

College voor Toetsen en Examens

**Onderzoek naar de
inhoudsvaliditeit,
betrouwbaarheid en
normering van het
centraal examen Engels B1
mbo 2022-2023**

College voor Toetsen en Examens

Onderzoek naar de inhoudsvaliditeit, betrouwbaarheid en normering van het centraal examen Engels B1 mbo 2022-2023

Apeldoorn, december 2023

De rapporten van RCEC B.V. worden alleen openbaar na toestemming van de opdrachtgever.

Rapporten met status openbaar zijn beschikbaar via de website van RCEC B.V.: <https://www.rcec.nl>

RCEC

RCEC, Research Center voor Examinering en Certificering, is dé toonaangevende kennis- en netwerkorganisatie voor het borgen en bevorderen van de kwaliteit van examinering.

Als certificerende instelling evalueren wij de kwaliteit van examinering en waarderen deze met het toekennen van het RCEC certificaat. Wij zijn betrokken bij de grote (inter)nationale examenstelsels, zowel binnen het reguliere onderwijs ressorterend onder de ministeries van onderwijs, als voor het niet-reguliere onderwijs, zoals de door de ministeries ingestelde zelfstandige bestuursorganen (ZBO's). Als expertisecentrum vervullen wij een onafhankelijke en onpartijdige loketfunctie. Wij ontwikkelen geen eigen examens, maar zijn er exclusief om instellingen in het regulier onderwijs en de beroepspraktijk te ondersteunen met onderzoek en innovatie. Wij richten ons op de volledige toetscyclus, van constructie en afname tot beoordeling en analyse. Vanuit onze state-of-the-art academie verzorgen wij training en opleiding op maat. Verder dragen wij bij aan kennisdeling via publicaties en presentaties.

Zo werken wij gepassioneerd en betrokken aan het vervullen van onze belangrijkste missie: het recht doen aan beslissingen door, over en voor talent.

RCEC is onder accreditatie van de Raad voor Accreditatie gecertificeerd voor de CIO Maatstaf 2016 en de ISO 9001:2015 norm voor het kwaliteitsmanagementsysteem. RCEC is tevens een CRKBO geregistreerde instelling. Daarnaast is RCEC Affiliate Institutional Member van de International Association for Educational Assessment.



RCEC B.V. • Villa de Regentes • Regentesselaan 5 • 7316 AA Apeldoorn

Postbus 71 • 8170 AB Vaassen

T 055 - 303 31 24 • I www.rcec.nl • E info@rcec.nl

Inhoudsopgave

1. Aanleiding	blz. 1
1.1. Aanleiding en probleemstelling	1
1.2. Centrale onderzoeksvraag	2
1.3. Aanvraag en proces	2
2. Onderzoekskader	3
2.1. Onderzoekskader centrale examens mbo	3
2.2. Inhoudsvaliditeit	3
2.3. Betrouwbaarheid	5
2.4. Normering	6
3. Onderzoeksmethode	7
3.1. Inhoudsvaliditeit van het centraal examen Engels B1 mbo 2022-2023	7
3.2. Betrouwbaarheid van het centraal examen Engels B1 mbo 2022-2023	10
3.3. Normering van het centraal examen Engels B1 mbo 2022-2023	11
4. Onderzoeksresultaten	13
4.1. Inhoudsvaliditeit van het centraal examen Engels B1 mbo 2022-2023	13
4.2. Domein definitie	13
4.2.1. Inhoud van het examen	13
4.2.2. Toetsspecificaties	15
4.3. Domein representatie	16
4.4. Domein relevantie	17
4.5. Adequate toetsconstructieprocedure	18
4.6. Betrouwbaarheid van het centraal examen Engels B1 mbo 2022-2023	20
4.7. Standaardbepaling en normhandhaving van het centraal examen Engels B1 mbo 2022-2023	26
5. Conclusies	29
5.1. Conclusies inhoudsvaliditeit	29
5.2. Conclusies betrouwbaarheid	29
5.3. Conclusies standaardbepaling en normhandhaving	30
6. Geraadpleegde literatuur	31

1. Inleiding

1.1. Aanleiding en probleemstelling

Het College voor Toetsen en Examens (CvTE) is een zelfstandig bestuursorgaan. Sinds 2009 is het de regisseur van centrale toetsing en examinering in ons land voor NT2, po, vo, staatsexamens vo en mbo. Krachtens de 'Wet College voor Toetsen en Examens'¹ heeft het CvTE een regie-voerende verantwoordelijkheid. Het CvTE heeft als taak om namens de overheid de kwaliteit en het niveau van centrale toetsen en examens te waarborgen en te bevorderen dat scholen en instellingen in staat gesteld worden de afname van centrale toetsen en examens vlekkeloos te laten verlopen. Het CvTE adviseert en/of besluit over beleid ten aanzien van de uitvoering van centrale toetsen en examens, regelingen en alle andere zaken die voortvloeien uit de wettelijke taken van het CvTE.

In de jaren 2017, 2018 en 2020 heeft RCEC voor het CvTE een viertal onderzoeken uitgevoerd naar verschillende kwaliteitsaspecten van de centrale examens in het vo en in het mbo. Het eerste onderzoek, uitgevoerd in het jaar 2017, was gericht op de examens Biologie vmbo GL/TL 2016 en Duits havo 2016. In deze studie is onderzoek gedaan naar de validiteit van de procedures die gehanteerd worden door het CvTE en Cito en naar de wijze van omgaan met onvolkomenheden in de twee genoemde centrale examens. In de tweede studie, uitgevoerd in het jaar 2018, heeft RCEC onderzoek gedaan naar de validiteit van de procedures die gehanteerd worden door het CvTE en Cito aan de hand van de examens Engels vwo 2017 en Wiskunde GL/TL 2017. Het derde onderzoek, uitgevoerd in het jaar 2020, betrof een studie naar de inhoudsvaliditeit van een tweetal beroepsgerichte centrale examens voortgezet onderwijs, zijnde Producteren, Installeren en Energie (PIE) vmbo KB 2019 en Dienstverlening en Producten (D&P) vmbo GL 2019. In de vierde studie, eveneens uitgevoerd in het jaar 2020, heeft RCEC onderzoek gedaan naar de inhoudsvaliditeit, betrouwbaarheid en normering van het centraal examen mbo Nederlands 3F 2018-2019.

Het CvTE heeft, vanuit zijn verantwoordelijkheid en zorgplicht, in navolging op de eerder uitgevoerde onderzoeken aan RCEC gevraagd om onderzoek te doen naar het centraal examen Engels B1 mbo 2022-2023. In deze studie worden de inhoudsvaliditeit, de betrouwbaarheid en de normering van het betreffende examen, zijnde variant P2-05, onderzocht. Aanvullend wordt de betrouwbaarheid van drie extra varianten: P2-08, P4-19 en P4-22 van het betreffende examen onderzocht.

¹ <https://wetten.overheid.nl/BWBR0025364/2023-01-01>

1.2. Centrale onderzoeksvraag

Het door RCEC uit te voeren onderzoek heeft als doel antwoord te geven op de volgende centrale onderzoeksvraag:

“Voldoen de procedures die het CvTE en Cito bij de borging van de inhoudsvaliditeit, de betrouwbaarheid en de normering van de centrale examens mbo hanteren aan (inter)nationale kwaliteitsstandaarden en resulteert dat in een inhoudsvalide en betrouwbaar examenproduct? Voldoet daarnaast het examenmateriaal aan het gewenste Europees Referentiekader voor moderne vreemde talen (ERK)?”

Ten behoeve van het onderzoek, is door RCEC een onderzoekskader ontwikkeld bestaande uit: (1) de vier centrale aspecten van de inhoudsvaliditeit voor centrale examens: domein definitie, domein representatie, domein relevantie en adequate toetsconstructieprocedure, (2) de betrouwbaarheid, en (3) de normering (standaardbepaling en normhandhaving).

1.3. Aanvraag en proces

Op 17 augustus 2023 heeft er een eerste gesprek plaatsgevonden tussen mevrouw Linda van Abeelen, clustermanager mbo en de heer Arnold Brouwer, directeur RCEC. Tijdens dit gesprek is de onderzoeksvraag besproken en toegelicht. Volgend op dit gesprek heeft RCEC op 11 september 2023 een onderzoeksvoorstel aangeboden aan het CvTE. Op 17 oktober 2023 heeft het CvTE ingestemd met dit voorstel. Voor de uitvoering van het onderzoeksvoorstel heeft RCEC een onderzoeksteam samengesteld. De leden van dit onderzoeksteam zijn de heer dr. Arnold Brouwer en de heer prof.dr.ir. Bernard Veldkamp. Beide onderzoekers hebben de opdracht via het vier-ogen-principe uitgevoerd.

In dit rapport worden de bevindingen van het onderzoek gepresenteerd. Hoofdstuk 2 en 3 bevatten achtereenvolgens het onderzoekskader en de onderzoeksmethode. De onderzoeksresultaten zijn beschreven in hoofdstuk 4 en hoofdstuk 5 sluit af met de conclusies van het onderzoek. De bevindingen in dit rapport zijn op 18 december 2023 met de opdrachtgever besproken, waarna het rapport door RCEC definitief is vastgesteld.

Namens RCEC,



Dr. Arnold J. Brouwer

Apeldoorn, 18 december 2023

2. Onderzoekskader

2.1. Onderzoekskader centrale examens mbo

Het onderzoekskader centrale examens mbo bestaat uit drie onderdelen. Het eerste onderdeel (paragraaf 2.2) onderzoekt de kwaliteit van de procedures voor het borgen van de inhoudsvaliditeit van een toets of een examen en is ontleend aan de Standards for Educational and Psychological Testing (AERA, 2014). De Standards for Educational and Psychological Testing (AERA, 2014) zijn een set van breed geaccepteerde standaarden die opgesteld zijn door de American Educational Research Association (AERA), de American Psychological Association (APA) en de National Council on Measurement in Education (NCME).

Het tweede en derde onderdeel van het onderzoekskader onderzoeken de betrouwbaarheid (paragraaf 2.3) en de normering (paragraaf 2.4) van het examen en zijn ontleend aan het RCEC Beoordelingssysteem voor de Kwaliteit van Toetsen en Examens (Sanders, Brouwer, Eggen, & Veldkamp, 2018). Het RCEC Beoordelingssysteem voor de Kwaliteit van Toetsen en Examens onderzoekt de kwaliteit van een toets of (praktijk)examen aan de hand van zes criteria, zijnde het doel en gebruik van een toets of examen, het toets- en examenmateriaal, de representativiteit/validiteit, de betrouwbaarheid, de standaardbepaling en normhandhaving, en de afname en beveiliging. Gezamenlijk geven deze criteria een gedetailleerd beeld van de inhoudelijke, organisatorische en toetstechnische kwaliteitsaspecten van een toets of een examen.

De drie onderdelen van het onderzoekskader gezamenlijk geven een antwoord op de in paragraaf 1.2 geformuleerde centrale onderzoeksvraag.

2.2. Inhoudsvaliditeit

Het eerste hoofdstuk van de Standards for Educational and Psychological Testing (AERA, 2014) betreft de validiteit en begint met: *'Validity refers to the degree to which evidence and theory support the interpretations for proposed uses of tests.'* Bij het toepassen van deze definitie op het valideren van toetsscores wordt onderscheid gemaakt tussen de interpretatie van toetsscores (welke betekenis heeft een score?) en het gebruik van toetsscores (welke beslissingen worden op grond van de toetsscores genomen?). Hoe dit van toepassing is op het middelbaar beroepsonderwijs, kan worden toegelicht aan de hand van het centraal examen Engels B1 mbo.

