

Normering met een vaardigheidsschaal, zoals bij rekentoets VO en COE

Zoals gebruikelijk bij examens, worden bij de centrale rekentoets voor het voortgezet onderwijs en de centraal ontwikkelde examens taal en rekenen in het middelbaar beroepsonderwijs cijfers toegekend. Deze cijfers worden bepaald volgens de methode 'normering met een vaardigheidsschaal'.

Dit artikel informeert docenten en andere betrokkenen over de normeringssystematiek bij de centrale rekentoets (vo) en bij de centraal ontwikkelde examens taal en rekenen (mbo).

Inhoud

- 1 | Doel van normering met een vaardigheidsschaal
 - 2 | Meten van vaardigheid
 - 3 | Rapportage van resultaten
 - 4 | Getallen bij de vaardigheidsschaal
 - 5 | Van score naar vaardigheid
 - 6 | Van vaardigheid naar cijfer
 - 7 | Toekomstperspectief: itembanken
- Literatuur
Colofon

Dit artikel is de printversie van de toetsspecial 'Normering bij de centrale examens in het voortgezet onderwijs' op de website Toetswijzer, zie:
http://toetswijzer.kennisnet.nl/html/normering_vaardigheidsschaal/default.shtml.

1 | Doel van normering met een vaardigheidsschaal


1

Zoals gebruikelijk bij examens, worden bij de centrale rekentoets in het voortgezet onderwijs en de centraal ontwikkelde examens taal en rekenen (COE) in het middelbaar beroepsonderwijs cijfers toegekend. Deze cijfers worden bepaald volgens de systematiek 'normering met een vaardigheidsschaal'.

Het gebruik van een vaardigheidsschaal is vooral geschikt als er meerdere varianten van een examen zijn. Dit is het geval bij COE en rekentoets VO. De ene variant heeft soms net iets makkelijkere opgaven dan de andere variant. Voordat een opgave is voorgelegd aan examenkandidaten kan niet altijd heel precies worden ingeschat wat de moeilijkheid van een opgave is. Bij het toekennen van cijfers houden we daar rekening mee. Op de makkelijkere varianten moet een kandidaat dan iets meer vragen goed beantwoord hebben om een 6 te krijgen dan bij de moeilijkere varianten. We kunnen ook zeggen: een kandidaat moet eenzelfde vaardigheid aantonen voor een 6, ongeacht de variant die hij of zij maakt. 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.

Verskil met normering centrale examens

De methode van normering met een vaardigheidsschaal wijkt af van de normeringsmethode die bij de centrale examens in het voortgezet onderwijs wordt gebruikt.

Meer informatie over het verschil tussen beide methodes leest u in het document [Verskil normering centrale examens vo en normering met vaardigheidsschaal](#) .

Meer informatie

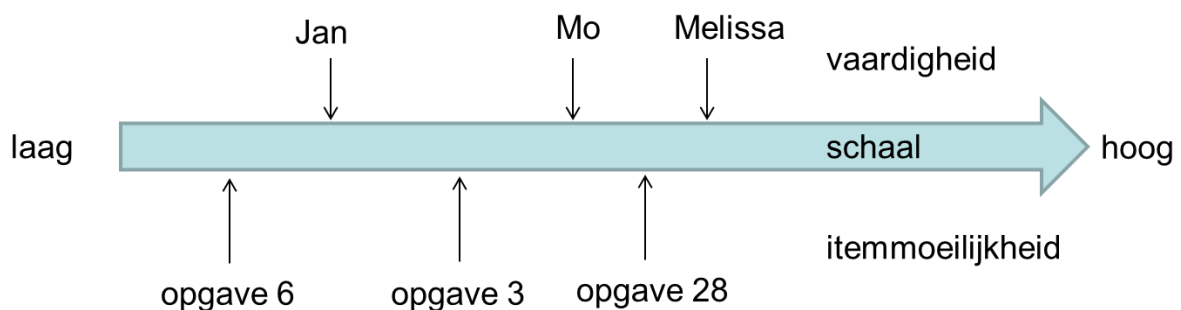
[Centraal Ontwikkelde Examens \(COE\)](#)
[Rekentoets VO](#)

2 | Meten van vaardigheid

In deel 1 hebben we kunnen lezen dat het cijfer wordt bepaald door de vaardigheid van een kandidaat. Deze vaardigheid kan hij of zij aantonen door opgaven van een examen goed te maken. Hiermee is het examen een meetinstrument van de vaardigheid van de kandidaat. Net zoals een thermometer een meetinstrument is van de temperatuur op de Celsius- of Fahrenheit-schaal, spreken we ook bij het examen 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 onderstaand figuur 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, moeilijkere opgaven worden minder vaak goed gemaakt dan makkelijkere opgaven.



