

RESEARCH CENTER VOOR  
EXAMINERING EN CERTIFICERING



## **KWALITEIT VAN TOETSEN BINNEN HANDBEREIK**

WELKE KWALITEITSASPECTEN ZIJN  
VAN BELANG BIJ EEN GOEDE TOETS?

BROCHURE VOOR DOCENTEN

ONDERZOEKSRESULTATEN EN AANGRIJPINGS  
PUNTEN  
VOOR DE PRAKTIJK



UNIVERSITEIT TWENTE.



# **Kwaliteit van toetsen binnen handbereik**

Welke kwaliteitsaspecten zijn  
van belang bij een goede toets?

Brochure voor docenten

Onderzoeksresultaten en aangrijpingspunten voor de praktijk

## INHOUDSOPGAVE

Voorwoord	3
Inleiding	5
1	Wat weten we over de kwaliteit van een toets? 8
1a	Welke aspecten van kwaliteit kwamen in de artikelen voor? 8
1b	Welke aspecten van kwaliteit zijn het meest genoemd? 10
2	Waar hangt de kwaliteit van een toets vanaf? 16
2a	Doel van de toets 16
2b	Onderwijssector 17
2c	Rollen in de beoordeling 17
2d	Toetscyclus 17
2e	In de praktijk 18
3	Hoe kan ik als docent een goede toetskwaliteit bereiken? 21
4	Waar moeten we nog meer over weten? 22
Literatuur	23

## VOORWOORD

Toetsing is overal in het onderwijs aan de orde van de dag. De uitslagen van deze toetsen kunnen zeer bepalend zijn voor de onderwijsloopbaan van de studenten. Studenten zullen daarom hun uiterste best doen om hun kennis en vaardigheid te demonstreren. Maar ook als docent gebruikt u toetsen om beslissingen te nemen over de precieze invulling van het onderwijsleerproces. Als schoolleider of beleidsmedewerker zult u erop gericht zijn om een toetsprogramma zo samen te stellen dat het eindniveau van de studenten is gewaarborgd. U wilt natuurlijk dat het diploma ook buiten uw school als waardevol wordt gezien.

Voor al deze doelen is het van belang dat de kwaliteit van de gebruikte toetsen goed is.

De afgelopen tijd zijn er echter vraagtekens gezet bij de in de onderwijspraktijk gerealiseerde kwaliteit van toetsing in alle onderwijssectoren<sup>1,2</sup>. Mede daardoor is de wetenschappelijke aandacht voor de toetskwaliteit verder toegenomen<sup>3</sup>. Meer dan ooit tevoren wordt er geïnvesteerd in kwaliteit van toetsing. Daarbij komen zowel in de onderwijspraktijk als in de wetenschap verschillende vragen op.

- [Wat weten we over de kwaliteit van een toets?](#)
- [Waar hangt de kwaliteit van een toets vanaf?](#)
- [Hoe kan ik als docent een goede toetskwaliteit bereiken?](#)
- [Waar moeten we nog meer over weten?](#)

Deze brochure is bedoeld voor docenten uit alle onderwijssectoren die zich willen oriënteren op de kwaliteit van toetsen. De bovenstaande vragen worden beantwoord aan de hand van de resultaten van een literatuuronderzoek. We hebben hiervoor in bijna 250 boeken, tijdschriften, wetenschappelijke en praktijkgerichte artikelen gezocht wat er over het onderwerp toetskwaliteit geschreven is. Een aantal van de lessen die daaruit geleerd kunnen worden hebben we in deze brochure beschreven. We hopen dat het voor u als docent inzicht geeft in het onderwerp kwaliteit van toetsen en dat het ook behulpzaam kan zijn bij het - daar waar mogelijk - verbeteren van de kwaliteit van toetsen.



## INLEIDING

Het belang van kwalitatief goede toetsen is helder, maar het blijkt moeilijk om te bepalen wat goede kwaliteit is. Er kan gezegd worden dat we in Nederland 16 miljoen deskundigen hebben op toetsgebied: bijna iedereen heeft ervaringen met toetsen en een beeld bij wat goede en slechte toetsen zijn<sup>4</sup>. Zoveel mensen als er zijn, zoveel meningen bestaan er over het begrip kwaliteit. Er kan vanuit veel verschillende invalshoeken naar worden gekeken en verschillende factoren hebben hier invloed op. Een paar voorbeelden:

<p><b>Doel van de toets:</b> In veel gevallen zult u de student willen toetsen voor een eindbeoordeling van zijn niveau. Dit heet <i>summatief toetsen</i>. Het kan echter ook zijn dat u wilt weten welke onderdelen van de leerstof een student beheerst en aan welke onderdelen u nog meer aandacht moet besteden. Dit heet <i>formatief toetsen</i><sup>5</sup>.</p>	<p><b>Onderwijssector:</b> Als docent van het Primair Onderwijs zult u op andere wijze, met een ander doel en andere inhoud toetsen dan als docent van het Hoger Onderwijs<sup>6</sup>.</p>
<p><b>Rollen in de beoordeling:</b> Misschien bent u als docent gewend om zelf een toets te maken, af te nemen en/of te beoordelen. Het kan daarentegen ook zijn dat uw studenten hun eigen werk (<i>self-assessment</i>), elkaars werk (<i>peer-assessment</i>) of in overleg met u als docent (<i>co-assessment</i>) hun werk beoordelen<sup>7,8</sup>.</p>	<p><b>Toetscyclus:</b> Op het moment dat u een toets wilt <i>construeren</i>, bent u op andere aspecten gericht dan op het moment dat u een toets <i>afneemt</i> of <i>beoordeelt</i><sup>9</sup>.</p>

U zult zich kunnen voorstellen dat deze factoren bepalend zijn voor de manier waarop naar kwaliteit gekeken wordt. Om te zien wat er wel en niet bekend is zijn 242 geselecteerde wetenschappelijke artikelen, artikelen uit praktijkvakbladen en boeken onderzocht. Ieder van deze teksten is geselecteerd omdat er wordt ingegaan op de kwaliteit van toetsen.

