

1. Uitgangspunten van de toetsconstructie

Bij onderstaande beoordeling van de kwaliteitsaspecten met bijbehorende codes van het voornoemde beoordelingskader worden passages uit de wetenschappelijke verantwoording en de Handleiding veelal letterlijk vermeld. De wetenschappelijke verantwoording heeft betrekking op de uitgangspunten van de toetsconstructie, de normen, de betrouwbaarheid en meetnauwkeurigheid en de validiteit. De Handleiding heeft betrekking op het gebruik van de toets, communicatie over de toetsgegevens en de inhoudsverantwoording.

Algemeen

Het Cito Volgstelsel primair en speciaal onderwijs beoogt de vorderingen van individuele leerlingen, groepen leerlingen en het onderwijs op school van groep 1 tot en met groep 8 te volgen en te evalueren. De toetsen Rekenen-Wiskunde 3.0 groep 5 zijn een onderdeel van het Cito Volgstelsel primair onderwijs en zijn bedoeld voor leerlingen in groep 5 van het primair onderwijs. De toetsen voor groep 5 zijn onderdeel van het Cito LVS Rekenen-Wiskunde 3.0, wat fungeert als een systeem om vast te stellen hoe goed leerlingen kunnen rekenen en hoe hun rekenvaardigheid zich in de basisschoolperiode ontwikkelt. Onderstaande beschrijving is gebaseerd op de Handleiding.

Meetpretentie

De toetsen Rekenen-Wiskunde 3.0 groep 5 meten het vaardigheidsniveau en de ontwikkeling van de leerlingen op het gebied van rekenen-wiskunde. De toetsen meten in hoeverre leerlingen kale rekenopgaven én rekenproblemen in contexten kunnen oplossen.

Doelgroep

De toetsen Rekenen-Wiskunde groep 5 zijn bedoeld voor leerlingen in groep 5 van het primair onderwijs en leerlingen in het speciaal (basis)onderwijs die functioneren op het niveau van groep 5 in het reguliere basisonderwijs. De toetsen zijn ook te gebruiken voor leerlingen in andere leerjaren die een rekenvaardigheid hebben op het niveau van groep 5. Voor leerlingen met een ontwikkelingsachterstand en/of extra onderwijs- behoeften zijn extra aanwijzingen opgenomen in de handleiding.

Gebruiksdoel en functie

Doel van de toetsen Rekenen-Wiskunde 3.0 voor groep 5 is het in kaart brengen van het vaardigheidsniveau en de ontwikkeling van de leerlingen op het gebied van rekenen-wiskunde. Hiervoor wordt de behaalde vaardigheidsscore normgericht geïnterpreteerd op basis van de vaardigheidsverdeling in een adequate, landelijke, referentiegroep. De vaardigheidsscore wordt uitgedrukt in de symmetrische niveau indeling I t/m V en in de asymmetrische niveau indeling A t/m E. De toetsen maken het mogelijk om:

- De vaardigheid rekenen-wiskunde van zowel individuele leerlingen als groepen leerlingen (groeps- en schoolniveau) te beoordelen via een vergelijking van de behaalde scores met de scores van een landelijke referentiegroep oftewel niveaubepaling.
- De ontwikkeling van de vaardigheid rekenen-wiskunde van zowel individuele leerlingen als groepen leerlingen (groeps- en schoolniveau) door de leerjaren heen te volgen oftewel progressiebepaling.

Inhoudelijke theoretische inkadering:

De inhoud van de toetsen sluit aan bij de kerndoelen Rekenen-Wiskunde van het primair onderwijs zoals die wettelijk zijn vastgesteld. De kerndoelen omvatten de onderwerpen 'wiskundig inzicht en handelen', 'getallen en bewerkingen' en 'meten en meetkunde'. Voor de uitwerking van de kerndoelen tot een domeinbeschrijving is gebruikgemaakt van de inhoud van de referentieniveaus, de tussendoelen van de SLO, de publicaties van het TAL-team en van de leerlijnen zoals die door veelgebruikte methodes zijn uitgewerkt.

De verschillende onderdelen van het domein rekenen-wiskunde vormen een samenhangend geheel van getalbegrip en rekenvaardigheid. Hierin staan inzicht in getallen, maatinzicht, ruimtelijk inzicht en het kunnen uitvoeren van operaties met getallen en het kunnen toepassen van die kennis en inzichten in uiteenlopende situaties centraal. Er worden in de domeinbeschrijving vier domeinen onderscheiden: 'Getallen', 'Verhoudingen', 'Meten en Meetkunde' en 'Verbanden. Deze komen overeen met de referentieniveaus. In de toetsen voor groep 5 zijn alleen opgaven opgenomen voor de domeinen die in die jaargroep op school aan de orde komen. Dit zijn de domeinen 'Getallen' en 'Meten en meetkunde'.

