



Regeling van het College voor toetsen en examens van 1 december 2025, nummer CvTE-25.012.12, houdende wijziging van de Regeling beoordelingskader doorstroomtoets PO in verband met de wijziging van de Regeling toetswijzer doorstroomtoets PO en enkele technische verbeteringen

Het College voor toetsen en examens,

Gelet op artikel 3a, eerste lid, onderdeel g, van de Wet College voor toetsen en examens;

Gezien de goedkeuring van de Staatssecretaris van Onderwijs, Cultuur en Wetenschap, gegeven op 15 december 2025, nummer 1783373,

Besluit:

ARTIKEL I

De regeling beoordelingskader doorstroomtoets PO wordt als volgt gewijzigd:

De bijlage wordt vervangen door de bijlage bij deze regeling.

ARTIKEL II INWERKINGTREDING

Deze regeling treedt in werking met ingang van de dag na de datum van uitgifte van de Staatscourant waarin de regeling wordt geplaatst.

Deze regeling zal met de toelichting in de Staatscourant worden geplaatst.

*Het College voor toetsen en examens,
de voorzitter,
J.H. van der Vegt*



BIJLAGE 1 BEOORDELINGSKADER VOOR DE DOORSTROOMTOETS

Bijlage behorende bij artikel 1 van de Regeling beoordelingskader doorstroomtoets PO.

1. Inleiding

1.1 Begrippen en definities

Aanbieder: aanbieder van een doorstroomtoets.

Absolute normen: standaarden waaraan de vaardigheidsniveaus van leerlingen gerelateerd kunnen worden. Deze standaarden worden vastgesteld op basis van inhoudelijke en/of toetsspecifieke criteria.

Afnemer: een afnemer van een toets zoals een leerkracht, intern-begeleider, directeur of observator.

Algoritme (beslisregel): systeem van regels, logische of wiskundige handelingen, die als stappenplan gelden en kunnen worden uitgevoerd om tot een bepaald resultaat te komen.

A-parameter: geeft bij een 2PL itemresponsmodel het discriminerend (onderscheidend) vermogen van een item weer. Hoe hoger het onderscheidend vermogen, hoe beter het item onderscheid maakt tussen lage en hoge vaardigheidsniveaus. In de praktijk worden de waarden relatief – d.w.z. tussen de items onderling – bekeken: hoe hoger de waarde van een bepaald item, hoe hoger het onderscheidend vermogen ten opzichte van de andere items.

A-waarde: is de proportie leerlingen die, of het percentage leerlingen dat, bij een meerkeuzevraag de desbetreffende afleider als antwoord heeft gekozen.

Beoordelaarsovereenstemming: de mate waarin verschillende beoordelaars tot identieke beoordelingen komen van (bijvoorbeeld) het werk van leerlingen. Twee veel gebruikte maten voor beoordelaarsovereenstemming zijn het percentage overeenstemming tussen twee beoordelaars en Cohen's Kappa.

Betrouwbaarheid: de mate waarin de toetsscores – op groepsniveau – vrij zijn van toevallige meetfouten.

B-parameter: verwijst over het algemeen naar de parameter in itemresponsmodellen die betrekking heeft op de moeilijkheidsgraad van een item.

CAT: computergestuurde adaptieve toets op itemniveau, waarbij een item wordt gekozen die qua moeilijkheid het beste aansluit op de vaardigheid zoals ingeschat op basis van de antwoorden op de eerdere opgaven.

Cesuur: minimale prestatie (gerepresenteerd als vaardigheid of toetsscore) die net indicatief is voor het halen van een (referentie)niveau en toetsadviescategorie.

Constructvaliditeit: eigenschap die een doorstroomtoets heeft als kan worden aangetoond dat de toets het door de constructeur beoogde kenmerk van de leerling (kennisniveau, vaardigheid) meet. Deze wordt per kenmerk binnen een doorstroomtoets bepaald.

Dichotome scoring: wijze van scoren (punten toekennen) waarbij slechts twee waarderingen van het antwoord worden onderscheiden: bijvoorbeeld 'goed' of 'fout'. Een meerkeuzevraag wordt dichotoom gescoord (0 = niet correct, 1 = correct).

Differentieel item functioneren (DIF): is een maat waarin items verschillend functioneren voor bepaalde (sub)groepen leerlingen, zelfs wanneer deze (sub)groepen leerlingen een vergelijkbaar prestatieniveau hebben. Het gaat hierbij over de vraagonzuiverheid, dat wil zeggen wanneer de verschillen in de prestaties op het item tussen groepen niet geheel verklaard kunnen worden door verschillen in de (latente) vaardigheid die aan de opgave ten grondslag ligt.

Discriminerend vermogen: mate waarin op grond van de scores op een item hoog- en laagscorende leerlingen (op de hele toets) onderscheiden kunnen worden.

Domein: verzameling van onderling samenhangende kennis en vaardigheden waarover getoetst wordt. Zo bestaat het terrein Rekenen uit de vier domeinen Getallen, Verhoudingen, Meten en Meetkunde en Verbanden.

Effectgrootte: statistische maat om de grootte van een effect – bijvoorbeeld het verschil in de gemiddelde rekenprestaties tussen jongens en meisjes – weer te geven. Er wordt een onderscheid gemaakt tussen niet gestandaardiseerde (ruwe) effectgroottes en gestandaardiseerde effectgroottes. De bekendste gestandaardiseerde effectmaat is Cohen's d. Deze maat geeft het verschil tussen twee gemiddelden in het aantal standaarddeviaties. Ook de correlatie kan worden gezien als een gestandaardiseerde maat voor de effectgrootte.

Gezamenlijk anker: verzameling van items die in alle doorstroomtoetsen wordt opgenomen om verschillen in de moeilijkheidsgraad tussen de betreffende doorstroomtoetsen vast te kunnen stellen.

Inhoudsgebied: vaardigheid die in het onderwijsaanbod in het primair onderwijs is opgenomen. Zoals vaardigheden waarover referentieniveaus kunnen worden gerapporteerd (rekenen, lezen, taalver-zorging) of andere vaardigheden waarop op hoofdniveau wordt gerapporteerd.

Inhoudsvaliditeit: mate waarin de inhoud van de doorstroomtoets een representatieve weergave is van het te meten referentieniveau.

Itemkalibratie: schatting (berekening) van de psychometrische eigenschappen (parameters) van een verzameling items met het doel om deze items op dezelfde schaal te brengen.

Itemtype: vraagvorm, zoals meerkeuze, open, dictee.



Jaarset: de selectie van gezamenlijke ankeritems uit de ankerbank die in een bepaald afnamejaar wordt afgenomen.

Leerdoel: beoogde resultaat van onderwijsleeractiviteiten. Als het beoogde leerresultaat wordt geformuleerd in termen van uiterlijke condities en waarneembaar leerlinggedrag spreekt men van 'concrete leerdoelen' of 'gedragsdoelen'. Als het beoogde leerresultaat wordt geformuleerd in termen van cognitief gedrag of van cognitieve structuren, spreekt men van cognitieve leerdoelen.

Meetnauwkeurigheid: mate waarin de toetsscores op een specifiek deel van de meetschaal vrij zijn van toevallige meetfouten.

Module: set van items in een multistage toets (MST), vaak van vergelijkbare moeilijkheid.

Moeilijkheidsgraad: in hoeverre leerlingen uit de doelpopulatie in staat zijn om de vraag of vragen goed te beantwoorden. De moeilijkheid wordt meestal uitgedrukt met een p-waarde.

Multistage toets (MST): adaptieve doorstroomtoets op moduleniveau. Een MST lijkt op een CAT, maar in plaats dat items een voor een worden geselecteerd, worden er sets van items afgenomen in modules. Nadat een leerling een set heeft afgerond, wordt een nieuwe inschatting van de vaardigheid gemaakt en op basis daarvan wordt een nieuwe set gekozen. Een MST kan een digitale toets zijn, of één op papier.

Normeren: toekennen van een waardering aan een score. Bij de landelijke (toetsoverstijgende) normering van de doorstroomtoets betreft dit het vaststellen van de cesuren van de referentieniveaus en de toetsadviescategorieën.

Onderwijstype: specificatie van onderwijssoort op basis van de mate van extra ondersteuning. Onderscheiden worden regulier basisonderwijs (BO), speciaal onderwijs (SO) en speciaal basis-onderwijs (SBO).

Plan, Do, Check, Act (PDCA) cyclus: is een model om de continue verbetering en vernieuwing in een organisatie te sturen. Het cyclisch karakter zorgt ervoor dat de kwaliteitsverbetering continu onder de aandacht is.

Po: primair onderwijs

Populatie: verzameling van vooraf gedefinieerde eenheden waarover men met behulp van statistisch onderzoek op steekproeven uit die populatie uitspraken wil doen.

Pretesten: afnemen van een aantal opgaven bij een groep leerlingen, meestal met het doel de eigenschappen van de doorstroomtoets te onderzoeken. Deze groep leerlingen is zoveel mogelijk vergelijkbaar met de populatie waarvoor de doorstroomtoets oorspronkelijk is ontwikkeld.

P-waarde: is een getal tussen 0 en 1 waarmee de moeilijkheidsgraad van een opgave kan worden weergegeven. De p-waarde wordt berekend door de gemiddelde score op een opgave te delen door de maximaal haalbare score op die opgave. De p-waarde voor een meerkeuzevraag is gelijk de proportie van de leerlingen die het juiste antwoord heeft gekozen.

Rar-waarde: correlatie tussen een afleider van een item en de totaalscore minus het betreffende item. Hiermee wordt de kwaliteit van de afleiders onderzocht voor meerkeuze-opgaven.

Referentiekader: beschrijving van wat leerlingen op verschillende momenten in hun schoolloopbaan op het gebied van Nederlandse taal en Rekenen moeten kennen en kunnen (zie ook: Besluit referentieniveaus Nederlandse taal en Rekenen).

Referentieniveaus: beschrijvingen van beheersingsdoelen voor Nederlandse taal en Rekenen die aangeven wat leerlingen moeten kennen en kunnen op verschillende momenten tijdens een schoolloopbaan. De referentieniveaus zijn wettelijk verankerd in de Wet referentieniveaus Nederlandse taal en Rekenen.

Rit-waarde: correlatie tussen het item en de totaalscore op alle items. De item-testcorrelatie is een maat voor het discriminerend vermogen van items. De item-testcorrelatie valt hoger uit dan de item-restcorrelatie omdat hier de itemscore in de totaalscore van een toets wordt meegeteld, waardoor sprake is van autocorrelatie. Daarom wordt aangeraden om de item-restcorrelatie te gebruiken.

Rir-waarde: correlatie tussen het item en de totaalscore minus het betreffende item. De item-restcorrelatie is een maat voor het discriminerend vermogen van een item.

Schaal: reeks getallen die volgens een bepaald voorschrift gekoppeld worden aan waarnemingen.

Steekproef: verzameling leerlingen met specifieke kenmerken die op aselecte wijze uit de te onderzoeken populatie zijn getrokken.

Stopcriterium: criteria op basis waarvan een CAT wordt stopgezet. Na iedere opgave in de adaptieve module wordt automatisch geëvalueerd of het criterium is bereikt (e.g., de standaardfout op de vaardigheidsschatting is onder een bepaalde grenswaarde gedaald, een vooraf vastgestelde maximale toetslengte is bereikt).

Subdomein: verzameling van onderling samenhangende kennis en vaardigheden waarover getoetst wordt. Zo bestaat het domein lezen uit de subdomeinen zakelijke en fictionele teksten.

Taaktype/taakuitvoering: classificering of combinatie van classificaties van de onderverdeling van een inhoudsgebied, zoals begrijpen, interpreteren, contextloos, morfologische spelling.

Terrein: centraal onderdeel dat in de doorstroomtoets aan bod moet komen. Het referentiekader onderscheidt twee centrale terreinen: het terrein Nederlandse taal en het terrein Rekenen. Beide terreinen worden uitgewerkt in onderliggende (sub)domeinen.

Toetsadvies: op basis van de behaalde resultaten op de doorstroomtoets berekende best passende niveau voor het volgen van aansluitend voorgezet onderwijs voor de leerling.

(Toets)item: is synoniem voor (toets)opgave.

Toetsmatrijs: een tabel die de blauwdruk is van een toets. Bevat o.a. het aantal vragen per toets verdeeld over aspecten en niveaus.

(Toets)opgave: is een onderdeel van een toets en bevat altijd een stimulus (een middel of prikkel om een bepaald gedrag van een kandidaat uit te lokken) waarop een respons (het uitgelokte gedrag) mogelijk moet zijn. Toetsreglement document waarin de wijze van afnemen wordt geregeld.

Toetsterm: concrete, eenduidige en meetbare operationalisering van het betreffende leerdoel zoals beschreven in het bijpassende referentieniveau Nederlandse taal en Rekenen.

Zaai-opgave: een toetsopgave die in een doorstroomtoets wordt gepretest om data over te verzamelen ten behoeve van parameterschattingen.

1.2 Reikwijdte

Sinds het schooljaar 2014–2015 zijn scholen in het primair onderwijs verplicht in groep 8 een eindtoets Nederlandse taal en Rekenen af te nemen. Per 1 januari 2023 is de eindtoets vervangen door de doorstroomtoets. De Wet doorstroomtoets po schrijft voor dat het College voor Toetsen en Examens (CvTE) als wettelijke taak heeft om doorstroomtoetsen te erkennen. Doorstroomtoetsen die het CvTE erkent, worden voor een periode van vier jaar toegelaten tot het primair onderwijs.

Het CvTE maakt voor het erkennen van een doorstroomtoets gebruik van een adviseur. De adviseur baseert diens advies op het beoordelingskader vastgesteld met deze regeling.

Op basis van het advies van de adviseur stelt het CvTE tevens jaarlijks vast of de erkende doorstroomtoets voor dat jaar voldoet aan de criteria van de Regeling beoordelingskader doorstroomtoetsen po.

Bij het beoordelen van extra kennisgebieden, buiten de verplichte en optionele domeinen van de terreinen Nederlandse taal en Rekenen, oordeelt het CvTE of ook dat deel van de toets inhoudelijk valide en betrouwbaar is en of het van een deugdelijke normering is voorzien. Voor de beoordeling van dit deel worden de toetsbare onderdelen van dit beoordelingskader toegepast.

Het CvTE beoordeelt ook de door de overheid aangeboden calamiteitentoets¹. Vanwege de bijzondere aard van de calamiteitentoets, bestaat de jaarlijkse vaststelling slechts uit een inhoudelijke check op actualiteiten en de kalibratie. Indien de calamiteitentoets in enig jaar moet worden ingezet, wordt een nieuwe toets voor het daarop volgende jaar ontwikkeld. Deze moet dan opnieuw erkend te worden door het CvTE.

1.3 Relatie beoordelingskader tot andere wetgeving

Deze regeling geeft een beoordelingskader voor de erkenning en jaarlijkse vaststelling van de doorstroomtoets primair onderwijs, ten aanzien van de toepassing van de psychometrische, onderwijskundige en organisatorische aspecten van de toets als bedoeld in artikel 3a, eerste lid aanhef en onder g van de Wet College voor toetsen en examens. Het CvTE betreft bij zijn beoordeling niet of de aanbieder voldoet aan andere wet- en regelgeving zoals die geldt voor het gebruik van intellectueel eigendom of die geldt op grond van de Artificial Intelligence-verordening.

1.4 Wat minimaal aan te leveren bij een indiening?

Het CvTE controleert of alle gegevens en bescheiden als bedoeld in de paragrafen 1.4.1 t/m 1.4.2 en eventueel paragrafen 1.4.3 t/m 1.4.4 zijn aangeleverd bij het indienen van de aanvraag. Indien dit het geval is, zal de adviseur worden ingeschakeld voor het opstellen van een advies op basis van de kwaliteitseisen in hoofdstuk 3, 4 en 5. De adviseur kan vaststellen dat hij nog informatie mist om de aanvraag te kunnen beoordelen. Het CvTE zal dan vragen aan de aanbieder het gebrek in de aanvraag (alsnog) te herstellen. Het CvTE kan dan de indiening buiten behandeling laten als de toetsaanbieder het gebrek in de aanvraag binnen de termijn die hij hiervoor krijgt niet herstelt.

1.4.1 Aan te leveren documentatie

De aanvrager moet minimaal de volgende gegevens en bescheiden verstrekken:

- Een aanbiederbrief met daarin het KvK-nummer van de toetsaanbieder en de officiële naam van de toetsaanbieder en het aangeboden instrument;

¹ https://www.eerstekamer.nl/behandeling/20210907/gewijzigd_amendement_van_het_lid/document3/f=/vImnj6ccx4n2.pdf

- Een ingevuld door het CvTE beschikbaar gestelde model voor een leeswijzer² met daarin uiteengezet in welk aangeleverd document en hoofdstuk een kwaliteitscriterium uit het beoordelingskader wordt verantwoord;
- Toets met bijbehorende ondersteunende materialen voor de gebruiker (inclusief beschikbare versie(s) voor leerlingen met ondersteuningsbehoeften). De aanbieder wordt gevraagd om alle toetsopgaven inzichtelijk aan te leveren. Daaronder vallen alle items van de doorstroomtoets zelf én ankeropgaven en eventuele zaai-opgaven;
- Itemlijst met itemgegevens (volgnummer, itemcode, (sub)domein, itemtype, sleutel en/of scoringsregels, kenmerk van de taakuitvoering, referentieniveau, ankertype (intern of gezamenlijk)) en indien van toepassing vermelden in welk jaartal item in de itembank is opgenomen, aangeven of en wanneer de items zijn aangeboden voor de onderwijskundige opgaven feedbackservice (OOF) en indien items al eerder zijn afgenomen in operationele setting deze jaartallen vermelden;
- Praktische handleiding voor leerkrachten;
- Wetenschappelijke en inhoudelijke verantwoording, inclusief de psychometrische aspecten van de doorstroomtoets, zoals beschreven in Hoofdstuk 5 van deze regeling en zoals beschreven in de Regeling Beoordelingsnormen Doorstroomtoetsen po.
- Toetsreglement van de betreffende toets;
- Informatie over de toetsinhoud en de afnamevorm;
- Voorbeeld leerlingrapportage en eventueel andere rapportages;
- Toetsveiligheidsplan;
- Actieplan hoe om te gaan met diefstal van toetsinhoud;
- Fraudepreventieplan;
- Calamiteitenplan;
- Indien van toepassing: Cd's, dvd's en/of usbsticks en log-ins en wachtwoorden;
- Indien van toepassing: Antwoordformulieren;
- Indien van toepassing: ISO-certificaat inclusief een verklaring van toepasbaarheid (c.q. de scope van het certificaat).

