

Janneke Lommertzen,
Marian Bruggink,
Ed Smeets



Onderzoek naar de validiteit van het toetsproces van examens

Economie vmbo kb flex 2023 en Kunst (algemeen) havo 2023 eerste tijdvak

ResearchNed

Juli 2024

en EXPERTISECENTRUM
NEDERLANDS

KBA Nijmegen

Onderzoek naar de validiteit van het toetsproces van examens - Economie vmbo kb flex 2023 en Kunst (algemeen) havo 2023 eerste tijdvak

Janneke Lommertzen, Marian Bruggink, Ed Smeets

Juli 2024

© 2024 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

1	Achtergrond	1
1.1	Aanleiding	1
1.2	De onderzochte examens in dit rapport	1
1.3	Validiteit van toetsen	2
1.4	Aanpak	3
1.5	Examencyclus Cito en CvTE	5
2	Onderzoekskader	8
2.1	Fase 1. Basisontwerp	9
2.2	Fase 2. Construeren toetsmatrijs	10
2.3	Fase 3. Construeren toets	11
2.4	Fase 4. Afnemen	12
2.5	Fase 5. Beoordelen, verwerken, analyseren en normeren	13
2.6	Fase 6. Registreren en communiceren	15
2.7	Fase 7. Evalueren en verbeteren	15
2.8	Relatie onderzoekskader en examencyclus Cito/CvTE	15
3	Onderzoek validiteit CE Kunst (algemeen) havo 2023 eerste tijdvak	16
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 Kunst (algemeen) havo 2023 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	20
3.2.4	Fase 5: Beoordeling, verwerking, analyse en normering	22
3.2.5	Fase 7: Evaluatie	23
3.3	Samenvattend	24
4	Onderzoek validiteit CE Economie 2023 vmbo kb flexibel digitaal examen	25
4.1	Beschrijving van het examen	25
4.1.1	Domeindefinitie	25
4.1.2	Domeinrepresentatie	26
4.1.3	Domeinrelevantie	26
4.2	Samenvatting onderzoek validiteit CE Economie vmbo kb flex digitaal 2023	27
4.2.1	Fase 1: Basisontwerp	27
4.2.2	Fase 2: Opstellen toetsspecificaties en toetsmatrijs	28
4.2.3	Fase 3: Constructie	29
4.2.4	Fase 5: Beoordeling, verwerking, analyse en normering	30
4.2.5	Fase 7: Evaluatie	31
4.3	Samenvattend	32

5	Conclusie en aanbevelingen	33
	Literatuur	35
	Bijlage 1: Overzicht van figuren en tabellen	37
	Bijlage 2: Checklist Centraal examen Kunst (algemeen) havo 2023 eerste tijdvak	38
	Bijlage 3: Checklist Centraal examen Economie vmbo kb flex 2023	41

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 is 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.¹

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.^{2,3} 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 en mede naar aanleiding van de Motie Jadnanansing: Transitie in onderwijstoezicht⁴, laat CvTE de validiteit van centrale examens periodiek onderzoeken.

CvTE heeft ResearchNed gevraagd om te onderzoeken of de procedures die het CvTE en Cito hanteren voor de borging van de validiteit van een selectie van examens voldoen aan (inter)nationale kwaliteitsstandaarden. ResearchNed voert dit onderzoek uit als hoofdaannemer, in samenwerking met Expertisecentrum Nederlands en KBA Nijmegen.

1.2 De onderzochte examens in dit rapport

Dit rapport beschrijft het onderzoek naar het centraal examen Kunst (algemeen) havo 2023 en het flexibele digitale examen economie voor vmbo kaderberoepsgerichte leerweg (vmbo kb) 2023.

ResearchNed heeft eerdere onderzoeken naar de validiteit van het toetsproces van centrale examens uitgevoerd: In 2023 onderzochten zij (in samenwerking met Expertisecentrum Nederlands en KBA Nijmegen) twee examens naar moderne talen (Nederlands vmbo gt en Duits vwo) uit 2022⁵, en in 2022 twee

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., Bruggink, M., Smeets, E. (2023). Validiteit van het toetsproces van examens Nederlands vmbo gl/tl en Duits vwo 2022 eerste tijdvak. Onderzoek in opdracht van CvTE. ResearchNed in samenwerking met Expertisecentrum Nederlands en KBA Nijmegen: Nijmegen.

aatschappijgerichte examens (Geschiedenis vmbo gl/tl en economie havo) uit het eerste tijdvak van 2021 (in samenwerking met CINOP)⁶. In de jaren 2016-2019 heeft RCEC enkele examens onderzocht.⁷

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 hebben gehanteerd aan (inter)nationale kwaliteitsstandaarden?’

Voor zowel het examen Kunst (algemeen) als het examen Economie kb-flex is gebruik gemaakt van het digitale platform Facet. Voor het examen Kunst (algemeen) werd het platform Facet gebruikt om de opgaven en bronnen te tonen⁸ en moesten de antwoorden op papier worden gegeven. Het onderzochte examen economie kb betrof een volledig digitaal examen, waarbij ook de antwoorden van de examenkandidaten digitaal zijn verwerkt. Er was ook een papieren examen economie kb, maar dat valt buiten de scope van dit onderzoek.

1.3 Validiteit van toetsen

Er bestaan verschillende opvattingen over validiteit van toetsen^{9,10}. *Inhoudsvaliditeit* 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 examens⁷. 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.¹¹ 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

6 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

7 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.

8 Centrale examens kunstvakken in facet - informatiebrochure 2023 | Versie 2.0, geraadpleegd op 1 februari 2024, van: https://www.examenblad.nl/system/files/2023/maart/centrale_examens_kunstvakken_in_facet_informatiebrochure_2023_versie_2_1.pdf

9 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.

10 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.

11 Wools, S., (2013). De validiteit van toetsscores. In: P. F. Sanders (Red.). *Toetsen op School* (pp. 69 - 83). Arnhem: Cito.

Council on Measurement in Education (NCME) zijn opgesteld¹². 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’¹³.

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 proces¹⁴. In paragraaf 2.3 gaan we hier nader op in.

1.4 Aanpak

Als basis voor dit onderzoek is het onderzoekskader gebruikt dat ResearchNed heeft ontwikkeld in samenwerking met CINOP¹⁵. 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.^{16,17}

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 onderzochte centrale examens voldoen aan de gestelde criteria.

Naar aanleiding van onze ervaringen bij eerdere onderzoeken naar de validiteit van het toetsproces van centrale examens is het analysekader aangescherpt. De volgende wijzigingen aan het kader zijn doorgevoerd:

- Criterium 1.7 is uitgebreid met de bijzin: ‘of er wordt verwezen naar de kaders waarop het examen is gebaseerd’;
- Het oorspronkelijke criterium 2.5: ‘In de constructieopdracht staat de voorgestelde lengte van de toets’ is verwijderd, omdat hij overlapt met criterium 2.2: ‘In de toetsspecificaties staan het aantal vragen met bijbehorende scorepunten’. Als gevolg daarvan is de nummering van de volgende criteria bij fase 2 gewijzigd.
- ‘De standaardbepalingsmethode is op de juiste wijze uitgevoerd’ (oorspronkelijk criterium 3.6a, nu 3.4a) is gewijzigd in ‘De standaardbepaling is op de juiste wijze uitgevoerd’.

12 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

13 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.

14 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.

15 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.

16 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

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

- Er is ‘indien van toepassing’ toegevoegd aan het oorspronkelijke criterium 3.7, nu 3.5: ‘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’.
- ‘Er is gedocumenteerd in hoeverre het inhoudsdomen van de toets het domein vertegenwoordigt, dat is gedefinieerd in de toetsspecificaties.’, (oorspronkelijk criterium 3.8, nu criterium 3.6) is gewijzigd in: ‘Er is gedocumenteerd in hoeverre de inhoud van de toets het domein vertegenwoordigt, dat is gedefinieerd in de toetsspecificaties.’
- De oorspronkelijke criteria 3.4 en 3.5 zijn nu opgenomen in fase 5: Constructie en Normering, als criterium 5.11: ‘Er wordt een norm/cesuur verstrekt die via een wetenschappelijke methodiek is bepaald’ en criterium 5.12: ‘De manier waarop de norm wordt bepaald is gedocumenteerd’. Omdat in de werkwijze van CvTE en Cito de normering pas in deze fase wordt vastgesteld via de N-term, na afname, beoordeling en analyse van het examen.

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 Cito¹⁸. Op basis hiervan is een lijst opgesteld met documenten die via CvTE is opgevraagd (Bijlage 2). In bijlage 2 is daarnaast ook 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 de examens van twee moderne talen in 2022 en van twee maatschappijgerichte examens in 2021¹⁹, dat aansluit bij de onderzoeken daarvoor. Nadat alle bronnen zijn geïnventariseerd en geanalyseerd is het CvTE gevraagd om feedback te leveren op het conceptrapport en waar nodig op enkele punten toelichting, c.q. aanvullende informatie aan te leveren. Waar nodig deed CvTE navraag bij betrokkenen bij Cito, waarna ook deze input is verwerkt in de beschrijving van de onderzoeksresultaten.



Figuur 1.1: Tijdslijn totstandkoming centrale examens 2023. Bron: CvTE.

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

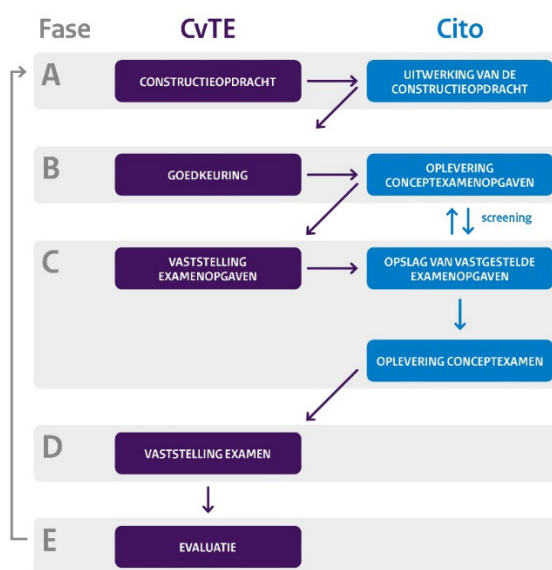
18 Schema samenwerking tussen CvTE en Cito voor de constructie en vaststelling van centrale examens vo. Geraadpleegd van: <https://www.examenblad.nl/onderwerp/examencyclus/2023>, op: 17-5-2023.

19 Lommertzen, J., Bruggink, M., Smeets, E. (2023). Validiteit van het toetsproces van examens Nederlands vmbo gl/tl en Duits vwo 2022 eerste tijdvak. Onderzoek in opdracht van CvTE. ResearchNed in samenwerking met Expertisecentrum en KBA Nijmegen: Nijmegen. En: Lommertzen, J., Janssen, J. & Heijsters, L. (2022). Onderzoek naar de inhoudsvaliditeit van het toetsproces van twee maatschappijgerichte centrale examens. Onderzoek in opdracht van CvTE.

discrepancies 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. Hierbij moet worden opgemerkt, dat het hele proces van totstandkoming van een examen van begin tot einde bijna twee jaar duurt (zie figuur 1.1), waardoor eventuele effecten van dergelijke heroverwegingen niet op korte termijn te verwachten zullen zijn.

1.5 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 1.2).²⁰ Deze examencyclus beschrijft vooral de constructie- en samenstellingsfase. Dit traject neemt doorgaans twee jaar in beslag.



