



Regeling van het College voor Toetsen en Examens van 9 februari 2015, nummer CvTE-15.00618, 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 2015)

Het College voor Toetsen en Examens,

Gelet op artikel 2, tweede lid, aanhef en onderdeel e, en artikel 2a van de Wet College voor toetsen en examens;

Gezien de goedkeuring van de Staatsecretaris van Onderwijs, Cultuur en Wetenschap, gegeven op 8 april 2015, nummer 748426;

Besluit:

Artikel 1 centrale examens VO

De omzetting van scores in cijfers bij centrale examens VO, bedoeld in artikel 2, tweede lid, onderdeel e, van de Wet College voor toetsen en examens geschiedt op de in de bijlage 1 bij deze regeling vastgestelde wijze.

Artikel 2 rekentoets VO

De omzetting van scores in cijfers bij de rekentoets VO, bedoeld in artikel 2a, van de Wet College voor toetsen en examens geschiedt op de in de bijlage 2 bij deze regeling vastgestelde wijze.

Artikel 3 pilot-profiel-cspe's

De omzetting van scores in cijfers bij centrale examens VO, bedoeld in artikel 2, tweede lid, onderdeel e, van de Wet College voor toetsen en examens geschiedt voor de pilot-profiel-cspe's op de in de bijlage 3 bij deze regeling vastgestelde wijze.

Artikel 4 inwerkingtreding

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

Artikel 5 intrekking

De Regeling omzetting scores in cijfers centrale examens en rekentoets VO 2014 van 18 maart 2014, wordt ingetrokken.

Artikel 6 citeertitel

Deze regeling wordt aangehaald als: Regeling omzetting scores in cijfers centrale examens en rekentoets VO 2015.

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

*Het College voor Toetsen en Examens,
de voorzitter,
P.J.J. Hendrikse*



BIJLAGE 1 BEHORENDE BIJ ARTIKEL 1 VAN DE REGELING OMZETTING SCORES IN CIJFERS CENTRALE EXAMENS EN REKENTOETS VO 2015, VAN 9 FEBRUARI 2015, NUMMER CVTE-15.00618

De instructies die het CvTE toepast om tot de N-termen te komen

Uitgangspunt

De normering van de centrale examens dient er zorg voor te dragen dat in opeenvolgende jaren aan kandidaten bij het zelfde vak in hetzelfde schooltype/dezelfde leerweg dezelfde eisen worden gesteld. Bij een 5,5 in het ene jaar hoort dezelfde prestatie van de kandidaat als bij een 5,5 in een ander jaar. Omdat het technisch onmogelijk is onze examens van jaar op jaar 'precies even moeilijk' te maken, zorgt de keuze van de N-term daarvoor. Het proces van normering is erop gericht gegeven dit uitgangspunt de juiste normeringsterm te bepalen.

In *supplement 1* staan de vergelijkingen die bij de omzetting van score naar cijfer worden gehanteerd.

Examens in het eerste tijdvak

Het examen wordt geijkt aan het door het CvTE vastgestelde referentie-examen. Een referentie-examen is een examen dat door CvTE en veld wordt gezien als een 'goed' examen.

Allereerst gaat het CvTE uit van de veronderstelling dat kandidaten niet ineens vaardiger of minder vaardig zijn geworden.

Als uitsluitend deze regel zou worden toegepast, dan zouden bij een dalend vaardigheidsniveau de cijfers gemiddeld gelijk blijven, en zou een vaardiger groep leerlingen gestraft worden met een strengere normering. Daarom worden aanvullende gegevens gehanteerd.

Bij elk examen is het oordeel van het CvTE over de zwaarte van het examen in relatie tot het referentie-examen in aanmerking genomen. Dit oordeel wordt bij de vaststelling van het examen door de vakdeskundigen gegeven.

Bij een aantal examens verzamelt Cito gegevens over de moeilijkheidsgraad van het examen in relatie met het referentie-examen¹. Dat kan door middel van pretest, posttest, standard setting of een andere methode van equivalering.

Als uit de gegevens blijkt dat de groep kandidaten in vaardigheid verschilt van die bij het referentie-examen, dan wordt de normeringsterm zo gekozen dat de prestatie-eisen gelijk blijven, en gemiddelde en percentage onvoldoendes afwijken van die bij het referentie-examen.

Als de groep kandidaten gemiddeld even vaardig is als de groep die destijds het referentie-examen maakte, dan wordt de normeringsterm zo gekozen dat gemiddeld cijfer en percentage onvoldoendes zo dicht mogelijk liggen bij die van het referentie-examen.

Als een examen een onvolkomenheid bevat die kandidaten aanwijsbaar kan benadelen, krijgen correctoren nadere instructies door een aanvulling op het correctievoorschrift.

Als een onvolkomenheid pas wordt ontdekt bij de vaststelling van de normeringsterm, worden kandidaten voor de nadelige gevolgen van de onvolkomenheid gecompenseerd in de vastgestelde normeringsterm.

Bijzondere situaties bij het eerste tijdvak

- a. *Nieuw examenprogramma, met een pilot in het centraal examen*
Als de oude en nieuwe stof voor het centraal examen elkaar deels overlappen wordt op grond van de resultaten van de kandidaten op de vragen over de overlappende stof een oordeel over de vaardigheid van de kandidaten bepaald.
Als er geen overlap is, wordt de moeilijkheidsgraad cq vaardigheid bepaald via standard setting.
- b. *Uit pretest/posttest blijkt dat de vaardigheid van de kandidaten in de loop van een aantal jaren zo veranderd is dat handhaven van de eisen niet mogelijk of reëel is.*
Als dit zich voordoet, benadert CvTE de Minister met een analyse waarin wordt onderzocht of, en op welke wijze, de eisen aan de veranderde vaardigheid van de kandidaten kunnen worden aangepast, of dat wellicht maatregelen kunnen worden genomen (bijvoorbeeld in het onderwijs) waardoor de vaardigheid van de kandidaten weer in lijn is met de eisen.
- c. *Door aanpassing in de structuur van het onderwijs is de positie van een vak of meerdere vakken dusdanig gewijzigd dat het niet reëel of niet mogelijk is om de eisen te handhaven.*
CvTE neemt een besluit dat ter goedkeuring aan de Minister wordt voorgelegd.

¹ Bij een pre- of posttest is de vergelijking gebaseerd op resultaten van leerlingen die onderdelen uit beide examens hebben gemaakt.

Bij standardsetting is de vergelijking gebaseerd op een door een groep experts uitgevoerde beoordeling van de moeilijkheidsgraad van de afzonderlijke items uit de beide examens



Examens in het tweede tijdvak

In het tweede tijdvak geldt in beginsel de normeringsterm van het eerste tijdvak. Aan de hand van de statistisch te verwachten resultaatverbetering van kandidaten die in het eerste tijdvak een onvolgende hadden, wordt nagegaan of deze normeringsterm zou leiden tot hogere eisen aan de kandidaat dan in het eerste tijdvak. Als dat het geval is, wordt de normeringsterm zo vastgesteld dat de eisen gelijk zijn.

Als een onvolkomenheid wordt ontdekt bij de vaststelling van de normeringsterm, worden kandidaten voor de nadelige gevolgen van de onvolkomenheid gecompenseerd in de vastgestelde normeringsterm.

Examens in het derde tijdvak

Bij de examens in het derde tijdvak wordt de normeringsterm bepaald aan de hand van een gewogen oordeel over de moeilijkheidsgraad, samengesteld uit de oordelen van Cito, van het CvTE en van de correctoren van deze examens en worden de gangbare normeringstermen voor het vak in de loop van een aantal recente jaren in aanmerking genomen.

Supplement

De formules voor de omzetting van score naar cijfer

De huidige regels zijn in 1999 vastgesteld en voor het eerst toegepast bij de examens van 2000. [CEVO -99/648 van 22 juni 1999, Gele katern 1999, nr. 18a].

Uitgangspunten

Het systeem voor de omzetting van score naar cijfer is gebaseerd op de volgende vier uitgangspunten:

1. Elk gescoord punt draagt altijd bij tot een hoger examencijfer (afrondding daargelaten);
2. Een score van 0% correspondeert altijd met examencijfer 1,0;
3. Een score van 100% correspondeert altijd met examencijfer 10,0;
4. Over een zo breed mogelijk centraal interval van de scoreschaal is er (afrondding daargelaten) sprake van een evenredige stijging van score- en cijferpunten die onafhankelijk is van de normering.