Bij het vergelijken van die teksten valt op dat er verschillende indelingen van kwaliteitsaspecten gebruikt worden<sup>10</sup>. Dit zorgt ervoor dat er voor eenzelfde kwaliteitsaspect ook nog verschillende definities bestaan. Kortom: het is duidelijk dat de kwaliteit van toetsen een lastig onderwerp is, waar op vele manieren naar gekeken kan worden.

In deze brochure staan nu twee kernvragen centraal:

- Wat weten we over de kwaliteit van een toets?
- Waar hangt de kwaliteit van een toets van af?

Bij het beantwoorden van deze vragen hopen we bij te dragen aan een meer gedeelde visie over de kwaliteit van toetsen. De lessen die we daaruit leren kunnen wellicht helpen met de voor de praktijk relevante vraag:

- Hoe kan ik als docent een goede toets kwaliteit bereiken?

De resultaten van dit onderzoek laten zien wat we tot nu toe weten. Over sommige onderdelen zouden we graag meer willen weten. De laatste paragraaf zal ingaan op de vraag:

- Waar moeten we nog meer over weten?



# 1 WAT WETEN WE OVER DE KWALITEIT VAN EEN TOETS?

Deze vraag is heel breed. Om die reden delen we deze op in twee onderliggende vragen. De eerste daarvan gaat over de aspecten van kwaliteit die uit het onderzoek naar voren zijn gekomen en de wijze waarop deze gerangschikt kunnen worden. Met de tweede vraag onderzoeken we hoe vaak de genoemde aspecten voorkomen.



## 1a Welke aspecten van kwaliteit kwamen in de artikelen voor?

Er worden zeer veel aspecten genoemd in de literatuur. De genoemde kwaliteitsaspecten hebben we samengebracht in één schema (zie pagina 14 en 15).

In dit schema zijn de kwaliteitsaspecten opgedeeld in vijf hoofdcategorieën. De hoofdcategorieën zijn: Betrouwbaarheid, generaliseerbaarheid, validiteit, het gebruik van het toetsresultaat en randvoorwaarden. De hoofdcategorie randvoorwaarden bestaat uit voorwaarden om te komen tot toetskwaliteit.

Hoofdcategorie	Voorbeeld
<p><b>Betrouwbaarheid:</b> De mate waarin de scores op een toets consistent, nauwkeurig en reproduceerbaar zijn. In dat geval is het meetresultaat vrij van meetfouten.</p>	<p>Als een student op maandag een toets maakt, zou het resultaat nagenoeg hetzelfde moeten zijn als wanneer hij op dinsdag de toets maakt.</p>
<p><b>Generaliseerbaarheid:</b> De mate waarin datgene wat een student in de toets laat zien (in deze specifieke omstandigheden), ook opgaat in andere omstandigheden.</p>	<p>Als een student Verpleegkunde aantoonde opgaven te beheersen die te maken hebben met het toedienen van vloeibare medicatie, mag er dan vanuit worden gegaan dat deze student voldoende vaardig is op het gebied van verpleegkundig rekenen?</p>
<p><b>Validiteit:</b> De mate waarin de toets meet wat de maker bedoeld heeft ermee te meten. Welke conclusie kunnen we nu trekken uit een toetsresultaat?</p>	<p>Een student die minder taalvaardig is maakt een rekentoets die uit veel verhaalsommen bestaat. Zijn lage score wordt verklaard door zijn slechte rekenvaardigheid. Of heeft hij de taal in de sommen niet goed begrepen en daardoor een lage score behaald?</p>
<p><b>Gebruik van het toetsresultaat:</b> Gaat over de vraag hoe het toetsresultaat wordt verwerkt en wat er vervolgens mee wordt gedaan.</p>	<p>Als een student een score van 50 of minder heeft gehaald, krijgt hij een onvoldoende. Hij krijgt hulp op de onderdelen die hij niet goed heeft gemaakt.</p>
<p><b>Randvoorwaarden:</b> Voorwaarden waaraan een toets moet voldoen om kwaliteit te bereiken.</p>	<p>De studenten moeten op de hoogte zijn van wat er van hen wordt verwacht. Ook moet de toets uitvoerbaar zijn qua tijd en kosten.</p>

Een hoofdcategorie geeft een algemene beschrijving van een aspect van toetskwaliteit. Daar kan echter ook in meer detail naar gekeken worden. Zo bevat een hoofdcategorie een aantal onderliggende begrippen (subcategorieën).

