

Janneke Lommertzen  
Marian Bruggink  
Ed Smeets



# Validiteit van het toetsproces van examens Nederlands vmbo gl/tl en Duits vwo 2022 eerste tijdvak

Uitgevoerd in opdracht van het College voor Toetsen en Examens

ResearchNed

en  
EXPERTISECENTRUM  
NEDERLANDS

KBA Nijmegen

November 2023

# Onderzoek naar de validiteit van het toetsproces van examens - Nederlands vmbo gl/tl en Duits vwo 2022 eerste tijdvak

Janneke Lommertzen, Marian Bruggink, Ed Smeets

November 2023

© 2023 ResearchNed Nijmegen in samenwerking met Expertisecentrum Nederlands en KBA Nijmegen in opdracht van CvTE

Alle rechten voorbehouden. Het is niet geoorloofd gegevens uit dit rapport te gebruiken in publicaties zonder nauwkeurige bronvermelding. ResearchNed werkt conform de kwaliteitsnormen NEN-EN-ISO 9001:2015 voor kwaliteitsmanagementsystemen, NEN-ISO 20252:2019 voor markt-, opinie- en maatschappelijk onderzoek en NEN-ISO 27001:2017 voor informatiebeveiliging

[Inhoudsopgave](#)

<b>1</b>	<b>Achtergrond</b>	<b>1</b>
1.1	Aanleiding	1
1.2	Validiteit van toetsen	2
1.3	Aanpak	3
1.3.1	Examencyclus Cito en CvTE	4
<b>2</b>	<b>Onderzoekskader</b>	<b>7</b>
2.1	Fase 1. Basisontwerp	8
2.2	Fase 2. Construeren toetsmatrijs	9
2.3	Fase 3. Construeren toets en normeren	10
2.4	Fase 4. Afnemen	12
2.5	Fase 5. Beoordelen, verwerken, analyseren	12
2.6	Fase 6. Registreren en communiceren	13
2.7	Fase 7. Evalueren en verbeteren	13
2.8	Relatie onderzoekskader en examencyclus Cito/CvTE	15
<b>3</b>	<b>Onderzoek validiteit CE Nederlands vmbo gl/tl 2022 eerste tijdvak</b>	<b>16</b>
3.1	Beschrijving van het examen	16
3.1.1	Domeindefinitie	16
3.1.2	Domeinrepresentatie	17
3.1.3	Domeinrelevantie	17
3.2	Samenvatting onderzoek validiteit CE Nederlands vmbo gl/tl 2022 eerste tijdvak	18
3.2.1	Fase 1: Basisontwerp	18
3.2.2	Fase 2: Opstellen toetsspecificaties en toetsmatrijs	18
3.2.3	Fase 3: Constructie en normering	19
3.2.4	Fase 5: Beoordeling, verwerking en analyse	21
3.2.5	Fase 7: Evaluatie	22
3.3	Samenvattend	23
<b>4</b>	<b>Onderzoek validiteit CE Duits vwo 2022 eerste tijdvak</b>	<b>24</b>
4.1	Beschrijving van het examen	24
4.1.1	Domeindefinitie	24
4.1.2	Domeinrepresentatie	24
4.1.3	Domeinrelevantie	25
4.2	Samenvatting onderzoek validiteit CE Duits vwo 2022 eerste tijdvak	25
4.2.1	Fase 1: Basisontwerp	25
4.2.2	Fase 2: Opstellen toetsspecificaties en toetsmatrijs	26
4.2.3	Fase 3: Constructie en Normering	27
4.2.4	Fase 5: Beoordeling, verwerking en analyse	28
4.2.5	Fase 7: Evaluatie	29
4.3	Samenvattend	29
<b>5</b>	<b>Conclusie en aanbevelingen</b>	<b>31</b>
	<b>Literatuur</b>	<b>34</b>
	<b>Bijlage 1: Overzicht van figuren en tabellen</b>	<b>35</b>
	<b>Bijlage 2: Overzicht van geanalyseerde documenten, algemeen en specifiek voor elk examen</b>	<b>36</b>
	<b>Bijlage 3: Checklist Centraal examen Nederlands vmbo gl/tl 2022 eerste tijdvak</b>	<b>38</b>
	<b>Bijlage 4: Checklist Centraal examen Duits vwo 2022 eerste tijdvak</b>	<b>41</b>

# 1 Achtergrond

## 1.1 Aanleiding

Het College voor Toetsen en Examens (CvTE) heeft regie over de examenketen in het voortgezet onderwijs. In de Wet College voor toetsen en examens staat vastgelegd dat Stichting Cito (Cito) en DUO wettelijke ketenpartners zijn. Stichting Cito is de ontwikkelaar van de opgaven en de examens en is verantwoordelijk voor het psychometrische onderzoek. DUO verzorgt de logistiek, doet het drukwerk, biedt de examensoftware Facet aan en is verantwoordelijk voor de distributie van de centrale examens. De scholen zijn verantwoordelijk voor de afnames en de beoordeling van de examens. Daarnaast wordt er nauw samengewerkt met organisaties als SLO (het nationaal expertisecentrum leerplanontwikkeling), de Inspectie van het Onderwijs, de Steunpunten taal en rekenen vo en mbo.<sup>1</sup>

Het College voor Toetsen en Examens (CvTE) bestaat uit een College en een bureau. Daarnaast zijn er veel docenten en vakdeskundigen die in deeltijd meewerken in verschillende vaststellings- of syllabuscommissies. Het College zet de koers uit van het CvTE, en legt verantwoording af aan de minister en/of staatssecretaris van onderwijs. De medewerkers van het bureau van het CvTE zorgen ervoor dat de centrale examens en toetsen tot stand komen zoals bedoeld in de wet, onder meer door sturing te geven aan vaststellingscommissies. In de verschillende vaststellingscommissies worden opgaven die door de verschillende constructiegroepen, onder leiding van Cito zijn gemaakt, op kwaliteit beoordeeld en wordt uiteindelijk het examen in zijn geheel vastgesteld. Deze commissies zijn uiteindelijk ook betrokken bij de normering.<sup>2,3</sup> Daarnaast stelt het CvTE syllabuscommissies samen, die de omschrijving van de te toetsen leerstof verzorgen. De jaarlijkse update van de syllabi wordt door de vaststellingscommissies uitgevoerd. Bij de noodzaak van grote veranderingen wordt een nieuwe syllabuscommissie ingesteld.

Vanuit deze regierol heeft het CvTE aan ResearchNed gevraagd te onderzoeken of de procedures die het CvTE en Cito bij de borging van de validiteit van twee examens moderne talen uit 2022 voldoen aan (inter)nationale kwaliteitsstandaarden. ResearchNed voert dit onderzoek uit als hoofdaannemer, in samenwerking met Expertisecentrum Nederlands en KBA Nijmegen.

Mede naar aanleiding van de Motie Transitie in onderwijstoezicht, ingediend door Jadnanansing<sup>4</sup>, laat CvTE de validiteit van centrale examens periodiek onderzoeken. Zo zijn twee maatschappijgerichte examens uit het eerste tijdvak van 2021 onderzocht door ResearchNed (Geschiedenis vmbo gl/tl en economie havo) in samenwerking met CINOP<sup>5</sup> en uit de jaren 2016-2019 heeft RCEC enkele examens onderzocht.<sup>6</sup>

---

1 <https://www.cvte.nl/over-het-cvte/samenwerking/ketenregie>

2 <https://www.cito.nl/-/media/files/over-cito/cito-informatie-docenten-rol-constructeur-examenopgaven.pdf>

3 <https://www.cvte.nl/over-het-cvte/wie-zijn-wij>

4 Kamerstukken II 2013/14 33905-7

5 Lommertzen, J., Janssen, J. & Heijsters, L. (2022). Onderzoek naar de inhoudsvaliditeit van het toetsproces van twee maatschappijgerichte centrale examens. Onderzoek in opdracht van CvTE. Geraadpleegd op 10 mei 2023 van: [https://www.examenblad.nl/document/inhoudsvaliditeit-van-het/2023/f=/inhoudsvaliditeit\\_toetsproces\\_twee\\_maatschappijgerichte\\_centrale\\_examens.pdf](https://www.examenblad.nl/document/inhoudsvaliditeit-van-het/2023/f=/inhoudsvaliditeit_toetsproces_twee_maatschappijgerichte_centrale_examens.pdf)

6 RCEC onderzocht tussen 2016 en 2019 de volgende examens: Biologie vmbo GL/TL 2016, Duits havo 2016, Engels vwo 2017, Wiskunde vmbo GL/TL 2017 en twee beroepsgerichte examens, te weten: Producteren en Installeren en Energie (PIE) vmbo KB 2019 en Dienstverlening en Producten (D&P) vmbo GL 2019. Zie o.a.: Brouwer, A., Sanders, P., & Veldkamp, B. (2019). Onderzoek naar de inhoudsvaliditeit van een tweetal beroepsgerichte centrale examens voortgezet onderwijs 2019. RCEC - Onderzoek in opdracht van het College voor Toetsen en Examens.



De onderzoeksvraag die CvTE heeft geformuleerd luidt als volgt:

*‘Voldoen de procedures die het CvTE en Cito bij de borging van de (inhouds)validiteit van de examens Nederlands (vmbo gl/tl) en Duits (vwo) hebben gehanteerd aan (inter)nationale kwaliteitsstandaarden?’*

## 1.2 Validiteit van toetsen

Er bestaan verschillende opvattingen over validiteit van toetsen<sup>7,8</sup>. *Inhouds*validiteit betreft de ‘mate waarin de inhoud van het examen overeenstemt met het doel van het examen’. Deze beperktere opvatting van validiteit werd tot voor kort ook gehanteerd in de eerder uitgevoerde onderzoeken naar de centrale examens<sup>6</sup>. ResearchNed heeft in 2021 een instrument ontwikkeld om de validiteit van het toetsproces te onderzoeken dat uitgaat van een bredere opvatting van validiteit. Validiteit gaat immers ook over het *gebruik* en de *interpretatie* van toetsscores. Dit verschil kan worden geïllustreerd aan de hand van een voorbeeld over een toets Franse luistervaardigheid, waar de toetsscore iets zegt over de beheersing van het luisteren naar in het Frans gesproken teksten. En waar de score gebruikt wordt om te beslissen of een leerling kan overgaan naar een volgende klas.<sup>9</sup> Dit laatste aspect van validiteit weegt zwaarder naarmate het belang van de toets toeneemt, zoals bij de centrale examens. Het centraal examen is een zogenaamde ‘*high stakes test*’.

Met andere woorden: validiteit is niet alleen een eigenschap van een toets of een examen, maar hangt nauw samen met waar de toetsscores voor gebruikt worden. Deze opvatting sluit aan bij de Amerikaanse standaarden, ‘*Standards for Educational and Psychological Testing*’ die in 1999 door de American Educational Research Association (AERA), de American Psychological Association (APA) en de National Council on Measurement in Education (NCME) zijn opgesteld<sup>10</sup>. Deze standaarden zeggen het volgende over validiteit:

*‘Because validity depends on the particular uses and interpretations of test scores, the Standards emphasize the need to elaborate the intended uses and interpretations of test scores and to provide evidence relevant to the intended uses and interpretations’<sup>11</sup>.*

In deze opvatting gaat het erom dat toetsontwikkelaars en hun opdrachtgevers de validiteit van de toetsscores voor het beoogde gebruik kunnen onderbouwen door voldoende bewijzen voor validiteit te verzamelen gedurende het toetsontwikkelp proces<sup>12</sup>. In paragraaf 1.2 gaan we hier nader op in.

- 
- 7 Sluijter, C., Hemker, B., & Eggen, Th. (2018b). Beoordelen van de kwaliteit van toetsen en examens, deel 2: de praktijk. Arnhem: Cito: [www.cito.nl/kennis-en-innovatie/kennisplein](http://www.cito.nl/kennis-en-innovatie/kennisplein).
  - 8 Newton, P. E., Shaw, S. D., (2014). Validity and validation. In: P. E. Newton & S. D. Shaw (Eds.) *Validity in Educational & Psychological Assessment* (pp. 1-26). Sage Publications Ltd.
  - 9 Wools, S., (2013). De validiteit van toetsscores. In: P. F. Sanders (Red.). *Toetsen op School* (pp. 69 – 83). Arnhem: Cito.
  - 10 American Educational Research Association, American Psychological Association, National Council on Measurement in Education. (2014). *1999 Standards for Educational and Psychological Testing, 2014 Edition*, Washington, DC: American Educational Research Association
  - 11 Linn, R. L. (2006). The Standards for Educational and Psychological Testing: Guidance in Test Development. In: S. M. Downing & T. M. Haladyna (Eds.). *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
  - 12 Voor een overzicht van historische visies op validiteit, zie: Newton, P. E. & Shaw, S. D. (2014). *Validity And Validation*. In: *Validity in Educational & Psychological Assessment* (pp. 1-26). London: Sage Publications.

### 1.3 Aanpak

Als basis voor dit onderzoek is het onderzoekskader gebruikt dat ResearchNed heeft ontwikkeld in samenwerking met CINOP<sup>13</sup>. Zij onderzochten de validiteit van het toetsproces van twee maatschappijgerichte examens uit 2021. Het gehanteerde onderzoekskader is gebaseerd op bestaande nationale en internationale kwaliteitsstandaarden voor toetsconstructie. Inhoudelijk zijn er veel overeenkomsten tussen deze verschillende kaders.<sup>14,15</sup>

Het onderzoekskader bestaat uit een lijst met criteria waaraan zowel het proces van construeren en vaststellen van toetsen, als de borging van de kwaliteit van toetsen aan moeten voldoen. Deze lijst wordt beschreven in hoofdstuk 2 en zal in hoofdstukken 3 en 4 worden gebruikt als checklist om per fase, per criterium na te gaan of de twee centrale examens voor de vakken Nederlands vmbo-tl/gl en Duits vwo die in dit onderzoek worden onderzocht, voldoen aan de gestelde criteria.

De criteria uit het gehanteerde onderzoekskader zijn niet altijd één op één te vertalen naar de Nederlandse context. Daar waar de internationale standaarden niet passend of voldoende duidelijk zijn, hebben we de criteria op kleine punten aangepast, zoals bijvoorbeeld bij criterium 2.11c (het woord ‘gebruikte’ is vervangen door ‘bestaande’) en criterium 3.9 (‘beoordelen van ontwikkelde items’ is toegevoegd om duidelijk te maken wat de beoordelaars doen in deze fase). Voor criteria die meer aanpassingen vereisen, geven we in hoofdstuk 5 aanbevelingen voor de aanscherping van het onderzoekskader.

De eerste stap in dit onderzoek bestond uit het inventariseren en verzamelen van relevante documenten over de examens. Als leidraad bij deze inventarisatiefase gebruiken we daarbij het schema van de examencyclus van Cito<sup>16</sup>. In bijlage 2 is een overzicht opgenomen van alle verzamelde en geanalyseerde documenten die in dit onderzoek zijn gebruikt.

Als tweede stap, zijn deze bronnen geanalyseerd aan de hand van het geschetste kader. De uitkomsten zijn per examen systematisch beschreven in hoofdstukken 3 en 4. Hierbij is rekening gehouden met de vergelijkbaarheid met eerdere onderzoeken. Er is hetzelfde onderzoekskader gebruikt als bij het onderzoek naar de validiteit van twee maatschappijgerichte examens in 2021, dat aansloot bij de onderzoeken daarvoor<sup>6</sup>. Nadat alle bronnen waren geïnventariseerd en geanalyseerd is het CvTE gevraagd om op enkele punten toelichting, c.q. aanvullende informatie aan te leveren, waarna ook deze input is verwerkt in de checklists en beschrijving van de onderzoeksresultaten.

Deze inventarisatie van de examens aan de hand van het internationale kader kan worden gezien als een formatieve evaluatie van de examens, waarbij is geïnventariseerd in hoeverre de onderzochte examens met de internationale kaders in overeenstemming zijn. Resultaten hiervan, vooral daar waar er mogelijk discrepanties zijn tussen de werkwijzen van het CvTE en Cito en de internationale kaders, kunnen door het CvTE en Cito als input worden gebruikt om eventueel processen te heroverwegen.

---

13 Lommertzen, J., Janssen, J., Heijsters, L. (2022). Onderzoek naar de inhoudsvaliditeit van het toetsproces van twee maatschappijgerichte centrale examens. Nijmegen: ResearchNed, in samenwerking met CINOP, onderzoek in opdracht van het CvTE.

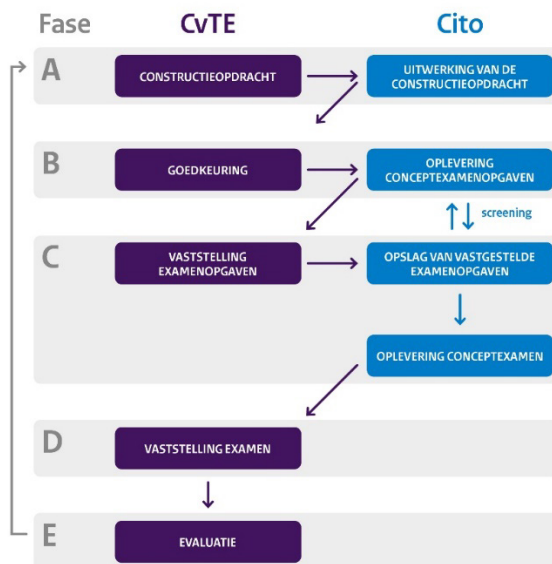
14 Sluijter, C., Hemker, B., & Eggen, Th. (2018a). Beoordelen van de kwaliteit van toetsen en examens, deel 1: Systemen en criteria (pp. 1-8). Arnhem: Cito

15 Wools, S. (2012). Towards a comprehensive Evaluation System for the quality of tests and assessments, in: Psychometrics in Practice at RCEC (pp. 95-106).

16 Schema samenwerking tussen CvTE en Cito voor de constructie en vaststelling van centrale examens vo. Geraadpleegd van: <https://www.examenblad.nl/onderwerp/examencyclus/2023>, op: 17-5-2023.

### 1.3.1 Examencyclus Cito en CvTE

In deze paragraaf worden alle stappen beschreven die CvTE en Cito doorlopen bij de totstandkoming van centrale examens. De totstandkoming van een centraal examen behelst ongeveer twee jaar, onderverdeeld in een cyclus van zes herhalende stappen (A t/m F, zie ook figuur 1.1).<sup>17</sup> Deze examencyclus beschrijft vooral de constructie- en samenstellingsfase.



Figuur 1.1: In schema: samenwerking tussen het CvTE en Cito voor constructie en vaststelling van centrale examens vo.

Bron: <https://www.examenblad.nl/onderwerp/examencyclus/2023>.