- De score die een student op het centraal examen Engels B1 mbo behaalt, kan men *interpreteren* als een maat voor wat de vaardigheid Engelse taal wordt genoemd.
- Die behaalde score kan men vervolgens *gebruiken* om te beslissen of de student een voldoende of onvoldoende krijgt voor het centraal examen Engels B1 mbo.

In navolging van voornoemde Standards onderscheidt en beschrijft Wools (2013) verschillende soorten bewijzen die gebruikt kunnen / moeten worden om de validiteit van toetsscores aan te tonen. Voor examens zijn vooral inhoudsbewijzen van belang. Bij inhoudsbewijzen gaat het volgens Wools *'om de keuzes die gemaakt worden ten aanzien van de onderwerpen of onderdelen uit de leerstof die in de toets opgenomen worden. Deze keuzes bepalen voor een groot deel of de inhoud van een toets representatief is voor het leerstofdomein of de vaardigheden waarover uitspraken gedaan moeten worden.'* In het huidige onderzoek wordt in plaats van de term inhoudsbewijzen de 'historische' term inhoudsvaliditeit gebruikt en wordt deze gedefinieerd als: *'de mate waarin de inhoud van het examen overeenstemt met het doel van het examen'*. Dat (gebruiks)doel is bij de centrale examens het certificeren van studenten.

Voor het beoordelen van de inhoudsvaliditeit van examens wordt gebruikgemaakt van een onderzoekskader dat gebaseerd is op publicaties van Sireci (1998a, 1998b) en Sireci en Faulkner-Bond (2014). In lijn met deze studies, wordt in het onderzoekskader voor het evalueren van de inhoudsvaliditeit van een toets of examen een onderscheid gemaakt tussen:

1. domein definitie
2. domein representatie
3. domein relevantie
4. adequate toetsconstructieprocedure

Domein definitie verwijst naar hoe het 'construct', bijvoorbeeld de vaardigheid Engelse taal B1, operationeel gedefinieerd wordt. Wat betreft de examens uit het middelbaar beroepsonderwijs komt die operationele definitie neer op het geven van (a) gedetailleerde beschrijvingen van de inhoud van het domein en de cognitieve vaardigheden waar een beroep op wordt gedaan, en (b) toetsspecificaties die zowel de specifieke inhoudscategorieën of inhoudsgebieden als de cognitieve niveaus benoemen. Evaluatie van de domein definitie houdt in dat de overeenstemming onderzocht wordt tussen de operationele definitie van het examen en de heersende opvattingen over het domein van vakdeskundigen uit het werkveld. Voor dit onderzoek worden veelal onafhankelijke vakdeskundigen ingezet bij de ontwikkeling en evaluatie van de toetsspecificaties. De mate waarin belangrijke aspecten van het construct of het curriculum niet voorkomen in de toetsspecificaties is een belangrijk criterium voor de evaluatie van de domein definitie.

Domein representatie verwijst naar de mate waarin een toets het domein zoals dat gedefinieerd is door de toetsspecificaties adequaat representeert en meet. Evaluatie van de domein representatie vereist de inzet van vakdeskundigen die de items/vragen/opdrachten bekijken en beoordelen. Het is de taak van de vakdeskundigen om te bepalen of de items in voldoende mate het beoogde domein representeren.

Domein relevantie betreft de mate waarin elk item van een examen relevant is voor het beoogde domein. Items die belangrijke aspecten van het domein meten, zouden hoge

beoordelingen voor domein representatie moeten krijgen en items die minder belangrijke aspecten meten lage beoordelingen. Ook hier zou men aan vakdeskundigen moeten vragen de relevantie van items voor bepaalde toetsspecificaties te evalueren en deze beoordelingen zou men dan binnen elke inhoudscategorie kunnen aggregeren om de domein relevantie te bepalen.

Adequate toetsconstructieprocedures zullen de inhoudsvaliditeit van examens bevorderen door bij alle onderscheiden stappen van het toetsconstructieproces kwaliteitscontrole te doen plaatsvinden. In de loop der jaren zijn er vele toetsconstructieprocedures voorgesteld die veelal grote overeenkomsten vertonen. Een van de bekendste toetsconstructieprocedures is het twaalf stappenplan dat beschreven staat in het eerste hoofdstuk van Downing en Haladyna (2006). In navolging hierop introduceerde Veldkamp (2016) een tien stappenplan ten behoeve van de (centrale) examens voortgezet onderwijs dat eveneens gebruikt kan worden voor de centrale examens middelbaar beroepsonderwijs zoals wij die in Nederland kennen. In het onderhavige onderzoek wordt nagegaan of de wijze waarop de centrale examens geconstrueerd worden, overeenkomt met de tien stappen uit deze toetsconstructieprocedure.

2.3. Betrouwbaarheid

Afhankelijk van het psychometrisch model dat bij de analyse van toetsscores gebruikt wordt, worden verschillende coëfficiënten berekend om de betrouwbaarheid van de toets of het examen mee aan te geven. In het voorgenomen onderzoek naar de betrouwbaarheid van het centraal examen Engels B1 mbo 2022-2023 wordt ervan uitgegaan dat de toetsscores geanalyseerd zijn met de toets- en item analyse uit de klassieke testtheorie die 'klassieke' betrouwbaarheidscoëfficiënten oplevert.

Bij de beoordeling van de betrouwbaarheid gaat het om de vraag of de betrouwbaarheid voldoende is gezien de beslissingen die met het examen genomen worden. Voor het antwoord op deze vraag hanteert het RCEC standaarden zoals die bijvoorbeeld ook door de COTAN gehanteerd worden. Aangeraden wordt om behalve de betrouwbaarheidscoëfficiënt ook het percentage misclassificaties (ten onrechte geslaagd/gezakt) te vermelden. Deze maat geeft een meer begrijpelijke interpretatie van de betrouwbaarheid dan de betrouwbaarheidscoëfficiënt. Daarnaast kan het ook zinvol zijn om het effect van individuele items op de betrouwbaarheid van het examen vast te stellen. Met behulp van een item-/toetsinformatietabel waarbij de item-/toetsinformatie van de items bij bepaalde vaardigheidsscores, met name ook bij de cesuur, berekend wordt, kan dit nagegaan worden.

2.4. Normering

Het onderzoek naar de normering richt zich op de aspecten standaardbepaling en normhandhaving. Standaardbepaling is het bepalen van de regels om prestaties op examens in cijfers of waarderingen om te zetten. Een belangrijk onderdeel hierbij is het vaststellen van de cesuur, de grens tussen voldoende of onvoldoende.

Normeren kan worden onderscheiden in absoluut normeren en relatief normeren. Kenmerkend voor absoluut normeren is dat de norm of standaard voor afname van de toets wordt bepaald. De absolute norm is gebaseerd op een minimaal acceptabel beheersingsniveau. Bij relatief normeren wordt de norm na afname van de toets bepaald. De relatieve norm is gebaseerd op een onderlinge vergelijking van de toetsprestaties van de studenten.

Methoden voor standaardbepaling kunnen onderscheiden worden in methoden die gebaseerd zijn op de beoordeling van vragen/opgaven/opdrachten van een toets of examen, bijvoorbeeld de absolute cesuurmethode van Angoff, en methoden die gebaseerd zijn op de beoordeling van studenten die een examen maken, bijvoorbeeld de relatieve cesuurmethode van contrasterende groepen.

De experts met vakdeskundigheid die de standaarden vaststellen dienen kennis te hebben van het vakgebied en de eisen waaraan de studenten moeten voldoen. Verder is het noodzakelijk dat de experts getraind zijn in het uitvoeren van de gebruikte standaardbepalingsmethode en het beoordelen van vragen.

Naast standaardbepaling gaat het ook om het handhaven van eenmaal bepaalde standaarden. De methoden om standaarden/normen te handhaven worden meestal met normhandhaving aangeduid. Idealiter worden aan studenten die verschillende varianten van hetzelfde examen doen dezelfde eisen gesteld. Dat betekent dat de varianten van het examen inhoudelijk gelijkwaardig moeten zijn en ook dat de ene variant niet moeilijker of gemakkelijker is dan de andere.

In het volgende hoofdstuk worden de drie onderdelen van het onderzoekskader – inhoudsvaliditeit, betrouwbaarheid en normering – toegepast op het centraal examen Engels B1 mbo 2022-2023.

3. Onderzoeksmethode

3.1. Inhoudsvaliditeit van het centraal examen Engels B1 mbo 2022-2023

Voor het onderzoek naar de inhoudsvaliditeit van het centraal examen Engels B1 mbo spelen het Europees Referentiekader (ERK), de syllabus en de constructieopdracht een cruciale rol. In deze paragraaf bespreken we deze drie instrumenten eerst in algemene zin. In de twee daaropvolgende paragrafen bespreken we de betrouwbaarheid en de normering van het centraal examen Engels B1 mbo 2022-2023.

Europees Referentiekader (ERK)

In Bijlage 2. Referentieniveaus moderne vreemde talen bij artikel 17a, vierde lid, en 17b van het Examen- en kwalificatiebesluit beroepsopleidingen WEB² zijn de voor mbo-4 studenten geldende referentieniveaus vastgesteld. Aanvullend is in Regeling nr. MBO/1045692³ vastgelegd dat in lijn met het Europees Referentiekader (ERK) het centraal examen Engels mbo zal bestaan uit de onderdelen lezen en luisteren. Het ERK is de Nederlandse benaming voor het Common European Framework of Reference for Languages (CEFR). Het is een systeem van niveauomschrijvingen voor de moderne vreemde talen dat een gemeenschappelijke basis verschaft voor de uitwerking van onder andere de examens in heel Europa. Het ERK beschrijft welke taalprestaties bij een bepaald taalbeheersingsniveau horen, met betrekking tot zowel de inhoud (in termen van taalhandelingen in sociale contexten) als de kwaliteit (in termen van grammaticale correctheid, woordenschatbeheersing, uitspraak, spelling). Het ERK onderscheidt zes stadia van taalprestatieniveaus: basisgebruiker A1 en A2, onafhankelijke gebruiker B1 en B2, en vaardige gebruiker C1 en C2. Deze zes niveaus worden nader uitgewerkt in beschrijvingen van wat men kan in de betreffende taal. De beschrijvingen zijn omschreven in descriptoren, de zogenaamde ik-kan stellingen of 'can-do statements'.