Hierbij wordt onder de score verstaan: de zuivere score, dus uitsluitend de punten die aan de kandidaat zijn toegekend voor goede antwoordelementen.

Er zal derhalve geen sprake zijn van scorepunten-vooraf en/of scorepunten-bijtelling (in geval van cesuuraanpassing).

Het normeringsvoorschrift

Het normeringsvoorschrift bestaat uit twee onderdelen:

- de hoofdrelatie: de formule die, voor de overgrote meerderheid van de kandidaten, het berekeningsvoorschrift geeft voor het omzetten van score naar cijfer;
- vier grensrelaties: vier formules die (bij andere N-termen dan 1,0) voorkomen dat kandidaten met zeer lage of zeer hoge scores een cijfer zouden krijgen dat in strijd is met bovengenoemde vier uitgangspunten.

De hoofdrelatie

De hoofdrelatie geeft aldus het examencijfer als functie van de score:

$$C = 9,0 * (S/L) + N \dots\dots\dots (1)$$

waarin:

C = het cijfer voor het centraal examen.

S = de score, dat wil zeggen de zuivere aan de kandidaat toegekende score.

L = de lengte van de scoreschaal, zoals vastgelegd in het correctievoorschrift;

N = de normeringsterm, liggend tussen de waarden: $N = 0,0$ en $N = 2,0$, vast te stellen door het College voor Toetsen en Examens middels een normeringsbeslissing.

Zijn zowel L als N bekend, dan leidt invullen van de score S direct tot het examencijfer C.

Voorbeeld:

Stel de lengte van de scoreschaal is $L = 90$ punten;

dan gaat formule (1) over in:

$$C = 9,0 * (S/90) + N.$$

Voordat hiermee uit score S examencijfer C kan worden berekend, moet het College voor Toetsen en Examens eerst een waarde voor normeringsterm N hebben vastgesteld.

Stel dat wordt: $N = 1,0$; dan krijgt formule (1) zijn definitieve vorm:

$$C = 9,0 * (S/90) + 1,0.$$

Deze is gevisualiseerd in Fig.1:

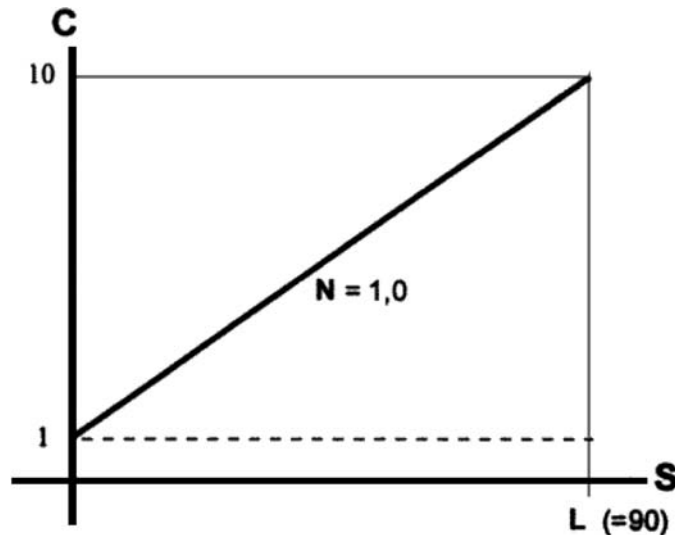


Fig.1

Concreet: drie kandidaten met scores van resp. 0%, 50% en 100% – wat bij deze L van 90 pt correspondeert met scores van 0, 45 en 90 pt – zouden achtereenvolgens de examencijfers: 1,0, 5,5 en 10,0 krijgen. Als was gekozen voor een andere schaallengte, bv $L = 68$, dan zou formule (1), – bij dezelfde N-term $N = 1,0$) overgaan in:

$$C = 9,0 * (S/68) + 1,0.$$

Nu zouden genoemde drie kandidaten voor dezelfde examencijfers (1,0, 5,5 en 10,0) respectievelijk de scores 0, 34 en 68 nodig hebben!

De grensrelaties

Deze zijn nodig om de boven gegeven vier uitgangspunten óók te kunnen eerbiedigen als de normeringsterm N groter of kleiner is dan 1,0.

Voorbeeld:

Bij een waarde voor de normeringsterm van $N = 1,3$, zouden de drie kandidaten met scores 0%, 50% en 100% op grond van de hoofdrelatie resp. de cijfers 1,3, 5,8 en 10,3 krijgen; daarvan is echter het eerste cijfer guller dan de bedoeling en is het derde cijfer hoger dan het toegestane maximum.

Iets dergelijks treedt op bij een normeringsterm lager dan 1,0, bijvoorbeeld: $N = 0,7$. Genoemde drie kandidaten zouden in dat geval de examencijfers 0,7, 5,2 en 9,7 krijgen, waarvan het eerste cijfer uitkomt onder het toegestane minimum en het derde cijfer lager is dan de verdiende 10,0!

Deze problematiek is in beeld gebracht in Fig.2:

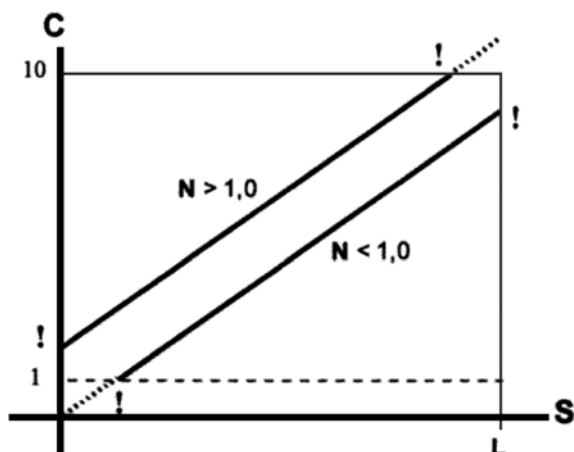


Fig.2

Deze ‘bijzonderheden’ worden verholpen door middel van een systeem van zogeheten grensrelaties.

Het principe van grensrelaties is gevisualiseerd in Fig.3. Bij voorbaat zullen alle score-cijfercombinaties liggen binnen het gebied dat begrensd wordt door de vier lijnen in deze figuur.

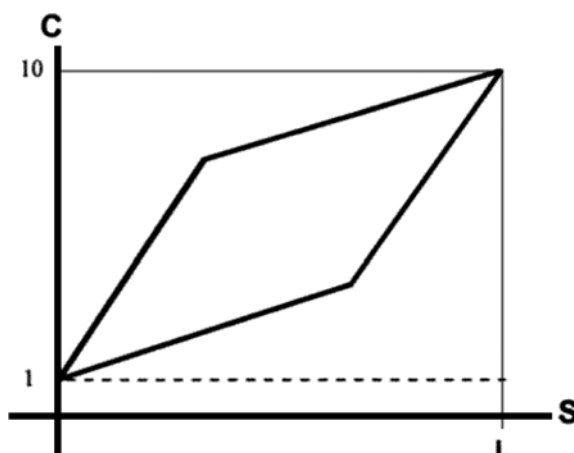


Fig. 3

Samen vormen deze vier lijnstukken een “venster” waarbinnen alle toegestane score-cijfercombinaties moeten liggen. Dreigt bij toepassing van de hoofdrelatie – formule (1) – een score-cijfercombinatie buiten deze grenzen te vallen, dan moet voor de desbetreffende score dat cijfer vervangen worden door het cijfer berekend met de corresponderende grensrelatie. Wat informeler gezegd: score-cijfercombinaties die buiten het “venster” dreigen te vallen, komen op het “kozijn” terecht. De grensrelaties worden gevormd door de volgende vier formules:

bij $N > 1,0$ geldt voor de laagste scores de formule:

$$C \leq 1,0 + S * (9/L) * 2 \dots\dots\dots (2a)$$

En voor de hoogste scores

$$C \leq 10,0 - (L-S) * 9/L * 0,5 \dots\dots\dots (2b)$$

bij $N < 1,0$ geldt voor de laagste scores de formule:

$$C \geq 1,0 + S * (9/L) * 0,5 \dots\dots\dots (3a)$$

en voor de hoogste scores

$$C \geq 10,0 - (L-S) * (9/L) * 2 \dots\dots\dots (3b)$$

Bij een waarde voor de normeringsterm van $N = 1,0$ treedt het systeem van grensrelaties niet in werking en resulteert een score-cijfertransformatie die grafisch wordt gerepresenteerd door de rechte lijn van Fig.1, de lijn die in Fig.4 is gelabeld met: “ $N=1,0$ ”.

