

NOTITIE

Omschrijving	Ontwikkelingsrichting normeringssystematiek na 2010
Datum	29 april 2010
Van	CvE

In 2009 is de normeringsprocedure voor de centrale examens in het voortgezet onderwijs onderwerp geweest van een audit door de hoogleraren Hoijtink en Sijtsma, in opdracht van de CEVO, voorganger van het College voor Examens. De conclusies zijn samengevat in een viertal punten. Op drie daarvan achten de onderzoekers een verbetering gewenst. Voor de centrale examens 2010 kon in de regels voor de omzetting van score naar cijfer door het College voor Examens nog geen rekening worden gehouden met de adviezen. Bij de regels voor de centrale examens in 2011 zullen de eerste stappen zichtbaar worden. In deze brief beschrijft het CvE de voornemens op grond van het advies van de hoogleraren, de wijze waarop omgegaan wordt met adviezen van derden en enkele ontwikkelingen ten aanzien van de normering die geen onderwerp van onderzoek door de hoogleraren waren.

1. Positief is het oordeel over wijze waarop met invalide vragen wordt omgegaan.
“De wijze waarop omgegaan wordt met foute en invalide vragen is goed. Een panel van experts (de vaksectie) bepaalt of er foute of invalide vragen in het examen zijn opgenomen, en de daarna toegepaste correctie van de N-term zorgt ervoor dat leerlingen niet door dergelijke vragen benadeeld worden.”
2. De methode van normering moet inzichtelijker en overzichtelijker worden beschreven en vastgelegd.
“Voor een keuze van een correct systeem van normering zijn verder nodig, een betere beschrijving, motivatie en uitleg van de ideeën en procedures die gebruikt worden.”
Het CvE werkt met Cito aan een Wikipedia-achtige integrale beschrijving van de systematiek. De beschrijving zal gereed zijn voor het centraal examen 2011.
3. Bij een deel van de examens wordt uitgegaan van gelijkblijvende vaardigheid van de kandidaten over de jaren heen. De hoogleraren vinden die aanname onvoldoende verantwoord.
“De vooronderstellingen van normering en normhandhaving met behulp van een referentie-examen gebaseerd op examenpopulaties met gelijkblijvende vaardigheid van jaar tot jaar zijn te streng. (...) Met behulp van deze procedure is het onmogelijk de cesuur zo te kiezen dat gelijke prestaties van jaar tot jaar gelijk gewaardeerd worden of om de ontwikkeling van vaardigheid over de jaren te volgen.”

Het College voor Examens wil beschikken over meer gegevens dan alleen de afnamegegevens en de vooronderstelling van gelijkblijvende vaardigheid. Het College breidt daarom het aantal examens met meer gegevens uit. In 2011 worden vier examens toegevoegd met een uitgebreide equivalering, met prioriteit voor het vmbo. Daarnaast wordt bij steeds meer examens standard setting gehanteerd, een eenvoudiger te realiseren methode met expertoordelen. In de toelichting wordt ingegaan op de vraag of uitbreiding tot alle examens nodig en haalbaar is.

4. Bij vakken met een pretest of posttest wordt niet uitgegaan van gelijkblijvende vaardigheid over de jaren heen. Het uitgangspunt achter pretest en posttest wordt in het advies onderschreven, de methodiek moet echter nader onderzocht, op onderdelen verbeterd of door alternatieven vervangen worden.
*“De principes waarop normering en normhandhaving met behulp van een pretest of een posttest zijn gebaseerd, zijn goed in de zin dat er expliciet rekening wordt gehouden met van jaar tot jaar variërende vaardigheden van leerlingen en moeilijkheden van examens.
 Echter, er is nader onderzoek nodig (...) om te kunnen bepalen of met behulp van benaderingen waarin extra gegevens worden verzameld daadwerkelijk de cesuur goed gekozen en de ontwikkeling van vaardigheid gevolgd kunnen worden.”*
 Het College voor Examens deelt deze conclusie. Nader onderzoek is nodig, en verbetering mogelijk. Bij de vier vanaf 2011 toegevoegde examens zal worden gewerkt met een schaduwexamen, zoals door de hoogleraren als alternatief genoemd.
5. Sinds de invoering van de huidige methodiek in 1999 is CEVO/CvE incidenteel benaderd door derden met voorstellen voor alternatieven. Deze zijn zorgvuldig bestudeerd maar waren geen aanleiding voor aanpassingen. Ook in de toekomst zijn adviezen van betrokkenen welkom en zullen deze op hun mogelijkheden worden beoordeeld.
6. De hoogleraren hebben zich in hun advies beperkt tot de normering van de papieren schriftelijke examens in het eerste tijdvak. Bij nieuwe en experimentele examenvormen, zoals het cspe in het vmbo, de flexibele en digitale centrale examens en de januari-examens in het kader van het Project Meerdere Examenmomenten, zijn verfijningen van de normeringssystematiek ontwikkeld die volledigheidshalve in de toelichting worden genoemd.