Ten behoeve van onder meer de centrale examinering in Nederland, zijn door het nationaal expertisecentrum leerplanontwikkeling (SLO) de Taalprofielen 2015⁴ opgesteld. In de Taalprofielen 2015 is het beheersingsniveau lezen B1 uit het ERK gedefinieerd als: *'de vaardigheid om teksten te begrijpen die hoofdzakelijk bestaan uit hoogfrequente, alledaagse, of aan werk gerelateerde taal en om beschrijvingen van gebeurtenissen, gevoelens en wensen in persoonlijke brieven te begrijpen.'* Dit beheersingsniveau is geoperationaliseerd in descriptoren voor de vier subvaardigheden: (1) correspondentie lezen, (2) oriënterend lezen, (3) lezen om informatie op te doen, en (4) instructies lezen. Het beheersingsniveau Luisteren B1 uit het ERK is gedefinieerd als: *'de vaardigheid om de hoofdpunten te begrijpen wanneer in duidelijk uitgesproken standaardtaal wordt gesproken over vertrouwde zaken die regelmatig aan bod*

² <https://wetten.overheid.nl/BWBR0027963/2023-01-01#Bijlage2>

³ <https://zoek.officielebekendmakingen.nl/stcrt-2016-58595.html>

⁴ <https://www.slo.nl/@4301/taalprofielen-2015/>

komen op werk, op school, en in de vrije tijd plus de vaardigheid om van radio- of tv-programma's over actuele zaken of over onderwerpen van persoonlijk of beroepsmatig belang het gesprek te begrijpen, wanneer er betrekkelijk langzaam en duidelijk gesproken wordt.' Dit beheersingsniveau is geoperationaliseerd in descriptoren voor vijf subvaardigheden: (1) een moedertaalspreker als gesprekspartner begrijpen, (2) gesprekken tussen moedertaalsprekers begrijpen, (3) luisteren als lid van het aanwezige publiek, (4) luisteren naar mededelingen en instructies, en (5) kijken en luisteren naar audiovisuele media.

Om houvast te geven aan examenconstructeurs, docenten en schoolboekenuitgevers, is het CvTE belast met de taak om deze genoemde eisen voor het centraal examen Engels B1 mbo te specificeren. Dat doet het CvTE door middel van een syllabus die periodiek, en voor het laatst op 1 augustus 2017, geactualiseerd wordt op de momenten dat dit om inhoudelijke en/of redactionele redenen noodzakelijk wordt geacht.

Syllabus

Behalve een beschrijving van de exameneisen voor een centraal examen kan een syllabus verdere informatie over het centraal examen bevatten, bijvoorbeeld over een of meer van de volgende onderwerpen: specificaties van examenstof, begrippenlijsten, bekend veronderstelde voorkennis, voorbeeldopgaven, en toelichting op de vraagstelling. De functie van een syllabus is docenten in staat te stellen zich een beeld te vormen van wat in het centraal examen wel en niet gevraagd kan worden. Een syllabus is zodoende een hulpmiddel voor de voorbereiding op een centraal examen. Daarnaast kan het behulpzaam zijn voor onder andere examenconstructeurs. Overige aspecten die de afname van de centrale examens betreffen, zoals protocollen voor afname, de toegestane hulpmiddelen, regels over aanpassingen voor kandidaten met een handicap, regels voor omzetting van scores in cijfers, de aard, de vorm en de tijdsduur van het examen worden door het CvTE bij regeling vastgesteld. Specifieke regelingen, referentiekaders en syllabi voor de centrale examens mbo kan men vinden op www.examenbladmbo.nl.

Constructieopdracht

De wijze waarop de examenstof uit de syllabus wordt getoetst in het centraal examen is vastgelegd in de constructieopdracht van het CvTE. Deze constructieopdracht is uitgewerkt in een algemeen gedeelte dat bestaat uit een viertal productspecificaties:

1. De keuze van tekstsoorten en tekstkenmerken voor de domeinen lezen en kijk-luisteren;
2. De samenstelling van de examenvarianten;
3. De examencondities, zoals tijdsduur, toegestane hulpmiddelen en examenafname-omgeving;
4. De toetsmatrijs, bestaande uit de verdeling van de vragen over de examenstof, de toe te passen vraagvormen en de regels voor de scoring.

De constructieopdracht bevat daarnaast een gedeelte dat de specifieke afspraken voor het studiejaar 2022-2023 beschrijft, waaronder (1) het aantal opgaven en het aantal varianten, (2) het te hanteren design, en (3) de overlap met andere examens.

Bij de vaststelling van vragen, teksten en fragmenten dient de preambule van het CvTE in acht te worden genomen. In de preambule in bijlage 2 van de constructieopdracht centraal examen Engels B1 mbo wordt er uitgebreid op ingegaan dat examens niet aanstootgevend mogen zijn voor bepaalde groepen in de samenleving waarbij 'bepaalde groepen' breed wordt opgevat en niet beperkt is tot politieke of religieuze stromingen.

De constructieopdracht wordt door de inhoudelijk opdrachtgever CvTE verstrekt aan de opdrachtnemer Cito en is beschreven in de procedures en werkwijzen in bijlage 3 van de constructieopdracht. Het centraal examen Engels B1 mbo bestaat uit het aantal varianten conform het vastgestelde design per afnameperiode (bijlage 4 van de constructieopdracht). Hiermee wordt geborgd dat elke variant voldoet aan de toetsmatrijs (bijlage 1 van de constructieopdracht). De verschillende teksten in de varianten hebben zoveel mogelijk dezelfde kenmerken, zoals het aantal scorepunten en de kennis en vaardigheden die getoetst worden. Daarnaast worden de tekstfragmenten getoetst op hun informatiewaarde, alvorens deze worden hergebruikt in een volgende examenvariant. Bij de oplevering van een variant van het centraal examen legt de opdrachtnemer Cito verantwoording af aan de opdrachtgever CvTE over de gevolgde procedure en de gemaakte keuzes.

De vaststellingscommissie van het CvTE beoordeelt conform de processchema's in bijlage 5 van de constructieopdracht of het door Cito gemaakte concept-examen geschikt en passend is als centraal examen en stelt het examen, inclusief de scoringsvoorschriften voor de opgaven vast. Een vaststellingscommissie bestaat uit een voorzitter die afkomstig is uit het middelbaar of het hoger beroepsonderwijs en enkele docenten die lesgeven in examenklassen. Deze leden worden na het doorlopen van een sollicitatieprocedure benoemd door het CvTE. Jaarlijks achteraf evalueert Cito de afgenomen examens aan de hand van de door het CvTE vastgestelde procedures en kwaliteitseisen en legt vervolgens verbetervoorstellen voor aan de vaststellingscommissie.

Zoals beschreven in de vaststellingsprocedures en de werkwijze vaststellingscommissie uit bijlage 3 van de constructieopdracht worden er samengevat vijf fasen onderscheiden:

1. Adviseren over de constructieopdracht. CvTE communiceert hierover met Cito.
2. Vaststellen van opgaven:
 - a. Beoordelen tekst/fragment. De vaststellingscommissie beoordeelt eerst of de teksten en programma's geschikt zijn, voordat de constructiegroep examenopgaven maakt.

- b. Vaststellen nieuwe opgaven. Voordat opgaven aan de vaststellingscommissie worden aangeboden, heeft Cito de inhoudelijke en redactionele controle van de opgaven/sleutel/label uitgevoerd.
3. Vaststellen van examens:
 - a. Vaststellen van ingevuld design. In deze fase worden de verschillende varianten vastgesteld, conform het geldende design.
 - b. Vaststellen van basisvarianten (preview exemplaar).
 - c. Vaststellen afnamevarianten door middel van het trekken van items uit de itembank.
4. Advisering bij de normering. Cito levert het technisch advies: een lijst met te bespreken items en een beargumenteerd voorstel voor het al dan niet neutraliseren van de opgave.
5. Evalueren van de informatiewaarde van de ingezette tekstfragmenten en de individuele items/opgaven. Aanvullend wordt de actualiteitscheck uitgevoerd voor de jaren waaruit teksten ingezet zullen worden in het komende studiejaar in het kader van hergebruik. Dit gebeurt volgens de notitie Procedure actualiteitscheck talen mbo.

Daarmee bestaat het constructie- en vaststellingsproces feitelijk uit twee onderdelen: (1) het construeren en vaststellen van tekstfragmenten en bijbehorende opgaven, en (2) het genereren van nieuwe varianten van het examen door middel van het trekken van items uit de itembank.

3.2. Betrouwbaarheid van het centraal examen Engels B1 mbo 2022-2023

Om de betrouwbaarheid van de verschillende varianten te onderzoeken is gebruik gemaakt van de door Cito uitgevoerde toets- en item analyses op basis van de klassieke testtheorie en de door Cito gekalibreerde toetsopgaven onder de item-responstheorie (IRT) in de Dexter software⁵. Door RCEC is ten eerste gekeken naar de door Cito geschatte gemiddelde moeilijkheidsgraad en het gemiddeld onderscheidend vermogen van de items en naar de door Cito geschatte betrouwbaarheid van de toetsvarianten. Vervolgens is door RCEC de op item-responstheorie gebaseerde toetsinformatiefunctie en standaardfout bij de cesuur samengesteld en beoordeeld op basis van de door Cito geschatte itemparameters. Dit laatste om ook inzicht te krijgen in de lokale betrouwbaarheid, dat wil zeggen de nauwkeurigheid van de onderscheiden vaardigheidsscores van in het bijzonder die van de zak/slaaggrens (cesuur).

In de beoordeling van de resultaten van de door Cito uitgevoerde toets- en item analyses op basis van de klassieke testtheorie worden de kwaliteitsnormen voor de gemiddelde p-waarde en gemiddelde rir-waarde uit het RCEC Beoordelingssysteem voor de Kwaliteit van Toetsen en

⁵ Dexter is software in de vorm van een R-pakket voor het beheren en analyseren van toetsgegevens. Met dit open source pakket, gepubliceerd op CRAN-R.-project, kunnen naast klassieke toets- en itemanalyses ook IRT-analyses worden uitgevoerd. Dexter is geïnspireerd door het One Parameter Logistic Model (OPLM).

Examens (Sanders et al., 2018) gehanteerd. In het onderzoek van de betrouwbaarheid van de toetsvarianten staat de greatest lower bound (glb) centraal, vanwege het feit dat deze coëfficiënt de betrouwbaarheid schat aan de hand van clusters van items die parallel zijn. Dit past bij de heterogeniteit van het examen, waarin naast lezen ook kijk-luisteren wordt getoetst. Daarnaast zijn de Cronbach's alpha en lambda2 gepresenteerd. Tevens is door RCEC het percentage niet-consistente beslissingen (de zogenaamde misclassificaties) geschat als functie van het percentage gezakte studenten en de toetsbetrouwbaarheid (Eggen & Sanders, 1993). In het RCEC Beoordelingssysteem voor de Kwaliteit van Toetsen en Examens wordt voor de betrouwbaarheid van een high-stakes examen aangegeven dat deze .80 of hoger dient te zijn. Het geschatte percentage misclassificaties is daarbij gelijk aan 15% of lager.

De lokale betrouwbaarheid rond de cesuur is onderzocht door op basis van de door Cito geschatte en met RCEC gedeelde item parameters de toets informatie functie te berekenen. Met deze functie is de standaardfout voor de verschillende waarden van de vaardigheidsparameter berekend en is in kaart gebracht wat de standaardfout is ter hoogte van de cesuur. Zo is af te leiden in welke mate het examen optimaal meet rond de cesuur.

3.3. Normering van het centraal examen Engels B1 mbo 2022-2023

Het onderzoek naar de normering van het centraal examen Engels B1 mbo richt zich op de beide in paragraaf 2.4 genoemde aspecten van de normering: standaardbepaling en normhandhaving.

Bij het centraal examen Engels B1 mbo 2022-2023 is er sprake van een eerder door CvTE vastgestelde absolute standaard/cesuur/norm, de zogenaamde referentiecesuur. De referentiecesuur is bepaald door de ERK-niveaus, zoals vastgesteld op de centrale examens VO Engels 2012, via een posttest over te brengen⁶.