Bij alle andere waarden van N zijn de grensrelaties wel van belang. In Fig. 4 zijn als voorbeelden de twee uiterste gevallen in beeld gebracht, die resp. corresponderen met de normeringsbeslissingen $N =$



2,0 en $N = 0,0$. Deze leveren als score-cijfertransformaties de twee dubbel-geknikte lijnen op (gelabeld met "N=2,0" en "N=0,0").

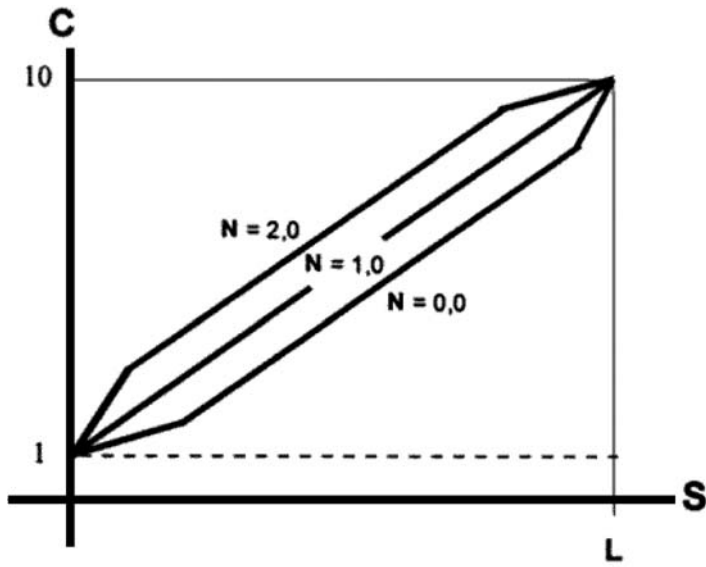


Fig. 4

BIJLAGE 2 BEHORENDE BIJ ARTIKEL 2 VAN DE REGELING OMZETTING SCORES IN CIJFERS CENTRALE EXAMENS EN REKENTOETS VO 2015, VAN 9 FEBRUARI 2015, NUMMER CVTE-15.00618

Normering met een vaardigheidsschaal bij de centrale examens Nederlandse taal en rekenen in het mbo en de Rekentoets VO (RVO)

Doel van normering met een vaardigheidsschaal

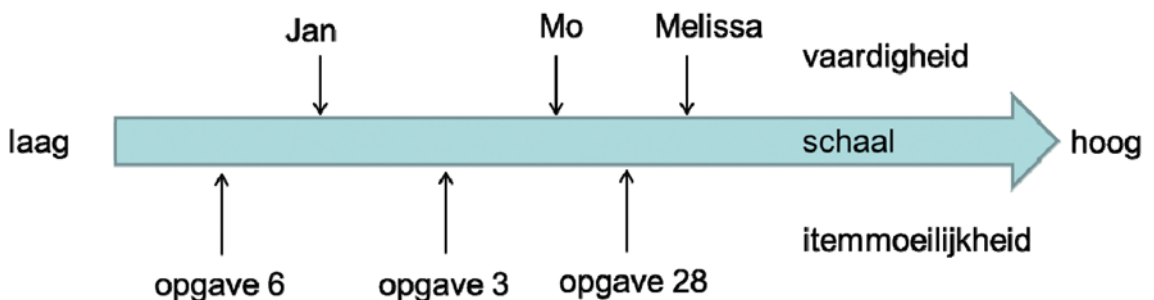
Zoals gebruikelijk bij toetsen en examens, worden bij de rekentoetsen VO en de centrale examens Nederlandse taal en rekenen in het MBO cijfers toegekend. De manier waarop deze cijfers bepaald worden, noemen we 'normering met een vaardigheidsschaal'.

Het gebruik van een vaardigheidsschaal is vooral geschikt als er meerdere varianten van een examen zijn. De ene variant kan net iets makkelijkere opgaven hebben dan de andere variant. Dit kan gebeuren omdat de moeilijkheid van een opgave niet heel precies ingeschat kan worden voordat deze is voorgelegd aan examenkandidaten. Bij het toekennen van cijfers moet daar rekening mee gehouden worden. Op de makkelijkere varianten moet een kandidaat dan iets meer vragen goed beantwoord hebben om een 6 te krijgen. We kunnen ook zeggen: een kandidaat moet eenzelfde vaardigheid aantonen voor een 6, ongeacht de variant die hij maakt, ongeacht het jaar waarin het examen wordt afgelegd dan wel de examenperiode binnen dat jaar. En dit moet ook gelden voor ieder ander cijfer. Ongeacht de variant die gemaakt wordt, moet eenzelfde aangetoonde vaardigheid steeds beloond worden met eenzelfde cijfer.

Metten van vaardigheid

Het cijfer dat toegekend wordt, wordt dus bepaald door de vaardigheid die een kandidaat heeft. Deze vaardigheid kan hij aantonen door opgaven van een examen goed te maken. Het examen is daarmee een meetinstrument geworden van de vaardigheid van de kandidaat. En net zoals een thermometer een meetinstrument is van de temperatuur, op de Celsius- of Fahrenheit-schaal, zo spreken we ook van een vaardigheidsschaal.

De lengte van mensen meten we normaal gesproken met een meetlat. Stel dat die niet uitgevonden was, dan zouden we de hoogte van muurtjes kunnen gebruiken om de lengte van mensen te meten: wie erover heen kijkt, is langer dan degene die dat niet kan. Op analoge wijze kunnen we opgaven gebruiken om de vaardigheid van kandidaten te meten: wie de opgave goed maakt, is vaardiger dan degene die dat niet kan. In Figuur 1 is een vaardigheidsschaal afgebeeld. De vaardigheden van de kandidaten en de moeilijkheden van de opgaven zijn streepjes op eenzelfde meetlat ofwel schaal. Kandidaten met een hoge vaardigheid hebben een grotere kans om opgaven goed te maken dan kandidaten met een lagere vaardigheid. Omgekeerd, opgaven met een hoge moeilijkheid worden minder vaak goed gemaakt dan opgaven met een lage moeilijkheid.



Figuur 1. Afbeelding van personen en opgaven op één vaardigheidsschaal

Een opgave heeft altijd dezelfde moeilijkheid; een examenvariant kan bestaan uit makkelijkere of moeilijker opgaven. Een mini-examenvariant met alleen opgave 6 en opgave 3 uit bovenstaand voorbeeld, is bijvoorbeeld gemakkelijker dan een mini-examenvariant met alleen opgave 3 en opgave 28. Uit het antwoordgedrag van voldoende leerlingen op de makkelijke variant, kunnen we afleiden hoeveel makkelijker opgave 6 is dan opgave 3. Uit het antwoordgedrag van andere leerlingen op de moeilijke variant, kunnen we afleiden hoeveel moeilijker opgave 28 is dan opgave 3. Door een koppeling van deze twee verschillen, kunnen we afleiden hoeveel moeilijker opgave 28 is dan opgave 6. Op deze manier kunnen de moeilijkheden van alle opgaven op dezelfde schaal worden afgebeeld. Meer details over de schatting van itemmoeilijkheden staan in de sectie Schattingsmethode.



Getallen bij de vaardigheidsschaal

De schattingen van punten op de vaardigheidsschaal liggen vast op een lineaire transformatie na. We mogen bij alle punten dus een willekeurig aantal optellen, of alle punten met een willekeurige constante vermenigvuldigen. Een vaardigheidsschaal wordt geschat om scores op verschillende examenvarianten met elkaar te vergelijken. De precieze getallen die eraan hangen, zijn in wezen keuzes. Het nulpunt komt vrij willekeurig ergens te liggen. Negatieve vaardigheidsscores zijn lager dan vaardigheidsscore 0. Dit heeft geen inhoudelijke betekenis. We hadden bij alle scores ook 100 punten op kunnen tellen. In technische termen: de vaardigheidsschaal is van interval-niveau. Vergelijk met een temperatuur-schaal: het nulpunt op de Fahrenheit-schaal ligt ergens anders dan op de Celsius-schaal, maar voor beide geldt: hoe hoger, hoe warmer.