*[De hoofdcategorie betrouwbaarheid bevat onder andere de subcategorie 'test-hertest betrouwbaarheid'. Betrouwbaarheid gaat namelijk over het feit dat de toets min of meer hetzelfde resultaat geeft als de student die op een ander moment nog eens maakt.]*

Elke subcategorie bestaat vervolgens uit een aantal onderdelen.

*[De subcategorie 'test-hertest betrouwbaarheid' bevat onder andere het onderdeel 'eerlijkheid van de toetsafname'. Als de omstandigheden waaronder de toets wordt afgenomen namelijk niet gelijk zijn, dan zal de toets ook niet hetzelfde resultaat geven.]*

Naast de hoofdcategorieën zijn er twee kwaliteitsaspecten die betrekking hebben op alle aspecten in het schema. Om die reden staan deze twee kwaliteitsaspecten aan de zijkanten van het schema. Dit zijn de kwaliteitsaspecten 'theoretische onderbouwing' en 'toetsanalyse'. Het eerste van deze twee aspecten gaat er over dat je moet weten wat, wie en waarom je wilt toetsen. Het tweede aspect gaat over de wijze waarop je alle kwaliteitsaspecten (cijfermatig) kan onderzoeken.

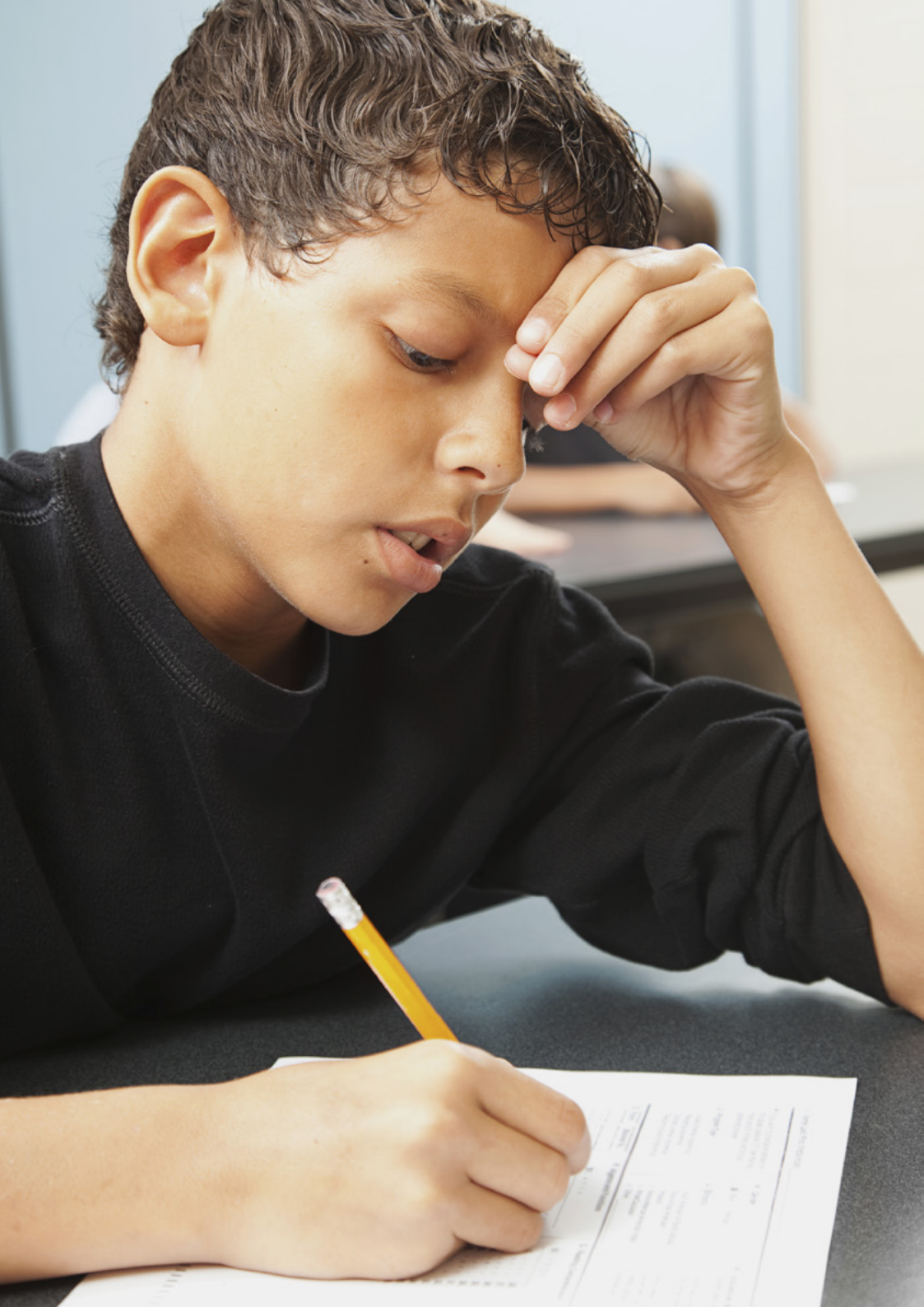
## **1b Welke aspecten van kwaliteit zijn het meest genoemd?**

In het literatuuronderzoek is niet alleen gekeken naar welke kwaliteitsaspecten er genoemd worden, maar ook hoe vaak. De resultaten zijn weergegeven in Tabel 1. In de eerste kolom van de tabel staan alle hoofdcategorieën nogmaals genoemd. Daarnaast staat een getal dat aangeeft hoe vaak deze hoofdcategorie (bestaande uit de verschillende subcategorieën en onderdelen) is genoemd in de verschillende artikelen. Het aantal genoemde hoofdcategorieën is groter dan het aantal artikelen omdat in de meeste artikelen meer dan één onderwerp behandeld wordt.