Het kunnen vaststellen of de standaard correct bepaald is, vereist conform het RCEC Beoordelingssysteem voor de Kwaliteit van Toetsen en Examens antwoord op drie vragen:

1. Is de gehanteerde standaardbepalingmethode op de juiste wijze uitgevoerd?
2. Zijn de experts met vakdeskundigheid die de standaard bepalen naar behoren geselecteerd en getraind?
3. Is er voldoende overeenstemming tussen de experts met vakdeskundigheid?

Zoals beschreven in paragraaf 3.1 is er bij het centraal examen Engels B1 mbo 2022-2023 sprake van meerdere varianten van hetzelfde examen. De eerder door het CvTE vastgestelde referentiecesuur wordt daarbij toegepast middels equivalering van de vaardigheidsschaal.

⁶ <https://wetten.overheid.nl/BWBR0036876/2017-08-01>

Het bepalen of de referentiecesuur op de juiste wijze gehandhaafd wordt, vereist conform het RCEC Beoordelingssysteem voor de Kwaliteit van Toetsen en Examens antwoord op twee vragen:

1. Is de gekozen normhandhavingsmethode van voldoende kwaliteit?
2. Is de normhandhavingsmethode correct uitgevoerd?

In hoofdstuk 4 wordt verslag gedaan van het door RCEC uitgevoerde onderzoek naar de inhoudsvaliditeit, betrouwbaarheid en normering van het centraal examen Engels B1 mbo 2022-2023.

4. Onderzoeksresultaten

4.1. Inhoudsvaliditeit van het centraal examen Engels B1 mbo 2022-2023

In dit hoofdstuk wordt het onderzoekskader centrale examens mbo uit hoofdstuk 2 toegepast op het centraal examen Engels B1 mbo 2022-2023. Het onderzoek start met de domein definitie (paragraaf 4.2), uitgewerkt in de inhoud van het examen (paragraaf 4.2.1) en de toetsspecificaties (paragraaf 4.2.2). Hierna volgen domein representatie (paragraaf 4.3), domein relevantie (paragraaf 4.4), en adequate toetsconstructieprocedures (paragraaf 4.5). De vier voornoemde onderdelen geven samen een beeld van de inhoudsvaliditeit van het centraal examen Engels B1 mbo 2022-2023.

4.2. Domein definitie

De domein definitie voor het centraal examen Engels B1 mbo 2022-2023 betreft een gedetailleerde beschrijving van:

1. De inhoud van het examen en de cognitieve vaardigheden waar een beroep op wordt gedaan;
2. De toetsspecificaties die zowel de specifieke inhoudscategorieën of inhoudsgebieden als de cognitieve niveaus benoemen.

Onderstaande beschrijving van de inhoud van het examen en de toetsspecificaties zijn nagenoeg letterlijk ontleend aan de bespreking van de syllabus 'Engels mbo, referentieniveau B1', versie 1 augustus 2017 en aan de constructieopdracht 'Engels mbo B1, 2022-2023', versie januari 2022.

4.2.1. Inhoud van het examen

Engelse taal is een verplicht examenonderdeel voor alle mbo-opleidingen op niveau 4. Het centraal examen Engels B1 mbo omvat de onderdelen lezen en kijk-luisteren. Bij het onderdeel lezen kunnen vier tekstsoorten worden opgenomen in een examen: (1) (zakelijke) correspondentie, (2) teksten t.b.v. oriënterend lezen van bijvoorbeeld nieuwsbrieven, contracten en brochures, (3) informatieve teksten zoals bijvoorbeeld krantenartikelen, voorlichtingsmateriaal en productinformatie, en (4) instructieve teksten zoals bijvoorbeeld gebruiksaanwijzingen en routebeschrijvingen.

Bij het onderdeel kijk-luisteren kunnen eveneens vier tekstsoorten worden opgenomen in een examen: (1) gesprekken tussen moedertaalsprekers zoals bijvoorbeeld een debat en een tweegesprek waarbij interactie is tussen beide sprekers, (2) monoloog voor een publiek zoals bijvoorbeeld een speech of een vlog, (3) instructieve teksten zoals bijvoorbeeld

aankondigingen en gebruiksaanwijzingen, en (4) reportages zoals bijvoorbeeld nieuwsberichten en documentaires. Bij het onderdeel kijk-luisteren kunnen geluidsopnamen met en zonder beeld gebruikt worden. In de praktijk is er voor gekozen om de verdeling van lees- en luisteropgaven vergelijkbaar te houden: beide ongeveer 50%. Het examen kent een tijdsduur van 90 minuten.

Bij de examinering van de vaardigheden lezen en kijk-luisteren in het centraal examen Engels B1 mbo worden teksten opgenomen die gaan over vertrouwde en alledaagse onderwerpen uit het eigen interessegebied van de student. Het kunnen ook alledaagse onderwerpen zijn die voortkomen uit de generieke kwalificatie-eisen op het gebied van loopbaan en de vier dimensies van burgerschap: (1) politiek-juridische dimensie, (2) economische dimensie, (3) sociaal-maatschappelijke dimensie, en (4) vitaal burgerschap. Hiermee wordt aangesloten bij de taalprestaties behorende bij beheersingsniveau B1, zoals gedocumenteerd in het ERK en zoals geoperationaliseerd in de Taalprofielen 2015.

Het B1-niveau is typerend voor een onafhankelijke gebruiker die: (1) de belangrijkste punten kan begrijpen uit duidelijke standaardteksten over vertrouwde zaken die regelmatig voorkomen op het werk, op school en in de vrije tijd, die (2) zich kan redden in de meeste situaties die kunnen optreden tijdens reizen in gebieden waar de taal wordt gesproken, die (3) een eenvoudige lopende tekst kan produceren over onderwerpen die vertrouwd of die van persoonlijk belang zijn, en die (4) een beschrijving kan geven van ervaringen en gebeurtenissen, dromen, verwachtingen en ambities en korte redenen en verklaringen kan geven voor meningen en plannen (Council of Europe, 2008).

Elke variant van het centraal examen Engels B1 mbo 2022-2023 biedt derhalve meerdere teksten of fragmenten aan, waarbij gestreefd wordt naar diversiteit in bovengenoemde onderwerpen. Daarmee kan beoordeeld worden of de student met teksten over verschillende onderwerpen uit de voeten kan, zoals op het door ERK gedefinieerde niveau B1 verwacht mag worden.

De vragen die bij de teksten en de (kijk-)luisterfragmenten worden gesteld, zijn bedoeld om te achterhalen in hoeverre de kandidaat de tekst of het luisterfragment heeft begrepen, uitgaande van het B1 referentieniveau. Daarbij gaat het er om te achterhalen wat zich in het hoofd van de kandidaat afspeelt door te testen hoe precies het begrip van de tekst, de alinea, de zin of het woord is. In het examen kunnen meerkeuzevragen met 3 of 4 alternatieven, matrixvragen en multiple-respons vragen voorkomen van verschillende orde, zoals:

- **Taakgerichte vragen:** Deze richten zich meestal op situaties uit het echte leven en testen hoe goed de kandidaat in staat is om taken uit te voeren of oplossingen voor problemen voor te stellen in een specifieke context.

- **Begrip van feitelijke informatie op woord- en zinsniveau:** Dit type vraag kijkt of de kandidaat de letterlijke betekenis van woorden en zinnen begrijpt, zonder dieper in te gaan op de bredere context van de tekst.
- **Begrip van hoofdpunten op tekst- en alinea-niveau:** Hierbij wordt gekeken naar het vermogen van de kandidaat om de belangrijkste informatie in een tekst of alinea te identificeren, zonder in te gaan op details.
- **Afleiden van minder frequente woorden uit de context:** Dit test het vermogen van de kandidaat om de betekenis van minder vaak voorkomende woorden te begrijpen door de context waarin ze worden gebruikt.
- **Begrip van minder feitelijke informatie, zoals wensen en bedoelingen:** Deze vragen kijken naar het begrip van meer impliciete of subjectieve aspecten van de tekst, zoals de bedoelingen van de schrijver of de gevoelens die worden uitgedrukt.
- **Onderscheid maken tussen significante en minder significante punten:** Dit type vraag test de vaardigheid van de kandidaat om belangrijke informatie van bijzaken te onderscheiden en de kernpunten van een tekst te identificeren.

Het taalgebruik is alledaags en duidelijk. Geschreven teksten zijn goed gestructureerd en kunnen uitgebreider zijn, met een lengte tussen de 600 en 1.000 woorden. Bij kijk- en luisterfragmenten wordt een normaal spreektempo gehanteerd, en de sprekers gebruiken helder en goed gearticuleerde standaardtaal, waaronder varianten van het Engels, zoals Amerikaans, Brits of Australisch (die allemaal zijn toegestaan). De fragmenten zijn niet te lang, meestal tussen de 2 en 6 minuten. De betekenis van onbekende woorden en zinnen kan worden afgeleid uit de context, op voorwaarde dat het onderwerp bekend is. Het examen omvat ook teksten en opgaven die variëren van onderaan niveau B1 tot bovenaan niveau B1. De teksten aan de onderkant van dit niveau focussen meer op het begrijpen van feitelijke informatie. Aan de bovenkant van niveau B1 betekent dit dat teksten en luisterfragmenten langer zijn, met meer abstracte onderwerpen die mogelijk minder vertrouwd zijn. De tekstindeling kan afwijkend of complexer zijn, en er wordt een uitgebreidere woordenschat gebruikt met meer ongebruikelijke woorden.

4.2.2. Toetsspecificaties

In de toetsmatrijs van het centraal examen Engels B1 mbo 2022-2023, zoals opgenomen in bijlage 1 van de constructieopdracht, komen de volgende toetsspecificaties aan de orde:

- **Lengte van de scoreschaal.** De student kan maximaal circa 40 punten behalen.
- **Vraagvormen.** Er is sprake van drie vraagvormen: (1) meerkeuzevraag met maximaal 4 antwoorden waarvan slechts 1 antwoord correct is, (2) multiple responsvraag, en (3) matrixvraag met een aantal juist/onjuist-beweringen.

- **Scoring van de vragen.** Voor alle vragen geldt dat bij het correct beantwoorden van de vraag 1 punt wordt toegekend, terwijl er geen punten worden toegekend bij een onjuist antwoord (dichotoom scoren).
- **Aantal teksten en bronnen.** In elke variant worden vijf tekstfragmenten voor lezen en vier teksten voor kijk-luisteren opgenomen. Daarbij dienen er gevarieerde bronnen gebruikt te worden.
- **Normering.** Bij de normering van het examen worden scores omgezet in cijfers waarbij rekening wordt gehouden met de verschillen in moeilijkheid van de verschillende examenvarianten.

Naast voornoemde specificaties zijn er nog de volgende aandachtspunten:

- **Spelling, taalconventies en diversiteit van de native standaarden.** Examenvragen volgen het taalgebruik en de spelling van de oorspronkelijke tekst: Amerikaans voor Amerikaanse teksten, Brits voor Britse teksten, en meer Brits in teksten uit landen zoals Canada, Australië en Nieuw-Zeeland. Verdere richtlijnen zijn te vinden in de vaststellingscommissie-besluitenlijst van Cito.
- **Authentieke teksten.** Het gebruik van echte teksten is cruciaal, aanpassingen zoals inkorten zijn toegestaan. Fouten worden gecorrigeerd indien deze hinderlijk zijn voor de examenkandidaten en leestekens voor leesbaarheid mogen toegevoegd of weggelaten worden. Moeilijke woorden/begrippen mogen waar nodig vervangen of verduidelijkt worden.
- **Bronnen.** Vanwege de geheimhouding worden bronteksten voorzien van een nieuwe titel, met uitzondering van de betogende leesteksten. Ook de naam van bijvoorbeeld het (televisie)programma waaruit de kijk- luisterfragmenten afkomstig zijn wordt niet vermeld. Cito beheert intern de licenties en bronnen behorende bij de examens.