Je kunt bij temperatuur niet spreken van 'het is vandaag twee keer zo warm als gisteren', want voor zo'n uitspraak is een absoluut nulpunt nodig. Dat is er op de Fahrenheit- en Celsius-schaal niet. Zo kun je dus ook niet zeggen dat een kandidaat twee keer zo vaardig is als een andere kandidaat. Er is wel een inhoudelijke betekenis van negatieve graden op de Celsius-schaal (het vriest), maar op de Fahrenheit-schaal vriest het al onder 32 graden en is deze inhoudelijke betekenis van negatieve temperaturen er dus niet. Zo'n inhoudelijke betekenis van het nulpunt is er bij de vaardigheidsschaal ook niet.

Om nog technischer te worden: de ratio van twee verschillen is bij een interval-schaal wel interpreteerbaar. Dus als Jan, Mo, Mieke en Melissa respectievelijk 80, 90, 100 en 120 als vaardigheidsscores hebben, dan kun je wel zeggen dat het verschil in vaardigheid tussen de meisjes Mieke en Melissa (20 punten) twee keer zo groot is als het verschil tussen de jongens Jan en Mo (10 punten).

Van score naar vaardigheid

Als de moeilijkheden van alle opgaven of items in een examenperiode geschat zijn, wordt een beste schatting van de vaardigheid van een kandidaat gemaakt² aan de hand van de items die de kandidaat daadwerkelijk zijn voorgelegd. Van alle antwoorden die een kandidaat geeft, wordt eerst de ruwe score berekend, ofwel het aantal behaalde punten. In combinatie met de itemkenmerken waarop deze ruwe score behaald is, wordt de vaardigheid geschat. Dit gebeurt per examenvariant voor iedere ruwe score apart, zie bijvoorbeeld tabel 1 voor een gedeelte van de omzettingstabel van score naar vaardigheid bij variant 2 van een toets of examen. De relatie tussen score en vaardigheid hoeft niet rechtlijnig te zijn. Dit hangt af van de onderlinge verschillen in moeilijkheid tussen de opgaven.

Tabel 1. Voorbeeld van een omzettingstabel score naar vaardigheid

Variant	Score	Vaardigheid
...
V2	30	66.5
V2	31	70.0
V2	32	73.7
V2	33	77.8
V2	34	82.3
V2	35	87.3
V2	36	93.0
...

Deze methode levert dus per examenvariant een tabel op met achter iedere mogelijke ruwe score een vaardigheidsschatting. De ruwe score loopt van 0 tot en met de maximale score. De vaardigheidsschatting heeft niet overal dezelfde nauwkeurigheid. Bij de allerlaagste en allerhoogste scores is de meetnauwkeurigheid lager dan ergens in het midden van de scores. De exacte score waarbij de meetnauwkeurigheid het hoogst is, hangt af van de moeilijkheid van de opgaven in de examenvariant.

Als er opgaven in een examenvariant zitten, waarover de normeringsvergadering beslist dat zij niet mee mogen tellen bij de beoordeling van een kandidaat, dan noemen we dit geneutraliseerde items. Alle kandidaten krijgen voor een geneutraliseerd item het maximale aantal punten dat behaald kan worden. Bij de schatting van de itemparameters doen de geneutraliseerde items niet mee, alleen de antwoorden op de overige items worden gebruikt voor het vaststellen van de vaardigheidsschaal. Bij de schatting worden de geneutraliseerde items buiten beschouwing gelaten.

² De gebruikte schatting is een weighted maximum likelihood (WML) schatter bij de ongewogen score. Verhelst en Engelen (1999) tonen aan dat dit een veralgemenisering is van de WML-schatter bij gewogen scores (Warm, 1989).



De vaardigheid van de kandidaat wordt dus eigenlijk geschat aan de hand van de antwoorden op een verkorte examenvariant, waarin de geneutraliseerde items niet opgenomen zijn. In tabel 2 staat een voorbeeld van de omzettingstabel van score naar vaardigheid als er in een variant, waarop maximaal 54 punten behaald kunnen worden, drie geneutraliseerde items van ieder maximaal 1 punt zijn. De schatting van de vaardigheid gaat uitsluitend over de verkorte variant, met scores 0 tot en met 51. De rapportage gaat over de gehele range 0 tot en met 54. De 'neutrale punten' worden bij de verkorte score opgeteld om tot de score op de volledige variant te komen. In de rapportage-tabel staat achter de scores 0, 1 en 2 dezelfde vaardigheid als bij de laagst mogelijke score 3, namelijk vaardigheid 75. Echter, in praktijk komen deze scores niet voor omdat iedere kandidaat minimaal 3 punten scoort, namelijk op de geneutraliseerde items.

Tabel 2. Vaardigheidsscores bij een examenvariant met 3 geneutraliseerde items van ieder 1 punt

Schatting		Rapportage	
Score verkorte variant	Vaardigheid	Score volledige variant	Vaardigheid
Nvt	nvt	0	75
Nvt	nvt	1	75
Nvt	nvt	2	75
0	75	3	75
1	83	4	83
2	86	5	86
..		..	
51	212	54	212

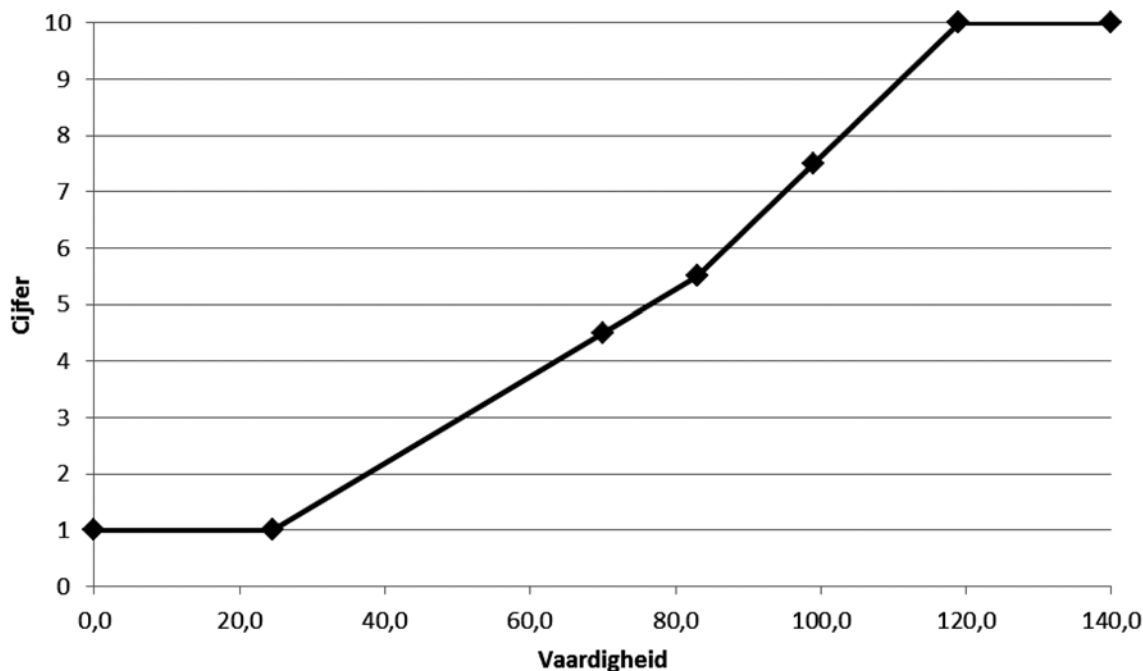
Noot. Gerapporteerde scores 0,1 en 2 komen niet voor, vanwege de geneutraliseerde items.

Van vaardigheid naar cijfer

Door middel van item respons theorie-schaling worden alle kandidaten en opgaven op dezelfde vaardigheidsschaal afgebeeld. Cijfers reflecteren de waardering voor behaalde vaardigheid. Het belangrijkste punt op de vaardigheidsschaal is de cesuur: vanaf de cesuur-vaardigheid wordt de vaardigheid van kandidaten als voldoende beoordeeld. Een vaardigheid onder de cesuur wordt als onvoldoende beoordeeld.