De hoofdcategorie **betrouwbaarheid** blijkt het meest genoemd te worden. Ongeveer 28% van de genoemde kwaliteitsaspecten valt onder deze hoofdcategorie. Uit de derde en vierde kolom van de tabel blijkt dat vooral **objectiviteit** (14.4%) veel is genoemd. Onderzoek naar objectiviteit is ook weer

Tabel 1. *Het aantal en percentage van de hoofd- en subcategorieën.*

Hoofdcategorie	Aantal (%)	Subcategorie	Aantal (%)
<b>Betrouwbaarheid</b>	188 (28.3%)	Betrouwbaarheid	54 (5.4%)
		Meetnauwkeurigheid	27 (2.7%)
		Objectiviteit	144 (14.4%)
		Test-hertest betrouwbaarheid	60 (6.0%)
<b>Generaliseerbaarheid</b>	76 (11.4%)	Generaliseerbaarheid	24 (2.4%)
		Context van taak/item	48 (4.8%)
		Cognitieve activiteit	48 (4.8%)
<b>Validiteit</b>	146 (22.0%)	Validiteit	73 (7.3%)
		Inhoudsvaliditeit	65 (6.5%)
		Begripsvaliditeit	73 (7.3%)
		Criterium validiteit	21 (2.1%)
<b>Gebruik toetsresultaat</b>	144 (21.7%)	Aanvaardbaarheid	41 (4.1%)
		Consequentiële validiteit	100 (10.0%)
		Normering en cesuur	47 (4.7%)
<b>Randvoorwaarden</b>	110 (16.6%)	Bruikbaarheid	47 (4.7%)
		Transparantie	94 (9.4%)
		Theoretische onderbouwing	20 (2.0%)
		Toetsanalyse	15 (1.5%)
<b>Totaal</b>	<b>664 (100%)</b>	<b>Totaal</b>	<b>1001 (100%)</b>



Item	Quantity	Unit Price	Total Price
1000	100	1.00	100.00
2000	200	2.00	400.00
3000	300	3.00	900.00
4000	400	4.00	1600.00
5000	500	5.00	2500.00
6000	600	6.00	3600.00
7000	700	7.00	4900.00
8000	800	8.00	6400.00
9000	900	9.00	8100.00
10000	1000	10.00	10000.00
Subtotal			50000.00
Tax			5000.00
Total			55000.00

uit te splitsen in verschillende onderwerpen (niet getoond in de tabel), waarbij het **beoordelingsvoorschrift** (9.8%) en de **deskundigheid van de beoordelaar** (7.8%) het meest genoemd zijn. Ook hier zijn er weer veel artikelen die meerdere onderwerpen behandelen.

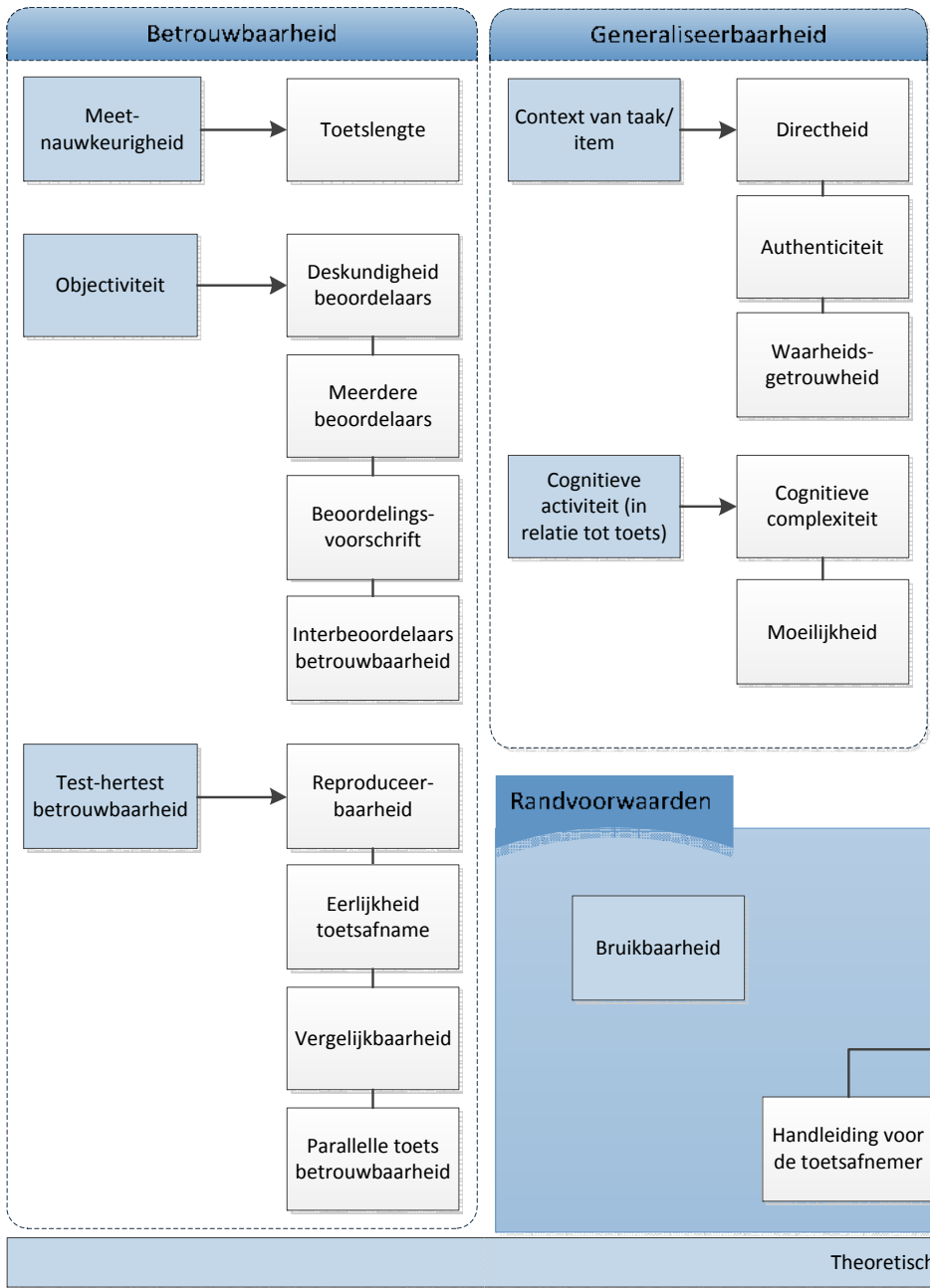
Een andere veelgenoemde subcategorie is **consequentiële validiteit** (10.0%). Dit zijn de bedoelde en onbedoelde gevolgen van een toetsresultaat op het gedrag van docenten en studenten. Ook komt **transparantie** (9.4%) veel voor, met name de **transparantie van het scoren van een toets** (6.8%).

