

## Beoordelen van de kwaliteit van toetsen en examens

Deel 1: Systemen en criteria



# Inhoud

<b>1. Inleiding</b>	<b>1</b>
<b>2. Inventarisatie en evaluatie van richtlijnen, standaarden en beoordelingssystemen</b>	<b>2</b>
2.1 Internationale systemen voor toetskwaliteit	3
2.2 Nederlandse systemen voor toetskwaliteit	4
2.3 Relevante criteria	5
<b>3. Conclusies</b>	<b>7</b>
<b>Bijlage: Literatuur</b>	<b>8</b>

# 1. Inleiding

In Nederland worden regelmatig belangrijke besluiten genomen op basis van resultaten op toetsen of examens. Zowel over onderwijs-carrières van leerlingen en studenten als over de loopbanen van werknemers. Om dit soort beslissingen goed onderbouwd te kunnen nemen, is het van groot belang dat de gebruikte instrumenten een hoge kwaliteit hebben. Discussies en klachten in de sociale media en daarbuiten over de kwaliteit van toetsen en examens komen echter regelmatig voor.

In het voortgezet onderwijs is bijvoorbeeld de klachtenregen rond de centrale examens een jaarlijks terugkerend fenomeen. En in het mbo geeft de onderwijsinspectie al verschillende jaren aan dat de kwaliteit van de examinering, ondanks een reeks verbeteringen, nog steeds te wensen overlaat (Inspectie van het Onderwijs, 2017). Verder hebben zich in het hbo de nodige incidenten voorgedaan rond examens die aanleiding waren voor een herziening van de wijze waarop de examinering in die onderwijssector vormgegeven wordt (Commissie externe validering examenkwaliteit hoger beroepsonderwijs, 2012).



Minder bekend voor het grote publiek zal zijn dat ook in de beroepssector het nodige te doen is over examenkwaliteit. Verschillende beroepsgroepen in de financiële sector moeten al sinds enkele jaren examen doen om een beroepskwalificatie te *verkrijgen* of te *behouden*. En ook daar is sprake van kritiek op de inhoud en kwaliteit.



Nieuws 9 apr 2014

## Hbo'ers keihard onderuit bij Wft-examens

Branche 169

**Hbo-studenten zijn massaal gezakt voor hun Wft-examens. De studenten dienden in groten getale een klacht in bij het klachtenloket van de Commissie Financiële Dienstverlening (CFD). De CFD trekt naar aanleiding van de grote klachtenstroom aan de bel bij het ministerie van Financiën.**

Er is dus meer dan voldoende reden om concreet in te gaan op twee belangrijke vragen:

1. Welke kwaliteitseisen gelden er eigenlijk voor toetsen en examens?
2. Waarom is de discussie over toetskwaliteit lastig te voeren?

We beantwoorden de eerste vraag in deze notitie aan de hand van een inventarisatie van richtlijnen en standaarden voor toetsontwikkeling en beoordelingskaders voor toetskwaliteit. De tweede vraag wordt in een aparte *notitie* beantwoord. De twee ToetsSpecials bieden samen een kader om de klachten en discussies over de kwaliteit van specifieke toetsen of examens op waarde te kunnen schatten.

De kwaliteit van een product is vanuit verschillende invalshoeken te beoordelen. Het meest direct en voor de hand liggend is natuurlijk de beoordeling van het product zélf. In sommige gevallen, bijvoorbeeld als de beoordeling van het product onmogelijk of te kostbaar is, kan de kwaliteit op een indirecte manier beoordeeld worden. Bijvoorbeeld door het bepalen van de deskundigheid van de productontwikkelaars. Een derde invalshoek is het evalueren van de kwaliteit van het ontwikkelingsproces van het product. Tot slot is het mogelijk om gebruikers van het product te bevragen over hun tevredenheid (Eggen, 2013). Dergelijke indirecte vormen van kwaliteitsevaluatie zijn ook in te zetten ter aanvulling of verdere versterking van de directe beoordeling van een toets of examen. In deze notitie ligt de focus echter volledig op de eerste invalshoek – evaluatie van het product zelf - en laten we de andere perspectieven buiten beschouwing. Voor het gemak zullen we in de rest van dit document vaak alleen de term ‘toetsen’ gebruiken in plaats van ‘toetsen en examens’.

