

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 6 zijn een onderdeel van het Cito Volgstelsel primair en speciaal onderwijs en zijn bedoeld voor leerlingen in groep 6 van het primair onderwijs. De toetsen voor groep 6 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 6 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 6 zijn bedoeld voor leerlingen in groep 6 van het primair onderwijs en leerlingen in het speciaal (basis)onderwijs die functioneren op het niveau van groep 6 in het reguliere basisonderwijs. De toetsen zijn ook te gebruiken voor leerlingen in andere leerjaren die een rekenvaardigheid hebben op het niveau van groep 6. 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 6 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.

Beoordeling van LOVS-toets Rekenen-Wiskunde 3.0 groep 6 – Cito B.V.

- 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, maatzicht, 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.

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 erom dat leerlingen structuur en samenhang van verhoudingen op hoofdlijnen leren te doorzien en dat zij er in praktische situaties mee rekenen.

Bij het domein 'Meten en meetkunde' ligt de nadruk op het beschrijven van en het greep krijgen op ruimtelijke aspecten van de werkelijkheid.

Bij het domein 'Verbanden' gaat het om het omgaan met tabellen, diagrammen en grafieken.

Inhoud van het toetspakket

Het toetspakket Rekenen-Wiskunde 3.0 groep 6 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 E5M6 (makkelijke variant van de toets M5)
 - Toets M6 (Medio groep 5)
 - Toets M6E6 (makkelijke variant van de toets E5)
 - Toets E6 (Eind groep 5)
- Afnamekaarten met aanwijzingen voor de 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 mevrouw Paulyn K. Berding-Oldersma MSc (secretaris).

De kwaliteit van de steekproef

S1.1. Is de steekproef representatief?

Bevindingen:

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

Dit onderzoek heeft in 2014 plaatsgevonden en de verschillende items zijn voorgelegd aan respectievelijk 2144 en 2807 leerlingen die representatief waren voor de doelgroep (zie ook de evaluatie van de criteria N1.2.1 en N1.2.2).

In januari en juni 2015 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. In januari zijn de items van M6 voorgelegd aan 637 leerlingen en in juni de items van E6 aan 556 leerlingen. Gemiddeld waren er 212 waarnemingen per item in januari en gemiddeld 185 waarnemingen per item in juni. Van deze groepen is niet bekend in hoeverre ze representatief waren voor de doelgroep.

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 als 'voldoende' beoordeeld.

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

Bevindingen:

Voor M6 waren 210 items verdeeld over 11 verschillende opgavenboekjes in een onvolledig, maar verbonden design. Elk boekje bestond uit de bestaande M6-toets van LVS II aangevuld met een taak met 32 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 322.

Voor M6 waren 345 items verdeeld over 12 verschillende opgavenboekjes in een onvolledig, maar verbonden design. Elk boekje bestond uit de bestaande E6-toets van LVS II aangevuld met een taak met 33 nieuwe items. De bestaande LVS II toets telt 97 items. Elk nieuw item kwam in twee boekjes voor. Het gemiddeld aantal leerling antwoorden per item was 380.

Beoordeling van LOVS-toets Rekenen-Wiskunde 3.0 groep 6 – Cito B.V.

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 als 'voldoende' beoordeeld.

Normering

N1.2.1. Zijn de normgroepen groot genoeg?

Bevindingen:

Voor het afnamemoment M6 hebben 2144 leerlingen meegedaan aan het normeringsonderzoek en voor het afname moment E6 2807 leerlingen. Deze aantallen zijn groot genoeg om schattingen te maken voor de populatie. 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 als 'voldoende' beoordeeld.

N1.2.2. Zijn de normgroepen representatief?

Bevindingen:

Voor de normeringsonderzoeken M6 en E6 werden scholen geworven na het trekken van een representatieve steekproef, waar bij rekening gehouden werd met verdeling over strata, regio en verstedelijking. Achteraf is de verdeling over sekse gecontroleerd.