#### Examenprogramma

Allereerst worden de examenprogramma's per vak en schooltype door de minister van OCW vastgesteld. Deze examenprogramma's bevatten op hoofdlijnen wat de leerlingen moeten kennen en kunnen.<sup>17</sup>

#### Syllabus

Het CvTE specificeert vervolgens de eisen voor het centraal examen uit het examenprogramma. Hiervoor geeft het CvTE syllabi uit die zijn samengesteld door een commissie van vakinhoudelijke experts (onder andere vakdocenten met ervaring in eindexamenklassen van het betreffende vak waarvoor de syllabus geschreven wordt). De syllabus biedt houvast aan examenconstructeurs, docenten en ontwikkelaars van leermiddelen. De syllabus kan naast de beschrijving van de exameneisen ook verdere informatie over het centraal examen bevatten, zoals bijvoorbeeld nadere specificaties van de examenstof, begrippenlijsten, bekend veronderstelde onderdelen van domeinen die verplicht zijn op het schoolexamen, bekend veronderstelde voorkennis uit de onderbouw, bijzondere vormen van examinering, voorbeeldopgaven, toelichting op de vraagstelling en hulpmiddelen. Nieuwe syllabi worden ontwikkeld door een syllabuscommissie die door het CvTE is ingesteld. Een syllabuscommissie bestaat uit een voorzitter, werkzaam in het naast liggende vervolgonderwijs, docenten uit het voortgezet onderwijs en adviseurs van Cito (toetsdeskundigen) en SLO (curriculumspecialisten).<sup>17, 18</sup>

<sup>17</sup> <https://www.examenblad.nl/onderwerp/examencyclus/2023>

<sup>18</sup> Actuele syllabi zijn te vinden op: [examenblad.nl](https://www.examenblad.nl)

## Constructieopdracht

De syllabus is de basis voor de constructieopdracht die het CvTE aan Cito geeft. Hierin zijn zaken vastgelegd als de verhouding tussen open vragen en gesloten vragen, de verdeling van hoofdthema's en vaardigheden en de hoeveelheid scorepunten die het examen moet tellen. De constructieopdracht bevat de toetsmatrijs (de verdeling van de vragen over de leerdoelen), de mate waarin het examenvragen een reproductief en/of productief karakter moet bevatten, de keuze van teksten en contextmateriaal, de toe te passen vraagvormen en vaardigheidsvragen en de wijze waarop de correctievoorschriften opgesteld dienen te zijn. Op basis van de constructieopdracht inclusief de toetsmatrijs maakt Cito de examenopgaven die in diverse rondes door vakexperts gescreend en beoordeeld worden. Na vaststelling van de examenopgaven door de vaststellingscommissie van het CvTE, doet Cito een voorstel voor de samenstelling van het examen op basis van de toetsmatrijs. Om de vaststelling te vergemakkelijken, levert Cito daartoe een toetsrationale aan. In de toetsrationale staat aangegeven hoe het examen voldoet aan wat in de toetsmatrijs staat.<sup>17</sup>

## Vaststelling

Vervolgens beoordeelt de vaststellingscommissie van het CvTE of dit examen geschikt en passend is als centraal examen en stelt vervolgens het examen en de correctievoorschriften vast onder verantwoordelijkheid van het college.

## Beoordeling en normering

Na afnames en na de beoordelingsrondes (door de examinator en gecommiteerde) van de centrale examens, zijn de scores binnen de scholen bekend. Ze kunnen nog niet aan de kandidaten worden meegedeeld: eerst moet de normering worden vastgesteld, op basis waarvan de directeur van de school vervolgens de cijfers kan vaststellen.

Voor een examen behalen leerlingen een bepaald aantal scorepunten. Door te normeren wordt het aantal scorepunten omgezet in een cijfer met behulp van de N-term die bepaalt hoeveel scorepunten er nodig zijn om een voldoende te behalen voor een examen. De N-term is ingevoerd om te corrigeren voor verschillen in de moeilijkheidsgraad van examens. Bij een moeilijk examen wordt een hoge N-term gehanteerd, en hoeven leerlingen minder scorepunten te behalen voor een voldoende dan bij een gemakkelijker examen. Het normeringsproces bestaat uit een technische fase en een inhoudelijke fase.

Als gevolg van de coronamaatregelen zijn de vaardigheden van de leerlingen die in 2021 en 2022 examens deden niet goed te vergelijken met die van de leerlingen uit de jaren daarvoor<sup>19</sup>. Daarom werd in deze jaren de N-term per examen bepaald aan de hand van de scoreverdeling van alle kandidaten voor dat vak; de gemiddelde cijfers uit de jaren 2015-2019 voor dat vak; een docentenoordeel over de moeilijkheidsgraad van het examen en de N-termen uit 2015-2019. Op basis van deze gegevens volgt een technisch advies voor de N-term die ervoor moet zorgen dat de eisen zoveel mogelijk gelijk, maar in ieder geval niet strenger zijn dan onder normale omstandigheden. Na deze technische analyse vindt vervolgens de inhoudelijke analyse plaats. In deze inhoudelijke fase adviseren docenten in de vaststellingscommissie over de N-term. Zij nemen in hun afweging de uitkomsten van de toets- en itemanalyse (TIA), evenals de feedback en reacties van docenten (via de QuickScan en de examenlijst van CvTE) en leerlingen (via LAKS) mee. In de normeringsvergadering wordt de N-term definitief vastgesteld door het normeringsteam. De uiteindelijke beslissingen worden gedocumenteerd in een verslagformulier. Met behulp van een procedure rondom normbepaling en normhandhaving, zoals

---

19 Normering centrale examens 2022, geraadpleegd op 26 oktober 2023, van: <https://www.examenblad.nl/2022/onderwerpen/normering-centrale-examens>.



gepubliceerd in de Staatscourant d.d. 09-04-2023<sup>20</sup> worden de toetsscores van kandidaten omgezet naar cijfers en wordt de zak/slaaggrens bepaald.<sup>21</sup>

Aan het einde van de cyclus, nadat de examenperiode is afgelopen worden de centrale examens geëvalueerd, samen met vertegenwoordigers van scholen. De QuickScan per vak wordt, evenals de toets- en itemanalyse, openbaar gemaakt aan het eind van de examencyclus en gepubliceerd door Cito.

---

20 Regeling van het College voor Toetsen en Examens van 15 maart 2021, nummer CvTE-21.00446, houdende wijziging van Regeling omzetting scores in cijfers centrale examens en rekentoets VO 2016 in verband met de gevolgen van COVID19 voor de wijze van normering in het examenjaar 2021. In Staatscourant. Nr. 18168, 9 april 2021. Geraadpleegd op 26 oktober 2023 van: <https://zoek.officielebekendmakingen.nl/stcrt-2021-18168.pdf>

21 Normeringsproces Centrale examens vo: geraadpleegd op 26-10-2023 van: [https://www.examenblad.nl/system/files/2020/Infographic\\_Normering\\_Centrale\\_Examens\\_VO.pdf](https://www.examenblad.nl/system/files/2020/Infographic_Normering_Centrale_Examens_VO.pdf)

## 2 Onderzoekskader

Het onderzoekskader bestaat uit een lijst met criteria, waarmee centrale examens kunnen worden beoordeeld op validiteit. Dit kader dient als checklist om te inventariseren welke bewijzen het CvTE en Cito leveren voor de (inhouds)validiteit van de twee onderzochte examens. In het onderzoekskader zijn de kwaliteitsstandaarden ingedeeld in zeven fasen van de toetscyclus (figuur 2.1)<sup>22</sup>. De cyclus begint met de beslissingen omtrent het ontwerp en de inhoud van het examen (doel, functie en inhoud, opgaven) en gaat via de afname, de beoordeling, de verwerking en analyse van scores en de registratie van resultaten (scores) naar het einde van de cyclus toe, waar kritisch wordt teruggeblikt op alle voorgaande fasen. Hierna kunnen eventuele noodzakelijke verbeteringen worden doorgevoerd in een nieuwe cyclus.

Elke fase van de toetscyclus kan op drie niveaus worden bekeken: 1) standaardisen of criteria waaraan moet worden voldaan; 2) methoden om aan deze eisen te voldoen; en 3) mogelijke bewijzen die kunnen worden gebruikt om te controleren of er aan de standaardisen wordt voldaan. De toetscyclus is een leidraad die zichtbaar maakt welke producten er per fase (zouden moeten) worden opgeleverd en met welke hulpmiddelen kwaliteit zichtbaar kan worden gemaakt. Het onderzoekskader in dit onderzoek volgt de toetscyclus om per fase te weer te kunnen geven hoe de toetscyclus van het centraal examen verloopt.



Figuur 2.1: Toetscyclus<sup>23</sup>

In paragrafen 2.1 tot en met 2.7 worden de zeven fasen van het toetsproces nader toegelicht. Bij centrale examens is het vooral relevant hoe er zorg wordt gedragen voor het goed definiëren van het domein, de domeinrepresentatie en -relevantie<sup>24</sup>. Voor die onderdelen richten we ons op de eerste drie fasen van de toetscyclus (Basisontwerp, Construeren toetsmatrijs, Construeren toets/normen) en de wijze waarop Cito in

22 Voor een uitgebreidere beschrijving van de totstandkoming van dit onderzoekskader zie: Lommertzen, J., Janssen, J. & Heijsters, L. (2022). Onderzoek naar de inhoudsvaliditeit van het toetsproces van twee maatschappijgerichte centrale examens. Onderzoek in opdracht van CvTE.

23 Overgenomen uit: Expertgroep BKE/SKE, geciteerd in Sanders 2016, p. 76, maar ook in het Framework of European Standards worden deze zeven stappen gehanteerd.

24 Het onderzoekskader gaat uit van examenontwikkelaars, waar in de Nederlandse situatie worden deze taken grotendeels door Cito en deels door CvTE uitgevoerd, zie paragraaf 1.3.1. CvTE is opdrachtgever aan Cito en stelt vast of de examens voldoen.

deze fasen de bewijzen voor validiteit aannemelijk maakt. Verder is het van belang dat een examen voldoende betrouwbaar is. De bewijzen die daarvoor worden geleverd komen vooral in fase 5 van de toetscyclus aan bod. Tot slot dienen in fase 7 van het toetsproces alle voorgaande stappen en ondersteunende bewijzen op een inzichtelijke wijze te worden gedocumenteerd.

## 2.1 Fase 1. Basisontwerp

Allereerst moet er een zogenaamd toetsplan worden opgesteld, met daarin in ieder geval beschreven: de doelgroep, het meetdoel en het gebruiksdoel. Het specificeren van de doelgroep kan onder andere van belang zijn bij het beoordelen van de inhoud van het examen zoals het taalgebruik en de gehanteerde normen of cesuren. Het meetdoel beschrijft wat de examenkandidaten geacht worden te beheersen. Dit kan de beheersing van een bepaald construct zijn of de beheersing van een exameneenheid van een examenprogramma. Door het formuleren van het meetdoel wordt duidelijk wat wel en wat niet tot het te meten domein wordt gerekend. Het gebruiksdoel bij centrale examens is diplomering. Afhankelijk van de toetsscores van een kandidaat wordt wel of niet een diploma verstrekt.

Volgens de Amerikaanse kwaliteitsstandaarden<sup>25</sup> is het daarnaast gebruikelijk om ook de interpretatie en het gebruik van toetsscores in het plan op te nemen. De examenmaker moet duidelijk aangeven hoe de scores geïnterpreteerd moeten worden door de contexten te beschrijven waarin de scores worden gebruikt. Voor elke beoogde interpretatie is nodig: een motivering, een samenvatting van het bewijs en de theorie met betrekking tot de beoogde interpretatie. Bij de centrale examens is dit eenvoudig, omdat er maar één interpretatie van de toetsscores is; er is een minimaal aantal scorepunten nodig voor een voldoende, waarmee wordt aangegeven dat de kandidaat de examenstof voldoende beheerst.

Verder zouden toetsontwikkelaars conform de Amerikaanse standaarden (standaard 1.2<sup>25</sup>) een toetsplan moeten opstellen waarin de beoogde interpretatie van de toetsscores (i.c. beheersing van een bepaald omschreven leerstofdomein) wordt onderbouwd, gegeven het beoogde gebruik (i.c. de beslissing zakken/slagen). In Nederland worden er verschillende begrippen gebruikt die gezien kunnen worden als onderdeel van het toetsplan. Bijvoorbeeld de *syllabus* en de *constructieopdracht* spelen bij het ontwikkelen van de centrale examens door het CvTE en Cito een cruciale rol. In de *syllabus* specificiert het CvTE op basis van het examenprogramma de eisen voor het centraal examen. In de *constructieopdracht* die het CvTE opstelt voor Cito, is vastgesteld op welke wijze de examenstof uit de syllabus wordt getoetst in het centraal examen. De *constructieopdracht* geeft nadere toetsspecificaties, waarbij de link tussen de examenstof en het examen wordt vastgelegd. De toetsspecificaties gaan onder meer over de inhoud, vorm, de lengte van de toets, psychometrie, scoring en afname van het examen.

Tabel 2.1: Checklist Fase 1: Basisontwerp.

	Checklist Fase 1	Bron
1.1	Er is een toetsplan.	Amerikaanse standaarden: st. 4.1
1.2	In het toetsplan is aangegeven wat de doelgroep van het examen is.	Amerikaanse standaarden: st. 4.1
1.3	In het toetsplan is aangegeven wat het meetdoel het examen is oftewel hoe het domein gedefinieerd is.	Amerikaanse standaarden: st. 4.1
1.4	In het toetsplan is aangegeven wat het gebruiksdoel van het examen is.	Amerikaanse standaarden: st. 4.1
1.5	In het toetsplan is in duidelijke taal gespecificeerd in welke contexten toetsscores gebruikt dienen te worden.	Amerikaanse standaarden: st. 4.1

25 American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014). 1999 Standards for Educational and Psychological Testing, 2014 Edition, Washington, DC: American Educational Research Association

1.6	In het toetsplan is in duidelijke taal gespecificeerd met welke processen de toets moet worden afgenomen en gescoord (bv. papieren of digitale toets, wijze van afname, wijze van beoordeling, hoe komt score tot stand?).	Amerikaanse standaarden: st. 4.2
1.7	In het toetsplan staat de theorie met betrekking tot de beoogde interpretatie beschreven.	Amerikaanse standaarden: st. 1.2

(Bron verwijst naar de herkomst van de criteria)

## 2.2 Fase 2. Construeren toetsmatrijs

Onderdeel van de toetsspecificaties is de toetsmatrijs. Een toetsmatrijs is een schematische weergave van de inhoud van het examen waarin voor de verschillende inhoudscategorieën is vastgelegd op welk niveau de toetsing ervan plaatsvindt. Het is een nuttig hulpmiddel om de beoogde leerdoelen uit te werken tot toetsbare leerdoelen. Voor examenmakers kan het prettig zijn als in de toetsmatrijs percentages staan in plaats van absolute aantallen, dus dat er sprake is van een relatieve toetsmatrijs.

Het is raadzaam om in de toetsmatrijs ook het vereiste beheersingsniveau van het betreffende meetdoel op te nemen. Dit kan worden verduidelijkt door het gebruik van werkwoorden, bijvoorbeeld: kennen, kunnen, toepassen of begrijpen. De taxonomieën van Bloom of Romizowsky zijn veelgebruikt<sup>26</sup>.

Conform de Amerikaanse kwaliteitsstandaarden is het noodzakelijk dat vakinhoudelijke experts de toetsspecificaties beoordelen op hun geschiktheid voor het beoogde gebruik van de toetsscores en de transparantie naar examenkandidaten. Hiermee wordt de inhoudsvaliditeit geborgd. Het doel van de beoordeling, het proces waarmee de beoordeling wordt uitgevoerd en de resultaten van de beoordeling moeten worden gedocumenteerd. Evenals de kwalificaties, relevante ervaringen en demografische kenmerken van de vakinhoudelijke experts.

Tabel 2.2: Checklist Fase 2: Construeren toetsmatrijs

	Checklist Fase 2	Bron
2.1	Er zijn toetsspecificaties (bv. in de vorm van een constructieopdracht)	Amerikaanse standaarden: st. 4.1
2.2	In de toetsspecificaties staan het aantal vragen met bijbehorende scorepunten	Amerikaanse standaarden: st. 4.2
2.3	In de toetsspecificaties staat de toetsvorm en/of het soort vragen (bijvoorbeeld gesloten en/of open)	Amerikaanse standaarden: st. 4.2
2.4	In de toetsspecificaties staan de toegestane hulpmiddelen.	Amerikaanse standaarden: st. 4.2
2.5	In de toetsspecificaties staat de voorgestelde lengte van de toets.	Amerikaanse standaarden: st. 4.2
2.6	In de toetsspecificaties staat de hoeveelheid tijd die is toegestaan voor de toets.	Amerikaanse standaarden: st. 4.2
2.7	In de toetsspecificaties staan de gewenste psychometrische eigenschappen van de items en de toets als geheel.	Amerikaanse standaarden: st. 4.2
2.8	In de toetsspecificaties wordt de volgorde van items gedefinieerd.	Amerikaanse standaarden: st. 4.2
2.9	Er is een toetsmatrijs.	Amerikaanse standaarden: st. 4.2
2.10	In de toetsmatrijs wordt de relevantie van de inhoud van de toets of het examen voor het beoogde meetdoel aannemelijk gemaakt.	Amerikaanse standaarden: st. 4.2
2.11	De toetsmatrijs is een adequate representatie van het meetdoel.	Amerikaanse standaarden: st. 4.2

26 Berkel, H. van, Bax, A. & Joosten-Ten Brinke, D. (Red.) (2013). Toetsen in het hoger onderwijs, 3e herziene druk (p. 6). Houten: Bohn Stafleu van Loghum.

2.11a	De eind- en toetstermen representeren het meetdoel.	Amerikaanse standaarden: st. 4.2
2.11b	De eind- en toetstermen sluiten aan op de inhoud en het vereiste beheersingsniveau van het betreffende meetdoel	Amerikaanse standaarden: st. 4.2
2.11c	Het werkwoordgebruik in de eind- en toetstermen is eenduidig en sluit goed aan bij bestaande taxonomieën (zoals bijvoorbeeld van Bloom of Romiszowsky)	Amerikaanse standaarden: st. 4.2
2.11d	Het aantal vragen geeft een voldoende dekking van het meetdoel.	Amerikaanse standaarden: st. 4.2
2.12	Relevante vakinhoudelijke experts beoordelen de toetsspecificaties op hun geschiktheid.	Amerikaanse standaarden: st. 4.6
2.13	Het doel van deze beoordeling van toetsspecificaties, het proces waarmee deze beoordeling wordt uitgevoerd en de resultaten van deze beoordeling worden gedocumenteerd.	Amerikaanse standaarden: st. 4.6
2.14	De kwalificaties, relevante ervaringen en demografische kenmerken van relevante vakinhoudelijke experts worden gedocumenteerd.	Amerikaanse standaarden: st. 4.6

### 2.3 Fase 3. Construeren toets en normeren

Bij het construeren van het examen moeten eerst opgaven worden ontwikkeld. Vervolgens kan het examen worden samengesteld. Dan moet ook de norm (ofwel cesuur, zak/slaaggrens genoemd) van het examen worden bepaald. Daarnaast dienen regels opgesteld te worden om de toetsscores in cijfers of een waardering om te zetten. Het bepalen van de norm dient volgens de Amerikaanse standaarden te gebeuren met een methode voor standaardbepaling (ook wel ‘standardsetting’) van een examen. Een standaardbepaling dient volgens een vooraf bepaalde procedure te worden uitgevoerd door een onafhankelijk orgaan<sup>27</sup>.