## 2. Inventarisatie en evaluatie van richtlijnen, standaarden en beoordelingssystemen

**Van Dale geeft als eerste betekenis van kwaliteit ‘mate waarin iets goed is; gesteldheid, hoedanigheid, aard’. Een in dit kader bruikbare definitie die daar goed bij aansluit is de volgende. Toetskwaliteit is de mate waarin een toets geschikt is voor het ermee beoogde doel (Wools, 2015). Naarmate een toets beter geschikt is voor het gebruiksdoel is de kwaliteit ervan dus beter.**

Om te kunnen bepalen in welke mate een toets geschikt is voor het ermee beoogde doel, zijn de nodige kwaliteitsbeoordelingssystemen ontwikkeld. Hierna wordt een reeks van vijftien verschillende systemen beschouwd. We maken hierbij een onderscheid tussen internationale systemen en Nederlandse systemen. Voor deze selectie is gekozen, omdat het gaat om systemen ‘van en voor professionals’ die zijn ontwikkeld binnen vakverenigingen. Of door wetenschappers die zich bewegen op het gebied van onderwijskundig meten. Naar vrijwel alle systemen zijn koppelingen opgenomen, zodat ze ook eenvoudig terug te vinden zijn.

## 2.1 Internationale systemen voor toetskwaliteit

### 2.1.1 *Standards for Educational and Psychological testing*

Deze ‘*Standards*’ worden sinds 1966 in verschillende edities gepubliceerd door drie Amerikaanse instituten: de American Educational Research Association (AERA), de American Psychological Association (APA) en de National Council on Measurement in Education (NCME). De meest recente uitgave is die uit 2014 (APA, AERA & NCME, 2014). Deze uitgave wordt in de Verenigde Staten en vele andere landen gezien als de gouden standaard op het gebied van het ontwikkelen van toetsen.

### 2.1.2 *EFPA Review model for the description and evaluation of psychological tests*

Dit door de European Federation of Psychological Associations (EFPA) ontwikkelde systeem is bedoeld als een beoordelingsinstrument. Het is onder meer gebaseerd op criteria van de British Psychological Society en die van de *COTAN*, de Commissie TestAangelegenheden Nederland, een onderdeel van het Nederlands Instituut van Psychologen (NIP). De oorsprong ervan ligt in 1989, maar een eerste formele versie ervan bestaat sinds 2002. De meest recente versie van de *review criteria* dateert van 2013 (EFPA, 2013).

### 2.1.3 *ITC Guidelines*

De International Test Commission (ITC) is een associatie van nationale verenigingen van psychologen, testcommissies, organisaties en individuen die de correcte ontwikkeling en het juiste gebruik van toetsen en psychologische tests willen bevorderen. De vereniging is ontstaan in de jaren zeventig van de vorige eeuw (zie Oakland, Poortinga, Schlegel, & Hambleton, 2001). De ITC heeft zes verschillende *richtlijnen* ontwikkeld, waaronder die voor testgebruik (International Test Commission, 2013), *testbeveiliging* (International Test Commission, 2014) en *computergestuurde en internettesten* (International Test Commission, 2005).

### 2.1.4 *Code of Fair Testing Practices in Education*

Deze *code* is consistent met de inhoud van de eerder genoemde Standards for Educational and Psychological Testing. Deze beknopte en toegankelijke notitie is uitgebracht door de JCTP (Joint Committee on Test Practices, 2004). De JCTP is in 1985 opgericht door de AERA, APA, en NCME en heeft bestaan tot 2007.