Om afrondingsproblemen te voorkomen, is het precieze cijfer dat wordt toegekend bij de cesuur-vaardigheid afhankelijk van het aantal gerapporteerde decimalen. Als hele cijfers gerapporteerd worden, zoals bij de Rekentoets VO of het centraal examen Rekenen, is het cijfer bij de cesuur een 5,5. Als cijfers met één decimaal worden gerapporteerd, zoals bij het centraal examen Taal, is het cijfer bij de cesuur gelijk aan 5,45. In Figuur 2 is dit bij vaardigheid 83,0.

Van vaardigheid naar cijfer



Figuur 2. Voorbeeld van omzetting van vaardigheidsscores naar cijfers

De omzetting van vaardigheid naar cijfer is lineair met een knik. Dat wil zeggen dat er een rechtlijnig verband is tussen cijfer en vaardigheid boven de cesuur, en dat er een ander rechtlijnig verband is onder de cesuur. Om de relaties exact vast te leggen, worden door de normeringsvergadering nog twee cijferpunten op de vaardigheidsschaal vastgelegd: één boven de cesuur en één onder de cesuur. Boven de cesuur is dit punt het cijfer 7,5, waarmee het cijfer 8 of hoger wordt toegekend aan kandidaten met een goede vaardigheid. Onder de cesuur wordt het cijfer 4,5 of 3,5 vastgesteld.

De drie punten op de vaardigheidsschaal die de omzetting van vaardigheid naar cijfer bepalen, noemen we ook wel standaarden. De rechte lijnen die door de drie punten bepaald worden, worden naar boven en beneden afgekapt. Cijfer boven 10,0 en onder 1,0 worden immers niet toegekend. Het bepalen van de waarden van standaarden wordt Standaardsetting genoemd. Meestal worden daarvoor de oordelen van experts gebruikt. Er zijn diverse methodes waarbij deze experts oordelen per opgave moeten geven, of waarbij zij oordelen over sets van opgaven moeten geven.

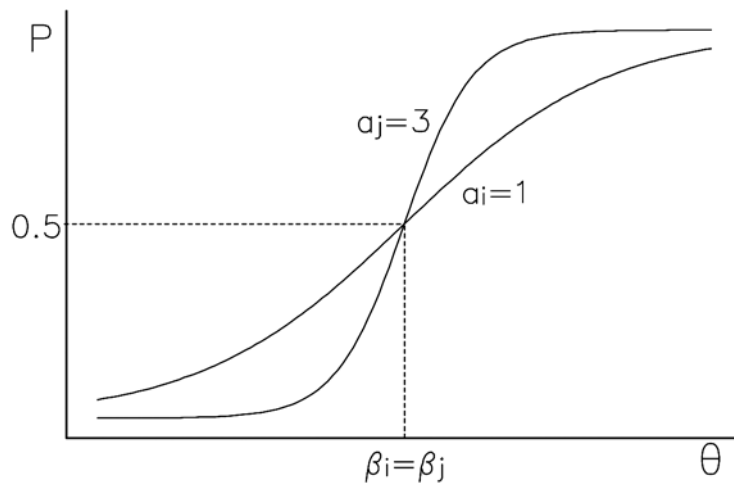
Schattingsmethode

Om de moeilijkheid van de opgaven precies te schatten, wordt gebruik gemaakt van item respons theorie (IRT). Opgaven worden binnen deze theorie items genoemd. De antwoorden van leerlingen worden responsen genoemd. De kans dat een bepaalde kandidaat een bepaald item goed beantwoordt, is afhankelijk van de vaardigheid van de persoon en van kenmerken van het item, zoals de moeilijkheid ervan. De vaardigheid van een kandidaat wordt met de Griekse letter θ aangeduid. De kans dat een kandidaat met een vaardigheid θ een item goed maakt, wordt omschreven met een wiskundige formule of functie. Van de familie van modellen die binnen de IRT vallen, wordt in dit geval het one parameter logistic model (OPLM, Glas & Verhelst, 1989, Verhelst & Glas, 1993; Verhelst, Glas & Verstralen, 1993; Verhelst & Eggen, 2011) gebruikt. De itemresponsfunctie van het OPLM is gegeven door

$$f_i(\theta) = \frac{\exp [a_i(\theta - \beta_i)]}{1 + \exp [a_i(\theta - \beta_i)]},$$

waarin a_i de zogenaamde discriminatie-index van het item is, en β_i de moeilijkheidsparameter van item i is. In Figuur 3 zijn de itemresponscurven weergegeven van twee items i en j , die even moeilijk zijn maar verschillend discrimineren. Als de vaardigheid van de kandidaat gelijk is aan de moeilijkheid van een opgave, dan heeft hij een kans van 50% om de opgave goed te maken. Opgaven met een hoge discriminatie-index onderscheiden beter tussen hoog- en laagvaardige kandidaten, ofwel de

kans om een opgave goed te maken, neemt hierbij snel toe met θ . De index a_i wordt ook wel de hellingsparameter genoemd.



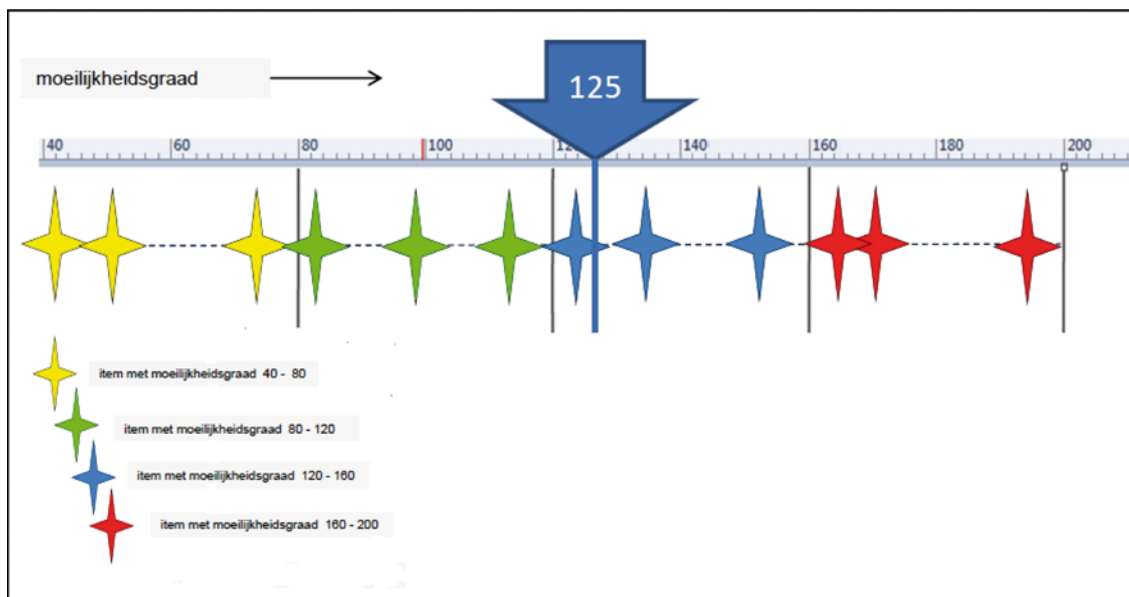
Figuur 3. Twee itemresponscurven in het OPLM: zelfde moeilijkheid, verschillende discriminatie-index

Een vaak toegepaste schattingsmethode voor de moeilijkheidsparameters β_i is de 'conditionele grootste aannemelijkheidsmethode' (in het Engels: Conditional Maximum Likelihood, verder aangeduid als CML). Die maakt gebruik van het feit dat in het model een afdoende steekproefgrootte ('sufficient statistic') bestaat voor de latente variabele θ , namelijk de ruwe score of het aantal correct beantwoorde items. Dat betekent grofweg dat, indien de itemparameters bekend zijn, alle informatie die het antwoordpatroon over de vaardigheid bevat, kan worden samengevat in de ruwe score; het doet er dan verder niet meer toe welke opgaven goed en welke fout zijn gemaakt. Hieruit vloeit voort dat de conditionele kans op een juist antwoord op item i , gegeven de ruwe score, een functie is die alleen afhankelijk is van de itemparameters en onafhankelijk van de waarde van θ^3 . De CML-schattingsmethode maakt van deze functie gebruik. Deze methode maakt geen enkele vooronderstelling over de verdeling van de vaardigheid in de populatie, en is ook onafhankelijk van de wijze waarop de steekproef is getrokken. Door de indices a_i te beperken tot (positieve) gehele getallen, en door ze a priori als constanten in te voeren, is het mogelijk CML-schattingen van de itemparameters β_i te maken.