Een aanbieder heeft de mogelijkheid om voor verschillende wijzen van afname (bijv. digitaal en papier) een erkenning of vaststelling aan te vragen (naast de optie om verschillende wijzen van afname in één aanvraag in te dienen bij het CvTE). Indien een aanbieder hiervoor kiest, moet de aanbieder ook daadwerkelijk meerdere aanvragen in te dienen, in separate mappen op de terminal server, met separaat ingevulde leeswijzers. Allebei de aanvragen krijgen dan ook separaat een erkenning of vaststelling. Deze aanvragen moeten volledig onafhankelijk van elkaar kunnen worden beoordeeld. Dat wil zeggen dat de ene afnamemodus bijvoorbeeld niet de terugvaloptie kan zijn van de andere afnamemodus. Dit zou namelijk betekenen dat als de ene afnamemodus niet wordt erkend of vastgesteld, dit ervoor zorgt dat door de afhankelijkheid de andere afnamemodus ook niet kan worden erkend of vastgesteld.

1.4.2 Het aanleveren van items

Om de items in een doorstroomtoets efficiënt en gestructureerd te kunnen beoordelen, zijn er eisen aan het leveren van items aan het CvTE:

In het geval van een papieren toets:

- Toetsboeken (digitaal aanleveren volstaat) inclusief ankeropgaven en eventuele zaai-opgaven.

In het geval van een lineaire digitale toets:

- Inlogmogelijkheid als leerling en als administrator (scrollen zonder antwoorden in te vullen, toets onderbreken en op later tijdstip te hervatten);
- Mogelijkheid alle items te kunnen inzien inclusief ankeropgaven en eventuele zaai-opgaven;
- Mogelijkheid om items per inhoudsgebied te bekijken.

In het geval van een MST:

- Inlogmogelijkheid als leerling en als administrator (scrollen zonder antwoorden in te vullen, toets onderbreken en op later tijdstip te hervatten);
- Mogelijkheid alle items te kunnen inzien inclusief ankeropgaven en eventuele zaai-opgaven;
- Mogelijkheid om items per inhoudsgebied te bekijken;
- Mogelijkheid om items per module te bekijken;
- Mogelijkheid om items in (bij voorkeur verschillende) toetspaden te bekijken.

In het geval van een CAT:

² Indien een aanbieder ervoor kiest om de volgorde van de kwaliteitscriteria uit het beoordelingskader als uitgangspunt voor één verantwoordingsdocument te nemen, hoeft de aanbieder geen aanvullende leeswijzer in te dienen.



- Inlogmogelijkheid als leerling en als administrator (scrollen zonder antwoorden in te vullen, toets onderbreken en op later tijdstip te hervatten);
- Mogelijkheid alle items te kunnen inzien inclusief ankeropgaven en eventuele zaai-opgaven;
- Mogelijkheid om items per inhoudsgebied te bekijken;
- Mogelijkheid om items in (bij voorkeur verschillende) toetspaden te bekijken.

In het geval van aparte varianten van de doorstroomtoets voor leerlingen met specifieke ondersteuningsbehoeften:

- Mogelijkheid om een aantal voorbeelditems te bekijken waarbij een duidelijk beeld wordt gegeven van de aanpassing, zoals items met gebarentaal of aanpassing van de kleuren en/of contrast.

1.4.3 Optioneel: het aanleveren van data t.b.v. landelijke normering

Het aanleveren van data t.b.v. de landelijke normering is alleen van toepassing voor al eerder afgenomen doorstroomtoetsen. De aanvrager moet minimaal de data verstrekken of heeft de data al verstrekt ten behoeve van de landelijke normering zoals uiteengezet in hoofdstuk 3.2 t/m 3.2.7 van de Regeling Beoordelingsnormen Doorstroomtoetsen PO.

1.4.4 Optioneel: het aanleveren van data t.b.v. toelatings- en doorstroomonderzoek

Het aanleveren van data t.b.v. toelatings- en doorstroomonderzoek is alleen van toepassing voor al eerder afgenomen doorstroomtoetsen. De aanvrager moet minimaal de data verstrekken of heeft de data al verstrekt ten behoeve van toelatings- en doorstroomonderzoek zoals uiteengezet in hoofdstuk 5.1 t/m 5.4.2 van de Regeling Beoordelingsnormen Doorstroomtoetsen PO.

2. Format Regeling beoordelingskader doorstroomtoetsen po

Het beoordelingsformat bevat de volgende drie inhoudelijke onderdelen, die enkele onderliggende thema's bevatten:

- H.3. Onderwijskundige aspecten
- H.4. Organisatorische aspecten
- H.5. Psychometrische aspecten

3. Onderwijskundige aspecten

3.1. Inleiding

In de volgende paragrafen worden de vakinhoudelijke kwaliteitseisen uitgewerkt voor de wettelijk verplichte domeinen Lezen en Begrippenlijst en Taalverzorging, daaronder niet begrepen het subdomein Begrippenlijst, van het terrein Nederlandse taal en de wettelijk verplichte domeinen Getallen, Verhoudingen, Meten en Meetkunde en Verbanden van het terrein Rekenen én de optionele domeinen Schrijven en Mondelinge taalvaardigheid en het eveneens optionele subdomein Begrippenlijst van het terrein Nederlandse taal, in lijn met het Toetsbesluit PO, de Wet doorstroomtoetsen po en zoals vastgesteld in het Besluit referentieniveaus Nederlandse taal en Rekenen. Voor alle wettelijk verplichte en optionele (sub)domeinen zijn in deze regeling kwaliteitseisen geformuleerd.

Naast de wettelijk verplichte én optionele terreinen en (sub)domeinen staat het de toetsaanbieder vrij om extra kennisgebieden aan de doorstroomtoets toe te voegen, bijv. aardrijkskunde. De score die de leerling op deze extra kennisgebieden haalt, mag de toetsaanbieder toevoegen aan het leerlingrapport. Deze extra kennisgebieden tellen niet mee voor het berekende toetsadvies en tellen ook niet mee voor de berekende score op de referentieniveaus. Alle toetsopgaven van de wettelijk verplichte en optionele (sub)domeinen en terreinen, én de extra kennisgebieden worden inhoudelijk door een adviseur van het CvTE beoordeeld. Zie in dit kader de kwaliteitseisen in Hoofdstuk 3.2. Dit beoordelingskader schrijft niet voor wat de minimale en maximale toetslengte in aantal toetsvragen van een doorstroomtoets moet zijn. Eveneens worden er geen richtlijnen gegeven over de verhouding van het aantal toetsvragen tussen de wettelijk verplichte en optionele (sub)domeinen en terreinen, én extra kennisgebieden. Verder worden er ook geen richtlijnen gegeven voor het aantal toetsvragen per referentieniveau 1F/2F/1S. Dit om de aanbieders de mogelijkheid te geven om zich van elkaar te onderscheiden en volgens de eigen zienswijze een compleet beeld van de leerlingen te kunnen geven.

3.2. Inhoudvaliditeit doorstroomtoets

Kwaliteitseisen

Beslisregel: Om aan deze categorie kwaliteitseisen te voldoen, moeten alle vragen met JA worden beantwoord.

Code	Vragen	Mogelijke antwoorden
V1	Is de toetsmatrijs een adequate representatie van het meetdoel voor in ieder geval het wettelijk verplichte domein Lezen en subdomein Taalverzorging van het terrein Nederlandse taal en het terrein Rekenen én de, indien van toepassing, optionele (sub)domeinen van Nederlandse taal?	ja/nee
V2	Zijn de toetsopgaven – inclusief eventuele zaai-opgaven – correct geconstrueerd en voldoen deze aan de kwaliteitscriteria voor de inhoudsvaliditeit (zijnde de relevantie van de inhoud van de toets)?	ja/nee

Toelichting V1: De toetsmatrijs is voor wat betreft het wettelijk verplichte domein Lezen en subdomein Taalverzorging van het terrein Nederlandse taal en het terrein Rekenen én de, indien van toepassing, optionele (sub)domeinen van Nederlandse taal een adequate representatie van het meetdoel. Dat betekent eveneens dat er via de toetsmatrijs wordt voldaan aan de eisen uit de Toetswijzer PO.

Er is sprake van een adequate representatie wanneer de toetstermen het meetdoel representeren. Dit blijkt uit het gegeven dat:

- de leerdoelen zijn geoperationaliseerd in observeerbare en meetbare toetstermen die ieder voor zich aansluiten bij de domeinbeschrijvingen en indeling van de referentieniveaus voor het wettelijk verplichte domein Lezen en subdomein Taalverzorging van het terrein Nederlandse taal en het terrein Rekenen én de, indien van toepassing, optionele (sub)domeinen van Nederlandse taal;
- de toetsmatrijs in ieder geval bevat: (1) het aantal vragen met bijbehorende scorepunten per vraag, (2) de toetsvorm en/of het type vragen, (3) de verdeling over de (sub)domeinen, (4) de verdeling van de verschillende vraag- en teksttypes in de toets, (5) de verdeling over de referentieniveaus.
- in het geval van een CAT dat de aanbieder verantwoordt hoe het wettelijk verplichte domein Lezen en subdomein Taalverzorging van het terrein Nederlandse taal en het terrein Rekenen én de, indien van toepassing, optionele (sub)domeinen van Nederlandse taal op representatieve wijze voor alle individuele leerlingen aan bod komen.

Toelichting V2: In de Checklist voor het beoordelen van de kwaliteit van observatie-categorieën en toetsopgaven staan de constructievoorschriften voor toetsvragen en de vijf kwaliteitscriteria (relevantie, objectiviteit, efficiëntie, specificiteit en neutraliteit) beschreven. De toetsopgaven moeten hieraan voldoen – inclusief eventuele zaai-opgaven – van in ieder geval het wettelijk verplichte domein Lezen en subdomein Taalverzorging van het terrein Nederlandse taal en het terrein Rekenen én de, indien van toepassing, optionele (sub)domeinen van Nederlandse taal. Indien in de doorstroomtoets ook optionele productieve vaardigheden worden getoetst, levert de aanbieder een beoordelaarsschema in, aangevuld met informatie over de beoordelaarsovereenstemming.

In het geval een item niet voldoet aan de kwaliteitscriteria voor inhoudsvaliditeit, moet het betreffende item te worden verwijderd of vervangen. Indien de adviseur twijfelt over de mate waarin een item voldoet aan de kwaliteitscriteria, wordt de aanbieder verzocht om te reflecteren op de kwaliteit. Daarnaast moet de aanbieder beargumenteren door middel van een inhoudelijke en psychometrische onderbouwing waarom de toetsopgave wel óf niet volledig voldoet aan de kwaliteitscriteria. Dit kan resulteren in één van de volgende opties:

1. de toetsopgave wordt verwijderd;
2. de toetsopgave wordt aangepast (minimale aanpassing met weinig impact);
3. de toetsopgave wordt – op basis van de inhoudelijke en/of psychometrische onderbouwing – goedgekeurd.

Indien een opgave op basis van punt 2 of 3 wordt behouden, moet deze toetsopgave bij de afname gemonitord te worden door de toetsaanbieder. Na de operationele afname analyseert de toetsaanbieder het functioneren van deze toetsopgave (KF2).

3.3. Terrein Nederlandse taal

In iedere doorstroomtoets worden voor wat betreft het terrein Nederlandse taal de wettelijk verplichte domeinen Lezen (paragraaf 3.3.1) en Begrippenlijst en Taalverzorging, daaronder niet begrepen het subdomein Begrippenlijst, (paragraaf 3.3.2) getoetst. Aanvullend hierop kunnen optioneel één of meerdere van de domeinen Schrijven (paragraaf 3.3.3) en Mondelinge taalvaardigheid (paragraaf 3.3.4) en het subdomein Begrippenlijst (paragraaf 3.3.5) worden getoetst. De gestelde vakinhoudelijke

kwaliteitseisen voor deze vijf onderdelen en de vereiste verdeling van de toetsopgaven over de (sub)domeinen in de toetsmatrijs en de toetssamenstelling zijn opgenomen in de Toetswijzer PO.

3.3.1. Wettelijk verplicht domein: Lezen

Beslisregel:

Om aan deze categorie kwaliteitseisen te voldoen, gelden de volgende beslisregels:

1. Relatieve verdeling: van RL1 t/m RL8 moeten alle vragen met JA of N.V.T. worden beantwoord.
2. Taken: van L1.1 t/m L1.3 moeten tenminste twee van de drie vragen met JA worden beantwoord en moet L4 met JA worden beantwoord.
3. Tekstkenmerken: vraag L2 en L5 moeten beide met JA worden beantwoord.
4. Kenmerken van de taakuitvoering: de vragen L3.3 t/m L3.7 en L6.2 t/m L6.4 moeten met JA of N.V.T. worden beantwoord. De vraag L3.1 mag met een NEE worden beantwoord. Van de vragen L3.2 en L6.1 moet minstens één vraag met een JA worden beantwoord.

3.3.1.1. Domein Lezen

Kwaliteitseisen

Relatieve verdeling:

Code	Vragen	Mogelijke antwoorden
RL1	Is het percentage aangeboden toetsopgaven van het domein lezen ten minste 50% van alle aangeboden toetsopgaven van de verplichte (sub)domeinen lezen en taalverzorging bij het terrein Nederlandse taal?	ja/nee/n.v.t.
RL2	Is het percentage aangeboden toetsopgaven van het domein lezen met het kenmerk van de taakuitvoering techniek en woordenschat maximaal 6% van alle aangeboden toetsopgaven van de verplichte (sub)domeinen lezen en taalverzorging bij het terrein Nederlandse taal?	ja/nee/n.v.t.
RL3	Is het percentage aangeboden toetsopgaven van het domein lezen met het kenmerk van de taakuitvoering begrijpen minimaal 10% van alle aangeboden toetsopgaven van de verplichte (sub)domeinen lezen en taalverzorging bij het terrein Nederlandse taal?	ja/nee/n.v.t.
RL4	Is het percentage aangeboden toetsopgaven van het domein lezen met het kenmerk van de taakuitvoering interpreteren minimaal 10% van alle aangeboden toetsopgaven van de verplichte (sub)domeinen lezen en taalverzorging bij het terrein Nederlandse taal?	ja/nee/n.v.t.
RL5	Is het percentage aangeboden toetsopgaven van het domein lezen met het kenmerk van de taakuitvoering samenvatten minimaal 6% van alle aangeboden toetsopgaven van de verplichte (sub)domeinen lezen en taalverzorging bij het terrein Nederlandse taal?	ja/nee/n.v.t.
RL6	Is het percentage aangeboden toetsopgaven van het domein lezen met het kenmerk van de taakuitvoering opzoeken minimaal 6% van alle aangeboden toetsopgaven van de verplichte (sub)domeinen lezen en taalverzorging bij het terrein Nederlandse taal?	ja/nee/n.v.t.
RL7	Is het percentage aangeboden toetsopgaven van het domein lezen met het kenmerk van de taakuitvoering evalueren minimaal 1% van alle aangeboden toetsopgaven van de verplichte (sub)domeinen lezen en taalverzorging bij het terrein Nederlandse taal?	ja/nee/n.v.t.
RL8	Verantwoordt de toetsaanbieder – in het geval van een adaptieve toets waarbij het aantal aangeboden toetsopgaven per leerling verschilt (CAT) – adequaat hoe deze streeft voor elke leerling de percentages uit de kwaliteitscriteria RL1 t/m RL7 te bewerkstelligen?	ja/nee/n.v.t.

Toelichting RL1 t/m RL7: Deze kwaliteitscriteria zijn van toepassing voor lineaire toetsen en adaptieve toetsen waarvan het aantal aangeboden toetsopgaven niet per leerling verschilt (MST). De toetsaanbieder toont aan hoe de percentages uit de kwaliteitscriteria RL1 t/m RL7 in een lineaire toets of in het geval van een MST per toetspad worden gegarandeerd.

Toelichting RL8: Dit kwaliteitscriterium is van toepassing op adaptieve toetsen waarvan het aantal aangeboden toetsopgaven per leerling verschilt (CAT). Voor deze adaptieve toetsen zijn de genoemde percentages uit kwaliteitscriteria RL1 t/m RL7 niet te garanderen voor elke leerling. De toetsaanbieder verantwoordt daarom adequaat hoe deze toch streeft voor elke leerling de percentages uit de kwaliteitscriteria RL1 t/m RL7 te bewerkstelligen. En ook uitgebreid in te gaan en te reflecteren op mogelijk situaties – in toetspaden – waarin deze percentages niet aan een leerling worden aangeboden.

3.3.1.2. Subdomein Zakelijke Teksten

Kwaliteitseisen

Taken:

Code	Vragen	Mogelijke antwoorden
L1.1	Lezen de leerlingen minimaal één informatieve tekst en maken ze daar opdrachten bij?	ja/nee
L1.2	Lezen de leerlingen minimaal één instructieve tekst en maken ze daar opdrachten bij?	ja/nee
L1.3	Lezen de leerlingen minimaal één betogende tekst en maken ze daar opdrachten bij?	ja/nee

Tekstkenmerken:

Code	Vragen	Mogelijke antwoorden
L2	Voldoen de teksten die de leerlingen lezen aan de tekstkenmerken inhoud (thema), vorm (taalgebruik en structuur) en bedoeling (doel), zoals nader omschreven in het Referentiekader?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
L3.1	Wordt in de toets het optionele kenmerk van de taakuitvoering techniek en woordenschat getoetst?	ja/nee
L3.2	Wordt in de toets als kenmerk van de taakuitvoering evalueren getoetst?	ja/nee
L3.3	Wordt in de toets als kenmerk van de taakuitvoering begrijpen getoetst?	ja/nee
L3.4	Wordt in de toets als kenmerk van de taakuitvoering interpreteren getoetst?	ja/nee
L3.5	Wordt in de toets als kenmerk van de taakuitvoering samenvatten getoetst?	ja/nee
L3.6	Wordt in de toets als kenmerk van de taakuitvoering opzoeken getoetst?	ja/nee
L3.7	Verantwoordt de toetsaanbieder (o.a. met het opleveren van enkele (voorbeeld)toetspaden) op basis van welke beslisregel (het algoritme in het geval van een CAT en het module-design in het geval van een MST) wordt gegarandeerd dat de vereiste verdeling tussen opgaven rond zakelijke teksten wordt gehaald?	ja/nee/n.v.t.