Omdat deze categorieën veelgenoemd worden, kunnen deze als belangrijk worden gezien. Dit hoeft echter niet zo te zijn. Misschien komen deze categorieën veel voor omdat ze makkelijk te onderzoeken zijn. Daarnaast blijkt de hoofdcategorie **betrouwbaarheid** meer subcategorieën te bevatten en is de kans op een hogere score dus groter. De hoofdcategorieën **generaliseerbaarheid** en **randvoorwaarden** worden het minst vaak genoemd, maar bevatten ook minder subcategorieën.

### **Samenvatting**

De voorkomende kwaliteitsaspecten zijn in de volgende hoofdcategorieën in te delen: Betrouwbaarheid, generaliseerbaarheid, validiteit, het gebruik van het toetsresultaat en randvoorwaarden.

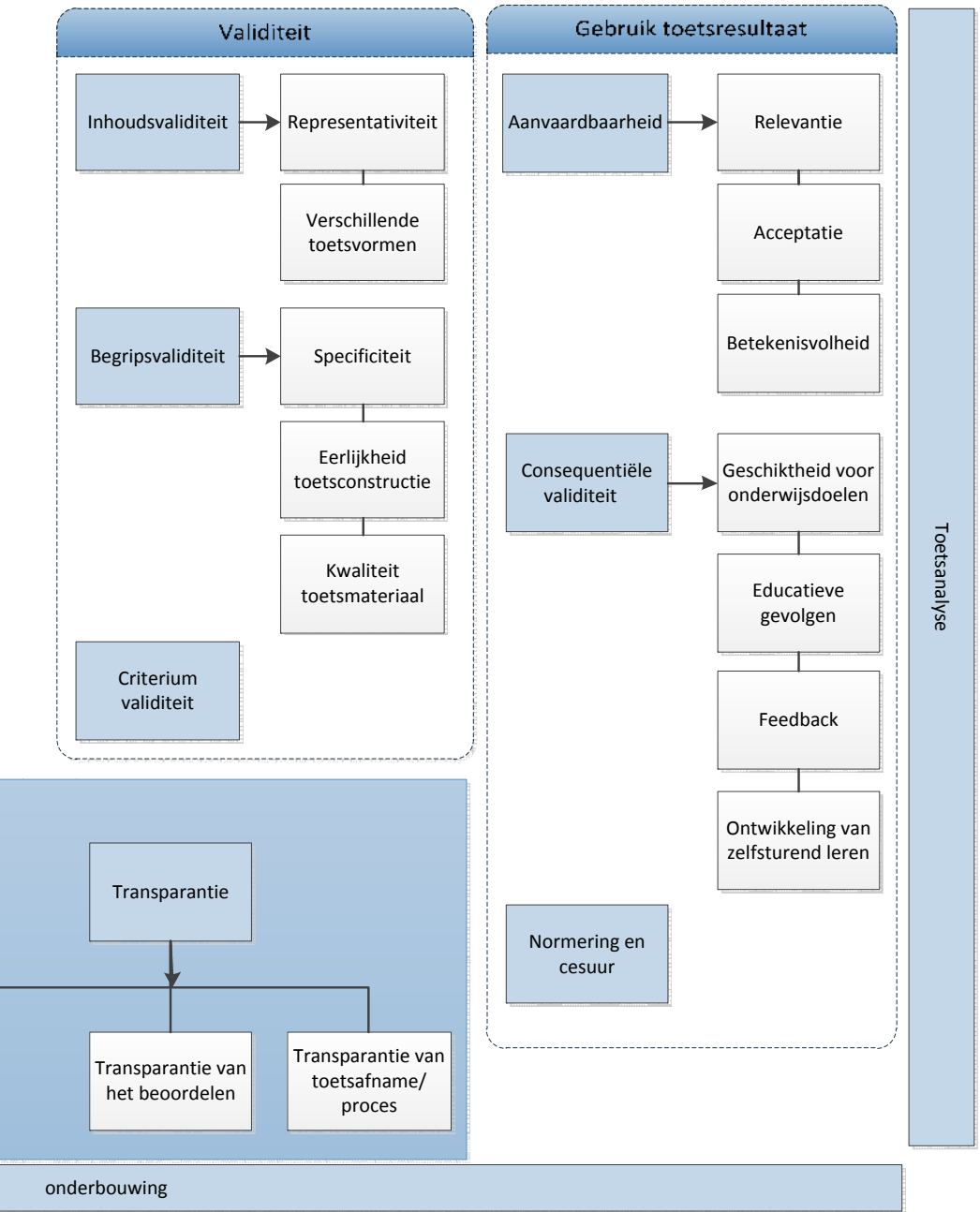
In de literatuur wordt de hoofdcategorie betrouwbaarheid het meest genoemd, met name de subcategorie objectiviteit en de onderdelen beoordelingsvoorschrift en deskundigheid van de beoordelaar.