Afbeelding van personen en opgaven op één vaardigheidsschaal

Itemmoeilijkheden

Een opgave heeft altijd dezelfde moeilijkheid; een examenvariant kan bestaan uit makkelijkere of moeilijkere opgaven. Een mini-examenvariant met alleen opgave 6 en opgave 3 uit bovenstaand voorbeeld, is bijvoorbeeld makkelijker 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, maken we op hoeveel moeilijker opgave 28 is dan opgave 3. Door deze twee verschillen te koppelen, leiden we af hoeveel moeilijker opgave 28 is dan opgave 6. Op deze manier kunnen de moeilijkheden van alle opgaven op dezelfde schaal worden afgebeeld.

Meer informatie over de schatting van itemmoeilijkheden leest u in het document [Schattingsmethode](#)



3 | Rapportage van resultaten

De rapportage van de prestaties van kandidaten heeft altijd een aantal kenmerken. Dit hoofdstuk geeft, aan de hand van onderstaande figuur, de vier belangrijkste kenmerken weer.

Naam instelling		School-	Landelijk
Instelling brincode		gemiddelde	gemiddelde
Examen	MBO COE 2012 Nederlands 3F pilot 11	84,8	86,2
Crebonummer			
Printdatum			

Kandidaatnummer	Naam kandidaat	Score	Cijfer	Schaalscore	Afnamedatum
155707		37	5,5	84,7	13-06-2012
154723		41	6,8	95,1	13-06-2012
157630		34	4,9	77,9	13-06-2012
157677		33	4,7	75,8	13-06-2012
157691		39	6,1	89,6	13-06-2012
157660		35	5,1	80,1	13-06-2012
157674		39	6,1	89,6	13-06-2012
157629		44	8,0	105,1	13-06-2012
157655		23	2,9	54,6	13-06-2012

Voorbeeld van een geanonimiseerde COE-rapportage

3

Kenmerken van de rapportage

De belangrijkste onderdelen van de rapportage zijn:

- De gemaakte examenvariant. Deze staat achter de naam van de kandidaat, of zoals in bovenstaand figuur, in de kop achter 'Examen'. Binnen een school worden aparte rapportages verstuurd per groep leerlingen. Bij de COE's worden deze groepen bepaald door de gemaakte examenvariant en het crebonummer.
- Het behaalde aantal punten op de gemaakte variant, ofwel de score per kandidaat.
- De vaardigheid bij deze score op deze variant. Deze staat in bovenstaand figuur onder het kopje 'Schaalscore'.
- Het cijfer behorende bij deze vaardigheid.

Lijsten van kandidaat-resultaten worden verstrekt per school en per groep. Bij COE wordt ook de gemiddelde vaardigheid van de kandidaten afgedrukt, zodat de school deze kan vergelijken met het landelijk gemiddelde.

4 | Getallen bij de vaardigheidsschaal

De schattingen van punten op de vaardigheidsschaal liggen vast op een lineaire transformatie na. We mogen bij alle punten 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 wordt willekeurig vastgesteld. 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 dit met een temperatuurschaal: het nulpunt op de Fahrenheit-schaal ligt ergens anders dan op de Celsius-schaal, maar voor beide geldt: hoe hoger, hoe warmer.

Nulpunt

Net als bij een thermometer heeft het nulpunt op een vaardigheidsschaal geen inhoudelijke betekenis. Zoals je bij temperatuur niet kunt spreken van 'het is vandaag twee keer zo warm als gisteren', zo kun je ook niet zeggen dat een kandidaat twee keer zo vaardig is als een andere kandidaat, want voor deze uitspraken is een absoluut nulpunt nodig. En die is er niet op de Fahrenheit-schaal, de Celsius-schaal en de vaardigheidsschaal. Op de Celsius-schaal is er wel een inhoudelijke betekenis van negatieve graden (het vriest), maar op de Fahrenheit-schaal vriest het al onder 32 graden en is deze inhoudelijke betekenis van negatieve temperaturen er dus niet.