Toelichting L1.1 t/m L1.3: Tenminste twee van de drie taken moeten in de doorstroomtoets worden opgenomen.

Toelichting L3.7: Alleen van toepassing bij een digitale adaptieve toets (CAT of MST). Naast het algoritme of module-design moet de werking van de beslisregel worden aangetoond door middel van het opleveren van enkele (voorbeeld)toetspaden, zoals die door de beslisregel worden gegenereerd.

Toelichting L1 t/m L3: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010, p. 12–14) leidend.

3.3.1.3. Subdomein Fictionele, narratieve en literaire teksten

Kwaliteitseisen

Taken:

Code	Vragen	Mogelijke antwoorden
L4	Lezen de leerlingen minimaal één fictionele, narratieve en/of literaire tekst en maken ze daar opdrachten bij?	ja/nee

Tekstkenmerken:

Code	Vragen	Mogelijke antwoorden
L5	Voldoen de teksten die de leerlingen lezen aan de tekstkenmerken inhoud (thema), vorm (taalgebruik en structuur) en bedoeling (doel), zoals nader omschreven in het Referentiekader?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
L6.1	Wordt in de toets als kenmerk van de taakuitvoering evalueren getoetst?	ja/nee
L6.2	Wordt in de toets als kenmerk van de taakuitvoering interpreteren getoetst?	ja/nee
L6.3	Wordt in de toets als kenmerk van de taakuitvoering begrijpen getoetst?	ja/nee
L6.4	Verantwoordt de toetsaanbieder (o.a. met het opleveren van enkele (voorbeeld)toetspaden) op basis van welke beslisregel (het algoritme in het geval van een CAT en het module-design in het geval van een MST) wordt gegarandeerd dat de vereiste verdeling tussen opgaven rond fictionele, narratieve en literaire teksten wordt gehaald?	ja/nee/n.v.t.

Toelichting L6.4: Alleen van toepassing bij een digitale adaptieve toets (CAT of MST). Naast het algoritme of module-design moet de werking van de beslisregel worden aangetoond door middel van het opleveren van enkele (voorbeeld)toetspaden, zoals die door de beslisregel worden gegenereerd.

3.3.2. Wettelijk verplicht subdomein: Taalverzorging

Beslisregel:

Om aan deze categorie kwaliteitseisen te voldoen, gelden de volgende beslisregels:

1. Relatieve verdeling en verschillende categorieën: Om aan deze categorie kwaliteitseisen te voldoen, moeten de vragen TV1 t/m TV7 met JA of N.V.T. worden beantwoord.

Kwaliteitseisen

Relatieve verdeling en verschillende categorieën:

Code	Vragen	Mogelijke antwoorden
TV1	Is het percentage aangeboden toetsopgaven van het subdomein taalverzorging ten minste 30% van alle aangeboden toetsopgaven van de verplichte (sub)domeinen lezen en taalverzorging bij het terrein Nederlandse taal?	ja/nee/n.v.t.
TV2	Is het percentage aangeboden toetsopgaven van het subdomein Taalverzorging met het aspect spelling werkwoorden ten minste 10% van alle aangeboden toetsopgaven van de verplichte (sub)domeinen lezen en taalverzorging bij het terrein Nederlandse taal?	ja/nee/n.v.t.
TV3	Is het percentage toetsopgaven van het subdomein Taalverzorging met het aspect spelling niet-werkwoorden ten minste 10% van alle aangeboden toetsopgaven van de verplichte (sub)domeinen lezen en taalverzorging bij het terrein Nederlandse taal?	ja/nee/n.v.t.
TV4	Is het percentage aangeboden toetsopgaven van het subdomein Taalverzorging met het aspect interpunctie ten minste 6% van alle aangeboden toetsopgaven van de verplichte (sub)domeinen lezen en taalverzorging bij het terrein Nederlandse taal?	ja/nee/n.v.t.
TV5	Verantwoordt de toetsaanbieder – in het geval van een adaptieve toets waarbij het aantal aangeboden toetsopgaven per leerling verschilt (CAT) – adequaat hoe deze streeft voor elke leerling de percentages uit de kwaliteitscriteria TV1 t/m TV4 te bewerkstelligen?	ja/nee/n.v.t.
TV6	Worden de leerlingen getoetst op verschillende categorieën van spellingsproblemen en -regels?	ja/nee
TV6.a	Worden de leerlingen getoetst op regels voor lettergreepgrenzen?	ja/nee
TV6.b	Worden de leerlingen getoetst op morfologische spelling?	ja/nee
TV6.c	Worden de leerlingen getoetst op regels voor werkwoordspelling?	ja/nee
TV6.d	Worden de leerlingen getoetst op overige regels?	ja/nee
TV6.e	Worden de leerlingen getoetst op interpunctie (hoofdletters en leestekens)?	ja/nee
TV7	Verantwoordt de toetsaanbieder (o.a. met het opleveren van enkele (voorbeeld)toetspaden) op basis van welke beslisregel (het algoritme in het geval van een CAT en het module-design in het geval van een MST) wordt gegarandeerd dat de vereiste verdeling tussen verschillende categorieën van spellingsproblemen en -regels wordt gehaald?	ja/nee/n.v.t.

Toelichting TV1 t/m TV4: Deze kwaliteitscriteria zijn van toepassing voor lineaire toetsen en adaptieve toetsen waarvan het aantal aangeboden toetsopgaven niet per leerling verschilt (MST). De toetsaanbieder toont aan hoe de percentages uit de kwaliteitscriteria TV1 t/m TV4 in een lineaire toets of in het geval van een MST per toetspad worden gegarandeerd.

Toelichting TV5: Dit kwaliteitscriterium is van toepassing op adaptieve toetsen waarvan het aantal aangeboden toetsopgaven per leerling verschilt (CAT). Voor deze adaptieve toetsen zijn de genoemde percentages uit kwaliteitscriteria TV1 t/m TV4 niet te garanderen voor elke leerling. De toetsaanbieder verantwoordt daarom adequaat hoe deze toch streeft voor elke leerling de percentages uit de kwaliteitscriteria TV1 t/m TV4 te bewerkstelligen. En ook uitgebreid in te gaan en te reflecteren op mogelijk situaties – in toetspaden – waarin deze percentages niet aan een leerling worden aangeboden.

Toelichting TV7: Alleen van toepassing bij een digitale adaptieve toets (CAT of MST). Naast het algoritme of module-design moet de werking van de beslisregel worden aangetoond door middel van het opleveren van enkele (voorbeeld)toetspaden, zoals die door de beslisregel worden gegenereerd.

3.3.3. Optioneel domein: Schrijven

Kwaliteitseisen schrijven via indirecte meting

Beslisregel schrijven indirecte meting:

Om aan deze categorie kwaliteitseisen te voldoen, gelden de volgende beslisregels:

1. Taken: van S1.1 t/m S1.3 moeten tenminste twee van de drie vragen met JA worden beantwoord.
2. Kenmerken van de taakuitvoering: de vragen bij S2 moeten met JA worden beantwoord.

Taken:

Code	Vragen	Mogelijke antwoorden
S1.1	Lezen de leerlingen tekstsoorten uit de schrijftakencategorie: Correspondentie (om te reviseren of te beoordelen)?	ja/nee
S1.2	Lezen de leerlingen tekstsoorten uit de schrijftakencategorie: Formulieren, berichten, advertenties, aantekeningen (om te reviseren of te beoordelen)?	ja/nee
S1.3	Lezen de leerlingen tekstsoorten uit de schrijftakencategorie: Verslagen, werkstukken, samenvattingen, artikelen (om te reviseren of te beoordelen)?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
S2.1	Maken leerlingen opgaven waarin ze teksten reviseren of beoordelen op het aspect: samenhang?	ja/nee
S2.2	Maken leerlingen opgaven waarin ze teksten reviseren of beoordelen op het aspect: afstemming op doel?	ja/nee
S2.3	Maken leerlingen opgaven waarin ze teksten reviseren of beoordelen op het aspect: afstemming op publiek?	ja/nee
S2.4	Maken leerlingen opgaven waarin ze teksten reviseren of beoordelen op het aspect: woordgebruik en woordenschat?	ja/nee

Toelichting S1 en S2: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010) leidend.

Kwaliteitseisen schrijven directe meting

Beslisregel schrijven directe meting:

Om aan deze categorie kwaliteitseisen te voldoen, gelden de volgende beslisregels:

1. Taken: De vragen bij S3.1 t/m S3.3 moeten alle met JA worden beantwoord (minimaal één taak per categorie).
2. Kenmerken van de taakuitvoering: de vragen bij S4 moeten met JA worden beantwoord.

Taken:

Code	Vragen	Mogelijke antwoorden
S3.1	Schrijven de leerlingen een tekst uit de schrijftakencategorie: Correspondentie?	ja/nee
S3.2	Schrijven de leerlingen een tekst uit de schrijftakencategorie: Formulieren, berichten, advertenties en/of aantekeningen?	ja/nee
S3.3	Schrijven de leerlingen een tekst uit de schrijftakencategorie: Verslagen, werkstukken, samenvattingen en/of artikelen?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
S4.1	Worden de teksten die de leerlingen schrijven beoordeeld op het aspect: samenhang?	ja/nee
S4.2	Worden de teksten die de leerlingen schrijven beoordeeld op het aspect: afstemming op doel?	ja/nee
S4.3	Worden de teksten die de leerlingen schrijven beoordeeld op het aspect: afstemming op publiek?	ja/nee



Code	Vragen	Mogelijke antwoorden
S4.4	Worden de teksten die de leerlingen schrijven beoordeeld op het aspect: woordgebruik en woordenschat?	ja/nee

Toelichting S3.1 t/m S3.3: Minimaal één taak per categorie in de doorstroomtoets moet worden opgenomen.

Toelichting S3 en S4: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010) leidend.

3.3.4. Optioneel domein: Mondelinge taalvaardigheid

Kwaliteitseisen mondelinge taalvaardigheid via indirecte meting

A. Subdomein Gesprekken

Kwaliteitseisen

Beslisregel mondelinge taalvaardigheid indirecte meting, subdomein Gesprekken:

Om aan deze categorie kwaliteitseisen te voldoen, gelden de volgende beslisregels:

1. Taken: van M1.1 en M1.2 moet één van de twee vragen met JA worden beantwoord (taak is minimaal één keer aanwezig).
2. Kenmerken van de taakuitvoering: de vragen bij M2 moeten met JA worden beantwoord.

Taken:

Code	Vragen	Mogelijke antwoorden
M1.1	Luisteren en beoordelen de leerlingen gesprekken uit de takencategorie: discussie en overleg?	ja/nee
M1.2	Luisteren en beoordelen de leerlingen gesprekken uit de takencategorie: informatie uitwisselen?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
M2.1	Maken leerlingen opgaven waarin ze gesprekken beoordelen op het aspect: beurten nemen en bijdragen aan samenhang?	ja/nee
M2.2	Maken leerlingen opgaven waarin ze gesprekken beoordelen op het aspect: afstemming op doel?	ja/nee
M2.3	Maken leerlingen opgaven waarin ze gesprekken beoordelen op het aspect: afstemming op de gesprekspartners?	ja/nee
M2.4	Maken leerlingen opgaven waarin ze gesprekken beoordelen op het aspect: woordgebruik en woordenschat?	ja/nee

Toelichting M1.1 en M1.2: Tenminste één van de twee taken moet in de doorstroomtoets worden opgenomen.

Toelichting M1 en m²: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010) leidend.

B. Subdomein Spreken

Kwaliteitseisen

Beslisregel mondelinge taalvaardigheid indirecte meting, subdomein Spreken:

Om aan deze categorie kwaliteitseisen te voldoen, gelden de volgende beslisregels:

1. Taken: vraag M3 moet met JA te worden beantwoord.
2. Kenmerken van de taakuitvoering: de vragen bij M4 moeten met JA worden beantwoord.

Taken:

Code	Vragen	Mogelijke antwoorden
M3	Luisteren en beoordelen de leerlingen een monoloog?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
M4.1	Maken leerlingen opgaven waarin ze een monoloog beoordelen op het aspect: samenhang?	ja/nee
M4.2	Maken leerlingen opgaven waarin ze een monoloog beoordelen op het aspect: afstemming op doel?	ja/nee
M4.3	Maken leerlingen opgaven waarin ze een monoloog beoordelen op het aspect: afstemming op publiek?	ja/nee
M4.4	Maken leerlingen opgaven waarin ze een monoloog beoordelen op het aspect: woordgebruik en woordenschat?	ja/nee

Toelichting M3 en M4: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010) leidend.

Kwaliteitseisen mondelinge taalvaardigheid via directe meting

A. Subdomein Gesprekken

Kwaliteitseisen

Beslisregel mondelinge taalvaardigheid directe meting, subdomein Gesprekken:

Om aan deze categorie kwaliteitseisen te voldoen, gelden de volgende beslisregels:

1. Taken: van M5.1 en M5.2 moet één van de twee vragen met JA worden beantwoord (taak is minimaal één keer aanwezig).
2. Kenmerken van de taakuitvoering: de vragen bij M6 moeten met JA worden beantwoord.

Taken:

Code	Vragen	Mogelijke antwoorden
M5.1	Nemen de leerlingen deel aan gesprekken uit de takencategorie: discussie en overleg?	ja/nee
M5.2	Nemen de leerlingen deel aan gesprekken uit de takencategorie: informatie uitwisselen?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
M6.1	Worden de gesprekken die leerlingen voeren beoordeeld op het aspect: beurten nemen en bijdragen aan samenhang?	ja/nee
M6.2	Worden de gesprekken die leerlingen voeren beoordeeld op het aspect: afstemming op doel?	ja/nee
M6.3	Worden de gesprekken die leerlingen voeren beoordeeld op het aspect: afstemming op de gesprekspartners?	ja/nee
M6.4	Worden de gesprekken die leerlingen voeren beoordeeld op het aspect: woordgebruik en woordenschat?	ja/nee

Toelichting M5.1 en M5.2: Tenminste één van de twee taken moet in de doorstroomtoets zijn opgenomen.

Toelichting M5 en M6: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010) leidend.

B. Subdomein Luisteren

Kwaliteitseisen

Beslisregel mondelinge taalvaardigheid directe meting, subdomein Luisteren:

Om aan deze categorie kwaliteitseisen te voldoen, gelden de volgende beslisregels:

1. Taken: van M7.1 t/m M7.3 moeten twee van de drie vragen met JA worden beantwoord (taak is minimaal één keer aanwezig).
2. Kenmerken van de taakuitvoering: de vragen bij M8 moeten met JA worden beantwoord.

Taken:

Code	Vragen	Mogelijke antwoorden
M7.1	Luisteren de leerlingen naar instructies?	ja/nee
M7.2	Luisteren de leerlingen als lid van een live publiek?	ja/nee
M7.3	Luisteren de leerlingen naar radio en/of televisie en/of naar gesproken tekst op internet en/of naar overige door de aanbieder gehanteerde communicatiemiddelen?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
M8.1	Wordt tijdens het luisteren getoetst of leerlingen de luistertekst kunnen begrijpen?	ja/nee
M8.2	Wordt tijdens het luisteren getoetst of leerlingen de luistertekst kunnen interpreteren?	ja/nee
M8.3	Wordt tijdens het luisteren getoetst of leerlingen de luistertekst kunnen samenvatten?	ja/nee

Toelichting M7.1 t/m M7.3: Twee van de drie taken moeten in de doorstroomtoets worden opgenomen.

Toelichting bij M7 en M8: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010) leidend.

C. Subdomein Spreken

Kwaliteitseisen

Beslisregel mondelinge taalvaardigheid directe meting, subdomein Spreken:

Om aan deze categorie kwaliteitseisen te voldoen, gelden de volgende beslisregels:

1. Taken: vraag M9 moet met JA worden beantwoord.
2. Kenmerken van de taakuitvoering: de vragen bij M10 moeten met JA worden beantwoord.

Taken:

Code	Vragen	Mogelijke antwoorden
M9	Houden de leerlingen een monoloog?	ja/nee

Kenmerken van de taakuitvoering:

Code	Vragen	Mogelijke antwoorden
M10.1	Wordt de monoloog die de leerlingen houden beoordeeld op het aspect: samenhang?	ja/nee
M10.2	Wordt de monoloog die de leerlingen houden beoordeeld op het aspect: afstemming op doel?	ja/nee
M10.3	Wordt de monoloog die de leerlingen houden beoordeeld op het aspect: afstemming op publiek?	ja/nee
M10.4	Wordt de monoloog die de leerlingen houden beoordeeld op het aspect: woordgebruik en woordenschat?	ja/nee

Toelichting M9 en M10: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010) leidend.

3.3.5. Optioneel subdomein: Begrippenlijst

Kwaliteitseis

Beslisregel Begrippenlijst:

Om aan deze categorie kwaliteitseis te voldoen, moet vraag B1 met JA worden beantwoord.

Code	Vragen	Mogelijke antwoorden
B1	Worden de leerlingen getoetst met opgaven die betrekking hebben op de begrippenlijst?	ja/nee

Toelichting B1: Hierbij zijn de beschrijvingen op niveau 1F en 2F in het Referentiekader (2010, p. 17–19) leidend.