### 2.1.5 *Classroom Assessment Standards for Pre K-12 Teachers*

De Joint Committee on Standards for Educational Evaluation (JCSEE) is een verbond van verenigingen en organisaties die zich bezighouden met beoordelingen en evaluatie. De JCSEE beheert een drietal standaarden, waaronder de *Classroom Assessment Standards* die bedoeld zijn voor leerkrachten in het basisonderwijs. De JCSEE is geaccrediteerd door het American National Standards Institute (ANSI). De standaarden van de JCSEE kunnen daarom beschouwd worden als dé Amerikaanse nationale standaarden.

### 2.1.6 *European framework of standards for educational assessment*

De Association for Educational Assessment Europe (AEA-Europe) bracht in 2012 een *document* uit dat standaarden voor toetsontwikkeling bevat. Dit document was ook bedoeld is om de kwaliteitsstandaarden die binnen Europa gebruikt worden voor onderwijskundig meten met elkaar te kunnen vergelijken om zo tot een internationaal geaccepteerde werkwijze te komen.

### 2.1.7 *ETS Standards for Quality and Fairness*

Educational Testing Service (ETS) is een van de grootste Amerikaanse toetsontwikkelaars en hanteert een eigen *systeem* met de voornoemde titel (Educational Testing Service, 2015). Het doel van dit systeem is ervoor te zorgen dat ETS kwalitatief hoogstaande producten en diensten oplevert. Daarnaast is het systeem bedoeld om de stakeholders van deze organisatie duidelijk te maken volgens welke voorschriften ETS zijn producten en diensten ontwikkelt.

### 2.1.8 Cambridge Approach to Assessment

Cambridge Assessment is een van de grootste Europese organisaties die zich bezighoudt met de ontwikkeling van toetsen. In het *document* 'The Cambridge Approach to Assessment' (Cambridge Assessment, 2017) geeft deze organisatie aan hoe men bewerkstelligt dat alle beoordelingsinstrumenten die dit bedrijf uitbrengt eerlijk en ethisch verantwoord zijn en voldoen aan de hoogste technische standaarden.

## 2.2 Nederlandse systemen voor toetskwaliteit

### 2.2.1 Beoordelingssysteem voor de kwaliteit van tests van de COTAN

Het *beoordelingssysteem* voor de kwaliteit van tests van de COTAN (Evers, Lucassen, Meijer, & Sijtsma, 2010) vindt zijn oorsprong in 1969 toen dit hierboven al genoemde onderdeel van het NIP begon met het publiceren van beoordelingen van tests. Het systeem is sindsdien verschillende malen herzien. De huidige versie is ten opzichte van de voorafgaande ingrijpend gewijzigd om beter te passen bij recente ontwikkelingen op het terrein van de toetstheorie en toetsconstructie.

### 2.2.2 RCEC beoordelingssysteem voor de kwaliteit van studietoetsen en examens

Het 'RCEC beoordelingssysteem van de kwaliteit van studietoetsen en examens' (Sanders, Van Dijk, Eggen, Den Otter & Veldkamp, 2016) is ontwikkeld om zowel de kwaliteit van toetsen als van (praktijk)examens te kunnen beoordelen. Het *RCEC-beoordelingssysteem* heeft de nodige raakvlakken met het systeem van de COTAN, maar is specifiek ontwikkeld vanuit het perspectief van toetsen en examens in onderwijs.

### 2.2.3 Beoordelingskaders van eindtoetsen en leerlingvolgsysteemtoetsen

Met ingang van het schooljaar 2014-2015 moeten alle schoolverlaters in het reguliere basisonderwijs een eindtoets maken die hun taal- en rekenvaardigheden meet. In 2017 konden scholen uit het basisonderwijs kiezen

tussen de Centrale Eindtoets en vijf andere eindtoetsen. Die vijf eindtoetsen zijn door de Expertgroep PO op kwaliteit beoordeeld aan de hand van een concreet *verantwoordingskader*. Daarnaast hanteert de Expertgroep PO een concreet *beoordelingskader* voor toetsen uit leerlingvolgsystemen.

