>Leerjaar 2</b>	8%	23%	22%	33%	14%
	<E1	E1	>E1		
<b>Leerjaar 1</b>	23%	35%	42%		
			<E2	E2	>E2
<b>Leerjaar 2</b>			53%	33%	14%

### Beginnende geletterdheid

De oordelen van de experts per cluster (subdomein) van Beginnende geletterdheid zijn te zien in Tabel 6.5. Te zien is dat bij de bovengrens van leerjaar 1 de oordelen van de experts het meeste uiteenlopen (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 = 0,85 – 0,94). De impact van de oordelen van de individuele experts op de domeingrenzen is gering. Als een individuele expert buiten beschouwing wordt gelaten, verandert in alle gevallen de domeingrens minder dan een punt (op een schaal van 0 tot 87). De gemiddelde interbeoordelaarscorrelatie (GIR = 0,93) en de Finn-coëfficiënt (Finn = 0,94) laten zien dat de overeenstemming tussen de experts goed is.

Figuur 6.3 Beginnende geletterdheid: Plot van de oordelen per subdomein en de standaarden gebaseerd op de oordelen





### Standaarden

De vier definitieve standaarden voor Beginnende geletterdheid zijn berekend door het gemiddelde van de expert domeingrenzen te nemen ( $\overline{G_T}$ ). De grenzen staan weergegeven in Figuur 6.3 en in Tabel 6.3. Nadat de standaarden zijn berekend, is geschat voor leerjaar 1 en leerjaar 2 welk percentage leerlingen in elk functioneringsniveau zou vallen (zie Tabel 6.3). Te zien is dat de bovengrens van leerjaar 1 vrij laag is gezet, omdat naar schatting meer dan de helft van de leerlingen midden leerjaar 1 in het functioneringsniveau Boven E1 zou vallen. Dit is ook de grens waarbij de oordelen meer uiteenliepen.

Tabel 6.3 Standaarden Beginnende geletterdheid en de verwachte impact (geschat percentage leerlingen dat onder E1, op E1, op M2, op E2 of boven E2 niveau valt, gegeven de standaarden)

Standaarden BG	Onder E1		Boven E1		Onder E2		Boven E2		
<b>Grens</b>	29,5		49,3		66,2		81,2		
<b>Stdev</b>	4,3		4,4		3,4		2,9		
	<b>&lt;E1</b>		<b>E1</b>		<b>M2</b>		<b>E2</b>		<b>&gt;E2</b>
<b>Bereik</b>	0-29		30-49		50-66		67-81		82-87
<b>Leerjaar 1</b>	18%		29%		30%		21%		2%
<b>Leerjaar 2</b>	3%		10%		22%		44%		21%
	<b>&lt;E1</b>		<b>E1</b>		<b>&gt;E1</b>				
<b>Leerjaar 1</b>	18%		29%		53%				
					<b>&lt;E2</b>		<b>E2</b>		<b>&gt;E2</b>
<b>Leerjaar 2</b>					35%		44%		21%

## 6.6 Conclusie

Onderwijsprofessionals hebben met de 3DC-methode de standaarden bepaald voor de grenzen tussen de vijf functioneringsniveaus bij Mondelinge taal en bij Beginnende geletterdheid. Deze standaarden worden gebruikt om het functioneringsniveau per subdomein te bepalen en te rapporteren op basis van de opdrachten van de niveauversies 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 & Sijtsma, 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 vijf jaar) zal derhalve een controle plaatsvinden van de inhoud en de impact.

Tabel 6.4 Expert oordelen per subdomein voor Mondelinge taal

Experts	Onder E1					Boven E1					Onder E2					Boven E2					Impact				RSI
	Luisteren	Woordenschat	Gesprekken	Spreken	Totaal	Luisteren	Woordenschat	Gesprekken	Spreken	Totaal	Luisteren	Woordenschat	Gesprekken	Spreken	Totaal	Luisteren	Woordenschat	Gesprekken	Spreken	Totaal	Ondergrens E1	Bovengrens E1	Ondergrens E2	Bovengrens E2	
1	8	12	10	9	39	14	18	12	22	66	19	20	14	23	76	23	23	16	27	89	0,6	0,1	0,1	0,2	0,88
2	13	16	8	14	51	16	19	12	19	66	18	20	13	21	72	21	22	15	25	83	0,8	0,1	0,7	0,6	0,91
3	13	13	10	7	43	15	18	13	19	65											0,1	0,1			0,79
4	15	15	10	13	53	17	19	12	21	69											1,0	0,4			0,84
5	13	14	8	10	45	16	19	11	19	65	18	20	13	21	72	21	24	15	24	84	0,1	0,1	0,7	0,5	0,93
6	6	14	8	8	36	14	18	14	19	65	18	19	14	20	71	21	21	15	24	81	0,9	0,1	0,8	0,9	0,84
7	14	16	6	12	48	17	20	11	19	67	21	23	13	25	82	24	25	14	29	92	0,4	0,2	0,7	0,6	0,86
8	12	13	8	11	44	16	19	11	17	63	21	22	14	22	79	23	24	15	29	91	0,0	0,3	0,3	0,5	0,93
9	11	12	8	11	42	20	18	10	17	65	22	23	13	26	84	24	24	14	29	91	0,2	0,1	1,0	0,5	0,85
10	13	13	5	10	41	16	19	11	18	64	20	22	13	24	79	22	24	15	28	89	0,4	0,2	0,3	0,2	0,91
Minimum	6	12	5	7	36	14	18	10	17	63	18	19	13	20	71	21	21	14	24	81					
Maximum	15	16	10	14	53	20	20	14	22	69	22	23	14	26	84	24	25	16	29	92					
Range	9	4	5	7	17	6	2	4	5	6	4	4	1	6	13	3	4	2	5	11					
Gemiddelde	11,8	13,8	8,1	10,5	44,2	16,1	18,7	11,7	19,0	65,5	19,6	21,1	13,4	22,8	76,9	22,4	23,4	14,9	26,9	87,5					
Maximum cluster	25	25	16	30	96	25	25	16	30	96	25	22	22	18	87	25	25	16	30	96					

Opmerking: Experts 3 en 4 moesten vanwege persoonlijke omstandigheden eerder weg en hebben geen oordelen gegeven voor de grenzen voor groep 2.

Tabel 6.5 Expert oordelen per subdomein voor Beginnende geletterdheid

Experts	Onder E1					Boven E1					Onder E2					Boven E2					Impact				RSI
	Fonologisch bewustzijn	Receptieve letterkennis	Oriëntatie op boek en verhaal	Oriëntatie op geschreven taal	Totaal	Fonologisch bewustzijn	Receptieve letterkennis	Oriëntatie op boek en verhaal	Oriëntatie op geschreven taal	Totaal	Fonologisch bewustzijn	Receptieve letterkennis	Oriëntatie op boek en verhaal	Oriëntatie op geschreven taal	Totaal	Fonologisch bewustzijn	Receptieve letterkennis	Oriëntatie op boek en verhaal	Oriëntatie op geschreven taal	Totaal	Ondergrens E1	Bovengrens E1	Ondergrens E2	Bovengrens E2	
1	7	6	8	9	30	12	10	11	11	44	19	15	17	15	66	25	20	22	16	83	0,1	0,6	0,0	0,2	0,91
2	11	6	9	8	34	14	10	13	11	48	20	15	18	14	67	23	19	21	16	79	0,5	0,1	0,1	0,2	0,94
3	7	5	7	4	23	15	11	12	10	48	18	16	18	13	65	22	20	20	17	79	0,7	0,1	0,1	0,2	0,94
4	11	8	8	6	33	17	15	10	9	51	21	19	14	12	66	25	22	22	18	87	0,4	0,2	0,0	0,6	0,86
5	9	5	6	3	23	15	11	14	11	51	20	16	17	15	68	22	20	20	17	79	0,7	0,2	0,2	0,2	0,93
6	9	7	8	7	31	16	11	12	10	49	20	18	18	13	69	23	20	21	16	80	0,2	0,0	0,3	0,1	0,94
7	9	7	8	6	30	16	14	13	10	53	19	17	18	13	67	23	20	20	16	79	0,1	0,4	0,1	0,2	0,93
8	12	5	8	7	32	17	11	17	11	56	21	16	19	15	71	24	19	21	18	82	0,3	0,7	0,5	0,1	0,93
9	11	4	10	9	34	15	5	18	14	52	21	10	19	15	65	25	21	21	18	85	0,5	0,3	0,1	0,4	0,85
10	9	4	9	3	25	13	6	15	7	41	16	10	18	14	58	23	19	20	17	79	0,5	0,9	0,9	0,2	0,89
Minimum	7	4	6	3	23	12	5	10	7	41	16	10	14	12	58	22	19	20	16	79					
Maximum	12	8	10	9	34	17	15	18	14	56	21	19	19	15	71	25	22	22	18	87					
Bereik	5	4	4	6	11	5	10	8	7	15	5	9	5	3	13	3	3	2	2	8					
Gemiddelde	9,5	5,7	8,1	6,2	29,5	15,0	10,4	13,5	10,4	49,3	19,5	15,2	17,6	13,9	66,2	23,5	20,0	20,8	16,9	81,2					
Maximum cluster	25	22	22	18	87	25	22	22	18	87	25	22	22	18	87	25	22	22	18	87					



## 7 Beschrijving algoritmes

Voor Kleuter in beeld - Taal wordt op grond van de observaties het functioneringsniveau van kleuters op subdomeinen gerapporteerd en een signaal gegeven over het hoofddomein Taal. 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

Domein	Subdomein		Leerkracht- route aantal (indirecte) observaties	Kindroute  aantal observatie- punten	aantal opdrachten per versie		
					Groep 1	Midden Groep 2	Eind Groep 2
<b>Mondelinge taal</b>	Woordenschat	(WS)	1		16	15	16
	Luisteren	(LU)	1		20	20	19
	<i>Spreken</i>	(SP)	4	30			
	<i>Gesprekken</i>	(GS)	2	16			
<b>Beginnende geletterdheid</b>	Fonologisch bewustzijn	(FB)	3		15	15	15
	Receptieve letterkennis	(LK)	1		22	22	26
	<i>Oriëntatie op boek en verhaal</i>	(BV)	2	22			
	<i>Oriëntatie op geschreven taal</i>	(GT)	2	18			

### 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 vijfpuntsschaal (zie Figuur 7.1). De oordelen worden gecodeerd van 0 tot 4. In Tabel 7.1 staat het aantal observaties dat de leerkracht per subdomein uitvoert in de leerkrachtroute.