3.4. Terrein Rekenen

Het Referentiekader onderscheidt voor het terrein Rekenen vier wettelijk verplichte domeinen, te weten Getallen (g), Verhoudingen (vh), Meten en meetkunde (m/mk) en Verbanden (vb). In de doorstroomtoets moeten alle domeinen getoetst worden. Dit betekent dat alle doorstroomtoetsen moeten voldoen aan de gestelde kwaliteitseisen aan het terrein Rekenen. De gestelde vakinhoudelijke kwaliteitseisen voor deze vier domeinen en de vereiste verdeling van de toetsopgaven over de domeinen in de toetsmatrijs en de toetssamenstelling zijn opgenomen in de Toetswijzer PO.

Een individuele rekenopgave kan betrekking hebben op meerdere domeinen van het terrein Rekenen. In dat geval moet de aanbieder dit duidelijk vermelden, bijvoorbeeld door te kiezen voor het domein dat het beste past bij het beoogde toetsdoel van de betreffende opgave. Tevens moet de aanbieder dan toelichten op welke wijze er wordt voldaan aan de voorgeschreven verdeling van toetsopgaven over de domeinen van het terrein Rekenen.

3.4.1. Verdeling toetsopgaven over domeinen in toetsmatrijs en toetssamenstelling

De doorstroomtoets bevat opgaven uit alle domeinen uit het referentiekader Rekenen. Het betreft de domeinen Getallen (g), Verhoudingen (vh), Meten en meetkunde (m/mk) en Verbanden (vb).

Kwaliteitseisen

Beslisregel: Om aan deze categorie kwaliteitseisen te voldoen, moeten alle vragen met JA of N.V.T. worden beantwoord.

De opgaven moeten in de toets als volgt over de domeinen zijn verdeeld:

Code	Vragen	Mogelijke antwoorden
D.G	Ligt het percentage aangeboden toetsopgaven uit het domein Getallen tussen de 30% en 40% van het totaal aantal aangeboden toetsopgaven van het terrein Rekenen?	ja/nee
D.Vh	Ligt het percentage aangeboden toetsopgaven uit het domein Verhoudingen tussen de 20% en 30% van het totaal aantal aangeboden toetsopgaven van het terrein Rekenen?	ja/nee
D.M/mk	Ligt het percentage aangeboden toetsopgaven uit het domein Meten & meetkunde tussen de 20% en 30% van het totaal aantal aangeboden toetsopgaven van het terrein Rekenen?	ja/nee
D.Vb	Ligt het percentage aangeboden toetsopgaven uit het domein Verbanden tussen de 15% en 20% van het totaal aantal aangeboden toetsopgaven van het terrein Rekenen?	ja/nee
D.1	Verantwoordt de toetsaanbieder (o.a. met het opleveren van enkele (voorbeeld)toetspaden) op basis van welke beslisregel (het algoritme in het geval van een CAT en het module-design in het geval van een MST) de vereiste verdeling van toetsopgaven over de domeinen wordt gehaald?	ja/nee/n.v.t.

Toelichting D.G t/m D.Vb: De gestelde percentages toetsopgaven per domein gelden voor zowel papieren toetsen, lineaire digitale toetsen, adaptieve toetsen op itemniveau (CAT) en adaptieve toetsen op moduleniveau (MST).

Toelichting D.1: Alleen van toepassing bij een digitale adaptieve toets (CAT of MST). Naast het algoritme of module-design moet de werking van de beslisregel worden aangetoond door middel van het opleveren van enkele (voorbeeld)toetspaden, zoals die door de beslisregel worden gegenereerd.

3.4.2. Verdeling toetsopgaven over onderdelen

Elk domein is opgebouwd uit drie onderdelen (zie hiervoor ook de Toetswijzer PO):

- Notatie, taal en betekenis, waarbij het gaat om de uitspraak, schrijfwijze en betekenis van getallen, symbolen en relaties en om het gebruik van wiskundetaal.
- Met elkaar in verband brengen, waarbij het gaat om het verband tussen begrippen, notaties, getallen en dagelijks spraakgebruik.
- Gebruiken, waarbij het er om gaat rekenkundige vaardigheden in te zetten bij het oplossen van problemen.

De toets bevat opgaven uit alle onderdelen.

Kwaliteitseisen

Beslisregel: Om aan deze categorie kwaliteitseisen te voldoen, moeten alle vragen met JA of N.V.T. worden beantwoord.

De relatieve (procentuele) verdeling van opgaven over de onderscheiden onderdelen is als volgt:

Code	Vragen	Mogelijke antwoorden
O.A	Is iedere toetsopgave, met uitzondering van onderdeel A, onder één onderdeel geplaatst?	ja/nee
O.B	Is het percentage aangeboden toetsopgaven uit onderdeel B ten minste 20 procent van het totaal aantal aangeboden toetsopgaven van het terrein Rekenen?	ja/nee
O.C	Is het percentage aangeboden toetsopgaven uit onderdeel C ten minste 30 procent van het totaal aantal aangeboden toetsopgaven van het terrein Rekenen?	ja/nee
O.D	Verantwoordt de toetsaanbieder (o.a. met het opleveren van enkele (voorbeeld)toetspaden) op basis van welke beslisregel (het algoritme in het geval van een CAT en het module-design in het geval van een MST) de vereiste verdeling van toetsopgaven over de onderdelen wordt gehaald?	ja/nee/n.v.t.

Toelichting O.A: Voor onderdeel A wordt geen minimum vereist omdat deze inhouden ook kunnen worden gebruikt en toegepast bij de inhouden van onderdeel B en onderdeel C.

Toelichting O.A t/m O.C: De gestelde percentages toetsopgaven per domein gelden voor zowel papieren toetsen, lineaire digitale toetsen, computergestuurde adaptieve toetsen op itemniveau (CAT) en adaptieve toetsen op moduleniveau (MST).

Toelichting O.D: Alleen van toepassing bij een digitale adaptieve toets (CAT of MST). Naast het algoritme of module-design moet de werking van de beslisregel worden aangetoond door middel van het opleveren van enkele (voorbeeld)toetspaden die voldoen aan de vereiste verdeling van toetsopgaven over de onderdelen, zoals die door de beslisregel worden gegenereerd.

3.4.3. Opgaven met en zonder context

In het terrein Rekenen van een doorstroomtoets moeten zowel opgaven met context als opgaven zonder context (zogenoemde 'kale opgaven') worden opgenomen.

Kwaliteitseisen

Beslisregel: Om aan deze categorie kwaliteitseisen te voldoen, moeten alle vragen met JA of N.V.T. worden beantwoord.

Code	Vragen	Mogelijke antwoorden
C1	Is minstens 20% en maximaal 35% van de aangeboden toetsopgaven van het terrein Rekenen een kale opgave?	ja/nee
C2	Is minstens 65% en maximaal 80% van de aangeboden toetsopgaven van het terrein Rekenen een contextopgave?	ja/nee
C3	Verantwoordt de toetsaanbieder (o.a. met het opleveren van enkele (voorbeeld)toetspaden) op basis van welke beslisregel (het algoritme in het geval van een CAT en het module-design in het geval van een MST) de vereiste procentuele verdelingen zoals in C1 t/m C2 geschetst wordt gehaald?	ja/nee/n.v.t.

Toelichting C1 t/m C2: De gestelde percentages toetsopgaven gelden voor zowel papieren toetsen, lineaire digitale toetsen, computergestuurde adaptieve toetsen op itemniveau (CAT) en adaptieve toetsen op moduleniveau (MST).

Toelichting C3: Alleen van toepassing bij een digitale adaptieve toets (CAT of MST). Naast het algoritme of module-design moet de werking van de beslisregel worden aangetoond door middel van het opleveren van enkele (voorbeeld)toetspaden met de vereiste procentuele verdelingen zoals in C1 t/m C2 geschetst, zoals die door de beslisregel worden gegenereerd.

3.4.4. Gebruik uitrekenpapier

Kwaliteitseisen

Beslisregel: Om aan deze categorie kwaliteitseisen te voldoen, moeten alle vragen met JA worden beantwoord.

Code	Vragen	Mogelijke antwoorden
K1	Mogen de leerlingen bij het beantwoorden van alle (digitale) toetsopgaven fysiek en/of niet programmeerbaar uitrekenpapier gebruiken?	ja/nee
K2	Wordt in de handleiding voor leraren en/of het toetsreglement aangegeven dat zij leerlingen nadrukkelijk moeten wijzen op het toegestane gebruik van fysiek en/of niet programmeerbaar uitrekenpapier bij het beantwoorden van alle (digitale) toetsopgaven?	ja/nee

3.4.5. Gebruik rekenmachine

Kwaliteitseis

Beslisregel: Om aan deze categorie kwaliteitseis te voldoen, moet deze vraag met JA worden beantwoord.

Code	Vragen	Mogelijke antwoorden
R1	Mogen de leerlingen bij maximaal 20% van de (digitale) toetsopgaven een niet-grafische en/of niet-programmeerbare rekenmachine gebruiken?	ja/nee

Toelichting: Gebruik van een niet-grafische en/of niet-programmeerbare rekenmachine is bij maximaal 20% van de toetsopgaven toegestaan. Gebruik van een grafische en/of programmeerbare rekenmachine is niet toegestaan.

3.5. Toetsopgaven Nederlandse taal en Rekenen op niveau 1F en 2F/1S

In het document *Referentiekader taal en Rekenen* (SLO, 2011) is vastgelegd wat leerlingen eind groep 8 van de basisschool moeten kennen en kunnen als het gaat om basisvaardigheden voor Nederlandse taal en Rekenen. Deze kennis en vaardigheden worden in het referentiekader gespecificeerd in een aantal referentieniveaus. De verschillende niveaus worden beschreven in twee kwaliteiten: fundamentele kwaliteit (F) en streefkwaliteit (S). De niveaus zijn cumulatief. Een leerling moet op een hoger niveau alle vaardigheden beheersen die op een lager niveau genoemd worden. Deze worden niet telkens herhaald. De beschrijvingen van de niveaus voor Nederlandse taal en Rekenen verschillen. Dit komt doordat de vakken verschillen.

Voor Nederlandse taal zijn vier niveaus beschreven: 1F, 2F, 3F en 4F en voor Rekenen zijn zes niveaus beschreven: 1F, 2F, 3F en 1S, 2S en 3S. Voor het po zijn voor het terrein Nederlandse taal de niveaus 1F en 2F wettelijk vastgelegd. Voor het terrein Rekenen zijn dit voor het po de niveaus 1F en 1S.

Dit betekent dat, in lijn met de Toetswijzer PO, de doorstroomtoets, ongeacht de toetsvorm, voor de terreinen Nederlandse taal en Rekenen toetsopgaven moet bevatten die de inhoud van de niveaus 1F (fundamentele kwaliteit) en 2F/1S (streefkwaliteit) meten.

Kwaliteitseis

Beslisregel: Om aan deze categorie kwaliteitseis te voldoen, moet vraag TN1 met JA worden beantwoord.

Code	Vragen	Mogelijke antwoorden
TN1	Zijn de opgaven van de toets voor de terreinen Nederlandse taal en Rekenen evenwichtig samengesteld qua referentieniveaus 1F en 2F/1S?	ja/nee

Toelichting TN1: Dit aspect moet uit de toetsmatrijs naar voren komen. Dat wil zeggen dat in de toetsmatrijs de verhouding 1F en 2F/1S vragen is gespecificeerd en dat in zowel de toetsmatrijs als in de varianten van de doorstroomtoets de verhouding 1F en 2F/1S vragen evenwichtig is. Evenwichtig wil zeggen dat het aantal en de verhouding 1F en 2F/1S vragen gezamenlijk inhoudelijk de referentieniveaus 1F en 2F/1S van de terreinen Nederlandse taal en Rekenen representeren, een en ander conform de voorschriften uit de Toetswijzer PO.

Een CAT moet voor elke individuele leerling voor de terreinen Nederlandse taal en Rekenen een uitspraak over het behaalde niveau 1F en/of 2F/1S geven. In het geval van een CAT is er verantwoord hoe het 1F en/of 2F/1S niveau van de terreinen Nederlandse taal en Rekenen op representatieve wijze

voor alle individuele leerlingen aan bod komt. De aanbieder kan dit verantwoorden aan de hand van de toetsmatrijs, aangevuld met de procesbeschrijving van hoe het algoritme werkt, aangevuld met enkele mogelijke itemsets inclusief de daarin aangeboden toetsopgaven en inclusief een toelichting op het daarbij geldende stopcriterium.

Elke routing binnen het MST design moet voor de terreinen Nederlandse taal en Rekenen een uitspraak over het behaalde niveau 1F en/of 2F/1S geven. In het geval van een MST is er verantwoord hoe het 1F en/of 2F/1S niveau van de terreinen Nederlandse taal en Rekenen op een bij de moeilijkheidsgraad van de betreffende route passende wijze voor alle individuele leerlingen aan bod komt. De aanbieder kan dit verantwoorden aan de hand van de toetsmatrijs, aangevuld met de beschrijving van de samenstelling van het MST design, aangevuld met enkele toetspaden uit het design inclusief de daarin aangeboden toetsopgaven.

4. Organisatorische aspecten

4.1. Leerlingrapport

Kwaliteitseisen

Beslisregel: Om aan deze categorie kwaliteitseisen te voldoen, moeten alle vragen met JA of N.V.T. worden beantwoord.

Code	Vragen	Mogelijke antwoorden
LR1	Stelt de toetsaanbieder voor elke individuele leerling die de doorstroomtoets aflegt een leerlingrapport op?	ja/nee
LR2	Bevat het leerlingrapport het toetsadvies, uitgedrukt in één van de toetsadviescategorieën?	ja/nee
LR3	Geeft het leerlingrapport aan welk referentieniveau Lezen de leerling beheerst?	ja/nee
LR4	Geeft het leerlingrapport aan welk referentieniveau Taalverzorging de leerling beheerst?	ja/nee
LR5	Geeft het leerlingrapport aan welk referentieniveau Rekenen de leerling beheerst?	ja/nee
LR6	Bevat het leerlingrapport een deugdelijke toelichting bij het toetsadvies en de beheersing van de referentieniveaus van de leerling die (ook) voor leerlingen, ouders, verzorgers, voogden en afnemers begrijpelijk is?	ja/nee
LR7	Zijn de andere uit te geven rapporten naast het leerlingrapport, en/of niet verplichte gegevens op het leerlingrapport, deugdelijk en helder voor de beoogde gebruiker?	ja/nee/n.v.t.

Toelichting LR1: De toetsaanbieder geeft één of meerdere (model)rapportages ter beoordeling. Wanneer er meer dan één variant van het leerlingrapport bestaat, moeten al deze varianten aangeboden worden, met een uitleg wanneer welke variant wordt gebruikt. Aangezien alle varianten aan de eisen moeten voldoen, worden alle voorbeeldrapporten beoordeeld op de kwaliteitseisen LR2 tot en met LR6. Mochten er rapporten ontbreken (d.w.z. LR1 = nee), dan krijgt de toetsaanbieder een onvoldoende voor 'leerlingrapport'. Wanneer het leerlingrapport wel wordt aangeboden, worden de kwaliteitseisen LR2 tot en met LR6 beoordeeld op basis van die versie(s) van het leerlingrapport.

Toelichting LR2: De toetsaanbieder moet in het leerlingrapport een toetsadvies geven. De aanbieder hanteert hierbij de door de Rijksoverheid vastgestelde toetsadviescategorieën. Dit toetsadvies is gebaseerd op in ieder geval de resultaten van het wettelijk verplichte domein Lezen en subdomein Taalverzorging van het terrein Nederlandse taal en het terrein Rekenen. Aanvullend mag de toetsaanbieder in de bepaling van het toetsadvies ook één of meerdere van de optionele (sub)domeinen van Nederlandse taal meenemen die in dit kader in Hoofdstuk 3 beschreven zijn. In de wetenschappelijke verantwoording moet zijn aangegeven op basis van metingen van welke (sub)domeinen het toetsadvies tot stand is gekomen.

Toelichting LR3: In het leerlingrapport moet aangegeven zijn op welk referentieniveau de leerling gepresteerd heeft op het domein Lezen. In het leerlingrapport moet in ieder geval aangegeven worden in welk van de volgende drie categorieën de leerlingprestatie past: onder niveau 1F, op niveau 1F, of op streefniveau 2F. Hoewel in de theorie de categorieën elkaar niet uitsluiten, en de leerling ook op een hoger niveau zou kunnen presteren dan het maximaal gegeven niveau, is in deze regeling vastgesteld dat in de rapportage een leerlingprestatie in één (niet meer en niet minder) van deze drie categorieën geclassificeerd wordt.

Toelichting LR4: In het leerlingrapport moet aangegeven zijn op welk referentieniveau de leerling gepresteerd heeft op het subdomein Taalverzorging. In het leerlingrapport moet in ieder geval aangegeven worden in welk van de volgende drie categorieën de leerlingprestatie past: onder niveau 1F, op niveau 1F, of op streefniveau 2F. Hoewel in de theorie de categorieën elkaar niet uitsluiten, en de

leerling ook op een hoger niveau zou kunnen presteren dan het maximaal gegeven niveau, is in deze regeling vastgesteld dat in de rapportage een leerlingprestatie in één (niet meer en niet minder) van deze drie categorieën geclassificeerd wordt.

Toelichting LR5: In het leerlingrapport moet aangegeven zijn op welk referentieniveau de leerling gepresteerd heeft op het terrein Rekenen. In het leerlingrapport moet in ieder geval aangegeven worden in welk van de volgende drie categorieën de leerlingprestatie past: onder niveau 1F, op niveau 1F, of op streefniveau 1S. Hoewel in de theorie de categorieën elkaar niet uitsluiten, en de leerling ook op een hoger niveau zou kunnen presteren dan het maximaal gegeven niveau, is in deze regeling vastgesteld dat in de rapportage een leerlingprestatie in één (niet meer en niet minder) van deze drie categorieën geclassificeerd wordt.

Toelichting LR6: Bij het rapport moet een toegankelijke toelichting worden geboden, waarin begrijpelijke handvatten gegeven worden voor de interpretatie van het toetsadvies en de beheersing van de referentieniveaus van de leerling door diverse betrokkenen. De aanbieder toont aan dat de rapportage begrijpelijk is voor verschillende doelgroepen, te weten leerlingen, ouders, verzorgers, voogden en afnemers.