Figuur 1.2: 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.²⁰

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,

²⁰ <https://www.examenblad.nl/onderwerp/examencyclus/2023>

werkzaam in het naast liggende vervolgonderwijs, docenten uit het voortgezet onderwijs en adviseurs van Cito (toetsdeskundigen) en SLO (curriculumspecialisten).^{20,21}

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 de examenvragen een reproductief en/of productief karakter moeten 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 ronden 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.²⁰

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 examinerator 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 examen deden niet goed te vergelijken met die van de leerlingen uit de jaren daarvoor. In deze jaren werd de N-term per examen bepaald aan de hand van de scoreverdeling van alle kandidaten voor dat vak; In 2023 wordt de score afgezet tegen de scores van andere leerlingen in 2023. De N-term wordt bepaald aan de hand van de verschillende technische gegevens. De scores van de leerlingen in 2023, het gemiddelde cijfer uit de periode 2015-2019 voor het vak, in combinatie met de historische N-term en het docentenoordeel geven een goede indicatie waar de grens ligt tussen voldoende en onvoldoende voor een examen.²²

21 Actuele syllabi zijn te vinden op: examenblad.nl

22 Normeren centrale examens 2023: technische analyse dit jaar anders, geraadpleegd op 12 maart 2024, van: <https://www.examenblad.nl/2023/onderwerpen/normering-centrale-examens#normeren-in-reguliere-jaren-uitgangspunt-is-gelijkblijvende-vaardigheid>

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 en de voorzitter²³ 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 examenlijn van CvTE) en leerlingen (via LAKS) mee. Hierbij worden zij ondersteund door de (vakinhoudelijke) toetsdeskundige van Cito. 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 gepubliceerd in de Staatscourant d.d. 09-04-2023²⁴ worden de toetsscores van kandidaten omgezet naar cijfers en wordt de zak/slaaggrens bepaald.²⁵

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 en CvTE²⁶.

23 Voorzitter vaststellingscommissie is altijd een vakexpert uit het hoger onderwijs

24 Regeling van het College voor Toetsen en Examens van 6 februari 2023, nummer CvTE-23.00408, houdende wijziging van Regeling omzetting scores in cijfers VO in verband met het vervallen van het derde tijdvak dat werd ingesteld om de gevolgen van COVID-19 voor de voorbereiding van de centrale examens weg te nemen. In Staatscourant. Nr. 6985, 6 maart 2023. Geraadpleegd op 24-6-2024 van: <https://zoek.officielebekendmakingen.nl/stcrt-2023-6985.html>

25 Normeringsproces Centrale examens vo: geraadpleegd op 24-6-2024 van: https://www.examenblad.nl/system/files/2024-03/infographic-normering-centrale-examens-vo_versie_interactief.pdf

26 Quickscans worden gepubliceerd op Examenblad.nl: zie o.a. https://www.examenblad.nl/system/files/exam-document/2023-10/23158_Quickscan_HAVO%20kunst%2028algemeen%29%202023%20tijdvak%201.pdf en https://www.examenblad.nl/system/files/exam-document/2023-11/63261_Quickscan_KB_economie%202023%20tijdvak_1.pdf, beide geraadpleegd op 24-6-2024

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)²⁷. 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: Toetscyclus²⁸

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

27 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.

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

domeinrepresentatie en -relevantie²⁹. 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 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 kwaliteitsstandaarden³⁰ 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.

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
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 of er wordt verwezen naar de kaders waarop het examen is gebaseerd.	Amerikaanse standaarden: st. 1.2

(Bron verwijst naar de herkomst van de criteria)

29 Het onderzoekskader gaat uit van examenontwikkelaars, waar in de Nederlandse situatie worden deze taken grotendeels door Cito en deels door CvTE uitgevoerd, zie paragraaf 1.5. CvTE is opdrachtgever aan Cito en stelt vast of de examens voldoen.

30 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

Verder zouden toetsontwikkelaars conform de Amerikaanse standaarden (standaard 1.2³⁰) een toetsplan moeten opstellen waarin de beoogde interpretatie van de toetsscores (i.c. beheersing van een bepaald omschreven leerstofdoel) 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 centraal examenstof en het examen wordt vastgelegd. Deze toetsspecificaties in de constructieopdracht gaan onder meer over de inhoud, vorm, de lengte van de toets, psychometrie, scoring en afname van het examen.

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, of varianten daarop zijn veelgebruikt³¹.

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 constructieopdracht staat de hoeveelheid tijd die is toegestaan voor de toets.	Amerikaanse standaarden: st. 4.2
2.6	In de toetsspecificaties staan de gewenste psychometrische eigenschappen van de items en de toets als geheel.	Amerikaanse standaarden: st. 4.2
2.7	In de toetsspecificaties wordt de volgorde van items gedefinieerd.	Amerikaanse standaarden: st. 4.2
2.8	Er is een toetsmatrijs.	Amerikaanse standaarden: st. 4.2

31 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.

Vervolg Tabel 2.2: Checklist Fase 2: Construeren toetsmatrijs

	Checklist Fase 2	Bron
2.9	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
2.10	De toetsmatrijs is een adequate representatie van het meetdoel.	Amerikaanse standaarden: st. 4.2
2.10a	De eind- en toetstermen representeren het meetdoel.	Amerikaanse standaarden: st. 4.2
2.10b	De eind- en toetstermen sluiten aan op de inhoud en het vereiste beheersingsniveau van het betreffende meetdoel	Amerikaanse standaarden: st. 4.2
2.10c	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
2.10d	Het aantal vragen geeft een voldoende dekking van het meetdoel.	Amerikaanse standaarden: st. 4.2
2.11	Relevante vakinhoudelijke experts beoordelen de toetsspecificaties op hun geschiktheid.	Amerikaanse standaarden: st. 4.6
2.12	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
2.13	De kwalificaties, relevante ervaringen en demografische kenmerken van relevante vakinhoudelijke experts worden gedocumenteerd.	Amerikaanse standaarden: st. 4.6

2.3 Fase 3. Construeren toets

Bij het construeren van het examen moeten eerst opgaven worden ontwikkeld. Vervolgens kan het examen worden samengesteld.

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. Daarnaast 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 worden. Ook dienen de processen die worden gebruikt om hen te trainen en te begeleiden bij deze activiteiten te worden vastgelegd.

In deze constructiefase is het belangrijk om 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

	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 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	De standaardbepaling is correct uitgevoerd.	Amerikaanse standaarden: st. 5.21
3.4a	De standaardbepaling is op de juiste wijze uitgevoerd.	Amerikaanse standaarden: st. 5.21
3.4b	De vakdeskundigen/experts die bij de standaardbepaling betrokken zijn, zijn naar behoren geselecteerd en getraind.	Amerikaanse standaarden: st. 5.21
3.4c	Er is voldoende overeenstemming tussen de vakdeskundige experts.	Amerikaanse standaarden: st. 5.22
3.5	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
3.6	Er is gedocumenteerd in hoeverre de inhoud van de toets het domein vertegenwoordigt, dat is gedefinieerd in de toetsspecificaties.	Amerikaanse standaarden: st. 4.12
3.7	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. ³²	Amerikaanse standaarden: st. 4.8
3.8	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

³² Bij het gebruik van empirische analyses gelden andere aanvullende criteria. Zie in de Amerikaanse standaarden de criteria 4.09 t/m 4.11.

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 en normeren

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 en daarna wordt de norm (ofwel cesuur, zak/slaaggrens genoemd) van het examen bepaald.

Met het psychometrisch onderzoek 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 kwaliteit³³ 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 tonen³⁴. 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 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.

Normering betreft de regels die worden opgesteld om de toetsscores in cijfers of een waardering (slagen/zakken) 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

33 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 opgaven 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.

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

standaardbepaling dient volgens een vooraf bepaalde procedure te worden uitgevoerd door een onafhankelijk orgaan³⁵.

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.³⁶ Tijdens de standaardbepaling wordt vastgesteld wat kandidaten zouden moeten beheersen en hoeveel punten er moeten worden behaald om te kunnen slagen.^{35,36} Vervolgens zijn er ook methoden nodig om de norm te handhaven. Bij de centrale eindexamens gebeurt dit via de procedure met de N-term.³⁷

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 en normeren

	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 ³⁸)	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
5.11	Er wordt een norm(cesuur) verstrekt die via een wetenschappelijke methodiek is bepaald.	Amerikaanse standaarden: st. 5.21
5.12	De manier waarop de norm wordt bepaald is gedocumenteerd.	Amerikaanse standaarden: st. 5.21

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

36 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.

37 Regeling van het College voor Toetsen en Examens van 6 februari 2023, nummer CvTE-23.00408, houdende wijziging van Regeling omzetting scores in cijfers VO in verband met het vervallen van het derde tijdvak dat werd ingesteld om de gevolgen van COVID-19 voor de voorbereiding van de centrale examens weg te nemen. In Staatscourant. Nr. 6985, 6 maart 2023. Geraadpleegd op 24-6-2024 van: <https://zoek.officielebekendmakingen.nl/stcrt-2023-6985.html>

38 Eindexamenbesluit VO art. 41 en 42.

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.

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 1.5 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	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 en normering	E: Evaluatie
6. Registratie en communicatie van resultaten	E: Evaluatie
7. Evaluatie en verbeteringen	E: Evaluatie

3 Onderzoek validiteit CE Kunst (algemeen) havo 2023 eerste tijdvak

3.1 Beschrijving van het examen

Vanaf 2009 is er voor de kunstvakken havo, naast de schoolexamens per discipline (beeldende vormgeving/dans/drama/muziek), samen één centraal examen. Dit centraal examen is voor al deze kunstvakken hetzelfde: Kunst (algemeen). Kunst (algemeen) kan ook als extra vak met uitsluitend een centraal examen worden gekozen.

Het centrale examen Kunst (algemeen) havo werd deels digitaal en deels op papier afgenomen: de vragen en antwoorden werden digitaal in Facet aangeboden en kandidaten gaven hun antwoorden op papier. Het examen bestond uit 31 vragen, waarvan één gesloten en 30 open vragen. Voor het examen waren maximaal 66 punten te halen. De afname van het examen vond plaats op dinsdag 23 mei 2023 tussen 9.00 en 12.00.

3.1.1 Domeindefinitie

Het centraal examen Kunst (algemeen) havo 2023 heeft betrekking op drie domeinen van het examenprogramma:

- A. *Vaardigheden*
- B. *Invalshoeken voor reflectie*
- C. *Onderwerpen*

Het centraal examen heeft betrekking op de domeinen B en C in combinatie met de vaardigheden uit domein A. In de toetsmatrijs is de verdeling van de scorepunten over de vaardigheden en subdomeinen weergegeven. In het examenprogramma zijn de drie domeinen uitgewerkt in acht eindtermen:

Domein A: Vaardigheden

1. De kandidaat kan:

- de belangrijkste termen en begrippen hanteren uit de kunstdisciplines beeldende vormgeving, dans, drama en muziek die voorwaardelijk zijn voor adequate receptie en reflectie, en noodzakelijk voor begrip van verbanden tussen kunst en cultuur;
- informatie over kunst en cultuur herkennen, benoemen en toepassen om verbanden aan te geven;
- bij het reflecteren bronnenmateriaal op een adequate wijze hanteren;

Domein B: Invalshoeken voor reflectie

Subdomein B1: Kunst en religie, levensbeschouwing

2. De kandidaat kan aangeven met welke visies, doelen, middelen en inhoud de kunsten religieuze en/of levensbeschouwelijke uitgangspunten vertolken.

Subdomein B2: Kunst en esthetica

3. De kandidaat kan aangeven welke ideeën over schoonheid in kunst en kunstwaardering een rol spelen.

Subdomein B3: Kunstenaar en opdrachtgever; politieke en economische macht

4. De kandidaat kan aangeven welke invloed opdrachtgevers en politieke ideeën hebben op de rol en de inhoud van kunst en op de positie van de kunstenaar.

Subdomein B4: Kunst en vermaak

5. De kandidaat kan aangeven hoe vorm en inhoud bepaald worden door de vermaaksfuncties van kunst in relatie tot het daarbij betrokken publiek.