Bij het domein 'Getallen' staan het doorzien van de structuur van de telrij, de structuur van getallen en de relaties tussen getallen centraal.

Bij het domein 'Meten en meetkunde' gaat het om het elementaire begrip van wat meten is, de basiskennis van klok en kalender en het rekenen met munten en biljetten.

Inhoud van het toetspakket

Het toetspakket Rekenen-Wiskunde 3.0 groep 5 bestaat uit de volgende documenten:

- Handleiding, deze bevat informatie over:
 - de afname van de toets (hfdst. 2),
 - nakijken en verwerken van toetsgegevens (hfdst. 3),
 - interpretatie van de toetsresultaten op leerling- en groepsniveau (hfdst 4),
 - algemene aandachtspunten voor het schoolplan (hfdst 5),
 - theoretisch kader en achtergronden van de toets (hfdst 6),
 - communiceren over toetsresultaten met leerling en ouders (hfdst 7),
 - achtergrondinformatie en veelgestelde vragen (hfdst 8) en
 - enkele bijlagen
- Vier toetsen:
 - Toets E4M5 (makkelijke variant van de toets M5)
 - Toets M5 (Medio groep 5)
 - Toets M5E5 (makkelijke variant van de toets E5)
 - Toets E5 (Eind groep 5)
- Afnamekaarten met aanwijzingen voor de papieren afname van de toetsen
- Nakijkaarten
- Antwoordbladen
- Tabellen voor de vier toetsen voor het bepalen van de vaardigheidsscore en -niveau.
- Wetenschappelijke verantwoording

2. Beoordeling van de kwaliteitsaspecten

De beoordeling vindt plaats volgens het 'Beoordelingskader voor de psychometrische aspecten van (reeksen van) toetsen uit leerlingvolgsystemen (LOVS)', zoals opgesteld door de Expertgroep Toetsen PO. De Expertgroep Toetsen PO wordt gevormd door Prof.

Dr. Cees Van der Vleuten (voorzitter), Prof. Dr. Cees Glas (psychometrisch expert), Dr. Desiree Joosten-Ten Brinke (onderwijskundig expert) en Paulyne K. Berding-Oldersma MSc (secretaris).

De kwaliteit van de steekproef

S1.1. Is de steekproef representatief?

Bevindingen:

Na proeftoetsing in januari en juni 2010 zijn voor M5 200 items en voor E5 240 items voorgelegd aan 100 tot 150 leerlingen. Naar aanleiding hiervan zijn items geselecteerd voor een normeringsonderzoek in 2013 (januari en juni). Verder heeft er een onderzoek plaatsgevonden met betrekking tot de relatie tussen de digitale en papieren versies van de toetsen. Dit onderzoek heeft in 2014 plaatsgevonden.

Het in 2013 uitgevoerde onderzoek (het normeringsonderzoek 2013) had verschillende doelstellingen: kalibratie van de items op een schaal, koppeling van die schaal aan de LVS-II schaal, koppeling aan voorgaande en vervolgoetsen en gegevens voor de samenstelling en normering van de nieuwe toetsen verzamelen.

Op grond van de proeftoetsing M5 is voor het normeringsonderzoek M5 een selectie gemaakt van 242 items, verdeeld over 9 boekjes met elk 87, 88 of 89 opgaven verdeeld over 3 taken. Deze zijn opgenomen in een embedded field normeringsonderzoek waarin nieuw ontwikkelde items voor LVS 3.0 meeliepen in de al bestaande en op scholen toegepaste LVS 2.0 toetscyclus. Het embedded field normeringsonderzoek M5 is toegepast op de resultaten van 1.975 leerlingen uit groep 5 van 219 scholen. Voor het bepalen van de normering zijn de gegevens aangevuld met gegevens van 2.148 leerlingen uit groep 5 uit Cito dataretour.

Op grond van de proeftoetsing E5 is voor het normeringsonderzoek E5 een selectie gemaakt van 266 items, verdeeld over 9 boekjes, met elk 110, 112 of 113 opgaven verdeeld over 4 taken. Deze zijn opgenomen in een embedded field normeringsonderzoek waarin nieuw ontwikkelde items voor LVS 3.0 meeliepen in de al bestaande en op scholen toegepaste LVS 2.0 toetscyclus. Het embedded field normeringsonderzoek E5 is toegepast op de resultaten van 1.731 leerlingen uit groep 5 van 264 scholen. Voor het bepalen van de normering zijn de gegevens aangevuld met gegevens van 1.901 leerlingen uit groep 5 uit Cito dataretour.