- 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 Luisteren, 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 Oriëntatie op boek en verhaal bestaat uit twee onderdelen: Boekoriëntatie

en Verhaalbegrip. Stel, in observatieperiode E1 schat de leerkracht het kind bij Boekoriëntatie in op E1 en bij Verhaalbegrip 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 vijfpuntsschaal in de leerkrachtroute*

Begrijpend luisteren				
<E1	E1	M2	E2	>E2
☐	☐ Begrijpt eenvoudige, enkelvoudige zinnen en tweevoudige instructies Toelichting: begrijpt instructies zoals 'Geef de schep aan Mats en de emmer aan Khalid'	☐	☐ Begrijpt samengestelde zinnen met veelvoorkomende voegwoorden, zoals toen en omdat, en complexere meervoudige instructies Toelichting: begrijpt zinnen zoals 'Anne neemt haar knuffel en rugtas mee naar oma, maar haar prentenboek laat ze thuis'	☐

**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.

Als  $\bar{o} - \lfloor \bar{o} \rfloor > 0,5$  dan  $f = \lceil \bar{o} \rceil$  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	
<E1	(0)	<E1	(0)	<E1	(0)
<E1	(0)	E1	(1)	<E1	(0)
<E1	(0)	M2	(2)	E1	(1)
<E1	(0)	E2	(3)	E1	(1)
<E1	(0)	>E2	(4)	M2	(2)
E1	(1)	<E1	(0)	<E1	(0)
E1	(1)	E1	(1)	E1	(1)
E1	(1)	M2	(2)	E1	(1)
E1	(1)	E2	(3)	M2	(2)
E1	(1)	>E2	(4)	M2	(2)
M2	(2)	<E1	(0)	E1	(1)
M2	(2)	E1	(1)	E1	(1)
M2	(2)	M2	(2)	M2	(2)
M2	(2)	E2	(3)	M2	(2)
M2	(2)	>E2	(4)	E2	(3)
E2	(3)	<E1	(0)	E1	(1)
E2	(3)	E1	(1)	M2	(2)
E2	(3)	M2	(2)	M2	(2)
E2	(3)	E2	(3)	E2	(3)
E2	(3)	>E2	(4)	E2	(3)
>E2	(4)	<E1	(0)	M2	(2)
>E2	(4)	E1	(1)	M2	(2)
>E2	(4)	M2	(2)	E2	(3)
>E2	(4)	E2	(3)	E2	(3)
>E2	(4)	>E2	(4)	>E2	(4)

### Kindroute

In de Kindroute bestaat zowel Mondelinge taal als Beginnende geletterdheid uit vier subdomeinen. Bij twee 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 vijf functioneringsniveaus voor Mondelinge taal en voor Beginnende geletterdheid. Deze grenzen worden gebruikt om per subdomein de functioneringsniveaus te bepalen.

### Activiteiten

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 ook uitgedrukt worden in bijbehorende vaardigheidsscores. Met behulp van deze vaardigheidsscores is vervolgens geschat hoeveel observatiepunten de kleuter naar verwachting beheerst op een cluster (subdomein), zie Tabel 7.3. Voor de activiteiten zijn dit direct de grenzen, omdat bij de activiteiten de observatiepunten uit de uitgave identiek zijn aan de observatiepunten die gebruikt zijn in de standaardbepaling.

### Opdrachten

Bij de opdrachten daarentegen zijn drie niveaueversies uitgegeven, daarom zijn de opdrachten die in de drie versies voorkomen niet identiek aan de 25 opdrachten die tijdens de standaardbepaling zijn gebruikt (zie hoofdstuk 6). Om het functioneringsniveau bij de opdrachten te bepalen, moet daarom nog een extra analysestap uitgevoerd worden in vergelijking met de activiteiten (zie stap 4):

- 1) Elke grens die de experts hebben gezet, is uitgedrukt in het aantal punten (opdrachten/observatiepunten) dat de kleuter beheerst per domein als zijn/haar vaardigheid precies op de grens ligt (zie Tabel 7.3).
- 2) Deze punten kunnen ook uitgedrukt worden in bijbehorende vaardigheidsscores.
- 3) Met behulp van deze vaardigheidsscores is vervolgens uitgerekend hoeveel van de 25 opdrachten de kleuter naar verwachting beheerst op een cluster (subdomein) dat is gebruikt in de standaardbepaling, zie Tabel 7.3.
- 4) Voor de extra stap is voor de opdrachten in de niveaueversies (uit de uitgave) een omzettingstabel gemaakt van het aantal correcte opdrachten (de ongewogen scores) naar de vaardigheidsscore. Vervolgens wordt bij de vaardigheidsscores die bij de gezette grenzen horen, (uit stap 2) het bijbehorende correcte aantal opdrachten<sup>6</sup> voor elk van de drie niveaueversies in de uitgave gezocht.

Tabel 7.3 Grenzen uit de standaardbepaling en op de clusters van de subdomeinen

Domein	Grens	<E1 -			E2 -
		E1	E1 - M2	M2 - E2	>E2
<b>Mondelinge taal</b>	Aantal beheerste punten	44,2	65,5	76,9	87,5
	Vaardigheidsscore	-0,014	0,302	0,515	0,857
	<b>Punten Clusters:</b>				
	<i>Gesprekken (GS)</i>	5,8	11,4	14,3	
	<i>Spreken (SP)</i>	9,8	18,5	23,6	28,5
	Woordenschat (WS)	15,3	19,3	21,2	23,1
	Luisteren (LU)	13,2	16,7	18,7	21,1
<b>Beginnende geletterdheid</b>	Aantal beheerste punten	29,5	49,3	66,2	81,2
	Vaardigheidsscore	-0,240	0,107	0,446	0,968
	<b>Punten Clusters:</b>				
	<i>Oriëntatie op boek en verhaal (BV)</i>	8,0	13,0	17,7	21,5
	<i>Oriëntatie op geschreven taal (GT)</i>	3,7	8,3	12,2	17,0
	Fonologisch bewustzijn (FB)	10,8	15,5	19,2	22,6
	Receptieve letterkennis (LK)	6,6	12,6	17,7	21,1

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

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

- Als de kleuter de observatiepunten/opdrachten ( $\sum o_i$ ) in mindere mate beheerst dan de eerste grens ( $g_{1i}$ ) dan is het gerapporteerde functioneringsniveau <E1 ( $f_{1i}$ ).  
IF ( $\sum o_i < g_{1i}$ ) THEN  $f_{1i}$

<sup>6</sup> Via interpolatie bepaald op 1 decimaal nauwkeurig



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 E1.

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 M2.

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 E2.

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 meerdere mate beheerst dan de vierde grens dan is het gerapporteerde functioneringsniveau  $>E2$ .

IF ( $\sum o_i \geq g_{4i}$ ) THEN  $f_{5j}$

In Tabel 7.4 staan de grenzen weergegeven die zijn toegepast bij de uitgave van 2019.

Tabel 7.4 Bereik aantal correcte observatiepunten/opdrachten en bijbehorend gerapporteerde functioneringsniveau

Schaal	Subdomein	Versie	Max	$g_1$	$g_2$	$g_3$	$g_4$	<E1 ( $f_{1i}$ )	E1 ( $f_{2i}$ )	M2 ( $f_{3i}$ )	E2 ( $f_{4i}$ )	>E2 ( $u_{5i}$ )
Beginnende geletterdheid	Oriëntatie op boek en verhaal (BV)		22	8	13	17,7	21,5	0-7	8-12	13-18*	19-21	22
	Oriëntatie op geschreven taal (GT)		18	3,7	8,3	12,2	17,0	0-3	4-8	9-12	13-16*	17-18
	Fonologisch bewustzijn (FB)	★	15	7,3	10,6	12,8	14,4	0-7	8-10	11-12	13-14	15
		★★	15	6,3	9,3	11,7	14,0	0-6	7-9	10-12*	13-14	15
		★★★	15	5,6	8,4	10,9	13,3	0-6*	7-8	9-11*	12-13	14-15
	Receptieve letterkennis (LK)	★	22	6,6	12,6	17,7	21,1	0-6	7-12	13-17	18-21	22
★★		22	6,6	12,6	17,7	21,1	0-6	7-12	13-17	18-21	22	
★★★		26	7,7	14,7	20,6	24,7	0-7	8-14	15-20	21-24	25-26	
Mondelinge taal	Gesprekken (GS)		16	5,8	11,4	14,3	**	0-6*	7-11	12-14	15**	16**
			30	9,8	18,5	23,6	28,5	0-10*	11-18	19-23	24-28	29-30
	Woordenschat (WS)	★	16	10,7	13,4	14,4	15,4	0-10	11-13	14	15	16
		★★	15	8,9	11,6	12,9	14,1	0-8	9-11	12-13*	14	15
		★★★	16	8,2	11,2	12,7	14,2	0-8	9-11	12-13*	14	15-16
	Luisteren (LU)	★	20	12,4	15,5	17,0	18,6	0-12	13-15	16-17*	18-19*	20
		★★	20	10,6	13,9	15,7	17,6	0-11*	12-13	14-16*	17-18*	19-20
		★★★	19	9,3	12,4	14,2	16,2	0-9	10-12	13-14	15-17*	18-19

Opmerkingen:

\* Kleiner of gelijk aan de grens (i.p.v. kleiner)

\*\* Op dit cluster kan een kleuter de observatiepunten niet in meerdere mate dan de vierde grens beheersen. Om deze reden wordt het vijfde functioneringsniveau gerapporteerde als de kleuter het maximale aantal punten (16) heeft bereikt en het vierde als de kleuter het maximum minus 1 heeft bereikt (15).

## 7.2 Algoritme voor een signaal voor Taal

Het onderwijsveld gaf aan dat het, naast het niveau voor elk van de subdomeinen, ook een signaal voor het hoofddomein Taal wil om zicht te krijgen op hoe elke kleuter het doet op Taal 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 taalsterke kinderen. Dit signaal wordt gegeven wanneer zes of meer subdomeinen boven verwachting zijn. Dat betekent dat de kleuter dus op beide domeinen (Mondelinge taal én Beginnende geletterdheid) op minimaal twee subdomeinen boven verwachting scoort. Het signaal ‘wees alert’ is bedoeld voor het signaleren van de (relatief) taalzwakke kinderen: kinderen met mogelijke taalproblemen en/of risico op latere leesproblemen. Hieronder vallen onder andere kinderen die volgens verwachting scoren op Mondelinge taal, maar onder verwachting op Beginnende geletterdheid (risico op leesproblemen) en andersom (mogelijke taalproblemen). Het is dus belangrijk een signaal te geven aan de leerkracht als een kind uitvalt op minimaal vier 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 mondelinge taal en lezen zoveel mogelijk voorkomen kunnen worden.