Subdomein B5: Kunst, wetenschap en techniek

6. De kandidaat kan aangeven hoe kunst en wetenschap/techniek op elkaar inwerken.

Subdomein B6: Kunst intercultureel

7. De kandidaat kan aangeven hoe Westerse en niet-Westerse kunst en cultuur elkaar wederzijds beïnvloeden.

Domein C: Onderwerpen

8. De kandidaat kan de eindtermen van domein A en B toepassen op voorbeelden uit de volgende onderwerpen:

- de burgerlijke cultuur van Nederland in de zeventiende eeuw;
- de cultuur van het moderne in de eerste helft van de twintigste eeuw;
- de massacultuur vanaf 1950

3.1.2 Domeinrepresentatie

Domeinrepresentatie betreft de mate waarin een examen het domein, zoals dat gedefinieerd is door de toetspecificaties 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. In de toetsmatrijs van het examen Kunst (algemeen) havo 2023 staat de beoogde verdeling van scorepunten over de algemene vaardigheden, over de vakinhoudelijke eindtermen en over het gedragsaspect. Een gerealiseerde verdeling van de scorepunten over de verschillende categorieën is te vinden in de toetsrationale.

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 door screeners en constructiegroepleden van Cito die betrokken zijn bij de constructie van de opgaven. En ook de vak- en toetsdeskundigen van de vaststellingscommissie kijken hier grondig naar: ze keuren opgaven die in hun ogen niet relevant zijn af. Voor het examen Kunst (algemeen) havo 2023 is het examen middels een vinkje in Facet en met een mail van de voorzitter van de vaststellingscommissie Kunst (algemeen) vastgesteld, onder voorbehoud van het verwerken van enkele aanpassingen.

Voor alle examens geldt dat er na afname ook gegevens van andere vakexperts beschikbaar zijn over het examen. Middels een enquête (de QuickScan) wordt aan docenten/ examinatoren hun waardering voor het examen Kunst (algemeen) havo 2023 gevraagd. 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 494 respondenten van 296 scholen waren als volgt: ruim de helft van de docenten vond het examen moeilijk (59%), gevolgd door 32 procent die het examen niet te moeilijk/niet te makkelijk vond en 9 procent die het examen te moeilijk

vond. Slechts 1 procent (4 docenten) vonden het examen makkelijk en geen enkele docent heeft aangegeven dat het examen te makkelijk was. Over de lengte van het examen zegt ongeveer de helft van de docenten het examen precies goed te vinden en de andere helft geeft aan dat het examen te lang was. Slechts 1 procent (3 docenten) vond het examen te kort. Een meerderheid van de docenten (54%) vond de inhoudelijke aansluiting van het examen bij het gegeven onderwijs voldoende. Bijna een kwart van de docenten vond de aansluiting onvoldoende (23%) en 18 procent vond de aansluiting goed. De vraag is door een kleine groep docenten met 'zeer goed' (2%) of 'slecht' (3%) beantwoord. De ruime meerderheid van de docenten gaf het examen een voldoende: 33% een 6; 39% een 7; 2% een 8; 1% een 9. De overige docenten gaven een onvoldoende: 19% een 5, 6% een 4 en 1% een 3.

In het evaluatieformulier van het examen Kunst (algemeen) havo 2023 wordt door de vaststellingscommissie aangegeven dat er geen echte kritiekpunten zijn op het examen. Als enig zwak punt wordt de hoeveelheid tekst en introductie van nieuwe informatie laat in het examen genoemd (bij vraag 30). Het correctievoorschrift werd als heel hanteerbaar en uitgebreid beoordeeld.

3.2 Samenvatting onderzoek validiteit CE Kunst (algemeen) havo 2023 eerste tijdvak

Aan de hand van het onderzoekskader, beschreven in hoofdstuk 2, is het centraal examen Kunst (algemeen) havo uit het eerste tijdvak van 2023 geanalyseerd. In deze paragraaf beschrijven we de belangrijkste bevindingen. Gedetailleerde bevindingen per criterium staan beschreven in Bijlage 3 'Checklist Centraal Examen Kunst (algemeen) havo 2023 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 en de syllabus, omvatten de informatie uit het toetsplan.

Aan bijna alle gestelde criteria voor een toetsplan (1.2 - 1.6) wordt voldaan. In de documenten, en ook deels in de wet vastgelegd, staat de doelgroep van het examen aangegeven, zijn het domein en het gebruiksdoel van het examen gedefinieerd en is duidelijk in welke contexten de toetsscores gebruikt moeten worden. De procedures voor het afnemen en scoren van het examen staan ook in diverse documenten beschreven. Wat betreft criterium 1.7 over de theorie of kaders waarop het examen gebaseerd is, kunnen we niets terugvinden. Met ingang van 2022 is de syllabus Kunst (algemeen) herzien. De nieuwe syllabus is samengesteld door een syllabuscommissie bestaande uit experts. Deze syllabuscommissie heeft de accenten en de specificaties van de cultuurhistorische periodes (zie de onderwerpen bij domein C) bestudeerd en heeft deze vervolgens aangepast, verruimd en/of geactualiseerd, maar er wordt daarbij niet verwezen naar bestaande kaders of theorieën, of waarom deze accenten en specificaties van de cultuurhistorische periodes zijn gekozen. Dit kan worden verklaard doordat er geen (inter)nationale theoretische kaders te lijken bestaan voor dit vak.

3.2.2 Fase 2: Opstellen toetsspecificaties en toetsmatrijs

Fase 2 van het onderzoekskader gaat over het opstellen van toetsspecificaties en een toetsmatrijs. In de constructieopdracht Kunst (algemeen) havo 2023 zijn toetsspecificaties opgenomen, zoals aantal vragen met bijbehorende scorepunten, de toetsvorm, het soort vragen en de hoeveelheid tijd die is toegestaan voor het examen. Daarmee wordt voldaan aan criteria 2.1, 2.2, 2.3 en 2.5. Vanuit de constructieopdracht wordt verwezen naar de regeling toegestane hulpmiddelen voor de centrale examens havo en vwo 2023 (zie

criterium 2.4). In deze regeling staan de toegestane hulpmiddelen uitgebreid beschreven. In de ‘Regeling eveneens toegestane hulpmiddelen voor leerlingen met een beperking 2023’ staat uitgebreid beschreven welke hulpmiddelen leerlingen met een beperking (zoals een visuele of auditieve beperking) mogen gebruiken. In de constructieopdracht staan de randvoorwaarden voor de opbouw van het examen beschreven. Hierin staat de eerste vraag van een opgavenblok niet te moeilijk, te uitgebreid of te tijdrovend dient te zijn en dat de laatste vragen in het laatste opgavenblok niet te moeilijk dienen te zijn. Daarmee wordt ook voldaan aan criterium 2.7 over het definiëren van de volgorde van de items. Aan criterium 2.6 wordt niet voldaan. Specifieke gewenste psychometrische eigenschappen per item en in totaal staan niet in de constructieopdracht. Er staat niet vermeld wat de gewenste betrouwbaarheid is, hoeveel gemakkelijke en hoeveel moeilijke vragen dienen te worden opgenomen en wat de gewenste gemiddelde p-waarde en de spreiding van p-waarden van de items is. Wel wordt er in verschillende documenten in algemene zin iets gezegd over de gewenste betrouwbaarheid. Cito verwijst hiervoor naar de publicatie *Toetsen op School*³⁹, waar vanuit het Handboek naar wordt verwezen. Hier worden richtlijnen voor en achtergronden van de betrouwbaarheid van centrale examens geschetst. Over de moeilijkheidsgraad van het examen staat in het algemene deel van de constructieopdracht (paragraaf 1.6.2) de algemene opmerking dat examens redelijke eisen aan de kandidaten dienen te stellen en van voldoende niveau dienen te zijn ‘om het civiel effect van het diploma te waarborgen.’ En: ‘Cito bepaalt op grond van technieken, die per toets kunnen variëren, een geschatte moeilijkheid van een toets, maar deelt die bij de oplevering niet mee.’ Dit heeft als reden om de vaststellingscommissie in de gelegenheid te stellen onbevooroordeeld het examen en de moeilijkheidsgraad te kunnen beoordelen.

Criteria 2.8 tot en met 2.10 (a-d) gaan over de toetsmatrijs. Deze is opgenomen in de constructieopdracht. In de toetsmatrijs staat de beoogde verdeling van de scorepunten van de domeinen A, B en C weergegeven. Het gaat hierbij over de verdeling van de scorepunten over informatievaardigheden en standpuntbepaling (domein A), over de subdomeinen (domein B) en over de onderwerpen (domein C). Ook is er een beoogde verdeling over gedragsaspecten in de matrijs opgenomen (i.e. de verhouding tussen het aantal scorepunten voor reproductie- en productievragen).

Afgaand op het oordeel van de vakexperts en toetsdeskundigen is de toetsmatrijs een adequate representatie van het meetdoel (criterium 2.10a, b en d). In de documenten worden geen specifieke taxonomieën genoemd (criterium 2.10c), maar het werkwoordgebruik in het examenprogramma sluit aan bij de taxonomie van Bloom (bijvoorbeeld werkwoorden als ‘herkennen’, ‘benoemen’, ‘toepassen’, ‘beargumenteren’, ‘aangeven hoe’). Over dat laatste werkwoord staat in de syllabus aangegeven dat dit werkwoord activiteiten omvat als ‘beschrijven’, ‘onderzoeken’ en ‘interpreteren’, maar ook ‘beargumenteren’ en ‘onderbouwen’. Aan de criteria voor het beoordelen van de toetsspecificaties door experts (criteria 2.11 - 2.13) wordt grotendeels voldaan. In de bijlage van de constructieopdracht en in het algemene gedeelte van de constructieopdracht staat het productiepad van het examen beschreven waarin ook de goedkeuring van de constructieopdracht is opgenomen. Op advies van de vaststellingscommissie (bestaande uit experts) geeft de clustermanager van het CvTE goedkeuring aan de uitwerking van de constructieopdracht en de toetsspecificaties daarin. Het resultaat is de definitieve constructieopdracht. Wat betreft de kwalificaties van de vakinhoudelijke experts die betrokken zijn bij de beoordeling van de toetsspecificaties, in dit geval de vaststellingscommissie, staan alleen algemene kenmerken globaal beschreven in het document ‘Vademecum - leidraad voor de vaststellingscommissies’, zoals de eis dat een lid lesgeeft aan examenklassen. Leden voor de vaststellingscommissie worden voorgedragen door de vakbond of vakvereniging. De voorzitter van de vaststellingscommissie beoordeelt of nieuwe leden passen binnen het functieprofiel. Tot dit functieprofiel, specifiek voor het vak Kunst (algemeen), hebben we geen beschikking. Wel zijn er voorbeelden beschikbaar gesteld van vacatureteksten voor leden van andere

39 Kuhlmeier, H., Til, van, A., Sanders, P. (2016). *Toetsen op School*. Stichting Cito Instituut voor Toetsontwikkeling Arnhem. Cito, geraadpleegd op: 17-4-2024, van: [toetsen-op-school-voortgezet-onderwijs.pdf](https://www.cito.nl/toetsen-op-school-voortgezet-onderwijs.pdf) (cito.nl).

vaststellingscommissies waarin de volgende (generieke) eisen staan benoemd: bevoegd voor het geven van het vak aan de betreffende doelgroep; goede vakkennis en vaardigheid; recente ervaring met het geven van onderwijs in het vak aan examenklassen van het betreffende schooltype. Ook strekt het tot de aanbeveling wanneer een kandidaat deskundig is op het gebied van examinering en toetsing. Daarnaast zijn er vereisten gericht op de onafhankelijkheid van de leden van de vaststellingscommissie: zo mag een lid van een vaststellingscommissie geen auteur zijn van een les- of onderwijsmethode; geen (recent) lid zijn van de constructiegroep van Cito en geen voorzitter of bestuurslid van een vakinhoudelijke vereniging. Bij andere nevenactiviteiten zal door CvTE een afweging worden gemaakt of die het onafhankelijk functioneren als lid van de vaststellingscommissie in de weg kunnen staan.

Wat betreft deze fase bevelen we aan om vooraf specifieke richtlijnen voor de gewenste psychometrische eigenschappen van de items en van het examen als geheel vast te leggen.

3.2.3 Fase 3: Constructie

Fase 3 betreft de constructie van het examen. Het eerste criterium ('De vragen zijn correct geformuleerd') bestaat uit een aantal richtlijnen (3.1a-3.1e). Aan vrijwel alle richtlijnen wordt voldaan. In bijlage 2 van het algemene gedeelte van de constructieopdracht staan de screeningsprocedures beschreven.