Toelichting LR7: Deze kwaliteitseis is niet van toepassing als er naast het leerlingrapport geen andere rapporten worden uitgegeven en/of het leerlingrapport geen aanvullende, niet verplichte, gegevens vermeldt. De kwaliteitseis is van toepassing als de toetsaanbieder één of meerdere (model)rapportages zoals bijvoorbeeld een groepsrapportage, schoolrapportage of bovenschoolse rapportage uitgeeft. Bij de uitgegeven rapporten moet een toegankelijke toelichting worden geboden aan de beoogde gebruiker, waarin begrijpelijke handvatten gegeven worden voor de interpretatie van gerapporteerde gegevens. De aanbieder toont aan dat de rapportage begrijpelijk is voor verschillende doelgroepen, te weten leerkrachten, intern begeleiders, en directeurs of vertegenwoordigers van schoolbesturen of samenwerkingsverbanden.

4.2 Leerlingen

De doorstroomtoets is in principe geschikt voor alle leerlingen, ook voor leerlingen met een specifieke ondersteuningsbehoefte, met uitzondering van leerlingen die onder de ontheffingsgronden vallen.

Kwaliteitseisen

Beslisregel: Om aan deze categorie kwaliteitseisen te voldoen, moeten alle vragen met JA worden beantwoord. In de toelichting wordt bij enkele kwaliteitseisen aangegeven waar een doorverwijzing naar www.Rijksoverheid.nl voor een toets van een aanbieder die voorziet in die behoefte voldoende kan zijn.

Code	Vragen	Mogelijke antwoorden
LL1	Kunnen alle leerlingen, ook met specifieke ondersteuningsbehoeften (uitgezonderd leerlingen die vallen onder de ontheffingsgronden), deelnemen aan de toets?	ja/nee
LL1.a	Is de toets aantoonbaar geschikt, al dan niet door extra ondersteuning, voor leerlingen met een visuele beperking, zoals leerlingen die slechtziend, blind of volledig kleurenblind zijn?	ja/nee
LL1.b	Is de toets aantoonbaar geschikt, al dan niet door extra ondersteuning, voor dove/slechthorende leerlingen?	ja/nee
LL1.c	Is de toets aantoonbaar geschikt, al dan niet door extra ondersteuning, voor leerlingen met een taalontwikkelingsstoornis (TOS)?	ja/nee
LL1.d	Is de toets aantoonbaar geschikt, al dan niet door extra ondersteuning, voor leerlingen met dyslexie?	ja/nee
LL1.e	Is de toets aantoonbaar geschikt, al dan niet door extra ondersteuning, voor leerlingen met ernstige rekenproblemen of dyscalculie?	ja/nee
LL1.f	Is de toets aantoonbaar geschikt, al dan niet door extra ondersteuning, voor leerlingen met een motorische beperking?	ja/nee/n.v.t.
LL1.g	Is de toets aantoonbaar geschikt, al dan niet door extra ondersteuning, voor leerlingen met ondersteuningsbehoeften voortkomend uit chronische/langdurige ziekte, een psychische aandoening en/of gedragsproblemen of voortkomend uit hier niet nader gespecificeerde beperkingen?	ja/nee/n.v.t.
LL2	Staat in de handleiding en/of toetsreglement voor de afnemer bij de toets in volledigheid beschreven op welke wijze de toets rekening houdt met leerlingen met ondersteuningsbehoeften?	ja/nee
LL2.a	Staat in de handleiding en/of toetsreglement omschreven welke hulpmiddelen tijdens de toetsafname zijn toegestaan voor leerlingen met specifieke ondersteuningsbehoeften?	ja/nee
LL2.b	Staat omschreven in de handleiding en/of toetsreglement welke aanpassingen in de afnamecondities mogelijk zijn?	ja/nee

Code	Vragen	Mogelijke antwoorden
LL2.c	Is de toets plaatsonafhankelijk af te nemen?	ja/nee
LL2.d	Wordt beschreven wat de mogelijkheden zijn als de geboden ondersteuning niet toereikend is?	ja/nee

Toelichting LL1: Er is automatisch voldaan aan kwaliteitseis LL1 wanneer is voldaan aan de kwaliteitseisen LL1a t/m LL1g.

Dit houdt het volgende in:

- De aanbieder neemt bij het maken van de toets zoveel mogelijk elementen op die zorgen dat de toets voor alle leerlingen (digitaal) toegankelijk is. Waar mogelijk worden de richtlijnen uit de WCAG 2.1 toegepast. Bij de itemconstructie wordt generiek minstens op de volgende wijze rekening gehouden met (digitale) toegankelijkheid: richtlijnen voor kleurgebruik en contrast bij items, richtlijnen met betrekking tot een ondubbelzinnige lay-out en tekstuele vormgeving, richtlijnen met betrekking tot figuurlijk taalgebruik (als dat geen toetsdoel is) en richtlijnen met betrekking tot het gebruik van ik, gebiedende wijs en vraagzinnen.
- De toets kent geen tijdslimiet en het is mogelijk om pauzes te nemen.
- De toets voorziet in een voorleesfunctie. De wijze waarop kan per toets(soort) verschillen.

Waar de bovengenoemde standaard onvoldoende tegemoet komt aan de doelgroepen genoemd onder LL1.a t/m LL1.g is aantoonbaar geprobeerd de belemmeringen die deze leerlingen ervaren bij het maken van de toets weg te nemen door de mogelijkheid tot inzet van hulpmiddelen of ondersteuning, aanpassing van de afnamecondities of door het aanbieden van aangepaste versies.

Toelichting LL1.a: De toets wordt door de aanbieder (op bestelling) geleverd in een format dat zich leent voor de hulpmiddelen die door leerlingen met een visuele beperking worden gebruikt en met inachtneming van de toetseisen inhoudelijk geschikt zijn gemaakt voor de leerlingen met een visuele beperking. De toetsaanbieder specificeert dit per type visuele beperking (kleurenblind, slechtziend, blind).

Leerlingen die volledig kleurenblind zijn.

Bij de constructie van de toets is rekening gehouden met leerlingen die volledig kleurenblind zijn. Dit kan onder andere door kleurgebruik alleen ondersteunend te laten zijn. Daarnaast is er bijvoorbeeld de mogelijkheid tot het aanbieden van een zwart/wit-variant. Relevant voor de normering is dat de aanbieder er rekening mee houdt dat het ontbreken van een kleuren(afbeelding) geen invloed heeft op de moeilijkheid van het item. Dit houdt in dat de inhoud en de structuur van het item duidelijk en begrijpelijk moeten zijn, ongeacht het gebruik van kleur.

Bij twijfel mag de leerling gebruik maken van de volgende procedure: voor kleurenblinde leerlingen kan een opzoekhulp worden ingezet die op verzoek van de leerling de kleur van een door de leerling aangewezen vlakdeel benoemt, of een door de leerling aangewezen kleur aanwijst in toets of hulpmiddel.

(Zeër) slechtziende en blinde leerlingen

Aanbieders houden bij het maken van de toets rekening met (digitale) toegankelijkheid voor leerlingen met een visuele beperking, middels bijvoorbeeld contrast, mogelijkheid tot inzoomen en mogelijkheid tot het gebruik van hulpmiddelen, waaronder audio-ondersteuning en gebruik van een brailleleesregel

Een aanbieder maakt waar mogelijk voor (zeer) slechtziende en blinde leerlingen aanvullende aanpassingen voor de toegankelijkheid van hun toets. Bijvoorbeeld door gebruik te maken van tekenboeken of het aanbieden van een braillevariant. Het uitgangspunt is dat zo min mogelijk wordt afgeweken van de reguliere toetsen. Daar waar gekozen wordt om de toets aan te passen, verantwoordt de aanbieder deze aanpassingen.

Het aanbieden van een brailleversie van de doorstroomtoets is facultatief. Wel moeten toetsaanbieders leerlingen met die behoefte in de verantwoordingsdocumenten en handleiding voor de afnemer doorverwijzen naar www.Rijksoverheid.nl voor een toets van een aanbieder die voorziet in die behoefte.

Afwijkende beoordeling calamiteitentoets: Van de calamiteitentoets moet een versie beschikbaar zijn in braille, in zwart-wit en voorzien zijn van een vorm van audio-ondersteuning om te voldoen aan LL1.a.

Toelichting LL1.b: De aanbieder levert bijvoorbeeld (op bestelling) een aangepaste toets indien de toets gebruikmaakt van functioneel geluid (zoals bijvoorbeeld bij dictee-opgaven). Indien een aanbieder hier niet in kan voorzien, moet deze de school in de verantwoordingsdocumenten en handleiding voor de afnemer verwijzen naar www.Rijksoverheid.nl waar is aangegeven welke toets van welke aanbieder voorziet in die behoefte.

Toelichting LL1.c: De aanbieder geeft aan hoe rekening is gehouden met leerlingen met een taalontwikkelingsstoornis (TOS). Hierbij kan gedacht worden aan de mogelijkheid tot het inzetten van de voorleesfunctie en mogelijke aanpassingen in de afnamecondities, zoals het werken met een afdeklad en het markeren van tekst.

Toelichting LL1.d: De aanbieder geeft aan hoe bij de lay-out en typografie met deze doelgroep rekening wordt gehouden. Denk hierbij bijvoorbeeld aan een schreefloos lettertype met minimale puntgrootte 12, afstand tussen letters, woorden, regels en alinea's, gebruik van witruimte en een eenduidige, rustige lay-out.

Als een leerling door dyslexie moeite heeft met het lezen van teksten op beeldscherm, moet de leerling gebruik kunnen maken van een door de aanbieder geleverde voorleesfunctie. Bij een papieren versie moet de aanbieder (op vraag) een versie leveren die geschikt is voor voorleessoftware. De wijze waarop kan per toets(soort) verschillen.

Andere voorbeelden van hulpmiddelen ter ondersteuning van deze doelgroep zijn de beschikbaarheid van een zoomfunctie en een markeerstift.

Toelichting bij LL1.e: Voor leerlingen met dyscalculie wordt geen aangepaste toets geleverd. In de afnamecondities kan (bijvoorbeeld door in de planning meer pauzes of spreiding van de toetsmomenten op te nemen) rekening worden gehouden met deze ondersteuningsbehoefte. De aanbieder benoemt de eventuele mogelijkheden in diens verantwoordingsdocumenten. Indien de aanbieder het gebruik van een rekenkaart als hulpmiddel toestaat, moet de aanbieder verantwoorden hoe dit hulpmiddel de toetsdoelstellingen beïnvloedt. Daarbij wordt toegelicht welke impact op de toetsresultaten verwacht wordt bij gebruik van de rekenkaart en onder welke voorwaarden deze tijdens de toets gebruikt mag worden.

Toelichting bij LL1.f: De aanbieder beschrijft hoe de toets toegankelijk is gemaakt voor leerlingen met motorische beperkingen. Digitale toetsen moeten afgenomen kunnen worden met alleen het toetsenbord in plaats van met de muis en bij schriftelijke toetsen moet de mogelijkheid geboden worden tot het typen van antwoorden op een antwoordblad. Indien het niet mogelijk is de toets volledig toetsenbordbedienbaar te maken, beargumenteert de aanbieder waarom dit niet mogelijk is. Indien het voor een leerling desondanks niet mogelijk is de fysieke handelingen uit te voeren die nodig zijn om de toets te maken, wordt in de handleiding beschreven dat de leerling dan gebruik kan maken van een hulp ter ondersteuning, zoals een schrijfhulp, die antwoorden noteert op aanwijzing van de leerling.

Afwijkende beoordeling calamiteitentoets: kwaliteitseis LL1.f is niet van toepassing voor de calamiteitentoets.

Toelichting bij LL1.g: De aanbieder toont aan welke ondersteuning wordt geboden voor andere beperkingen zoals hier genoemd. Voor bijvoorbeeld leerlingen met ondersteuningsbehoeften voortkomend uit ziekte, psychische aandoeningen en/of gedragsproblemen geeft de aanbieder aan welke aanpassingen mogelijk zijn in de wijze van afnemen en van welke hulpmiddelen zij gebruik kunnen maken tijdens de afname. De aanbieder benoemt daarbij expliciet welke aanpassingen en hulpmiddelen voor deze doelgroepen extra aandacht behoeven, waarbij gedacht kan worden aan het bieden van voldoende tijd, een prikkelarme werkomgeving of individuele afname, extra structureren van wijze van afname en regels die daarbij gelden.

Afwijkende beoordeling calamiteitentoets: kwaliteitseis LL1.g is niet van toepassing voor de calamiteitentoets.

Toelichting LL2: De vorm van de ondersteuning staat beschreven in de handleiding en/of toetsreglement, waaronder de wijze waarop rekening gehouden wordt met doelgroepen in de itemconstructie, door het aanbod van varianten van de toets, de beschikbaarheid van hulpmiddelen en/of via de afnamecondities.

De aanbieder schrijft voor dat bij alle toets(vorm)en alle onderdelen de hulpmiddelen zijn toegestaan binnen de gegeven kaders van de vigerende regelingen. In de Toetswijzer PO staat van een aantal hulpmiddelen voorgescreven dat deze altijd zijn toegestaan of wanneer al dan niet deze zijn toegestaan. Aanvullend kan een aanbieder andere hulpmiddelen toestaan.

De aanbieder wijst de afnemer er in de handleiding en/of toetsreglement op dat er geen tijdslimiet bestaat en dat de afnemer zelf kan besluiten wat een passende afnameduur (spreiding van de toetsonderdelen over de dag, pauzes e.d.) is bij leerlingen met ondersteuningsbehoeften. Daarnaast wijst de aanbieder de afnemer op de mogelijkheid van de voorleesfunctie voor alle leerlingen die op basis van hun ondersteuningsbehoefte hier gebruik van moeten kunnen maken.

Toelichting LL2.a: Hulpmiddelen voor leerlingen met specifieke ondersteuningsbehoeften kunnen in twee categorieën opgesplitst worden. Enerzijds gaat het om hulpmiddelen die al in de toets(omgeving) zijn opgenomen of ingebouwd. Anderzijds gaat het om de mogelijkheid tot het inzetten van hulpmiddelen van de school of leerling zelf. Bij deze tweede categorie is het belangrijk dat de toets toegankelijk is voor deze hulpmiddelen. De aanbieder beschrijft beide categorieën in de handleiding voor de afnemer en/of toetsreglement.

Toelichting LL2.b: De aanbieder geeft aan welke aanpassingen in de afnamecondities mogelijk zijn en beschrijft hier ten minste de mogelijkheden voor het bieden van voldoende tijd (langer dan de gegeven tijdsindicatie), de inzet van pauzes en de mogelijkheid tot individuele afname. Voor veel aanpassingen in de afnamecondities geldt dat deze voor alle leerlingen ingezet zouden kunnen worden. Voor leerlingen met ondersteuningsbehoeften vragen deze afnamecondities echter extra aandacht. De aanbieder beschrijft de mogelijke aanpassingen in de handleiding voor de afnemer en/of toetsreglement en wijst de afnemer erop dat de afnamecondities voor leerlingen met ondersteuningsbehoeften extra aandacht vragen.

Toelichting LL2.c: Wanneer een leerling aantoonbaar niet in de gelegenheid is om de toets op school te maken, kan de toets op een andere locatie worden afgenomen. Denk hierbij aan leerlingen die in het ziekenhuis liggen, of thuiszittende leerlingen. De aanbieder beschrijft in de handleiding welke (technische) randvoorwaarden hieraan verbonden zijn en hoe de school afname op een andere locatie kan organiseren.

Toelichting LL2.d: De aanbieder toont aan wat er aan maatwerk mogelijk is indien het beschreven aanbod aan hulpmiddelen, toetsvarianten en aanpassingen van afnamecondities niet toereikend zijn voor een leerling met specifieke ondersteuningsbehoeften.

4.3 Afname doorstroomtoets

Kwaliteitseisen

Beslisregel: Om aan deze categorie kwaliteitseisen te voldoen, moeten alle vragen met JA of N.V.T. worden beantwoord.

Code	Vragen	Mogelijke antwoorden
A1	Staat in het toetsreglement/de afnamehandleiding adequaat beschreven hoe een leerling tijdens een tweede afnamemoment (omdat die bij de eerste afname was verhinderd) binnen de wettelijke gestelde termijn de toets alsnog kan maken?	ja/nee/n.v.t.
A2.1	Staat in het toetsreglement/de afnamehandleiding vermeld hoe de directeur de leerlingen voor de toets moet aanmelden?	ja/nee
A2.2	Staat in het toetsreglement/de afnamehandleiding vermeld welke hulpmiddelen leerlingen wel én niet mogen gebruiken?	ja/nee
A2.3	Staat in het toetsreglement/de afnamehandleiding vermeld hoe en wanneer scholen incidenten voor, tijdens en na de afname moeten en kunnen melden bij de toetsaanbieder?	ja/nee
A2.4	Staan in het toetsreglement/de afnamehandleiding de voorwaarden aan de (wijze van) geheimhouding van de toetsopgaven door de deelnemende scholen beschreven?	ja/nee
A2.5	Staan in het toetsreglement/de afnamehandleiding de voorwaarden aan de wijze van waarop toezicht op de leerlingen wordt gehouden beschreven?	ja/nee
A2.6	Staat in het toetsreglement/de afnamehandleiding vermeld op welke wijze de toetsopgaven aan het bevoegd gezag ter beschikking worden gesteld?	ja/nee
A2.7	Staan in het toetsreglement/de afnamehandleiding de eisen beschreven waar de ICT omgeving (computerconfiguratie) op de school aan moet voldoen om de digitale toets af te kunnen nemen?	ja/nee/n.v.t.
A3	Staat in het toetsreglement/de afnamehandleiding vermeld op welke wijze de afnamecondities moeten bijdragen aan optimale prestaties van de leerlingen?	ja/nee

Toelichting A1: In het toetsreglement en/of de afnamehandleiding is beschreven hoe leerlingen, die bij de eerste afname waren verhinderd, alsnog bij een tweede afnamemoment binnen de wettelijk gestelde termijn een doorstroomtoets kunnen maken.