De representativiteit van de steekproeven voor de normeringsonderzoeken M5 en E5 is onderzocht met betrekking tot regio, urbanisatiegraad, schooltype en sekse. Bij regio is uitgegaan van de vier landsdelen / regio's van de CBS-indeling. Bij urbanisatiegraad is uitgegaan van een tweedeling in stad en platteland, afgeleid van de CBS-indeling naar vijf niveaus van verstedelijking. Bij schooltype is uitgegaan van de formatiegewichten volgens OCW. Hierin worden drie niveaus onderscheiden die gebaseerd zijn op het opleidingsniveau van de ouders. Bij sekse is een tweedeling gemaakt naar jongens en meisjes. De steekproefverdeling wijkt weinig af van de populatieverdeling. De effectgroottes phi liggen ver onder de 0.10 en zijn daarmee zeer klein. De effectgrootte phi is het grootst voor de variabele schooltype voor het afnamemoment E5 (.041).

Uit de ruwe scores van de individuele leerlingen uit het embedded field normeringsonderzoek en Cito dataretour werden plausible values gegenereerd op de nieuw

ontwikkelde vaardigheidsschaal. De normering werd vervolgens gebaseerd op de plausible values van de leerlingen in de normeringssteekproef. De plausible values voor de afnamemomenten M5 en E5 bleken een normale verdeling te vormen. De schoolverdeling werd bepaald met het intercept-only multilevel model. Dit model werd geschat via een bootstrap procedure. De intraklassecorrelatie (ICC) lag boven de 0.04, wat inhoudt dat een multilevelanalyse zinvol is. Ondanks dat de percentielen van de normgegevens op schoolniveau dichter bij elkaar kwamen te liggen dan in de leerlingverdeling, waren de afstanden groot genoeg om scholen zinvol te classificeren in de verschillende niveaus.

Het onderzoek naar de relatie tussen de papieren en digitale afnamen in 2014 werd voor M5 werd uitgevoerd in een design met 612 leerlingen, verdeeld over 4 boekjes waarbij ieder boekje steeds zowel papieren als digitale items bevatten. Het analoge onderzoek voor E5 omvatte 851 leerlingen en 5 boekjes. Na analyses werd de conclusie getrokken dat de twee soorten opgaven dezelfde vaardigheid meten, c.q. dat ze op dezelfde latente schaal passen.

Conclusie:

De steekproeven zijn representatief, zijn adequaat gestratificeerd naar sekse, regio, schooltype en urbanisatiegraad en geven informatie over hoe de steekproeven zich verhouden tot de populatiewaarden. De procedure voor het samenstellen van de steekproeven is onderbouwd en de omstandigheden waaronder data is verzameld, is redelijk vergelijkbaar met de omstandigheden waaronder de toets wordt afgenomen. Daarmee wordt aan aspect S1.1. het oordeel '**voldoende**' toegekend.

S1.2. In geval van een onvolledig dataverzamelingsdesign: is het design adequaat?

Bevindingen:

Om te komen tot een set van psychometrisch en inhoudelijk geschikte items zijn de opgaven uit het proefonderzoek en de opgaven uit de daaropvolgende normeringsonderzoeken van januari 2013 (M5) en juni 2013 (E5) gekalibreerd. Hiervoor is gebruik gemaakt van het IRT model OPLM. Met dit statistische model zijn de psychometrische kenmerken (moeilijkheidsparameters en discriminatie indices) van de items geschat.

Voor de kalibratie en normering van M5 is uitgegaan van een onvolledig maar 'verbonden' design. Daartoe zijn van januari 2013 zijn 242 items voorgelegd aan 1.975 leerlingen van groep 5. De 242 items waren verdeeld over 9 boekjes (booklets). Elke boekje bestond uit 87, 88 of 89 opgaven verdeeld over 3 taken. Bij leerlingen zijn digitale taken met nieuwe LVS 3.0 opgaven afgenomen in combinatie met de twee papieren taken van LVS 2.0 en in combinatie met de twee digitale taken van LVS 2.0. De opgaven werden gemiddeld door 339 leerlingen gemaakt (minimaal 202, maximaal 459) wat redelijk in de buurt komt van het door het COTAN voor het Rasch model (en dus voor het OPLM) vereiste minimum van 300.

Voor de kalibratie en normering van E5 van juni 2013 zijn 266 items voorgelegd aan 1.731 leerlingen van groep 5. De 266 items waren verdeeld over 9 boekjes (booklets). Elke boekje bestond uit 110, 112 of 113 opgaven verdeeld over 4 taken. Bij leerlingen zijn digitale taken met nieuwe LVS 3.0 opgaven afgenomen in combinatie met de twee

papieren taken van LVS 2.0 en in combinatie met de twee digitale taken van LVS 2.0. De opgaven werden gemiddeld door 296 leerlingen gemaakt (minimaal 157, maximaal 411) wat ook weer redelijk in de buurt komt van het door het COTAN voor het Rasch model (en dus voor het OPLM) vereiste minimum van 300.