De representativiteit van de steekproef is onderzocht met betrekking tot strata, regio, urbanisatiegraad en sekse. Bij de definitie van strata 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 steekproef ten aanzien van het schooltype statistisch gezien afweek van de populatie voor beide afname momenten. De

Beoordeling van LOVS-toets Rekenen-Wiskunde 3.0 groep 6 – Cito B.V.

effectgroottes zijn echter klein. De conclusie is daarmee dat de steekproeven een goede afspiegeling vormen van de populatie.

Conclusie:

Op aspect N1.2.2 wordt de toets Rekenen-Wiskunde 3.0 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 3.0 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 gebruik, mag er van uitgegaan worden dat voornoemde gegevens correct berekend zijn.

Conclusie:

Op aspect B1.1. wordt de toets Rekenen-Wiskunde 3.0 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 5.9a 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 72 tot 81 procent.

Conclusie:

Op aspect B1.2. wordt de toets Rekenen-Wiskunde 3.0 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.

Beoordeling van LOVS-toets Rekenen-Wiskunde 3.0 groep 6 – Cito B.V.

In de toetsen voor groep 6 wordt de rekenvaardigheid van alle vier domeinen getoetst. Binnen het domein Getallen maken de toetsontwikkelaars een onderscheid tussen het onderdeel 'Getallen en getalrelaties' en het onderdeel 'Bewerkingen'.

Een verschil van de nieuwe toetsen ten opzichte van de oude toetsen is dat de nieuwe toetsen meer kale bewerkingen bevatten. De opgaven van de toetsen bestaan voor een deel uit opgaven met een context en voor een deel uit opgaven zonder context.

De opgavenboekjes van de reguliere toetsen M6 en E6, en van de tussentoetsen E5M6 en M6E6 bevatten elk 96 opgaven.

De toetsen bevatten zowel open als een beperkt aantal meerkeuzeopgaven.

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 papieren en digitale toetsen bevatten bijna exact dezelfde inhoud en meten dezelfde vaardigheden. De resultaten van de digitale toetsen kunnen vergeleken worden met de papieren versie en andersom.

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 leerlingen met een vertraagde ontwikkeling/zwakke 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 aandachtspanne/concentratieproblemen, leerlingen die langzamer werken dan gemiddeld, leerlingen met visusproblemen, leerlingen met dyslexie, en leerlingen met behoefte aan voorspelbaarheid doet de handleiding bruikbare aanpassingsvoorstellen in de handleiding (Tabel 2).

Voor leerlingen met dyscalculie worden (terecht) geen aanpassingsvoorstellen gedaan. Wel wordt erop gewezen dat het mogelijk is om een leerling een makkelijkere versie van de toets voor te leggen.

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

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.

Beoordeling van LOVS-toets Rekenen-Wiskunde 3.0 groep 6 – Cito B.V.

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.

3. *Convergente en divergente validiteit.* Correlatieonderzoek 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 significant effect gevonden. Geconcludeerd kan worden dat er nauwelijks sprake is van DIF met betrekking tot sekse.

5. *Prestaties van relevante subgroepen.* Er is geen effect gevonden voor de verschillende methoden. 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 groep 6 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 6 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 6 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 dient 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.

Beoordeling van LOVS-toets Rekenen-Wiskunde 3.0 groep 6 – Cito B.V.

In de handleiding worden de leerkrachten 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 6 als 'voldoende' beoordeeld.

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. wordt de toets Rekenen-Wiskunde 3.0 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	I1.1	Voldoende

4. Literatuurlijst

- Hop, M., & Engelen, R. (2017). *Wetenschappelijke verantwoording Rekenen-Wiskunde 3.0 voor groep 6*. Arnhem: Cito.
- Cito (2016). *Rekenen-Wiskunde 3.0 voor groep 6*. Arnhem: Cito.