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

Er zijn twee domeinen (Mondelinge taal en Beginnende geletterdheid) met elk vier subdomeinen, dus in totaal acht subdomeinen (A, B, C, D, E, F, G, H). Per subdomein is bepaald welk functioneringsniveau ( $f$ ) het kind heeft: 0 (<E1); 1 (E2); 2 (M2); 3 (E2); 4 (>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 ( $p = 1$ ) en Medio leerjaar 1 ( $p=2$ ) is het verwachte niveau <E1 ( $v = 0$ )
- Eind leerjaar 1 ( $p = 3$ ) en Begin leerjaar 2 ( $p=4$ ) is het verwachte niveau E1 ( $v = 1$ )
- Medio leerjaar 2 ( $p = 5$ ) is het verwachte niveau M2 ( $v = 2$ )
- Eind leerjaar 2 ( $p = 6$ ) is het verwachte niveau E2 ( $v = 3$ )

#### Regel:

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

- zes of meer subdomeinen (van de acht) een functioneringsniveau boven verwacht niveau hebben, dan krijgt het kind het signaal “uitstekend”.
- vier of meer subdomeinen (van de acht) 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 negen 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 16 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
Oriëntatie op boek en verhaal	23	18	41
Oriëntatie op geschreven taal	16	25	41
Spreken	20	23	43
Gesprekken	21	37	58
	80	103	183

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 de subdomeinen voor Beginnende geletterdheid is de interbeoordelaarsbetrouwbaarheid en -overeenstemming uitstekend (zie Tabel 8.2). Bij Mondelinge taalvaardigheid zijn de betrouwbaarheid en overeenstemming bij Spreken wederom uitstekend en bij Gesprekken redelijk. Bij Gesprekken waren de correlatie en de absolute G-coëfficiënt weliswaar redelijk, maar duidelijk minder goed dan bij de andere activiteiten.

Een nadere analyse waarbij telkens een geobserveerde groep eruit is gelaten, laat zien dat de oordelen van één groep de correlatie bij gesprekken aanzienlijk verlaagd hebben (Tabel 8.3). Aangezien bij de activiteit Gesprekken kinderen in twee- of drietallen werken en geobserveerd worden, is het denkbaar dat waarnemingen (kinderen) verwisseld zijn.

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
Oriëntatie boek en verhaal	0,87	0,87	12	22	41
Oriëntatie geschreven taal	0,93	0,92	11	18	41
Spreken	0,95	0,95	21	30	41
Gesprekken	0,68	0,68	8	16	57

Tabel 8.3 Correlatie tussen de observaties van de leerkracht en de observaties van een Cito-expert bij Gesprekken

	Correlatie	N
Totaal	0,68	57
Totaal - Groep 1	0,72	46
Totaal - Groep 2	<b>0,77</b>	30
Totaal - Groep 3	0,68	54
Totaal - Groep 4	0,64	54
Totaal - Groep 5	0,60	50
Totaal - Groep 6	0,66	51

Vervolgens is de standaard toegepast op de leerkrachtoordelen en de expertoordelen om het functioneringsniveau te bepalen. In 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 0,70 en 0,80 goed te noemen (zie Tabel 8.6). In de overeenstemmingstabellen 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 8.4 Overeenstemmingstabel voor Beginnende geletterdheid

Oriëntatie op boek en verhaal						Oriëntatie op geschreven taal					
Functioneringsniveau op grond van leerkrachtoordeel						Functioneringsniveau op grond van leerkrachtoordeel					
Cito-expert oordeel	1	2	3	4	5	Cito-expert oordeel	1	2	3	4	5
	(<E1)	(E1)	(M2)	(E2)	(>E2)		(<E1)	(E1)	(M2)	(E2)	(>E2)
1 (<E1)	50,0	50,0	0,0	0,0	0,0	1 (<E1)					
2 (E1)	9,1	81,8	9,1	0,0	0,0	2 (E1)		100,0	0,0	0,0	0,0
3 (M2)	0,0	15,8	73,7	10,5	0,0	3 (M2)		16,7	75,0	8,3	0,0
4 (E2)	0,0	0,0	28,6	71,4	0,0	4 (E2)		0,0	13,3	86,7	0,0
5 (>E2)	0,0	0,0	0,0	0,0	100,0	5 (>E2)		0,0	0,0	37,5	62,5

Opmerking: Bij Oriëntatie op geschreven taal kwam na toepassen van de standaard het niveau <E1 niet voor bij zowel de leerkracht als de Cito-expert.

Tabel 8.5 Overeenstemmingstabel voor Mondelinge taalvaardigheid

Spreken						Gesprekken*					
Functioneringsniveau op grond van leerkrachtoordeel						Functioneringsniveau op grond van leerkrachtoordeel					
Cito-experto oordeel						Cito-experto oordeel					
	1 (<E1)	2 (E1)	3 (M2)	4 (E2)	5 (>E2)		1 (<E1)	2 (E1)	3 (M2)	4 (E2)	5 (>E2)
1 (<E1)	78,6	21,4	0,0	0,0	0,0	1 (<E1)	100,0	0,0	0,0	0,0	0,0
2 (E1)	40,0	40,0	20,0	0,0	0,0	2 (E1)	0,0	81,8	0,0	9,1	9,1
3 (M2)	0,0	14,3	57,1	28,6	0,0	3 (M2)	0,0	12,5	75,0	0,0	12,5
4 (E2)	0,0	0,0	0,0	75,0	25,0	4 (E2)	0,0	0,0	0,0	66,7	33,3
5 (>E2)	0,0	0,0	14,3	0,0	85,7	5 (>E2)	0,0	0,0	33,3	33,3	33,3

\* De afwijkende groep is hierbij niet meegenomen.

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

	Absolute overeenstemming	Gewogen kappa
Oriëntatie boek en verhaal	0,74	0,73
Oriëntatie geschreven taal	0,80	0,81
Spreken	0,71	0,81
Gesprekken*	0,70	0,63

\* De afwijkende groep is hierbij niet meegenomen.

## 8.2 Betrouwbaarheid

Door het onvolledige design kan bij de subdomeinen Luisteren, Woordenschat en Fonologisch bewustzijn 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  $\text{Var}(\tau)$ . 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 toetscore  $\tau(\theta)$  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 toetscore wordt aangegeven met  $\text{Var}(t|\tau(\theta))$ , en door weer gebruik te maken van de distributie van  $\theta$  in de populatie kan ook de gemiddelde variantie van de geobserveerde toetscores berekend worden.

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

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

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

waarin MAcc staat voor 'Accuracy of Measurement'.

Tabel 8.7 en Tabel 8.8 bevatten een beschrijving van de meeteigenschappen van de niveaueversies van de subdomeinen. Voor elk subdomein van Beginnende geletterdheid en Mondelinge taal worden de eigenschappen beschreven in de groep waarvoor de niveaueversie is bedoeld. Bij de vier 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 voldoende tot goed te noemen.

Tabel 8.7 Beschrijvende gegevens van de niveauseries van de subdomeinen van Beginnende geletterdheid in de kindroute

Subdomein	Versie	Groep	Aantal	Gemiddelde	St.dev	Maximum	MAcc
Oriëntatie op geschreven taal	–	Groep 1	11	8,50	4,28	18	<b>0,86</b>
	–	Groep 2	11	13,40	3,84	18	<b>0,87</b>
Oriëntatie op boek en verhaal	–	Groep 1	12	12,20	5,12	22	<b>0,88</b>
	–	Groep 2	12	16,40	4,38	22	<b>0,87</b>
Fonologisch bewustzijn	★	Groep 1	15	9,80	3,36	15	<b>0,76</b>
	★★	Groep 2	15	12,40	3,06	15	<b>0,83</b>
	★★★	Groep 2	15	11,70	3,13	15	<b>0,82</b>
Receptieve letterkennis	★	Groep 1	22	13,20	5,18	22	<b>0,86</b>
	★★	Groep 2	22	17,50	4,34	22	<b>0,86</b>
	★★★	Groep 2	26	20,50	5,10	26	<b>0,88</b>

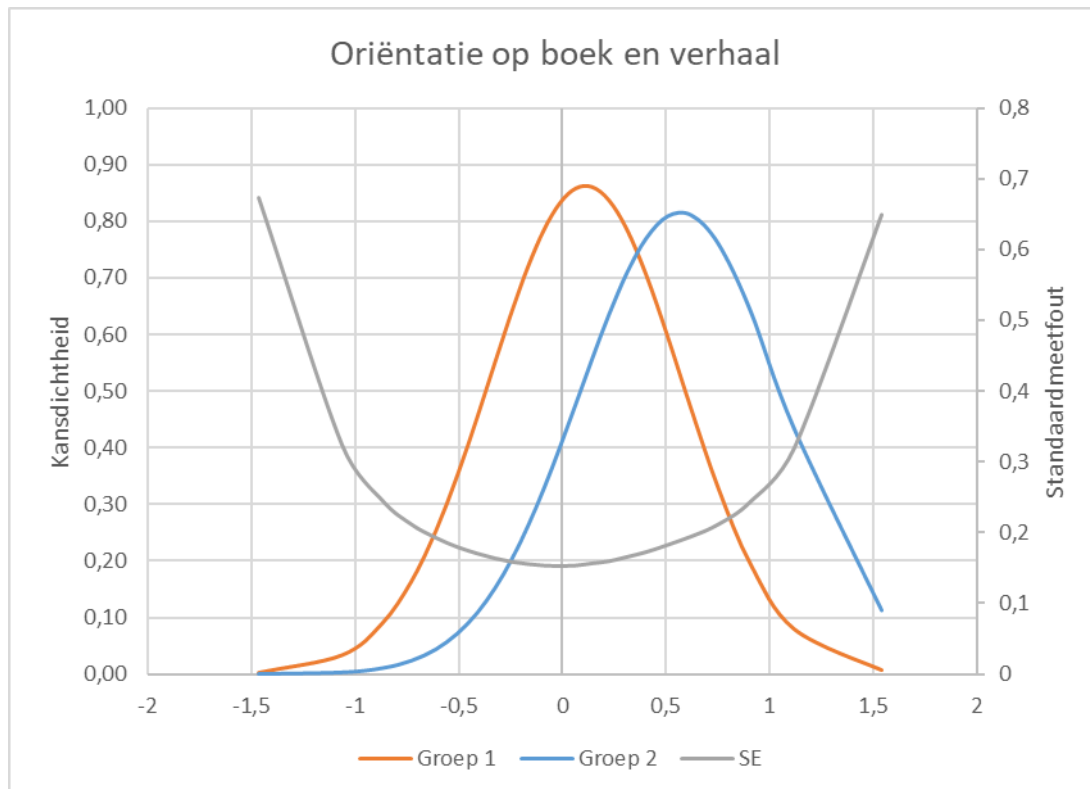
Tabel 8.8 Beschrijvende gegevens van de niveauseries van de subdomeinen van Mondelinge taal in de kindroute

Subdomein	Versie	Groep	Aantal	Gemiddelde	St.dev	Maximum	MAcc
Spreken	–	Groep 1	21	18,10	9,48	30	<b>0,96</b>
	–	Groep 2	21	22,90	6,75	30	<b>0,93</b>
Gesprekken	–	Groep 1	8	10,90	3,75	16	<b>0,84</b>
	–	Groep 2	8	12,50	3,57	16	<b>0,87</b>
Woordenschat	★	Groep 1	16	11,3	2,99	16	<b>0,70</b>
	★★	Groep 2	15	12,10	2,59	15	<b>0,72</b>
	★★★	Groep 2	16	11,9	2,86	16	<b>0,71</b>
Luisteren	★	Groep 1	20	14,4	3,41	20	<b>0,70</b>
	★★	Groep 2	20	14,6	3,36	20	<b>0,71</b>
	★★★	Groep 2	19	13,2	3,29	19	<b>0,71</b>

### 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 Oriëntatie op boek en verhaal. 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 Oriëntatie op boek en verhaal 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 niveaoversie nagegaan hoe vaak het functieniveau dat volgt uit de resultaten gelijk is aan het ware (gesimuleerde) functieniveau op het domein.

