

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 Volgsysteem 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 7 zijn een onderdeel van het Cito Volgsysteem primair en speciaal onderwijs en zijn bedoeld voor leerlingen in groep 7 van het primair onderwijs. De toetsen voor groep 7 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 7 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 3.0 groep 7 zijn bedoeld voor leerlingen in groep 7 van het primair onderwijs en leerlingen in het speciaal (basis)onderwijs die functioneren op het niveau van groep 7 in het reguliere basisonderwijs. De toetsen zijn ook te gebruiken voor leerlingen in andere leerjaren die een rekenvaardigheid hebben op het niveau van groep 7. Voor leerlingen met een ontwikkelingsachterstand en/of extra onderwijsbehoeften zijn extra aanwijzingen opgenomen in de handleiding.

Gebruiksdoel en functie

Doel van de toetsen Rekenen-Wiskunde 3.0 voor groep 7 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 7 zijn opgaven opgenomen voor deze vier domeinen.

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 'Verhoudingen' gaat het om het doorzien van de structuur en samenhang van verhoudingen op hoofdlijnen en er in praktische situatie mee te rekenen.

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.

Bij het domein 'Verbanden' staat het omgaan met tabellen, diagrammen en grafieken centraal.

Inhoud van het toetspakket

Het toetspakket Rekenen-Wiskunde 3.0 groep 7 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 groepsniveaus (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
- Twee toetsen:
 - Toets M7 (Medio groep 7)
 - Toets E7 (Eind groep 7)
- Afnamekaarten met aanwijzingen voor de papieren of de digitale afname van de toetsen
- Nakijkaarten
- Antwoordbladen
- Tabellen voor de twee 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 mevrouw Pauly K. Berding-Oldersma MSc (secretaris).

De kwaliteit van de steekproef

S1.1. Is de steekproef representatief?

Bevindingen

De proeftoetsing van de items heeft in 2011 en in 2013 plaatsgevonden. Op basis van de geschatte moeilijkheidsgraad en discriminerend vermogen zijn items geselecteerd voor het kalibratie- en normeringsonderzoek.

Dit onderzoek heeft in 2015 plaatsgevonden en de verschillende items zijn voorgelegd aan respectievelijk 2279 en 1187 leerlingen die representatief waren voor de doelgroep (zie ook de evaluatie van de criteria N1.2.1 en N1.2.2). Gemiddeld werden de items door respectievelijk 301 en 160 leerlingen beantwoord.

We merken wel op dat het gemiddelde aantal antwoorden voor E7 per item aanzienlijk lager was dan het gemiddelde aantal antwoorden per item in M7.

In januari en juni 2016 is bovendien een onderzoek uitgevoerd om de itemparameters voor de gedigitaliseerde variant van de toetsen te schatten en om de digitale items op eenzelfde schaal te brengen als de papieren items. Gemiddeld waren er 232 waarnemingen per item in januari en gemiddeld 185 waarnemingen per item in juni.

De steekproef is geanalyseerd in relatie tot de variabelen regio, urbanisatiegraad, schooltype en sekse en vormde een zeer goede afspiegeling van de populatie.

Geconstateerd kan worden dat de steekproefgroottes voldoen aan het doel en dat de steekproef van het kalibratie- en normeringsonderzoek representatief was.

Conclusie

Op aspect S1.1 wordt de toets Rekenen-Wiskunde 3.0 voor groep 7 als 'voldoende' beoordeeld.

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

Bevindingen

Voor M7 waren 243 items verdeeld over 12 verschillende opgavenboekjes in een onvolledig, maar verbonden design. Elk boekje bestond uit de bestaande M7-toets van LVS II aangevuld met een taak met 32 of 33 nieuwe items. De bestaande LVS II toets telt 96 items. Elk nieuw item kwam in twee boekjes voor. Het gemiddeld aantal leerling antwoorden per item was 301.