### 2.2.4 Kwaliteitscriteria voor competentie-assessmentprogramma's

Deze publicatie betreft een *proefschrift* (Baartman, 2008) dat een overzicht geeft van de criteria waar de examinering binnen competentiegericht onderwijs zouden moeten voldoen. Competentiegericht onderwijs is een term die een tijd in zwang is geweest om middelbaar en hoger beroepsonderwijs te beschrijven. Toetsing en examinering moet in dit soort onderwijs plaatsvinden in de vorm van een methodenmix (Straetmans & Sanders, 2001) of een competentie-assessmentprogramma zoals Baartman dat beschrijft.

### 2.2.5 Regeling standaarden examenkwaliteit mbo

De beoordeling van de kwaliteit van de examens in het middelbaar beroepsonderwijs kent een complexe geschiedenis (Sluijter, 2016). In 2012 werd met het *Toezichtkader 2012* (Inspectie van het Onderwijs, 2011) een nieuw integraal waarderingskader geïntroduceerd. Het huidige toezichtkader dat de Inspectie van het Onderwijs hanteert, is ten opzichte van de versie van 2012 niet principieel gewijzigd.

### 2.2.6 Kwaliteitspiramide voor eigentijds toetsen en beoordelen

De *kwaliteitspiramide* voor eigentijds toetsen en beoordelen is ontwikkeld door Joosten-ten Brinke (2011) en richt zich op het borgen van de kwaliteit van toetsing en examinering binnen een instelling op verschillende met elkaar samenhangende niveaus: toetsonderdelen, toetsen, toetsprogramma's en de visie op toetsing en examinering van een onderwijsinstelling.

### 2.2.7 Checklist toetskwaliteit Eerste hulp bij toetsen

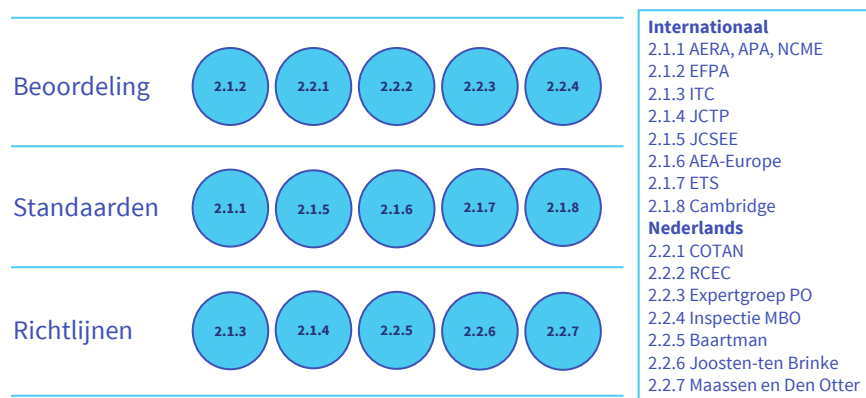
Meer gericht op docenten in het voortgezet onderwijs is de publicatie *Eerste hulp bij toetsen* (Maassen & Den Otter, 2014). Op basis van herkenbare vragen en opmerkingen van leerlingen wordt de kwaliteit van toetsen beschreven in de

vorm van een checklist. De kwaliteitscriteria die aan bod komen zijn vaak dezelfde als in andere publicaties, alleen worden deze in deze checklist gezien vanuit de praktijk van de docent. Zo moet deze publicatie voor hen als hulpmiddel dienen voor het maken en evalueren van toetsen.

De vijftien genoemde systemen verschillen wat karakter betreft. Sommige systemen ademen een zekere vrijblijvendheid en sommen een reeks criteria op waar men aandacht aan zou *kunnen* besteden bij het ontwikkelen van toetsen. Andere systemen leveren richtlijnen: regels die je zou *moeten* volgen bij het ontwikkelen van toetsen. Tot slot zijn er ook systemen die werken met voorschriften: regels waar je je aan *moet* houden bij het ontwikkelen van toetsen. Deze systemen maken het ook mogelijk om tot een concreet kwaliteitsoordeel te komen.

Figuur 1 geeft grafisch weer in welke categorie de verschillende genoemde systemen zich bevinden. De systemen van de EFPA, COTAN, het RCEC, de Expertgroep PO en de onderwijsinspectie voor het mbo kunnen beschouwd worden als beoordelingssystemen.