Tabel 8.8, Tabel 8.10, Tabel 8.11 en Tabel 8.12 laten voor alle subdomeinen en niveaoversies zien hoe vaak het werkelijke functieniveau overeenkomt met het waargenomen functieniveau. Zo is in Tabel 8.9 te zien dat 86% van de kleuters die volgens de resultaten in functieniveau 1 (<E1) vallen op het subdomein Oriëntatie op boek en verhaal, ook werkelijk in functieniveau 1 zitten. Aan de rechterkant van de tabel zien we dat 73% van de kleuters waarbij het waargenomen functieniveau 4 (E2) is, ook werkelijk het functieniveau E2 hebben voor Oriëntatie op boek en verhaal.



Tabel 8.9 *Betrouwbaarheidstabellen voor de subdomeinen van Beginnende geletterdheid met activiteiten*

Subdomein		Ware functieniveau					
		1 (<E1)	2 (E1)	3 (M2)	4 (E2)	5 (>E2)	
Oriëntatie op boek en verhaal	Waargenomen functieniveau	1 (<E1)	<b>0,86</b>	<b>0,14</b>	0,00	0,00	0,00
		2 (E1)	<b>0,15</b>	<b>0,71</b>	<b>0,14</b>	0,00	0,00
		3 (M2)	0,00	<b>0,16</b>	<b>0,72</b>	<b>0,12</b>	0,00
		4 (E2)	0,00	0,00	<b>0,19</b>	<b>0,73</b>	<b>0,08</b>
		5					
		(>E2)	0,00	0,00	0,00	<b>0,44</b>	<b>0,56</b>
Oriëntatie op geschreven taal	Waargenomen functieniveau	1 (<E1)	<b>0,73</b>	<b>0,27</b>	0,00	0,00	0,00
		2 (E1)	<b>0,10</b>	<b>0,71</b>	<b>0,19</b>	0,00	0,00
		3 (M2)	0,00	<b>0,13</b>	<b>0,68</b>	<b>0,19</b>	0,00
		4 (E2)	0,00	0,00	<b>0,14</b>	<b>0,81</b>	<b>0,05</b>
		5					
		(>E2)	0,00	0,00	0,00	<b>0,48</b>	<b>0,52</b>

Tabel 8.10 *Betrouwbaarheidstabellen voor de subdomeinen en niveaueversies van Beginnende geletterdheid met opdrachten*

Subdomein			1 (<E1)	2 (E1)	3 (M2)	4 (E2)	5 (>E2)	
Fonologisch bewustzijn	★	Waargenomen functieniveau	1 (<E1)	<b>0,98</b>	<b>0,02</b>	0,00	0,00	0,00
			2 (E1)	<b>0,69</b>	<b>0,30</b>	<b>0,01</b>	0,00	0,00
			3 (M2)	0,15	<b>0,60</b>	<b>0,22</b>	<b>0,02</b>	0,00
			4 (E2)	0,00	0,21	<b>0,51</b>	<b>0,24</b>	<b>0,03</b>
			5 (>E2)	0,00	0,01	0,20	<b>0,54</b>	<b>0,25</b>
	★★	Waargenomen functieniveau	1 (<E1)	<b>0,94</b>	<b>0,06</b>	0,00	0,00	0,00
			2 (E1)	<b>0,49</b>	<b>0,43</b>	<b>0,07</b>	0,00	0,00
			3 (M2)	0,07	<b>0,45</b>	<b>0,37</b>	<b>0,11</b>	0,00
			4 (E2)	0,00	0,07	<b>0,31</b>	<b>0,55</b>	<b>0,07</b>
			5 (>E2)	0,00	0,00	0,02	<b>0,44</b>	<b>0,54</b>
	★★★	Waargenomen functieniveau	1 (<E1)	<b>0,91</b>	<b>0,09</b>	0,00	0,00	0,00
			2 (E1)	<b>0,41</b>	<b>0,49</b>	<b>0,09</b>	0,01	0,00
			3 (M2)	0,06	<b>0,42</b>	<b>0,40</b>	<b>0,12</b>	0,00
			4 (E2)	0,00	0,06	<b>0,30</b>	<b>0,52</b>	<b>0,13</b>
			5 (>E2)	0,00	0,00	0,02	<b>0,31</b>	<b>0,67</b>
Receptieve letterkennis	★	Waargenomen functieniveau	1 (<E1)	<b>0,96</b>	<b>0,04</b>	0,00	0,00	0,00
			2 (E1)	<b>0,40</b>	<b>0,57</b>	<b>0,03</b>	0,00	0,00
			3 (M2)	0,01	<b>0,43</b>	<b>0,53</b>	<b>0,03</b>	0,00
			4 (E2)	0,00	0,01	<b>0,45</b>	<b>0,51</b>	<b>0,03</b>
			5 (>E2)	0,00	0,00	0,01	<b>0,66</b>	<b>0,33</b>
	★★	Waargenomen functieniveau	1 (<E1)	<b>0,96</b>	<b>0,04</b>	0,00	0,00	0,00
			2 (E1)	<b>0,36</b>	<b>0,61</b>	<b>0,03</b>	0,00	0,00
			3 (M2)	0,01	<b>0,37</b>	<b>0,58</b>	<b>0,04</b>	0,00
			4 (E2)	0,00	0,01	<b>0,35</b>	<b>0,59</b>	<b>0,05</b>
			5 (>E2)	0,00	0,00	0,01	<b>0,55</b>	<b>0,44</b>
	★★★	Waargenomen functieniveau	1 (<E1)	<b>0,96</b>	<b>0,04</b>	0,00	0,00	0,00
			2 (E1)	<b>0,35</b>	<b>0,62</b>	<b>0,03</b>	0,00	0,00
			3 (M2)	0,01	<b>0,35</b>	<b>0,60</b>	<b>0,04</b>	0,00
			4 (E2)	0,00	0,00	<b>0,33</b>	<b>0,61</b>	<b>0,06</b>
			5 (>E2)	0,00	0,00	0,00	<b>0,49</b>	<b>0,51</b>

Tabel 8.11 Betrouwbaarheidstabellen voor de subdomeinen van Mondelinge taal met activiteiten

Subdomein			1 (<E1)	2 (E1)	3 (M2)	4 (E2)	5 (>E2)
Spreken	Waargenomen functieniveau	1 (<E1)	0,89	0,11	0,00	0,00	0,00
		2 (E1)	0,10	0,78	0,12	0,00	0,00
		3 (M2)	0,00	0,17	0,65	0,18	0,00
		4 (E2)	0,00	0,00	0,12	0,77	0,11
		5 (>E2)	0,00	0,00	0,00	0,24	0,76
Gesprekken	Waargenomen functieniveau	1 (<E1)	0,86	0,14	0,00	0,00	
		2 (E1)	0,10	0,75	0,14	0,00	
		3 (M2)	0,00	0,13	0,77	0,10	
		4 (E2)	0,00	0,00	0,31	0,69	

\* Opmerking: Bij Gesprekken is er geen niveau 5, omdat de laatste grens op het maximum is gezet (zie ook hoofdstuk 7).  
Op dit cluster kan een kleuter de observatiepunten dus niet in meerdere mate dan de vierde grens beheersen.

Tabel 8.12 *Betrouwbaarheidstabellen voor subdomeinen en niveaoversies van Mondelinge taal met opdrachten*

Subdomein			1 (<E1)	2 (E1)	3 (M2)	4 (E2)	5 (>E2)	
Woordenschat	★	Waargenomen functieniveau	1 (<E1)	<b>0,99</b>	<b>0,01</b>	0,00	0,00	0,00
			2 (E1)	<b>0,73</b>	<b>0,24</b>	<b>0,02</b>	0,00	0,00
			3 (M2)	0,24	<b>0,53</b>	<b>0,16</b>	<b>0,05</b>	0,01
			4 (E2)	0,04	0,37	<b>0,31</b>	<b>0,19</b>	<b>0,10</b>
			5 (>E2)	0,00	0,09	0,24	<b>0,32</b>	<b>0,36</b>
	★★	Waargenomen functieniveau	1 (<E1)	<b>0,95</b>	<b>0,05</b>	0,00	0,00	0,00
			2 (E1)	<b>0,48</b>	<b>0,44</b>	<b>0,06</b>	0,02	0,00
			3 (M2)	0,09	<b>0,53</b>	<b>0,24</b>	<b>0,11</b>	0,04
			4 (E2)	0,01	0,22	<b>0,30</b>	<b>0,28</b>	<b>0,19</b>
			5 (>E2)	0,00	0,03	0,12	<b>0,28</b>	<b>0,57</b>
	★★★	Waargenomen functieniveau	1 (<E1)	<b>0,88</b>	<b>0,12</b>	0,00	0,00	0,00
			2 (E1)	<b>0,35</b>	<b>0,51</b>	<b>0,12</b>	0,02	0,00
			3 (M2)	0,05	<b>0,43</b>	<b>0,34</b>	<b>0,15</b>	0,03
			4 (E2)	0,00	0,14	<b>0,34</b>	<b>0,36</b>	<b>0,16</b>
			5 (>E2)	0,00	0,01	0,10	<b>0,35</b>	<b>0,54</b>
Luisteren	★	Waargenomen functieniveau	1 (<E1)	<b>0,98</b>	<b>0,02</b>	0,00	0,00	0,00
			2 (E1)	<b>0,73</b>	<b>0,25</b>	<b>0,02</b>	0,00	0,00
			3 (M2)	0,29	<b>0,55</b>	<b>0,12</b>	<b>0,03</b>	0,00
			4 (E2)	0,06	0,44	<b>0,29</b>	<b>0,18</b>	<b>0,04</b>
			5 (>E2)	0,00	0,12	0,25	<b>0,40</b>	<b>0,23</b>
	★★	Waargenomen functieniveau	1 (<E1)	<b>0,92</b>	<b>0,07</b>	0,00	0,00	0,00
			2 (E1)	<b>0,47</b>	<b>0,45</b>	<b>0,07</b>	0,01	0,00
			3 (M2)	0,10	<b>0,53</b>	<b>0,26</b>	<b>0,11</b>	0,01
			4 (E2)	0,01	0,22	<b>0,33</b>	<b>0,35</b>	<b>0,09</b>
			5 (>E2)	0,00	0,02	0,12	<b>0,44</b>	<b>0,42</b>
	★★★	Waargenomen functieniveau	1 (<E1)	<b>0,92</b>	<b>0,07</b>	0,00	0,00	0,00
			2 (E1)	<b>0,47</b>	<b>0,43</b>	<b>0,09</b>	0,01	0,00
			3 (M2)	0,10	<b>0,47</b>	<b>0,32</b>	<b>0,10</b>	0,02
			4 (E2)	0,01	0,17	<b>0,36</b>	<b>0,32</b>	<b>0,13</b>
			5 (>E2)	0,00	0,01	0,12	<b>0,34</b>	<b>0,53</b>

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 8.13 zien we dat alle subdomeinen met activiteiten boven dit ambitieniveau liggen. Bij de niveaoversie voor groep 1 (★) voor Fonologisch bewustzijn en de niveaoversies voor Luisteren en Woordenschat ligt de plus/minus 1 niveau-index iets onder dit ambitieniveau.