Voor E7 waren 346 items verdeeld over 12 verschillende opgavenboekjes in een onvolledig, maar verbonden design. Elk boekje bestond uit de bestaande E7-toets van LVS II aangevuld met een taak met 34 nieuwe items. De bestaande LVS II toets telt 96 items. Elk nieuw item kwam in twee boekjes voor. Het gemiddeld aantal leerling antwoorden per item was 160.

De items zijn gekalibreerd met OPLM en de modelfit is getoetst. Zowel op individueel item als op het niveau van de gehele verzameling items past het model.

Conclusie

Op aspect S1.2 wordt de toets Rekenen-Wiskunde 3.0 groep 7 als 'voldoende' beoordeeld.

Normering

Algemene opmerking vooraf

Hoewel de psychometrische afhandeling van de kalibratie en normering van hoog niveau zijn, is de rapportage hier en daar slordig. Zo springt meteen in het oog dat formule (4.2) op bladzijde 37 verminkt is. Ook boven de formule staan een aantal uitdrukkingen voor kansen en proporties die niet goed zijn doorgekomen. Dat doet wat afbreuk aan de verder goede presentatie.

N1.2.1. Zijn de normgroepen groot genoeg?

Bevindingen

Voor het afnamemoment M7 hebben 2262 leerlingen meegedaan aan het normeringsonderzoek en voor het afname moment E7 1175 leerlingen. Daar zijn nog leerling resultaten uit de dataretour aan toegevoegd. Deze aantallen zijn groot genoeg om schattingen te maken voor de populatie. Opvallend is wel dat deze aantallen licht afwijken van de eerder genoemde aantallen bij de beschrijving van de streekproef. Daar wordt echter geen verklaring voor gegeven. Uit de gegevens blijkt dat de gemiddelde vaardigheid in algemene rekenvaardigheid in de periode tussen de afnamemomenten toeneemt, terwijl de spreiding in de scores licht toeneemt.

Conclusie

Op aspect N1.2.1 wordt de toets Rekenen-Wiskunde 3.0 groep 7 als 'voldoende' beoordeeld.

N1.2.2. Zijn de normgroepen representatief?

Bevindingen

Voor de normeringsonderzoeken M7 en E7 werden scholen geworven na het trekken van een representatieve steekproef, waar bij rekening gehouden werd met verdeling over schooltype, regio en verstedelijking. Achteraf is de verdeling over sekse gecontroleerd. De representativiteit van de steekproef is onderzocht met betrekking tot schooltype, regio, urbanisatiegraad en sekse. Bij de definitie van schooltype is gebruikgemaakt van de formatiegewichten van de leerlingen binnen een school. In navolging van OCW worden op basis van het opleidingsniveau van de ouders drie opleidingsniveaus onderscheiden:

- 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 heeft of hebben een opleiding uit categorie 2 gehad;
- 1.2 = één van de ouders heeft een opleiding gehad uit categorie 1 en de ander een opleiding uit categorie 1 of 2.

In deze indeling wordt verwezen naar de volgende categorieën in het opleidingsniveau van de ouders: categorie 1 is maximaal basisonderwijs, categorie 2 is maximaal lbo/vbo en categorie 3 is voortgezet onderwijs en hoger. Voor elke school wordt het percentage leerlingen met een gewicht afwijkend van 0,0 bepaald. Gebaseerd op deze gewichten zijn vier groepen scholen gevormd. Bij regio is uitgegaan van de CBS-indeling naar landsdeel waarbij de regio's noord, oost, west en zuid onderscheiden worden. Bij urbanisatiegraad is gebruik gemaakt van de twee niveaus: Stad of Land. Tenslotte werd bij de variabele sekse een tweedeling naar jongens en meisjes gehanteerd.

Uit de resultaten van de representativiteitsanalyse bleek dat de steekproeven een goede afspiegeling vormen van de populatie.

Conclusie

Op aspect N1.2.2 wordt de toets Rekenen-Wiskunde 3.0 voor groep 7 als 'voldoende' beoordeeld.