Toetsdeskundigen en/of vakinhoudelijke deskundigen screenen de items op verschillende manieren: inhoudelijk, toetstechnisch, taalkundig, en screening ter voorkoming van errata. Voor de vragen gelden de kwaliteitseisen zoals uitgewerkt in de handboeken. In bijlage 4 van het algemene gedeelte van de constructieopdracht staan de criteria voor uitgangsmateriaal beschreven. De lay-out van digitale examens is conform de specificaties in de handboeken productie digitale examens (zie ook het document met Richtlijnen voor Facet). De instructie voor de kandidaat (zie criterium 3.1d) bevat het aantal vragen, de totale maximale score en per vraag de hoeveelheid punten die behaald kunnen worden met een goed antwoord. De cesuur volgt na afname van het examen (zie fase 5). De beschikbare tijd voor het maken van het examen staat vermeld. Er is geen uitwerkbijlage bij het examen; kandidaten schrijven hun antwoorden op door de school beschikbaar gesteld examenpapier. De toegestane hulpmiddelen, zoals een Nederlands woordenboek, staan niet vermeld in de instructie voor de kandidaat in de Facet-omgeving. Deze toegestane hulpmiddelen staan wel in andere documenten beschreven, zoals de 'Regeling toegestane hulpmiddelen voor de centrale examens havo en vwo 2023' en de informatie voor kandidaten op www.mijneindexamen.nl⁴⁰. Bij het digitale examen zelf (gebaseerd op de oefenomgeving) in Facet worden enkel de hulpmiddelen die beschikbaar zijn in Facet genoemd, zoals het digitale vergrootglas en een digitale markeerstift. CvTE geeft aan bewust geen uitgebreide lijst met toegestane hulpmiddelen op het (papieren en digitale) examen zelf te plaatsen, omdat dit afleidend/storend kan zijn voor kandidaten en er in het geval een kandidaat een hulpmiddel vergeten is, er tijdens de afname van het examen toch niets meer aan gedaan kan worden.

In criterium 3.2 staat dat de vragen voldoende gedetailleerd moeten zijn. Dit is volgens de experts die betrokken zijn bij de constructie en screening van de items het geval. Criteria 3.3a en b gaan over de moeilijkheidsgraad van het examen. Aan deze eisen wordt niet of deels voldaan. Bij de inhoudelijke en toetstechnische screening wordt de moeilijkheidsgraad van de items meegenomen. Een toetsdeskundige of een vakinhoudelijke deskundige screent of het examen aansluit bij het onderwijsniveau en kijkt naar de moeilijkheidsgraad, ook in relatie tot de lengte van het examen. Er heeft echter geen pretest plaatsgevonden om informatie te verkrijgen over de moeilijkheidsgraad. De deskundigen van de

40 Te weten: schrijfmateriaal, inclusief millimeterpapier; tekenpotlood (en puntenslijper); vlakgum; blauw en rood kleurpotlood; geodriehoek; rekenmachine; liniaal met millimeterverdeling; passer (met potloodstiftje); eendelig verklarend woordenboek Nederlands of eendelig woordenboek Nederlands-thuis taal en thuis taal-Nederlands).

vaststellingscommissie zijn, gezien de vaststelling van het examen, akkoord met de moeilijkheidsgraad, maar we hebben geen documentatie over de vergelijking van de moeilijkheid van het examen met een referentie-examen door deskundigen. Het fase D-formulier zou hier meer inzage in kunnen geven, echter, voor het examen Kunst (algemeen) havo 2023 lijkt dit formulier er niet te zijn. Dit zou volgens bijlage 6 van het algemene gedeelte van de constructieopdracht wel gebeurd moeten zijn. Hierin staat dat de eerste stap in het vaststellen van een examen in Facet is dat de vaststellingscommissie de examenvragen en bronnen vaststelt in Facet. De uitwerkbijlage (indien van toepassing) en het correctievoorschrift worden op papier vastgesteld. De voorzitter zet dan nog geen vinkje in Facet, maar hij geeft wel via een fase D formulier aan met in de bestandsnaam 'Facet' dat deze fase in goede orde is gepasseerd. In een mail van de voorzitter van de vaststellingscommissie naar Cito staat dat de voorzitter akkoord is met de handtekeningenversie 'onder voorbehoud van het verwerken van de aanpassingen'. Onduidelijk is welke aanpassingen dat zijn en of deze inderdaad verwerkt zijn.

Voor het examen Kunst (algemeen) havo 2023 is voorafgaand aan het examen geen standaardbepaling uitgevoerd (criterium 3.4). Aan criterium 3.5 wordt voldaan; in bijlage 1 van het algemene gedeelte van de constructieopdracht staat het proces van de ontwikkeling en de vaststelling van een examen beschreven. Er staat ook beschreven wie waar verantwoordelijk voor is en in bijlage 2 staan de screeningsprocedures beschreven (inhoudelijk, toetstechnisch, taalkundig, opmaak etc). Ook aan criterium 3.6 wordt voldaan: in de toetsrationale is vastgelegd in hoeverre het examen de verschillende domeinen vertegenwoordigt. In deze toetsrationale is ook de gerealiseerde samenstelling van het examen in kaart gebracht, zie criterium 3.8. Bij de toetsrationale ontbreekt echter de vergelijking met de beoogde verdeling zoals in de toetsmatrijs is opgenomen. Als we deze twee documenten zelf vergelijken dan zien we verschillen tussen de beoogde verdeling en de gerealiseerde verdeling. Uit de rationale is af te leiden dat een evenwichtige verdeling van de 6 subdomeinen van domein B niet (helemaal) bereikt is en ook de beoogde verdeling wat betreft het gedragsaspect (met toegestane marge) is niet bereikt. Over deze verschillen tussen toetsmatrijs en rationale is geen onderbouwing of verklaring gegeven. Bovendien is in de rationale niet terug te vinden hoe de vragen/ punten zijn verdeeld over de informatievaardigheden en de standpuntbepaling. De gewenste percentages per onderdeel bij domein C zijn wel gehaald.

Criterium 3.7 gaat over de kwalificaties van beoordelaars van de reeds ontwikkelde items en over de instructies en training die deze beoordelaars t.b.v. het beoordelingsproces krijgen. Cito is verantwoordelijk voor het bewaken van de expertise van de constructeurs/screeners en toetsdeskundigen. In bijlage 2 van het algemene gedeelte van de constructieopdracht staat aangegeven door wie (type deskundige) een bepaalde screening wordt gedaan, bijvoorbeeld door de vakdeskundige, de vakspecialist, de toetsdeskundige of de neerlandicus. Op de website van Cito staan vacatureteksten voor constructeurs en worden criteria genoemd als recente ervaring in het gevraagde vak en schooltype, talent voor analyseren en structureren, goede schriftelijke uitdrukkingsvaardigheden en feedback kunnen geven en ontvangen. Als voorwaarde wordt gesteld dat een constructeur geen auteur mag zijn van een onderwijsmethode. Cito geeft aan dat het bij het examen Kunst (algemeen) gaat om twee screeners die ieder werkzaam zijn bij een instelling voor hoger onderwijs, ieder op een ander deelonderwerp van Kunst (algemeen). De screeners zijn benaderd door de toetsdeskundigen van Cito. De kwalificaties op basis waarvan deze screeners zijn geselecteerd zijn o.a. aantoonbare affiniteit met examinering en toetsing én bekendheid met het examenprogramma van Kunst (algemeen). Cito geeft aan dat het van belang is dat een screener in staat is om inhoudelijke sterke feedback te geven. Voor het construeren en screenen van items kunnen deskundigen gebruik maken van de volgende hoofdstukken uit het Handboek 'H7. Checklist open vragen', 'H8. Checklist meerkeuzevragen' en 'H5. handreiking taalgebruik in Centrale Examens'.

Voor deze fase adviseren we om het fase D formulier inclusief de vergelijking met een referentie-examen, ook in te vullen wanneer het examen in Facet wordt vastgesteld. Onze inschatting is dat het niet reëel is om elk jaar een pretest af te nemen, maar dat een vergelijking met een referentie-examen wel haalbaar moet zijn als kwalitatieve evaluatie om voorafgaand aan het examen informatie te krijgen over de moeilijkheidsgraad. Wat betreft de verantwoording over de samenstelling van het examen bevelen we aan om in de toetsrationale de vergelijking te maken met de beoogde verdeling uit de toetsmatrijs en daarbij te beschrijven hoe te grote verschillen tussen beide verdelingen verklaard kunnen worden en hoe men dit een volgende keer probeert te voorkomen. Tot slot adviseren voor deze fase om te overwegen om de *belangrijkste* toegestane hulpmiddelen op een beknopte wijze op de eerste pagina in Facet op te nemen. Hoewel we de redenen begrijpen om geen uitvoerige lijst met hulpmiddelen te vermelden, kan met een compacte opsomming in de instructie voor de kandidaat eventuele discussies tijdens het examen worden voorkomen.

3.2.4 Fase 5: Beoordeling, verwerking, analyse en normering

Fase 5 betreft de beoordeling, verwerking, analyse en normering van het examen na de afname. Het eerste criterium (5.1) gaat over het documenteren van de procedures voor het scoren en de scorecriteria. Het examen Kunst (algemeen) wordt op papier ingeleverd en nagekeken. In het 'correctievoorschrift Havo 2023- Kunst algemeen' worden de procedures, algemene regels en het beoordelingsmodel beschreven. Tijdens de correctie kan de docent het examen inzien via de oefenomgeving van Facet. Het proces van beoordelen door twee onafhankelijke beoordelaars is bij wettelijk besluit vastgelegd (criteria 5.2 en 5.3). Het examenwerk wordt door de eigen docent nagekeken en door een tweede corrector van een andere school. De twee beoordelaars mogen daarbij niet afwijken van het correctievoorschrift.

Docenten worden bij het nakijken van het examen ondersteund door de vakvereniging Onderwijs Kunst en Cultuur (VONKC). VONKC organiseert na de afname van het centraal examen besprekingen van de opgaven en het correctievoorschrift. Beginnende docenten hebben de mogelijkheid om voorafgaand aan de examenperiode examencorrectietrainingen te volgen. Hierin wordt uitgelegd hoe het correctievoorschrift gelezen moet worden en waar de ruimte zit voor de corrector. Dit gebeurt aan de hand van diverse typen examenvragen en antwoorden uit eerdere jaren (zie verslag Centraal Examens 2023). De twee beoordelaars moeten per kandidaat samen tot een score komen. Als dit niet lukt wordt door de Inspectie een derde corrector ingeschakeld die de definitieve score van de betreffende kandidaat bepaalt.

De twee onafhankelijke scores worden niet beide gerapporteerd (criterium 5.4). Beoordelaars (docenten) voeren de scores van de kandidaten in via de applicatie Wolf. Het is niet verplicht om de tweede correctie ook in Wolf in te voeren. De wettelijke procedure wordt hierbij gevolgd. Na invoering van de scores wordt in de applicatie overigens ook aan docenten gevraagd om aan te geven waar volgens hen de grens tussen voldoende en onvoldoende prestatie ligt op dit examen, wanneer ze het vergelijken met de jaren voor de coronapandemie.

Aan de criteria 5.5 tot en met 5.9 wordt voldaan. In het 'correctievoorschrift Havo 2023- Kunst algemeen' is per vraag aangegeven hoe de punten moeten worden toegekend. Op examenblad.nl is de omzettingstabel te vinden waarmee het behaalde aantal punten in een score kan worden omgezet naar een cijfer (criterium 5.5). In het beoordelingsmodel worden per vraag meerdere antwoorden gegeven die goed gerekend dienen te worden (criterium 5.6). Verder is er een toets- en itemanalyse uitgevoerd over 6551 kandidaten (criterium 5.7). Dat is 85 procent van het totale aantal kandidaten havo Kunst (algemeen). In de toets- en itemanalyse staat de Coëfficiënt Alpha per blok van het examen en in zijn geheel gerapporteerd. Per item worden de p-waarde en de rit-waarde gerapporteerd (criterium 5.8). De betrouwbaarheid van de vier blokken van het examen is vrij laag (van 0.40 tot 0.55). Dit heeft waarschijnlijk te maken met het feit dat er relatief weinig items per blok zijn opgenomen. De betrouwbaarheid van het totale examen is goed, namelijk 0.74 (Coëfficiënt Alpha) (criterium 5.9).