Standaardsetting

Bij een standaardsetting wordt een punt op de vaardigheidsschaal bepaald (figuur 4). Uitgangspunt voor de standaardbepaling was de gehanteerde procedure en de bijbehorende uitkomsten van het standaardsettingsonderzoek bij de referentiesets. Bij de referentiesets is gebruik gemaakt van standaardsetting-procedures met beoordelaars, in casu de **Angoff-procedure** en de **Bookmark-procedure**. Merk op dat deze methoden niet beperkt hoeven te zijn tot de cesuur voor een voldoende, maar ook toegepast kunnen worden voor het cijfer 5 of 8.

³ Een gedetailleerde uiteenzetting hierover kan men vinden in Verhelst, 1992.



Figuur 4. Standaard (125) op een vaardigheidsschaal

Angoff-procedure

Bij een Angoff-procedure moeten de experts een grenskandidaat in gedachten nemen. Een grenskandidaat is een kandidaat die het vereiste niveau net beheerst. De experts geven per item aan of een grenskandidaat deze goed maakt of niet. Bij een gemodificeerde Angoff-procedure wordt per item door de experts de kans ingeschat dat een grenskandidaat de opgave goed maakt. Dit is hetzelfde als inschatten hoeveel van 100 grenskandidaten de opgave goed maken. Optellen van de kansen van alle opgaven van een bepaald examen of toets geeft de grenscore die gehaald moet worden op het betreffende examen of toets om te voldoen aan minimale eisen: de cesuur. Deze cesuur wordt vervolgens afgebeeld op de vaardigheidsschaal, waarmee de standaard bepaald is.

Bookmark-procedure

Als er ook afnamegegevens beschikbaar zijn, kan ook gewerkt worden met een Bookmark-procedure, waarbij gebruik gemaakt wordt van de empirische ordening van de items naar moeilijkheid (Van der Schoot, 2001, 2008). Deze procedure voor het vaststellen van een standaard voor een bepaald niveau maakt gebruik van een reeks opgaven die naar moeilijkheid zijn gerangschikt. Deskundigen geven aan welke opgave nog wel en welke niet meer beheerst zouden moeten worden door een grenskandidaat. In de reeks opgaven plaatsen zij een 'bookmark'. Experts geven dus aan hoeveel items van een set opgaven beheerst moeten worden op een bepaald niveau. Dit oordeel wordt weer omgezet in een punt op de vaardigheidsschaal.

Hieronder wordt voor rekenen en Nederlandse taal beschreven hoe de standaardsettingsprocedure is verlopen.

Procedure rekenen

Gedurende de standaardbepalingsprocedure van de referentiesets rekenen is gebruik gemaakt van de Extended Angoff methode. Voor deze procedure zijn de panelleden in drie groepen verdeeld. Bij de groepsindeling is rekening gehouden met de achtergrond van panelleden. Elke groep heeft een gedeelte van het materiaal beoordeeld. Na een individuele beoordelingsronde (ronde 1) hebben de panelleden in kleine groepen de opgaven besproken. Tijdens deze discussieronde was het mogelijk om het oordeel dat in de individuele ronde gegeven was aan te passen (ronde 2). De discussieronde is zodanig georganiseerd dat de opgaven in kleine groepen van wisselende samenstelling zijn besproken. Aangezien slechts een gedeelte van de opgaven door een panellid beoordeeld is, is door middel van een omzettingstabel (gebaseerd op een vaardigheidsschaal) de grenscore op het beoordeelde gedeelte geëxtrapoleerd naar de volledige referentieset.

Naast de Extended Angoff procedure is ook een Bookmarkprocedure uitgevoerd. Voor deze procedure



is op basis van de verzamelde data de moeilijkheid van opgaven bepaald⁴. De opgaven zijn vervolgens geordend van makkelijk naar moeilijk. Panelleden konden aangeven tot en met welke opgave een grenskandidaat een kans van 67% of minder had om de opgave goed te maken.

De uitkomsten van beide procedures zijn in een plenaire discussie voorgelegd aan de panelleden. Hierbij werd weergegeven wat het percentage leerlingen zou zijn dat het referentieniveau zou halen. Daarnaast is deze voorlopige cesuur vergeleken met reeds vastgestelde cesuren, namelijk de tot nu toe gehanteerde rekenen 2F en 3F standaarden voor de centrale examens en de Rekentoets VO. Vervolgens hebben alle experts op schrift een definitief oordeel gegeven. Het gemiddelde van deze oordelen geldt vervolgens als uiteindelijk advies aan het College voor Toetsen en Examens.

Procedure Nederlandse taal

Gedurende de standaardbepalingsprocedure is gebruik gemaakt van de Extended Angoff methode. Voor deze procedure zijn de panelleden in vier groepen verdeeld. Elke groep heeft een gedeelte van het materiaal beoordeeld. Na een individuele beoordelingsronde (ronde 1) hebben de panelleden in kleine groepen de opgaven besproken. Tijdens deze discussieronde was het mogelijk om het oordeel dat in de individuele ronde gegeven was aan te passen (ronde 2).

Na afloop van deze discussieronde konden experts hun oordelen intekenen in een figuur. In deze figuur werden de opgaven die een expert beoordeeld had afgebeeld op de volledige schaal van de referentieset. Door zorgvuldig de oordelen uit de inhoudelijke rondes in de figuur in te tekenen kon het oordeel op een gedeelte van de opgaven daarmee geëxtrapoleerd worden naar de volledige referentieset. Zo konden experts tot een eerste advies over de cesuur op de volledige referentieset komen.

Naar aanleiding van dit eerste advies is nagegaan wat de gemiddelde cesuur van alle panelleden op de referentieset zou zijn. Voor deze cesuur is bekeken wat het percentage kandidaten zou zijn dat het referentieniveau zou halen. Daarnaast is deze voorlopige cesuur vergeleken met reeds vastgestelde cesuren, namelijk de CENTRALE EXAMENS Taal 2F en 3F-cesuur. Vervolgens hebben alle experts op schrift een definitief oordeel gegeven. Het gemiddelde van deze oordelen geldt vervolgens als uiteindelijk advies aan het College voor Toetsen en Examens. Het College heeft de adviezen overgenomen en vastgesteld.

Omzetting referentiecesuur naar cesuur centraal examen taal en rekenen en rekentoets VO-cesuur

Door de items uit de referentieset rekenen en die uit de varianten van de centrale examens rekenen en de rekentoets VO af te beelden op dezelfde vaardigheidsschaal (zie figuur 1 en de beschrijving in paragraaf 4) wordt de referentiecesuur (= de cesuur van de referentieset) overgebracht op iedere variant van het centraal examen rekenen en de rekentoets VO. Geheel analoog wordt de referentiecesuur Nederlandse taal op de varianten van de centrale examens Nederlandse taal overgebracht.

Omzetting van scores in cijfers

Van iedere toets- en examenvariant wordt de omzetting van scores in cijfers bepaald door de scores om te zetten in vaardigheidsniveaus op de wijze waarop dat in paragraaf 4 is beschreven en door de vaardigheidsniveaus om te zetten in cijfers volgens paragraaf 5. De omzetting van scores verloopt daardoor via een vaardigheidsschaal.

Tabel 3a is een voorbeeld van een omzettingstabel van scores -via vaardigheidsniveaus – in cijfers.

Mogelijkheid van maatwerk bij de normering

Het is mogelijk om bij de normering met een andere cesuur dan de referentiecesuur te werken. In 2013 is dit voor het eerst gebeurd bij de normering van de rekentoets 2F. Toen is de cesuur voor de kandidaten in de gemengde en theoretische leerweg één cijferpunt hoger gelegd dan de referentiecesuur, terwijl de cesuur voor de kandidaten in de basisberoepsgerichte leerweg één cijferpunt lager lag. Tabel 3b respectievelijk 3c is een voorbeeld van een omzettingstabel waarbij de cesuur één cijferpunt lager respectievelijk hoger ligt dan bij de omzetting van scores in cijfers volgens tabel 3a.