Voor kandidaten met een auditieve beperking (doof of slechthorend) worden voor elke afnameperiode twee speciale dos-varianten samengesteld. Deze varianten bestaan uit al eerder bij de reguliere doelgroep afgenomen materiaal. Speciaal is dat het kijk-luistergedeelte bestaat uit bij de doelgroep passende fragmenten met beeld die ondertiteld zijn. De voorzitter stelt deze varianten vast zonder tussenkomst van de overige vaststellingscommissieleden.

4.3. Domein representatie

Domein representatie betreft de mate waarin een examen het domein zoals dat gedefinieerd is door de toetsspecificaties adequaat representeert en meet. Het behoort tot de expertise van vakdeskundigen om te bepalen of de opgaven in voldoende mate het beoogde domein representeren. In paragraaf 3.1 is beschreven hoe de inhoudelijk opdrachtgever CvTE voor ieder afzonderlijk vak een constructieopdracht verstrekt aan de opdrachtnemer Cito.

Hierna bespreken we het design van de variant P2-05 van het centraal examen Engels B1 mbo 2022-2023. De bevindingen zijn gepresenteerd in [Tabel 1](#).

Tabel 1.

Design van variant P2-05 van het centraal examen Engels B1 mbo 2022-2023

Algemene gegevens: examen voor schoolsoort mbo-4 met tijdsduur van 90 minuten

Lezen:

1. 1 tekstfragment + illustratie
 - a. 3x 3-keuzevraag
 - b. 1x matrixvraag
2. 1 tekstfragment + illustratie
 - a. 3x 3-keuzevraag
 - b. 1x matrixvraag
3. 1 tekstfragment + illustratie
 - a. 3x 3-keuzevraag
 - b. 1x matrixvraag
4. 1 tekstfragment + illustratie
 - a. 3x 3-keuzevraag
 - b. 1x matrixvraag
5. 1 tekstfragment + illustratie
 - a. 3x 3-keuzevraag
 - b. 1x matrixvraag

Kijk-luisteren:

1. 1 videofragment
 - a. 5x 3-keuzevraag
2. 2 videofragmenten
 - a. 4x 3-keuzevraag
 - b. 1x matrixvraag
3. 1 videofragment
 - a. 5x 3-keuzevraag
4. 1 videofragment
 - a. 4x 3-keuzevraag
 - b. 1x matrixvraag

De inhoudelijke indeling van de onderzochte examenvariant P2-05 is in lijn met de gewenste verdeling van lees- en luisteropgaven van beide 50%, is conform de toetsmatrijs uit bijlage 1 van de constructieopdracht, conform de mogelijke vraagtypes, en conform het generiek design uit bijlage 4 van de constructieopdracht.

4.4. Domein relevantie

Domein relevantie betreft de mate waarin elke opgave van een examen relevant is voor het beoogde domein. Om de relevantie van opgaven te bepalen, is de inzet van vakdeskundigen (docenten Engels binnen het mbo) nodig. Bij de constructie van alle centrale examens, dus ook van het centraal examen Engels B1 mbo 2022-2023, zijn altijd twee groepen toets- en

vakdeskundigen betrokken: leden van de constructiegroep en leden van de vaststellingscommissie.

Bij de constructie van de opgaven en antwoordalternatieven zijn docenten als constructiegroeplid betrokken. Deze docenten selecteren de contexten en leveren de opgaven daarbij aan bij de Cito-toetsdeskundige. Een constructiegroeplid is steevast een docent met ervaring in examenklassen. Per examen wordt in de regel door een drietal constructiegroepleden materiaal aangeleverd. De Cito-toetsdeskundige stelt het examen samen en bespreekt het concept met de CvTE-vaststellingscommissie. De vaststellingscommissie is verantwoordelijk voor de vaststelling van de examenopgaven.

Voor de vaststellingsprocedure van het centraal examen Engels B1 mbo 2022-2023 wordt verwezen naar paragraaf 3.1 en de in bijlage 3 van de constructieopdracht beschreven vaststellingsprocedures en werkwijze vaststellingscommissie.

In paragraaf 3.1 is de samenstelling en vakdeskundigheid van de vaststellingscommissie besproken. Van de constructiegroepleden en de vaststellingscommissie wordt aangenomen dat zij in staat zijn de relevantie van de opgaven voor het meten van de taalvaardigheid te kunnen beoordelen.

4.5. Adequate toetsconstructieprocedure

Het gebruik van een adequate toetsconstructieprocedure zal de inhoudsvaliditeit van alle examens in het algemeen ten goede komen. Zoals beschreven in paragraaf 2.1., wordt aan de hand van het tienstappenplan van Veldkamp (2016) de constructieprocedure die bij het centraal examen Engels B1 mbo 2022-2023 gehanteerd wordt tegen het licht gehouden. De toetsconstructieprocedure van Veldkamp (2016) onderscheidt de volgende tien stappen:

1. Maak een toetsplan
2. Definieer de inhoud
3. Stel de toetsmatrijs op
4. Ontwikkel de items
5. Stel de toets samen
6. Neem de toets af
7. Scoor de antwoorden
8. Bepaal de zak/slaaggrens
9. Rapporteer de scores
10. Documenteer de voorgaande stappen

Hieronder wordt nagegaan of deze stappen ook bij de constructie van het centraal examen Engels B1 mbo 2022-2023 gehanteerd worden.

Deelstap	Beoordeling
Ad 1. Maak een toetsplan Bij het centraal examen Engels B1 mbo 2022-2023 is het duidelijk welke constructen gemeten worden, namelijk de vaardigheden lezen en luisteren op niveau B1. Het examen is bedoeld om studenten te certificeren.	✓
Ad 2. Definieer de inhoud Het definiëren van de inhoud is beschreven in paragraaf 4.2.1 van de domein definitie.	✓
Ad 3. Stel de toetsmatrijs op Voor deze stap zie paragraaf 4.2.2 waar de toetsspecificaties vermeld worden.	✓
Ad 4. Ontwikkel de items Deze stap betreft de constructieopdracht op basis waarvan de opdrachtnemer Cito de opgaven voor een concept-examen construeert.	✓
Ad 5. Stel de toets samen De vaststellingscommissie van het CvTE stelt de examenvarianten vast.	✓
Ad 6. Neem de toets af Het centraal examen Engels B1 mbo 2022-2023 wordt onder regie van het CvTE op scholen afgenomen.	✓
Ad 7. Scoor de antwoorden Het examen wordt digitaal afgenomen en bestaat enkel uit geautomatiseerd scorebare vragen.	✓

<p>Ad 8. Bepaal de zak/slaaggrens</p> <p>Voor de centrale examens mbo Engels is de referentiecesuur bepaald door de ERK-niveaus, zoals vastgesteld op de centrale examens VO Engels 2012, via een posttest over te brengen. Met behulp van equivalering wordt voor elke variant van het centraal examen Engels B1 mbo 2022-2023 de bijpassende vaardigheidsscore vastgesteld. Zo kunnen de verschillen in zwaarte binnen de verschillende varianten gecompenseerd worden.</p>	✓
<p>Ad 9. Rapporteer de scores</p> <p>De kandidaten ontvangen korte tijd na het afleggen van het examen hun cijfer. Het cijfer wordt vastgesteld door de examencommissie op de scholen op basis van de door het CvTE vastgestelde normering.</p>	✓
<p>Ad 10. Documenteer de voorgaande stappen</p> <p>Voorgaande stappen zijn in verschillende documenten beschreven en veelal digitaal beschikbaar via de websites van CvTE en Cito.</p>	✓

4.6. Betrouwbaarheid van het centraal examen Engels B1 mbo 2022-2023

Van vier varianten van het centraal examen Engels B1 mbo 2022-2023 is een betrouwbaarheidsanalyse uitgevoerd: de varianten P2-05 en P2-08 uit afnameperiode 2 (17 oktober t/m 16 december 2022) en de varianten P4-19 en P4-22 uit afnameperiode 4 (6 maart t/m 28 april 2023).

Tabel 2 geeft een samenvatting van de door Cito uitgevoerde toets- en item analyses op basis van de klassieke testtheorie. Per examenvariant is het totaal aantal respondenten (N) en de verdeling van de items over de verschillende toetsonderdelen gepresenteerd. Vervolgens zijn een tweetal itemparameters, zoals berekend door Cito, gepresenteerd. Ten eerste is de gemiddelde moeilijkheidsgraad van alle items per toetsvariant uitgedrukt in de gemiddelde p-waarde. Deze bevindt zich overal binnen de voor deze toets optimale waarde tussen 0,20 en 0,80 op een schaal van 0 – 1. Het gemiddeld onderscheidend vermogen van alle items per toetsvariant is uitgedrukt in de rir-waarde (c.q. item-rest correlatie). De rir-waarde van een item is een indicatie voor de mate waarin het item onderscheid maakt tussen studenten met hoge toetsscores en studenten met lage toetsscores. Op basis van Ebel en Frisbie (1986) en zoals gepresenteerd in Tabel 3.12 in Eggen & Sanders (1993), is de gemiddelde rir-waarde van deze vier toetsvarianten [0,35 – 0,38] te beoordelen als goed.

De toetsbetrouwbaarheid is gepresenteerd in een drietal coëfficiënten, de Cronbach's alpha, de lambda2 en de greatest lower bound (glb). Ook dit betreft de waarden zoals berekend door Cito. De betrouwbaarheidscoëfficiënt bevindt zich overal boven de voor een summatieve toets vereiste ondergrens van 0,80. Met behulp van de verdeling van de somscores over de studenten en de omzettingstabellen CE per variant is door RCEC het percentage gezakte kandidaten per examenvariant berekend. Van daaruit is door RCEC het percentage niet-consistente beslissingen (de zogenaamde misclassificaties) geschat als functie van het percentage gezakte studenten en de toetsbetrouwbaarheid (Eggen & Sanders, 1993).

Tabel 2.