Bij vakken met een nieuw examenprogramma wordt bij Cito een standaardbepaling uitgevoerd door vakexperts (docenten uit het voortgezet onderwijs en/of uit het hoger onderwijs) onder begeleiding van een toetsdeskundige van Cito.<sup>28</sup> Tijdens de standaardbepaling wordt vastgesteld wat kandidaten zouden moeten beheersen en hoeveel punten er moeten worden behaald om te kunnen slagen.<sup>27,28</sup> Vervolgens zijn er ook methoden nodig om de norm te handhaven. Bij de centrale eindexamens gebeurt dit via de procedure met de N-term.<sup>29</sup>

In de (inter)nationale kwaliteitsstandaarden staan aandachtspunten voor het ontwikkelen en evalueren van vragen beschreven. Bewijzen die tonen dat hierop is gelet, tonen vooral de betrouwbaarheid van het examen aan. Daarentegen hangen criteria zoals “De stam bevat geen overbodige informatie (behalve als selectie van informatie het doel is)” en “Het taalniveau is afgestemd op de doelgroep” samen met de vraag of er getoetst wordt wat getoetst moet worden (domeinrepresentatie). Verder is het belangrijk dat kandidaten de vraag kunnen beantwoorden zoals bedoeld. Daarvoor is het essentieel dat de vragen voldoende gedetailleerd zijn en dat er indien nodig vooraf oefenvragen worden verstrekt.

De Amerikaanse standaarden schrijven voor dat de kwalificaties van de personen die de vragen ontwikkelen en op kwaliteit beoordelen, gedocumenteerd moeten worden. Ook moeten de processen die worden gebruikt om hen te trainen en te begeleiden bij deze activiteiten worden vastgelegd.

27 Cizek, J. K. (2006). Standard Setting. In S. M. Downing, & T. M. Haladyna (Eds.), *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

28 Cito: Een kijkje achter de schermen van het normeren: geraadpleegd op 2 november 2023 van: [https://cito.nl/media/0z2imaao/cito\\_cte\\_publicatie-een-kijkje-achter-de-schermen-van-het-normeren.pdf](https://cito.nl/media/0z2imaao/cito_cte_publicatie-een-kijkje-achter-de-schermen-van-het-normeren.pdf).

29 Regeling van het College voor Toetsen en Examens van 30 november 2015, nummer CvTE-15.02159. In: *Staatscourant*, nr. 4817, 4 februari 2016. Geraadpleegd op 1 november van <http://www.stilus.nl/examen/n-term-2016.pdf>



In deze constructiefase is het belangrijk de geconstrueerde vragen op kwaliteit te screenen. Dit kan door de opgaven aan relevante onafhankelijke vakexperts voor te leggen ter beoordeling. Het kan ook (en de Amerikaanse standaarden adviseren dat) door het organiseren van een pretest bij leerlingen, om informatie te verkrijgen over de moeilijkheidsgraad. De Amerikaanse standaarden benadrukken dat het betrekken van vakexperts in deze fase bijdraagt aan de validiteit van een examen. De standaarden beschrijven waar experts in het beoordelingsproces allemaal voor kunnen worden ingezet. Ze kunnen gevraagd worden om de score van de items te controleren en materiaal te identificeren dat mogelijk ongepast, verwarrend of aanstootgevend is voor bepaalde groepen examenkandidaten. Onafhankelijke deskundigen kunnen beoordelen in welke mate de inhoud van de items overeenkomt met de inhoudscategorieën in de toetsspecificaties en of de toets in zijn geheel een evenwichtige dekking van de beoogde inhoud vormt.

Tenslotte adviseren de Amerikaanse standaarden dat er een toetsrationale wordt gemaakt waarin verantwoording over de samenstelling van het examen wordt afgelegd. Deze wordt mogelijk aan het eind van de constructiefase, dus na de samenstelling van het examen, geschreven.

Tabel 2.3: Checklist fase 3: Construeren van de toets en normeren

	Checklist Fase 3	Bron
3.1	De vragen zijn correct geformuleerd	Amerikaanse standaarden: st. 4.7
3.1a	De vragen of opdrachten zijn gestandaardiseerd.	Amerikaanse standaarden: st. 4.7
3.1b	De vragen of opdrachten zijn zodanig ontworpen dat fouten bij invulling voorkomen worden.	Amerikaanse standaarden: st. 4.7
3.1c	De instructie (vaak eerste bladzijde toetsboekje) voor de kandidaat is gestandaardiseerd, volledig en duidelijk.	Amerikaanse standaarden: st. 4.7
3.1d	De instructie voor de kandidaat dient minimaal te bevatten: aantal vragen, wijze waarop antwoorden worden gegeven, (deel) score per vraag of opdracht, de maximaal te behalen score en de cesuur; toegestane hulpmiddelen; beschikbare tijd en wat ingeleverd moet worden bij afronding; beoordelingspunten bij open vragen.	Amerikaanse standaarden: st. 4.16
3.1e	De kwaliteit van de lay-out en vormgeving is in orde.	Amerikaanse standaarden: st. 4.16
3.2	De vragen zijn voldoende gedetailleerd, zodat examenkandidaten de vraag kunnen beantwoorden zoals bedoeld.	Amerikaanse standaarden: st. 4.16
3.3	De moeilijkheidsgraad van de vraag is afgestemd op de beoogde doelgroep.	Amerikaanse standaarden: st. 4.10
3.3a	Er heeft een pretest plaatsgevonden om informatie te verkrijgen over de moeilijkheidsgraad (kwantitatieve evaluatie).	Amerikaanse standaarden: st. 4.10
3.3b	Er heeft een kwalitatieve evaluatie met deskundigen plaatsgevonden om informatie te verkrijgen over de moeilijkheidsgraad.	Amerikaanse standaarden: st. 4.8
3.4	Er wordt een norm(cesuur) verstrekt die via een wetenschappelijke methodiek (standaardbepaling) is bepaald.	Amerikaanse standaarden: st. 5.21
3.5	De manier waarop de norm wordt bepaald is gedocumenteerd.	Amerikaanse standaarden: st. 5.21
3.6	De standaardbepaling is correct uitgevoerd.	Amerikaanse standaarden: st. 5.21
3.6a	De standaardbepalingmethode is op de juiste wijze uitgevoerd.	Amerikaanse standaarden: st. 5.21
3.6b	De vakdeskundigen/experts die bij de standaardbepaling betrokken zijn, zijn naar behoren geselecteerd en getraind.	Amerikaanse standaarden: st. 5.21
3.6c	Er is voldoende overeenstemming tussen de vakdeskundige experts.	Amerikaanse standaarden: st. 5.22
3.7	De procedures die worden gebruikt om items te ontwikkelen, beoordelen en uitproberen en om -indien van toepassing- items uit de vragenbank te selecteren, zijn gedocumenteerd.	Amerikaanse standaarden: st. 4.7
3.8	Er is gedocumenteerd in hoeverre het inhoudsdomen van de toets het domein vertegenwoordigt, dat is gedefinieerd in de toetsspecificaties.	Amerikaanse standaarden: st. 4.12

3.9	Er is gedocumenteerd wat de kwalificaties, relevante ervaringen en demografische kenmerken zijn van degenen die de ontwikkelde items beoordelen, evenals de instructies en training die zij t.b.v. het beoordelingsproces krijgen. <sup>30</sup>	Amerikaanse standaarden: st. 4.8
3.10	Er is een rationale voor de toets waarin verantwoording wordt afgelegd over de samenstelling van de toets of het examen.	Amerikaanse standaarden: st. 1.2

## 2.4 Fase 4. Afnemen

In de (inter)nationale kwaliteitsstandaarden staan verschillende criteria beschreven voor het afnemen van een examen. Deze criteria hebben voornamelijk te maken met de betrouwbaarheid en de transparantie naar examenkandidaten. Maar daar waar het *high stakes* examens betreft, kunnen er in deze fase ook aanzienlijke bedreigingen voor de validiteit van de toetsscores optreden. Bijvoorbeeld wanneer de geheime opgaven voortijdig uitlekken of als er mogelijkheden zijn tot fraude, dan is de validiteit van de toetsscores in het geding. Immers, als een kandidaat van tevoren op de hoogte is van de opgaven, kan hij/zij zich met of zonder externe hulpmiddelen perfect voorbereiden op het geven van een goed antwoord, zonder dat hij/zij daadwerkelijk de betreffende leerstof beheerst. Op basis van de aan ons geleverde documenten is van dergelijke incidenten bij de afname van deze twee examens echter geen sprake geweest. We concluderen daarom dat een check op de criteria voor deze fase verder buiten beschouwing kan blijven.

## 2.5 Fase 5. Beoordelen, verwerken, analyseren

In deze fase van de toetscyclus worden de antwoorden van open vragen beoordeeld door beoordelaar(s), worden de antwoorden op de gesloten meerkeuze vragen (automatisch in het geval van digitale afname of met behulp van schrapkaarten) gescoord. De scores op de examens worden vervolgens geanalyseerd met behulp van psychometrisch onderzoek. Hiermee wordt onderzocht hoe nauwkeurig de score voor kandidaten op het examen is. Bij de centrale examens betekent dit dat de standaardmeetfout wordt berekend en dat gekeken wordt naar de kwaliteit<sup>31</sup> van de individuele opgaven en dat de kwaliteit van het examen als geheel wordt onderzocht op basis van de resultaten van kandidaten. Hoe meer kandidaten, hoe betrouwbaarder dit psychometrische onderzoek is. Het voert te ver om hier diep in te gaan op de verschillende methoden van onderzoek. Belangrijk is dat in alle internationale standaarden wordt genoemd dat de nauwkeurigheid van de toetsscores moet worden onderzocht in de empirie. Afhankelijk van het toetsdoel wordt gekozen voor de meest geschikte uit verschillende methoden om betrouwbaarheid aan te tonen<sup>32</sup>. In de Amerikaanse standaarden (met name standaard 4.13) wordt erop gewezen dat een toetsontwikkelaar, voor zover mogelijk, moet onderzoeken hoe construct-irrelevante variatie in de examens kan worden voorkomen en gereduceerd. Construct-irrelevante variatie betekent dat opgaven niet meten wat ze beogen te meten, en dit is een bedreiging voor de validiteit van de toetsscore. Dit dient eigenlijk alleen te gebeuren als in fase 5 de statistische data aanleiding geven tot verdenking van construct-irrelevante variatie. Het is ook gebruikelijk om in deze fase te analyseren of de hoeveelheid examentijd voldoende was voor de kandidaten.

Volgens de Amerikaanse standaarden moeten er door de examenmaker duidelijke en voldoende gedetailleerde procedures voor het scoren en beoordelingscriteria worden opgesteld. Met het oog op de betrouwbaarheid van de score is het van belang om ervoor te zorgen dat het scoren op een zo duidelijk en nauwkeurig mogelijke manier gebeurt. De beoordeling van de antwoorden op de open opgaven dient op een zo objectief mogelijke manier plaats te vinden en bij *high stakes* examens bovendien door meer dan één

30 Bij het gebruik van empirische analyses gelden andere aanvullende criteria. Zie in de Amerikaanse standaarden de criteria 4.09 t/m 4.11.

31 De diverse parameters van opgaven blijken uit de Toets- en Itemanalyse via het programma TIA-plus dat Cito gebruikt in fase 5. Dit zijn onder meer: de moeilijkheid van een opgave voor de populatie en het vermogen van een opgave om te onderscheiden tussen sterke en zwakke kandidaten. Daarnaast worden de gemiddelde moeilijkheid en de betrouwbaarheid van het examen berekend.

32 Sluijter, C., Hemker, B., & Eggen, Th. (2018b). Beoordelen van de kwaliteit van toetsen en examens, deel 2: de praktijk. Arnhem: Cito.

beoordelaar, onafhankelijk van elkaar. De Amerikaanse standaarden noemen het gebruikelijk om in beoordelingsvoorschriften meerdere voorbeelden van antwoorden op elk scoreniveau te geven (standaard 4.18). Indien van toepassing dienen examenmakers selectiecriteria op te stellen voor degene die verantwoordelijk zijn voor het scoren van de examens. Ook de procedures voor training, kwalificatie en monitoring van beoordelaars moeten worden gedocumenteerd.

In het onderhavige onderzoek met de focus om na te gaan of de examenontwikkelaars bewijzen voor validiteit leveren, is het van belang om in fase 5 na te gaan of en hoe de betrouwbaarheid van het examen wordt berekend en hoe er is omgegaan met de beoordeling van de open vragen.

Tabel 2.4: Checklist Fase 5: Beoordelen, verwerken, analyseren

	Checklist Fase 5	Bron
5.1	Procedures voor het scoren en scorecriteria zijn gedocumenteerd	Amerikaanse standaarden: st. 4.18
5.2	Het proces van het selecteren, trainen, kwalificeren en monitoren van de beoordelaar(s) die scoort (scoren) is gedocumenteerd.	Amerikaanse standaarden: st. 4.20
5.3	Er zijn twee onafhankelijke beoordelaars die het examen apart van elkaar scoren.	Amerikaanse standaarden: st. 4.18
5.4	De twee onafhankelijke scores worden allebei gerapporteerd (n.b. huidige procedure is wettelijke procedure <sup>33</sup> )	Amerikaanse standaarden: st. 6.9
5.5	Er is gedocumenteerd hoe de uiteindelijke score voor het examen wordt berekend.	Amerikaanse standaarden: st. 6.10
5.6	Op elk scoreniveau worden in de beoordelingsvoorschriften meerdere voorbeelden van antwoorden gegeven.	Amerikaanse standaarden: st. 4.18
5.7	Er wordt een toets- en itemanalyse uitgevoerd onder voldoende kandidaten	Amerikaanse standaarden: st. 4.10
5.8	De toets- en itemanalyse maakt de betrouwbaarheid van de toetsscores zichtbaar met behulp van Cronbach's alfa of een andere algemeen geaccepteerde maat voor betrouwbaarheid.	Amerikaanse standaarden: st. 2.3
5.9	De betrouwbaarheid die berekend is geeft geen aanleiding om te twijfelen aan de validiteit van de toetsscores.	Amerikaanse standaarden: st. 2.3
5.10	Indien sprake is van open vragen, wordt via statistisch onderzoek de mate van beoordelaarsovereenstemming onderzocht.	Amerikaanse standaarden: st. 4.20

## 2.6 Fase 6. Registreren en communiceren

In de (inter)nationale kwaliteitsstandaarden staan verschillende criteria beschreven voor het registreren van en communiceren over examenresultaten. Deze criteria hebben echter voornamelijk te maken met de transparantie naar examenkandidaten. Volgens de (inter)nationale standaarden worden in deze fase geen bewijzen geleverd voor validiteit.

## 2.7 Fase 7. Evalueren en verbeteren

Voor het evalueren en verbeteren is het belangrijk dat de stappen die tijdens het ontwerp- en ontwikkelingsproces zijn genomen om bewijs te leveren voor validiteit en betrouwbaarheid, worden gedocumenteerd. Volgens de Amerikaanse standaarden zijn de ondersteunende documenten voor examens het belangrijkste middel waarmee examenmakers met kandidaten communiceren. Het is van belang dat deze documenten worden beoordeeld op hun volledigheid, nauwkeurigheid en duidelijkheid.

33 Eindexamenbesluit VO art. 41 en 42.

Tabel 2.5: Checklist Fase 7: Evalueren en verbeteren

	Checklist Fase 7	Bron
7.1	Toetsspecificaties worden aangepast als er significante veranderingen zijn in het vertegenwoordigde domein.	Amerikaanse standaarden: st. 4.24
7.2	Bewijzen voor (inhouds)validiteit zijn vastgelegd	Amerikaanse standaarden: st. 4.0
7.3	Bewijzen voor betrouwbaarheid (waaronder de mate van beoordelaarsovereenstemming) zijn vastgelegd	Amerikaanse standaarden: st. 4.0
7.4	Procedures rondom toetsontwikkeling zijn vastgelegd	Amerikaanse standaarden: st. 4.0
7.5	Rollen, verantwoordelijkheden en taakverdeling rondom toetsontwikkeling zijn vastgelegd	Amerikaanse standaarden: st. 4.0
7.6	Er wordt onderzoek gedaan onder de stakeholders (belangenvertegenwoordigers van kandidaten en/of van opleiders) die zijn betrokken bij het gebruik van de toets.	Amerikaanse standaarden: st. 1.2
7.7	Er is in een verslag (jaarrapport, evaluatieverslag) waarin verantwoording is afgelegd over de ontwikkeling, samenstelling en afname van de toets.	Amerikaanse standaarden: st. 7.4

## 2.8 Relatie onderzoekskader en examencyclus Cito/CvTE

Zoals eerder beschreven in paragraaf 1.3.1 hanteren Cito en CvTE een examencyclus die uit vijf fasen bestaat. In tabel 2.6 staat weergegeven hoe de zeven fasen uit de toetscyclus, waar het onderzoekskader op is gebaseerd, zich verhouden tot de examencyclus van Cito en CvTE.

Tabel 2.6: Relatie tussen de fasen van de toetscyclus en examencyclus van Cito en het CvTE.

Toetscyclus	Examencyclus Cito/CvTE
1. Basisontwerp	A: Constructieopdracht
2. Constructie toetsmatrijs	A & B: uitwerking constructieopdracht
3. Constructie opgaven, samenstelling van examen en normering	B & C: Screening concept examenopgaven, C: vaststelling examenopgaven, oplevering conceptexamen en D: vaststelling examen en evt. een standaardbepaling op een referentie examen.
4. Afname	
5. Beoordeling, verwerking analyse	E: Evaluatie
6. Registratie en communicatie van resultaten	E: Evaluatie
7. Evaluatie en verbeteringen	E: Evaluatie



## 3 Onderzoek validiteit CE Nederlands vmbo gl/tl 2022 eerste tijdvak

### 3.1 Beschrijving van het examen

Het centraal examen Nederlands vmbo gl/tl 2022 eerste tijdvak is een papieren examen, bestaande uit een leesvaardigheidsgedeelte en een schrijfvaardigheidsgedeelte. Het leesvaardigheidsgedeelte van het examen bestaat uit vier teksten met in totaal 36 vragen (5 open, 22 meerkeuze en 9 voorgestructureerde vragen). Het schrijfvaardigheidsgedeelte van het examen bestaat uit één schrijfopdracht met 11 inhoudselementen. Voor dit examen zijn maximaal 53 punten te behalen. De afname van het examen vond plaats op maandag 16 mei 2022 tussen 13:30 en 15:30 uur.