⁴ Moeilijkheid is hier uitgedrukt in p-waarde. Hierbij is een gewogen p-waarde gehanteerd waarbij alle beschikbare data van alle schooltypen meegenomen is.

tabel 3a omzettingstabel (voorbeeld waarbij de cesuur ligt bij een fictief vaardigheidsniveau van 105,7)

score	vaardig-	cijfer
0	-16,0	1
1	23,1	1
2	38,8	1
3	48,3	1
4	55,0	1
5	60,1	1
6	64,2	1
7	67,7	2
8	70,7	2
9	73,4	2
10	75,8	3
11	78,0	3
12	80,0	3
13	81,9	3
14	83,6	3
15	85,3	4
16	86,8	4
17	88,3	4
18	89,7	4
19	91,1	4
20	92,4	4
21	93,6	4
22	94,8	4
23	96,0	5
24	97,1	5
25	98,3	5
26	99,4	5
27	100,4	5
28	101,5	5
29	102,5	5
30	103,6	5
31	104,6	5
32	105,6	5
33	106,6	6
34	107,6	6
35	108,7	6
36	109,7	6
37	110,7	6
38	111,7	6
39	112,8	6
40	113,9	7
41	115,0	7
42	116,1	7
43	117,2	7
44	118,4	7
45	119,6	7
46	120,9	7
47	122,3	8
48	123,7	8
49	125,2	8
50	126,8	8
51	128,5	8
52	130,5	9
53	132,6	9
54	135,0	9
55	137,9	9
56	141,4	10
57	145,8	10
58	151,9	10
59	161,4	10
60	182,9	10

tabel 3a

tabel 3b omzettingstabel (voorbeeld waarbij de cesuur één cijferpunt lager ligt dan in tabel 3a)

score	vaardig-	cijfer
0	-16,0	1
1	23,1	1
2	38,8	1
3	48,3	1
4	55,0	2
5	60,1	2
6	64,2	2
7	67,7	3
8	70,7	3
9	73,4	3
10	75,8	4
11	78,0	4
12	80,0	4
13	81,9	4
14	83,6	4
15	85,3	5
16	86,8	5
17	88,3	5
18	89,7	5
19	91,1	5
20	92,4	5
21	93,6	5
22	94,8	5
23	96,0	6
24	97,1	6
25	98,3	6
26	99,4	6
27	100,4	6
28	101,5	6
29	102,5	6
30	103,6	6
31	104,6	6
32	105,6	6
33	106,6	7
34	107,6	7
35	108,7	7
36	109,7	7
37	110,7	7
38	111,7	7
39	112,8	7
40	113,9	8
41	115,0	8
42	116,1	8
43	117,2	8
44	118,4	8
45	119,6	8
46	120,9	8
47	122,3	9
48	123,7	9
49	125,2	9
50	126,8	9
51	128,5	9
52	130,5	10
53	132,6	10
54	135,0	10
55	137,9	10
56	141,4	10
57	145,8	10
58	151,9	10
59	161,4	10
60	182,9	10

tabel 3b

tabel 3c omzettingstabel (voorbeeld waarbij de cesuur één cijferpunt hoger ligt dan in tabel 3a)

score	vaardig-	cijfer
0	-16,0	1
1	23,1	1
2	38,8	1
3	48,3	1
4	55,0	1
5	60,1	1
6	64,2	1
7	67,7	1
8	70,7	1
9	73,4	1
10	75,8	2
11	78,0	2
12	80,0	2
13	81,9	2
14	83,6	2
15	85,3	3
16	86,8	3
17	88,3	3
18	89,7	3
19	91,1	3
20	92,4	3
21	93,6	3
22	94,8	3
23	96,0	4
24	97,1	4
25	98,3	4
26	99,4	4
27	100,4	4
28	101,5	4
29	102,5	4
30	103,6	4
31	104,6	4
32	105,6	4
33	106,6	5
34	107,6	5
35	108,7	5
36	109,7	5
37	110,7	5
38	111,7	5
39	112,8	5
40	113,9	6
41	115,0	6
42	116,1	6
43	117,2	6
44	118,4	6
45	119,6	6
46	120,9	6
47	122,3	7
48	123,7	7
49	125,2	7
50	126,8	7
51	128,5	7
52	130,5	8
53	132,6	8
54	135,0	8
55	137,9	8
56	141,4	9
57	145,8	9
58	151,9	10
59	161,4	10
60	182,9	10

tabel 3c

Literatuur

Engen, T. J. H. M., & Verhelst, N. D. (2011). Item calibration in incomplete testing designs. *Psicologica: International Journal of Methodology and Experimental Psychology*, 32, 107-132.

Glas, C. A. W., & Verhelst, N.D. (1989). Extensions of the partial credit model, *Psychometrika*, 54, 635-659.

Van der Schoot, F. (2001). *Standaarden voor kerndoelen basisonderwijs. De ontwikkeling van standaarden voor kerndoelen basisonderwijs op basis van resultaten uit peilingsonderzoek.* (Proefschrift Universiteit van Amsterdam). Arnhem, Cito.

Van der Schoot, F. (2008). *Onderwijs op peil? Een samenvattend overzicht van 20 jaar PPO.* Arnhem, Cito.

Verhelst, N.D., & Engelen, R.J.H. (1999). *An ability estimator in the two parameter logistic model based on raw scores.* Research memorandum. Arnhem: Cito.



Verhelst, N.D., & Glas, C.A.W. (1993). A dynamic generalization of the Rasch model. *Psychometrika*, 58, 395-415.

Verhelst, N.D., Glas, C.A.W., & Verstralen, H.H.F.M. (1993). OPLM: One parameter logistic model. Computer program and manual. Arnhem: Cito.

Warm, T.A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika*, 54, 427-450.



BIJLAGE 3 BEHORENDE BIJ ARTIKEL 3 VAN DE REGELING OMZETTING SCORES IN CIJFERS CENTRALE EXAMENS EN REKENTOETS VO 2015, VAN 9 FEBRUARI 2015, NUMMER CVTE-15.00618

Niet-bindende N-termen voor de pilot-profiel-cspe's

Bij de omzetting van scores in cijfers van de pilot-profiel-cspe's

- Bouwen, Wonen, Interieur 2015;
- Producten, Installatie en Energie 2015;
- Mobiliteit & Transport 2015;
- Zorg en Welzijn 2015;
- Economie en Ondernemen 2015;
- Horeca, Bakkerij en Recreatie 2015

wordt gewerkt met niet-bindende N-termen.

Dit houdt in dat:

- het CvTE de N-termen van de pilot-profiel-cspe's bepaalt volgens bijlage 1 bij deze regeling;
- pilotscholen per pilot-profiel-cspe een hogere N-term mogen hanteren dan de N-term die het College voor Toetsen en Examens heeft vastgesteld;
- de N-term van de pilotschool maximaal 0,5 hoger mag liggen dan de N-term van het College voor Toetsen en Examens;
- de pilotschool deze keuze mag maken op grond van door het bevoegd gezag opgestelde regels;
- een hogere N-term alsmede de reden daarvan gemeld wordt aan het College voor Toetsen en examens en aan de Inspectie van het Onderwijs.



TOELICHTING BEHORENDE BIJ ARTIKEL 3 VAN DE REGELING OMZETTING SCORES IN CIJFERS CENTRALE EXAMENS EN REKENTOETS VO 2015, VAN 9 FEBRUARI 2015, NUMMER CVTE-15.00618

Deze Regeling omzetting scores in cijfers centrale examens en rekentoets VO 2015 is ten opzichte van de regeling van 18 maart 2014 geactualiseerd ten aanzien van de wijziging van de Wet College voor examens, in de Wet College voor toetsen en examens. Daarnaast is er een bijlage toegevoegd die aangeeft op welke wijze de omzetting van scores in cijfers van de pilot-profiel-cspe's kan worden gehanteerd (zie onder 'Niet-bindende N-termen voor de pilot-profiel-cspe's' in deze toelichting).

Met deze regeling, die op 8 april 2015 is goedgekeurd door de Staatsecretaris van Onderwijs, Cultuur en Wetenschap, voldoet het College voor Toetsen en Examens aan de hem in artikel 2, tweede lid, aanhef en onderdeel e, en artikel 2 van de Wet College voor Toetsen en examens opgedragen taak om regels te stellen voor de omzetting van scores voor centrale examens en de rekentoets VO in cijfers. Gelet op de aard van betreffende voorschriften zijn deze niet opgenomen in de regeling zelf, maar in de daarbij behorende bijlage.