Op basis van inhoudelijke en psychometrische criteria werden 60 items voor elk van de papieren en digitale toetsen M5, E4M5 en M5E5 Rekenen-Wiskunde 3.0 groep 5 geselecteerd en werden 84 items voor de papieren en digitale toets E5 Rekenen-Wiskunde 3.0 groep 5 geselecteerd. De 60 respectievelijk 84 items zijn verdeeld naar vaardigheid en inhoudsaspecten van de elementen getallen en getal-relaties, optellen en aftrekken (kaal en context), vermenigvuldigen en delen en naar meten, tijd en geld als de componenten van de latente vaardigheid Rekenen. Van alle opgaven die zijn meegegaan in het normeringsonderzoek zijn de klassieke p-waarde en de rit-waarde bepaald en eveneens de IRT-indices. Voor de normeringsonderzoeken M5 en E5 werden na het trekken van een representatieve steekproef, waarbij rekening werd gehouden met verdeling naar regio, urbanisatiegraad, schooltype en sekse, zowel scholen geworven als data gehaald uit Cito dataretour. Voor het normeringsonderzoek M5 werd van in totaal 219 scholen gebruik gemaakt van resultaten van 1.975 leerlingen uit groep 5 en van dataretour van 2.148 leerlingen uit groep 5. De verdeling van het aantal scholen over de steekproef en de dataretour worden niet vermeld. Voor het normeringsonderzoek E5 werd van in totaal 264 scholen gebruik gemaakt van resultaten van 1.731 leerlingen uit groep 5 en van dataretour van 1.901 leerlingen uit groep 5. De verdeling van het aantal scholen over de steekproef en de dataretour worden niet vermeld.

Uit het kalibratieonderzoek (S-toetsing, R1c-waarden en de constante 'c') blijkt dat de items passen bij voornoemd IRT model en dat het model ook past voor de toets als geheel. Dit betekent dat er sprake is van één unidimensionele vaardigheidsschaal waar items en leerlingen op afgebeeld kunnen worden.

Conclusie:

Het onvolledige maar 'verbonden' design van de proefonderzoeken is adequaat. Het volledige design van de toets M5 en E5 zijn eveneens adequaat. Aan aspect S1.2 wordt het oordeel '**voldoende**' toegekend.

Normering

N1.2.1. Zijn de normgroepen groot genoeg?

Bevindingen:

De toets is genormeerd voor de afnamemomenten M5 en E5.

Op grond van het kalibratieonderzoek M5 (januari 2013) is een selectie gemaakt van 242 items, verdeeld over 9 boekjes met elk 87, 77 of 89 opgaven verdeeld over 3 taken. Deze zijn opgenomen in een embedded field normeringsonderzoek waarin nieuw ontwikkelde items voor LVS 3.0 meeliepen in de al bestaande en op scholen toegepaste LVS 2.0 toetscyclus. Het embedded field normeringsonderzoek M5 is toegepast op de resultaten van 1.975 leerlingen uit groep 5 van 219 scholen. Voor het bepalen van de normering zijn de gegevens aangevuld met gegevens van 2.148 leerlingen uit groep 5 uit Cito dataretour.

De verdeling van het aantal scholen over de steekproef en de dataretour worden niet vermeld.

Op grond van het kalibratieonderzoek E5 (juni 2013) is een selectie gemaakt van 266 items, verdeeld over 9 boekjes, met elk 110, 112 of 113 opgaven verdeeld over 4 taken. Deze zijn opgenomen in een embedded field normeringsonderzoek waarin nieuw ontwikkelde items voor LVS 3.0 meeliepen in de al bestaande en op scholen toegepaste LVS 2.0 toetscyclus. Het embedded field normeringsonderzoek E5 is toegepast op de resultaten van 1.731 leerlingen uit groep 5 van 264 scholen. Voor het bepalen van de normering zijn de gegevens aangevuld met gegevens van 1.901 leerlingen uit groep 5 uit Cito dataretour. De verdeling van het aantal scholen over de steekproef en de dataretour worden niet vermeld.

Voor de afnamemomenten M5 en E5 werden vaardigheidsverdelingen gepresenteerd op leerlingniveau en op schoolniveau. Dit betreft de gemiddelde score, standaarddeviatie en de percentielen P10, P20, P25, P40, P50, P60, P75 en P80. Van hieruit kunnen de beide niveau indelingen (de symmetrische niveau indeling I t/m V en de asymmetrische niveau indeling A t/m E) worden bepaald.

De normen voor de toetsen Rekenen-Wiskunde 3.0 groep 5 zijn geldig tot en met 2023.

Conclusie:

Er is sprake van relatieve normen, de steekproeven zijn representatief en groot genoeg. Daarmee wordt aan aspect N1.2.1. het oordeel '**voldoende**' toegekend.

N1.2.2. Zijn de normgroepen representatief?