Betrouwbaarheid

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

Bevindingen

In tabel 5.1 op pagina 52 van de wetenschappelijke verantwoording worden van de twee afnamemomenten van de toets Rekenen-Wiskunde Standaard meetfout, de MAcc en de gesimuleerde Test-hertest betrouwbaarheid vermeld. Er wordt omschreven hoe de simulatie is uitgevoerd. Gegeven de software die er voor deze berekeningen is gebruikt, mag er van uitgegaan worden dat voornoemde gegevens correct berekend zijn.

Conclusie

Op aspect B1.1. wordt de toets Rekenen-Wiskunde 3.0 voor groep 7 als 'voldoende' beoordeeld.

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

Bevindingen

De betrouwbaarheidsscores waren hoger dan 0,70 en mogen daarmee als voldoende beoordeeld worden volgens de richtlijnen van de COTAN.

Wat betreft de lokale betrouwbaarheid, wordt in Tabel 5.2 tot en met Tabel 5.5a weergegeven hoe vaak de werkelijke vaardigheidsscore en de geschatte vaardigheidsscore in dezelfde categorie vallen. Voor alle categorieën is dit percentage voldoende hoog. De marginal classification accuracy loopt uiteen van 77,7 tot 80,4 procent.

Conclusie

Op aspect B1.2. wordt de toets Rekenen-Wiskunde 3.0 voor groep 7 als 'voldoende' beoordeeld.

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

De toetsontwikkelaars onderscheiden, overeenkomstig de referentieniveaus, vier domeinen:

- Getallen
- Verhoudingen
- Meten en meetkunde
- Verbanden.

In de toetsen voor groep 7 komen alle bovengenoemde domeinen voor. Deze onderwerpen komen ook aan bod in de meest gebruikte methodes voor rekenen-wiskunde van groep 7. Binnen het domein Getallen wordt een onderscheid gemaakt tussen het onderdeel Getallen en getalrelaties en het onderdeel Bewerkingen.

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 naast meerkeuzeopgaven ook open opgaven. De opgaven bestaan voor een deel uit opgaven met context en voor een deel uit opgaven zonder context. De opgaven en de vraagstelling zijn helder en eenduidig. Het antwoordmodel laat geen ruimte voor interpretatie.

De moeilijkheidsgraad is, gezien het niveau van de doelgroep, aanvaardbaar.

Wanneer een leerling onvoldoende leesvaardig is, kan ervoor gekozen worden om de opgaven voor te lezen. Het heeft dan overigens de voorkeur om de leerling de toets digitaal te laten maken en de computer de opgaven te laten voorlezen. Aanpassingen door bijvoorbeeld het toestaan van hulpmiddelen (voor leerlingen met dyscalculie) zijn niet toegestaan.

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

Inhoudsvaliditeit

Volgens de auteurs van de wetenschappelijke verantwoording wordt de inhoudsvaliditeit van de toets Rekenen-Wiskunde 3.0 gewaarborgd door de wijze waarop het instrument is ontwikkeld. In hoofdstuk 2 wordt uitvoerig verslag gedaan van de theoretische en praktische uitgangspunten die ten grondslag hebben gelegen aan de constructie van de toets. De inhoudsvaliditeit wordt beoordeeld door de onderwijskundige expert(s).

Begripsvaliditeit

In het kader van de begripsvaliditeit is onderzoek gedaan naar:

1. Unidimensionaliteit: Op basis van de grafische voorstellingen van de S-toetsen, de R1c toets en de 'c' uit het COTAN systeem kan geconcludeerd worden dat er

bij de toets Rekenen-Wiskunde 3.0 een beroep gedaan wordt op één onderliggende vaardigheid.

