

# De kwaliteit van toetsing onder de loep

Dit artikel is het vijfenvijftigste in een serie praktische artikelen over onderwijsinnovatie. Deze serie heeft de bedoeling om mensen die werkzaam zijn in het hoger onderwijs handreikingen en aandachtspunten te bieden voor eigen initiatieven in onderwijsinnovatie. De onderwerpen van deze reeks kunnen uiteenlopen, maar zullen altijd gaan over het maken van onderwijs en dus over toepassingen van onderwijskundige en onderwijstechnologische inzichten in het dagelijks werk van de docent, het onderwijsteam of de opleidingsmanager.

## Auteur

Dominique Sluijsmans  
Annemieke Peeters  
Linda Jakobs  
Saskia Weijzen

Sluijsmans was tot 1 mei 2012 lector Duurzaam beoordelen bij de Hogeschool van Arnhem en Nijmegen en is nu lector Professioneel beoordelen bij Zuyd Hogeschool. Peeters en Jakobs zijn verbonden aan het Service Centrum Onderwijs van de Hogeschool van Arnhem en Nijmegen (HAN), Weijzen is verbonden aan Maatschappelijk Werk en Dienstverlening van de HAN.

Reacties op dit artikel kunnen gemaild worden naar:  
dominique.sluijsmans@zuyd.nl of  
annemieke.peeters@han.nl.

## Inhoud

- \_ Samenvatting
- \_ Aanleiding voor het project De toetsing getoetst
- \_ Waarom is goed toetsen zo belangrijk?
- \_ Wat betekent kwaliteit van toetsing?
- \_ Hoe ziet het fundament van de methodiek er uit?
- \_ Hoe ziet een eerste ontwerp van een methodiek er uit?
- \_ Welke stappen zijn nodig voor een onderbouwde analyse van toetskwaliteit?
- \_ Hoe kunnen we de validiteit, betrouwbaarheid, bruikbaarheid en transparantie van de methodiek vaststellen?
- \_ Wat is ons beoogde ideaal?
- \_ Tot slot

- Box 1: De kwaliteitspiramide voor eigentijds toetsen en beoordelen
- Box 2: Kwaliteitscriteria voor toetsitems
- Box 3: Kwaliteitscriteria voor toetsen
- Box 4: Kwaliteitscriteria voor het toetsprogramma
- Box 5: Kwaliteitscriteria voor het toetsbeleid
- Box 6: Toetscompetenties van docenten
- Box 7: Het ontwerp kader van de methodiek
- Box 8: Uitwerking van de bouwsteen toetsbekwaamheid

## Samenvatting

Het zal niemand zijn ontgaan dat de kwaliteit van toetsing in het hoger onderwijs ter discussie staat. Opleidingen zijn zich meer dan ooit bewust dat de kwaliteit van toetsing op orde moet zijn om het bestaansrecht van de opleiding te kunnen verzekeren.

In deze bijdrage wordt het project 'De toetsing getoetst' uiteengezet. In dit project wordt een methodiek ontwikkeld en gevalideerd die opleidingen kan ondersteunen bij het analyseren, verbeteren en borgen van toetskwaliteit. Met deze methodiek kunnen opleidingen vaststellen hoe de toetskwaliteit nu is (positiebepaling) en waar men met toetskwaliteit naar toe wil (ambitiebepaling). We hopen dat deze methodiek opleidingen houvast biedt in het formuleren van een eigen visie op toetskwaliteit.

## Aanleiding voor het project 'De toetsing getoetst'

De afgelopen jaren wordt de kwaliteit van toetsing in het hoger onderwijs aan een kritisch oog onderworpen. Vanuit de politiek en de maatschappij worden steeds meer vraagtekens gezet bij de kwaliteit van toetsing (ministerie van OC&W, 2012; Onderwijsraad, 2006). Het feit dat de Inspectie van het Onderwijs (2009) heeft vastgesteld dat bij een aantal opleidingen de wet- en regelgeving rond het afstuderen onvoldoende wordt nageleefd, heeft een extra impuls gegeven aan de noodzaak voor kwaliteitsverbetering van toet-