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 niveauversies voor groep 1 (★) een 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 vooral in positieve zin zijn: als sprake is van een misclassificatie vallen kleuters vaker in een hoger functieniveau dan hun werkelijke functieniveau en niet in een lager functieniveau.

Tabel 8.13 Samenvattende indices voor de nauwkeurigheid van de subdomeinen en niveauversies

Subdomein	Niveau- versie	+/- index	Foutpercentage		Aantal schaal- punten	Aantal items
			Vals negatief (P-)	Valse positief (P+)		
<b>Oriëntatie boek en verhaal</b>	–	1,00	0,10	0,17	22	12
<b>Oriëntatie geschreven taal</b>	–	1,00	0,11	0,16	18	11
<b>Fonologisch bewustzijn</b>	★	0,91	0,02	0,57	15	15
	★★	0,96	0,04	0,42	15	15
	★★★	0,97	0,06	0,34	15	15
<b>(Receptieve) letterkennis</b>	★	0,99	0,03	0,38	22	22
	★★	0,99	0,03	0,40	22	22
	★★★	1,00	0,04	0,37	26	26
<b>Spreken</b>	–	1,00	0,10	0,13	30	21
<b>Gesprekken</b>	–	1,00	0,09	0,15	16	8
<b>Woordenschat</b>	★	0,91	0,04	0,49	16	16
	★★	0,87	0,10	0,47	15	15
	★★★	0,91	0,12	0,41	16	16
<b>Luisteren</b>	★	0,83	0,03	0,60	20	20
	★★	0,88	0,08	0,50	20	20
	★★★	0,90	0,10	0,47	19	19



## 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 - Taal heeft geen voorspellende pretentie, daarom is de criteriumvaliditeit niet van belang. Bij een meetinstrument zoals Kleuter in beeld - Taal, 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 - Taal is uitgebreid beschreven in hoofdstuk 2 en 3. Bij Kleuter in beeld - Taal hebben vier informatiebronnen de ontwikkeling gestuurd: de input van onderwijsprofessionals, wetenschappelijke kennis over mondeling taal en beginnende geletterdheid, de expertise van toetsdeskundigen en (psychometrische) informatie verkregen uit empirisch onderzoek. Zoals in hoofdstuk 3 beschreven, zijn eerst met behulp van focusgroepen, expertbijeenkomsten en een online community, de wensen vanuit het veld geïnventariseerd en de globale opzet van het instrument bepaald.

De inhoudskaart Taal (Nederlands) jonge kind (fase 1) van SLO (2018) heeft een belangrijke rol gespeeld bij het bepalen van de domeinen, 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 uiteindelijke te meten leerdoelen geselecteerd en geplaatst onder de subdomeinen. Dit proces heeft geleid tot een toetsmatrijs bestaande uit twee domeinen, acht subdomeinen en 16 onderdelen (zie hoofdstuk 3 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 uitgeprobeerd 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 Domeinen, subdomeinen en onderdelen van *Kleuter in beeld - Taal*

Domein	Subdomein	Onderdelen
<b>Mondelinge taal</b>	Spreken	- Verstaanbaarheid - Productieve woordenschat - Grammaticale beheersing - Verhalen vertellen
	Gesprekken	- Vasthouden van gespreksonderwerp - Eenvoudige gespreksregels
	Woordenschat	- Receptieve woordenschat
	Luisteren	- Begrijpend luisteren
<b>Beginnende geletterdheid</b>	Oriëntatie op boek en verhaal	- Boekoriëntatie - Verhaalbegrip
	Oriëntatie op geschreven taal	- Voorbereidend lezen - Voorbereidend schrijven
	Fonologisch bewustzijn	- Rijm - Beginfoneem - Synthese
	Letterkennis	- Receptieve letterkennis

## 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, itembias, soortgenoten validiteit en gegevens over verschillen tussen relevante groepen.

### Structuur

Het algemene uitgangspunt van het meetmodel is dat elk van de subdomeinen van taalvaardigheid 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. Met uitzondering van het subdomein Gesprekken geldt voor de subdomeinen dat *R1c* niet groter is dan ongeveer anderhalf maal het aantal vrijheidsgraden. De subdomeinen Spreken en Receptieve letterkennis 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
Woordenschat	0,89
Luisteren	0,89
Spreken – verstaanbaarheid	0,85
Spreken – productieve woordenschat	0,88
Spreken – grammatica	0,89
Spreken – verhalen vertellen	0,88
Gesprekken – vasthouden onderwerp	0,88
Gesprekken – gespreksregels	0,86
Fonologisch bewustzijn – rijm	0,88
Fonologisch bewustzijn – beginfoneem	0,89
Fonologisch bewustzijn – synthese	0,88
Receptieve letterkennis	0,86
Oriëntatie op boek en verhaal – boekoriëntatie	0,89
Oriëntatie op boek en verhaal – verhaalbegrip	0,89
Oriëntatie op geschreven taal – lezen	0,86
Oriëntatie op geschreven taal – schrijven	0,86

### 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 de ranges 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
BV	–	Groep 1	12	22	0,27	0,57	0,76	0,54	0,65	0,74
		Groep 2	12	22	0,47	0,76	0,91	0,49	0,63	0,75
GT	–	Groep 1	11	18	0,10	0,48	0,78	0,47	0,62	0,76
		Groep 2	11	18	0,39	0,76	0,96	0,45	0,62	0,76
FB	★	Groep 1	15	15	0,50	0,65	0,78	0,36	0,48	0,60
	★★	Groep 2	15	15	0,72	0,83	0,90	0,44	0,55	0,63
	★★★	Groep 2	15	15	0,48	0,78	0,91	0,35	0,54	0,61
LK	★	Groep 1	22	22	0,33	0,60	0,83	0,24	0,50	0,59
	★★	Groep 2	22	22	0,50	0,79	0,92	0,27	0,51	0,63
	★★★	Groep 2	26	26	0,50	0,79	0,92	0,25	0,50	0,62
SP	–	Groep 1	21	30	0,33	0,58	0,86	0,61	0,73	0,86
		Groep 2	21	30	0,48	0,74	0,95	0,43	0,63	0,78
GS	–	Groep 1	8	16	0,53	0,68	0,89	0,41	0,67	0,77
		Groep 2	8	16	0,66	0,78	0,93	0,43	0,69	0,81
WS	★	Groep 1	16	16	0,42	0,70	0,87	0,23	0,43	0,51
	★★	Groep 2	15	15	0,63	0,81	0,95	0,36	0,46	0,55
	★★★	Groep 2	16	16	0,47	0,75	0,91	0,31	0,43	0,54
LU	★	Groep 1	20	20	0,57	0,72	0,88	0,31	0,39	0,52
	★★	Groep 2	20	20	0,48	0,73	0,87	0,30	0,39	0,48
	★★★	Groep 2	19	19	0,34	0,69	0,88	0,28	0,40	0,48

Voor de subdomeinen met opdrachten zien we dat de p-waarden liggen tussen 0,33 (LK ★) en 0,95 (WS ★★). Omdat met alle niveaueversies 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 makkelijker opdrachten. De gemiddelde moeilijkheid past goed bij de groepen (tussen de ,60 en ,83).

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 groter voor de observatiepunten en de gemiddelde p-waarden lager. 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 niveaueversies. Bij de grote meerderheid van alle itemparameters (86%) is de nauwkeurigheid ook goed (een  $c \leq ,2$ ).

### Itembias

Er is onderzoek uitgevoerd naar differentieel itemfunctioneren (*Differential Item Functioning*, DIF) met betrekking tot sekse. Voor alle opdrachten en observatiepunten zijn de itemresponsen van zowel jongens als meisjes berekend. Vervolgens is hier een S-statistiek voor berekend met het OPLM programma (Verhelst, Glas en Verstralen, 1995, p. 17), analoog aan hoe dit gebeurt tijdens de kalibratie (zie hoofdstuk 5). Bij de

subdomeinen met opdrachten (Fonologisch bewustzijn, Luisteren, Receptieve letterkennis en Woordenschat) werden in totaal vier items met een significante *S*-statistiek gevonden (bij  $\alpha = .01$ ). Er is dus nauwelijks sprake van itembias met betrekking tot sekse bij de opdrachten. Het aantal van vier significante *S*-statistieken (op in totaal 227 toetsingen) ligt immers dicht in de buurt van de twee à drie fouten van de eerste soort die men bij het gekozen significantieniveau mag verwachten.

Bij de observaties van Gesprekken functioneren de meeste observatiepunten verschillend voor jongens en meisjes. Dit komt overeen met eerdere bevindingen uit onderzoeken. In een onderzoek naar de spreekvaardigheid in het basisonderwijs werden diverse verschillen tussen jongens en meisjes gevonden (Krämer, Kuhlemeier, Knoop, Hemker, & van Weerden, 2010). Het onderzoek laat bijvoorbeeld zien dat meisjes hoger scoren op het aspect beleefdheidsvormen. Bij Kleuter in beeld - Taal scoorden meisjes significant hoger op het observatiepunt: 'Laat de leerkracht en andere kinderen uitpraten in een groepsgesprek'. Alle acht observatiepunten zijn polytoom (Nee, Met hulp, Ja) en in totaal zijn er van de 16 *S*-toetsen zeven significant. In vijf van de zeven gevallen heeft de tweede dichotomisering ("Ja" tegenover "Nee" of "Met hulp") een significante *S*-test. Bij de overige activiteiten is nauwelijks of helemaal geen sprake van DIF. Bij Spreken en Oriëntatie op boek en verhaal zijn geen van de *S*-statistieken significant en bij Oriëntatie op geschreven taal slechts twee.

Door bestaande verschillen tussen scholen is een onderzoek naar differentieel itemfunctioneren (*Differential Item Functioning*, DIF) met betrekking tot de wijze van afname niet mogelijk. 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 (Woordenschat, Luisteren en Letterkennis) 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 Letterkennis hebben een gemiddelde vaardigheid in Letterkennis van 0,02 zoals gemeten in de kindroute. Kleuters die volgens de leerkrachtroute in het functioneringsniveau E1 vallen bij Letterkennis hebben een hogere gemiddelde vaardigheid in Letterkennis, namelijk 0,17. Alleen bij Oriëntatie op boek en verhaal is er tussen functioneringsniveau E2 en boven E2 geen significante toename in vaardigheid in de kindroute.