= hoofdcategorie

= subcategorie

= onderdeel



Figuur 1. Aspecten van toetskwaliteit volgens de geselecteerde literatuurbronnen.

## 2 WAAR HANGT DE KWALITEIT VAN EEN TOETS VANAF?

Moet een toets onder alle omstandigheden aan dezelfde kwaliteitsaspecten voldoen? Of maakt het uit waar je de toets voor gebruikt (toetsdoel), in welke onderwijssector je toetst, wat de rol van de beoordelaar is of van welke fase in de toetscyclus sprake is?

In dit hoofdstuk gaan we hier verder op in. Daarnaast is er onderzocht of er verschil is tussen artikelen die een onderzoek hebben gedaan in de praktijk of artikelen die een theoretische beschrijving geven. Dit resultaat vindt u als laatste in dit hoofdstuk.

### 2a Doel van de toets

Een toets kan om verschillende redenen ingezet worden. Als u een eindoordeel wilt geven over het niveau van de student (summatief toetsdoel), vindt u misschien wel dat de toets van andere kwaliteit moet zijn dan als u tussentijds wilt weten op welk niveau de student zich nu bevindt en waar u nog verdere instructie over moet geven (formatief toetsdoel).

Er was inderdaad verschil tussen de genoemde kwaliteitsaspecten bij de twee toetsdoelen. Zo is **validiteit** vaker genoemd bij summatieve toetsing. Dat is logisch, want bij summatieve toetsen zijn de gevolgen van verkeerde beslissingen vaak ernstiger dan bij formatieve toetsen en bovendien vaak onomkeerbaar. Denk hierbij bijvoorbeeld aan de eindexamens op het Voortgezet Onderwijs<sup>11</sup>. Aangezien er belangrijke beslissingen worden genomen, is het verder van belang dat de **normering en cesuur** op een juiste manier vastgesteld en gehanteerd worden. Hoeveel punten moet een student bijvoorbeeld halen voor een voldoende om zo aan te tonen dat hij bekwaam is?

*PS. Niet elk artikel heeft vermeld wat het doel van de toets in dit onderzoek was. Er zijn 88 artikelen die op één van beide doelen ingingen, waarvan 64 artikelen summatief.*

Bij formatieve toetsing lijkt **consequentiële validiteit** vooral van belang te zijn. De toetskwaliteit wordt bepaald door wat de student en docent met het resultaat doen: Geeft u feedback? Wat, hoe en wanneer? Gaat de student er mee aan de slag? Of wordt het toetsresultaat slechts voor kennisgeving aangenomen en verdwijnt het vervolgens onder in een (digitale) lade?

Kortom: Het toetsdoel is mede bepalend voor de aspecten waar naar gekeken wordt en aan welke eisen een goede toets moet voldoen.

## **2b      Onderwijssector**

In dit onderzoek kwamen geen verschillen naar voren tussen de onderwijssectoren als het gaat om aspecten van de toetskwaliteit. Over het algemeen spelen in de verschillende onderwijssectoren dus dezelfde kwaliteitsaspecten een rol. Wel blijkt uit de artikelen en uit gesprekken met docenten dat de aandacht voor toetskwaliteit enorm verschilt: van relatief weinig aandacht in het Primair Onderwijs tot veel aandacht in het Hoger Onderwijs.