sing. Ook studenten roeren zich meer en constateren dat toetsing en examinering een onderschoven kindje zijn bij gesprekken en debatten over onderwijskwaliteit (ISO, 2009). Studenten ervaren direct de gevolgen van een toets van voldoende of onvoldoende kwaliteit (Mager & Nowak, 2012). Dit beïnvloedt hun tevredenheid en juist deze tevredenheid wordt door onderwijsinstellingen gebruikt als indicator voor de kwaliteit van hun toetsing (Gerritsen-Van Leeuwenkamp, 2012). De behoefte aan hogere toetskwaliteit is inmiddels omgezet in een aantal concrete acties. In september 2010 is de wet Versterking Besturing in werking getreden. Deze houdt in dat de examencommissie het orgaan is dat op objectieve en deskundige wijze vaststelt of een student voldoet aan de voorwaarden die de onderwijs- en examenregeling (OER) stelt ten aanzien van kennis, inzicht en vaardigheden die nodig zijn voor het verkrijgen van een graad. Daarmee is de rol van examencommissies verschoven van het uitvoeren van procedures rondom toetsing naar het borgen van de kwaliteit van de toetsing. Ook is het nieuwe accreditatiestelsel in januari 2011 in werking getreden waarbij 'Toetsing en gerealiseerde eindkwalificaties' als aparte standaard is opgenomen (standaard 3 van het accreditatiestelsel beperkte opleidingsbeoordeling): *'Opleidingen beschikken over een adequaat systeem van toetsing en tonen aan dat de beoogde eindkwalificaties worden gerealiseerd'*. Opleidingen die onvoldoende scores op toetsing en gerealiseerde eindkwalificaties kunnen een herstelperiode van een jaar toegewezen krijgen. Voor standaard 1 (doelen van de opleiding of eindkwalificaties) geldt dat niet: onvoldoende beoordeling is einde verhaal. Examencommissies moeten zich bij de accreditatie kunnen verantwoorden over het aspect toetsen en beoordelen (Smits, 2011). Recent is het rapport 'Vreemde ogen' gepubliceerd onder leiding van commissie Bruijn. Deze commissie heeft een aantal adviezen geformuleerd over de externe validering van de kwaliteit van

toetsen. Enkele adviezen zijn het gezamenlijk ontwikkelen van landelijke toetsen, certificeren van examinatoren en het borgen van toetsdeskundigheid in visitatiecommissies.

Ondanks de vele plannen en acties zijn er ook veel vragen: Wat betekent kwaliteit van toetsing? Wanneer is het systeem van toetsing adequaat? Hoe goed is ons systeem van toetsen? Waar liggen verbeterpunten? Hoe houden we eigenaarschap over ons systeem van kwaliteitszorg? In dit praktisch artikel delen wij de eerste ervaringen van het project 'De toetsing getoetst', een project van de Hogeschool van Arnhem en Nijmegen (HAN) dat is geïnitieerd door Kenniscentrum Kwaliteit van Leren en Service Centrum Onderwijs en gefinancierd door de faculteit Educatie van de HAN. Het project stelt tot doel via praktijkgericht onderzoek een methodiek te ontwerpen waarmee betrokkenen in het hoger onderwijs (management, examencommissies, ontwikkelaars, docenten, studenten en werkveld) de kwaliteit van toetsing kunnen analyseren, verbeteren en borgen. Op basis van wetenschappelijke en praktische bevindingen over toetskwaliteit en analysemodellen voor kwaliteit is gewerkt aan een prototype van de methodiek. In dit praktisch artikel zijn de volgende vragen leidend:

1. Waarom is goed toetsen zo belangrijk?
2. Wat betekent kwaliteit van toetsing?
3. Hoe ziet het fundament van de methodiek er uit?
4. Hoe ziet een eerste ontwerp van een methodiek er uit?
5. Welke stappen worden gezet in het werken met de methodiek?
6. Hoe kunnen we de validiteit, betrouwbaarheid, bruikbaarheid en transparantie van de methodiek vaststellen?
7. Wat is ons beoogde ideaal?