Aan criterium 5.10 wordt niet voldaan. Zoals gezegd worden de twee onafhankelijke scores niet beide apart gerapporteerd. De mate van beoordelaarsovereenstemming kan in dat geval niet onderzocht worden. Criteria 5.11 en 5.12 gaan over de bepaling van de norm. De normhandhaving gebeurt via de procedure met de N-term, zoals beschreven in paragraaf 1.5. Het geheel aan regels en procedures rondom de vaststelling van de normeringsterm staat beschreven in een interne notitie over de normering, de informatie op [examenblad.nl](https://www.examenblad.nl)⁴¹ en de Regeling van het College voor toetsen en examens van 6 februari 2023⁴².

Bij deze fase is ons advies om beide onafhankelijke scores in te laten voeren en daarmee de beoordelaarsovereenstemming te onderzoeken en vast te stellen. Juist bij een examen dat vrijwel alleen maar bestaat uit open vragen is het onderzoeken van de beoordelaarsovereenstemming van belang.

3.2.5 Fase 7: Evaluatie

De laatste fase omvat de evaluatie en vastlegging van het hele toetsproces. Na de examenperiode evalueren de vaststellingscommissies en de toetsdeskundigen het examen. Op basis van de evaluatie stelt het CvTE vast of de constructieopdracht en/ of de syllabus moet worden aangepast (criterium 7.1). Bewijzen voor de (inhouds)validiteit (criterium 7.2) zijn terug te vinden in verschillende documenten. In de toetsrationale is de gerealiseerde samenstelling van de toets in kaart gebracht, waarbij de examenonderdelen zoals ze in het examenprogramma en de syllabus beschreven staan, terugkomen. Daarnaast is het examen door verschillende deskundigen gescreend op verschillende aspecten en is het examen door experts vastgesteld. Bewijzen voor de betrouwbaarheid van het examen zijn deels vastgelegd (criterium 7.3). Zoals gezegd is de betrouwbaarheid van het examen uit de TIA goed gebleken, maar is er geen beoordelaarsovereenstemming onderzocht.

Aan criteria 7.4 tot en 7.5 wordt voldaan. In het algemene deel van de constructieopdracht worden procedures rondom toetsontwikkeling beschreven en staat beschreven (wie van) welke organisaties waar verantwoordelijk voor zijn (zie ook paragraaf 1.1. van het verslag Centrale Examens 2023). Ook is er onderzoek gedaan onder de stakeholders die betrokken zijn bij het gebruik van het examen (criterium 7.6). Na afloop van het examen is de QuickScan uitgezet onder docenten. Daarnaast worden de meldingen van docenten bij de Examenlijn, de reacties van docenten tijdens examenbesprekingen en de input vanuit het LAKS meegenomen bij de evaluatie.

Aan criterium 7.7 wordt deels voldaan. Er is niet één compleet verslag beschikbaar met de verantwoording rondom het examen Kunst (algemeen) havo 2023. De formulieren voor fase B, C, D zijn niet ingevuld voor Kunst (algemeen) als verslaglegging van het toetsproces. Er is wel een evaluatieformulier ingevuld waarin kritiekpunten van docenten en leerlingen worden meegenomen, de opvallendheden uit de toets- en itemanalyse en de hanteerbaarheid van het correctievoorschrift. Ook wordt er in dit formulier gevraagd of de vastgestelde moeilijkheid van het examen overeenkomt met vooraf ingeschatte moeilijkheid door de vaststellingscommissie. Hierop is 'ja' geantwoord. Echter kunnen wij de vooraf ingeschatte moeilijkheid niet terugvinden, aangezien er geen D-formulier lijkt te zijn ingevuld.

Onze aanbeveling bij deze fase is om de verantwoording over de ontwikkeling, samenstelling, afname en beoordeling van het examen Kunst (algemeen) havo vollediger en in één document vast te leggen.

41 Zie ook: <https://www.examenblad.nl/2023/onderwerpen/normering-centrale-examens>, geraadpleegd op 26 april 2024

42 <https://zoek.officielebekendmakingen.nl/stcrt-2023-6985.html>, geraadpleegd op 24-6-2024

3.3 Samenvattend

Het examen Kunst (algemeen) havo 2023 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. Aan een aantal criteria wordt echter niet of slechts deels voldaan. Zo worden vooraf niet de gewenste psychometrische eigenschappen per item en in totaal gespecificeerd in de constructieopdracht. Verder ontbreekt het fase D formulier en is daarmee het examen niet vergeleken met een referentie-examen. Een ander belangrijk criterium waar niet aan voldaan wordt, is de verantwoording over de te grote verschillen tussen de toetsmatrijs en de toetsrationale. Tot slot ontbreekt er onderzoek naar de beoordelaarsovereenstemming. Dit is bij een examen met bijna alleen maar open vragen wel wenselijk.

Op basis van deze gevonden hiaten in het toetsproces van het examen Kunst (algemeen) havo 2023 hebben we hierboven enkele aanbevelingen geformuleerd. Samengevat luiden deze aanbevelingen als volgt:

- Leg vooraf de gewenste psychometrische eigenschappen van de items en van de toets als geheel vast.
- Vermeld de *belangrijkste* toegestane hulpmiddelen op een beknopte wijze in de instructie voor de kandidaat, om discussie te voorkomen tijdens het examen.
- Vul -naast de procedures voor het vaststellen van examen in Facet- ook het fase D formulier in, inclusief de vergelijking met een referentie-examen.
- Maak in de toetsrationale de vergelijking tussen de beoogde verdelingen uit de toetsmatrijs en de gerealiseerde verdelingen van het examen. Beschrijf hoe grote verschillen tussen beide verdelingen verklaard kunnen worden en hoe dit een volgende keer voorkomen kan worden.
- Laat beide examinatoren hun eigen score invoeren in Wolf en gebruik deze informatie om de beoordelaarsovereenstemming te onderzoeken.
- Leg de verantwoording over de ontwikkeling, samenstelling, afname en beoordeling van het examen vollediger en in één document vast.

4 Onderzoek validiteit CE Economie 2023 vmbo kb flexibel digitaal examen

4.1 Beschrijving van het examen

Het examen Economie vmbo kb flexibel digitaal 2023 is een digitaal examen⁴³, waarmee maximaal 48 punten te behalen zijn. Het is een 'flexibel' examen. Dat betekent dat er geen tijdvakken zijn, maar dat elke school zelf bepaalt wanneer het examen wordt afgenomen, binnen een vooraf aangegeven periode. In 2023 kon dit van maandag 3 april t/m vrijdag 21 juli 2023. Dit examen wordt ook wel aangeduid met 'digitaal examen' of cbt (computer based test).

Het examen is afgenomen met behulp van Facet, het digitale systeem van DUO, waarin CvTE de examens aanbiedt. In het voortgezet onderwijs worden alle centrale digitale examens en toetsen afgenomen met behulp van Facet.⁴⁴

In het vakspecifieke gedeelte 'Economie kb Digitaal 2023' van de constructieopdracht⁴⁵ is te lezen dat het examen uit negen varianten bestaat, waarbij elke variant drie blokken omvat, een A-, een B- en een C-blok. Per blok bestaan verschillende versies (bijvoorbeeld Blok A1, blok A2, blok A3). Elke versie van een blok voldoet aan bepaalde eisen. Elk blok heeft bijvoorbeeld hetzelfde aantal scorepunten en dezelfde toetstechnische kenmerken (zoals: moeilijkheidsgraad en spreiding van de moeilijkheidsgraad over de opgaven). Een aantal blokken bevat (geheim gebleven) opgaven uit voorafgaande jaren, die worden gebruikt om de prestaties te vergelijken met de prestaties van kandidaten uit voorgaande jaren.

4.1.1 Domeindefinitie

In de flexibele en digitale examens Economie kb 2023 worden drie domeinen van het examenprogramma getoetst:

1. consumptie;
2. arbeid en productie;
3. internationale ontwikkelingen.

Binnen deze drie exameneenheden worden volgens de toetsmatrijs de volgende leervaardigheden getoetst:

- Economische verschijnselen, vraagstukken en ontwikkelingen beschrijven;
- Relevante economische vragen herkennen en zelfstandig vragen formuleren met het oog op het verwerven van kennis en inzicht;
- Verworven en/of aangeboden informatie over economische verschijnselen, ontwikkelingen en vraagstukken verwerken;
- Een beargumenteerd standpunt bepalen over een gegeven of zelf geformuleerde economische vraag;
- De rekenvaardigheden binnen economie toepassen:
 - tabellen, grafieken en diagrammen kunnen lezen enz.;
 - een kosten- en batenanalyse maken.

43 Bij sommige scholen werd in 2023 een papieren examen afgenomen voor economie kb, maar dat valt buiten het bereik van dit onderzoek.

44 DUO: geraadpleegd op 15-2-2024, van: <https://duo.nl/zakelijk/voortgezet-onderwijs/examens-en-diplomas/facet/over-facet.jsp>

45 CvTE (2021). Constructieopdrachten 2023 Mens en maatschappij vmbo. Utrecht: CvTE.

In de Syllabus Centraal Examen 2023 Economie vmbo wordt een nadere invulling gegeven van de verschillende onderdelen in de eenheden die in het CE worden getoetst; hiermee worden de globale eindtermen gespecificeerd. De specificaties voor het CE zijn onder verantwoordelijkheid van het CvTE tot stand gekomen en hebben een verplichtend karakter voor de inhoud en de onderwijsniveaus. In de Syllabus Centraal Examen 2023 Economie vmbo staat de verdeling van de examenstof over het SE en het CE. In het overzicht van de tien exameneenheden voor het vmbo kb is per eenheid aangegeven of deze in het CE moet worden getoetst (vier eenheden), of deze op het SE moet (zeven eenheden) of mag (drie eenheden) worden getoetst.

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 document 'Constructieopdrachten 2023 Mens en maatschappij vmbo' (par. 1.6.1) staat dat er bij de oplevering van de toets een toetsrationale aanwezig is, waarin is vermeld welke eindtermen van de syllabus met welke vragen getoetst worden, volgens welke procedure de zwaarte van de vragen is bepaald en wat daarvan de uitkomst was. Volgens de constructieopdracht staat in de toetsrationale in hoeverre de toetsmatrijzen (en overige aanwijzingen met percentages) uit het vakspecifieke deel van de constructieopdracht zijn gerealiseerd. Bovendien staat (volgens de constructieopdracht) in de toetsrationale een inschatting van de tijd die de modale kandidaat nodig heeft om de toets te maken. Ook staan in de toetsrationale de overige relevante constructiespecificaties die in het vakspecifieke deel van het betreffende examen opgenomen kunnen zijn. Deze toetsrationale wordt opgesteld door de toetsdeskundige van Cito, zodat de vaststellingscommissie kan vergelijken of het daadwerkelijk samengestelde examen voldoet aan de toetsmatrijs qua verdeling van de vragen over de leerstofonderdelen. In de toetsmatrijs van het examen Economie kb digitaal 2023 staat de beoogde bandbreedte van de percentuele verdeling van scorepunten over de drie exameneenheden en over de leervaardigheden. Een gerealiseerde verdeling van de scorepunten over de verschillende categorieën is te vinden in de toetsrationales. Er zijn verschillende toetsrationales, voor drie verschillende varianten van het examen. Hierin staat de percentuele verdeling van items over domeinen / exameneenheden, over (vak)vaardigheden, over gedrag (productief of receptief) en over vraagvormen.