Afwijkende beoordeling calamiteitentoets: kwaliteitseis A1 is niet van toepassing voor de calamiteiten-toets.

Toelichting A2.1 t/m A2.7:

In het toetsreglement en/of de afnamehandleiding is de volgende informatie gegeven:

- A2.1: Hoe moet de schooldirecteur (of andere verantwoordelijke) de leerlingen voor de toets aanmelden?
- A2.2: Welke hulpmiddelen mogen de leerlingen wel én niet gebruiken?
- A2.3: Hoe en wanneer scholen incidenten voor, tijdens en na de afname bij de toetsaanbieder kunnen en moeten melden?
- A2.4: Welke voorwaarden worden er gesteld aan de (wijze van) geheimhouding van de toetsopgaven door de schooldirecteur (of andere verantwoordelijke) en de afnemers?
- A2.5: Welke voorwaarden zijn er verbonden aan de wijze waarop toezicht op de leerlingen moet worden gehouden?
- A2.6: Op welke wijze worden de toetsopgaven aan het bevoegd gezag ter beschikking gesteld?
- A2.7: Wat zijn de eisen waar de ICT omgeving (computerconfiguratie) op de school aan moet voldoen om de digitale toets af te kunnen nemen? Deze kwaliteitseis is niet van toepassing indien de ingediende toets niet digitaal wordt afgenomen.

Bij het evalueren van deze vragen kijkt de adviseur niet alleen of deze informatie er is, maar ook of deze duidelijk is en geen tegenstrijdigheden bevat. Mocht achteraf blijken dat in de praktijk de door de aanbieders verstrekte informatie en communicatie dusdanig verwarrend en/of tegenstrijdig zijn dat dit een negatieve impact heeft op de afname en/of de resultaten van de leerlingen, dan kan dit in de jaarlijkse evaluatie van de toets een negatief oordeel opleveren.

Toelichting A3: Staat in het toetsreglement en/of de afnamehandleiding vermeld op welke wijze de afnamecondities moeten bijdragen aan optimale prestaties van de leerlingen? Denk aan voldoende tijd om de toets te kunnen maken en een goede voorbereiding van leerlingen op de afname van de toets.

4.4 Beveiligingsaspecten

Kwaliteitseisen

Indien de aanbieder ISO 27001 is gecertificeerd, volstaat het om dit certificaat inclusief een verklaring van toepasbaarheid (c.q. de scope van het certificaat) aan te leveren. Indien voorgaande wordt aangeleverd zijn kwaliteitseisen RB1 t/m RB5 alvast voldoende verantwoord.

Beslisregel: Om aan deze categorie kwaliteitseisen te voldoen, moeten alle vragen met JA of N.V.T. worden beantwoord.

Code	Vragen	Mogelijke antwoorden
RB1	Heeft de toetsaanbieder een toetsveiligheidsplan opgesteld met daarin opgenomen een risicoanalyse en een PDCA cyclus voor het beschermen van de beschikbaarheid, integriteit en vertrouwelijkheid van informatie (BIV)?	ja/nee
RB1.1	Zijn de functies en verantwoordelijkheden van de functionarissen van de toetsaanbieder die betrokken zijn bij de inhoud en de beveiliging van de doorstroomtoets beschreven?	ja/nee
RB1.2	Hebben de functionarissen van de toetsaanbieder die betrokken zijn bij de inhoud en de beveiliging van de doorstroomtoets aantoonbaar een geheimhoudingsverklaring ondertekend?	ja/nee
RB1.3	Beschermde de toetsaanbieder het toetsconstructieproces door middel van een PDCA cyclus van beveiligingsprocedures, inclusief aantoonbare maatregelen ter voorkoming van het voortijdig uitlekken van de toetsinhoud?	ja/nee
RB1.4	Is voor het toetsafnameproces een PDCA cyclus van beveiligingsmaatregelen geïmplementeerd, inclusief aantoonbare maatregelen ter voorkoming van het voortijdig uitlekken van de toetsinhoud?	ja/nee
RB1.5	Zijn de fysieke en elektronische locaties waar de toetsaanbieder examens, systemen en materialen bewaart, aantoonbaar voldoende beveiligd?	ja/nee
RB1.6	Is de digitale en fysieke informatie omtrent de examenketen zowel tijdens opslag als verzending aantoonbaar voldoende beveiligd?	ja/nee
RB1.7	Verzorgt de toetsaanbieder periodiek training over beveiligingsbewustheid aan haar functionarissen die betrokken zijn bij de inhoud en de beveiliging van de doorstroomtoets?	ja/nee
RB2	Waarborgt de toetsaanbieder, met behulp van passende technische en organisatorische beveiligingsmaatregelen, de privacy en persoonsgegevens van de afnemer, leerling en ouder conform de Algemene Verordening Gegevensbescherming (AVG)?	ja/nee

Code	Vragen	Mogelijke antwoorden
RB3	Heeft de toetsaanbieder een gedegen en compleet actieplan voor het omgaan met diefstal van (delen van) de toetsinhoud en voor het doen verwijderen van internetpagina's die (delen van) de toetsinhoud onthullen?	ja/nee
RB4	Heeft de toetsaanbieder een gedegen en compleet fraudepreventieplan opgesteld met daarin opgenomen een PDCA cyclus voor de beveiliging van het toetsproces en de toetsinhoud?	ja/nee
RB4.1	Zijn er gedegen en complete procedures voor het ontdekken en evalueren van verdachte toetsresultaten en wijzigingen in kenmerken van toetsonderdelen en toetsscores in de loop van de tijd?	ja/nee
RB4.2	Controleert de toetsaanbieder regelmatig het internet en andere media op het bekend raken van (delen) van de inhoud van de toets of andere niet-openbare informatie over de doorstroomtoets?	ja/nee
RB5	Heeft de toetsaanbieder voldoende verantwoord hoe de toets(omgeving) beveiligd wordt?	ja/nee
RB6	Heeft de toetsaanbieder voldoende verantwoord hoe een leerling tijdens een tweede afnamemoment (omdat die bij de eerste afname was verhinderd) binnen de wettelijke gestelde termijn de toets alsnog kan maken en hoe deze eventueel verschilt van de toets van het eerste afnamemoment?	ja/nee/n.v.t.
RB7.1	Heeft de toetsaanbieder voldoende verantwoord hoe wordt voorzien in een terugvaloptie, die in het geval van kleine incidenten op leerling-, klas- of schoolniveau volgens een bijbehorend calamiteitenplan kan worden ingezet?	ja/nee/n.v.t.
RB7.2	Heeft de toetsaanbieder voldoende verantwoord hoe wordt omgegaan met binnengekomen incidenten van scholen en hoe deze meldingen worden geëvalueerd?	ja/nee
RB7.3	Heeft de toetsaanbieder voldoende verantwoord hoe er richting het onderwijsveld gecommuniceerd wordt over eventuele incidenten (d.w.z. via welke kanalen en welke inhoud)?	ja/nee
RB8	Heeft de toetsaanbieder voldoende verantwoord hoe het de benodigde gegevensoverdracht realiseert indien de landelijke calamiteitentoets van het Ministerie van Onderwijs, Cultuur en Wetenschap ingezet moet worden?	ja/nee/n.v.t.

Toelichting RB1 t/m RB5: De aanbieder moet overzichtelijk per kwaliteitseis aangeven hoe daar aan is voldaan. Een voorbeeld daarvan is de verwijzing naar een verwerkersovereenkomst, waar op de kwaliteitseisen RB1 t/m RB5 wordt ingegaan. Daarbij moet bewijslast worden overlegd van de bij Bl.2, Bl.3 en Bl.4 gevraagde documenten, processen, contracten en protocollen.

De verantwoordelijkheid voor het goed gebruik van de toets en de daarmee verzamelde informatie in de dagelijkse praktijk ligt bij de school / de eindgebruiker zelf.

Toelichting RB6: De toetsaanbieder beschrijft helder hoe een leerling tijdens een tweede afnamemoment (omdat die bij de eerste afname was verhinderd) binnen de wettelijke gestelde termijn alsnog de toets kan maken. De toetsaanbieder vermeldt daarnaast adequaat hoe eventueel deze toets verschilt van de toets van het eerste afnamemoment en waarom de toetsaanbieder hiervoor wel/niet kiest. Dit is informatie die niet per se in het toetsreglement en/of de afnamehandleiding wordt opgenomen en daarom extra is voor de adviseur t.o.v. de verantwoording van kwaliteitseis A1.

Afwijkende beoordeling calamiteitentoets: kwaliteitseis RB6 is niet van toepassing voor de calamiteitentoets.

Toelichting RB7.1: Het doel van de terugvaloptie is dat in het geval van een incident of onregelmatigheid de toetsaanbieder tijdig een toetsversie beschikbaar heeft zodat de leerling toch een doorstroomtoets van de initiële toetsaanbieder kan afnemen tijdens de afname- of inhaalperiode. De toetsaanbieder beschrijft in een bijbehorend calamiteitenplan de procedure ter voorbereiding op incidenten en calamiteiten. Hierin moet in ieder geval vastgelegd worden hoe de door de aanbieder geboden terugvaloptie van de doorstroomtoets eruitziet: bijvoorbeeld een itembank waaruit kan worden gekozen voor verschillende digitale versies of verschillende versies van een papieren toets. Ook moet het risico dat iteminhoud openbaar komt worden afgedekt.

Toetsaanbieders zijn bovendien zelf verantwoordelijk voor het in de gaten houden van signalen vanuit scholen, het onderwijsveld en sociale media dat hun doorstroomtoets mogelijk niet aan de voorwaarden voor doorgang vinden voldoet.

Afwijkende beoordeling calamiteitentoets: kwaliteitseis RB7.1 is niet van toepassing voor de calamiteitentoets.

Toelichting RB7.2: De toetsaanbieder beschrijft in een calamiteitenplan hoe scholen incidenten bij de toetsaanbieder kunnen melden en welke procedures klaarliggen om hiermee om te gaan. En hoe eventuele incidenten intern beoordeeld worden. Daarbij wordt ook vermeld hoe deze meldingen gearchiveerd worden, in het kader van evaluatie (en indien mogelijk openbaarheid).

Toelichting RB7.3: De toetsaanbieder beschrijft in een calamiteitenplan hoe er richting het onderwijsveld gecommuniceerd wordt over eventuele incidenten (d.w.z. via welke kanalen en welke inhoud)?

Toelichting RB8: De toetsaanbieder beschrijft in een calamiteitenplan hoe de benodigde gegevensoverdracht wordt gerealiseerd in het geval de landelijke calamiteitentoets van het Ministerie van Onderwijs, Cultuur en Wetenschap ingezet moet worden.

Afwijkende beoordeling calamiteitentoets: kwaliteitseis RB8 is niet van toepassing voor de calamiteitentoets.

5. Psychometrische aspecten

5.1. Inleiding

De toetsaanbieder moet nieuw geconstrueerde items pretesten. Dit om de psychometrische kwaliteit van de items te schatten, om van daaruit te kunnen bepalen welke items mogen worden gebruikt bij de definitieve toetssamenstelling.

De psychometrische kwaliteitsbeoordeling van de doorstroomtoets door de adviseur in opdracht van het CvTE bestaat uit de volgende stappen:

1. Pretestprocedure voor het valideren van nieuw geconstrueerde items, uit te voeren door de toetsaanbieder en bestaande uit:
 - a. Steekproefkader en samenstelling steekproef (paragraaf 5.2.1);
 - b. Kalibratie en kwaliteit van items (paragraaf 5.2.2);
 - c. Toetssamenstelling na pretest (paragraaf 5.2.3).
2. Landelijke itemgerelateerde (toetsoverstijgende) normering, uit te voeren door een adviseur in opdracht van het CvTE, zoals beschreven in de separate Regeling Beoordelingsnormen Doorstroomtoetsen po, en bestaande uit:
 - a. Het door de toetsaanbieder aanleveren van de ruwe itemscores, waarna een adviseur deze controleert en analyseert;
 - b. Kalibratie van de gezamenlijk ankeritems;
 - c. Vaststellen van cesuren van referentieniveaus en toetsadviezen;
 - d. Het door de toetsaanbieder vaststellen van de behaalde referentieniveaus en het toetsadvies per leerling.
3. Beoordeling van de kwaliteit en het functioneren van de doorstroomtoets (rapportage achteraf) zoals beschreven in de eisen KF1 t/m KF5, Hoofdstuk 5.4, te verwerken tot één rapportage.

In het geval van een MST, geldt dat voor wat betreft de psychometrische kwaliteitseisen uit Hoofdstuk 5 het CvTE per kwaliteitseis mag beoordelen of en welke van de eisen voor een lineaire doorstroomtoets en/of en welke van de eisen voor een computergestuurde adaptieve doorstroomtoets op itemniveau (CAT) van toepassing zijn.

De psychometrische kwaliteitsbeoordeling van de calamiteitentoets wijkt af van bovengenoemd kader. De calamiteitentoets bevat geen items uit het gezamenlijk anker. Ook de normering van de calamiteitentoets wijkt af van de reguliere normering. Dit is nader uitgewerkt in de Regeling Beoordelingsnormen doorstroomtoetsen po.

5.2 Pretestprocedure

Om nieuwe items in te kunnen zetten in een operationele afname, moeten deze eerst gepretest worden. De dataverzameling van deze nieuwe items kan twee doelen dienen:

- Doel A (basis-kwaliteitscontrole): dataverzameling van nieuwe items ten behoeve van een basis-kwaliteitscontrole. Hierbij wordt onder andere de itemkwaliteit geëvalueerd aan de hand van itemstatistieken uit de klassieke testtheorie. Na de dataverzameling selecteert de toetsaanbieder de items die in aanmerking komen voor gebruik in de operationele toets.
- Doel B (basis-kwaliteitscontrole + additionele dataverzameling): in aanvulling op bovenstaande beschrijving van doel A worden bij dit doel aanvullend de verzamelde data gebruikt in de kalibratie na de operationele afname. Dat is vooral relevant bij een MST of CAT, waarbij de afname niet zo ingericht kan zijn dat van tevoren bepaald kan worden dat de opgaven door voldoende leerlingen gemaakt zullen worden. Voor een MST of een CAT wordt deze dataverzameling gebruikt voor het schatten van itemparameters ten behoeve van de adaptieve module.

De dataverzameling kan worden uitgevoerd aan de hand van twee scenario's:

- Scenario 1: de aanbieder neemt een aparte proeftoets af bij een representatieve steekproef waarin de items worden afgenomen onder omstandigheden die vergelijkbaar zijn met de operationele afname.

- Scenario 2: nieuwe items worden gezaaid in de operationele afname. Items mogen voor een leerling niet herkenbaar zijn als zaai-item.

De pretestprocedure bestaat in beide scenario's uit drie stappen: 1) Steekproefkader en samenstelling steekproef vaststellen (paragraaf 5.2.1), 2) Kalibratie en kwaliteit van items vaststellen met de verzamelde pretestdata (paragraaf 5.2.2), en 3) Definitieve doorstroomtoets samenstellen (paragraaf 5.2.3).

5.2.1 Steekproefkader en samenstelling steekproef

De onderwijskundig goed geconstrueerde items worden gepretest volgens de kwaliteitseisen voor het steekproefkader en de samenstelling van de steekproef.

Kwaliteitseisen

Beslisregel: Om aan deze categorie kwaliteitseisen te voldoen, moeten alle vragen met JA worden beantwoord.

Code	Vragen	Mogelijke antwoorden
N1	Is de steekproef groot genoeg?	ja/nee
N2	Is de steekproef relevant?	ja/nee
N3	Wordt het afnamedesign dat is gebruikt in de pretest adequaat beschreven en verantwoord?	ja/nee

Toelichting N1: De aanbieder geeft aan via welk scenario en met welke doel de nieuwe items zijn gepretest. Vervolgens geeft de aanbieder aan door hoeveel leerlingen ieder nieuw item is gepretest. Het aantal observaties wordt als voldoende gezien wanneer ieder item door minstens 200 leerlingen gemaakt is. Uiteraard geven meer gegevens meer zekerheid. Indien doel B wordt nagestreefd moet rekening worden gehouden met het minimumaantal observaties dat per item wordt vereist voor de kalibratie na operationele afname (zie kwaliteitscriterium NR1). Het staat de aanbieder vrij om zowel de gegevens van de reguliere bo-leerlingen als de s(b)o-leerlingen te gebruiken, mits er geen sprake is van DIF (zie kwaliteitscriterium KA2).

Gekloonde opgaven zijn in beginsel nieuwe opgaven, en worden zo ook beoordeeld. Wanneer er sprake is van (zeer) kleine wijzigingen, zoals de aanpassing van een munteenheid (van euro naar dollar), of de aanpassing van een naam (bijvoorbeeld ten behoeve van inclusie), waarbij de opgave inhoudelijk hetzelfde blijft, dan kunnen dergelijke items als onveranderd gezien worden. De toetsaanbieder moet dan wel duidelijk aangeven welke (inhoudelijke) aanpassing gedaan is, bij voorkeur met een rationale waaruit blijkt waarom de aanpassing geen impact heeft.

Wanneer de gekloonde opgave in termen van relevantie en/of objectiviteit verandert ten opzichte van het originele item, dan is er wel sprake van een nieuwe opgave. Meer specifiek is er sprake van een nieuwe opgave wanneer:

- de gekloonde opgave een relatie heeft met een ander of gewijzigd leerdoel;
- de gekloonde opgave naar verwachting een andere of gewijzigde moeilijkheidsgraad heeft;
- de gekloonde opgave naar een ander of gewijzigd modelantwoord leidt.

Wanneer eerder al data verzameld zijn bij het item in een andere vorm moet ook bij kleine wijzigingen het gekloonde item een nieuw label krijgen.