Resultaten van de toets- en itemanalyse van de vier examenvarianten

examenvariant	P2-05	P2-08	P4-19	P4-22
toetssamenstelling				
N	1.471	1.500	1.124	1.149
totaal items	40	40	40	40
lezen (5 teksten)	20	20	20	20
kijk-luisteren (4 teksten)	20	20	20	20
item parameters				
gem. p-waarde	0,71	0,72	0,76	0,74
gem. rir-waarde	0,37	0,35	0,38	0,37
betrouwbaarheid				
alpha	0,88	0,86	0,88	0,88
lambda2	0,88	0,86	0,88	0,88
glb	0,91	0,90	0,91	0,91
%onvoldoende*	9,79%	11,26%	11,48%	10,27%
%misclassificaties**	[4 - 6%]	[6 - 8%]	[6 - 8%]	[6 - 8%]

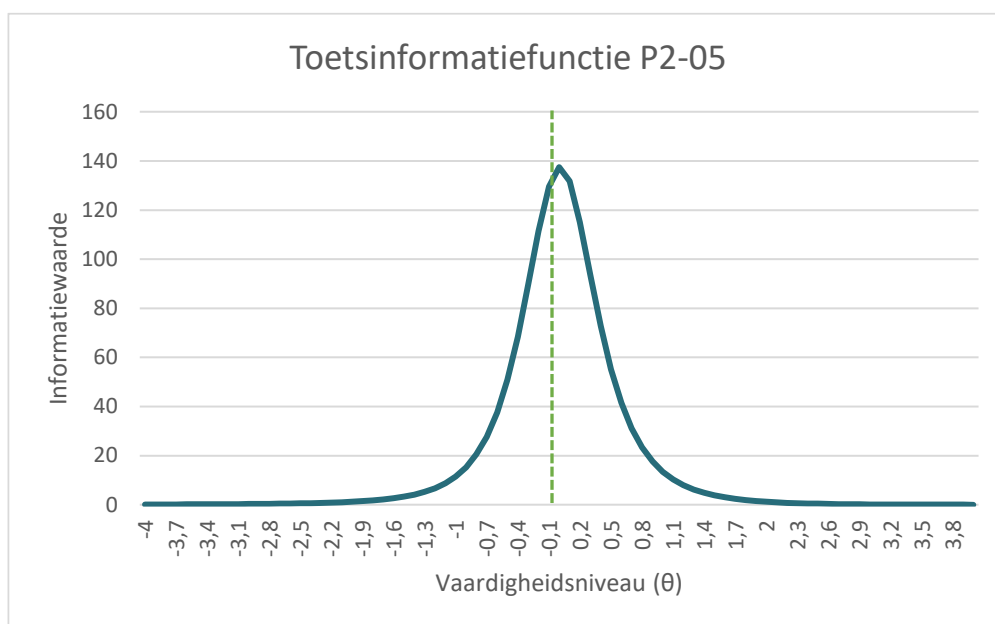
* %kandidaten dat een onvoldoende heeft behaald op het centraal examen, berekend op basis van de somscores en de omzettingstabellen CE per variant.

** Percentage niet-consistente beslissingen als functie van %kandidaten met een onvoldoende en de betrouwbaarheid (Tabel 3.13 Psychometrie in de Praktijk, Eggen & Sanders, 1993).

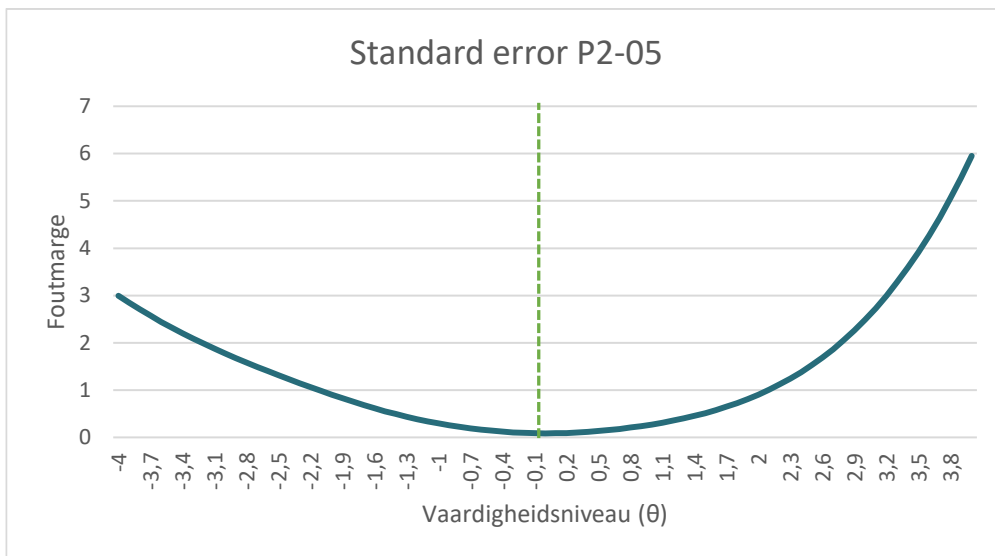
Aanvullend op de psychometrische analyses, uitgevoerd onder de klassieke testtheorie, zijn door Cito psychometrische analyses uitgevoerd onder de item-responstheorie (IRT). IRT is een theorie binnen de testleer waarbij op het niveau van items op een wiskundige manier een relatie wordt gelegd tussen de antwoorden die studenten geven en een onderliggende latente vaardigheid. Voor de verschillende psychometrische analyses heeft Cito gebruik gemaakt van de Dexter software. Dexter is geïnspireerd door het One Parameter Logistic Model (OPLM). In dit psychometrisch model worden per item (c.q. toetsvraag) de populatie onafhankelijke moeilijkheidsgraad, uitgedrukt in de geschatte b-parameter, en het onderscheidend

vermogen, binnen OPLM uitgedrukt in de geïmputeerde a-parameter, bepaald. De beide itemparameters van de twee varianten uit afnameperiode 4 (6 maart t/m 28 april 2023) zijn daarbij door Cito gekalibreerd op basis van de totaalkalibratie uit afnameperiode 2 (17 oktober t/m 16 december 2022).

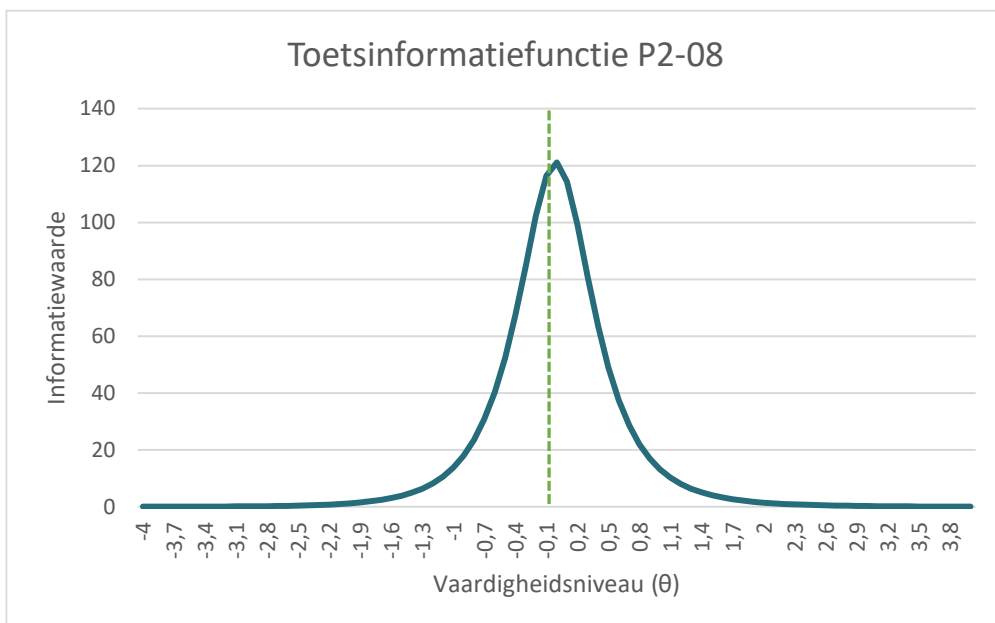
Figuur 1 visualiseert de toetsinformatiefunctie en de standaardfout van de vier varianten, op basis van de in de Dexter software door Cito gekalibreerde items. Aan onderstaande grafieken is met de groen gestreepte verticale lijn de referentiecesuur, uitgedrukt in de theta-waarde -0,15 als omzetting van het rapportcijfer 5,5 toegevoegd (zie het Cito Excel bestand met referentiecijfers).



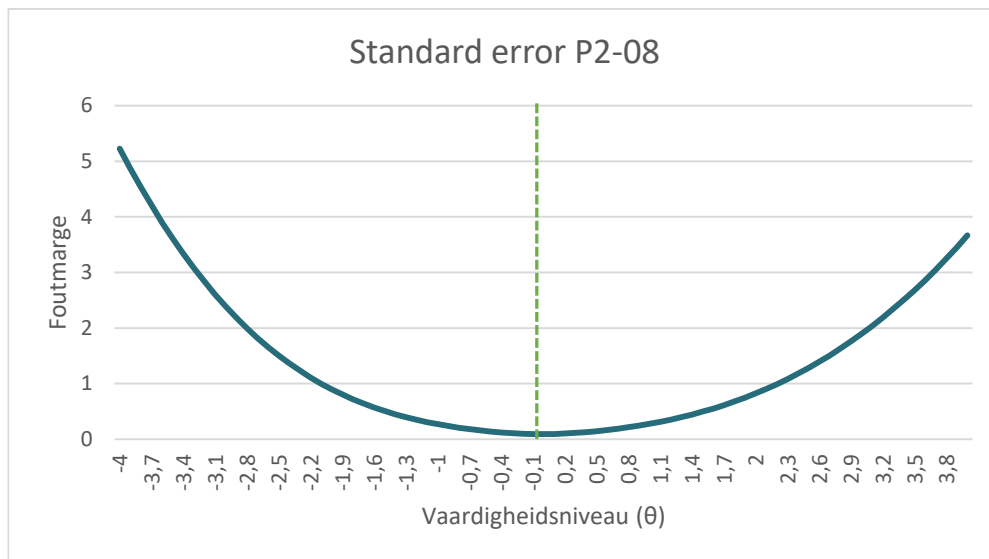
- Informatiewaarde P2-05 bij cesuur $\Theta = -0,15$ is $TIF(\Theta) = 129,57$
- Maximale informatiewaarde P2-05 is $TIF(\Theta) = 137,43$ bij cesuur van $\Theta = 0,00$



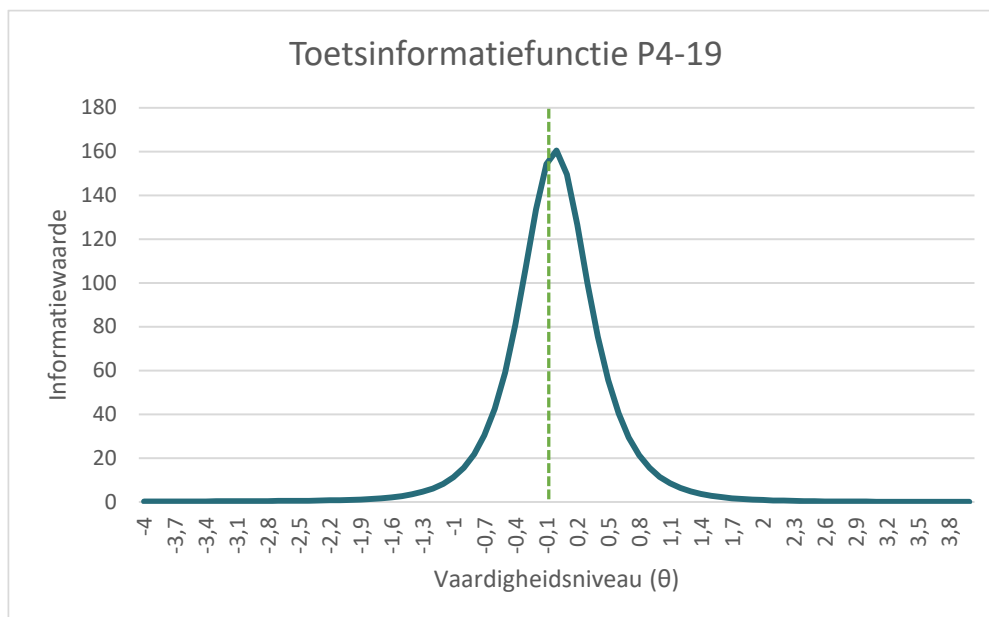
- Foutmarge P2-05 bij cesuur $\Theta = -0,15$ is $SE(\Theta) = 0,088$
- Kleinste foutmarge P2-05 is $SE(\Theta) = 0,085$ bij cesuur van $\Theta = 0,00$