Interval-schaal

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

5 | 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 gemaakt aan de hand van de items die daadwerkelijk aan de kandidaat zijn voorgelegd *).

Van alle antwoorden die een kandidaat geeft, wordt eerst de ruwe score berekend, ofwel het aantal behaalde punten. In het geval van een digitale toets met uitsluitend automatisch scorebare vragen, bepaalt de examensoftware de ruwe score. 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 hieronder het voorbeeld van een gedeelte van de omzettingstabel van score naar vaardigheid bij COE Nederlands 2F 2012, versie 2. De relatie tussen score en vaardigheid hoeft niet rechtlijnig te zijn. Dit hangt af van de onderlinge verschillen in moeilijkheid tussen de opgaven.

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

...
-----	-----	-----

Voorbeeld van een omzettingstabel van score naar vaardigheid

Deze methode levert 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 is niet overal even nauwkeurig. Bij de allerlaagste en allerhoogste scores is deze het minst nauwkeurig. De exacte score waarbij de meetnauwkeurigheid het hoogst is, hangt af van de moeilijkheid van de opgaven in de examenvariant.

Geneutraliseerde items

Als er opgaven in een examenvariant zitten, waarvan de normeringsvergadering beslist dat zij niet meewegen bij de beoordeling van een kandidaat, dan noemen we dit geneutraliseerde items. Met deze technische term maken we onderscheid in items die wel en niet meetellen. Alle kandidaten krijgen voor een geneutraliseerd item het maximale aantal te behalen punten. 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 doen we net of de geneutraliseerde items niet bestaan.

De vaardigheid van de kandidaat wordt geschat aan de hand van de antwoorden op een verkorte examenvariant, waarin de geneutraliseerde items niet opgenomen zijn. Hieronder 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 geneutraliseerde 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. In de praktijk komen deze scores niet voor omdat iedere kandidaat minimaal 3 punten scoort, namelijk op de geneutraliseerde items.

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

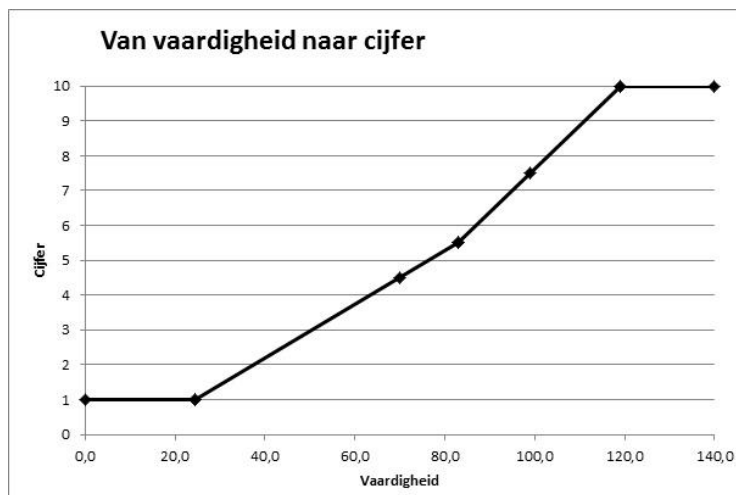
Vaardigheidsscores bij een examenvariant met 3 geneutraliseerde items van ieder 1 punt

*) 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).

6 | Van vaardigheid naar cijfer

Alle kandidaten en opgaven worden op dezelfde vaardigheidsschaal afgebeeld. Cijfers reflecteren de waardering voor behaalde vaardigheid. De cesuur is het belangrijkste punt op de vaardigheidsschaal: vanaf de cesuur-vaardigheid wordt de vaardigheid van kandidaten als voldoende beoordeeld. Een vaardigheid onder de cesuur wordt als onvoldoende beoordeeld.

Het precieze cijfer dat wordt toegekend bij de cesuur-vaardigheid is afhankelijk van het aantal gerapporteerde decimalen. Als hele cijfers gerapporteerd worden, zoals bij de rekentoets VO of het COE Rekenen, is het cijfer bij de cesuur een 5,5. Dit cesuur-cijfer heeft, om afrondingsproblemen te voorkomen, één decimaal meer dan er gerapporteerd wordt. Als cijfers met één decimaal worden gerapporteerd, zoals bij het COE Taal, is het cijfer bij de cesuur gelijk aan 5,45. In onderstaand figuur is dit bij vaardigheid 83,0.