#### **Waarom is goed toetsen zo belangrijk?**

Je kunt je afvragen waarom er zoveel ophef is over de kwaliteit van toetsen, er is toch jarenlang ook niet over gesproken? Het antwoord is vrij simpel: toetsen worden gebruikt om beslissingen te

nemen over de student. Deze beslissingen kunnen klein zijn, bijvoorbeeld het geven van een extra opdracht, maar ook heel groot, zoals een bindend negatief studieadvies aan het einde van het propeusejaar. Omdat toetsen de functie heeft misclassificatie te voorkomen - voorkomen dat studenten die dat niet verdienen een diploma krijgen en andersom - is het van groot belang dat toetsen met de grootste zorgvuldigheid worden ontworpen, uitgevoerd en geïnterpreteerd zodat de juiste beslissingen worden genomen. Daarbij kan als een soort stelregel worden genoemd dat hoe meer er op het spel staat, hoe meer er in de kwaliteit van toetsing moet worden geïnvesteerd (Van der Vleuten & Schuwirth, 2011).

Om tot goede beslissingen te komen waarbij het uitreiken van het diploma de belangrijkste is, is het louter verzamelen van een reeks toetsbriefjes per docent en per vak ter afronding niet meer toereikend om het hbo-niveau aan te tonen. Er zal er sprake moeten zijn van een sterk samenhangend en gebalanceerd toetsprogramma dat aansluit bij het onderwijsprogramma. Dit heeft behoorlijke implicaties voor de inrichting van het onderwijs. Een eerste implicatie is een sterke uitbreiding aan toetsvormen.

Voorbeelden zijn de generieke toepassing van (digitale) portfolio's, proeven van bekwaamheid en simulaties. Een tweede implicatie is dat toetsing een herijking van de kwaliteitscriteria vraagt. Omdat toetsing integraal onderdeel wordt van een leertraject zal naast de gangbare psychometrische criteria als betrouwbaarheid en validiteit een uitbreiding van criteria nodig zijn. De laatste jaren is dan ook (inter)nationaal geïnvesteerd in deze uitbreiding waarbij aansluiting met het leerproces een prominentere plek krijgt. Drie voorbeelden hiervan zijn het kwaliteitskader van Baartman (2008), de zeven principes van Assessment 2020 (Boud and associates, 2010) en de tien kenmerken van duurzaam toetsen (Sluijsmans, 2010). Een derde implicatie is dat de inzichten over de toetsing een uitbreiding

van kennis en kunde over toetsing vraagt van opleiders die zij vanuit hun eigen opleiding niet hebben vergaard. Een schriftelijke toets kun je nog valide en betrouwbaar maken door itemanalyse of een goede handleiding, maar zeker bij de nieuwe vormen van toetsen hangt de kwaliteit veel meer af van de uitvoerder. Omdat toetsing niet meer los staat van het onderwijs en de kwaliteit van de toetsing steeds meer moet worden verantwoord, zal de opleider zich moeten professionaliseren in de nieuwe benaderingen van toetsing, bijvoorbeeld het analyseren van een portfolio en het construeren van goede toetsitems. Ook zal de opleider moeten worden voorzien van handvatten om de toetsing beheersbaar, transparant en eerlijk te houden.