Tabel 9.4 De gemiddelde vaardigheid per subdomein van Beginnende geletterdheid binnen elk functioneringsniveau van de leerkrachtroute

Functioneringsniveau leerkrachtroute	Vaardigheid uit kindroute				
	Gemiddelde	N	Standaard- deviatie	F	Sig.
<b>Oriëntatie op boek en verhaal</b>					
0 (<E1)	-0,28	32	0,38	45,02	0,000
1 (E1)	0,11	39	0,31		
2 (M2)	0,39	67	0,37		
3 (E2)	0,80	37	0,37		
4 (>E2)	0,76	5	0,36		
<b>Totaal</b>	<b>0,31</b>	<b>180</b>	<b>0,51</b>		
<b>Oriëntatie op geschreven taal</b>					
0 (<E1)	-0,34	33	0,46	60,32	0,000
1 (E1)	0,21	44	0,39		
2 (M2)	0,53	60	0,37		
3 (E2)	1,04	33	0,32		
4 (>E2)	1,66	1	.		
<b>Totaal</b>	<b>0,39</b>	<b>171</b>	<b>0,59</b>		
<b>Fonologisch bewustzijn</b>					
0 (<E1)	-0,04	235	0,29	126,34	0,000
1 (E1)	0,15	292	0,32		
2 (M2)	0,42	314	0,35		
3 (E2)	0,59	223	0,39		
4 (>E2)	0,62	21	0,49		
<b>Totaal</b>	<b>0,29</b>	<b>1085</b>	<b>0,41</b>		
<b>Receptieve letterkennis</b>					
0 (<E1)	0,02	50	0,32	54,78	0,000
1 (E1)	0,17	88	0,30		
2 (M2)	0,41	102	0,34		
3 (E2)	0,71	67	0,33		
4 (>E2)	0,88	25	0,38		
<b>Totaal</b>	<b>0,38</b>	<b>332</b>	<b>0,42</b>		

Tabel 9.5 De gemiddelde vaardigheid per subdomein van Mondelinge taal binnen elk functioneringsniveau van de leerkrachtroute

Functioneringsniveau leerkrachtroute	Vaardigheid uit kindroute				
	Gemiddelde	N	Standaard- deviatie	F	Sig.
<b>Spreken</b>					
0 (<E1)	-0,43	10	0,35	26,84	0,000
1 (E1)	0,18	43	0,45		
2 (M2)	0,62	56	0,40		
3 (E2)	0,90	27	0,42		
4 (>E2)	1,40	1	.		
<b>Totaal</b>	0,46	137	0,56		
<b>Gesprekken</b>					
0 (<E1)	-0,04	45	0,56	35,62	0,000
1 (E1)	0,43	119	0,52		
2 (M2)	0,67	165	0,45		
3 (E2)	1,01	53	0,40		
4 (>E2)	1,32	3	0,23		
<b>Totaal</b>	0,56	385	0,56		
<b>Woordenschat</b>					
0 (<E1)	-0,01	96	0,34	108,44	0,000
1 (E1)	0,21	387	0,33		
2 (M2)	0,46	433	0,36		
3 (E2)	0,62	322	0,37		
4 (>E2)	0,76	48	0,39		
<b>Totaal</b>	0,40	1286	0,41		
<b>Luisteren</b>					
0 (<E1)	0,01	68	0,33	42,37	0,000
1 (E1)	0,20	376	0,33		
2 (M2)	0,35	444	0,35		
3 (E2)	0,43	415	0,33		
4 (>E2)	0,50	54	0,32		
<b>Totaal</b>	0,32	1357	0,35		

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. We zien dat de samenhang bij Beginnende geletterdheid hoger is dan bij Mondelinge taal (Tabel 9.6). Daarnaast zien we dat de samenhang sterker is tussen de indirecte observaties en de vaardigheidsschatting gebaseerd op directe observaties dan tussen de indirecte observaties en de vaardigheidsschatting gebaseerd op opdrachten. Waarschijnlijk zijn vaardigheden zoals luisteren en woordenschat lastiger in de dagelijkse lespraktijk te observeren en in te schatten dan bijvoorbeeld spreekvaardigheid.

Om twee redenen is de divergente validiteit niet onderzocht. Ten eerste is er nog geen valide observatie-instrument beschikbaar voor een andere vaardigheid (bijvoorbeeld rekvaardigheid). Belangrijker nog is

dat het kwaliteitsonderzoek met de leerkrachtroute, acht domeinen 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 - Taal.

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

	Spearman's rho	p	N
Oriëntatie op boek en verhaal	0,71	0,000	180
Oriëntatie op geschreven taal	0,77	0,000	171
Fonologisch bewustzijn	0,57	0,000	1085
Receptieve letterkennis	0,65	0,000	332
Spreken	0,64	0,000	137
Gesprekken	0,49	0,000	385
Woordenschat	0,51	0,000	1286
Luisteren	0,31	0,000	1357

### Verschillen tussen relevante subgroepen

In het kwaliteitsonderzoek zijn geboortedatum, geslacht en leerlinggewicht van de leerlingen opgevraagd. Er zijn drie categorieën voor het leerlinggewicht:

- Gewicht 0,0: Ouders/verzorgers die meer dan twee klassen of leerjaren mavo (c- of d-niveau), vmbo gemengde leerweg, vmbo theoretische leerweg, havo of vwo of ouders/verzorgers die een mbo, hbo of universitaire opleiding hebben afgerond
- Gewicht 0,3: Ouders/verzorgers die maximaal lager beroepsonderwijs, voorbereidend beroepsonderwijs, kaderberoepsgerichte leerweg, leerwegondersteunend beroepsonderwijs of vmbo basis hebben genoten
- Gewicht 1,2: Ouders/verzorgers die geen onderwijs of maximaal basisonderwijs hebben genoten. Bij meer dan de helft van de leerlingen is geen gewicht opgegeven (1258). Bij de meeste subdomeinen hebben minder dan 15 kleuters met een gewicht 0,3 of 1,2 meegedaan of zelfs helemaal geen. 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 2019. Vervolgens zijn de kleuters in vier leeftijdscategorieën ingedeeld: (0) ≤ 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 9.7. Alleen bij Spreken vinden we geen significant effect ( $\alpha = 0,01$ ). In Tabel 0.1 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.

Tabel 9.7 Effect van leeftijdscategorie op de vaardigheid van de subdomeinen

Effect leeftijd	F	df1	df2	Sig.	Eta
Oriëntatie op boek en verhaal	7,70	3	173	0,000	0,34
Oriëntatie op geschreven taal	6,55	3	142	0,000	0,35
Fonologisch bewustzijn	87,52	3	1167	0,000	0,43
(Receptieve) letterkennis	19,98	3	372	0,000	0,37
Spreken	2,93	3	127	0,036	0,26
Gesprekken	7,02	3	390	0,000	0,23
Woordenschat	88,35	3	1334	0,000	0,41
Luisteren	41,57	3	1430	0,000	0,28

### Geslacht

We vinden bij de meeste subdomeinen geen significante verschillen tussen jongens en meisjes, zie Tabel 9.8. Alleen bij Luisteren vinden we een significant effect van geslacht, maar het effect is klein.

De Luistervaardigheid van de meisjes is iets hoger dan van de jongens (zie Tabel 0.2).

Tabel 9.8 Effect van geslacht op de vaardigheid van de subdomeinen

Effect geslacht	F	df1	df2	Sig.	Eta
Gesprekken	2,28	1	393	0,132	0,08
Oriëntatie op geschreven taal	1,15	1	131	0,285	0,09
Luisteren	9,49	1	1432	0,002	0,08
Fonologisch bewustzijn	3,43	1	1168	0,064	0,05
(Receptieve) letterkennis	2,21	1	375	0,138	0,08
Woordenschat	2,09	1	1336	0,148	0,04
Oriëntatie op boek en verhaal	0,38	1	200	0,540	0,04
Spreken	0,02	1	124	0,885	0,01

### Leerkrachtroute

#### Leeftijd

Het functioneringsniveau is significant verschillend tussen de leeftijdscategorieën, net zoals het geval was in de kindroute (zie Tabel 9.9). Er is een positieve samenhang tussen functioneringsniveau en leeftijdscategorie. In Appendix A zijn de tabellen te vinden met het percentage kinderen in elke leeftijdscategorie per functioneringsniveau (Tabel 0.3 en Tabel 0.4). Het patroon in de tabel is naar verwachting. Hoe ouder het kind, hoe hoger het functioneringsniveau.

Tabel 9.9 *Chikwadraattoets voor de verschillen in functioneringsniveau tussen leeftijdscategorieën en de samenhang tussen functioneringsniveau en leeftijdscategorie*

Subdomein	Chi-Square Tests	df	Sig	Eta	Gamma
Oriëntatie op boek en verhaal	805,49	12	0,000	0,64	0,67
Oriëntatie op geschreven taal	762,72	12	0,000	0,63	0,66
Fonologisch bewustzijn	748,46	12	0,000	0,65	0,65
(Receptieve) letterkennis	703,59	12	0,000	0,61	0,65
Spreken	622,75	12	0,000	0,56	0,60
Gesprekken	647,05	12	0,000	0,58	0,62
Woordenschat	635,81	12	0,000	0,56	0,60
Luisteren	702,71	12	0,000	0,60	0,64

### **Geslacht**

Bij twee van de acht subdomeinen vinden we een significant verschil tussen jongens en meisjes, alhoewel het effect klein is (zie Tabel 9.10). Als we kijken naar de percentages jongens en meisjes in elk functioneringsniveau (Appendix A) zien we bij Oriëntatie op geschreven taal en bij Gesprekken dat een iets lager percentage meisjes onder E1 valt en een iets hoger percentage in M2 dan bij jongens het geval is, maar de verschillen zijn klein (zie Tabel 0.5 in de Appendix).