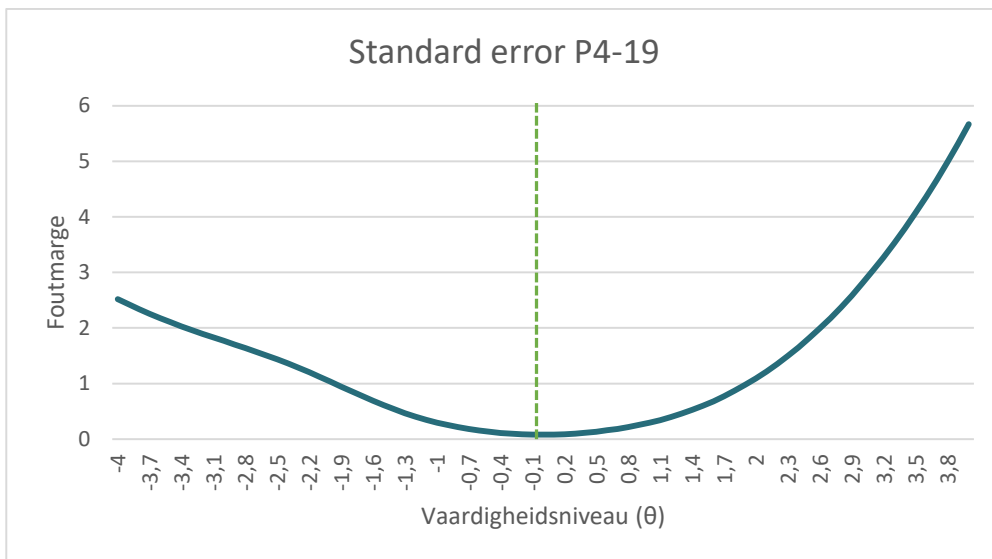
- Informatiewaarde P2-08 bij cesuur $\Theta = -0,15$ is $TIF(\Theta) = 116,44$
- Maximale informatiewaarde P2-08 is $TIF(\Theta) = 121,15$ bij cesuur van $\Theta = 0,00$



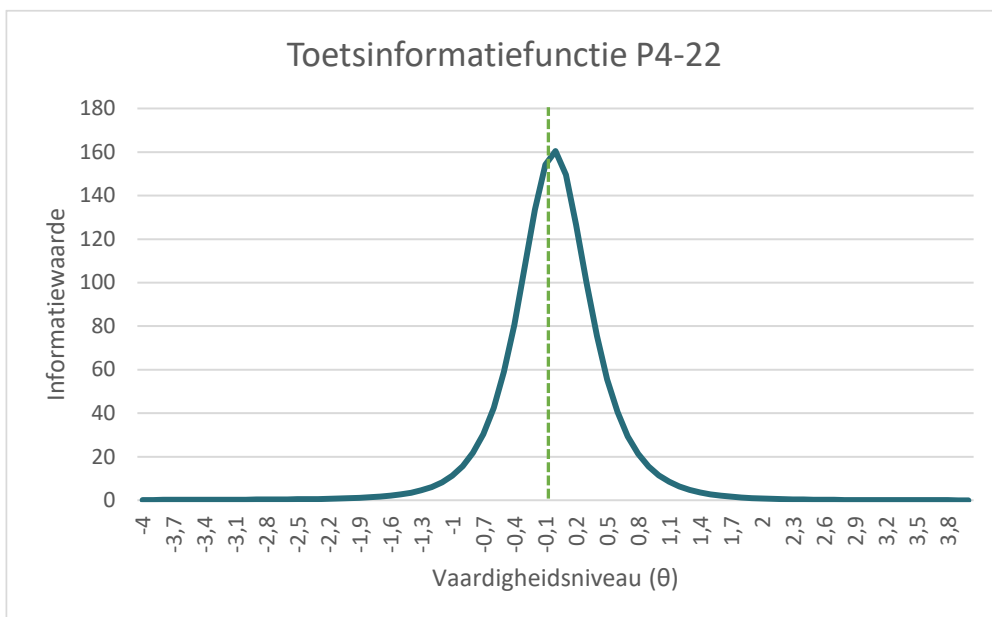
- Foutmarge P2-08 bij cesuur $\Theta = -0,15$ is $SE(\Theta) = 0,093$
- Kleinste foutmarge P2-08 is $SE(\Theta) = 0,091$ bij cesuur van $\Theta = 0,00$



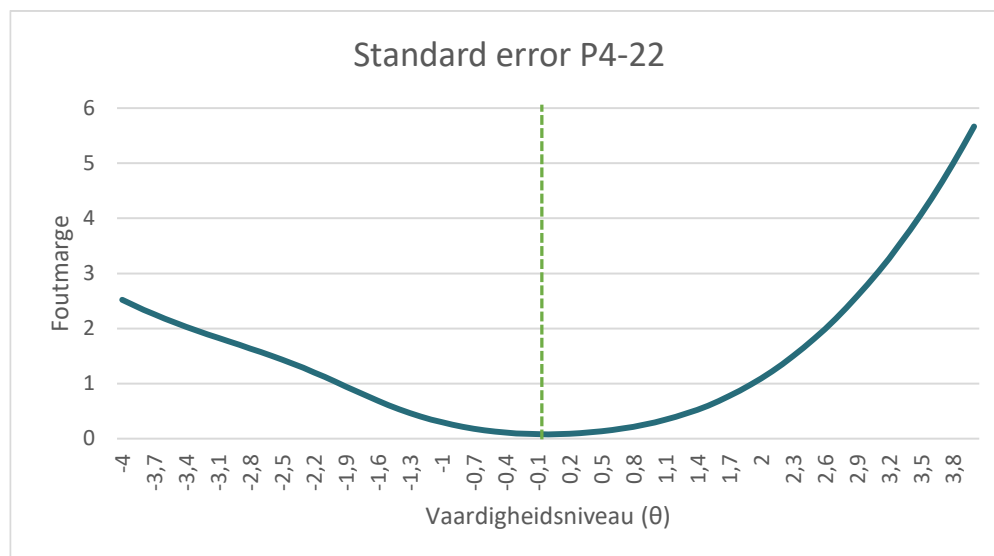
- Informatiewaarde P4-19 bij cesuur $\Theta = -0,15$ is $TIF(\Theta) = 154,27$
- Maximale informatiewaarde P4-19 is $TIF(\Theta) = 160,52$ bij cesuur van $\Theta = 0,00$



- Foutmarge P4-19 bij cesuur $\Theta = -0,15$ is $SE(\Theta) = 0,081$
- Kleinste foutmarge P4-19 is $SE(\Theta) = 0,079$ bij cesuur van $\Theta = 0,00$



- Informatiewaarde P4-22 bij cesuur $\Theta = -0,15$ is $TIF(\Theta) = 137,11$
- Maximale informatiewaarde P4-22 is $TIF(\Theta) = 142,24$ bij cesuur van $\Theta = 0,00$



- Foutmarge P4-22 bij cesuur $\Theta = -0,15$ is $SE(\Theta) = 0,085$
- Kleinste foutmarge P4-22 is $SE(\Theta) = 0,084$ bij cesuur van $\Theta = 0,00$

Figuur 1. De toetsinformatiefunctie en standaardfout versus de cesuur van de vier examenvarianten

Uit de verschillende psychometrische analyses blijkt dat de door Cito berekende gemiddelde p-waardes optimaal zijn voor de gehanteerde itemtypes. Het door Cito berekende onderscheidend vermogen is eveneens goed. De door Cito berekende toetsbetrouwbaarheid van de verschillende varianten is goed, ondanks dat er een zekere mate van heterogeniteit in de onderliggende toetsdomeinen zit als gevolg van het combineren van lezen en kijk-luisteren in het examen. Het daarbij door RCEC geschatte percentage misclassificaties, berekend als de functie van het percentage kandidaten dat een onvoldoende heeft behaald op het centraal examen en de toetsbetrouwbaarheid, is passend. De door RCEC opgestelde toetsinformatiefunctie en standaardfout op basis van de door Cito gekalibreerde itemparameters bij de door Cito bepaalde cesuur laten zien dat de lokale toetsbetrouwbaarheid rond de cesuur hoog is. RCEC beschikte tijdens dit onderzoek niet over de ruwe data. Met deze ruwe data had RCEC de door Cito gepresenteerde itemparameters en toetsbetrouwbaarheid ter controle kunnen narekenen.

4.7. Standaardbepaling en normhandhaving van het centraal examen Engels B1 mbo 2022-2023

Bij de centrale examens Engels B1 in het mbo worden cijfers toegekend via 'normering met een vaardigheidsschaal'. Deze methode past zich aan als er verschillende examenvarianten zijn met lichte verschillen in moeilijkheid. Om eerlijk te beoordelen, moet een kandidaat op

makkelijkere varianten meer vragen correct beantwoorden voor een bepaald cijfer, terwijl de vereiste vaardigheid voor elk cijfer constant blijft, ongeacht de variant of het examenjaar. Dit geldt voor alle cijfers, waarbij een gelijke vaardigheid consistent beloond wordt met hetzelfde cijfer, ongeacht de examenvariant.

De conversie van vaardigheid naar cijfer verloopt derhalve lineair met een knik. Dit betekent dat er twee afzonderlijke, lineaire verbanden zijn tussen cijfers en vaardigheid, afhankelijk van het ERK-niveau. Voorbij het ERK-niveau B2 bestaat er een specifiek rechtlijnig verband tussen cijfers en vaardigheid, terwijl er onder dit niveau een ander rechtlijnig verband geldt. In de normeringsprocedure worden drie cruciale punten vastgelegd op de vaardigheidsschaal, die overeenkomen met ERK-niveaus. Bij de centrale examens Engels B1 en B2 zijn deze punten gerelateerd aan respectievelijk de ERK-niveaus B1, B2 en boven-B2. In het B1-Engels examen komen deze overeen met de cijfers 5,5, 7,5 en 9,5. Deze drie punten op de vaardigheidsschaal, die de conversie van vaardigheid naar cijfer beïnvloeden, worden standaarden genoemd. De rechte lijnen die door deze punten worden gedefinieerd, worden aan beide uiteinden afgekapt, omdat cijfers boven 10,0 en onder 1,0 niet worden toegekend. Voor de centrale examens mbo Engels zijn deze standaarden bepaald door de ERK-niveaus, zoals vastgesteld op de centrale examens vo Engels 2012, via een posttest met behulp van de Data-Driven Direct Consensus (3DC) standaardbepalingsmethode over te brengen.

De standaard (c.q. referentiecesuur) voor de centrale examens vo Engels 2012 is op zijn beurt door experts met vakdeskundigheid, verdeeld over panels, met behulp van de Extended Angoff methode bepaald. Daarmee wordt genoegzaam aangenomen dat de gehanteerde standaardbepalingsmethode van het centraal examen Engels B1 mbo 2022-2023 op de juiste wijze is uitgevoerd, dat de betrokken experts met vakdeskundigheid naar behoren zijn geselecteerd en getraind, en dat er voldoende overeenstemming is tussen de experts over de vastgestelde standaard.