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 1.2) 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, later gevolgd door de evaluatie (fase E). Na de vaststelling van het examen wordt bij de papieren examens het formulier Fase D ingevuld. Bij de digitale examens wordt de goedkeuring door de vaststellingscommissie gegeven in de vorm van een 'vinkje', door de voorzitter van de vaststellingscommissie gezet in de digitale omgeving 'Facet'. Daarmee is bekrachtigd dat door het CvTE is vastgesteld dat de domeinrepresentatie voldoende is.

4.1.3 Domeinrelevantie

Domeinrelevantie betreft de mate waarin elk item 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 vak- en toetsdeskundigen 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. Hieraan hebben 146 docenten van 137 scholen meegewerkt. Zij gaven in totaal 222 reacties. Qua moeilijkheidsgraad vond 34 procent het digitale examen precies goed; 50 procent vond het moeilijk en 13 procent vond het te moeilijk. Slechts enkelen (3%) vonden het te makkelijk. Een ruime meerderheid (78%) vond het examen precies goed van lengte, terwijl 22 procent het te lang vond. De inhoudelijke aansluiting van het digitale examen bij het gegeven onderwijs was volgens 54 procent voldoende en volgens 26 procent goed. Een op de vijf docenten vond de aansluiting onvoldoende (18%) of slecht (2%)⁴⁶. De docenten is ook gevraagd een cijfer te geven voor het examen. De gemiddelde score die zij geven is 5,9. Als we de gegeven cijfers nader bekijken, blijkt dat 33 procent van de docenten het examen een onvoldoende gaf: 5 procent gaf het examen het cijfer 2 of 3, 9 procent gaf het cijfer 4 en 19 procent gaf het cijfer 5. Relatief veel docenten gaven het cijfer 6 (29%) of 7 (30%); 7 procent gaf het examen het cijfer 8 of 9.

In de evaluatie van het examen Economie kb digitaal 2023 door de vaststellingscommissie van het CvTE (Evaluatieformulier fase E) wordt aangegeven dat er geen echte kritiekpunten zijn. Er zijn volgens de vaststellingscommissie geen opmerkingen via het beveiligde forum (BBF) ontvangen, er zijn geen klachten van LAKS en er zijn slechts drie meldingen gedaan bij de examenlijn. Ook concludeert de vaststellingscommissie dat de toets- en itemanalyse vergelijkbaar is met die van het jaar ervoor.

4.2 Samenvatting onderzoek validiteit CE Economie vmbo kb flex digitaal 2023

Aan de hand van het onderzoekskader, beschreven in hoofdstuk 2, is het centraal examen Economie vmbo kb flex digitaal 2023 geanalyseerd. In deze paragraaf beschrijven we de belangrijkste resultaten. Gedetailleerde uitkomsten per criterium staan beschreven in Bijlage 4: Checklist Centraal examen Economie kb flex 2023. 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 (criterium 1.1). De term 'toetsplan' wordt door het CvTE niet gebruikt, maar er zijn wel twee documenten die samen als zodanig kunnen worden aangemerkt: de constructieopdracht met het vakspecifieke gedeelte voor het digitale examen en de syllabus.

Aan de criteria voor een toetsplan wordt bijna volledig voldaan. De doelgroep van het examen is vastgesteld, het domein is gedefinieerd en het gebruiksdoel van het examen is gedefinieerd. In het examenprogramma, dat door de minister op hoofdlijnen wordt vastgesteld, zijn de exameneenheden voor het CE aangewezen. In de syllabus worden tien exameneenheden en de bijbehorende vaardigheden beschreven. Ook wordt aangegeven hoe de exameneenheden over het centraal examen en het schoolexamen verdeeld zijn. In de syllabus is aangegeven dat deze als hulpmiddel dient voor degenen die anderen of zichzelf op een centraal examen voorbereiden. Aanvullende informatie over het basisontwerp van het CE staat in de constructieopdracht.

Een van de criteria in deze fase is het specificeren van hoe de toets moet worden afgenomen en gescoord (criterium 1.6). Deze informatie is grotendeels beschikbaar, maar is verdeeld over verschillende documenten: de constructieopdracht, de vakspecifieke informatie bij de september-/maartmededeling en het correctievoorschrift. Er is een document met algemene regels bij de correctie van digitale CE's bb en

46 Dit zou zowel aan het examen als aan het onderwijs kunnen liggen.

kb. Daarin staat dat in de digitale examens de antwoorden op de gesloten vragen automatisch worden gecorrigeerd. De examiner beoordeelt alleen de antwoorden op de open vragen in de correctiemodule van Facet. Dit dient te gebeuren in overeenstemming met het beoordelingsmodel. Een voorbeeld correctievoorschrift met uitwerkingen van de antwoorden op de vragen was voor de onderzoekers beschikbaar in de oefenomgeving van Facet. Theoretische uitgangspunten zijn niet beschreven in de documenten die samen het toetsplan vormen (criterium 1.7). Dit kan worden verklaard doordat er geen (inter)nationale theoretische kaders lijken te bestaan voor dit vak.

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, wordt een globale verdeling over drie vraagsoorten (meerkeuzevragen, voorgestructureerde vragen en open vragen) aangegeven. In de constructieopdracht staat dat de tijdsduur voor de afname van de toets 90 minuten is. Voor de toegestane hulpmiddelen wordt verwezen naar de desbetreffende regeling op examenblad.nl.

In de constructieopdracht is niet aangegeven wat de psychometrische eigenschappen van de items en van de toets als geheel moeten zijn (criterium 2.6). 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 wordt er in verschillende documenten in algemene zin iets gezegd over de gewenste betrouwbaarheid. Cito verwijst hiervoor naar de publicatie *Toetsen op School*⁴⁷, waar vanuit het Handboek naar wordt verwezen. Hier worden richtlijnen voor en achtergronden van de betrouwbaarheid van centrale examens geschetst. Over de moeilijkheidsgraad van het examen staat in het algemene deel van de constructieopdracht (paragraaf 1.6.2) de algemene opmerking dat examens redelijke eisen aan de kandidaten dienen te stellen en van voldoende niveau dienen te zijn om het civiel effect van het diploma te waarborgen. Ook wordt er beschreven dat Cito een inschatting maakt van de moeilijkheid van een toets, waar zij verschillende technieken voor hanteert. De volgorde van de vragen wordt niet specifiek aangegeven (criterium 2.7). Dit hangt waarschijnlijk samen met het feit dat er meerdere varianten van dit examen bestaan die uit verschillende vragenblokken worden samengesteld.

In de toetsmatrijs wordt de relevantie van de inhoud van de toets voor het beoogde meetdoel aannemelijk gemaakt. Afgaand op het oordeel van de experts in de vaststellingscommissie is de toetsmatrijs een adequate representatie van het meetdoel. Het werkwoordgebruik in de eind- en toetstermen, zoals beschreven in de syllabus, is eenduidig en sluit goed aan bij bestaande taxonomieën, al wordt er niet naar een taxonomie verwezen.

De procedure van vakinhoudelijke en toetstechnische screening is beschreven in de constructieopdracht. De relevante experts zijn toetsdeskundigen en leden van de constructiegroep van het Cito en de leden van de vaststellingscommissie van het CvTE. De vaststellingscommissie van het CvTE beoordeelt het examen in verschillende stappen. Kenmerken van de leden van de vaststellingscommissie zijn globaal beschreven in het vademecum. Leden voor de vaststellingscommissie worden voorgedragen door de vakbond of vakvereniging. De voorzitter van de vaststellingscommissie beoordeelt of nieuwe leden passen binnen het functieprofiel. Het specifieke functieprofiel voor dit examen hebben we niet ter beschikking. Wel zijn er,

47 Kuhlmeier, H., Til, van, A., Sanders, P. (2016). *Toetsen op School*. Stichting Cito Instituut voor Toetsontwikkeling Arnhem. Cito, geraadpleegd op: 17-4-2024, van: [toetsen-op-school-voortgezet-onderwijs.pdf](https://www.cito.nl/toetsen-op-school-voortgezet-onderwijs.pdf) (cito.nl).

zoals hierboven beschreven voorbeelden beschikbaar gesteld van vacatureteksten voor leden van andere vaststellingscommissies waarin (generieke) eisen staan beschreven qua kwalificaties, relevante ervaringen en onafhankelijkheid. Demografische kenmerken staan echter niet beschreven (criterium 2.13).

Wat betreft deze fase bevelen we aan om vooraf richtlijnen voor de gewenste psychometrische eigenschappen van de items en van het examen als geheel vast te leggen.

4.2.3 Fase 3: Constructie

Fase 3 gaat over het construeren en normeren van examenopgaven. In aanvulling op de informatie in de toetsmatrijs (zie paragraaf 4.2.1) wordt in de constructieopdracht ingegaan op de examendocumenten en op criteria voor het uitgangsmateriaal. Bovendien worden algemene aanvullende eisen voor de digitaal afgenomen examens gegeven, waaronder richtlijnen voor de inhoud en de vraagstelling, voor het gebruik van het vraagscherm, voor standaardtaalgebruik en voor het gebruik van multimedia.

Er heeft een kwalitatieve evaluatie plaatsgevonden door de vaststellingscommissie. In bijlage 6 van de constructieopdracht staat dat er door de toetsdeskundige van Cito en de voorzitter van de vaststellingscommissie een logboek wordt bijgehouden tijdens de verschillende fasen van het ontwikkelingsproces. Cito licht verder toe dat: *'een examen niet verder in het constructieproces komt als het logboek niet is ingevuld. Procesvoering (afdeling Cito) monitort dat scherp. Ook beschikken we over de eerdere versies van het examen, dus de 1ste keer bespreking, 2de keer bespreking etc.'* De formulieren die bij de papieren toetsversies na de verschillende fasen worden ingevuld (fase A t/m D), zijn niet beschikbaar voor de digitale versie. Ook hadden wij geen toegang tot het logboek. Het CvTE geeft aan dat de goedkeuring van het examen wordt vastgelegd met een 'vinkje' in Facet. Uit die goedkeuring leiden wij af dat de vragen correct zijn geformuleerd, dat de vragen of opdrachten zijn gestandaardiseerd en zodanig ontworpen dat fouten bij het invullen worden voorkomen (criteria 3.1, 3.1a, 3.1b), dat de kwaliteit van de lay-out in orde is (criterium 3.1e), dat de vragen voldoende gedetailleerd zijn en dat de moeilijkheidsgraad is afgestemd op de doelgroep (criteria 3.2 en 3.3).

Voor leerlingen staan op www.mijneindexamen.nl alle toegestane hulpmiddelen bij het examen beschreven⁴⁸. Bij het digitale examen zelf (gebaseerd op de oefenomgeving) in Facet is enkel een lijst van digitale hulpmiddelen te vinden die voor de kandidaat beschikbaar zijn in Facet (een markeerfunctie, vergrootglas, rekenmachine, en letterplankje met 'speciale letters'). De overige toegestane middelen worden hier niet genoemd. Hiermee lijkt de beschikbare informatie voor de kandidaat in de fase *voorbereidend op het examen* wel volledig en duidelijk, maar *tijdens het examen* niet compleet (criterium 3.1c en 3.1d).

In de Amerikaanse standaarden wordt geadviseerd dat een examen wordt samengesteld uit gepreteste opgaven (criterium 3.3a). Bij het digitale examen Economie kb wordt geen pretest uitgevoerd. In de constructieopdracht staat dat er negen combinaties met drie vragenblokken zijn. Een deel van het examen bestaat uit ankeropgaven. Deze zijn eerder afgenomen, waardoor de moeilijkheidsgraad bekend is en de resultaten op die opgaven kunnen worden vergeleken met de resultaten van een eerdere jaargroep. Bij het bepalen van de normering, wordt dit gebruikt als aanvullend gegeven. 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 3.8).

⁴⁸ Te weten: schrijfmateriaal, inclusief millimeterpapier; tekenpotlood (en puntenslijper); vlakgum; blauw en rood kleurpotlood; geodriehoek; rekenmachine; liniaal met millimeterverdeling; passer (met potloodstiftje); eendelig verklarend woordenboek Nederlands of eendelig woordenboek Nederlands-thuis taal en thuis taal-Nederlands).