#### 3.1.1 Domeindefinitie

De examenstof voor het centraal examen Nederlands vmbo gl/tl 2022 bestaat uit de volgende exameneenheden: leervaardigheden voor het vak Nederlands (NE/K/3), leesvaardigheid (NE/K/6) en schrijfvaardigheid (NE/K/7). In de syllabus worden deze exameneenheden verder gespecificeerd met de volgende eindtermen:

- NE/K/3 Leervaardigheden voor het vak Nederlands:

De kandidaat beheerst een aantal strategische vaardigheden die bijdragen tot de ontwikkeling van het eigen leervermogen. De kandidaat kan: 1) strategieën kiezen en hanteren die afgestemd zijn op het bereiken van verschillende lees-, schrijf-, luister- en kijk- en spreek- en gespreksdoelen, 2) strategieën kiezen die het eigen taalproces bevorderen, 3) compenserende strategieën kiezen en hanteren wanneer de eigen taal- of communicatieve kennis tekortschiet.

- NE/K/6 Leesvaardigheid

- Onderwerpen van de teksten: de teksten hebben betrekking op alledaagse onderwerpen en onderwerpen die iets verder van de leerling afstaan. Ze sluiten aan bij de intellectuele en emotionele ontwikkeling van de kandidaten. Bij de tekstkeuze is rekening gehouden met de variatie in culturele achtergrond van de kandidaten.
- Tekstkenmerken: de teksten voor gl/tl-kandidaten zijn ten opzichte van de teksten voor de kb-kandidaten wat complexer qua woordenschat, zinsbouw en opbouw. De teksten hebben een heldere structuur: verbanden worden duidelijk aangegeven. De teksten hebben overwegend een lage informatiedichtheid en ze zijn niet te lang.
- Tekstsoorten: informatieve teksten, instructieve teksten en betogende teksten.
- Opdrachten: de opdrachten bieden de kandidaten steun door een zekere mate van voorstructurering.
- De kandidaat kan: 1) leesstrategieën hanteren, 2) compenserende strategieën gebruiken, 3) de functie van beeld en opmaak in een tekst herkennen, 4) het schrijfdoel van de auteur aangeven, 5) een tekst indelen in betekenisvolle eenheden en de relaties tussen die eenheden benoemen, 6) van een tekst het hoofdonderwerp en de hoofdgedachte aangeven, 7) uit een tekst de betekenis afleiden van een voor de strekking van de tekst belangrijk woord, belangrijke woordgroep of zin 8) in verschillende tekstsoorten de hoofdzaken herkennen, onderscheiden en ordenen, 9) op eenvoudig niveau verschillende tekstrelaties herkennen, 10) talige middelen herkennen die een schrijver hanteert om zijn of haar doel te bereiken en 11) een oordeel geven over de tekst en dit toelichten op grond van aanwijsbare tekstgegevens.

- NE/K/7 Schrijfvaardigheid
  - Opdrachten: de opdrachten hebben betrekking op herkenbare onderwerpen van maatschappelijke en beroepsmatige aard. De opdrachten bieden de kandidaten steun door een zekere mate van sturing met betrekking tot doel, publiek, vorm en inhoud van de te produceren teksten.
  - Kenmerken van de taakuitvoering: de teksten van de kandidaten brengen het beoogde schrijfdoel tot uitdrukking en zijn qua woordkeuze, toon en zinsbouw afgestemd op het beoogde publiek. In de teksten is structuur aangebracht zowel op tekst- als op alinea-niveau. De teksten voldoen aan de eisen op het gebied van spelling, interpunctie, grammaticaliteit en uiterlijke verzorging.
  - Tekstsoorten: de kandidaten schrijven informatieve, overtuigende, opiniërende en activerende teksten.
  - De kandidaat kan: 1) schrijfstrategieën hanteren 2) compenserende strategieën kiezen en hanteren 3) het schrijfdoel in teksten tot uitdrukking brengen 4) het schrijfdoel en taalgebruik richten op verschillende soorten lezerspubliek, 5) conventies hanteren met betrekking tot tekstsoorten, tekst- en alineaopbouw, spelling en interpunctie en uiterlijke verzorging en 6) beschikbare elektronische hulpmiddelen in het schrijfproces gebruiken.

### 3.1.2 Domeinrepresentatie

Domeinrepresentatie betreft de mate waarin een examen het domein zoals dat gedefinieerd is door de toetsspecificaties en in de toetsmatrijs adequaat representeert en meet. Het behoort tot de expertise van de vakdeskundigen in de vaststellingscommissie om te bepalen of de opgaven in voldoende mate het beoogde domein representeren. De toetsdeskundige van Cito stelt voorafgaand aan deze beoordeling de toetsrationale op, zodat de vaststellingscommissie kan vergelijken of het daadwerkelijk samengestelde examen voldoet aan de toetsmatrijs qua verdeling van de vragen over de leerstofonderdelen. Het examen Nederlands vmbo gl/tl 2022 bestaat uit vier teksten met gesloten en open vragen en een schrijfopdracht. In de toetsmatrijs voor het onderdeel lezen is te zien hoe de beoogde verdeling van scorepunten is over vier verschillende categorieën. Een gerealiseerde verdeling van items over de verschillende categorieën is te vinden in de toetsrationale. Voor het onderdeel schrijven zijn ook categorieën gedefinieerd, maar wordt vanwege de aard van de opdracht geen beoogde -en dus ook geen gerealiseerde- verdeling over de specificaties gegeven.

### 3.1.3 Domeinrelevantie

Domeinrelevantie betreft de mate waarin elke opgave van een examen relevant is voor het beoogde domein. Ook om de relevantie van opgaven te bepalen, is de inzet van vakdeskundigen nodig. In de constructiefase gebeurt dit doordat screeners en constructiegroepleden van Cito zijn betrokken bij de constructie van de opgaven. En ook de vakdeskundigen van de vaststellingscommissie kijken hier grondig naar: ze keuren opgaven die in hun ogen niet relevant zijn af. Voor het examen Nederlands vmbo gl/tl 2022 zijn alle opgaven door de vaststellingcommissie zonder voorbehoud vastgesteld.

Voor alle examens geldt dat er na afname ook gegevens van andere vakexperts beschikbaar zijn over het examen. Middels een enquête (de QuickScan) hebben docenten/ examinatoren Nederlands hun waardering voor het examen Nederlands vmbo gl/tl 2022 uitgesproken. In deze QuickScan werd hun oordeel gevraagd over de moeilijkheidsgraad van het examen, de lengte van het examen en de aansluiting bij het gegeven onderwijs. Ook werd er gevraagd een cijfer te geven aan het examen.

De resultaten van 1672 respondenten van 743 scholen waren als volgt: Een meerderheid vond het examen niet te moeilijk/niet te makkelijk (58%), gevolgd door 37 procent die het examen moeilijk vond. De groepen docenten die aangaven dat het examen te moeilijk, of te makkelijk was, waren even groot (beide 2%). Ruim de helft (53%) vond het examen te lang en de rest (46%) vond het examen qua lengte precies goed. De ruime meerderheid van de docenten vond de aansluiting van het examen bij het gegeven onderwijs

voldoende tot (zeer) goed (54% voldoende, 39% goed, 2% zeer goed); vijf procent vond de aansluiting onvoldoende. Ook gaf de meerderheid (91%) het examen een 6 of hoger (35% een 6; 45% een 7; 11% een 8); zeven procent van de docenten gaf het examen een 5 en de overige respondenten gaven het examen een 4 (2%) of 3 (1%).

## 3.2 Samenvatting onderzoek validiteit CE Nederlands vmbo gl/tl 2022 eerste tijdvak

Aan de hand van het onderzoekskader, beschreven in hoofdstuk 2, is het centraal examen Nederlands vmbo gl/tl uit het eerste tijdvak van 2022 geanalyseerd. In deze paragraaf beschrijven we de belangrijkste bevindingen. Gedetailleerde bevindingen per criterium staan beschreven in Bijlage 3 'Checklist Centraal Examen Nederlands vmbo gl/tl 2022 eerste tijdvak'. Fasen 4 en 6 uit het toetsproces ('Afname' en 'Registratie en communicatie') worden hier buiten beschouwing gelaten.

### 3.2.1 Fase 1: Basisontwerp

In fase 1 van het onderzoekskader staat dat er een toetsplan moet zijn en er staat beschreven aan welke kenmerken het toetsplan moet voldoen. De term 'toetsplan' wordt door het CvTE echter niet gebruikt, maar verschillende documenten, zoals het examenprogramma, de syllabus en de constructieopdracht, omvatten de informatie uit het toetsplan. Aan alle gestelde criteria voor een toetsplan (1.2 - 1.7) wordt voldaan. De doelgroep van het examen staat aangegeven, het domein en het gebruiksdoel van het examen is gedefinieerd en het is duidelijk in welke contexten de toetsscores gebruikt moeten worden. In het Examenbesluit is terug te vinden in welke context scores gebruikt moeten worden en in diverse documenten, zoals de constructie-opdracht en de Septembermededeling, wordt gespecificeerd met welke processen het examen moet worden afgenomen en gescoord. In de syllabus staan de specificaties van de globale eindtermen beschreven. In dit document wordt verwezen naar het wettelijk kader Referentieniveaus taal, die bij de uitvoering van het eindexamenprogramma in acht worden genomen.

### 3.2.2 Fase 2: Opstellen toetsspecificaties en toetsmatrijs

Fase 2 van het onderzoekskader gaat over het opstellen van toetsspecificaties en een toetsmatrijs. De constructieopdracht bevat de toetsmatrijs voor lees- en schrijfvaardigheid en beschrijft omvang, tijdsduur en hulpmiddelen, vakspecifieke scoringsregels (criterium 2.1). In de syllabus Nederlands VMBO (versie 2, 2020) staan de exameneenheden beschreven. De exameneenheid 'leervaardigheid' (NE/K/3) wordt alleen impliciet getoetst. Dat wil zeggen dat er geen opdrachten zijn die hier expliciet naar vragen, maar dat de leerling die vaardigheden moet gebruiken om andere opdrachten op het centraal examen te kunnen uitvoeren.

Aan de criteria 2.2 tot en met 2.6 wordt voldaan. Specificaties zoals het aantal vragen met bijbehorende scoringspunten, de toetsvorm, het soort vragen, de toegestane hulpmiddelen, de lengte van de toets en de toegestane hoeveelheid tijd voor het maken van de toets staan allemaal beschreven in de constructieopdracht.

Aan criterium 2.7 wordt niet voldaan, de gewenste psychometrische eigenschappen van de items en de toets als geheel staan niet beschreven in de toetsspecificaties. Hier is niet vermeld wat de gewenste betrouwbaarheid is, hoeveel gemakkelijke en hoeveel moeilijke vragen dienen te worden opgenomen en wat de gewenste p-waarde van de items is. Wel geeft Cito geeft aan dat bij de constructie van een examen een p-waarde van .50 het uitgangspunt is.<sup>34</sup> Wat betreft criterium 2.8, het definiëren van de volgorde van items, geldt dat dit voor schrijven niet van toepassing is. Er wordt immers gevraagd om één grotere schrijfopdracht te maken. Voor het onderdeel lezen geeft CvTE aan dat de volgorde van teksten bepaald

34 Cito: Waaraan moet een goede toetsvraag voldoen? Geraadpleegd op 9-11-2023 van: <https://cito.nl/centrale-toetsen-en-examens/centrale-examens-vo/onze-expertise/waaraan-moet-een-goede-toetsvraag-voldoen/>

wordt aan de hand van de moeilijkheidsgraad van teksten (niet eerst de moeilijkste tekst bijvoorbeeld) of het onderwerp van de tekst. Het lijkt ons gewenst om deze keuze in volgorde te documenteren. Wat betreft de volgorde van de items, lijkt het ons logisch om vragen over tekstdelen te stellen in de volgorde van de tekst, maar verder is een volgorde definiëren volgens ons bij dit domein niet nodig.

Criteria 2.9 tot en met 2.11 (a-d) gaan over de toetsmatrijs. Deze is opgenomen in de constructieopdracht. Hierin zijn specificaties van lees- en schrijfvaardigheid opgenomen die -volgens de experts- passen bij het referentiekader en het te meten einddoel. Het werkwoordgebruik in de eindtermen is eenduidig en hoewel niet expliciet verwezen wordt naar een taxonomie, zoals bijvoorbeeld van Bloom of Romiszowsky, zien we in de toetsmatrijs werkwoorden terug die passend zijn bij de verschillende niveaus (taxonomieën) van Bloom, (zoals benoemen, herkennen, toelichten, afleiden, hanteren, gebruiken, oordeel geven over en onderscheiden)

Aan de criteria rondom het beoordelen van de toetsspecificaties door experts (criteria 2.12 - 2.14) wordt grotendeels voldaan. In het algemene gedeelte van de constructieopdracht staat beschreven hoe de constructieopdracht met de specificaties tot stand komt. Deze wordt opgesteld door de clustermanager van het CvTE. De clustermanager vraagt advies aan zowel de (voorzitter van de) vaststellingscommissie als de betrokken toetsdeskundigen en clustermanagers van Cito. Cito voert tevens een controle uit op de uitvoerbaarheid van de constructieopdracht. Uiteindelijk stelt de clustermanager CvTE de constructieopdrachten vast. Het CvTE geeft aan dat er jaarlijks een check plaatsvindt of leden van de vaststellingscommissie nog voldoen aan de formele eis dat ze les geven aan de betreffende examenklassen. De beoordeling of een docent relevante ervaring heeft, vindt plaats bij de werving door de vakvereniging of de vakbond. De kwalificaties van de vaststellingscommissie staan beschreven in een zeer verouderd document ('huishoudelijk reglement CEVO'). Daarnaast worden vereiste kwalificaties van leden van de vaststellingscommissie beschreven in een brief aan het bestuur van de sectie Nederlands van de VLLT met het verzoek om een lid voor de vaststellingscommissie Nederlands vmbo voor te dragen.

Wat betreft deze fase bevelen we aan om de gewenste psychometrische eigenschappen van de items en van de toets als geheel vooraf vast te leggen. Ook is de aanbeveling om de keuze in tekstvolgorde bij het onderdeel lezen, te documenteren. Verder is ons advies om de kwalificaties van alle betrokken experts bij het opstellen van de toetsspecificaties vast te leggen in een centraal competentieprofiel.

### 3.2.3 Fase 3: Constructie en normering

Fase 3 betreft de constructie van het examen en de normering. Het eerste criterium (de vragen zijn correct geformuleerd) bestaat uit een aantal richtlijnen (3.1a-3.1e). Aan vrijwel alle richtlijnen wordt voldaan. In het algemene deel van de constructieopdracht staat hoe de constructieopdracht wordt uitgewerkt en staat uitgebreid beschreven hoe de screening van het examen voor en na de vaststelling plaatsvindt om tot een definitief examen te komen. In het 'Handboek centrale examens vo' zijn activiteiten en afspraken vastgelegd, die betrekking hebben op het leveren en verbeteren van de kwaliteit van producten, processen en organisatie rondom examens. Ook zijn er checklists beschikbaar voor het construeren van open en gesloten vragen. De instructie voor de kandidaat is gestandaardiseerd en helder; het enige dat lijkt te ontbreken op het opgavenboekje is een beschrijving van de toegestane hulpmiddelen. Dit staat wel in andere documenten beschreven zoals het document "Regeling toegestane hulpmiddelen voor de centrale examens van de eindexamens en de staatsexamens vwo, havo en vmbo" (zie Examenblad.nl) en ook op mijnneindexamen.nl. Daarnaast is de cesuur niet opgenomen in de instructie voor de kandidaat; deze wordt om begrijpelijke redenen na afloop bekend gemaakt. De lay-out en vormgeving van de examenmaterialen zijn conform de specificaties in het document 'richtlijnen lay-out opgavenboekje'.

Volgens de betrokken experts zijn de vragen in het opgavenboekje voldoende gedetailleerd (criterium 3.2). Criterium 3.3 gaat over de moeilijkheidsgraad van het examen. Er heeft voor dit examen geen pretest plaatsgevonden. CvTE geeft aan dat een pretest bij het examen Nederlands vmbo gl/tl ook wat complexer

is vanwege de schrijfpodracht. Wel heeft er een kwalitatieve evaluatie met deskundigen plaatsgevonden om informatie te verkrijgen over de moeilijkheidsgraad. In het document 'Fase D: de vaststelling van een heel examen' is door vier leden van de vaststellingscommissie de moeilijkheid van het examen vergeleken met het referentie-examen. Drie van de vier leden stellen daarin een gelijk niveau vast; één lid stelt een makkelijker niveau vast. Onze inschatting is dat het niet reëel is om elk jaar een pretest af te nemen en dat een vergelijking met een referentie-examen voldoende is om vooraf de moeilijkheidsgraad van het examen te bepalen.

Criteria 3.4 en 3.5 gaan over de norm/cesuur. Volgens de Amerikaanse standaarden wordt deze voor afname bepaald en in het opgavenboekje vermeld, zodat de kandidaat weet hoeveel vragen hij of zij goed moet hebben om het examen te behalen. De normhandhaving van het centraal examen in Nederland gebeurt echter na afname van het examen via de procedure met de N-term.<sup>35</sup> In het document 'normering vmbo' wordt medegedeeld wat de definitieve normeringstermen (N-termen) voor de algemene vakken vmbo zijn. Voor tijdvak 1 is voor Nederlands vmbo gl/tl geen standaardbepaling uitgevoerd (criterium 3.6). De moeilijkheidsschatting door de vaststellingcommissie (ook een vorm van standaardbepaling) is wel uitgevoerd (zie criterium 3.3). CvTE geeft aan dat de referentiepopulaties en de examenpopulaties van 2022 niet goed vergelijkbaar waren, vanwege de onderwijsomstandigheden en voorbereiding op het examen tijdens de coronapandemie (zoals beschreven in paragraaf 1.3.1). Daarom zijn in 2022 ook de historische N-termen meegenomen bij het vaststellen van de N-term (zie document 'Nadere uitwerking normering 2022') en een docentenoordeel.