Toelichting N2: De steekproef moet relevant zijn voor de toepassing. In het geval van scenario 1 geldt dat aan de volgende zes voorwaarden moet worden voldaan:

- 1) Op alle in de onderstaande tabel genoemde variabelen zitten er op minimaal twee niveaus per variabele en op minimaal drie niveaus van de variabele schoolweging gegevens in de steekproef;
- 2) Er is aangegeven hoe de steekproef getrokken is;
- 3) Er is aangegeven onder welke omstandigheden de data verzameld zijn;
- 4) Er is aangegeven hoe de steekproef zich verhoudt tot de verwachte populatie die de operationele toets afneemt met betrekking tot de in de tabel genoemde niveaus per variabele;
- 5) Er is aangegeven hoe de steekproef zich verhoudt tot de landelijke populatiewaarden van leerlingen in groep 8 met betrekking tot de in de tabel genoemde niveaus per variabele;
- 6) Er is aangegeven of en hoeveel S(B)O scholen hebben mee gedaan.

Als bij (3) de omstandigheden afwijken van de werkelijke afname wordt aangegeven welke impact verwacht wordt en hoe daar in de selectie van opgaven voor de operationele afname rekening mee is gehouden.

Variabele	#N	Welke niveaus (NB: #N = aantal niveaus)
regio	4	noord, zuid, oost, west
stedelijkheid	2	verstedelijkt, landelijk ¹
schoolgrootte	3	aantal leerlingen: <100, 100–300, >300 leerlingen
schoolweging	5	gewichtscategorieën: <23 / 23–26,99 / 27–32,99 / 33–36,99 / ≥ 37

¹ De mate van verstedelijking is terug te voeren op de bij het CBS gebruikelijke indeling naar vijf niveaus, namelijk zeer sterk, sterk, matig, weinig en niet verstedelijkt. De vijfdeling is hier teruggebracht naar een tweedeling in enerzijds verstedelijkt (zeer sterk en sterk) en anderzijds landelijk (matig, weinig en niet verstedelijkt). De toetsaanbieder kan ervoor kiezen om van deze indeling af te wijken, mits deze keuze is voldoende onderbouwd.

In het geval van scenario 2 moet de aanbieder aantonen dat de populatie voor de zaai-items niet structureel anders is dan de populatie van de niet-zaai-items.

Afwijkende beoordeling calamiteitentoets: De eis dat de steekproef van de pretest overeen moet komen met de populatie die de toets van de aanbieder regulier afneemt, geldt niet voor de calamiteitentoets. Voor de calamiteitentoets is 'leerlingen uit groep 8' de best passende operationalisering van de beoogde doelgroep.

Toelichting N3: Het afnamedesign van de pretest wordt door de aanbieder uiteengezet. Wanneer er sprake is van doel B is het van belang dat de parameters per (sub)domein op dezelfde schaal liggen. Hiervoor is het van belang dat er sprake is van een afnamedesign waarbij dit mogelijk is. Dit design wordt aangeleverd en bevat – indien van toepassing – in ieder geval informatie over:

- het totaal aantal boekjes ('booklets');
- aantal observaties per boekje;
- het aantal items dat tussen de opgavenboekjes onderling overlapt;
- het aantal items, en de kwaliteit van de items, dat de pretest verbindt met de bestaande schaal.

Wanneer de pretest gedaan wordt door middel van zaai-opgaven in de afname dan wordt aangegeven hoe dat zaai-design er uitziet.

Wanneer de pretestdata zijn verzameld met een CAT is het niet zinvol om het hele onvolledige design weer te geven, omdat iedere leerling dan in principe een unieke verzameling items maakt. Het moet – indien sprake is van doel B – wel duidelijk zijn hoe deze data worden verbonden aan de landelijke schaal na kalibratie na afgelopen operationele afname.

Ook moet duidelijk zijn op welke posities de verschillende opgaven ingezet zijn, zodat door de toetsaanbieder op gedegen en complete wijze beargumenteerd kan worden op welke wijze er rekening is gehouden met vermoeidheids- en/of volgorde effecten.

5.2.2 Itemkalibratie en toetskwaliteit

Na het verzamelen van de observaties voert de toetsaanbieder een kalibratie uit met de verzamelde pretestdata. In het geval dat opgaven al eerder zijn afgenomen (bij een itembank of bij pretestgegevens) kan er een schaal gekalibreerd worden die op basis van de afnamegegevens bestendig moet worden. Op basis van de afnamegegevens kunnen de (goed functionerende) zaai-items op de schaal geplaatst worden. In de kalibratie schat de toetsaanbieder de itemparameters, zodat de kwaliteit van de gepreteste items kan worden vastgesteld.

Kwaliteitseisen

Beslisregel: Om aan deze categorie kwaliteitseisen te voldoen, moeten alle vragen met JA of N.V.T. worden beantwoord.

Code	Vragen	Mogelijke antwoorden
KA1	Zijn de nieuwe items psychometrisch gezien van voldoende kwaliteit?	ja/nee
KA2	Heeft er gedegen DIF-onderzoek plaatsgevonden ten aanzien van de nieuwe gepreteste items?	ja/nee
KA3	Worden beslisregels voor adaptieve toetsen geëxpliciteerd?	ja/nee/n.v.t.
KA4	Zijn de metingen van de individuele leerlingen voldoende betrouwbaar om uitspraken op individueel niveau te kunnen doen?	ja/nee
KA5	Past het psychometrisch model wat gebruikt wordt voor de toetssamenstelling bij de data?	ja/nee

Toelichting KA1: De toetsaanbieder selecteert de gepreteste items die psychometrisch gezien van

voldoende kwaliteit zijn om te kunnen gebruiken in de samen te stellen doorstroomtoets. Dat betekent dat de toetsaanbieder op gedegen en complete wijze onderbouwt dat de geselecteerde items aan minimaal de volgende eisen voldoen:

- de moeilijkheidsgraad is passend voor het meetdoel en de doelgroep (d.w.z. voor lineaire toetsen vergelijkbaar met een p-waarde tussen de gokkans en 0,95 en voor adaptieve toetsen passend bij de moeilijkheidsgraad van het betreffende design en/of toetspad);
- het discriminerend vermogen is vergelijkbaar met een rir-waarde $> 0,15$, rekening houdend met de beperkte vaardigheidsrange van de leerlingen in het geval van een adaptieve toets;
- de kwaliteit van de afleiders van meerkeuze-opgaven wordt geëvalueerd middels a- en rar-waarden. Deze moet de aanbieder presenteren. Bij het selecteren van een item met een rar-waarde $> 0,05$ moet de aanbieder voor dat item aangeven waarom dit item is meegenomen. Voor items die worden gepretest in een adaptieve pretest is het de aanbieder toegestaan om afleider-thetacorrelaties (IRT-variant op de klassieke rar-waarde waarbij theta is gebaseerd op de scores items minus het betreffende item) te berekenen, waarvoor 0,05 eveneens als grenswaarde geldt.
- de interbeoordelaarsovereenstemming, in het geval van open vragen, vergelijkbaar is met een interbeoordelaarsovereenstemmingscoëfficiënt $> 0,80$.

De toetsaanbieder verantwoordt op gedegen en complete wijze de keuze voor de geselecteerde items die in aanmerking komen voor de doorstroomtoets en waarom items die niet aan deze eisen voldoen onder gegeven omstandigheden toch zijn mee genomen.

Toelichting KA2: De toetsaanbieder beschrijft op gedegen en complete wijze hoe ze onderzoek naar DIF heeft uitgevoerd. Dit wordt gedaan door – indien van toepassing en indien voldoende observaties per categorie – een DIF-analyse op itemniveau uit te voeren ten aanzien van:

- Proeftoetsboekjes;
- Schoolweging. Het staat de aanbieder vrij hoe ze de leerlingen indeelt in een x-aantal schoolwegingscategorieën;
- Schooltypes (regulier bo vs. s(b)o).

Hiervoor geldt:

- De DIF-analyses worden grafisch onderbouwd;
- Een DIF-statistiek naar eigen keuze wordt toegepast en de resultaten worden onderbouwd.

De grafische weergaven en DIF-statistieken moeten aan worden geleverd voor ieder item, dus ook voor de items waarbij geen indicatie is gevonden van DIF. Wanneer een item met een indicatie van DIF desondanks wordt geselecteerd voor de operationele afname moet de aanbieder aantonen dat het verwachte effect op de behaalde referentieniveaus en behaalde toetsadviezen verwaarloosbaar is.

Toelichting KA3 (alleen van toepassing voor adaptieve toetsen): Een adaptieve toets is een toets die zich aanpast aan het individuele vaardigheidsniveau van de persoon die de toets maakt. In plaats van alle deelnemers dezelfde set vragen voor te leggen, past een adaptieve toets de complexiteit van de vragen aan op basis van de correctheid van de eerdere antwoorden van de deelnemer. Het idee hierachter is dat hierdoor een efficiënte schatting van het vaardigheidsniveau van de leerling kan worden verkregen. Derhalve moeten van dit type doorstroomtoets de beslisregels of de algoritmes voor de samenstelling van de toets zijn geëxpliciteerd.

Voor een CAT moet de toetsaanbieder op gedegen en complete wijze verantwoorden:

1. Hoe de toets wordt gestart;
2. Hoe de keuze voor een volgend item en wordt gemaakt;
3. Wanneer de toets wordt beëindigd (maximum aantal items en standard error).

Voor een MST moet de toetsaanbieder op gedegen en complete wijze verantwoorden:

1. Hoe de toets wordt gestart;
2. Hoe de keuze voor het volgende blokje wordt gemaakt;
3. Wanneer de toets wordt beëindigd.

Naast dat de beslisregels voor de samenstelling van de toets geëxpliciteerd moeten zijn, is het belangrijk dat tevens beslisregels geëxpliciteerd zijn om dekking van het inhoudelijke domein te waarborgen per individuele afname. Dit houdt in dat indien er bijvoorbeeld vier verschillende domeinen van rekenen worden bevraagd, de beslisregels van de adaptieve toets ervoor moeten zorgen dat ook uit alle domeinen voldoende opgaven worden afgenomen.

Toelichting KA4: De schattingsfout van iedere geschatte vaardigheid moet beperkt zijn voor de nieuw in te zetten operationele toets. Het evalueren van de meetnauwkeurigheid op verschillende posities op de latente vaardigheidsschaal wordt aan de hand van onderstaande vragen beoordeeld:

- Is er een gedegen en complete beschrijving van de procedure, waaruit blijkt dat de meetnauwkeurigheid op de juiste manier wordt berekend?

- Is de verdeling van schattingsfouten op de vaardigheidsschattingen grafisch weergegeven, zowel voor de schaal waarop de referentieniveaus gegeven worden, als de schaal waarop de toetsadviezen gegeven worden?
- Is de hoogte van de meetnauwkeurigheid rond de cesuurpunten van de referentieniveaus groter dan 0,70?
- Is de hoogte van de meetnauwkeurigheid rond de cesuurpunten van de toetsadviezen groter dan 0,90?³ Indien de hoogte van de meetnauwkeurigheid rond een cesuurpunt van een toetsadvies niet groter is dan 0,90, moet de toetsaanbieder voldoende verantwoorden waarom het behalen van een meetnauwkeurigheid groter dan 0,90 niet mogelijk was en hoe zij desondanks hebben getracht toch een meetnauwkeurigheid zo dicht als mogelijk richting 0,90 te realiseren.⁴

Toelichting KA5: Alle psychometrische modellen werken met aannames en voorwaarden waaraan voldaan moet worden, wil het model bruikbaar zijn. Past het – voor de toetsamenstelling gebruikte model – bij de data? De aanbieder moet de itemfit grafisch onderbouwen. Daarnaast moet de aanbieder een passingsmaat hanteren en de resultaten onderbouwen.

5.2.3. Toetsamenstelling

Na afronding van de kalibratie van de nieuwe items, stelt de toetsaanbieder de definitieve doorstroomtoets samen.

Kwaliteitseisen

Beslisregel: Om aan deze categorie kwaliteitseisen te voldoen, moeten alle vragen met JA of N.V.T. worden beantwoord.

Code	Vragen	Mogelijke antwoorden
T1	Bevat de doorstroomtoets voor de terreinen Nederlandse taal en Rekenen in ieder geval vragen met een moeilijkheidsgraad tussen de cesuur van referentieniveau 1F en 2F/1S?	ja/nee
T2	Bevat de doorstroomtoets de door het CvTE voorgeschreven subset van gezamenlijke ankeritems voor het wettelijk verplichte domein Lezen en subdomein Taalverzorging van het terrein Nederlandse taal en het terrein Rekenen op de in de Regeling Beoordelingsnormen Doorstroomtoetsen po beschreven wijze?	ja/nee/n.v.t.
T3	Wordt het afnamedesign dat is gebruikt in de operationele afname adequaat beschreven en verantwoord?	ja/nee/n.v.t.
T4	Bestaat een papieren toets of een lineaire digitale toets jaarlijks volledig uit nieuwe items voor het wettelijk verplichte domein Lezen en subdomein Taalverzorging van het terrein Nederlandse taal en het terrein Rekenen, met uitzondering van de items van het gezamenlijk anker?	ja/nee/n.v.t.
T5	Is er in het geval van een CAT of MST sprake van een adequate verversingsstrategie en wordt de toets dus voldoende ververs over de jaren?	ja/nee/n.v.t.
T6	Heeft de itembank van een CAT separaat voor het wettelijk verplichte domein Lezen en subdomein Taalverzorging van het terrein Nederlandse taal en het terrein Rekenen een omvang van minimaal 6x het aantal unieke items van een maximale operationele toetsversie?	ja/nee/n.v.t.

Toelichting T1: Voor zowel een MST als papieren en digitale lineaire doorstroomtoetsen geldt dat de toets voor de terreinen Nederlandse taal en Rekenen in ieder geval vragen moet bevatten die inhoudelijk aansluiten bij de referentieniveaus 1F en 2F/1S en die een moeilijkheidsgraad tussen de cesuur van referentieniveau 1F en 2F/1S hebben. In het geval van een CAT is de itemselectie geoptimaliseerd rond de opeenvolgende vaardigheidsschattingen. De aanbieder presenteert hiertoe per domein grafisch een verdeling van moeilijkheidsparameters samen met de cesuren.

Toelichting T2: De door het CvTE toegewezen subset van gezamenlijke ankeritems is opgenomen in (de varianten van) de doorstroomtoets. Eventuele (minimale) aanpassingen op deze items t.o.v. het moederitem zijn alleen toegestaan met akkoord en na beoordeling van het CvTE.

De door het CvTE toegewezen subset van gezamenlijke ankeritems moet op een zodanige wijze over de toets worden verspreid, dat de positie in de toets zo min mogelijk invloed op de prestatie van de leerling heeft en de ankeritems qua lay-out en tekstuele vormgeving niet als opvallend afwijkend in de toets herkenbaar zijn, een en ander zoals gespecificeerd in de Regeling beoordelingsnormen door-

³ Voor de berekening hiervan mag worden uitgegaan van een ongecorreleerde meetfout tussen de domeinen, zie ter illustratie de rapportage 'Secundaire analyses eindtoetsen 2022' van de Expertgroep toetsen PO (2022).

⁴ In het beoordelingskader voor doorstroomtoetsen voor 2028 is het CvTE voornemens op te nemen dat de hoogte van de meetnauwkeurigheid rond de cesuurpunten van de toetsadviezen altijd groter moet zijn dan 0,90.

stroomtoetsen PO. De aanbieder legt duidelijk uit hoe de ankeritems over de toets zijn verspreid. Omdat de vergelijkbaarheid van de ankeritems tussen verschillende toetsaanbieders gewaarborgd moet worden, is het belangrijk dat de ankeritems nagenoeg onveranderd in de toets worden opgenomen. Daarom moet de toetsaanbieder de voorgenomen veranderingen aan de ankeritems uitleggen en rechtvaardigen.

Afwijkende beoordeling calamiteitentoets: kwaliteitseis T2 is niet van toepassing voor de calamiteitentoets.

Toelichting T3 (alleen van toepassing voor een lineaire toets of MST): De aanbieder beschrijft het afnamesdesign dat wordt gebruikt in de operationele afname. Hiertoe wordt beschreven hoe de items zijn verdeeld over verschillende toetsversies/toetspaden. Het is mogelijk dat de toetsversies/toetspaden anders ingericht of verminderd moeten worden door een terugloop in aanmeldingen. De aanbieder beschrijft dergelijke afnamescenario's die afhankelijk zijn van het aantal aanmeldingen, of ze geeft aan waarom verschillende afnamescenario's niet nodig worden geacht.

Toelichting T4 (alleen van toepassing voor lineaire toetsen): Een lineaire toets bestaat jaarlijks volledig uit nieuwe gekalibreerde items voor het wettelijk verplichte domein Lezen en subdomein Taalverzorging van het terrein Nederlandse taal en het terrein Rekenen, met uitzondering van de items van het gezamenlijk anker en een intern anker. Als een intern anker wordt gebruikt moet de noodzakelijkheid ervan voldoende worden uitgelegd.

Afwijkende beoordeling calamiteitentoets: kwaliteitseis T4 is niet van toepassing voor de calamiteitentoets.

Toelichting T5 (alleen van toepassing bij al eerder erkende of vastgestelde adaptieve toetsen): In het geval van een MST of een CAT, moet er sprake zijn van een adequate jaarlijkse verversingsstrategie en moet de toets ook over de jaren voldoende worden verversd. De verversing moet op itemniveau gedocumenteerd zijn. Daarbij geldt het volgende:

- De aanbieder hanteert jaarlijks een adequate verversingsstrategie voor de opgaven van de itembank. De toetsaanbieder geeft op gedegen en complete wijze aan hoeveel procent van de opgaven is verversd en onderbouwt dat de toegepaste vernieuwing voldoende is om overmatige blootstelling van items te voorkomen (exposure control).
- Items die uitgelekt lijken, moeten verwijderd worden. De toetsaanbieder geeft op gedegen en complete wijze aan hoeveel items er vermoedelijk zijn uitgelekt en hoeveel van deze items uit de itembank zijn verwijderd.
- Voor items die in het vorige jaar door het algoritme niet geselecteerd zijn, geeft de toetsaanbieder op gedegen en complete wijze aan, waarom deze niet of wel uit de itembank zijn verwijderd.

Afwijkende beoordeling calamiteitentoets: kwaliteitseis T5 is niet van toepassing voor de calamiteitentoets.

Toelichting T6: Deze kwaliteitseis is alleen van toepassing voor een CAT.