Bevindingen:

De representativiteit van de steekproeven is besproken bij punt S1.1. Hier werd reeds geconstateerd dat deze representatief zijn.

Conclusie:

Aan aspect N1.2.2. wordt het oordeel '**voldoende**' toegekend.

Betrouwbaarheid

B1.1. Zijn of worden de betrouwbaarheidsgegevens correct berekend?

Bevindingen:

Om relevante gegevens bij de toets te genereren, is gebruik gemaakt van het programma OPLAT. Binnen dit programma wordt de coëfficiënt MAcc ('Accuracy of Measurement') berekend. Deze coëfficiënt vertoont qua interpretatie grote overeenkomst met de betrouwbaarheidscoëfficiënt uit de KTT. Deze coëfficiënt wordt in de psychometrische literatuur beschreven en als correct aangemaakt.

Conclusie:

Aan aspect B1.1 wordt het oordeel '**voldoende**' toegekend.

B1.2. Zijn de betrouwbaarheidsgegevens voldoende gezien de beslissingen die met de toets genomen worden?

Bevindingen:

Er wordt verwezen naar de COTAN criteria voor toetsen voor minder belangrijke beslissingen. De interne consistentie betrouwbaarheid is, volgens deze criteria, voldoende bij een betrouwbaarheidscoëfficiënt tussen 0,70 en 0,80. Voor de toets Rekenen-Wiskunde 3.0 groep 5 wordt deze coëfficiënt berekend als MAcc (zie B1.1) voor de afnamemomenten M5, E5, M5E5, E4M5 van zowel de papieren als de digitale toetvariant. Aanvullend hierop wordt de standaardmeetfout vermeld. De afnamecontext van de toets leent zich, dankzij een OPLM kalibratie, voor een gesimuleerd test-hertest onderzoek onder ideale condities. De test-hertest coëfficiënt is identiek aan de MAcc voor alle afnamemomenten.

De resultaten laten zich lezen als standaardmeetfout (papieren versie: E4M5 2,94; M5 3,11; M5E5 3,01; E5 3,69; digitale versie: E4M5 12,94; M5 14,79; M5E5 12,89; E5 15,87), MAcc (papieren versie: E4M5 0,92; M5 0,94; M5E5 0,93; E5 0,95; digitale versie: E4M5 0,93; M5 0,95; M5E5 0,93; E5 0,95) en test-hertest simulatie (papieren versie: E4M5 0,92; M5 0,94; M5E5 0,93; E5 0,95; digitale versie: E4M5 0,93; M5 0,95; M5E5 0,93; E5 0,95) en worden aangemerkt als 'voldoende'. Hierbij wordt opgemerkt dat de digitale toetsen werken met gewogen scores in het computerprogramma LOVS.

Aanvullend hierop is de lokale meetnauwkeurigheid weergegeven in betrouwbaarheidstabellen (misclassificaties). Uitgaande van de betrouwbaarheids tabellen worden twee indices voor de nauwkeurigheid van de classificaties gerapporteerd: de plus/minus 1 niveau-index en de marginal classification index. Uit de hoogte van de indices blijkt dat de laagst en de hoogst scorende leerlingen accuraat te classificeren zijn, maar dat over het algemeen tussen leerlingen in de niveaugroepen B, C en D, respectievelijk II, III en IV, er een licht minder duidelijk onderscheid te maken is. De marginal classification indices lopen uiteen van 67 tot 78 procent en de resultaten stemmen hiermee eveneens tot tevredenheid.

Conclusie:

De betrouwbaarheid van de toetsen Rekenen-Wiskunde 3.0 groep 5 is 'voldoende' als aangenomen mag worden dat de toets geen zware consequenties voor de leerlingen heeft en ingestemd wordt met de beoordelingscriteria voor de betrouwbaarheid van de COTAN.

Op aspect B.1.2. wordt aan de toets Rekenen-Wiskunde groep 5 het oordeel '**voldoende**' toegekend.

Validiteit

V1. Dragen de items in de toets bij aan de validiteit van de toets (hierbij gaat het om aspecten als relevantie, objectiviteit en efficiëntie van de items)

Bevindingen:

Het toetspakket Rekenen-Wiskunde voor groep 5 bevat vier papieren toetsen en vier digitale toetsen. De reguliere toetsen M5 en E5 zijn bedoeld voor de reguliere afnamemomenten (halverwege en aan het einde van het schooljaar). Naast deze reguliere toetsen bevat het pakket ook de tussentoetsen E4M5 en M5E5. De toets E4M5 kan aan het einde van groep 4 worden afgenomen voor leerlingen waarbij de rekenvaardigheid zich sneller ontwikkeld heeft, maar kan ook medio groep 5 worden ingezet voor leerlingen waarbij de rekenvaardigheid zich minder snel ontwikkeld heeft. De tussentoetsen E4M5 en M5E5 worden afgenomen op de reguliere afnamemomenten.