Aan criteria 3.7 en 3.8 wordt voldaan; in het algemene deel van de constructieopdracht en in het handboek worden de procedures om items te ontwikkelen en te beoordelen met daarbij het productiepad uitgebreid toegelicht. In de toetsmatrijs wordt duidelijk gemaakt hoe het inhoudsdomen van de toets het domein vertegenwoordigt.

Criterium 3.9 gaat over de kwalificaties, relevante ervaringen en demografische kenmerken van beoordelaars van de ontwikkelde items en over de instructies en training die deze beoordelaars t.b.v. het beoordelingsproces krijgen. In de vacaturetekst voor een nieuw lid van de vaststellingscommissie Nederlands vmbo worden enkele vereisten genoemd, zoals 'is bevoegd voor het geven van onderwijs in het vak Nederlands vmbo, beschikt over een goede vakkennis en -vaardigheid op het gebied van het vak Nederlands vmbo en heeft (recente) ervaring met het geven van onderwijs in het vak Nederlands in examenklassen in het vmbo, bij voorkeur ook in bb en kb'. Leden worden voorgedragen door de vakbond of vakvereniging. CvTE gaat daarbij af op de deskundigheid van de vakbond of vakvereniging. CvTE geeft aan dat de leden jaarlijks worden gemonitord en er voortgangsgesprekken worden gevoerd. Verder laat CvTE weten dat de vakdeskundigen/experts van Cito toetsdeskundigen zijn en bij Stichting Cito in dienst zijn. De deskundigen Nederlands vmbo gl/tl zijn afgestudeerd in de taal en hebben een opleiding gekregen als toetsdeskundige. Over de instructie en training van deze deskundigen zijn documenten beschikbaar zoals 'Handreiking Taalgebruik in Centrale Examens' en de checklists voor open vragen en meerkeuzevragen. Cito is verantwoordelijk voor het bewaken van de expertise van de constructeurs en toetsdeskundigen. De samenwerking met, en de kwaliteit van de toetsdeskundigen is onderdeel van het jaarlijkse gesprek van de clustermanager met de voorzitter van de vaststellingscommissie. Ook in de loop van het jaar verwacht het CvTE dat de voorzitter zorgen doorgeeft aan de clustermanager, die vervolgens contact opneemt met de leidinggevende van de toetsdeskundige.

Aan het laatste criterium in deze fase (criterium 3.10) over de aanwezigheid van een toetsrationale, wordt ook voldaan. In de toetsrationale staat aangegeven welke vragen bij welke categorie van het domein lezen horen. De vergelijking met de beoogde verdeling uit de toetsmatrijs ontbreekt daarbij, maar als we de toetsmatrijs naast de toetsrationale leggen, dan komen de verdelingen over de categorieën redelijk goed overeen (< minder dan 10% verschil).

---

35 We bevelen dan ook aan het onderzoekskader op dit punt aan te passen en dit criterium voor toekomstige analyses op te nemen in fase 5 van de toetscyclus.



Hoewel de toegestane hulpmiddelen in andere documenten helder worden beschreven, is ons advies om deze ook op het opgavenboekje/de uitwerkbijlage te vermelden om er zeker van te zijn dat alle examenkandidaten dezelfde hulpmiddelen tot hun beschikking kunnen hebben. Daarnaast bevelen we aan om net als voor de experts die betrokken zijn bij het opstellen van de toetsspecificaties ook de kenmerken van de vak- en toetsdeskundigen die betrokken zijn bij het construeren van het examen vast te leggen in een centraal competentieprofiel.

### 3.2.4 Fase 5: Beoordeling, verwerking en analyse

Fase 5 betreft de beoordeling, verwerking en analyse van de examens na de afname. Het eerste criterium (5.1) gaat over het documenteren van de procedures voor het scoren en de scorecriteria. De procedures en scorecriteria zijn duidelijk vastgelegd in het correctievoorschrift. In de Septembermededeling staat dat in 2022 een testcorrectie zal plaatsvinden voor Nederlands vmbo gl/tl. Dat betekent dat een aantal daartoe uitgenodigde docenten hun leerlingenwerk direct na het centraal examen nakijkt met behulp van een voorlopig correctievoorschrift, dat alleen aan hen op de dag van afname beschikbaar wordt gesteld. De volgende dag komen deze docenten samen om met elkaar het leerlingenwerk te bespreken en op basis daarvan aanpassingen te adviseren op het voorlopig correctievoorschrift. De vaststellingscommissie van het CvTE bepaalt daarna welke aanpassingen wel of niet worden overgenomen. Dit is vastgelegd in het document 'Wijzigingen correctievoorschrift naar aanleiding van vaststellingscommissie vergadering'. Het proces van beoordelen door twee onafhankelijke beoordelaars is bij wettelijk besluit vastgelegd (criteria 5.2 en 5.3). De twee onafhankelijke scores worden echter niet beide gerapporteerd (criterium 5.4). Het CvTE geeft aan dat voor alle examens geldt dat de beoordelaarsovereenstemming niet wordt vastgelegd. Het is namelijk niet verplicht om de tweede correctie in het uitwisselingsprogramma 'Wolf' te doen. Hoewel Cito wel de service aanbiedt om dit in Wolf te doen, kan het CvTE die informatie nog niet gebruiken in verband met de privacywetgeving. Vooralsnog wordt eens in de zoveel jaar aan de hand van de scores die scholen in BRON invoeren, onderzocht in hoeverre de tweede correctie een significant effect gehad zou hebben op de normering. Tot nu toe is dat niet het geval volgens CvTE.

Aan de criteria 5.5 tot en met 5.8 wordt voldaan. Het gaat hier om het documenteren van de berekening van de uiteindelijke score in de omzettingstabel voor normering (criterium 5.5). In het beoordelingsmodel voor de open vragen wordt in het correctievoorschrift de kern van een goed antwoord beschreven en worden er elementen gegeven die in het antwoord moeten terugkomen. Voor schrijven staat in het correctievoorschrift beschreven welke criteria voor puntenaftrek er gelden voor wat betreft inhoud, taalgebruik en presentatie/conventies (criterium 5.6). Er is een Toets- en itemanalyse (TIA) uitgevoerd onder 43.835 kandidaten (de kandidaten waarvan de resultaten zijn ingevoerd in Wolf) (criterium 5.7). In de TIA staat de betrouwbaarheid als volgt vermeld: gemiddelde P-waarde: 60,59; Gemiddelde Rit: 0,23; Coëfficiënt Alpha: 0,54 (criterium 5.8).

Wat betreft criterium 5.9 (betrouwbaarheid van de toetsscores), is de coëfficiënt Alpha vrij laag te noemen (0,54). In de evaluatie worden echter geen aanwijsbare zwakke punten genoemd en er zijn bij het examen Nederlands vmbo gl/tl zeer weinig vragen met extreem lage p-waarden. Het CvTE geeft aan dat verschillende pilots om de betrouwbaarheid van het examen Nederlands vmbo te verbeteren tot niets hebben geleid en dat dit alles te maken heeft met het feit dat er binnen één examen twee verschillende vaardigheden worden getoetst. Daardoor bevat het examen een beperkt aantal teksten met opgaven voor lezen en slechts één opdracht voor schrijven.

Bij criterium 5.10 gaat het over het onderzoeken van de beoordelaarsovereenstemming. Zoals eerder aangegeven bij criterium 5.4 worden de twee onafhankelijke scores niet beide gerapporteerd en denkt het CvTE na in hoeverre het vaststellen van de beoordelaarsovereenstemming wenselijk/mogelijk is. De beoordelaarsovereenstemming is vooral bij een open taak als de schrijftaak van belang. De beoordeling van de schrijfpodracht blijft echter jaarlijks discussie opleveren. Uit een enquête onder docenten die CvTE

heeft gehouden, concludeert het CvTE dat het opstellen van nadere richtlijnen de onvrede onder docenten niet zal doen afnemen en hoogstens zal verplaatsen naar andere docenten.

Het CvTE is zich bewust van de lage betrouwbaarheid van het examen Nederlands en weet de oorzaak daarvoor aan te wijzen. Het toetsen van twee vaardigheden in één examen heeft als gevolg dat er een beperking zit aan het aantal teksten en opdrachten. Deze beperking maakt het lastig om de betrouwbaarheid hoger te krijgen volgens het CvTE. Een mogelijke oplossing zou zijn om de twee vaardigheden in aparte examens te toetsen, één voor Nederlands Lezen en één voor Nederlands Schrijven. Hierdoor kan zowel het aantal teksten en opgaven van het examen Lezen als het aantal schrijfoopdrachten uitgebreid worden. Dit zal de betrouwbaarheid ten goede komen. Wat betreft het betrouwbaar beoordelen van de schrijftaak, beseffen we ons dat het domein 'schrijven' een zeer lastig te beoordelen vaardigheid is. Toch is een goede en eenduidige beoordeling erg belangrijk. Daarom zou onze aanbeveling zijn om te onderzoeken welke mogelijkheden er zijn om de schrijfoopdracht beter en eenduidiger te beoordelen. Een mogelijkheid die overwogen kan worden is het inzetten van onafhankelijke getrainde beoordelaars. Een andere optie is om in een e-learning voor docenten het correctievoorschrift verder toe te lichten en docenten ermee te laten oefenen, zodat docenten uitgebreider getraind worden in het beoordelen van de schrijftaak op een eenduidige manier. Ook adviseren we om de beoordelaarsovereenstemming te onderzoeken en vast te stellen. Dit kan inzicht geven in welke typen vragen gemakkelijk te beoordelen zijn en welke niet, zodat correctievoorschriften daarvoor volgende examens op kunnen worden aangepast en er aandacht aan kan worden geschonken in een training voor correctoren.

### 3.2.5 Fase 7: Evaluatie

De laatste fase omvat de evaluatie. Criterium 7.1 gaat over het aanpassen van toetsspecificaties als er significante veranderingen zijn in het vertegenwoordigde domein. In het document 'Fase E' wordt o.a. geconcludeerd dat voor het examen Nederlands vmbo gl/tl een ander referentie-examen is gewenst, omdat in het huidige referentie-examen nog een geleide samenvattingsopdracht zat. Verder zijn er enkele conclusies getrokken over het correctievoorschrift van de schrijfoopdrachten.

Aan criterium 7.2 wordt voldaan. De bewijzen voor inhoudsvaliditeit zijn vastgelegd in een toetsmatrijs met de beoogde samenstelling en in een toetsrationale met degerealiseerde samenstelling op item-niveau. Aan criterium 7.3 wordt deels voldaan. De bewijzen voor de betrouwbaarheid zijn vastgelegd in de TIA; de beoordelaarsovereenstemming is niet vastgelegd (zie ook eerder advies hierover).

Aan criteria 7.4 tot en met 7.6 wordt voldaan. In het algemene deel van de constructieopdracht worden procedures rondom toetsontwikkeling beschreven en wordt aangegeven wie waar verantwoordelijk voor is (zie ook het Handboek Centrale Examens). Ook is er onderzoek gedaan onder de stakeholders die betrokken zijn bij het gebruik van het examen. Er is een QuickScan uitgezet onder scholen en er heeft een evaluatie via het LAKS plaatsgevonden.

Aan het laatste criterium in deze fase (criterium 7.7) wordt deels voldaan. Met het evaluatieformulier wordt er teruggeblikt op het examen. Daarin worden ervaringen met de testcorrectie, de bevindingen van het LAKS, de QuickScan en de resultaten uit de TIA meegenomen. Het CvTE geeft daarnaast aan dat er een verslag wordt gemaakt van het jaargesprek met de voorzitter van de vaststellingscommissie, maar onduidelijk is of daarin de verantwoording over de ontwikkeling, samenstelling, afname en beoordeling aan bod komt.

Onze aanbeveling bij deze fase is om de verantwoording over de ontwikkeling, samenstelling, afname en beoordeling vollediger en in één document vast te leggen.

### 3.3 Samenvattend

Het examen Nederlands vmbo gl/tl 2022 eerste tijdvak voldoet grotendeels aan het onderzoekskader dat op internationale standaarden is gebaseerd, dat wil zeggen dat aan de meeste criteria van het onderzoekskader wordt voldaan. Met uitzondering van fase 1, zijn er in de verschillende fasen echter ook criteria aan te wijzen waar niet of niet helemaal aan voldaan wordt. Veel van deze ontbrekende criteria betreffen het completer en/of duidelijker vastleggen van verschillende zaken rondom het toetsproces.

Zoals hierboven beschreven hebben we op basis van de gevonden hiaten in het toetsproces van het examen Nederlands vmbo gl/tl enkele aanbevelingen geformuleerd. Samengevat luiden deze aanbevelingen als volgt:

- leg vooraf de gewenste psychometrische eigenschappen per item en per examen als geheel vast;
- leg de keuze voor de volgorde van teksten voor het onderdeel 'lezen' vast;
- leg de kwalificaties, relevante ervaringen en demografische kenmerken van vakinhoudelijke experts die betrokken zijn bij het opstellen van de toetspecificaties en bij de constructie van het examen vast in een competentieprofiel;
- neem in het opgavenboekje of de uitwerkbijlage voor kandidaten op welke hulpmiddelen zijn toegestaan;
- onderzoek hoe schrijfvaardigheid eenduidiger beoordeeld kan worden;
- onderzoek de beoordelaarsovereenstemming tussen de eerste en tweede corrector;
- leg de verantwoording over de ontwikkeling, samenstelling, afname en beoordeling van het examen vollediger en in één document vast.

## 4 Onderzoek validiteit CE Duits vwo 2022 eerste tijdvak

### 4.1 Beschrijving van het examen

Het examen Duits vwo 2022 eerste tijdvak is een papieren examen dat uit 40 vragen over elf korte teksten bestaat. Het bevat 28 meerkeuzevragen en twaalf open vragen. Voor dit examen kunnen leerlingen maximaal 51 punten behalen. De afname van het examen vond plaats op maandag 23 mei 2022, van 9:00 tot 11:30 uur.

#### 4.1.1 Domeindefinitie

De examenstof voor het centraal examen Duits vwo 2022 is gericht op het meten van de leesvaardigheid. Dit is domein A van het examenprogramma moderne vreemde talen. De overige domeinen zijn kijk- en luistervaardigheid (B), gespreksvaardigheid (C), schrijfvaardigheden (D), literatuur (E) en oriëntatie op studie en beroep (F). De domeinen B t/m F worden in het schoolexamen geëxamineerd en zijn geen onderdeel van het centraal examen.

Het centraal examen bepaalt de helft van het eindcijfer van de kandidaat. De toetsing van de overige vaardigheden, in het schoolexamen, bepaalt de andere helft van het eindcijfer. De specificaties voor het centraal examen zijn onder verantwoordelijkheid van het CvTE tot stand gekomen en hebben een verplichtend karakter voor de inhoud en de onderwijsniveaus. Deze staan in de syllabus. Voor het toetsen van vaardigheden in het schoolexamen zijn handreikingen opgesteld door de SLO. Deze hebben geen verplichtend karakter.

Voor het domein leesvaardigheid zijn in de syllabus de volgende globale eindtermen geformuleerd: De kandidaat kan:

- aangeven welke informatie relevant is, gegeven een vaststaande behoefte;
- de hoofdgedachte van een tekst(gedeelte) aangeven;
- de betekenis van belangrijke elementen van een tekst aangeven;
- relaties tussen delen van een tekst aangeven;
- conclusies trekken met betrekking tot intenties, opvattingen en gevoelens van de auteur.

#### 4.1.2 Domeinrepresentatie

Zoals in 3.1.2 beschreven, betreft de domeinrepresentatie de mate waarin een examen het domein zoals dat gedefinieerd is door de toetsspecificaties en in de toetsmatrijs adequaat representeert en meet. Het behoort tot de expertise van de vakdeskundigen in de vaststellingscommissie om te bepalen of de opgaven in voldoende mate het beoogde domein representeren. In het algemene deel van de constructieopdracht is aangegeven dat de experts van het CvTE in fase B van de examencyclus van Cito en CvTE (Figuur 1.1) van het vaststellen van de examens goedkeuring geven aan de uitwerking van de constructieopdracht, in fase C gevolgd door de vaststelling van de examenopgaven. In fase D volgt dan de vaststelling van het gehele examen. Voor het examen Duits vwo 2022 eerste tijdvak geldt dat deze stappen zijn doorlopen, waarmee dus door het CvTE is vastgesteld dat de domeinrepresentatie voldoende is. Daarbij moet wel worden aangetekend dat de bij fase B en fase C behorende documenten ontbreken. Fase C kan bij het examen Duits samenvallen met fase D.

### 4.1.3 Domeinrelevantie

Domeinrelevantie betreft de mate waarin elke opgave van een examen relevant is voor het beoogde domein. Ook om de relevantie van opgaven te bepalen, is de inzet van vakdeskundigen nodig. In de constructiefase gebeurt dit doordat screeners en constructiegroepleden van Cito zijn betrokken bij de constructie van de opgaven. En ook de vakdeskundigen van de vaststellingscommissie kijken hier grondig naar: ze keuren opgaven die in hun ogen niet relevant zijn af.

Met behulp van een QuickScan is nagegaan hoe docenten/examinatoren oordelen over de moeilijkheidsgraad en lengte van het examen en over de aansluiting van het examen bij het gegeven onderwijs. Docenten van tachtig procent van de scholen hebben hieraan meegewerkt (573 docenten van 463 scholen). De meerderheid (62%) vond het examen niet te moeilijk en niet te makkelijk; bijna een derde (31%) vond het moeilijk. Iets meer dan de helft (53%) vond de lengte van het examen in verhouding tot de beschikbare tijd precies goed; bijna de helft (47%) vond het te lang). Iets meer dan de helft vond de inhoudelijke aansluiting bij het gegeven onderwijs voldoende; ruim een derde (37%) vond de aansluiting goed. De docenten is ook gevraagd een cijfer te geven voor het examen. De gemiddelde score die zij geven is 6,6. Als we de gegeven cijfers nader bekijken, blijkt dat 13 procent van de docenten het examen een onvoldoende gaf: 1 procent gaf het examen het cijfer 3, 2 procent gaf het cijfer 4 en 10 procent gaf het cijfer 5. Relatief veel docenten gaven het cijfer 6 (29%) of 7 (45%); 13 procent gaf het examen het cijfer 8.

In de evaluatie door de vaststellingscommissie van het CvTE wordt verwezen naar de resultaten van de bovengenoemde QuickScan, het verslag van Levende Talen, LAKS, scholieren.com en de vaststellingscommissie. De conclusie van de vaststellingscommissie over dit examen is positief, maar er wordt op gewezen dat zowel docenten als LAKS minder tevreden zijn over de lengte van het examen. Dit zou volgens velen korter moeten zijn. Tevens verwijst de vaststellingscommissie naar positieve resultaten van de toets- en itemanalyse. Een aandachtspunt is dat er volgens de commissie onvoldoende duidelijkheid bij examinatoren is over wat 'de kern van een goed antwoord' inhoudt, zoals opgenomen in het correctievoorschrift.