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 3.5). In de constructieopdracht is het proces van ontwikkeling en vaststelling van een examen beschreven. Daarnaast zijn er verschillende checklists en handreikingen die door het Cito zijn ontwikkeld en er wordt verwezen naar handboeken. Ook zijn er documenten met informatie over de normeringssystematiek en over de normering.

Criterium 3.7 gaat over de kwalificaties van beoordelaars van de reeds ontwikkelde items en over de instructies en training die deze beoordelaars ten behoeve van het beoordelingsproces krijgen. Hier wordt deels aan voldaan. Zoals bij het examen Kunst (algemeen) havo 2023 is beschreven, staat in bijlage 2 van het algemene gedeelte van de constructieopdracht door welke typen deskundigen de screenings worden gedaan. Cito is verantwoordelijk voor het bewaken van de expertise van de constructeurs/screeners en toetsdeskundigen. Voor het construeren en screenen van items kunnen deskundigen gebruik maken van de volgende hoofdstukken uit het Handboek 'H7. checklist open vragen', 'H8. checklist meerkeuzevragen' en 'H5. handreiking taalgebruik in Centrale Examens'.

Zoals aangegeven, konden wij niet over het logboek (of het D-formulier) beschikken waarin alle relevante fasen in het ontwikkelproces zouden moeten worden bijgehouden. Met betrekking tot constructie en normering bevelen wij aan om relevante documenten te bundelen en ervoor te zorgen dat de informatie volledig is. Hoewel we de redenen begrijpen om geen uitvoerige lijst hulpmiddelen op te nemen in de instructies voor de kandidaat, raden we wel aan de *belangrijkste* toegestane hulpmiddelen op een beknopte wijze te vermelden. Dit om eventuele discussies tijdens het examen te voorkomen.

4.2.4 Fase 5: Beoordeling, verwerking, analyse en normering

Op het gebied van de beoordeling, verwerking en analyse is aan de meeste criteria voldaan. Bij de digitale examens in Facet worden de gesloten vragen automatisch gecorrigeerd. Het beoordelingsmodel per item is opgenomen in Facet (criterium 5.1). Wij hadden inzage in de oefenexamens en bijbehorende beoordelingsmodellen, daar worden ook voorbeelden van beoordeling van gedeeltelijk goed beantwoorde vragen gegeven (criterium 5.6). De manier waarop de norm wordt bepaald, is gedocumenteerd, op examenblad.nl en in een notitie over de normeringssystematiek. Er is een omzettingstabel waarmee de uiteindelijke scores worden omgezet in examencijfers (criterium 5.5).

Door Cito is 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. Uit de toets- en itemanalyse blijkt dat de gemiddelde p-waarde (d.w.z. de proportie goed beantwoorde vragen) voor de negen varianten varieert van 0,45 tot 0,57. De betrouwbaarheid, uitgedrukt in de coëfficiënt Cronbachs alpha, varieert van 0,64 tot 0,71 voor de negen verschillende examenvarianten.

Er zijn twee onafhankelijke beoordelaars die de open vragen in het examen apart van elkaar scoren (criterium 5.3): de eerste corrector (een docent van de examenkandidaat) en een tweede corrector. Het proces van selecteren, trainen, kwalificeren en monitoren van degenen die scoren, is grotendeels gedocumenteerd (criterium 5.2). De school wijst de tweede corrector aan. Dit kan een docent van de eigen locatie zijn of een docent van een andere locatie of een andere school. Hoe die selectie precies plaatsvindt, is niet gedocumenteerd. Er is ook niet bekend hoe het scoren wordt gemonitord. Na het uitvoeren van de correctie kunnen de beide correctoren in de overeenstemmingsmodule van Facet elkaars scores zien en overeenstemming proberen te bereiken. Het invoeren van cijfers in de applicatie 'Wolf' (van het Cito) is niet nodig. De twee onafhankelijke scores worden opgeslagen in de correctiemodule in Facet en

worden beide gerapporteerd in de overeenstemmingsmodule (criterium 5.4). De mate van beoordelaarsovereenstemming tussen de eerste en tweede corrector bij de correctie van de antwoorden op de open vragen wordt niet met behulp van psychometrisch onderzoek getoetst. Daardoor wordt niet voldaan aan criterium 5.10. 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. Het (kunnen) onderzoeken van de interbeoordelaarscorrelatie komt ten goede aan de betrouwbaarheid van de uitslag voor kandidaten. In de 'Brochure Kaders flexibele en digitale examens bb en kb flex' is wel aangegeven dat er in het najaar een controle is of de correctie (door de eerste corrector) conform het correctievoorschrift heeft plaatsgevonden. Die controle wordt uitgevoerd door gecommiteerden van de Commissie Staatsexamens VO. De resultaten worden gerapporteerd aan de Inspectie van het Onderwijs. De scholen ontvangen in het najaar een rapportage van het Cito waarin hun resultaten worden vergeleken met het landelijke beeld en waarin tevens informatie wordt gegeven over verschillen tussen de automatisch door de computer gecorrigeerde (gesloten) opgaven en de handmatig gecorrigeerde (open) opgaven.

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/of de constructieopdracht (laten) aanpassen (criterium 7.1). Daartoe kan worden besloten op grond van de grondige evaluatie, waarover wordt gerapporteerd in formulier E.

Bewijzen voor de (inhouds)validiteit (criterium 7.2) zijn in verschillende documenten vastgelegd. In de toetsrationale is de gerealiseerde samenstelling van de toets in kaart gebracht, wat de verdeling over de in het examenprogramma en de syllabus beschreven exameneenheden betreft. Daarnaast is het examen door verschillende deskundigen gescreend op verschillende aspecten en is het examen door experts vastgesteld. Bewijzen voor de betrouwbaarheid van het examen zijn deels vastgelegd (criterium 7.3). Zoals gezegd is de betrouwbaarheid van het examen uit de TIA goed gebleken, maar is er geen beoordelaarsovereenstemming onderzocht bij de correctie van de antwoorden op de open vragen.

Procedures rondom de toetsontwikkeling zijn vastgelegd in de constructieopdracht ('Taakverdeling en werkwijze CvTE-Cito') en in het handboek productie examens vmbo van Cito (criterium 7.4). Ook zijn daarin de rollen, verantwoordelijkheden en taakverdeling vastgelegd (criterium 7.5). Er wordt onderzoek gedaan onder stakeholders (via de QuickScan bij docenten, via de Examenlijn en via de BBF ('besloten beveiligde fora'), waar docenten opmerkingen over het examen kunnen maken. Ook LAKS kan opmerkingen maken over het examen. Daarmee is ook aan criterium 7.6 voldaan.

Criterium 7.7 vereist dat er een verslag is (jaarrapport, evaluatieverslag) waarin verantwoording is afgelegd over de ontwikkeling, samenstelling, afname, beoordeling, scores en uitgevoerde analyses van de toets. Aan dit criterium is deels voldaan. Er is niet één compleet verslag beschikbaar met de verantwoording rondom het examen Economie kb. Er is wel een algemeen verslag van de centrale examens van 2023, maar dat geeft geen specifieke informatie. De formulieren voor de verslaglegging van het toetsproces uit de constructiefase van het examen (fase B, C en D) ontbreken. Zoals in paragraaf 4.2.3 is aangegeven, is in de constructieopdracht aangegeven dat de toetsdeskundige van Cito en de voorzitter van de vaststellingscommissie van CvTE een logboek bijhouden tijdens de verschillende fasen van het

ontwikkelingsproces. Wij hadden geen toegang tot een dergelijk logboek en kunnen niet met zekerheid zeggen of dit alle fasen in het toetsproces beschrijft. Wel is er een evaluatieformulier E bij het digitale examen. Daarin wordt de reactie van stakeholders samengevat en door de vaststellingscommissie wordt aangegeven of er opvallende punten zijn gebleken uit de toets- en itemanalyse. Ook geeft de commissie in dat formulier haar mening over de hanteerbaarheid van het correctievoorschrift, een inschatting van de moeilijkheidsgraad, haar mening over eventuele nieuwe aspecten in het examen en aandachtspunten en aanbevelingen voor het volgende jaar.

Onze aanbeveling bij deze fase om de verantwoording over de ontwikkeling, samenstelling, afname en beoordeling van het examen Economie kb digitaal vollediger en in één document vast te leggen en beschikbaar te hebben

4.3 Samenvattend

Samenvattend kan worden geconcludeerd dat het eindexamen Economie vmbo kb digitaal 2023 qua procedures rond ontwikkeling, afname en beoordeling in het algemeen voldoet aan de criteria die in het beoordelingskader zijn opgenomen. Echter, de verantwoording is over tal van documenten verspreid en ontbreekt deels of is onvolledig. Zo ontbreken 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 van de items en van de toets als geheel moeten 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.
- Vermeld de *belangrijkste* toegestane hulpmiddelen op een beknopte wijze in de instructie voor de kandidaat, om discussie te voorkomen tijdens het examen.
- Onderzoek de beoordelaarsovereenstemming tussen de eerste en tweede corrector.

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 Kunst (algemeen) havo en Economie vmbo kb flexibel digitaal examen 2023. 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 Kunst (algemeen) havo en Economie vmbo kb flex digitaal hanteren aan (inter)nationale kwaliteitsstandaarden?’

In hoofdstuk 3 en hoofdstuk 4 hebben we beschreven in hoeverre het toetsproces van de examens Kunst (algemeen) havo en Economie vmbo kb voldoet aan de criteria van het onderzoekskader zoals beschreven in hoofdstuk 2. Op basis hiervan geven we hieronder per examen antwoord op de onderzoeksvraag.

Het examen Kunst (algemeen) havo 2023 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. Aan een aantal criteria wordt echter niet of slechts deels voldaan. Zo worden vooraf niet de gewenste psychometrische eigenschappen per item en in totaal gespecificeerd in de constructieopdracht. Verder ontbreekt het fase D formulier en is daarmee het examen niet vergeleken met een referentie-examen. Een ander belangrijk criterium waar niet aan voldaan wordt, is een verantwoording over de te grote verschillen tussen de toetsmatrijs en de toetsrationale. Tot slot ontbreekt er onderzoek naar de beoordelaarsovereenstemming. Dit is bij een examen met bijna alleen maar open vragen wel wenselijk.

Op basis van deze gevonden hiaten in het toetsproces van het examen Kunst (algemeen) havo 2023 hebben we hierboven enkele aanbevelingen geformuleerd. Samengevat luiden deze aanbevelingen als volgt:

- Leg vooraf de gewenste psychometrische eigenschappen van de items en van de toets als geheel vast.
- Vermeld de *belangrijkste* toegestane hulpmiddelen op een beknopte wijze in de instructie voor de kandidaat, om discussie te voorkomen tijdens het examen.
- Vul -naast de procedures voor het vaststellen van examen in Facet- ook het fase D formulier in, inclusief de vergelijking met een referentie-examen.
- Maak in de toetsrationale de vergelijking tussen de beoogde verdelingen uit de toetsmatrijs en de gerealiseerde verdelingen van het examen. Beschrijf hoe grote verschillen tussen beide verdelingen verklaard kunnen worden en hoe dit een volgende keer voorkomen kan worden.
- Laat beide examinatoren hun eigen score invoeren in Wolf en gebruik deze informatie om de beoordelaarsovereenstemming te onderzoeken.
- Leg de verantwoording over de ontwikkeling, samenstelling, afname en beoordeling van het examen vollediger en in één document vast.

Voor het examen Economie vmbo kb flexibel digitaal examen 2023 geldt eveneens dat de procedures die het CvTE en Cito hebben gevolgd over het algemeen voldoen aan de (inter)nationale kwaliteitsstandaarden. Echter, de verantwoording is over tal van documenten verspreid en ontbreekt deels of is onvolledig. Zo ontbreken documenten die betrekking hebben op het proces van vaststelling van de examenopgaven. Ook wordt aan sommige criteria niet voldaan. Zo wordt ook bij dit examen niet vooraf aangegeven wat de gewenste psychometrische eigenschappen van de items en van de toets als geheel moeten 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.
- Vermeld de belangrijkste toegestane hulpmiddelen op een beknopte wijze in de instructie voor de kandidaat, om discussie te voorkomen tijdens het examen.
- 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.