De reguliere toets M5 bestaat uit twee taken van elk 30 opgaven. De reguliere toets E5 bestaat uit drie taken van elk 28 opgaven. De tussentoetsen E4M5 en M5E5 bestaan uit drie taken van elk 20 opgaven. De taken zijn een combinatie van meerkeuzevragen en kort antwoordvragen. De diverse toetsen kunnen worden afgenomen op 2 respectievelijk 3 verschillende dagdelen. De toets kan zowel handmatig als via de computer via het computerprogramma LOVS nagekeken worden.

In de handleiding wordt door toetsmatrijzen inzicht geboden in hoe de toets is samengesteld. In de toets worden twee deelvaardigheden getoetst: (1) getallen en (2) meten en meetkunde. Dit zijn ook twee van de kenmerken van de taakuitvoering die uitgewerkt zijn bij de referentieniveaus voor rekenvaardigheid.

De toetsontwikkelaars onderscheiden, overeenkomstig de referentieniveaus, vier domeinen:

- Getallen
- Verhoudingen
- Meten en meetkunde
- Verbanden.

In de toetsen voor groep 5 komen alleen de domeinen Getallen en Meten en meetkunde voor. Deze onderwerpen komen aan bod in de meest gebruikte methodes voor rekenen-wiskunde van groep 5. Het is een voor de hand liggende keuze om de domeinen verhoudingen en verbanden in groep 5 onderbelicht te laten.

De toetsontwikkelaars onderscheiden binnen het domein Getallen de onderdelen Getalbegrip en Bewerkingen.

Binnen het domein Meten worden de volgende onderdelen onderscheiden: Meten (lengte, oppervlakte, inhoud, gewicht), Meetkunde, Tijd, Geld.

De vaardigheden die getoetst worden zijn gezamenlijk een goede afspiegeling van de rekenvaardigheden waarmee leerlingen in aanraking zijn geweest tijdens het rekenonderwijs dat zij hebben gekregen.

De toetsen bevatten een beperkt aantal meerkeuzeopgaven en korte antwoordvragen. De opgaven van de toetsen bestaan voor een deel uit opgaven met een context en voor een deel uit opgaven zonder context.

De leerlingen lezen zelf de opgaven en maken de toets zelfstandig. Bij de digitale taken kan de leerling ervoor kiezen de opgaven te laten voorlezen door de computer. De inhouden van de papieren en digitale toetsen zijn uitwisselbaar. De toetsen meten dezelfde vaardigheden.

De opgaven zijn duidelijk en ook qua taal voldoende toegankelijk. De opgaven en de vraagstelling zijn helder en eenduidig. Het antwoordmodel laat geen ruimte voor interpretatie.

De rekenvaardigheid van leerlingen in een groep loopt vaak sterk uiteen. De toetsontwikkelaars houden hier op een goede wijze rekening mee doordat ze extra toetsen hebben ontwikkeld voor zowel leerlingen met een vertraagde ontwikkeling/zwakke leerlingen, als voor extreem vaardige leerlingen. Tevens geeft de handleiding duidelijke aanwijzingen voor de leerkracht voor het maken van een goede inschatting van de best passende toets (zie Handleiding, Bijlage 2 Toetsen op maat). Door te 'toetsen op maat' wordt rekening gehouden met het niveau van de leerling en wordt op een zo'n betrouwbaar mogelijke wijze de vaardigheid van een leerling in kaart gebracht.

Voor leerlingen met een beperkte aandachtsspanne/concentratieproblemen, leerlingen die langzamer werken dan gemiddeld, leerlingen met visusproblemen, leerlingen met dyslexie, leerlingen met dyscalculie en leerlingen met behoefte aan voorspelbaarheid doet de handleiding bruikbare aanpassingsvoorstellen in de handleiding.

Merk op dat de beoordeling van de validiteit zich hieronder beperkt tot het statistisch/psychometrisch onderzoek dat is verricht.

De toets Rekenen-Wiskunde 3.0 groep 5 is niet bedoeld voor voorspellend gebruik. Daarmee is de criteriumvaliditeit niet van toepassing. De (psychometrische) begripsvaliditeit wordt uitgewerkt in unidimensionaliteit, itemkwaliteit, itembias, convergente en divergente validiteit en in verschillen tussen relevante subgroepen.

De resultaten van de uitgevoerde kalibratie maken het aannemelijk dat er sprake is van unidimensionaliteit. Dit betekent dat met elke willekeurige subset van items uit de gekalibreerde itembank dezelfde onderliggende rekenvaardigheid kan worden vastgesteld. Dit wordt tevens bevestigd door de nauwkeurigheid van de itemparameterschattingen (op vijf van de 60 items uit de papieren toets M5 en twee items uit M5 digitaal na, is de constante c voor alle items hoger dan 0.20 maar lager dan 0.40).