## 4.2 Samenvatting onderzoek validiteit CE Duits vwo 2022 eerste tijdvak

Aan de hand van het onderzoekskader, beschreven in hoofdstuk 1.3.1, is het centraal examen Duits vwo uit het eerste tijdvak van 2022 geanalyseerd. In deze paragraaf beschrijven we de belangrijkste bevindingen. Gedetailleerde bevindingen per criterium staan beschreven in Bijlage 4: Checklist Centraal examen Duits vwo 2022 eerste tijdvak. Fasen 4 (Afname) en 6 (Registratie en communicatie) uit het toetsproces worden hier buiten beschouwing gelaten.

### 4.2.1 Fase 1: Basisontwerp

In het onderzoekskader staat dat er een toetsplan moet zijn en wat daarin moet worden gespecificeerd. De term 'toetsplan' wordt door het CvTE niet gebruikt, maar er zijn wel drie documenten die samen als zodanig kunnen worden aangemerkt: het examenprogramma moderne vreemde talen en literatuur havo/vwo, met bijlage vwo Duits, de syllabus moderne vreemde talen vwo en de constructieopdracht.

Aan de criteria voor een toetsplan wordt grotendeels voldaan. De doelgroep van het examen is vastgesteld, het domein is gedefinieerd en het gebruiksdoel van het examen is gedefinieerd. In het examenprogramma en in de syllabus worden zes domeinen en de bijbehorende subdomeinen en vaardigheden beschreven. Ook wordt aangegeven hoe de domeinen over het centraal examen en het schoolexamen verdeeld zijn. In het centraal examen komt alleen leesvaardigheid aan bod. In de syllabus wordt een overzicht gegeven van de algemene beheersingsniveaus bij lezen en er wordt een koppeling gelegd met het Europees referentiekader



(ERK). In verband hiermee wordt verwezen naar een internationaal ijkingsonderzoek waaruit bleek dat examenkandidaten (in 2010) 57 procent van de scorepunten moesten halen om aan het ERK-niveau C1 te voldoen (hetgeen overeenkwam met iets hoger dan een voldoende voor het examen).

Een van de criteria in deze fase is het specificeren van hoe de toets moet worden afgenomen en gescoord. Deze informatie is beschikbaar, maar is verdeeld over verschillende documenten: de constructieopdracht, de vakspecifieke informatie bij de september-/maartmededeling en het correctievoorschrift. Qua theoretische uitgangspunten wordt verwezen naar het Europees Referentiekader (ERK).

Wij bevelen aan informatie over het afnemen en scoren zoveel mogelijk te bundelen in één document.

#### 4.2.2 Fase 2: Opstellen toetsspecificaties en toetsmatrijs

Aan de meeste criteria met betrekking tot het opstellen van de toetsspecificaties en de toetsmatrijs is geheel of deels voldaan.

In de constructieopdracht worden de algemene toetsspecificaties toegelicht. In de toetsmatrijs, die deel uitmaakt van de constructieopdracht, is aangegeven wat het minimale en maximale aantal vragen is, met bijbehorende maximumscore (40 tot 45 vragen; 45 tot 55 scorepunten) en er wordt een globale verdeling over vier vraagsoorten (meerkeuzevragen, voorgestructureerde vragen, open vragen, citeervragen) aangegeven. De meerkeuzevragen dienen in dit examen samen tussen 53 en 67 procent van de scorepunten op te leveren. Binnen de vier vraagsoorten worden één of meer vraagvormen onderscheiden: traditioneel meerkeuze, meerkeuze invul (cloze), in- of aanvulvraag, combinatievraag, beweringenvraag, ordeningsvraag, kort-antwoordvraag en vraag waarbij een woord, woordgroep of zin(nen) moeten worden geciteerd. Ook wordt het minimale en maximale aantal teksten in het examen aangegeven en er wordt ingegaan op de gewenste afwisseling in tekstsoorten, tekstfuncties en thema's en op de linguïstische en/of cognitieve complexiteit. In het onderdeel 'examencondities' van de constructieopdracht wordt een specificatie gegeven van de tijdens de afname van het examen toegestane hulpmiddelen en van de maximaal beschikbare tijd.

In de constructieopdracht is niet aangegeven wat de psychometrische eigenschappen van de items en van de toets als geheel moeten zijn (criterium 2.7). Hier is niet vermeld wat de gewenste betrouwbaarheid is, hoeveel gemakkelijke en hoeveel moeilijke vragen dienen te worden opgenomen en wat de gewenste p-waarde van de items is. Cito geeft op haar website wel in het algemeen aan dat wordt gestreefd naar een p-waarde van 0.50 per vraag. Er zijn enkele richtlijnen voor de opbouw van het examen, maar de volgorde van de vragen wordt niet specifiek aangegeven (criterium 2.8).

In de syllabus wordt eveneens ingegaan op vraagvormen, het mogelijke aantal teksten, tekstsoorten en thema's. De relevantie van de inhoud van het examen voor het beoogde meetdoel wordt slechts gedeeltelijk aannemelijk gemaakt. Er wordt weliswaar verwezen naar de vijf eindtermen leesvaardigheid die getoetst worden in het centraal examen en er wordt verwezen naar de relatie met het ERK, maar er wordt geen duidelijk verband gelegd tussen de inhoud van de toets en de eindtermen (criterium 2.10).

Of de toetsmatrijs een adequate representatie is van het meetdoel (criterium 2.11), is niet duidelijk. De beoordeling van de toetsspecificaties door deskundigen is slechts deels gedocumenteerd. Documenten die betrekking hebben op fase B (goedkeuring van de uitwerking van de constructieopdracht) en fase C (vaststelling examenopgaven) ontbreken. Het werkwoordgebruik in de eindtermen is eenduidig en hoewel niet expliciet verwezen wordt naar een taxonomie, zoals bijvoorbeeld van Bloom of Romiszowsky, zien we

in de toetsmatrijs werkwoorden terug die passend zijn bij de verschillende niveaus (taxonomieën) van Bloom, (zoals ‘de kandidaat kan aangeven’ en ‘de kandidaat kan conclusies trekken’).

De toetsspecificaties worden door relevante experts op hun geschiktheid beoordeeld. De relevante experts zijn toetsdeskundigen van het Cito en de leden van de vaststellingscommissie van de CvTE. De kwalificaties waaraan zij moeten voldoen, zijn slechts deels vastgelegd in de beschikbare documenten. In het Huishoudelijk Reglement CEVO staat aan welke voorwaarden de voorzitter en de twee leden van de vaststellingscommissie moeten voldoen. Dit reglement is echter erg gedateerd. Het stamt uit de jaren ‘90. Daarnaast worden de vereiste kwalificaties van leden van de vaststellingscommissie beschreven in het schriftelijke verzoek om leden voor die commissie voor te dragen dat wordt gedaan aan de vakbonden. Hiermee wordt niet volledig voldaan aan criterium 2.14.

Wij bevelen aan de gewenste psychometrische eigenschappen van de items en van de toets als geheel vast te leggen en een duidelijk verband te leggen tussen de inhoud van de toets en de eindtermen. Ook bevelen wij aan te documenteren in hoeverre de toetsmatrijs een adequate representatie is van het meetdoel en de beoordeling van de toetsspecificaties door deskundigen volledig te documenteren. Tevens bevelen wij aan de kwalificaties waaraan de constructeurs van de examenopgaven en de vakexperts die de examenopgaven moeten vaststellen, beter vast te leggen in een actueel document.

### 4.2.3 Fase 3: Constructie en Normering

Fase 3 gaat over het construeren en normeren van examenopgaven.

In aanvulling op de informatie in de toetsmatrijs (zie paragraaf 4.2.2) wordt in de constructieopdracht ingegaan op de examendocumenten, bronvermelding bij de teksten en het lettertype en er worden scoringsregels gegeven. Ook zijn in de constructieopdracht afspraken en aandachtspunten opgenomen die betrekking hebben op het beoordelingsmodel (bijvoorbeeld formuleringen met het oog op de correctie van open vragen en het beoordelen van citeervragen).

Er heeft een kwalitatieve evaluatie plaatsgevonden door de vaststellingscommissie. Deze uit drie leden bestaande commissie heeft het examen vóór afname vergeleken met het referentie-examen en heeft voor het examen Duits vwo 2022 eerste tijdvak bepaald dat twee items moesten worden bijgesteld. Dit is vastgelegd in het document fase D.

De instructie aan de examenkandidaat is bijna volledig; er wordt niet aangegeven welke hulpmiddelen toegestaan zijn, maar dit is wel te vinden op [mijneindexamen.nl](http://mijneindexamen.nl) (criterium 3.1d).

In de Amerikaanse standaarden wordt geadviseerd dat een examen wordt samengesteld uit gepreteste opgaven (criterium 3.3a). Bij het examen Duits wordt geen pretest uitgevoerd, zo geeft het CvTE aan. De moeilijkheidsgraad wordt vastgesteld met behulp van een posttest. Dit is een vergelijkbare procedure als bij een pretest, maar deze wordt uitgevoerd nadat het examen is afgenomen. De informatie uit de posttest is beschikbaar voordat de N-term wordt bepaald en kan daarbij dus meegenomen worden.

De norm wordt na afloop van het examen bepaald met behulp van de N-termprocedure. De manier waarop de norm (of cesuur) wordt bepaald, is gedocumenteerd. In de constructieopdracht is vastgelegd dat er bij de oplevering van het examen een toetsrationale aanwezig moet zijn. Dat is inderdaad het geval (criterium 3.10).

In het referentiekader is opgenomen dat de procedures die worden gebruikt om items te ontwikkelen, beoordelen en uitproberen en om, indien van toepassing, items uit de vragenbank te selecteren, moeten zijn gedocumenteerd (criterium 3.7). In de constructieopdracht wordt verwezen naar de syllabi moderne vreemde talen havo en vwo 2022. Daarnaast zijn er verschillende checklists en handreikingen die door het Cito zijn ontwikkeld. Ook zijn er documenten met informatie over de normeringssystematiek en over de normering. Het gaat om een groot aantal afzonderlijke documenten.

Wij bevelen aan in de instructie aan de kandidaat informatie toe te voegen over de bij het examen toegestane hulpmiddelen. Met betrekking tot constructie en normering bevelen wij aan om relevante documenten te bundelen.

#### 4.2.4 Fase 5: Beoordeling, verwerking en analyse

Op het gebied van de beoordeling, verwerking en analyse is aan bijna alle criteria voldaan.

De procedures voor het scoren van de opgaven en de beoordelingscriteria bij opgaven zijn vastgelegd in het correctievoorschrift. Het correctievoorschrift bevat de regels voor de beoordeling, algemene regels, vakspecifieke regels, een beoordelingsmodel, regels over het aanleveren van scores en bronvermeldingen. Daarnaast is door het Cito een toets- en itemanalyse (TIA) uitgevoerd waarin de betrouwbaarheid is bepaald. Deze TIA wordt met het CvTE gedeeld. Vervolgens vult de vaststellingscommissie, die deel uitmaakt van het CvTE, een evaluatieformulier (fase E) in en deelt dit met de clustermanager van het CvTE. De gemiddelde p-waarde die in de toets- en itemanalyse wordt vermeld, is 62,1. Uit de toets- en itemanalyse blijkt volgens de vaststellingscommissie dat dit examen een betrouwbaar beeld oplevert; de coëfficiënt alpha is 0,81. Er is daarmee dus ook geen aanleiding om aan de validiteit van de toetscores te twijfelen.

Er zijn twee beoordelaars die het examen apart van elkaar scoren: de examiner (een docent van de examenkandidaat) en een 'gecommitteerde' (tweede corrector) van een andere school. De examiner en de tweede corrector dienen in onderling overleg het behaalde aantal scorepunten vast te stellen, waarna het bevoegd gezag op basis van de omzettingstabel het behaalde cijfer vaststelt. De examiner voert vervolgens het cijfer in de applicatie 'Wolf' (van het Cito) in of op een standaardformulier. 'Wolf' biedt ook de mogelijkheid aan de eerste en tweede beoordelaar om de score in te voeren, maar dit is niet verplicht. De twee onafhankelijke scores worden dus niet beide gerapporteerd (criterium 5.4). Daardoor kan de mate van beoordelaarsovereenstemming bij de correctie van de antwoorden op de open vragen tussen de examiner en gecommitteerde niet (voorafgaand aan de uitslag) door psychometrisch onderzoek onderzocht worden. De monitoring van het beoordelingsproces verloopt daarmee niet volgens de adviezen van de Amerikaanse standaarden, waarin grote waarde gehecht wordt aan een zo objectief mogelijk beoordelingsproces met monitoring van de overeenstemming. Dit komt ten goede aan de betrouwbaarheid van de uitslag voor kandidaten. In een toelichting van het CvTE wordt opgemerkt dat de praktijk van de eerste en tweede correctie door Cito wel na afloop van een afnameperiode van de examens periodiek onderzocht wordt.

Wij bevelen aan dat met behulp van psychometrisch onderzoek wordt vastgesteld in hoeverre de beoordeling van de antwoorden op de open vragen door de eerste en tweede corrector met elkaar overeenstemt.

#### 4.2.5 Fase 7: Evaluatie

De wijze waarop de toetscyclus wordt geëvalueerd voldoet grotendeels aan de criteria uit het onderzoekskader. Als dit nodig wordt gevonden, kan de vaststellingscommissie de syllabus en de constructieopdracht (laten) aanpassen. Ook de clustermanager kan hiertoe suggesties doen. Er is jaarlijks overleg met Levende Talen. Het CvTE geeft aan dat grote wijzigingen leiden tot de instelling van een syllabuscommissie. Of het examenprogramma wordt aangepast, is een politiek besluit, zo geeft het CvTE aan.

Het is niet duidelijk of bewijzen voor (inhouds)validiteit zijn vastgelegd (criterium 7.2). De toets- en itemanalyse geeft informatie over de betrouwbaarheid van de toets, maar niet over de mate van overeenstemming tussen de examinerator en de gecommitteerde (criterium 7.3). Procedures rondom de toetsontwikkeling zijn vastgelegd, evenals de rollen, verantwoordelijkheden en taakverdeling. Er wordt onderzoek gedaan onder stakeholders (via de QuickScan bij docenten en verslagen en informatie van Levende Talen, Laks en scholieren.com).

Er zijn verschillende fasen in de beoordeling en vaststelling van het eindexamen. In fase B worden de bronnen (teksten) beoordeeld op geschiktheid en niveau. In fase C worden (blokken van) opgaven beoordeeld op correctheid, geschiktheid en niveau. In fase D wordt het examen als geheel beoordeeld, waarbij ook wordt gekeken naar de overeenkomst met de toetsspecificaties. Zoals aangegeven, is er daarnaast de evaluatie (fase E). Er is geen document waarin de resultaten van fase B t/m E samen worden gerapporteerd. Bovendien ontbraken documenten die fase B en C betreffen. Overigens vallen bij Duits fase C en D samen en is een afzonderlijk document voor fase C niet nodig.

Wij bevelen aan een aantal zaken rond het examen en de evaluatie daarvan vollediger te documenteren. Daarbij gaat het om het bewijs van (inhouds)validiteit, de in de verschillende fasen van de ontwikkeling van het examen uitgebrachte oordelen van de vaststellingscommissie en de mate van overeenstemming tussen de eerste en tweede corrector.

### 4.3 Samenvattend

Samenvattend kan worden geconcludeerd dat het eindexamen Duits vwo 2022 eerste tijdvak qua procedures rond ontwikkeling, afname en beoordeling in het algemeen voldoet aan de criteria die in het beoordelingskader zijn opgenomen. Bij de manier waarop een en ander is gedocumenteerd, kunnen echter kanttekeningen worden geplaatst. De verantwoording is over tal van documenten verspreid en ontbreekt deels of is onvolledig. Zo ontbreken documenten waarin de (inhouds)validiteit is vastgelegd en ontbreekt een deel van de documenten die betrekking hebben op het proces van vaststelling van de examenopgaven. Ook aan sommige criteria wordt niet voldaan. Er wordt bijvoorbeeld niet vooraf aangegeven wat de gewenste psychometrische eigenschappen zijn en de overeenstemming tussen de eerste en tweede corrector bij het beoordelen van antwoorden op open vragen wordt niet vastgesteld.

In de voorgaande paragrafen hebben wij een aantal aanbevelingen geformuleerd met betrekking tot dit eindexamen. Samengevat luiden deze als volgt:

- leg vooraf de gewenste psychometrische eigenschappen van de items en van de toets als geheel vast;
- leg een duidelijk verband tussen de inhoud van de toets en de eindtermen;
- leg de kwalificaties, relevante ervaringen en demografische kenmerken van vakinhoudelijke experts die betrokken zijn bij het opstellen van de toetsspecificaties en bij de constructie van het examen vast in een competentieprofiel;

- neem in het opgavenboekje of de uitwerkbijlage voor kandidaten op welke hulpmiddelen zijn toegestaan;
- onderzoek de beoordelaarsovereenstemming tussen de eerste en tweede corrector;
- leg de verantwoording over de ontwikkeling, samenstelling, afname en beoordeling van het examen vollediger en in één document vast.

## 5 Conclusie en aanbevelingen

In dit hoofdstuk beschrijven we de conclusies die naar voren komen uit het hierboven beschreven onderzoek naar de procedures rondom de borging van de inhoudsvaliditeit van de eindexamens Nederlands vmbo gl/tl en Duits vwo 2022. Vervolgens gaan we in op de aanbevelingen per examen, en aanbevelingen in het algemeen.

In het onderzoek is de volgende vraag onderzocht:

*‘Voldoen de procedures die het CvTE en Cito bij de borging van de inhoudsvaliditeit van de examens Duits (vwo) en Nederlands (vmbo gl/tl) hanteren aan (inter)nationale kwaliteitsstandaarden?’*

In hoofdstuk 3 en hoofdstuk 4 hebben we beschreven in hoeverre het toetsproces van de examens Duits (vwo) en Nederlands (vmbo gl/tl) voldoet aan de criteria van het onderzoekskader zoals beschreven in hoofdstuk 2. Op basis hiervan geven we hieronder per examen antwoord op de onderzoeksvraag. Zoals gezegd, realiseren we ons dat de internationale standaarden niet voor alle criteria één op één te vertalen zijn naar de Nederlands context. Daar waar de Amerikaanse standaarden niet passend zijn, hebben we de criteria op kleine punten aangepast. Voor criteria die meer aanpassingen vereisen, geven we in dit hoofdstuk aanbevelingen voor aanscherping van het onderzoekskader.