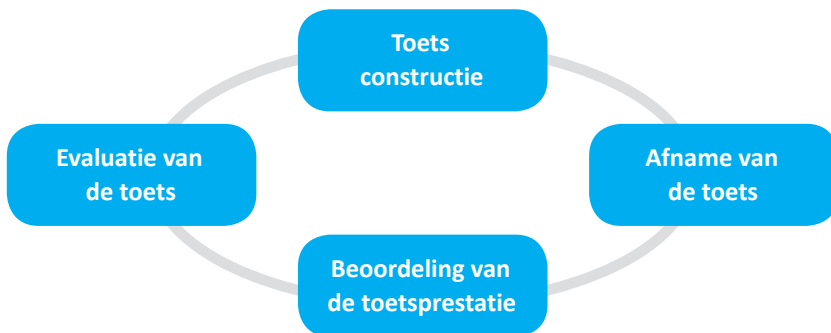
## **2c      Rollen in de beoordeling**

In de meeste gevallen is nog steeds de docent diegene die het werk van studenten beoordeelt (de docent in de rol van de beoordelaar). Er is in dit onderzoek weinig gevonden over alternatieve beoordelingsmethodieken zoals peer-assessment (studenten beoordelen elkaar), self-assessment (studenten beoordelen zichzelf) of co-assessment (student en docent beoordelen samen de student). De enkele gevonden artikelen hierover gaan over toetsing in het Hoger Onderwijs, maar uit gesprekken met verschillende HBO-docenten blijkt dit nog geen grote rol te spelen. Op dit moment zijn daarom nog geen conclusies te trekken over de invloed van de rol in de beoordeling op de toetskwaliteit.

## **2d      Toetscyclus**

Voordat een toets kan worden afgenomen, moet deze eerst geconstrueerd worden. Na de afname wordt de toetsprestatie beoordeeld. Dan zijn er nog een aantal beoordelingssystemen die de kwaliteit van een toets voor- of achteraf evalueren. De toetsing bestaat dus uit verschillende fasen (zie Figuur 2).

Het meeste onderzoek uit de geselecteerde artikelen heeft zich gericht op de fase van afname van een toets en beoordeling van de toetsprestatie. In deze fase is **betrouwbaarheid**, en vooral de **objectiviteit** van de beoordelaar vaak genoemd. De beoordelaar moet **deskundig** zijn en een helder **beoordelings-schema** gebruiken. Deze aspecten zijn tijdens de evaluatiefase minder van belang.



*Figuur 2.* Toetscyclus

*PS. Bij de constructiefase van de toetscyclus zijn geen opvallende verschillen gevonden. Dit kan komen doordat slechts weinig artikelen aandacht hebben besteed aan deze fase.*

## 2e In de praktijk

Ruim 60% van de artikelen uit dit onderzoek gaf een theoretische beschrijving over toetskwaliteit, zonder de praktijk te onderzoeken. Deze artikelen kijken naar andere kwaliteitsaspecten dan de artikelen waarbij sprake is van observaties, toetsanalyses, interviews of vragenlijsten, waarbij er wel gekeken wordt naar wat er werkelijk in een school gebeurt.

De theoretische beschrijvingen noemden vooral **validiteit**: de toets meet wat deze zou moeten meten. Ook **generaliseerbaarheid**, **criteriumvaliditeit** en **theoretische onderbouwing** zijn vaak genoemd.

Bij artikelen waarin praktijkonderzoek was uitgevoerd, ging het vooral over de **aanvaardbaarheid** van een toets. Is deze geaccepteerd door de studenten? En is



deze betekenisvol voor hen en de docenten? Voor een optimaal toetsresultaat is het van belang dat de toets door de betrokkenen als nuttig ervaren wordt.

Beide soorten artikelen gaan in op **betrouwbaarheid** en aspecten die daarmee samenhangen, maar deze worden in de praktijkartikelen het meest genoemd. Validiteit blijkt daarentegen vooral een theoretisch begrip te zijn.

### **Samenvatting**

Welke aspecten van toetskwaliteit vooral belangrijk zijn, hangt af van:

- Het toetsdoel
- De fase in de toetscyclus
- Of het artikel een theoretische beschrijving of een praktijkonderzoek is

Er blijken geen kenmerkende verschillen tussen:

- De onderwijssectoren
- De rollen in de beoordeling



### 3 HOE KAN IK ALS DOCENT EEN GOEDE TOETSKWALITEIT BEREIKEN?

We hebben nu de antwoorden op de gestelde vragen, maar wat betekent dat voor een docent? In het beste geval geeft het een aantal aangrijpingspunten om rekening mee te houden bij het maken, afnemen en beoordelen van toetsen. Een paar voorbeelden:

Resultaat	Aangrijpingspunten voor de docent
1 Betrouwbaarheid en objectiviteit zijn belangrijke aspecten.	<ul style="list-style-type: none"><li>- Zorg ervoor dat u gebruik maakt van een beoordelingsvoorschrift bij het nakijken.</li><li>- Laat uw collega ook eens een paar toetsen nakijken en controleer of jullie op dezelfde score uitkomen.</li></ul>
2 Transparantie van een toets is veelgenoemd.	<ul style="list-style-type: none"><li>- Stel uw studenten op de hoogte van de inhoud, vorm en moment van de toets. Geef ook informatie over de manier van beoordelen.</li></ul>
3 De kwaliteitsaspecten hangen af van het toetsdoel.	<ul style="list-style-type: none"><li>- Bepaal met welk doel u toetst: summatief of formatief. Bij summatief toetsen bepaalt u secuur de normering en cesuur: wanneer is de student geslaagd? Bij formatief toetsen houdt u vooral rekening met consequentiële validiteit: geef feedback en stel het onderwijs af op het resultaat van de toets.</li></ul>

Om docenten te ondersteunen bij het ontwerpen, afnemen en beoordelen van toetsen, wordt een hulpmiddel ontwikkeld. Zodra dit beschikbaar is, vindt u het op onze website.