De gemiddelde moeilijkheidsgraden van de items (als criterium voor de itemkwaliteit) voldoen (papieren versie: p-waarden E4M5: range van .54 - .92 en gemiddelde .76; M5: range .42 - .90 en gemiddelde .68; M5E5: range .51 - .91 en gemiddelde .74; E5: range .42 - .91 en gemiddelde .74; digitale versie: p-waarden E4M5: range van .54 - .92 en gemiddelde .74; M5: range .42 - .91 en gemiddelde .66; M5E5: range .43 - .92 en gemiddelde .72; E5: range .36 - .92 en gemiddelde .67).

De gemiddelde Rit-waarde (ook een criterium voor de itemkwaliteit) voldoet eveneens. Voor alle acht de toetsen is de gemiddelde Rit waarde .40 of hoger en is de ondergrens E5: 0.21 (de Rit-waarde moet idealiter boven de 0.20 liggen). De gemiddelde Rit-waarde bij deze toets E5 is met 0.42 uitstekend. De gemiddelden komen uit in de range van .40 tot .46.

In het onderzoek naar itembias is er met een lichte vorm van Differential Item Functioning (DIF) bij 42 van de 433 items nauwelijks sprake van DIF naar sekse.

De constructvaliditeit is uitgewerkt in convergente validiteit door de samenhang te onderzoeken met de voorgaande versie Rekenen-Wiskunde 2.0. De correlaties waren voor zowel M5 ($r=0,97$) en E5 ($r=0,96$) hoog.

De divergente validiteit is onderzocht door de samenhang met de Schoolvaardigheidstoets rekenen-wiskunde en met de Tempotoets rekenen (beide van Boom testuitgevers) te onderzoeken. Daarnaast is de samenhang met diverse LVS leestoetsen onderzocht. De hoge correlatie met de schoolvaardigheidstoets (M5: $r=0,70$; E5: $r=0,78$) en de lagere correlatie met de diverse LVS leestoetsen (r in de range van $0,32 - 0,62$) sterken de eerdere COTAN waardering 'goed' of 'voldoende' voor de begripsvaliditeit. De hoge correlaties tussen Rekenen-Wiskunde 3.0 en vergelijkbare rekentoetsen en de lagere correlatie tussen Rekenen-Wiskunde 3.0 en diverse taaltoetsen vormen een ondersteuning voor de validiteit van de toets.

Wat betreft de verschillen tussen relevante subgroepen scores jongere leerlingen (de versnelde leerlingen) naar verwachting iets beter dan oudere leerlingen en scores jongens iets hoger dan meisjes. In termen van effectgrootte is er sprake van een klein effect (M5: $0,37$ en E5: $0,36$).

Conclusie:

Op zowel onderwijskundig en psychometrisch gebied wordt de validiteit voldoende onderbouwd. Op aspect V1.1 wordt aan de toetsen Rekenen-Wiskunde 3.0 groep 5 het oordeel '**voldoende**' toegekend.

Het volg-aspect

VA1.1. Is er een voldoende empirische onderbouwing van de schaal waarop de groei van een leerling wordt uitgedrukt? Wordt groei op een adequate manier gemeten?

Bevindingen:

Het algemene (inhoudelijke) uitgangspunt voor de toets Rekenen-Wiskunde 3.0 groep 5 is dat de (latente) vaardigheid Rekenen kan worden opgevat als een unidimensioneel continuüm en dat elke leerling kan worden voorgesteld als een punt op dit continuüm. Hierbij wordt opgemerkt dat de elementen (1) getallen en (2) meten en meetkunde niet opgevat kunnen worden als te isoleren vaardigheden. Daarmee wordt de rekenvaardigheid, als interactie tussen deze componenten, beschouwd als één unidimensionele vaardigheid.

Uit het kalibratieonderzoek blijkt dat de items passen bij het gehanteerde IRT model en dat het model ook past voor de toetsen M5 en E5 als geheel. Dit betekent dat er sprake is van één unidimensionele vaardigheidsschaal waar items en leerlingen op afgebeeld kunnen worden.

Afhankelijk van het aantal items dat een leerling goed maakt, wordt er een vaardigheidsscore toegekend. Jongere leerlingen scoren iets beter dan oudere leerlingen. Tevens scoren jongens iets hoger dan meisjes.

Voor Rekenen-Wiskunde 3.0 is een nieuwe vaardigheidsschaal ontwikkeld, waarop alle uitgebrachte en nog uit te brengen toetsen uit het Cito Volgstelsel primair en speciaal basisonderwijs Rekenen-Wiskunde 3.0 worden gekalibreerd. Vanwege het volgmodel en de gevolgde dataverzamelingsstrategie voor de normering, worden de nieuw ontwikkelde toetsen gefaseerd uitgebracht, d.w.z. in elk schooljaar een toetspakket voor een hogere groep.