Voorbeeld van omzetting van vaardigheidsscores naar cijfers

Verband tussen cijfer en cesuur-vaardigheid

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: een boven de cesuur en een 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. Een vaardigheid die resulteert in het cijfer 4 of lager is zodanig laag dat de kandidaat geen diploma waardig is.

Standaardsetting

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. Cijfers boven 10,0 en onder 1,0 worden niet toegekend. Het bepalen van de waarden van standaarden wordt [Standaardsetting](#) genoemd. Bij COE en rekentoets VO worden daarvoor de oordelen van zo'n vijftien experts gebruikt. Deze experts zijn niet betrokken bij de constructie van het examen of de toets. Er zijn diverse methoden waarbij deze experts oordelen per opgave moeten geven, of oordelen over sets van opgaven. De experts baseren zich voor het vereiste niveau op de [Syllabi COE en Toetswijzers Rekentoets VO](#).

7 | Toekomstperspectief

Na een aantal jaren van examenontwikkeling en -afnames kan een groot aantal opgaven op dezelfde schaal worden afgebeeld. We spreken dan van een itembank. Uit deze itembank kunnen varianten worden samengesteld, waarbij de normering vóór afname vastgesteld wordt. De itemparameters en de standaarden zijn immers al bekend.

Een voordeel van het werken met itembanken is dat de standaarden of cesuren die bij diverse afnames gehanteerd worden, goed met elkaar vergeleken kunnen worden. We kunnen dan binnen referentieniveaus diverse toetsen en de prestaties daarop vergelijken, bijvoorbeeld de rekentoets VO 2F vergelijken met de COE 2F rekenen. Ook de verticale vergelijking tussen niveaus is mogelijk, bijvoorbeeld tussen de rekentoets VO 2F en de rekentoets VO 3F. Dergelijke vergelijkingen bieden perspectief voor de verdere ontwikkeling van doorlopende toetslijnen.

Adaptieve toetsing

In het meest geavanceerde geval kan in de toekomst het examen of de toets op elk gewenst tijdstip en elke locatie worden afgenomen. Niet iedereen maakt hetzelfde examen. Mogelijk maakt iedere kandidaat een ander examen, dat wil zeggen een examen met andere opgaven. Misschien wordt het examen zelfs als adaptieve toets afgenomen. Dit laatste betekent dat de moeilijkheid van de opgaven in de loop van het examen wordt aangepast aan het niveau van de kandidaat. De exacte samenstelling van de examenvariant wordt dan gedurende de afname bepaald, in plaats van vooraf. Een nauwkeuriger meting van de vaardigheid van een kandidaat is een voordeel van adaptieve toetsing.

Een voorwaarde voor adaptieve toetsing is dat de opgaven niet algemeen bekend zijn, zodat een kandidaat steeds voor hem/haar nieuwe opgaven maakt. Om de geheimhouding van opgaven te waarborgen, is het inzagerecht bij COE en rekentoets VO bijvoorbeeld beperkt.

Literatuur

Eggen, 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 PPON](#). Arnhem: Cito.

Verhelst, N.D. (1992). *Het één parameter model (OPLM). Een theoretische inleiding en handleiding bij het computerprogramma*. 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.

Colofon

De toetsspecial 'Normering met een vaardigheidsschaal' is een uitgave van [Cito](#) in samenwerking met het [College voor Examens](#), gepubliceerd in april 2013.

De toetsspecial informeert docenten en andere betrokkenen over de normeringssystematiek bij de centrale rekentoets VO en bij de centraal ontwikkelde examens taal en rekenen mbo (COE).

De special is geschreven door M. van Onna, Cito. De special heeft de instemming van het College voor Examens.

Copyright © Stichting Cito Instituut voor Toetsontwikkeling

Het auteursrecht op de toetsspecial 'Normering met een vaardigheidsschaal' berust bij Cito.

Overname van de inhoud van de special is uitsluitend toegestaan na toestemming van de redactie van Toetswijzer. Verzoeken tot overname van de inhoud dienen te worden gericht aan de [redactie van Toetswijzer](#), onder vermelding van het medium, de oplage en de doelgroep.

Het is toegestaan om via de weblink

http://toetswijzer.kennisnet.nl/html/normering_vaardigheids

[schaal/default.shtm](http://toetswijzer.kennisnet.nl/html/normering_vaardigheids_schaal/default.shtm) te verwijzen naar de toetsspecial 'Normering met een vaardigheidsschaal'.