Afwijkende beoordeling calamiteitentoets: kwaliteitseis T6 is niet van toepassing voor de calamiteitentoets.

5.3 Normering ten behoeve van leerlingrapportage

Kwaliteitseisen

Beslisregel: Om aan deze categorie kwaliteitseisen te voldoen, moeten alle vragen met JA of N.V.T. worden beantwoord.

Code	Vragen	Mogelijke antwoorden
NR1	Zijn er voldoende observaties per toetsopgave?	ja/nee
NR2	Is de gekozen normeringsmethode voldoende omschreven?	ja/nee
NR3	Zijn de gerapporteerde scores voldoende nauwkeurig?	ja/nee/n.v.t.

Toelichting NR1: Iedere toetsopgave heeft een minimumaantal observaties van 600, eventueel aangevuld met pretestdata die verkregen zijn met doel B voor ogen of eerdere afnamejaren. Observaties uit alle onderwijstypen kunnen gebruikt worden. Voor ankeropgaven kunnen andere vereiste aantallen observaties gelden volgend uit de regeling beoordelingsnormen doorstroomtoetsen po.

Toelichting NR2: Vanuit de landelijke normering worden aan alle aanbieders 2PL parameterschattingen geleverd. Hoe past de toetsaanbieder deze toe bij diens normering? Zodra de gehanteerde methode afwijkt van de landelijke normering zoals omschreven in de Regeling beoordelingsnormen doorstroomtoetsen PO (bijvoorbeeld een normering via een discrete (standaard)scoreschaal of verschillen in de meegekalibreerde onderwijstypen), moet de toetsaanbieder de gehanteerde normering duiden en moet de aanbieder aangeven hoe de scores op de eigen scoreschaal zijn berekend uit de latente vaardigheden op de afzonderlijke domeinen en/of uit de GLV-scores.

Toelichting NR3: Indien de uiteindelijke rapportageschaal afwijkt van de latente schaal waarop de cesuren zijn vastgesteld, moet de toetsaanbieder de nauwkeurigheid van de gekozen normeringsmethode ten opzichte van een 2PL-vaardigheidsschatter en de landelijk gehanteerde cesuurpunten van de toetsadviezen en referentieniveaus via een tabel weergeven (verwarringsmatrix) en verantwoordt mogelijk verschillen. Hierbij geldt dat classificatie van behaalde toetsadviezen en behaalde referentieniveaus op basis van de getransformeerde toetsscores niet meer dan 5% mag afwijken van de classificatie van behaalde toetsadviezen en behaalde referentieniveaus op basis van de landelijke normering. De 5% geldt zowel voor de referentieniveaus als bij de toetsadviezen voor het totaal. Bij de verwarringsmatrices is het dus niet toegestaan dat meer dan 5% van de observaties buiten de diagonaal valt.

Indien de aanbieder niet werkt met getransformeerde toetseigen scores is deze eis voor de betreffende aanbieder niet van toepassing.

5.4 Kwaliteit en functioneren van de doorstroomtoets (rapportage achteraf)

Deze kwaliteitseisen zijn alleen van toepassing voor al eerder afgenomen toetsen. De beoordeling van de kwaliteit en het functioneren van de calamiteitentoets vindt alleen plaats nadat de operationele toets daadwerkelijk is ingezet. Het CvTE beslist of de calamiteitentoets meedoet aan het Toelatings- en doorstroomonderzoek.

Kwaliteitseisen

Beslisregel: Om aan deze categorie kwaliteitseisen te voldoen, moeten alle vragen met JA of N.V.T. worden beantwoord.

Code	Vragen	Mogelijke antwoorden
KF1	Verantwoordt de toetsaanbieder op een gedegen manier of en waarom opgaven met inhoudelijke fouten, psychometrisch disfunctioneren, DIF of te weinig waarnemingen al dan niet zijn meegenomen voor de scorerapportage?	ja/nee/n.v.t.
KF2	Reflecteert de toetsaanbieder via een analyse op het functioneren van de toetsopgaven die op basis van adequate onderbouwing zijn opgenomen in de operationele afname?	ja/nee/n.v.t.
KF3	Zijn verschillen in percentage referentieniveaus voldoende verantwoord?	ja/nee/n.v.t.
KF4	Zijn verschillen in de verdeling van populatievaardigheden ten opzichte van de verantwoording van het afgelopen jaar en de onderdelen onderling voldoende verantwoord?	ja/nee/n.v.t.
KF5	Worden afwijkingen van het landelijke gemiddelde in de relatie tussen toets- en schooladviezen voldoende in kaart gebracht en waar mogelijk voldoende verantwoord?	ja/nee/n.v.t.

De toetsaanbieder verantwoordt jaarlijks na de toetsafname-periode de kwaliteit en het functioneren van de afgenomen doorstroomtoets op basis van de bovengenoemde vijf kwaliteitscriteria. Er wordt beoordeeld of de inhoud van de informatie die door de toetsaanbieder is aangeleverd voldoet aan de vooraf gestelde eisen. Voor een nieuwe aanbieder zijn deze eisen bij de eerste evaluatie niet van toepassing.

Toelichting KF1: De aanbieder presenteert het aantal wel en het aantal niet meegenomen items voor de scorerapportage naar de leerlingen wegens:

1. een inhoudelijke fout;
2. psychometrisch disfunctioneren;
3. DIF;
4. te weinig waarnemingen.

Daarbij wordt verantwoord waarom de meegenomen items in de scorerapportage zijn meegenomen.

Voor (3) geldt: indien er voldoende observaties zijn worden DIF-analyses uitgevoerd tussen:

- Pretest en operationele afname, indien pretestdata worden meegenomen in de kalibratie na operationele afname;
- Operationele afnames, indien data van operationele afnames van de jaren hiervoor worden meegenomen in de kalibratie na operationele afname;

- Schooltypes (regulier bo vs. s(b)o) indien van toepassing.

Verder geldt:

- De DIF-analyses worden grafisch onderbouwd.
- Een DIF-statistiek naar eigen keuze wordt toegepast en de resultaten worden onderbouwd.

De grafische weergaven en DIF-statistieken moeten aan worden geleverd voor ieder item, dus ook voor de items waarbij geen indicatie is gevonden van DIF. Wanneer een item met een indicatie van DIF desondanks wordt meegenomen ten behoeve van de scorerapportage moet de aanbieder aantonen wat het effect is van de inzet van het betreffende item op de behaalde referentieniveaus en behaalde toetsadviezen.

Toelichting KF2: Indien bij de beoordeling van de afgelopen indiening er bij kwaliteitseis V2 afspraken zijn gemaakt over het monitoren van het functioneren van specifieke toetsopgaven, moet de aanbieder de resultaten van de inhoudelijke en psychometrische analyse aanleveren.

Toelichting KF3: De toetsaanbieder moet informatie verstrekken over de percentages referentieniveaus van het wettelijk verplichte domein Lezen en subdomein Taalverzorging van het terrein Nederlandse taal en het terrein Rekenen plus een overzicht van de percentages toetsadviezen (de zes categorieën) van dit jaar ten opzichte van het vorige jaar (in tabelformaat). Verschillen tussen de beide jaren worden door de toetsaanbieder getoetst. Betekenisvolle verschillen tussen beide jaargangen worden door de aanbieder verklaard.

Toelichting KF4: De aanbieder moet inzicht geven in de populatieverdeling van de vaardigheid. Hiertoe worden populatieparameters (gemiddelde en standaarddeviatie) en de schattingsfout op de populatieparameters gegeven. De verdelingen worden grafisch inzichtelijk gemaakt. In deze informatie wordt onderscheid gemaakt tussen de verschillende onderwijstypen. Eventuele verschillen ten opzichte van de rapportage achteraf van het afgelopen jaar worden getoetst. Betekenisvolle verschillen tussen beide jaargangen worden door de aanbieder verklaard. Eventuele verschillen ten opzichte van de landelijke normering worden verklaard.

Toelichting KF5: De toetsaanbieder moet informatie verstrekken over de relatie tussen de toetsadviezen en de eerder gegeven (voorlopige) schooladviezen voor zover deze bekend zijn (in tabelformaat). Als de relatie tussen toets- en schooladviezen sterk afwijkt van de landelijke gemiddelde verdeling, moet de toetsaanbieder mogelijke redenen van de afwijking in kaart brengen. Deze redenen moeten in ieder geval (maar niet uitsluitend) worden gezocht op basis van specifieke aspecten van de eigen populatie en de gebruikte toets (zoals afnamevorm, algoritme ect...), en moeten waar mogelijk voldoende worden verantwoord.

6. Vierjaarlijkse erkenning en jaarlijkse vaststelling door het CvTE

6.1. De vierjaarlijkse erkenning

Een adviseur beoordeelt iedere vier jaar in opdracht van het CvTE de onderwijskundige, organisatorische en psychometrische aspecten van de doorstroomtoets. Na de beoordeling stelt een adviseur een advies op. Dit advies wordt uitgebracht aan het CvTE dat, een besluit neemt over het al dan niet erkennen van de betreffende doorstroomtoets. Wanneer de toets is erkend kan deze toets vier achtereenvolgende jaren worden ingezet als het CvTE vaststelt dat de toets voldoet aan de voorwaarden waaronder hij is erkend.

6.2. De jaarlijkse vaststelling

Na de erkenning van vier jaar, dienen de toetsaanbieders opnieuw jaarlijks een toelatingsaanvraag in bij het CvTE. Voor deze jaarlijkse vaststelling vindt opnieuw een beoordeling plaats. Voor deze beoordeling van de onderwijskundige inhoud, organisatorische aspecten en psychometrische aspecten doet het CvTE eveneens een beroep op een adviseur. Deze adviseur beoordeelt jaarlijks aan de hand van het beoordelingskader de kwaliteit van de doorstroomtoets, en stelt nu een advies op over of de toets nog voldoet aan de wet- en regelgeving. Dit advies wordt uitgebracht aan het CvTE, dat een besluit neemt en vaststelt of de toets voldoet.

Voor de calamiteitentoets wordt op een aantal punten van het beoordelingskader afgeweken.



Model leeswijzer

Verantwoording Doorstroomtoets

Kwaliteitseis		Niet van toepassing (n.v.t) ofwel mag met nee beantwoord worden (m.m.n.b.w.)	Beschreven in document: (naam document)	Hoofdstuk + paragraaf	Evt. opmerkingen ¹
Paragraaf kader	Code				
3.2	V1				
	V2				
3.3.1.1	RL1				
	RL2				
	RL3				
	RL4				
	RL5				
	RL6				
	RL7				
	RL8				
3.3.1.2	L1.1				
	L1.2				
	L1.3				
	L2				
	L3.1				
	L3.2				
	L3.3				
	L3.4				
	L3.5				
	L3.6				
	L3.7				
3.3.1.3	L4				
	L5				
	L6.1				
	L6.2				
	L6.3				
	L6.4				
3.3.2	TV1				
	TV2				
	TV3				
	TV4				
	TV5				
	TV6				



Kwaliteitseis		Niet van toepassing (n.v.t) ofwel mag met nee beantwoord worden (m.m.n.b.w.)	Beschreven in document: (naam document)	Hoofdstuk + paragraaf	Evt. opmerkingen ¹
Paragraaf kader	Code				
	TV6.a				
	TV6.b				
	TV6.c				
	TV6.d				
	TV6.e				
	TV7				
3.4.1	D.G				
	D.Vh				
	D.M				
	D.Vb				
	D1				
3.4.2	O.A				
	O.B				
	O.C				
	O.D				
3.4.3	C1				
	C2				
	C3				
3.4.4	K1				
	K2				
3.4.5	R1				
3.5	TN1				

¹ Denk bijvoorbeeld aan paginanummers. Indien het document geen paginanummers bevat, kan de aanbieder alsnog specificeren waar de informatie gevonden kan worden. Voorbeeld: bij een Excel document aangeven welke rijen en kolommen het betreft.

Kwaliteitseis		Niet van toepassing (n.v.t)	Beschreven in document: (naam document)	Hoofdstuk + paragraaf	Evt. opmerkingen ¹
Paragraaf kader	Code				
4.1	LR1				
	LR2				
	LR3				
	LR4				
	LR5				
	LR6				
	LR7				
4.2	LL1				



Kwaliteitseis		Niet van toepassing (n.v.t)	Beschreven in document: (naam document)	Hoofdstuk + paragraaf	Evt. opmerkingen ¹
Paragraaf kader	Code				
	LL1.a				
	LL1.b				
	LL1.c				
	LL1.d				
	LL1.e				
	LL1.f				
	LL1.g				
	LL2				
	LL2.a				
	LL2.b				
	LL2.c				
	LL2.d				
4.3	A1				
	A2.1				
	A2.2				
	A2.3				
	A2.4				
	A2.5				
	A2.6				
	A2.7				
	A3				
4.4	RB1				
	RB1.1				
	RB1.2				
	RB1.3				
	RB1.4				
	RB1.5				
	RB1.6				
	RB1.7				
	RB2				
	RB3				
	RB4				
	RB4.1				
	RB4.2				
	RB5				
	RB6				
	RB7.1				



Kwaliteitseis		Niet van toepassing (n.v.t)	Beschreven in document: (naam document)	Hoofdstuk + paragraaf	Evt. opmerkingen ¹
Paragraaf kader	Code				
	RB7.2				
	RB7.3				
	RB8				

¹ Denk bijvoorbeeld aan paginanummers. Indien het document geen paginanummers bevat, kan de aanbieder alsnog specificeren waar de informatie gevonden kan worden. Voorbeeld: bij een Excel document aangeven welke rijen en kolommen het betreft.

Kwaliteitseis		Niet van toepassing (n.v.t)	Beschreven in document: (naam document)	Hoofdstuk + paragraaf	Evt. opmerkingen ¹
Paragraaf kader	Code				
5.2.1	N1				
	N2				
	N3				
5.2.2	KA1				
	KA2				
	KA3				
	KA4				
	KA5				
5.2.3	T1				
	T2				
	T3				
	T4				
	T5				
	T6				
5.3	NR1				
	NR2				
	NR3				
5.4	KF1				
	KF2				
	KF3				
	KF4				
	KF5				

¹ Denk bijvoorbeeld aan paginanummers. Indien het document geen paginanummers bevat, kan de aanbieder alsnog specificeren waar de informatie gevonden kan worden. Voorbeeld: bij een Excel document aangeven welke rijen en kolommen het betreft.



Optionele domeinen				
Kwaliteitseis		Beschreven in document: (naam document)	Hoofdstuk + paragraaf	Evt. opmerkingen ¹
3.3.3	S1.1			
	S1.2			
	S1.3			
	S2.1			
	S2.2			
	S2.3			
	S2.4			
	S3.1			
	S3.2			
	S3.3			
	S4.1			
	S4.2			
	S4.3			
	S4.4			
3.3.4	M1.1			
	M1.2			
	M2.1			
	M2.2			
	M2.3			
	M2.4			
	M3			
	M4.1			
	M4.2			
	M4.3			
	M4.4			
	M5.1			
	M5.2			
	M6.1			
	M6.2			
	M6.3			
	M6.4			
	M7.1			
	M7.2			
	M7.3			
	M8.1			
	M8.2			
	M8.3			
	M9			



Optionele domeinen				
Kwaliteitseis		Beschreven in document: (naam document)	Hoofdstuk + paragraaf	Evt. opmerkingen ¹
	M10.1			
	M10.2			
	M10.3			
	M10.4			
3.3.5	B1			

¹ Denk bijvoorbeeld aan paginanummers. Indien het document geen paginanummers bevat, kan de aanbieder alsnog specificeren waar de informatie gevonden kan worden. Voorbeeld: bij een Excel document aangeven welke rijen en kolommen het betreft.



TOELICHTING

Op 1 januari 2023 trad de Regeling beoordelingskader doorstroomtoets PO in werking. Deze Regeling is in 2024 en 2025 gewijzigd na ervaringen met de eerste twee beoordelingsperiodes. Nu wordt de Regeling opnieuw gewijzigd omdat de regeling toetswijzer doorstroomtoets po zal worden gewijzigd per 1 april 2026 (Stcrt. 2023, 33376). Om aan te sluiten bij de gewijzigde toetswijzer moeten onderwijskundige kwaliteitscriteria in het beoordelingskader worden aangepast. Zo zijn bij het terrein Rekenen de vereiste percentages m.b.t. aangeboden context en kale toetsopgaven aangepast. Tevens wordt nu gegarandeerd dat elke leerling bij het terrein Nederlandse taal een minimaal percentage Lezen én een minimaal percentage Taalverzorging toetsopgaven krijgt aangeboden. Ook wordt gegarandeerd dat elke leerling binnen het domein Lezen minimale percentages van de verschillende kenmerken van de taakuitvoering aangeboden krijgt én binnen het domein Taalverzorging minimale percentages van de verschillende aspecten aangeboden krijgt. Vereiste aangeboden percentages werden voor het terrein Rekenen al langer vereist en het terrein Nederlandse taal is hiermee meer synchroon gebracht.

Met deze wijziging wordt verder het beoordelingskader verbeterd naar aanleiding van ervaringen uit de derde beoordelingsperiode uit de zomer van 2025. Zo is nu duidelijker vereist – door één kwaliteits-eis op te delen in meerdere kwaliteitseisen – hoe een toetsaanbieder moet omgaan met calamiteiten én leerlingen die de toets moeten inhalen. De meeste wijzigingen hebben echter betrekking op psychometrische kwaliteitseisen. Zo zijn veel kwaliteitseisen tekstueel verbeterd, terwijl enkele psychometrische kwaliteitseisen ook lichtelijk inhoudelijk zijn aangepast. Tevens is er een extra kwaliteitseis opgenomen die vereist dat toetsaanbieders een inhoudelijke en psychometrische analyse leveren over het functioneren van toetsopgaven tijdens de afgelopen afname, nadat bij de vorige indiening afspraken zijn gemaakt met de toetsaanbieder dat deze toetsopgaven mogen worden afgenomen onder overeengekomen voorwaarden.

De wijzigingen van het beoordelingskader zijn van zodanige omvang, dat ervoor gekozen is om de bijlage te vervangen.

*Het College voor toetsen en examens,
de voorzitter,
J.H. van der Vegt*