Het toetsproces van beide examens voldoet op veel punten aan de gestelde criteria. Voor beide examens zijn er echter ook criteria waar niet aan wordt voldaan. Hieronder wordt daar per examen nader op ingegaan.

Voor het examen Nederlands vmbo gl/tl eerste tijdvak 2022 zijn er met uitzondering van fase 1 in alle fasen criteria aan te wijzen waar niet of niet helemaal aan voldaan wordt (zie bijlage 3 voor een uitgebreid overzicht). Deze criteria gaan vooral over het duidelijk en compleet vastleggen van de verschillende aspecten van het toetsproces. Uit onze documentenanalyse blijkt dat dit niet voor alle aspecten even goed gebeurt. De aanbevelingen die hieronder beschreven zijn, betreffen dan ook met name het beter documenteren van de verschillende onderdelen van het toetsproces. Verder concluderen we voor het examen Nederlands vmbo gl/tl dat de betrouwbaarheid van het examen vrij laag is (coëfficiënt Alpha is 0,54) en dat de beoordelaarsovereenstemming niet onderzocht is. Tot slot hebben we vastgesteld dat er enkele details ontbreken zoals het vermelden van de hulpmiddelen op de examenmaterialen voor de kandidaat.

Op basis van deze conclusies hebben we voor het examen Nederlands vmbo gl/tl de volgende aanbevelingen geformuleerd:

- leg vooraf de gewenste psychometrische eigenschappen van de items en de toets als geheel vast;
- leg de keuze voor de volgorde van teksten voor het onderdeel 'lezen' vast;
- leg de kwalificaties, relevante ervaringen en demografische kenmerken van vakinhoudelijke experts die betrokken zijn bij het opstellen van de toetsspecificaties en bij de constructie van het examen vast in een competentieprofiel;
- neem in het opgavenboekje of de uitwerkbijlage voor kandidaten op welke hulpmiddelen zijn toegestaan;
- onderzoek hoe schrijfvaardigheid eenduidiger beoordeeld kan worden;
- onderzoek de beoordelaarsovereenstemming tussen de eerste en tweede corrector;

- leg de verantwoording over de ontwikkeling, samenstelling, afname en beoordeling van het examen vollediger en in één document vast.

Voor het examen Duits vwo eerste tijdvak 2022 geldt eveneens dat de procedures die het CvTE en Cito hebben gevolgd over het algemeen voldoen aan de (inter)nationale kwaliteitsstandaarden. Aan de meeste criteria van het onderzoekskader wordt voldaan, maar ook hier zijn er in de verschillende fasen ook criteria aan te wijzen waar niet of niet helemaal aan voldaan wordt (zie bijlage 4 voor een uitgebreid overzicht). Deels heeft dit te maken met documentatie die ontbreekt, tekortschiet, gefragmenteerd of verouderd is. Zo ontbreken documenten waarin de ontwikkeling en samenstelling van het examen zijn gedocumenteerd. Deels gaat het ook om technische aspecten zoals het ontbreken van een indicatie van de gewenste psychometrische eigenschappen en de controle op overeenstemming in beoordeling van open vragen door de eerste en tweede corrector.

Op basis van deze conclusies hebben we voor het examen Duits vwo de volgende aanbevelingen geformuleerd:

- leg vooraf de gewenste psychometrische eigenschappen van de items en de toets als geheel vast;
- leg een duidelijk verband tussen de inhoud van de toets en de eindtermen;
- leg de kwalificaties, relevante ervaringen en demografische kenmerken van vakinhoudelijke experts die betrokken zijn bij het opstellen van de toetsspecificaties en bij de constructie van het examen vast in een competentieprofiel;
- neem in het opgavenboekje of de uitwerkbijlage voor kandidaten op welke hulpmiddelen zijn toegestaan;
- onderzoek de beoordelaarsovereenstemming tussen de eerste en tweede corrector;
- leg de verantwoording over de ontwikkeling, samenstelling, afname en beoordeling van het examen vollediger en in één document vast.

In vergelijking met het eerdere onderzoek naar de validiteit van het toetsproces<sup>36</sup> is er een opvallende overlap in de geformuleerde aanbevelingen. Dat doet vermoeden dat toen gesignaleerde afwijkingen ten opzichte van de internationale kwaliteitsstandaarden van structurele aard zijn. Hierbij is het goed in ogenschouw te nemen dat de toetscyclus van de examens in het huidige onderzoek al vergevorderd was op het moment dat het vorige rapport met aanbevelingen verscheen.

Het hele proces van toetsontwikkeling staat beschreven in een breed scala aan documenten, variërend in stijl en status. Sommige documenten lijken werkdocumenten; sommige documenten zijn weer erg formeel opgesteld. Deze variatie geeft al met al een rommelig beeld. Dit maakt het omslachtig om een overkoepelend beeld te krijgen van het hele toetsproces. Het is de vraag of dat passend en wenselijk is bij high stakes examens die jaarlijks worden afgenomen onder grote aantallen vo-scholieren.

Onze algemene aanbevelingen voor de procedures met betrekking tot de centrale examens overlappen dan ook sterk met die uit 2022:

- zorgvuldiger en systematischer documenteren: er is een groot aantal documenten om het werk aan de ontwikkeling van de opgaven en examens te ondersteunen. Deze documenten bestrijken elk een deel van de toetsontwikkeling en de evaluatie maar overlappen op sommige punten ook. Dit maakt het lastig om een duidelijk en compleet beeld van de gemaakte keuzes in de toetscyclus te vormen;

36 Lommertzen, J, Janssen, J., & Heijsters, L. (2022). Onderzoek naar de inhoudsvaliditeit van het toetsproces van twee maatschappijgerichte centrale examens. Geschiedenis vmbo gl/tl 2021-1 en Economie havo 2021-1. Nijmegen / 's-Hertogenbosch: ResearchNed / CINOP.



- zorgen voor een systeem dat het mogelijk maakt om (psychometrisch) onderzoek te doen naar de overeenstemming tussen eerste en tweede beoordelingen. Hoewel het systeem nu voldoet aan wettelijke regelingen, is het nu niet mogelijk om de beoordelaarsovereenstemming te onderzoeken;
- Het afdrukken van de toegestane hulpmiddelen op het opgavenboekje. Hierdoor beschikken alle kandidaten op dit punt over dezelfde informatie en zijn zij niet afhankelijk van hoe hun school er vooraf over informeert en ermee omgaat.

Aanbevelingen voor aanscherping van het onderzoekskader:

- criterium 1.7, dat nu als volgt is geformuleerd: “In het toetsplan staat de theorie met betrekking tot de beoogde interpretatie beschreven” uitbreiden met de bijzin: “of er wordt verwezen naar de kaders waarop het examen is gebaseerd.”;
- de huidige criteria 3.4 en 3.5 zijn nu opgenomen in fase 3: Constructie en Normering. Criterium 3.4 luidt: “Er wordt een norm/cesuur verstrekt die via een wetenschappelijke methodiek (standaardbepaling) is bepaald”. In de werkwijze van CvTE en Cito wordt de normering via de N-term echter pas vastgesteld in fase 5, na afname, beoordeling en analyse van het examen. Daarom stellen wij voor om het analysekader op dit punt aan te passen en criteria 3.4 en 3.5 op te nemen in fase 5; fase 3 te hernoemen naar Constructie en fase 5 naar Beoordeling, analyse en normering.

## Literatuur

- AEA-Europe. (2017) European Framework of Standards for Educational Assessment 1.0: [https://www.aea-europe.net/wp-content/uploads/2017/07/SW\\_Framework\\_of\\_European\\_Standards.pdf](https://www.aea-europe.net/wp-content/uploads/2017/07/SW_Framework_of_European_Standards.pdf)
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014). *1999 Standards for Educational and Psychological Testing, 2014 Edition*, Washington, DC: American Educational Research Association. Geraadpleegd van: [https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards\\_2014edition.pdf](https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf)
- Berkel, H. van, Bax, A. & Joosten-Ten Brinke, D. (Red.) (2013) *Toetsen in het hoger onderwijs, 3e herziene druk*, Houten: Bohn Stafleu van Loghum.
- Brouwer, A., Sanders, P., Veldkamp, B. (2019) Onderzoek naar de inhoudsvaliditeit van een tweetal beroepsgerichte centrale examens voortgezet onderwijs 2019. RCEC - Onderzoek in opdracht van het College voor Toetsen en Examens.
- Cizek, J. K. (2006). Standard Setting. In S. M. Downing, & T. M. Haladyna (Eds.), *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Downing, S. M. (2006) (Online edition 2011). Twelve steps for effective test development. In S. M. Downing, & T. M. Haladyna (Eds.), *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Linn, R. L. (2006). The Standards for Educational and Psychological Testing: Guidance in Test Development. In: S. M. Downing, & T. M. Haladyna (Eds.), *Handbook of Test Development*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Lommertzen, J, Janssen, J., & Heijsters, L. (2022). *Onderzoek naar de inhoudsvaliditeit van het toetsproces van twee maatschappijgerichte centrale examens. Geschiedenis vmbo gl/tl 2021-1 en Economie havo 2021-1*. Nijmegen / 's-Hertogenbosch: ResearchNed / CINOP.
- Newton, P.E., Shaw, S.D., (2014). Validity and validation. In: P. E. Newton, & S. D. Shaw (Eds.), *Validity in Educational & Psychological Assessment*. Sage Publications Ltd. Geraadpleegd van: <https://sk.sagepub.com/books/validity-and-educational-assessment/n1.i400.xml>
- Regeling van het College voor Toetsen en Examens van 30 november 2015, nummer CvTE-15.02159, houdende vaststelling van regels voor de omzetting van scores in cijfers bij centrale examens en de rekentoets in het voortgezet onderwijs (Regeling omzetting scores in cijfers centrale examens en rekentoets VO 2016). In: Staatscourant, nr. 4817, 4 februari 2016. Geraadpleegd van <http://www.stilus.nl/examen/n-term-2016.pdf>
- Sanders, P. (Red.). (2016). *Toetsen op school. Hoger Onderwijs*. Arnhem: Cito. Geraadpleegd van <https://www.cito.nl/kennis-en-innovatie/kennisplein>
- Sanders, P., Brouwer, A.J., Veldkamp, B.P. (2018). *Onderzoek naar de inhoudsvaliditeit van een tweetal centrale examens voortgezet onderwijs 2017*. Enschede: RCEC
- Sanders, P. (Red.) (2017). *Toetsen op school*. Arnhem: Cito. 1e druk 2011, herziene versie 2017. Geraadpleegd van <https://www.cito.nl/kennis-en-innovatie/kennisplein>
- Sluijter, C., Hemker, B., & Eggen, Th. (2018a). *Beoordelen van de kwaliteit van toetsen en examens, deel 1: Systemen en criteria*. Arnhem: Cito. Geraadpleegd van [www.cito.nl/kennis-en-innovatie/kennisplein](http://www.cito.nl/kennis-en-innovatie/kennisplein)
- Sluijter, C., Hemker, B., & Eggen, Th. (2018b). *Beoordelen van de kwaliteit van toetsen en examens, deel 2: de praktijk*. Arnhem: Cito. Geraadpleegd van [www.cito.nl/kennis-en-innovatie/kennisplein](http://www.cito.nl/kennis-en-innovatie/kennisplein)
- Wools, S. (2012). Towards a comprehensive Evaluation System for the quality of tests and assessments, in: *Psychometrics in Practice at RCEC* (pp. 95 - 106).
- Wools, S., (2013). De validiteit van toetsscores. In: P. F. Sanders (Red.). *Toetsen op School* (pp. 69 - 83). Arnhem: Cito.
- Wools, S. (2015). *All About Validity: towards a new evaluation system for educational assessments*. [Dissertatie. Universiteit van Twente]. Geraadpleegd van [https://www.researchgate.net/publication/281434844\\_All\\_About\\_Validity\\_An\\_evaluation\\_system\\_for\\_the\\_quality\\_of\\_educational\\_assessment](https://www.researchgate.net/publication/281434844_All_About_Validity_An_evaluation_system_for_the_quality_of_educational_assessment)

## Bijlage 1: Overzicht van figuren en tabellen

### Figuren

Figuur 1.1: In schema: samenwerking tussen het CvTE en Cito voor constructie en vaststelling van centrale examens vo.....	4
Figuur 2.1: Toetscyclus .....	7

### Tabellen

Tabel 2.1: Checklist Fase 1: Basisontwerp. ....	8
Tabel 2.2: Checklist Fase 2: Construeren toetsmatrijs .....	9
Tabel 2.3: Checklist fase 3: Construeren van de toets en normeren .....	11
Tabel 2.4: Checklist Fase 5: Beoordelen, verwerken, analyseren .....	13
Tabel 2.5: Checklist Fase 7: Evalueren en verbeteren .....	14
Tabel 2.6: Relatie tussen de fasen van de toetscyclus en examencyclus van Cito en het CvTE. ....	15

## Bijlage 2: Overzicht van geanalyseerde documenten, algemeen en specifiek voor elk examen

### Algemeen

- Afspraken over vakken met een pre-test of try-out - Cito
- Begrippen uit de TIA: geraadpleegd van Cito: [https://www2.cito.nl/vo/share/Begrippen uit de TIA.pdf](https://www2.cito.nl/vo/share/Begrippen%20uit%20de%20TIA.pdf)
- Beijering, M. & Hielkema, M. (2016). Hoe komen de examens tot stand? *Tijdschrift voor het economisch onderwijs*. ([https://www.examenblad.nl/document/artikel-teo-vecon-hoe-komen-de/2017/f=/TEO\\_Vecon\\_mei\\_2016\\_hoe\\_komen\\_de\\_ce\\_tot\\_stand.pdf](https://www.examenblad.nl/document/artikel-teo-vecon-hoe-komen-de/2017/f=/TEO_Vecon_mei_2016_hoe_komen_de_ce_tot_stand.pdf))
- Bij centrale toetsen en examens betrokken instanties - Cito
- Checklist meerkeuzevragen - Cito
- Checklist open vragen - Cito
- Checklist uitgangsmateriaal examens - Cito
- Constructieopdracht 2022 AVO-vakken vmbo havo en vwo - algemeen deel.
- Een kijkje achter de schermen van het normeren, geraadpleegd van Cito: <https://www.cito.nl/-/media/files/voortgezet-onderwijs/centrale-examens/achtergrondinformatie-artikelen-discussie-over-examens/cito-achtergrondinformatie-ce-publicatie-een-kijkje-achter-de-schermen-van-het-normeren.pdf>
- Eindexamen besluit (<https://wetten.overheid.nl/BWBR0004593/2022-04-06>)
- Functies van examens en kwaliteitseisen - Cito
- Handboek Centrale Examens VO - Cito
- Handreiking Taalgebruik in Centrale Examens - Cito
- Huishoudelijk Reglement CEVO
- Nadere uitwerking normering 2022
- Normeringssystematiek van de Centraal Examens - CvTE
- Opbouw toets- en itemanalyse: geraadpleegd van Cito: [https://www2.cito.nl/vo/share/Opbouw toets-en itemanalyse.pdf](https://www2.cito.nl/vo/share/Opbouw%20toets-en%20itemanalyse.pdf)
- Over de normering bij de Centrale Eindexamens: <https://www.examenblad.nl/onderwerp/normering-centrale-examens-2022/2022>
- Overzicht vaststellingsformulieren - Cito
- Richtlijnen lay-out opgavenboekje papier
- Technische uitwerking normering 2022.
- Wolf - uitwisselprogramma voor examengegevens: <https://www.cito.nl/onderwijs/voortgezet-onderwijs/centrale-examens-voortgezet-onderwijs/tools-en-informatie-voor-docenten/wolf>

### Specifiek voor het examen Nederlands vmbo gl/tl 2022-1

- Bijlage 1A: Regeling toegestane hulpmiddelen voor de centrale examens vmbo 2022 als bedoeld in artikel 1 Examenprogramma Nederlandse taal vmbo vanaf het CE 2014.
- College voor Toetsen en Examens (2020). Nederlands vmbo | syllabus BB, KB en GT centraal examen 2022, versie 2, juli 2020. Utrecht: College voor Toetsen en Examens.
- College voor Toetsen en Examens (2020). Vacature lid vaststellingscommissie - Nederlands VMBO 2021.

- College voor Toetsen en Examens (2021). Septembermededeling over de eindexamens 2022 + Bijlage bij septembermededeling - NEDERLANDS VMBO: VAKSPECIEKE INFORMATIE 2022. Utrecht: College voor Toetsen en Examens.
- College voor Toetsen en Examens (2022). Normering vmbo algemene vakken.
- Constructieopdracht Nederlands vmbo KB en G/T papier 2022.
- Correctievoorschrift VMBO GL en TL 2022 tijdvak 1 Nederlands CSE GL en TL.
- Evaluatie (fase E), Nederlands, vmbo gl/tl, 2022.
- Fase c: de vaststelling van opgaven papieren examens Nederlands GL/TL 2022-1
- Fase d: de vaststelling van een heel examen vmbo gt 2022-1
- Opmerkingen examenbespreking Nederlands vmbo gl/tl 2022-1
- QuickScan GL/TL Nederlands CSE 2022 tijdvak 1
- Referentiegegevens en historische N-termen 2022 vmbo avo
- Regeling toegestane hulpmiddelen voor de centrale examens vmbo 2022 als bedoeld in artikel 1 ([www.examenblad.nl](http://www.examenblad.nl)).
- Toets- en itemanalyse gl/tl Nederlands CSE 2022 tijdvak 1. Cito, Psychometrisch Onderzoek- en Kenniscentrum.
- Toetsrationale Nederlands vmbo gl/tl 2022. Cito.

#### Specifiek voor het examen Duits vwo 2022-1

- Brief aan het bestuur van de AOOb, betreffende de voordracht van leden vaststellingscommissie CvTE, d.d. 23-11-2021.
- College voor Toetsen en Examens (2020). Moderne vreemde talen vwo. Syllabus Centraal Examen 2022 Arabisch, Duits, Engels, Frans, Russisch, Spaans, Turks. Versie 2, juli 2020. Utrecht: College voor Toetsen en Examens.
- Constructieopdracht Engels, Frans, Duits, Spaans, Turks, Arabisch en Russisch havo/vwo 2022.
- Correctievoorschrift VWO 2022, tijdvak 1, Duits.
- Duits VWO: Vakspecifieke informatie 2022. Bijlage bij de september-/maartmededeling centrale examens 2022.
- Evaluatie (fase E), Duits, vwo, 2022.
- Examenprogramma moderne vreemde talen en literatuur havo/vwo, met bijlage VWO 2022, tijdvak 1, Duits.
- Fase D: de vaststelling van een heel examen, Duits, vwo 2022-1.
- Omzettingstabel normering, Duits vwo CSE 1e tijdvak; zitting 23 mei 2022; bekendmaking 9 juni 2022 (op [www.examenblad.nl](http://www.examenblad.nl)).
- Opbouw examen; VWO Duits 2022 tijdvak 1 (<https://www2.cito.nl/vo/ex2022/VW-1004-a-22-1-s.pdf>).
- QuickScan 22106 VWO Duits 2022 tijdvak 1.
- Referentiegegevens en historische N-termen 2022 havo-vwo
- Regeling toegestane hulpmiddelen voor de centrale examens havo en vwo 2022 als bedoeld in artikel 1 ([www.examenblad.nl](http://www.examenblad.nl)).
- Toets- en itemanalyse VWO Duits 2022 tijdvak 1. Cito, Psychometrisch Onderzoek- en Kenniscentrum. (<https://www2.cito.nl/vo/ex2022/VW-1004-a-22-1-tia.docx>)
- Toetsrationale Duits vwo 2022-1. Cito.