Tabel 9.10 *Chikwadraattoets voor de verschillen in functioneringsniveau tussen meisjes en jongens en de samenhang tussen functioneringsniveau en geslacht*

Geslacht	Chi-kwadraat	df	Sig.	Eta	Gamma
Oriëntatie op boek en verhaal	5,26	4	0,262	0,03	0,05
Oriëntatie op geschreven taal	17,71	4	0,001	0,10	0,15
Fonologisch bewustzijn	6,57	4	0,161	0,03	0,04
(Receptieve) letterkennis	5,74	4	0,220	0,05	0,08
Spreken	8,45	4	0,076	0,03	0,06
Gesprekken	13,57	4	0,009	0,06	0,09
Woordenschat	6,00	4	0,199	0,02	0,04
Luisteren	4,74	4	0,315	0,02	0,04



## Appendix A

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

Leeftijdscategorie	(0) ≤ 5 jaar				(1) 5 - 5,5				(2) 5,5 - 6				(3) > 6			
Subdomein	N	Gem	Sig	Std. Dev.	N	Gem	Sig	Std. Dev.	N	Gem	Sig	Std. Dev.	N	Gem	Sig	Std. Dev.
Oriëntatie op boek en verhaal	64	0,16	<sup>12</sup>	0,42	56	0,45	<sup>0</sup>	0,39	44	0,53	<sup>0</sup>	0,50	13	0,42		0,46
Oriëntatie op geschreven taal	44	-0,02	<sup>123</sup>	0,65	42	0,38	<sup>0</sup>	0,57	43	0,50	<sup>0</sup>	0,59	17	0,47	<sup>0</sup>	0,43
Fonologisch bewustzijn	399	0,05	<sup>123</sup>	0,30	337	0,33	<sup>023</sup>	0,41	311	0,48	<sup>01</sup>	0,41	124	0,45	<sup>01</sup>	0,41
(Receptieve) letterkennis	78	0,13	<sup>123</sup>	0,39	114	0,42	<sup>02</sup>	0,41	136	0,56	<sup>01</sup>	0,39	48	0,47	<sup>0</sup>	0,42
Spreken	50	0,32		0,61	32	0,28		0,45	36	0,62		0,61	13	0,60		0,66
Gesprekken	139	0,40	<sup>123</sup>	0,56	113	0,59	<sup>0</sup>	0,56	100	0,61	<sup>0</sup>	0,61	42	0,81	<sup>0</sup>	0,48
Woordenschat	416	0,19	<sup>023</sup>	0,33	395	0,44	<sup>02</sup>	0,38	381	0,61	<sup>013</sup>	0,40	146	0,51	<sup>02</sup>	0,43
Luisteren	403	0,19	<sup>123</sup>	0,30	426	0,31	<sup>023</sup>	0,35	432	0,42	<sup>01</sup>	0,33	173	0,44	<sup>01</sup>	0,33

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

Geslacht	Jongens				Meisjes			
Subdomein	Gem.	Stdev.	N	Gem.	Stdev.	N		
Oriëntatie op boek en verhaal	0,33	0,52	97	0,29	0,46	105		
Oriëntatie op geschreven taal	0,24	0,61	73	0,36	0,66	60		
Fonologisch bewustzijn	0,27	0,42	619	0,32	0,42	551		
(Receptieve) letterkennis	0,38	0,43	201	0,45	0,43	176		
Spreken	0,43	0,56	61	0,41	0,61	65		
Gesprekken	0,51	0,60	200	0,60	0,55	195		
Woordenschat	0,40	0,41	703	0,43	0,42	635		
Luisteren	0,30	0,35	757	0,35	0,34	677		

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

Subdomein	Leeftijdscategorie	0 (<E1)	1 (E1)	2 (M2)	3 (E2)	4 (>E2)	Totaal	Aantal	
<b>Oriëntatie op boek en verhaal</b>	0	≤ 5 jaar	39,80%	46,00%	12,00%	2,20%	100,00%	417	
	1	5 – 5,5 jaar	5,50%	21,70%	45,70%	24,80%	2,30%	100,00%	383
	2	5,5 – 6 jaar	2,30%	11,60%	49,90%	33,50%	2,80%	100,00%	397
	3	> 6 jaar	1,40%	9,30%	36,90%	27,60%	24,80%	100,00%	214
<b>Oriëntatie op geschreven taal</b>	0	≤ 5 jaar	49,00%	41,00%	8,60%	1,40%	100,00%	420	
	1	5 – 5,5 jaar	9,90%	25,00%	39,10%	21,40%	4,70%	100,00%	384
	2	5,5 – 6 jaar	3,00%	16,90%	44,30%	30,00%	5,80%	100,00%	397
	3	> 6 jaar	1,90%	11,60%	36,30%	20,50%	29,80%	100,00%	215
<b>Fonologisch bewustzijn</b>	0	≤ 5 jaar	47,30%	39,60%	9,50%	3,40%	0,20%	100,00%	412
	1	5 – 5,5 jaar	9,10%	18,90%	33,40%	33,20%	5,40%	100,00%	386
	2	5,5 – 6 jaar	3,70%	10,70%	35,40%	43,60%	6,50%	100,00%	401
	3	> 6 jaar	1,80%	6,90%	28,60%	35,50%	27,20%	100,00%	217
<b>(Receptieve) letterkennis</b>	0	≤ 5 jaar	35,30%	53,70%	8,60%	1,90%	0,50%	100,00%	417
	1	5 – 5,5 jaar	4,90%	30,20%	32,60%	24,50%	7,80%	100,00%	384
	2	5,5 – 6 jaar	3,00%	16,90%	37,90%	32,80%	9,30%	100,00%	396
	3	> 6 jaar	0,90%	12,60%	28,80%	23,30%	34,40%	100,00%	215

Tabel 0.4 Percentage kinderen van elke leeftijdscategorie in elk functioneringsniveau van Mondelinge taal volgens de leerkrachtroute

Subdomein	Leeftijdscategorie	0 (<E1)	1 (E1)	2 (M2)	3 (E2)	4 (>E2)	Totaal	Aantal	
<b>Spreeken</b>	0	≤ 5 jaar	28,60%	48,70%	19,60%	2,80%	0,20%	100,00%	433
	1	5 – 5,5 jaar	5,00%	24,40%	43,30%	25,70%	1,50%	100,00%	397
	2	5,5 – 6 jaar	3,60%	13,60%	43,20%	36,20%	3,40%	100,00%	412
	3	> 6 jaar	0,90%	14,60%	34,20%	27,90%	22,40%	100,00%	219
<b>Gesprekken</b>	0	≤ 5 jaar	27,30%	51,30%	18,60%	2,80%		100,00%	429
	1	5 – 5,5 jaar	5,50%	23,90%	41,60%	27,00%	2,00%	100,00%	397
	2	5,5 – 6 jaar	2,90%	11,20%	48,40%	33,60%	3,90%	100,00%	411
	3	> 6 jaar	1,40%	10,50%	35,60%	30,10%	22,40%	100,00%	219
<b>Woordenschat</b>	0	≤ 5 jaar	16,20%	57,00%	17,80%	8,50%	0,50%	100,00%	433
	1	5 – 5,5 jaar	2,80%	21,90%	38,30%	33,80%	3,30%	100,00%	397
	2	5,5 – 6 jaar	2,70%	9,20%	38,80%	42,00%	7,30%	100,00%	412
	3	> 6 jaar	0,50%	7,70%	32,70%	29,50%	29,50%	100,00%	220
<b>Luisteren</b>	0	≤ 5 jaar	13,20%	61,90%	18,70%	6,00%	0,20%	100,00%	433
	1	5 – 5,5 jaar	2,00%	21,00%	36,70%	36,70%	3,50%	100,00%	395
	2	5,5 – 6 jaar	2,20%	6,80%	37,10%	46,80%	7,00%	100,00%	412
	3	> 6 jaar		6,80%	30,10%	35,20%	27,90%	100,00%	219

Tabel 0.5 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>Oriëntatie op boek en verhaal</b>	Jongens	17,70%	24,40%	35,00%	21,20%	1,70%	100,00%	712
	Meisjes	13,50%	25,30%	38,50%	21,10%	1,50%	100,00%	667
	Totaal	15,70%	24,90%	36,70%	21,20%	1,60%	100,00%	1379
<b>Oriëntatie op geschreven taal</b>	Jongens	24,00%	26,80%	29,50%	16,50%	3,20%	100,00%	713
	Meisjes	15,80%	25,80%	33,70%	21,20%	3,60%	100,00%	671
	Totaal	20,00%	26,30%	31,50%	18,80%	3,40%	100,00%	1384
<b>Fonologisch bewustzijn</b>	Jongens	19,70%	22,80%	26,90%	25,60%	4,90%	100,00%	710
	Meisjes	18,10%	21,50%	26,60%	30,60%	3,10%	100,00%	673
	Totaal	18,90%	22,20%	26,80%	28,10%	4,00%	100,00%	1383
<b>(Receptieve) letterkennis</b>	Jongens	15,80%	32,50%	25,90%	19,20%	6,60%	100,00%	710
	Meisjes	12,10%	31,50%	27,50%	22,60%	6,30%	100,00%	669
	Totaal	14,00%	32,10%	26,70%	20,80%	6,50%	100,00%	1379
<b>Spreken</b>	Jongens	13,40%	29,00%	33,90%	21,10%	2,60%	100,00%	738
	Meisjes	10,10%	27,90%	37,90%	22,80%	1,30%	100,00%	692
	Totaal	11,80%	28,50%	35,80%	22,00%	2,00%	100,00%	1430
<b>Gesprekken</b>	Jongens	14,00%	28,30%	34,50%	20,50%	2,70%	100,00%	736
	Meisjes	9,00%	27,30%	39,10%	23,10%	1,50%	100,00%	688
	Totaal	11,60%	27,80%	36,70%	21,80%	2,10%	100,00%	1424
<b>Woordenschat</b>	Jongens	8,00%	28,70%	30,90%	27,50%	5,00%	100,00%	739
	Meisjes	5,90%	27,20%	33,80%	29,60%	3,50%	100,00%	692
	Totaal	7,00%	28,00%	32,30%	28,50%	4,30%	100,00%	1431
<b>Luisteren</b>	Jongens	6,10%	29,90%	29,40%	30,30%	4,30%	100,00%	737
	Meisjes	4,80%	27,10%	33,40%	31,30%	3,50%	100,00%	691
	Totaal	5,50%	28,50%	31,40%	30,70%	3,90%	100,00%	1428

## 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 - Taal. De handleiding bevat bijvoorbeeld een stappenplan voor het gebruik van Kleuter in beeld - Taal. 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 - Taal 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 acht 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 vijf 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 autootje). Hiermee kan de leerkracht de kinderen naar de juiste bladwijze verwijzen. Per opdracht staat de voorleeszin vermeld die de leerkracht voorleest.



**Wat rijmt op pet? net, pot, pen**

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 het hokje onder het juiste plaatje aan te kruisen of in te kleuren. Bij Luisteren zijn er ook enkele opdrachten waarbij het kind een of twee lijnen 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 is dit uitgeprobeerd, maar het bleek te veel nadelen te hebben. Bij het voorlezen van het prentenboek worden bijvoorbeeld inhoudelijke vragen gesteld, zoals over de omslag van het prentenboek. 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. De activiteit Gesprekken is hierop een uitzondering. Deze activiteit vraagt juist om de afname in een klein groepje. Op deze manier kunnen dan observatiepunten als 'Kan korte tijd bij het gespreksonderwerp blijven in een kleine groep' en 'Laat de leerkracht en andere kinderen uitpraten in een groepsgeprek' goed in beeld gebracht worden. Daarom is er bij Gesprekken voor gekozen de activiteit in een groepje uit te voeren en bij de overige activiteiten de activiteit een-op-een uit te voeren.

Alle vier de activiteiten hebben een speels karakter die aansluiten bij de belevingswereld van kleuters. Bij Spreken vertelt het kind iets bij een stripje van Anne en Mo. Bij Gesprekken gaan twee of drie kinderen in gesprek over een voorwerp. Bij Oriëntatie op boek en verhaal leest de leerkracht een prentenboek voor. Bij Oriëntatie op geschreven taal gaat het om speelse activiteiten gericht op letters. Bij elk van de vier

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 de strips, een prentenboek en kaartjes met pictogrammen. Deze zitten handig opgeborgen in de leerkrachtmap. 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 en ja. Ook staan daar indien van toepassing een toelichting en de scoringsregels vermeld, zodat duidelijk is welke antwoordmogelijkheid de leerkracht moet kiezen. Zie hieronder voor twee voorbeelden.