## 4 WAAR MOETEN WE NOG MEER OVER WETEN?

De resultaten van dit onderzoek laten zien wat we tot nu toe weten. Over sommige dingen zouden we juist graag meer willen weten. Een paar aanbevelingen voor de onderzoekers:

- De fase van de toetscyclus blijkt bepalend voor welke kwaliteitsaspecten relevant zijn. Kan de toetskwaliteit verbeterd worden als bekend is welke kwaliteitsaspecten precies passen binnen de verschillende fasen in de toetscyclus?
- Validiteit komt nauwelijks voor in de praktijkgerichte artikelen, terwijl dit kwaliteitsaspect in theoretische beschrijvingen juist heel vaak wordt beschreven. Hoe kan het validiteitsaspect vertaald worden naar de praktijk? Hoe kunnen problemen met betrekking tot dit aspect in de praktijk verdwijnen?
- In het meeste onderzoek wordt er gekeken naar de gehele toets (het toetsniveau). Er kan ook meer gedetailleerd worden ingezoomd op de opgaven (itemniveau) of juist worden uitgezoomd naar het toetsprogramma-niveau, omdat toetsen of metingen vaak worden gecombineerd om tot een beoordeling te komen. Daarnaast geldt dat toetsen verweven kunnen worden in het leerproces door ze bijvoorbeeld meer formatief in te zetten. Hoe kan een toetsprogramma zo effectief mogelijk worden ingericht om dit op een juiste manier uit te voeren?

### *PS. Wilt u meer weten?*

- *Lees de hele rapportage: meer achtergrondinformatie over de methode en een uitgebreider verslag van de resultaten.*
- *Houd onze website in de gaten voor een praktisch hulpmiddel ter ondersteuning bij het ontwerpen, afnemen en beoordelen van toetsen.*
- *Contact met RCEC?*  
*[www.rcec.nl](http://www.rcec.nl) of [rcec@utwente.nl](mailto:rcec@utwente.nl)*

## LITERATUUR

- <sup>1</sup> Inspectie van het Onderwijs (2009). *Boekhouder of wakend oog. Verslag van een onderzoek bij examencommissies in het hoger onderwijs over de garantie van het niveau*. Inspectierapport 2009-16 (april).
- <sup>2</sup> Onderwijsraad (2006). *Advies examinering: draagvlak en toegankelijkheid, uitgebracht aan de staatssecretaris van Onderwijs, Cultuur en Wetenschap. Nr. 20060320/865*. Den Haag (november).
- <sup>3</sup> Joosten-ten Brinke, D., & Sluijsmans, D. M. A. (2012). Tijd voor toetskwaliteit: het borgen van toetsdeskundigheid van examencommissies. *TH&MA*, 19(4), 16-21. Verkregen van: <http://hdl.handle.net/1820/4759>
- <sup>4</sup> Eggen, T. J. H. M. (2009). *De kwaliteit van Toetsen*. Oratie Universiteit Twente, 9 april 2009.
- <sup>5</sup> Eggen, T. J. H. M. (2013). Computerized adaptive testing serving educational testing purposes. Paper presented at IAEA Conference, Tel Aviv, Israel.
- <sup>6</sup> Straetmans, G. J. J. M. (2006). *Bekwaam beoordelen en beslissen*. Lectorale rede. Deventer: Saxion Hogescholen.
- <sup>7</sup> Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer- and co-assessment in higher education: A review. *Studies in Higher Education*, 24(3), 331-350. doi:10.1080/03075079912331379935
- <sup>8</sup> Ploegh, K., Tillema, H. H., & Segers, M. S. R. (2009). In search of quality criteria in peer assessment practices. *Studies in Educational Evaluation*, 35, 102-109. doi:10.1016/j.stueduc.2009.05.001
- <sup>9</sup> Tillema, H., Leenknecht, M., & Segers, M. (2011). Assessing assessment quality: Criteria for quality assurance in design of (peer)assessment for learning – A review of research studies. *Studies in Educational Evaluation*, 37, 25-34. doi:10.1016/j.stueduc.2011.03.004
- <sup>10</sup> Poldner, E., Simons, P. R. J., Wijngaards, G., & Schaaf, M. F. van der (2012). Quantitative content analysis of procedures to analyse students' reflective essays: A methodological review of psychometric and edumetric aspects. *Educational Research Review*, 7(1), 19-37. doi:10.1016/j.edurev.2011.11.002
- <sup>11</sup> Tanilon, J., Segers, M., Vedder, P., & Tillema, H. (2009). Development and validation of an admission test designed to assess samples of performance on academic tasks. *Studies in Educational Evaluation*, 35, 168-173. doi:10.1016/j.stueduc.2009.12.003

## COLOFON

De literatuurreview *'Kwaliteit van toetsen binnen handbereik, een reviewstudie van onderzoek naar kwaliteit van toetsen'* is gefinancierd door de NRO Programmaraad voor Praktijkgericht Onderwijsonderzoek.

Nathalie Maassen, RCEC, Universiteit Twente

Dorien den Otter, RCEC, Universiteit Twente

Saskia Wools, Cito

Bas Hemker, Cito

Gerard Straetmans, Saxion Hogescholen & Cito

Theo Eggen, RCEC, Universiteit Twente & Cito

[www.rcec.nl](http://www.rcec.nl)