## Bijlage 3: Checklist Centraal examen Nederlands vmbo gl/tl 2022 eerste tijdvak

Fase 1: Basisontwerp Nederlands vmbo gl/tl 2022-1		Aanwezig?
1.1	Er is een toetsplan.	ja
1.2	In het toetsplan is aangegeven wat de doelgroep van het examen is.	ja
1.3	In het toetsplan is aangegeven wat het meetdoel van het examen is oftewel hoe het domein gedefinieerd is.	ja
1.4	In het toetsplan is aangegeven wat het gebruiksdoel van het examen is.	ja
1.5	In het toetsplan is in duidelijke taal gespecificeerd in welke contexten toetsscores gebruikt dienen te worden.	ja
1.6	In het toetsplan is in duidelijke taal gespecificeerd met welke processen (papieren toets of digitale toets, welke wijze van afname, welke wijze van beoordeling, hoe komt score tot stand) de toets moet worden afgenomen en gescoord.	ja
1.7	In het toetsplan staat de theorie met betrekking tot de beoogde interpretatie beschreven.	ja

Fase 2: Opstellen toetsspecificaties en toetsmatrijs Nederlands vmbo gl/tl 2022-1		Aanwezig?
2.1	Er zijn toetsspecificaties (bv. in de vorm van een constructieopdracht).	ja
2.2	In de constructieopdracht staat het aantal vragen met bijbehorende scorepunten.	ja
2.3	In de constructieopdracht staat de toetsvorm en/of het soort vragen (bijvoorbeeld gesloten en/of open).	ja
2.4	In de constructieopdracht staat de toegestane hulpmiddelen.	ja
2.5	In de constructieopdracht staat de voorgestelde lengte van de toets.	ja
2.6	In de constructieopdracht staat de hoeveelheid tijd die is toegestaan voor de toets.	ja
2.7	In de constructieopdracht staan de gewenste psychometrische eigenschappen van de items en de toets als geheel.	nee
2.8	In de constructieopdracht wordt de volgorde van items gedefinieerd.	nvt/nee
2.9	Er is een toetsmatrijs.	ja
2.10	In de toetsmatrijs wordt de relevantie van de inhoud van de toets of het examen voor het beoogde meetdoel aannemelijk gemaakt.	ja
2.11	De toetsmatrijs is een adequate representatie van het meetdoel.	
	2.11a) De eind- en toetstermen representeren het meetdoel.	ja
	2.11b) De eind- en toetstermen sluiten aan op de inhoud en het vereiste beheersingsniveau van het betreffende meetdoel.	ja
	2.11c) Het werkwoordgebruik in de eind- en toetstermen is eenduidig en sluit goed aan bij bestaande taxonomieën (zoals bijvoorbeeld van Bloom of Romiszowsky).	ja
	2.11d) Het aantal vragen geeft een voldoende dekking van het meetdoel.	ja
2.12	Relevante vakinhoudelijke experts beoordelen de toetsspecificaties op hun geschiktheid.	ja
2.13	Het doel van deze beoordeling, het proces waarmee de beoordeling wordt uitgevoerd en de resultaten van de beoordeling worden gedocumenteerd.	ja
2.14	De kwalificaties, relevante ervaringen en demografische kenmerken van relevante vakinhoudelijke experts worden gedocumenteerd.	deels

Fase 3 Constructie en normering Nederlands vmbo gl/tl 2022-1		Aanwezig?
3.1	De vragen zijn correct geformuleerd (zie richtlijnen in RCEC; die richtlijnen zijn hieronder ingevoegd als subcriteria):	
	3.1a) De vragen of opdrachten zijn gestandaardiseerd.	ja
	3.1b) De vragen of opdrachten zijn zodanig ontworpen dat fouten bij invulling voorkomen worden.	ja
	3.1c) De instructie voor de kandidaat (vaak eerste bladzijde van toetsboekje) is gestandaardiseerd, volledig en duidelijk.	ja
	3.1d) De instructie voor de kandidaat dient minimaal te bevatten: aantal vragen, wijze waarop antwoorden worden gegeven, (deel) score per vraag of opdracht, de maximaal te behalen score en de cesuur; toegestane hulpmiddelen; beschikbare tijd en wat ingeleverd moet worden bij afronding; beoordelingspunten bij open vragen.	deels
	3.1e) De kwaliteit van de lay-out en vormgeving is in orde.	ja
3.2	De vragen zijn voldoende gedetailleerd, zodat examenkandidaten de vraag kunnen beantwoorden zoals bedoeld.	ja

<b>Fase 3 Constructie en normering Nederlands vmbo gl/tl 2022-1</b>		<b>Aanwezig?</b>
3.3	De moeilijkheidsgraad van de vraag is afgestemd op de beoogde doelgroep.	ja
	3.3a) Er heeft een pretest plaatsgevonden om informatie te verkrijgen over de moeilijkheidsgraad (kwantitatieve evaluatie).	nee
	3.3b) Er heeft een kwalitatieve evaluatie met deskundigen plaatsgevonden om informatie te verkrijgen over de moeilijkheidsgraad.	ja
3.4	Er wordt een norm/cesuur verstrekt die via een wetenschappelijke methodiek (standaardbepaling) is bepaald.	Ja achteraf
3.5	De manier waarop de norm wordt bepaald is gedocumenteerd.	ja
3.6	De standaardbepaling is correct uitgevoerd.	nvt
	3.6a) De standaardbepalings-methode is op de juiste wijze uitgevoerd.	nvt
	3.6b) De vakdeskundigen/experts die bij de standaardbepaling betrokken zijn, zijn naar behoren geselecteerd en getraind.	nvt
	3.6c) Er is voldoende overeenstemming tussen de beoordelaars.	nvt
3.7	De procedures die worden gebruikt om items te ontwikkelen, beoordelen en uitproberen en - indien van toepassing om items uit de vragenbank te selecteren, zijn gedocumenteerd.	ja
3.8	Er is gedocumenteerd in hoeverre het inhoudsdomen van de toets het domein vertegenwoordigt, dat is gedefinieerd in de toets specificaties.	ja
3.9	Er is gedocumenteerd wat de kwalificaties, relevante ervaringen en demografische kenmerken zijn van degenen die de ontwikkelde items beoordelen, evenals de instructies en training die zij t.b.v. het beoordelingsproces krijgen.	deels
3.10	Er is een rationale voor de toets waarin verantwoording wordt afgelegd over de samenstelling van het examen.	ja
<b>Fase 4: Afname (n.v.t.)</b>		

<b>Fase 5: Beoordeling, verwerking en analyse Nederlands vmbo gl/tl 2022-1</b>		<b>Aanwezig?</b>
5.1	Procedures voor het scoren en scorecriteria zijn gedocumenteerd.	ja
5.2	Het proces van het selecteren, trainen, kwalificeren en monitoren van degene die scoren is gedocumenteerd.	ja
5.3	Er zijn twee onafhankelijke beoordelaars die het examen apart van elkaar scoren.	ja
5.4	De twee onafhankelijke scores worden allebei gerapporteerd.	nee
5.5	Er is gedocumenteerd hoe de uiteindelijke score voor het examen wordt berekend.	ja
5.6	Op elk scoreniveau worden in de beoordelingsvoorschriften meerdere voorbeelden van antwoorden gegeven.	ja
5.7	Er wordt een Toets- en itemanalyse uitgevoerd onder voldoende kandidaten.	ja
5.8	De Toets- en Itemanalyse maakt de betrouwbaarheid van de toetscores zichtbaar met behulp van Cronbachs coëfficiënt alfa of een andere algemeen geaccepteerde maat voor betrouwbaarheid.	ja
5.9	De betrouwbaarheid die berekend is geeft geen aanleiding om te twijfelen aan de validiteit van de toetsscores.	nee
5.10	Indien sprake is van open vragen wordt via statistisch onderzoek de mate van beoordelaarsovereenstemming onderzocht.	nee
<b>Fase 6: Registreren en communiceren: n.v.t.</b>		



	<b>Fase 7: Evaluatie Nederlands vmbo gl/tl 2022-1</b>	<b>Aanwezig?</b>
7.1	Toetsspecificaties worden aangepast als er significante veranderingen zijn in het vertegenwoordigde domein.	ja
7.2	Bewijzen voor (inhouds)validiteit en betrouwbaarheid zijn vastgelegd.	ja
7.3	Bewijzen voor betrouwbaarheid (waaronder de mate van beoordelaarsovereenstemming) zijn vastgelegd.	deels
7.4	Procedures rondom toetsontwikkeling zijn vastgelegd.	ja
7.5	Rollen, verantwoordelijkheden en taakverdeling rondom toetsontwikkeling zijn vastgelegd.	ja
7.6	Er wordt onderzoek gedaan onder de stakeholders (belangenvertegenwoordigers van kandidaten en/of van opleiders) die zijn betrokken bij het gebruik van de toets.	ja
7.7	Er is een verslag (jaarrapport, evaluatieverslag) waarin verantwoording is afgelegd over de ontwikkeling, samenstelling, afname, beoordeling, scores en uitgevoerde analyses van de toets.	deels

## Bijlage 4: Checklist Centraal examen Duits vwo 2022 eerste tijdvak

Fase 1: Basisontwerp Duits vwo 2022-1		Aanwezig?
1.1	Er is een toetsplan.	ja
1.2	In het toetsplan is aangegeven wat de doelgroep van het examen is.	ja
1.3	In het toetsplan is aangegeven wat het meetdoel van het examen is oftewel hoe het domein gedefinieerd is.	ja
1.4	In het toetsplan is aangegeven wat het gebruiksdoel van het examen is.	ja
1.5	In het toetsplan is in duidelijke taal gespecificeerd in welke contexten toetsscores gebruikt dienen te worden.	deels
1.6	In het toetsplan is in duidelijke taal gespecificeerd met welke processen (papieren toets of digitale toets, welke wijze van afname, welke wijze van beoordeling, hoe komt score tot stand) de toets moet worden afgenomen en gescoord.	ja
1.7	In het toetsplan staat de theorie met betrekking tot de beoogde interpretatie beschreven.	ja

Fase 2: Opstellen toetsspecificaties en toetsmatrijs Duits vwo 2022-1		Aanwezig?
2.1	Er zijn toetsspecificaties (bv. in de vorm van een constructieopdracht).	ja
2.2	In de constructieopdracht staat het aantal vragen met bijbehorende scorepunten.	ja
2.3	In de constructieopdracht staat de toetsvorm en/of het soort vragen (bijvoorbeeld gesloten en/of open).	ja
2.4	In de constructieopdracht staat de toegestane hulpmiddelen.	ja
2.5	In de constructieopdracht staat de voorgestelde lengte van de toets.	ja
2.6	In de constructieopdracht staat de hoeveelheid tijd die is toegestaan voor de toets.	ja
2.7	In de constructieopdracht staan de gewenste psychometrische eigenschappen van de items en de toets als geheel.	nee
2.8	In de constructieopdracht wordt de volgorde van items gedefinieerd.	deels
2.9	Er is een toetsmatrijs.	ja
2.10	In de toetsmatrijs wordt de relevantie van de inhoud van de toets of het examen voor het beoogde meetdoel aannemelijk gemaakt.	deels
2.11	De toetsmatrijs is een adequate representatie van het meetdoel.	ja
	2.11a) De eind- en toetstermen representeren het meetdoel.	ja
	2.11b) De eind- en toetstermen sluiten aan op de inhoud en het vereiste beheersingsniveau van het betreffende meetdoel.	ja
	2.11c) Het werkwoordgebruik in de eind- en toetstermen is eenduidig en sluit goed aan bij bestaande taxonomieën (zoals bijvoorbeeld van Bloom of Romiszowsky).	ja
	2.11d) Het aantal vragen geeft een voldoende dekking van het meetdoel.	ja
2.12	Relevante vakinhoudelijke experts beoordelen de toetsspecificaties op hun geschiktheid.	ja
2.13	Het doel van deze beoordeling, het proces waarmee de beoordeling wordt uitgevoerd en de resultaten van de beoordeling worden gedocumenteerd.	deels
2.14	De kwalificaties, relevante ervaringen en demografische kenmerken van relevante vakinhoudelijke experts worden gedocumenteerd.	deels

Fase 3: Constructie en normering Duits vwo 2022-1		Aanwezig?
3.1	De vragen zijn correct geformuleerd (zie richtlijnen in RCEC; die richtlijnen zijn hieronder ingevoegd als subcriteria).	deels
	3.1a) De vragen of opdrachten zijn gestandaardiseerd.	deels
	3.1b) De vragen of opdrachten zijn zodanig ontworpen dat fouten bij invulling voorkomen worden.	ja
	3.1c) De instructie voor de kandidaat (vaak eerste bladzijde van toetsboekje) is gestandaardiseerd, volledig en duidelijk.	deels
	3.1d) De instructie voor de kandidaat dient minimaal te bevatten: aantal vragen, wijze waarop antwoorden worden gegeven, (deel) score per vraag of opdracht, de maximaal te behalen score en de cesuur; toegestane hulpmiddelen; beschikbare tijd en wat ingeleverd moet worden bij afronding; beoordelingspunten bij open vragen.	deels
	3.1e) De kwaliteit van de lay-out en vormgeving is in orde.	ja
3.2	De vragen zijn voldoende gedetailleerd, zodat examenkandidaten de vraag kunnen beantwoorden zoals bedoeld.	ja

<b>Fase 3: Constructie en normering Duits vwo 2022-1</b>		<b>Aanwezig?</b>
3.3	De moeilijkheidsgraad van de vraag is afgestemd op de beoogde doelgroep.	ja
	3.3a) Er heeft een pretest plaatsgevonden om informatie te verkrijgen over de moeilijkheidsgraad (kwantitatieve evaluatie).	nee
	3.3b) Er heeft een kwalitatieve evaluatie met deskundigen plaatsgevonden om informatie te verkrijgen over de moeilijkheidsgraad.	ja
3.4	Er wordt een norm/cesuur verstrekt die via een wetenschappelijke methodiek (standaardbepaling) is bepaald.	ja
3.5	De manier waarop de norm wordt bepaald is gedocumenteerd.	ja
3.6	De standaardbepaling is correct uitgevoerd.	n.v.t.
	3.6a) De standaardbepalingsmethode is op de juiste wijze uitgevoerd.	n.v.t.
	3.6b) De vakdeskundigen/experts die bij de standaardbepaling betrokken zijn, zijn naar behoren geselecteerd en getraind.	n.v.t.
	3.6c) Er is voldoende overeenstemming tussen de beoordelaars.	n.v.t.
3.7	De procedures die worden gebruikt om items te ontwikkelen, beoordelen en uitproberen en om (indien van toepassing) items uit de vragenbank te selecteren, zijn gedocumenteerd.	ja
3.8	Er is gedocumenteerd in hoeverre het inhoudsdomen van de toets het domein vertegenwoordigt, dat is gedefinieerd in de toets specificaties.	ja
3.9	Er is gedocumenteerd wat de kwalificaties, relevante ervaringen en demografische kenmerken zijn van degenen die de ontwikkelde items beoordelen, evenals de instructies en training die zij t.b.v. het beoordelingsproces krijgen.	deels
3.10	Er is een rationale voor de toets waarin verantwoording wordt afgelegd over de samenstelling van het examen.	ja
<b>Fase 4: Afname (n.v.t.)</b>		

<b>Fase 5: Beoordeling, verwerking en analyse Duits vwo 2022-1</b>		<b>Aanwezig?</b>
5.1	Procedures voor het scoren en scorecriteria zijn gedocumenteerd.	ja
5.2	Het proces van het selecteren, trainen, kwalificeren en monitoren van degene die scoren is gedocumenteerd.	ja
5.3	Er zijn twee onafhankelijke beoordelaars die het examen apart van elkaar scoren.	ja
5.4	De twee onafhankelijke scores worden allebei gerapporteerd.	nee
5.5	Er is gedocumenteerd hoe de uiteindelijke score voor het examen wordt berekend.	ja
5.6	Op elk scoreniveau worden in de beoordelingsvoorschriften meerdere voorbeelden van antwoorden gegeven.	ja
5.7	Er wordt een Toets- en itemanalyse uitgevoerd onder voldoende kandidaten.	ja
5.8	De Toets- en Itemanalyse maakt de betrouwbaarheid van de toetscores zichtbaar met behulp van Cronbachs coëfficiënt alfa of een andere algemeen geaccepteerde maat voor betrouwbaarheid.	ja
5.9	De betrouwbaarheid die berekend is geeft geen aanleiding om te twijfelen aan de validiteit van de toetsscores.	ja
5.10	Indien sprake is van open vragen wordt via statistisch onderzoek de mate van beoordelaarsovereenstemming onderzocht.	nee

**Fase 6: Registreren en communiceren: n.v.t.**

<b>Fase 7: Evaluatie Duits vwo 2022-1</b>		<b>Aanwezig?</b>
7.1	Toetsspecificaties worden aangepast als er significante veranderingen zijn in het vertegenwoordigde domein.	ja
7.2	Bewijzen voor (inhouds)validiteit en betrouwbaarheid zijn vastgelegd.	deels
7.3	Bewijzen voor betrouwbaarheid (waaronder de mate van beoordelaarsovereenstemming) zijn vastgelegd.	deels
7.4	Procedures rondom toetsontwikkeling zijn vastgelegd.	ja
7.5	Rollen, verantwoordelijkheden en taakverdeling rondom toetsontwikkeling zijn vastgelegd.	ja
7.6	Er wordt onderzoek gedaan onder de stakeholders (belangenvertegenwoordigers van kandidaten en/of van opleiders) die zijn betrokken bij het gebruik van de toets.	ja
7.7	Er is een verslag (jaarrapport, evaluatieverslag) waarin verantwoording is afgelegd over de ontwikkeling, samenstelling, afname, beoordeling, scores en uitgevoerde analyses van de toets.	deels