Figuur 2.1 | Categorieënindeling van de verschillende systemen



## 2.3 Relevante criteria

Een beschouwing van al deze systemen maakt duidelijk dat er een grote overlap bestaat voor wat betreft de criteria die specialisten op het vlak van onderwijskundig meten van belang achten. De vijf systemen waarbij sprake is van een beoordelingskader vertonen grote overeenkomsten. Dat mag ook geen verbazing wekken, omdat zowel het systeem van de EFPA, het systeem van het RCEC als het beoordelingskader van de Expertgroep PO allemaal mede het beoordelingssysteem van de COTAN als uitgangsbasis hebben genomen. Het beoordelingskader voor de kwaliteit van examens van de onderwijsinspectie voor het mbo wijkt af van de overige vier, maar ook in dit systeem herkennen we veel elementen uit de andere.

We concentreren ons hier verder op het systeem van het RCEC. Het RCEC-systeem is namelijk specifiek ontwikkeld voor het beoordelen van toetsen én examens. Dat geldt niet voor de andere drie systemen. De systemen van de EFPA en de COTAN focussen op het beoordelen van tests die voor een langere tijd bruikbaar zijn, veelal psychologische tests. Deze tests blijven jaren, soms wel decennia onveranderd. Dat staat in duidelijk contrast met onderwijskundige toetsen. Met name examens kennen een beperkte, soms zeer beperkte, afnameperiode. Zo worden de centrale eindexamens slechts op één moment afgenomen, en daarna niet meer. Het beoordelingskader van de Inspectie van het Onderwijs richt zich volledig op examenkwaliteit binnen een enkel onderwijsniveau. Het beoordelingskader van de Expertgroep PO concentreert zich volledig op toetsen en dan ook nog eens twee specifieke typen toetsen.

Het RCEC beoordeelt de kwaliteit van toetsen en examens op basis van de volgende criteria:

- Doel en gebruik
- Kwaliteit van het toets- of examenmateriaal
- Representativiteit
- Betrouwbaarheid
- Standaardbepaling en normhandhaving
- Afname en beveiliging

Hieronder gaan we kort in op deze criteria.

#### ► **Doel en gebruik**

Bij dit criterium draait het om een drietal vragen: voor wie; wat en waarvoor?

Allereerst moet duidelijk zijn voor wie de toets bedoeld is. Concrete beschrijving van de doelgroep is nodig om te kunnen bepalen of de moeilijkheid van het meetinstrument en het taalgebruik passend zijn. Daarnaast moet helder zijn hoe de inhoud van de toets zich verhoudt tot het meetdoel, en moet voldoende gedetailleerd beschreven zijn wat de toets meet. Tot slot moet het doel van de toets concreet omschreven zijn: selectie, classificatie, plaatsing, certificering of diplomering, of een ander doel.

Zie voor meer informatie hierover de aansluitende ToetsSpecial *Beoordelen van de kwaliteit van toetsen en examens in de praktijk*.

#### ► **Kwaliteit van het toets- of examenmateriaal**

Bij dit criterium gaat het om de vraag of het afnemen van de toets leidt tot resultaten die voor alle leden van de doelgroep vergelijkbaar zijn. De werkelijke vaardigheid van leden van de doelgroep moet zo goed mogelijk weerspiegeld worden door de toetsresultaten en niet afhankelijk zijn van toevallige factoren. Standaardisatie van de toetsinhoud en het beoordelen van de prestaties van de leerlingen is daarvoor van belang.

#### ► **Representativiteit**

Dit criterium richt zich op de vraag of de inhoud van de toets past bij hetgeen men wil meten en voldoende toegesneden is op het niveau van de leden van de doelgroep. Een toets mag niet eenzijdig zijn en de onderdelen die aan bod horen te komen moeten ook daadwerkelijk bevestigd worden.