Toelichting:

Ad 1: normering bij fouten

Het is onvermijdelijk dat incidenteel pas bij de afname wordt geconstateerd dat een opgave niet “werkt”. Het CvE zal de daartoe bij de normering gehanteerde systematiek conform het advies van de hoogleraren handhaven.

Ad 2: betere beschrijving

Alle stappen in de normeringsprocedure zijn grondig en gedetailleerd beschreven en vastgelegd. Door voortdurende ontwikkeling is de beschrijving inmiddels een lappendeken die hoewel inhoudelijk consistent en intern goed hanteerbaar, voor externen niet inzichtelijk is.

Cito maakt in opdracht van het CvE een integrale, hiërarchisch gestructureerde wikipedia-achtige beschrijving van het (huidige) systeem, gebruikmakend van de aanwezige documentatie. Deze integrale beschrijving wordt jaarlijks geactualiseerd.

Het is daarnaast het voornemen om te komen tot jaarlijks één wetenschappelijke publicatie over (een aspect van) de normering, en (tenminste) één publicatie gericht op een breder publiek. Dat kan zijn gericht (bijvoorbeeld) op de docent moderne vreemde talen (via het vakblad van de Vereniging Levende Talen), of op de docent in het algemeen (via de bladen van AO_b en CNV Onderwijs), of via eigen breed verspreide publicaties van het College voor Examens en/of Cito. Een beknopt voorbeeld van zo'n publicatie voor een breder publiek is te vinden in de onlangs verschenen brochure "Examen en werkelijkheid", ter gelegenheid van het feit dat de CEVO opging in het College voor Examens. Juist bij de normering is soms sprake van vooroordelen in het veld en in de samenleving die aan de werkelijke kwaliteit geen recht doen.

Ad 3: gelijke vaardigheid over de jaren heen

Bij een deel van de centrale examens wordt uitgegaan van een gelijke vaardigheid van de populaties (over de jaren heen). De hoogleraren geven in hun rapport aan dat er onvoldoende zekerheid is dat dit uitgangspunt deugdelijk is. Het College voor Examens onderschrijft dit. In zijn brief bij het rapport van de hoogleraren geeft het CvE aan dat op dit moment met name op het vmbo het aantal examens met aanvullende equivaleringsgegevens te beperkt is. Al voor het advies van beide hoogleraren zijn stappen gezet naar uitbreiding.

Met de gewenste uitbreiding is niet alleen geld gemoeid maar ook tijd. Desondanks moeten op korte termijn, al bij het centraal examen 2011, stappen zichtbaar zijn. Gekozen is voor een uitbreiding bij het vmbo, in eerste instantie met vier examens (biologie en Engels BB en GT), met een schaduwexamen, zoals door de hoogleraren gesuggereerd is.

Een bredere basis met meer pretest, posttest of alternatieven is gewenst. Dat hoeft er niet toe te leiden dat op termijn alle centrale examens "hard" geëquivalereerd worden. Het "hard" equivaleren van de honderden huidige centrale examens is een nauwelijks haalbare en dure logistieke operatie. Per examen kost een goede pretest of posttest ruim €15.000,-, en jaarlijks stelt het CvE meer dan 400 verschillende centrale examens voor het voortgezet onderwijs vast. Een alternatief als het schaduwexamen zal niet goedkoper zijn. Daarbij is het rendement soms betrekkelijk; als het gaat om vakken die door maar weinig kandidaten worden gekozen, zeggen de resultaten weinig over het integrale onderwijspeil. Noodzakelijk is een voldoende grote kern van stevig geëquivalereerde centrale examens. Als bij die vakken een structurele verandering in vaardigheid wordt geconstateerd, kan indien nodig achteraf bij de overige vakken ten behoeve van het beleid nader onderzoek plaatsvinden. Hoe groot zo'n kern moet zijn is deels ook een maatschappelijke vraag (afweging van baten en lasten). Wel staat voor het College voor Examen vast dat de kern groter moet zijn dan thans het geval is, vooral bij het vmbo maar ook bij havo en vwo.

Ad 4: verbeteringen van, en alternatieven voor pretest en posttest.

Een door de hoogleraren genoemd alternatief voor de huidige pretest en posttest is het schaduwexamen (ankertoets). Een goed systeem met schaduwexamens kan goede en betrouwbare equivaleringsgegevens opleveren. De nauwkeurigheid van deze procedure moet worden onderzocht. Meer specifiek van belang is het vaststellen van het aantal kandidaten dat nodig is om met voldoende zekerheid uit het schaduwexamen conclusies te trekken.

Een schaduwexamen kan voorafgaand aan het centraal examen worden afgenomen. Een alternatieve vorm is een intern schaduwexamen waarbij bij een deel van de kandidaten een deel van het examen wordt vervangen door een schaduwdeel. Dit alternatief heeft onmiskenbare voordelen. Zo is de maximale, examengerichte inzet van kandidaten gegarandeerd. Ook blijft de uitvoeringslast voor deelnemende scholen beperkt. Er kunnen echter ook deels gevoelsmatige bezwaren aan kleven. Het is nieuw dat een deel van de kandidaten een afwijkend examen krijgt met inzage maar zonder publicatie. Bij de flexibele digitale examens in BB en KB, die al minder publiek zijn en varianten kennen, is een en ander eenvoudiger uit te voeren dan bij papieren examens.