Doordat in elke variant van het examen de opgaven uit de referentiesets (c.q. de zogenaamde ankers uit teksten die voor het eerst zijn ingezet in de periodes 2019-2020 en 2020-2021) samen met de reguliere opgaven worden afgenomen is het mogelijk om alle opgaven op één vaardigheidsschaal te plaatsen (door middel van IRT). De standaard op de referentieset wordt daarbij omgezet in een cesuur op de vaardigheidsschaal. Deze standaard wordt vervolgens gebruikt voor de normering van toetsen en examens door middel van de reguliere procedure 'normering met een vaardigheidsschaal'⁷. Bij het centraal examen Engels B1 mbo is er sprake van een eerste afnameperiode van één dag (16 september 2022), gevolgd door vier opeenvolgende afnameperiodes van elk zeven weken. Per afnameperiode worden meerdere basisvarianten van het examen geconstrueerd conform het generiek design uit bijlage 4 van de constructieopdracht. De verschillende teksten in de varianten hebben zoveel mogelijk

⁷ https://www.toetspecials.nl/html/normering_vaardigheidsschaal/vaardigheid-cijfer.shtml

dezelfde toetstechnische kenmerken. Daarnaast worden de tekstfragmenten getoetst op hun informatiewaarde, alvorens deze worden hergebruikt in een nieuwe examenvariant.

Voor de normhandhaving wordt het design van de basisvarianten uit de afnameperiodes 2 (17 oktober t/m 16 december 2022) en 4 (6 maart t/m 28 april 2023) verbonden. Hetzelfde procedé geldt voor de afnameperiodes 3 (9 januari t/m 23 februari 2023) en 5 (8 mei 2023 t/m 7 juli 2023). Na de eerste drie weken van afnameperiode 2 vindt een eerste normeringsronde plaats. Cito analyseert alle opgaven psychometrisch en selecteert de opgaven die minder goed functioneren. Deze betreffende opgaven worden voorgelegd aan de vaststellingscommissie. De vaststellingscommissie beoordeelt de mate waarin deze opgaven inhoudelijk juist en volledig en daarmee bruikbaar zijn. Op basis van de bevindingen van de vaststellingscommissie kan worden besloten om specifieke opgaven te neutraliseren en zo buiten de normering te houden. Hierna worden de examenresultaten uit week 3 en 4 van afnameperiode 2 verzameld, waarmee er voldoende data is om alle items te kalibreren met behulp van het item-responsmodel in Dexter. Derhalve is na de eerste vier weken van afnameperiode 2 de definitieve normering bepaald. Deze definitieve norm geldt eveneens voor de basisvarianten uit afnameperiode 4, welke inhoudelijk en toetstechnisch zijn gelinkt met de basisvarianten uit afnameperiode 2. Een en ander conform de toetsmatrijs (bijlage 1 uit de constructieopdracht) en conform het generiek design (bijlage 4 uit de constructieopdracht). De overlap in tekstfragmenten en bijbehorende items, zoals ook weergegeven in [Tabel 1](#), draagt eraan bij dat met behulp van het item-responsmodel in Dexter de scores van de verschillende basisvarianten op één en dezelfde schaal afgebeeld kunnen worden.

Daarmee worden aan kandidaten die het examen doen dezelfde eisen gesteld, onafhankelijk van de aangeboden basisvariant. Dit betekent dat de basisvarianten van het examen inhoudelijk gelijkwaardig zijn en van eenzelfde moeilijkheidsgraad zijn, wat tevens wordt bevestigd door het design van variant P2-05 van het centraal examen Engels B1 mbo 2022-2023, zoals gepresenteerd in [Tabel 1](#) en door de vergelijkbare gemiddelde p-waarde van de vier onderzochte varianten uit [Tabel 2](#). Daarmee wordt aangenomen dat de gekozen normhandhavingsmethode van voldoende kwaliteit is en correct is uitgevoerd.

5. Conclusies

5.1. Conclusies inhoudsvaliditeit

In de paragrafen 4.1 tot en met 4.5 zijn voor het centraal examen Engels B1 mbo 2022-2023 de vier aspecten van het onderdeel inhoudsvaliditeit van het onderzoekskader centrale examens mbo onderzocht. Op basis van de analyse van de vier aspecten kan geconcludeerd worden dat de procedures voor het realiseren van de inhoudsvaliditeit van het voornoemde examen voldoen aan de eisen die daaraan volgens nationale en internationale richtlijnen gesteld worden.

Voor wat betreft het onderzochte examen is er sprake van een domein definitie die genoegzaam als een volledige beschrijving van beide aspecten (zie paragraaf 4.2) aangemerkt kan worden. De gevolgde procedures resulteren, op basis van de onderzochte examenvariant P2-05, in voldoende informatie met betrekking tot de domein representatie. Het proces van totstandkoming van de examens is beschreven in paragraaf 3.1 en in de daarbij genoemde vaststellingsprocedures en de werkwijze vaststellingscommissie. Dit proces wordt als zorgvuldig gekwalificeerd. Het examen voldoet tevens aan de eisen die aan het aspect domein relevantie worden gesteld. De wijze van construeren van het examen volgt het tien stappenplan (Veldkamp, 2016) voor een adequate toetsconstructieprocedure.

5.2. Conclusies betrouwbaarheid

De betrouwbaarheidsanalyse is uitgevoerd op de varianten P2-05 en P2-08 uit afnameperiode 2 (17 oktober t/m 16 december 2022) en de varianten P4-19 en P4-22 uit afnameperiode 4 (6 maart t/m 28 april 2023).

Uit de analyse volgt dat de moeilijkheidsgraad van de vier varianten, uitgedrukt in de gemiddelde p-waarde, zich overal binnen de optimale waarde tussen 0,20 en 0,80 bevindt. Het gemiddeld onderscheidend vermogen van alle items per toetsvariant, uitgedrukt in de rir-waarde (c.q. item-rest correlatie), is te beoordelen als goed.

De toetsbetrouwbaarheid is gepresenteerd in een drietal coëfficiënten, de Cronbach's alpha, de lambda2 en de greatest lower bound (glb). De betrouwbaarheidscoëfficiënten bevinden zich overal boven de voor een summatieve toets vereiste ondergrens van 0,80. Het daarbij geschatte percentage niet-consistente beslissingen (de zogenaamde misclassificaties) is passend. De door RCEC opgestelde toetsinformatiefunctie en standaardfout op basis van de door Cito met het IRT model gekalibreerde itemparameters bij de door Cito bepaalde cesuur laten zien dat de lokale toetsbetrouwbaarheid rond de cesuur hoog is.

Daarmee voldoen de vier onderzochte examenvarianten aan de criteria voor betrouwbaarheid zoals gesteld in het RCEC Beoordelingssysteem voor de Kwaliteit van Toetsen en Examens (Sanders et al., 2018). Wel wordt opgemerkt dat RCEC tijdens dit onderzoek niet over de ruwe data beschikte. Met deze ruwe data had RCEC de door Cito gepresenteerde item- en toetsparameters ter controle kunnen narekenen.

5.3. Conclusies standaardbepaling en normhandhaving

Voor de centrale examens mbo Engels is de referentiecesuur bepaald door de ERK-niveaus, zoals vastgesteld op de centrale examens vo Engels 2012, via een posttest met behulp van de Data-Driven Direct Consensus (3DC) standaardbepalingsmethode over te brengen. De standaard (c.q. referentiecesuur) voor de centrale examens vo Engels 2012 is op zijn beurt door experts met vakdeskundigheid, verdeeld over panels, met behulp van de Extended Angoff methode bepaald. Daarmee wordt genoegzaam aangenomen dat de gehanteerde standaardbepalingsmethode van het centraal examen Engels B1 mbo 2022-2023 op de juiste wijze is uitgevoerd, dat de betrokken experts met vakdeskundigheid naar behoren zijn geselecteerd en getraind, en dat er voldoende overeenstemming is tussen de experts over de vastgestelde standaard.

Voor de normhandhaving worden de basisvarianten uit de afnameperiode 2 (17 oktober t/m 16 december 2022) inhoudelijk en toetstechnisch gelinkt met de basisvarianten uit de afnameperiode 4 (6 maart t/m 28 april 2023). Daarmee worden aan studenten die het examen doen dezelfde eisen gesteld, onafhankelijk van de aangeboden basisvariant. Dit betekent dat de basisvarianten van het examen inhoudelijk gelijkwaardig zijn en van eenzelfde moeilijkheidsgraad zijn, wat tevens wordt bevestigd door de het design van variant P2-05 van het centraal examen Engels B1 mbo 2022-2023, zoals gepresenteerd in [Tabel 1](#) en door de vergelijkbare gemiddelde p-waarde van de vier onderzochte varianten uit [Tabel 2](#). Daarmee wordt aangenomen dat de gekozen normhandhavingmethode van voldoende kwaliteit is en correct is uitgevoerd.

Geconcludeerd kan worden dat de procedures die het CvTE en Cito hanteren bij de ontwikkeling van de centrale examens Engels B1 mbo leiden tot examenproducten die voldoen aan (inter)nationale kwaliteitsstandaarden met betrekking tot inhoudsvaliditeit, betrouwbaarheid en normering.

6. Geraadpleegde literatuur

- AERA (2014). *Standards for educational and psychological testing. Joint Committee on Standards for Educational and Psychological Testing*. Washington D.C.: American Educational Research Association.
- Council of Europe (2008). *Gemeenschappelijk Europees Referentiekader voor Moderne Vreemde Talen: Leren, Onderwijzen, Beoordelen*. (Nederlandse Taalunie, vert.). Den Haag: Nederlandse Taalunie (Origineel werk gepubliceerd in 2001).
- Downing, S.M., & Haladyna, T.M. (2006). *Handbook of test development*. Lawrence Erlbaum Associates.
- Ebel, R.L., & Frisbie, D.A. (1986). *Essentials of educational measurement*. Englewood Cliffs: Prentice Hall.
- Eggen, T.J.H.M., & Sanders, P.F. (1993). *Psychometrie in de praktijk*. Arnhem: Cito.
- Sanders, P.F., Brouwer, A.J., Eggen, T.J.H.M., & Veldkamp, B.P. (2018). *RCEC beoordelingssysteem voor de kwaliteit van studietoetsen en (praktijk)examens*. Apeldoorn: RCEC.
- Sireci, S.G. (1998a). Gathering and analyzing content validity data. *Educational Assessment*, 5, 299-321.
- Sireci, S.G. (1998b). The construct of content validity. *Social indicators Research*, 45, 83 – 117.
- Sireci, S.G., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26, 100 – 107.
- Veldkamp B.P. (2016). De inhoud en constructie van toetsen. In Sanders, P.F. (Red.), *Toetsen op School: Hoger onderwijs* (pp. 21 – 30). Arnhem: Cito. Geraadpleegd 10 juni 2020 van: <https://www.rcec.nl/publicaties>.
- Wools, S. (2013). De validiteit van toetsscores In: Sanders, P.F. (Red.), *Toetsen op School* (pp. 69 – 83). Arnhem: Cito. Geraadpleegd 10 juni 2020 van: <https://www.rcec.nl/publicaties>.

Apeldoorn, december 2023

In opdracht van:



College voor Toetsen en Examens

RCEC • Research Center voor Examinering en Certificering
Villa de Regentes • Regentesselaan 5 • 7316 AA Apeldoorn
T 055 - 303 31 24 • I www.rcec.nl • E info@rcec.nl