#### ► **Betrouwbaarheid**

Dit criterium richt zich op de vraag of het de ontwikkelaars van de toets ook voldoende gelukt is toevallige elementen uit te bannen die effect zouden kunnen hebben op de scores op de toets. Dit kunnen toevalligheden zijn die samenhangen met de samenstelling van de toets, de beoordelaars of de afname-omstandigheden. Voor ieder van deze toevalligheden zijn verschillende maten beschikbaar. Hoe betrouwbaarder een toets, hoe consistent, nauwkeuriger en beter reproduceerbaar de uitkomsten ervan zijn.

#### ► **Standaardbepaling en normhandhaving**

Een toetsscore zoals bijvoorbeeld het aantal goede antwoorden zegt op zich weinig. De toetsscore moet te interpreteren zijn om tot een beslissing te kunnen leiden. Bij relatieve normen gebeurt dat door de score te vergelijken met de scores van andere leden van de doelgroep. Bij absolute normen is er vooraf een grens vastgelegd op de scoreschaal, bijvoorbeeld om te bepalen of een kandidaat gezakt of geslaagd is. Als er verschillende toetsen of examens zijn, bijvoorbeeld voor verschillende meetmomenten, dan is ook vereist dat de genomen beslissing niet afhangt van de toevallige samenstelling van het instrument. Dit criterium richt zich dus op de kwaliteit van de vertaling van toetsscore naar interpretatie of beslissing.

#### ► **Afname en beveiliging**

Bij dit criterium gaat het erom dat het voor de personen die de toets afnemen duidelijk moet zijn hoe ze horen te handelen. Een andere element is dat de inhoud van de toets én de kandidaat- of leerlinggegevens voldoende beveiligd moeten zijn, zodat fraude niet mogelijk is en de privacy van personen niet in het geding komt.





## 3. Conclusies

We hebben in het voorafgaande laten zien dat er een reeks voldoende concreet omschreven kwaliteitseisen voor toetsen en examens bestaat. En dat er een grote mate van overlap bestaat tussen de inhoud van de verschillende systemen die in omloop zijn voor het beoordelen van de kwaliteit van toetsen en examens. Dat zou het doen van een uitspraak over de kwaliteit van een specifieke toets of een specifiek examen redelijk eenvoudig moeten maken. In de praktijk is dat echter niet zo. Waarom dat zo is, is het onderwerp van het *tweede deel* van deze ToetsSpecial.

# Bijlage: Literatuur

American Psychological Association, American Educational Research Association & National Council on Measurement in Education. (2004). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Baartman, L. (2008). 'Assessing the assessment'. *Development and use of quality criteria for Competence Assessment Programmes*. Proefschrift: Dutch Interuniversity Centre for Educational Research.

Cambridge Assessment (2017). *The Cambridge Approach to Assessment. Principles for designing, administering and evaluating assessment*. Revised Edition. <http://www.cambridgeassessment.org.uk/Images/cambridge-approach-to-assessment.pdf>.

Commissie externe validering examenkwaliteit hoger beroepsonderwijs. (2012). *Vreemde ogen dwingen. Eindrapport Commissie externe validering examenkwaliteit hoger beroepsonderwijs*. Den Haag: HBO-raad, vereniging van hogescholen.

Educational Testing Service (2014). *ETS standards for Quality and Fairness*. Princeton: Educational Testing Service.

Eggen, T.J.H.M. (November 2013). Doelbewuste kwaliteit van toetsten getoetst. Presentatie op het negende nationale examencongres van de Nederlandse Vereniging voor Examen. Nunspeet.

European Federation of Psychological Associations (2013). *EFPA Review Model for the Description and Evaluation of Psychological and Educational Tests. Test Review Form and Notes for Reviewers. Version 4.2.6*. Geraadpleegd op 27-07-2017 op: <http://www.efpa.eu/professional-development/assessment>.

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, S. (2010). *COTAN Beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP. [[www.psynip.nl](http://www.psynip.nl): <https://www.psynip.nl/wp-content/uploads/2016/07/COTAN-Beoordelingssysteem-2010.pdf>].

Oakland, T., Poortinga, Y. H., Schlegel, J., & Hambleton, R. K. (2001). International Test Commission: Its History, Current Status, and Future Directions. *International Journal of Testing*, 1, 3-32.