Uit bovenstaande adviezen per examen kunnen we de volgende algemene aanbevelingen doen:

- 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;
- 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.

Onze aanbevelingen voor de procedures met betrekking tot de centrale examens overlappen sterk met de eerdere onderzoeken naar de validiteit van centrale examens uit 2022 en 2023. Zoals eerder beschreven bestrijkt het hele toetsconstructieproces ongeveer twee jaar. Hierdoor zijn niet alle effecten van eventuele nieuwe maatregelen op korte termijn te meten.

Literatuur

- AEA-Europe. (2017) European Framework of Standards for Educational Assessment 1.0: https://aea-europe.net/wp-content/uploads/2022/11/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
- 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. (Red.) (2017). *Toetsen op school*. Arnhem: Cito. 1e druk 2011, herziene versie 2017. 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.
- 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
- 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
- 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

Bijlage 1: Overzicht van figuren en tabellen

Figuren

Figuur 1.1: Tijdlijn totstandkoming centrale examens 2023. Bron: CvTE.	4
Figuur 1.2: In schema: samenwerking tussen het CvTE en Cito voor constructie en vaststelling van centrale examens vo. Bron: https://www.examenblad.nl/onderwerp/examencyclus/2023 .	5
Figuur 2.1: Toetscyclus	8

Tabellen

Tabel 2.1: Checklist Fase 1: Basisontwerp.	9
Tabel 2.2: Checklist Fase 2: Construeren toetsmatrijs	10
Tabel 2.3: Checklist fase 3: Construeren van de toets	12
Tabel 2.4: Checklist Fase 5: Beoordelen, verwerken, analyseren en normeren	14
Tabel 2.5: Checklist Fase 7: Evalueren en verbeteren	15
Tabel 2.6: Relatie tussen de fasen van de toetscyclus en examencyclus van Cito en het CvTE.	15

Bijlage 2: Checklist Centraal examen Kunst (algemeen) havo 2023 eerste tijdvak

	Fase 1: Basisontwerp Kunst (algemeen) havo 2023-1	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 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 de toets moet worden afgenomen en gescoord (bv. papieren of digitale toets, wijze van afname, wijze van beoordeling, hoe komt score tot stand?).	ja
1.7	In het toetsplan staat de theorie met betrekking tot de beoogde interpretatie beschreven of er wordt verwezen naar de kaders waarop het examen is gebaseerd.	nee

	Fase 2: Opstellen toetsspecificaties en toetsmatrijs Kunst (algemeen) havo 2023-1	Aanwezig?
2.1	Er zijn toetsspecificaties (bv. in de vorm van een constructieopdracht)	ja
2.2	In de toetsspecificaties staan het aantal vragen met bijbehorende scorepunten	ja
2.3	In de toetsspecificaties staat de toetsvorm en/of het soort vragen (bijvoorbeeld gesloten en/of open)	ja
2.4	In de toetsspecificaties staan de toegestane hulpmiddelen.	ja
2.5	In de constructieopdracht staat de hoeveelheid tijd die is toegestaan voor de toets.	ja
2.6	In de toetsspecificaties staan de gewenste psychometrische eigenschappen van de items en de toets als geheel.	nee
2.7	In de toetsspecificaties wordt de volgorde van items gedefinieerd.	ja
2.8	Er is een toetsmatrijs.	ja
2.9	In de toetsmatrijs wordt de relevantie van de inhoud van de toets of het examen voor het beoogde meetdoel aannemelijk gemaakt.	ja
2.10	De toetsmatrijs is een adequate representatie van het meetdoel.	ja
2.10a	De eind- en toetstermen representeren het meetdoel.	ja
2.10b	De eind- en toetstermen sluiten aan op de inhoud en het vereiste beheersingsniveau van het betreffende meetdoel	ja
2.10c	Het werkwoordgebruik in de eind- en toetstermen is eenduidig en sluit goed aan bij bestaande taxonomieën (zoals bijvoorbeeld van Bloom of Romiszowsky)	ja
2.10d	Het aantal vragen geeft een voldoende dekking van het meetdoel.	ja
2.11	Relevante vakinhoudelijke experts beoordelen de toetsspecificaties op hun geschiktheid.	ja
2.12	Het doel van deze beoordeling van toetsspecificaties, het proces waarmee deze beoordeling wordt uitgevoerd en de resultaten van deze beoordeling worden gedocumenteerd.	ja
2.13	De kwalificaties, relevante ervaringen en demografische kenmerken van relevante vakinhoudelijke experts worden gedocumenteerd.	ja

	Fase 3 Constructie en normering Kunst (algemeen) havo 2023-1	Aanwezig?
3.1	De vragen zijn correct geformuleerd	ja
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 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
3.3	De moeilijkheidsgraad van de vraag is afgestemd op de beoogde doelgroep.	deels
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.	deels
3.4	De standaardbepaling is correct uitgevoerd.	n.v.t.
3.4a	De standaardbepaling is op de juiste wijze uitgevoerd.	n.v.t.
3.4b	De vakdeskundigen/experts die bij de standaardbepaling betrokken zijn, zijn naar behoren geselecteerd en getraind.	n.v.t.
3.4c	Er is voldoende overeenstemming tussen de vakdeskundige experts.	n.v.t.
3.5	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
3.6	Er is gedocumenteerd in hoeverre de inhoud van de toets het domein vertegenwoordigt, dat is gedefinieerd in de toetsspecificaties.	ja
3.7	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. ⁴⁹	ja
3.8	Er is een rationale voor de toets waarin verantwoording wordt afgelegd over de samenstelling van de toets of het examen.	deels
	Fase 4: Afname (n.v.t.)	

⁴⁹ Bij het gebruik van empirische analyses gelden andere aanvullende criteria. Zie in de Amerikaanse standaarden de criteria 4.09 t/m 4.11.

	Fase 5: Beoordeling, verwerking en analyse en normering Kunst (algemeen) havo 2023-1	Aanwezig?
5.1	Procedures voor het scoren en scorecriteria zijn gedocumenteerd	ja
5.2	Het proces van het selecteren, trainen, kwalificeren en monitoren van de beoordelaar(s) die scoort (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 (n.b. huidige procedure is wettelijke procedure)	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 toetsscores zichtbaar met behulp van Cronbach's 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
5.11	Er wordt een norm(cesuur) verstrekt die via een wetenschappelijke methodiek is bepaald.	ja
5.12	De manier waarop de norm wordt bepaald is gedocumenteerd.	ja
	Fase 6: Registreren en communiceren: n.v.t.	

	Fase 7: Evaluatie Kunst (algemeen) havo 2023-1	Aanwezig?
7.1	Toetsspecificaties worden aangepast als er significante veranderingen zijn in het vertegenwoordigde domein.	ja
7.2	Bewijzen voor (inhouds)validiteit 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 in een verslag (jaarrapport, evaluatieverslag) waarin verantwoording is afgelegd over de ontwikkeling, samenstelling en afname van de toets.	deels

Bijlage 3: Checklist Centraal examen Economie vmbo kb flex 2023

	Fase 1: Basisontwerp Economie vmbo kb-flex 2023	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 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 de toets moet worden afgenomen en gescoord (bv. papieren of digitale toets, wijze van afname, wijze van beoordeling, hoe komt score tot stand?).	ja
1.7	In het toetsplan staat de theorie met betrekking tot de beoogde interpretatie beschreven of er wordt verwezen naar de kaders waarop het examen is gebaseerd.	nee

	Fase 2: Opstellen toetsspecificaties en toetsmatrijs Economie vmbo kb-flex 2023	Aanwezig?
2.1	Er zijn toetsspecificaties (bv. in de vorm van een constructieopdracht)	ja
2.2	In de toetsspecificaties staan het aantal vragen met bijbehorende scorepunten	ja
2.3	In de toetsspecificaties staat de toetsvorm en/of het soort vragen (bijvoorbeeld gesloten en/of open)	ja
2.4	In de toetsspecificaties staan de toegestane hulpmiddelen.	ja
2.5	In de constructieopdracht staat de hoeveelheid tijd die is toegestaan voor de toets.	ja
2.6	In de toetsspecificaties staan de gewenste psychometrische eigenschappen van de items en de toets als geheel.	nee
2.7	In de toetsspecificaties wordt de volgorde van items gedefinieerd.	nee
2.8	Er is een toetsmatrijs.	ja
2.9	In de toetsmatrijs wordt de relevantie van de inhoud van de toets of het examen voor het beoogde meetdoel aannemelijk gemaakt.	ja
2.10	De toetsmatrijs is een adequate representatie van het meetdoel.	ja
2.10a	De eind- en toetstermen representeren het meetdoel.	ja
2.10b	De eind- en toetstermen sluiten aan op de inhoud en het vereiste beheersingsniveau van het betreffende meetdoel	ja
2.10c	Het werkwoordgebruik in de eind- en toetstermen is eenduidig en sluit goed aan bij bestaande taxonomieën (zoals bijvoorbeeld van Bloom of Romiszowsky)	ja
2.10d	Het aantal vragen geeft een voldoende dekking van het meetdoel.	ja
2.11	Relevante vakinhoudelijke experts beoordelen de toetsspecificaties op hun geschiktheid.	ja
2.12	Het doel van deze beoordeling van toetsspecificaties, het proces waarmee deze beoordeling wordt uitgevoerd en de resultaten van deze beoordeling worden gedocumenteerd.	ja
2.13	De kwalificaties, relevante ervaringen en demografische kenmerken van relevante vakinhoudelijke experts worden gedocumenteerd.	ja

	Fase 3 Constructie en normering Economie vmbo kb-flex 2023	Aanwezig ?
3.1	De vragen zijn correct geformuleerd	ja
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 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
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).	ja
3.3b	Er heeft een kwalitatieve evaluatie met deskundigen plaatsgevonden om informatie te verkrijgen over de moeilijkheidsgraad.	ja
3.4	De standaardbepaling is correct uitgevoerd.	n.v.t.
3.4a	De standaardbepaling is op de juiste wijze uitgevoerd.	n.v.t.
3.4b	De vakdeskundigen/experts die bij de standaardbepaling betrokken zijn, zijn naar behoren geselecteerd en getraind.	n.v.t.
3.4c	Er is voldoende overeenstemming tussen de vakdeskundige experts.	n.v.t.
3.5	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
3.6	Er is gedocumenteerd in hoeverre de inhoud van de toets het domein vertegenwoordigt, dat is gedefinieerd in de toetsspecificaties.	ja
3.7	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. ⁵⁰	deels
3.8	Er is een rationale voor de toets waarin verantwoording wordt afgelegd over de samenstelling van de toets of het examen.	ja
	Fase 4: Afname (n.v.t.)	

⁵⁰ Bij het gebruik van empirische analyses gelden andere aanvullende criteria. Zie in de Amerikaanse standaarden de criteria 4.09 t/m 4.11.

	Fase 5: Beoordeling, verwerking en analyse en normering Economie vmbo kb-flex 2023	Aanwezig?
5.1	Procedures voor het scoren en scorecriteria zijn gedocumenteerd	ja
5.2	Het proces van het selecteren, trainen, kwalificeren en monitoren van de beoordelaar(s) die scoort (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 (n.b. huidige procedure is wettelijke procedure)	ja
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 toetsscores zichtbaar met behulp van Cronbach's 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
5.11	Er wordt een norm(cesuur) verstrekt die via een wetenschappelijke methodiek is bepaald.	ja
5.12	De manier waarop de norm wordt bepaald is gedocumenteerd.	ja
	Fase 6: Registreren en communiceren: n.v.t.	

	Fase 7: Evaluatie Economie vmbo kb-flex 2023	Aanwezig?
7.1	Toetsspecificaties worden aangepast als er significante veranderingen zijn in het vertegenwoordigde domein.	ja
7.2	Bewijzen voor (inhouds)validiteit 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 in een verslag (jaarrapport, evaluatieverslag) waarin verantwoording is afgelegd over de ontwikkeling, samenstelling en afname van de toets.	deels