Conclusie:

Aan aspect VA1.1. wordt het oordeel '**voldoende**' toegekend.

VA1.2. Worden er gegevens verstrekt over hoe groei geïnterpreteerd dient te worden? Wordt de betrouwbaarheid van de groei op die schaal adequaat weergegeven?

Bevindingen:

In hoofdstuk 7 van de handleiding ('Communiceren over toetsresultaten met leerling en ouders') wordt beschreven hoe er met de verschillende gebruikers over de toetsresultaten kan worden gecommuniceerd. Hierin wordt onderscheid gemaakt tussen 'niveau' en 'groei', wat wordt onderbouwd met diverse rapportage mogelijkheden.

In de wetenschappelijke verantwoording wordt toegelicht hoe de toetsen ingezet kunnen worden om de ontwikkeling van leerlingen te volgen in de tijd, namelijk door het toetsresultaat van een leerling te vergelijken met andere leerlingen en door het toetsresultaat van een leerling te vergelijken met diens andere toetsresultaten. Voor alle vergelijkingen geldt dat uitspraken over de voortgang van leerlingen gerelativeerd moeten worden vanwege de (on)betrouwbaarheid van de toetsen. Door betrokkenen bij de toetsen Rekenen-Wiskunde moet beseft worden dat vaardigheidsgroei zich langzaam in de tijd voltrekt.

Conclusie:

Aan aspect VA1.2. wordt het oordeel '**voldoende**' toegekend.

Inzicht in leervorderingen

I1. Levert de toetsaanbieder een format voor een geschreven toelichting bij de leervorderingen van de leerling die (ook) voor ouders/voogden/verzorgers begrijpelijk is?

Bevindingen:

Via de portal van Cito B.V. kan gebruik worden gemaakt van rapportage-/registratieformulieren voor een leerlingrapport, groepsrapport, groepsoverzicht (overzicht van één groep leerlingen tijdens hun schoolperiode) en een alternatief leerlingrapport (voor leerlingen die op een eigen niveau werken). Voor ouders is met name het leerlingrapport of alternatief leerlingrapport informatief omdat deze rapporten van hun kind individueel de vaardigheid en de groei weergeven.

In de Handleiding wordt in hoofdstuk 7 aandacht besteed aan de wijze waarop met ouders over de toetsresultaten gecommuniceerd kan/moet worden. Met name wordt daarbij gewezen op het leerlingrapport, waarin zowel het niveau van de leerling als de progressie van de leerling numeriek en grafisch gepresenteerd worden.

Daarnaast wordt de leraar gewezen op misverstanden die zich bij de interpretatie van de niveau-indelingen bij de ouders kunnen voordoen. Ook moeten zij aan ouders het verschil tussen methode-onafhankelijke en methodegebonden toetsen duidelijk maken en erop

wijzen dat deze toetsen leerlingen anders (kunnen) beoordelen. De informatie biedt goede handvatten voor de gesprekken met ouders. In hoofdstuk 8 worden veelgestelde vragen behandeld die weliswaar voor de leraren bestemd zijn maar waar de antwoorden voor een deel ook informatief zijn tijdens bijvoorbeeld de tienminutengesprekken. Over de interpretatie van toetsresultaten is ook een folder ouderinformatie beschikbaar die men via de website van het Cito kan downloaden.

Vermeldenswaard is dat de toetsontwikkelaars ook aanwijzingen hebben opgenomen voor scholen omtrent de wijze van communiceren met leerlingen over de toetsresultaten.

Conclusie:

Op aspect I1.1 wordt aan de toetsen Rekenen-Wiskunde 3.0 groep 5 het oordeel '**voldoende**' toegekend.

3. Verzamelstaat

| Kwaliteitsaspect | Code | Oordeel |
|--------------------------------|-------|------------------|
| De kwaliteit van de steekproef | S1.1 | Voldoende |
| | S1.2 | Voldoende |
| Normering | N1.1 | Voldoende |
| | N1.2 | Voldoende |
| Betrouwbaarheid | B1.1 | Voldoende |
| | B1.2 | Voldoende |
| Validiteit | V1.1 | Voldoende |
| Volg-aspect | VA1.1 | Voldoende |
| | VA1.2 | Voldoende |
| Inzicht in leervorderingen | I1.1 | Voldoende |

4. Literatuurlijst

- Hop, M., Janssen, J. & Engelen, R.(2016). *Wetenschappelijke verantwoording Rekenen-Wiskunde 3.0 voor groep 5*. Arnhem: Cito.
- Cito (2015). *Rekenen-Wiskunde 3.0 voor groep 5*. Arnhem: Cito.