International Test Commission (2005). *International Guidelines on Computer-Based and Internet Delivered Testing*. [[www.intestcom.org](http://www.intestcom.org): [https://www.intestcom.org/files/guideline\\_computer\\_based\\_testing.pdf](https://www.intestcom.org/files/guideline_computer_based_testing.pdf)]

International Test Commission (2013). *ITC Guidelines on Test Use. Version 1.2*. [[www.intestcom.org](http://www.intestcom.org): [https://www.intestcom.org/files/guideline\\_test\\_use.pdf](https://www.intestcom.org/files/guideline_test_use.pdf)]

International Test Commission (2014). *International Guidelines on the Security of Tests, Examinations, and Other Assessments*. [[www.intestcom.org](http://www.intestcom.org): [https://www.intestcom.org/files/guideline\\_test\\_security.pdf](https://www.intestcom.org/files/guideline_test_security.pdf)]

Inspectie van het Onderwijs (2011). *Toezichtkader BVE 2012*. Utrecht: Inspectie van het Onderwijs.

Joosten-ten Brinke, D. (2011). *Eigentijds toetsen en beoordelen*. Lectorale rede. Tilburg: Fontys Lerarenopleiding Tilburg.

Maassen, N. & Den Otter, D. (2014). *Eerste hulp bij toetsen: grip op toetskwaliteit*. Enschede: RCEC. [[https://www.nro.nl/wp-content/uploads/2014/12/RCEC\\_Kwaliteit\\_Toets\\_Checklist\\_2014.pdf](https://www.nro.nl/wp-content/uploads/2014/12/RCEC_Kwaliteit_Toets_Checklist_2014.pdf)]

Sanders, Van Dijk, Eggen, Den Otter & Veldkamp, 2016 RCEC Beoordelingssysteem voor de kwaliteit van studietoetsen en examens. Enschede: RCEC.

Sluijter, C. (2016). De kwaliteit van toetsen en examens in het middelbaar beroepsonderwijs. In P.F. Sanders (Red.) *Toetsen op School. Middelbaar Beroepsonderwijs*. (pp. 102-116). Arnhem: Cito.

Straetmans, & P.F. Sanders (2001). *Beoordelen van competenties van docenten*. Den Haag: Programmamanagement EPS/HBO-raad.

Wools, S. (2015). All about Validity. An evaluation system for the quality of educational assessment. Proefschrift. Enschede: Universiteit Twente. [[https://www.researchgate.net/publication/281434844\\_All\\_About\\_Validity\\_%20An\\_evaluation\\_system\\_for\\_the\\_quality\\_of\\_educational\\_assessment](https://www.researchgate.net/publication/281434844_All_About_Validity_%20An_evaluation_system_for_the_quality_of_educational_assessment)]

**Over deze ToetsSpecial** | Resultaten op toetsen en examens beïnvloeden het verloop van onderwijsloopbanen en carrières daarna in sterke mate. Toetsen en examens moeten daarom van onbesproken kwaliteit zijn. Maar het bepalen van de kwaliteit van toetsen en examens is niet eenvoudig. Deze special is de eerste in een reeks van twee die zich richten op de kwaliteit van toetsen en examens. *Dit eerste deel bevat een overzicht en evaluatie van kwaliteitscriteria en kwaliteitssystemen.*

De notitie is bedoeld voor toetsdeskundigen en onderwijsgeevenden in alle sectoren, maar kan ook inzichtelijk zijn voor iedereen die zich een eigen oordeel wil vormen over de mate waarin de huidige kritiek op de kwaliteit van toetsen en examens terecht is.

## Colofon

**Uitgave:** ToetsWijzer | Stichting Cito Instituut voor Toetsontwikkeling

**Auteur(s):** Cor Sluijter, Bas Hemker, Theo Eggen (Cito)

**Datum:** mei 2018

e [info@toetswijzer.nl](mailto:info@toetswijzer.nl) | i [www.toetswijzer.nl](http://www.toetswijzer.nl)