Grammaticale beheersing   Werkwoorden vervoegen		nee	ja
SP14	Vervoegt klankvaste (zwakke) werkwoorden die veel voorkomen in het dagelijks taalgebruik goed. • Toelichting: Bijvoorbeeld 'Anne pakt een appel en stopt 'm in de tas.'		

Voorbeeld observatiepunt uit Spreken

Vorbereidend lezen		nee	beetje	ja
GT01	Kan alle pictogrammen benoemen. Het goede antwoord staat op de achterzijde van elk kaartje. nee: 1-2 goed, beetje: 3-4 goed, ja: 5-6 goed			

Voorbeeld observatiepunt uit Oriëntatie op geschreven taal

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**

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 Privacybijsluiter PO-VSO en de Beveiligingsbijlage. Deze gelden voor de uitgevoerde onderzoeken en het uiteindelijke product Kleuter in beeld - Taal.

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 Taal 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 Taal), 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 acht subdomeinen van Taal 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 is bijvoorbeeld bij Fonologisch bewustzijn snel te zien of een kind bijvoorbeeld meer moeite heeft met een bepaald onderdeel (bijvoorbeeld synthese). Bij Letterkennis kun je zien welke letters het kind al wel en niet kent of dat het kind met een bepaalde categorie, zoals de lange klanken, meer moeite heeft. Dit is belangrijke informatie om het onderwijsaanbod beter op het kind te kunnen afstemmen.







### Voorbeeld Kindrapportage

Er zijn drie signalen voor het totaal van Taal als alle acht 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 8 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 envelopje 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.

Taal	GROEPSRAPPORTAGE	GROEPSVOORTGANG	PDF	BEGIN 2019-2020					
<p>Resultaat</p> <ul style="list-style-type: none"> <li>0% 0 van 14</li> <li>14% 2 van 14</li> <li>0% 0 van 14</li> <li>86% 12 van 14</li> </ul> <p>Bekijk niet afgeronde observaties en opdrachten bij <a href="#">Groepsvoortgang</a>.</p> <p>Sorteer: ↑ Voornaam z-a</p>	Mondelinge taal				Beginnende geletterheid				Legenda
	Woordenschat	Luisteren	Spreeken	Gesprekken	Fonologisch bewustzijn	Letterkennis	Oriëntatie op boek en verhaal	Oriëntatie op geschreven taal	<ul style="list-style-type: none"> <li>Leerkracht-route</li> <li>Kindroute</li> </ul>
	WS	LU	SP	GS	FB	LK	BV	GT	notities
Aun Mohamed 5 jaar 1 maand	<E1	E2	<E1	M2	<E1	<E1	E1	E1	notities
Basera Ay 4 jaar 9 maanden	M2	>E2	<E1	<E1	<E1	<E1	M2	M2	notities
Dania Riedl 5 jaar 1 maand	<E1	M2	E1			E1			notities

Voorbeeld Groepsrapportage

### Groepsoverzicht

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

Sorteer: ↓ Voornaam a-z	Sociaal gedrag	Welbevinden	Speelwerk-houding	Taal	Rekenen	Grove motoriek	Fijne motoriek
Aun Mohamed 5 jaar 1 maand	👁️	✍️	✍️	🌱		✍️	✍️
Basera Ay 4 jaar 9 maanden	🔔	👁️	👁️	🌱		✍️	✍️
Dania Riedl 5 jaar 1 maand						✍️	
Delano Kisters 5 jaar 5 maanden						✍️	
Ibtissa Ganesh 5 jaar 4 maanden						✍️	

**Groepsoverzicht**

Het groepsoverzicht laat in één oogopslag voor alle kinderen in een groep zien welke signalen ze hebben voor alle (sub)domeinen die zijn ingevuld. Voor Taal en Rekenen zijn de signalen net iets anders dan voor Sociaal-Emotioneel en Motoriek.

Voorbeeld Groepsoverzicht

## Literatuur

- Anastasi, A. (1986). Evolving concepts of test validation. *Annual Review of Psychology*, 37, 1-15.
- Boets, B., Smedt, B., Cleuren, L., Vandewalle, E., Wouters, J., & Ghesquiere, P. (2010). Towards a further characterization of phonological and literacy problems in Dutch speaking children with dyslexia. *British Journal of Developmental Psychology*, 28, 5-31.
- 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.
- Carmiggelt, E.C., Uilenburg, N.N., Romeijn, J.E., Stam-van den Doel, H.H., & Pijpers, F.I.M (2013). *Handreiking uniforme signalering van taalachterstanden bij jonge kinderen*. Geraadpleegd van website Nederlands Centrum Jeugdgezondheid
- De Blauw, A., Boland, T., Vernooy, K., & Van het Zandt, R. (2013). *Tussendoelen beginnende geletterdheid – een leerlijn voor groep 1 tot en met 3*. Nijmegen: Expertisecentrum Nederlands.
- 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 (2008b). *Over de drempels met taal. De niveaus voor de taalvaardigheid*. Enschede: SLO.
- Feskens, R., Keuning, J., Van Til, A., & 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
- Gamer, M., Lemon, J., Fellows, I., & Singh, P. (2019). Package 'irr' [Computer software]. Geraadpleegd via <https://cran.r-project.org/web/packages/irr/irr.pdf>
- Gijssel, M., & Van Druenen, M. (2011). *Opbrengstgericht werken aan mondelinge taalvaardigheid*. Nijmegen: Expertisecentrum Nederlands.
- Glas, C.A.W. (1988). The Rasch model and multi-stage testing. *Journal of Educational Statistics*, 13, 45-52.
- 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.
- Goorhuis, S.M., & Schaerlakens, A.M. (2000). *Handboek taalontwikkeling, taalpathologie en taaltherapie bij Nederlandssprekende kinderen*. Utrecht: De Tijdstroom Uitgeverij.
- Guddemi, M., & Case, B.J. (2004). *Assessing young children*. Assessment report, Pearson Education.

- 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.
- Krämer, I., Kuhlemeier, H., Knoop, H., Hemker, B., & Van Weerden, J. (2014). *Balans van de spreekvaardigheid in het basis- en speciaal basisonderwijs 2. Uitkomsten van de peiling in 2010 in groep 8, groep 5 en de eindgroep van het SBO*. Arnhem: Cito.
- Landis, J.R., & Koch, G.G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33, 159-174.
- Lee, W., Hanson, B.A., & Brennan, R.L. (2002). Estimating Consistency and Accuracy Indices for Multiple Classifications. *Applied Psychological Measurement*, 26, 412-432.
- Ministerie van Onderwijs, Cultuur en Wetenschap (2006). *Kerdoelenboekje basisonderwijs*. Geraadpleegd van [www.rijksoverheid.nl](http://www.rijksoverheid.nl)
- Mulder, F., Timman, Y., & Verhallen, S. (2009). *Handreiking bij de Basiswoordenlijst Amsterdamse Kleuters (BAK)*. Amsterdam: ITTA.
- 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, Denmark: Nielsen & Lydiche.
- Schatschneider, C., Fletcher, J.M., Francis, D.J., Carlson, C., & Foorman, B.R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal of Educational Psychology*, 96, 265-282.
- 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.
- Schrooten, W., & Vermeer, A. (1994). *Woorden in het basisonderwijs. 15.000 woorden aangeboden aan leerlingen*. Tilburg: Tilburg University Press.
- Shanahan, T., MacArthur, C.A., Graham, S., & Fitzgerald, J. (2006). *Relations among oral language, reading, and writing development. Handbook of writing research*. New York: Guilford Press.
- SLO (2014). *Praatjes peilen: het spelonderwijs peilen en stimuleren van mondelinge taalontwikkeling bij jonge kinderen*. Geraadpleegd van [www.slo.nl](http://www.slo.nl)
- SLO (2018). *Inhoudskaart Taal (Nederlands) fase 1 (jonge kind)*. Geraadpleegd van [www.slo.nl](http://www.slo.nl)
- Staphorsius, G. (1994). *Leesbaarheid en leesvaardigheid: de ontwikkeling van een domeingericht meetinstrument*. Enschede: Universiteit Twente.
- Stichting Digiwak, UvA, ITTA UvA. *Digiwak*. Geraadpleegd van [www.digiwak.nl](http://www.digiwak.nl)

- Tellings, A.E.J.M. (2015). *Basilex Lexicon (Version 1.0.1)*. Available at the Dutch Language Institute: <http://hdl.handle.net/10032/tm-a2-k7>
- Tinsley, H.E.A., & Weiss, D.J. (1975). Interrater Reliability and Agreement of Subjective Judgments. *Journal of Counseling Psychology*, 22 (4), 358-376.
- Tomesen, M., Weekers, A., Hilte, M., Jolink, A., & Engelen, R. (2016). *Wetenschappelijke verantwoording Begrijpend lezen 3.0 voor groep 5*. Arnhem: Cito.
- Van Boxtel, H., & Hemker, B.T. (2009). *Wetenschappelijke verantwoording van de Intelligentietest Eindtoets Basisonderwijs*. Arnhem: Cito.
- Van der Beek, A., Hoogeveen, M., & Prenger, J. (2015). *Leerstoflijnen mondelinge taalvaardigheid beschreven: Uitwerking van het referentiekader Nederlandse taal voor het onderwijs in mondelinge taalvaardigheid op de basisschool*. Enschede: SLO.
- Van der Leij, A. (2013). Dyslexia and early intervention: what did we learn from the Dutch Dyslexia Programme? *Dyslexia*, 19, 241-255.
- Van der Leij, A. (2016). *Dit is dyslexie*. Houten: LannooCampus.
- Van Druenen, M., Scheltinga, F., Wentink, H., & Verhoeven, L. (2017). *Protocol preventie van leesproblemen – groep 1 en 2*. Nijmegen: Expertisecentrum Nederlands.
- Van Elsäcker, W., Van der Beek, A., Hillen, J., & Peters, S. (2013). *De Taallijn – Interactief taalonderwijs in groep 1 en 2*. Nijmegen: Expertisecentrum Nederlands.
- 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. (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., Glas, C.A.W., & Verstralen, H.H.F.M. (1995). *OPLM: One Parameter Logistic Model. Computer program and manual*. Arnhem: Cito.
- Verhelst, N.D., & Kleintjes, F.G.M. (1993). Toepassingen van itemresponstheorie. In: T.J.H.M. Eggen en P.F. Sanders (red.). *Psychometrie in de praktijk*. (pp. 285-310). 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.
- Vloedgraven, J. (2008). *Development of phonological awareness in relation to literacy: An item response theory perspective*. Proefschrift. Nijmegen: Expertisecentrum Nederlands.
- 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. Available: <https://pdfs.semanticscholar.org/d725/366c7c074d4dff4faa9f8bf115d825b900968.pdf>

Zijlstra, H. (2015). *Early grade learning: The role of teacher-child interaction and tutor-assisted intervention*. Proefschrift. Amsterdam: Universiteit van Amsterdam.



**Cito**

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

Fotografie: Gijs Versteeg