2. De itemkwaliteit: De gemiddelde moeilijkheidsgraad van de toetsen ligt op het gewenste niveau en de gemiddelde samenhang tussen de items uitgedrukt in Rit en Rir is goed. Ik adviseer wel om het ene item met een Rit waarde lager dan 0.20 uit de itembank te verwijderen.
3. Convergente en divergente validiteit Soortgenootonderzoek met verschillende andere LVS-onderdelen en de Schoolvaardigheidstoets Rekenen-Wiskunde van Boom liet zien dat de verwachtingen werden bevestigd.
4. Item bias Er is onderzoek gedaan naar DIF met betrekking tot sekse. Bij enkele items werd een lichte vorm van DIF gevonden. Geconcludeerd kan worden dat er nauwelijks sprake is van DIF met betrekking tot sekse.
5. Prestaties van relevante subgroepen Er is een effect gevonden voor de verschillende methoden. Bij 3 van de 5 methoden was het aantal leerlingen te laag om daar uitspraken over te kunnen doen. De effectgrootte voor de 2 resterende methoden was klein. Het patroon voor de verschillende halfjaargroepen is naar verwachting. Jongens scoren enigszins hoger dan meisjes. In termen van effectgrootte is er sprake van een klein effect. Dit is in lijn met wat verwacht wordt op basis van de literatuur. Er is een matig effect voor leerling gewicht.

Conclusie

Op aspect V1. wordt de toets Rekenen-Wiskunde 3.0 voor groep 7 als 'voldoende' beoordeeld.

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

In paragraaf 2.3 van de wetenschappelijke verantwoording wordt beschreven dat er sprake is van een bescheiden gemiddelde vaardigheidsgroei voor rekenen-wiskunde. De gemiddelde groei in groep 7 is kleiner dan die in groep 3 en 4. Bij het gebruik van de toets moet wel rekening gehouden worden met de meetfouten.

In de verschillende leerjaren wordt met verschillende toetsen gemeten die met behulp van IRT op één en dezelfde vaardigheidsschaal zijn gebracht.

Conclusie

Op aspect VA1.1 wordt de toets Rekenen-Wiskunde 3.0 voor groep 7 als 'voldoende' beoordeeld.

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 de handleiding wordt in hoofdstuk 4 ingegaan op hoe de resultaten van de toets geïnterpreteerd dienen te worden. Daarin wordt beschreven hoe gekeken kan worden of

een leerling zich ontwikkelt in lijn met de gehele populatie. De gerapporteerde score is een vaardigheidsscore en door deze te vergelijken met voorafgaande periodes kan de groei van een leerling in kaart gebracht worden.

De leerkrachten worden ook gewezen op de betrouwbaarheid van de gerapporteerde scores. Die betrouwbaarheid wordt ook vermeld op het leerling rapport.

Conclusie

Op aspect VA1.2 wordt de toets Rekenen-Wiskunde 3.0 voor groep 7 als 'voldoende' beoordeeld.

Inzicht in leervorderingen

I1.2 Levert de toetsaanbieder een format voor een geschreven toelichting bij de leervorderingen van de leerling die (ook) voor ouders/voogden/verzorgers begrijpelijk is?

Bevindingen

Hoofdstuk 7 van de handleiding heeft als titel 'Communiceren over resultaten'. Hier staat de interpretatie van de toetsresultaten centraal en krijgen de leerkrachten adviezen over hoe zij ouders kunnen helpen bij het interpreteren van de resultaten en hoe de eventuele hulp eruit zou kunnen zien.

De handleiding beschrijft uitvoerig hoe scholen met ouders over de toetsresultaten kunnen communiceren. Daarbij worden ook veelvoorkomende misverstanden rondom de interpretatie van volgsysteem toetsen besproken. Er kunnen verschillende rapporten worden gemaakt om de resultaten te laten zien waarbij zowel gekeken wordt naar het actuele niveau van de leerling als naar de ontwikkeling.

Conclusie

Op aspect I1.2 wordt de toets Rekenen-Wiskunde 3.0 voor groep 7 als 'voldoende' beoordeeld.

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	I1.2	Voldoende

4. Literatuurlijst

- Hop, M., Jansen J. & Engelen, R. (2017). *Wetenschappelijke verantwoording Rekenen-Wiskunde 3.0 voor groep 7*. Arnhem: Cito.
- Cito (2017). *Rekenen-Wiskunde 3.0 voor groep 7*. Arnhem: Cito.