Deze regeling komt in de plaats van de Regeling omzetting scores in cijfers centrale examens en rekentoets VO 2014. De regeling van 2014 wordt ingetrokken door artikel 5. Artikel 1 en 2, met respectievelijk bijlage 1 en 2, waarin de normeringssystematiek voor de centrale examens en de rekentoets worden vastgesteld, zijn inhoudelijk ongewijzigd ten opzichte van de Regeling omzetting scores in cijfers centrale examens en rekentoets VO 2014. Deze twee bijlagen zijn aangepast op de naamswijziging van het College voor Examens in het College voor Toetsen en Examens. Aan de omzetting van scores in cijfers bij de centrale examens en de rekentoets verandert in schooljaar 2014-2015 dus niets.

Centrale examens VO

In bijlage 1 wordt beschreven op welke wijze bij een gegeven schaallengte (het maximaal aantal punten voor een examen), een door de correctoren vastgestelde score (het aantal door de kandidaat behaalde punten), en de door het College voor Toetsen en Examens vastgestelde normeringsterm van het examen, door de directeur een cijfer dient te worden vastgesteld. Er zijn vijf formules: één hoofdrelatie en vier grensrelaties. De normeringsterm, ook N-term genoemd en in de hoofdrelatie aangeduid met 'N', wordt aangepast aan de moeilijkheidsgraad van het examen, de grensrelaties dragen er zorg voor dat de cijferschaal begint bij 1,0 (voor de kandidaat die geen scorepunten heeft behaald) en eindigt bij 10,0 (voor de kandidaat die alle scorepunten heeft behaald). Dit systeem is zonder technische wijzigingen in gebruik sinds de centrale examens in het jaar 2000.

De normeringsterm van het examen wordt door het College voor Toetsen en Examens vastgesteld op grond van een specifiek vakinhoudelijke analyse van de door kandidaten per vraag behaalde score. Tevens wordt rekening gehouden met eventuele onvolkomenheden in de opgaven. Op het havo en vwo wordt bij de examens in mei van negen vakken door middel van equivalering (pretest of posttest) de moeilijkheidsgraad van het examen geïkt aan die van voorgaande jaren.

De vakken zijn: Frans, Duits, Engels, wiskunde A, wiskunde B, natuurkunde, scheikunde, biologie en economie. Op het vmbo is het aantal vakken met een pretest of posttest op dit moment beperkt tot Frans, Duits, Engels, wiskunde en economie in de theoretische en gemengde leerweg, en daardoor indirect ook voor de kaderberoepsgerichte leerweg, omdat de examens in beide leerwegen gedeeltematig (voor ongeveer de helft) identiek zijn.

In 2009 hebben de hoogleraren Hoijtink en Sijsma een rapport geschreven over de normeringssystematiek. Zij gaven aan dat het wenselijk is het aantal examens met equivalering uit te breiden en dat het eveneens wenselijk is andere vormen van equivalering dan de posttest en de pretest te overwegen. In 2011 is een eerste stap gezet naar uitbreiding van het aantal vakken met equivalering en naar toepassing van andere equivaleringstechnieken naast of in de plaats van pretest en posttest. Zo wordt standard setting toegepast bij het centraal schriftelijk en praktisch examen van tien beroepsgerichte vakken. Standard setting is een equivaleringstechniek waarbij experts de moeilijkheid van de afzonderlijke opgaven uit een centraal examen vergelijken met die van een referentie-examen. De ontwikkeling van digitale centrale examens gebaseerd op het principe van opgavenbanken heeft nieuwe mogelijkheden voor equivalering geopend. De digitale centrale examens in de algemene vakken van de basisberoepsgerichte leerweg zijn geëquivalerd, doordat geheim gehouden opgaven uit een eerder examenjaar deel uitmaakten van verschillende varianten van het digitale centrale examen.

Op basis van het grote aantal geëquivalerde centrale examens heeft het College voor Toetsen en Examens de vaardigheidsverschillen kunnen constateren tussen kandidaten van voor en na het invoeren van de nieuwe uitslagregels (2012).



Rekentoets VO

De systematiek die wordt toegepast bij de normering van de rekentoets VO wordt aangeduid als 'Normering met een vaardigheidsschaal'. In bijlage 2 wordt deze systematiek beschreven.

Moelijkheidsgraad van examenvarianten

Normering met een vaardigheidsschaal is vooral geschikt als er meer varianten van een toets zijn, waarbij de ene variant net iets makkelijkere opgaven kan hebben dan de andere variant. Dat is bij de rekentoets het geval.

De moeilijkheidsgraad van een variant kan worden bepaald doordat de moeilijkheid van de opgaven waaruit de variant bestaat psychometrisch is bepaald. De moeilijkheid van een opgave wordt bepaald door deze aan kandidaten voor te leggen.

Bij het toekennen van cijfers moet rekening gehouden worden met de verschillen in moeilijkheidsgraad tussen varianten. Op de makkelijkere varianten moet een kandidaat dan iets meer vragen goed beantwoord hebben om een zes te krijgen. Anders gezegd: een kandidaat moet eenzelfde vaardigheid aantonen voor een zes, ongeacht de variant die hij maakt, ongeacht het jaar waarin de toets wordt afgelegd dan wel de afnameperiode binnen dat schooljaar. En dit moet ook gelden voor ieder ander cijfer. Ongeacht de variant die gemaakt wordt, moet eenzelfde aangetoonde vaardigheid steeds beloond worden met eenzelfde cijfer. De methodiek van het normeren met een vaardigheidsschaal voorziet hierin.

Omzetting van scores – via vaardigheidsniveaus – in cijfers

Iedere toets en examen is te zien als een meting van het vaardigheidsniveau van de desbetreffende kandidaat. De score van de kandidaat op een toetsvariant is het resultaat van die meting. Iedere mogelijke score op een toetsvariant wordt uitgedrukt in een vaardigheidsniveau op een vaardigheidsschaal. De vaardigheidsniveaus van de vaardigheidsschaal worden omgezet in een cijfer. Zo wordt de score van een kandidaat omgezet in een cijfer. De wijze waarop dit gebeurt, is beschreven in bijlage 2. Daarbij wordt specifiek ingegaan op de wijze waarop de cesuur bepaald wordt: het vaardigheidsniveau dat de grens tussen een voldoende en een onvoldoende markeert.

Bij de normering van de centrale examens rekenen in het mbo wordt dezelfde systematiek toegepast. Daardoor wordt bewerkstelligd dat de cijfers voor de rekentoets VO dezelfde betekenis hebben als de cijfers voor de centrale examens rekenen in het mbo. Dat is van belang omdat dezelfde referentieniveaus rekenen worden getoetst.

Niet-bindende N-termen voor de pilot-profiel-cspe's

De mogelijkheden van hantering van de N-termen bij de omzetting van scores in cijfers van de pilot-profiel-cspe's zijn opgenomen in bijlage 3 bij deze regeling.

De profielvakken gaan de CE-stof vormen van de geheel vernieuwde beroepsgerichte examenprogramma's. De keuzevakken vormen de overige gedeelten van deze nieuwe examenprogramma's. Bij de meeste profielen worden drie of meer 'oude' beroepsgerichte vakken geïntegreerd. Bij het profiel PIE (Producten, Installatie en Energie) bijvoorbeeld zijn dat metaal-, elektro- en installatietechniek. De combinatie van een geheel nieuwe structuur en nieuwe inhoud maakt dat de vernieuwing van de beroepsgerichte programma's een zeer ingrijpende verandering is. De pilot-profiel-cspe's zijn daarvan het slotstuk.

De pilot-profiel-cspe's zijn ontwikkeld op basis van syllabi die de nieuwe CE-stof beschrijven en worden door de pilotscholen geheel volgens 'de regelen der kunst' afgenomen. Het enige verschil is dat de school bij de omzetting van score naar cijfer in 2015 enige speelruimte heeft. Deze invoeringsmaatregel is genomen om leerlingen en scholen niet te duperen, als in de pilot-profiel-cspe's van 2015 nog andere accenten gelegd zijn dan in de lessen die eraan voorafgegaan zijn.

*Het College voor Toetsen en Examens,
de voorzitter,
P.J.J. Hendrikse*