Ook een meer traditioneel schaduwexamen, voorafgaand aan het gewone examen, heeft nadelen en risico's. Die liggen vooral in de uitvoeringslast voor scholen (er moet een extra toets onder zuivere condities worden afgenomen).

Alleen proefondervindelijk kan worden vastgesteld hoe het zit met mogelijkheden en risico's. Het College voor Examen start met Cito een pilot, in het examenjaar 2011. Daarbij wordt gedacht aan Engels en biologie, BB en GL/TL, en aan een pilot zowel met een intern als met een extern schaduwexamen. Cito heeft aangegeven dat een pilot van deze omvang op deze termijn logistiek mogelijk is.

Naast het ontwikkelen van deze nieuwe equivaleringsmethoden zal worden gewerkt aan verbetering van de bestaande methoden van pretest, posttest en standard setting, en aan nader onderzoek naar de robuustheid van de verschillende methodes. Wellicht kunnen elders beschikbare gegevens over de vaardigheid van populaties in de normering worden meegewogen. De methodologen van Cito sluiten niet uit dat langs die weg met betrekkelijk geringe kosten een stap vooruit kan worden gezet. De stappen zijn van wezenlijk belang, al beslaan zij niet een omvangrijk aantal vakken. Het belang van een goede normering vereist zorgvuldigheid. Evaluaties moeten uitwijzen in welke richting de verdere ontwikkeling zal gaan. Daarbij lijkt het waarschijnlijker dat bij de normering uit een rijkere scala aan mogelijkheden zal worden geput zodat beter kan worden aangesloten bij het karakter van soms heel verschillende examens, dan dat over enkele jaren elke normering totaal anders gaat dan nu het geval is.

Ad 5: Adviezen van derden

De huidige systematiek met de van jaar tot jaar variërende normeringsterm dateert van 1999. In de loop van deze jaren heeft CEVO/CvE enkele adviezen gekregen van derden. De adviezen vallen in twee categorieën uiteen.

Allereerst meenden sommigen dat het mogelijk en gewenst is om de normeringsterm vooraf vast te stellen. Dat is mogelijk, maar alleen op zeer lange termijn met een zeer grote gecalibreerde toetsenbank.

De tweede categorie betreft alternatieven voor de knik in de grafiek voor omrekening van score naar cijfer. De knik ontstaat omdat een kandidaat die alles goed heeft een 10,0 hoort te krijgen, en een kandidaat die niets goed heeft een 1,0. De knik heeft in de praktijk eigenlijk alleen een klein effect op cijfers boven de 9 en onder de 2, en zorgt ervoor dat 10,0 en 1,0 echt het maximum en het minimum zijn. In de adviezen wordt gesuggereerd de knikgrafiek te vervangen door een soepel ogende kromme, bijvoorbeeld tweedegraads of exponentieel. Dat oogt fraaier, maar er is geen wetenschappelijke onderbouwing voor. Een lineair verband is het eenvoudigst en daarom te prefereren. Alleen bij een lineair verband heeft bijvoorbeeld een verhoging van de normeringsterm met 0,2 punt vanwege een onvolkomenheid in de opgave hetzelfde effect op vrijwel alle cijfers. Bij andere verbanden varieert het effect afhankelijk van het cijfer, en dat is niet uit te leggen. Zowel de hoogleraren als de werkgroep examens van de VO-Raad achten de optisch merkwaardige knik een marginaal onderdeel van de normering en vinden de gekozen oplossing pragmatisch juist.

Ad 6: andere methoden van normering

In het advies van de hoogleraren zijn niet alle lopende ontwikkelingen belicht. De andere ontwikkelingen passen in de gewenste richting en kunnen daarom in dit kader worden genoemd. Bij de digitale centrale examens BB wordt de overlap tussen verschillende varianten in hetzelfde jaar benut om alle varianten een eigen normeringsterm te geven zo dat verschillen in moeilijkheidsgraad tussen de varianten door de normeringsterm worden geneutraliseerd. Het is naar verwachting een betrekkelijk kleine stap om deze ijking ook over de jaren heen te doen plaatsvinden, en daar wordt aan gewerkt. In feite is dat het ingebouwde schaduwexamen, zie Ad 4. Het uiteindelijke doel is een gecalibreerde itembank zoals bekend van de staatsexamens NT2. Een andere methode die nu al bij verschillende examens wordt toegepast, is de standard setting in de vorm van expert oordelen. In de vorm van een pilot wordt dit toegepast bij diverse cspe's van de beroepsgerichte vakken. En daarnaast wordt standard setting structureel toegepast bij de examens in januari en augustus, als onderdeel van het Project Meerdere Examenmomenten, en in bijzondere situaties zoals bij kunst (algemeen) dat in het eerste tijdvak twee zittingen kent. Standard setting kan een bruikbare aanvulling zijn met name in die gevallen waarin pretest, posttest of schaduwexamen niet goed mogelijk is of te kostbaar.