#### Wat betekent kwaliteit van toetsing?

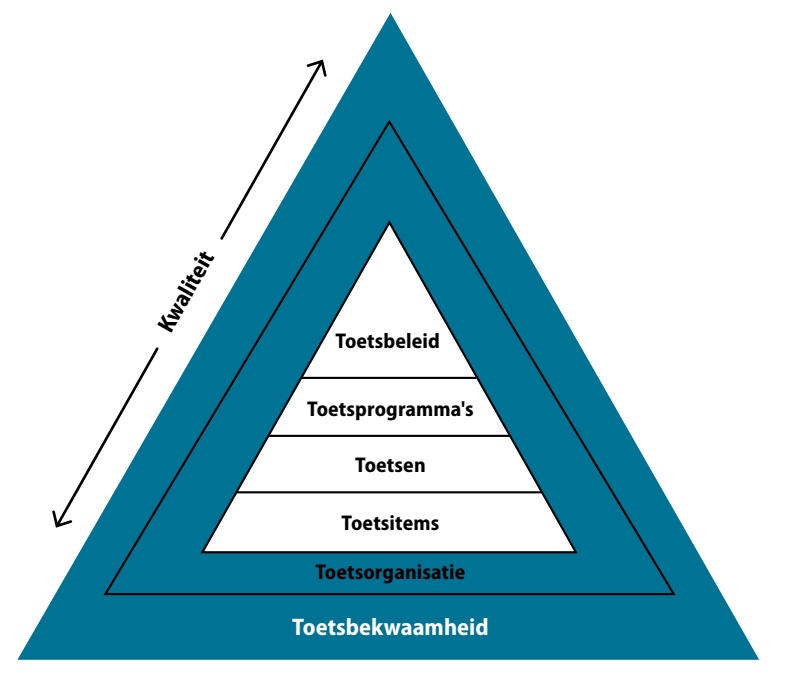
Hoewel kwaliteit van toetsing wordt gezien als essentieel, is het voor velen niet duidelijk wat dat dan precies inhoudt. Kwaliteit is een veelgehoord maar vaak slecht gedefinieerd begrip. In het algemeen wordt kwaliteit gebruikt in de betekenis van 'goede hoedanigheid', 'deugdelijkheid' of 'zorgvuldigheid'. Om grip te krijgen op de vraag wat toetskwaliteit betekent, nemen we de kwaliteitspiramide van eigentijds toetsen en beoordelen van Joosten-ten Brinke, Sluijsmans & Peeters (2012) als uitgangspunt, zie box 1.

In deze piramide wordt kwaliteit van toetsing in zes belangrijke bouwstenen vertaald: kwaliteit van toetsitems, kwaliteit van toetsen, kwaliteit van toetsprogramma's, kwaliteit van het toetsbeleid, kwaliteit van de toetsorganisatie en kwaliteit van toetsbekwaamheid. Vervolgens hebben we bronnen geraadpleegd die invulling geven aan deugdelijkheid en zorgvuldigheid als parameters voor kwaliteit. In het volgende wordt per bouwsteen een aantal kwaliteitscriteria benoemd.

#### Kwaliteit van toetsitems

Als het gaat om de kwaliteit van toetsitems is veel informatie voorhanden om de kwaliteit van open en gesloten items

### BOX 1: DE KWALITEITSPIRAMIDE VOOR EIGENTIJD'S TOETSEN EN BEOORDELEN



### BOX 2: KWALITEITSCRITEIA VOOR TOETSITEMS

- Relevantie:** De relevantie van een item heeft betrekking op de mate waarin een item toetst wat het beoogt te toetsen. Met andere woorden: op de mate waarin het item een directe relatie heeft met het leerdoel.
- Objectiviteit:** Een item is objectief indien de score van het item onafhankelijk is van degene die de scoring uitvoert. Dit impliceert dat de procedure voor het beoordelen van het antwoord op een item zo is geregeld dat geen beslissingen aan het oordeel van de beoordelaar worden overgelaten en het antwoordmodel geen ruimte voor interpretatie openlaat en deskundigen het eens zijn over het juiste antwoord of over de criteria waaraan een juist antwoord moet voldoen.
- Efficiëntie:** Wanneer men bij het ontwikkelen van een item kan kiezen uit verschillende soorten items die in alle opzichten gelijkwaardig zijn, dan verdient het item dat het minste tijd kost in termen van afname en correctie de voorkeur.
- Moeilijkheid:** De moeilijkheidsgraad van een item moet aanvaardbaar zijn, gegeven het niveau van de studenten. Daarom willen we vooral de zogenaamde grensstudenten nauwkeurig van elkaar onderscheiden. Dit bereiken we door deze studenten opgaven voor te leggen die goed bij hun niveau aansluiten.

### BOX 3: KWALITEITSCRITERIA VAN TOETSEN

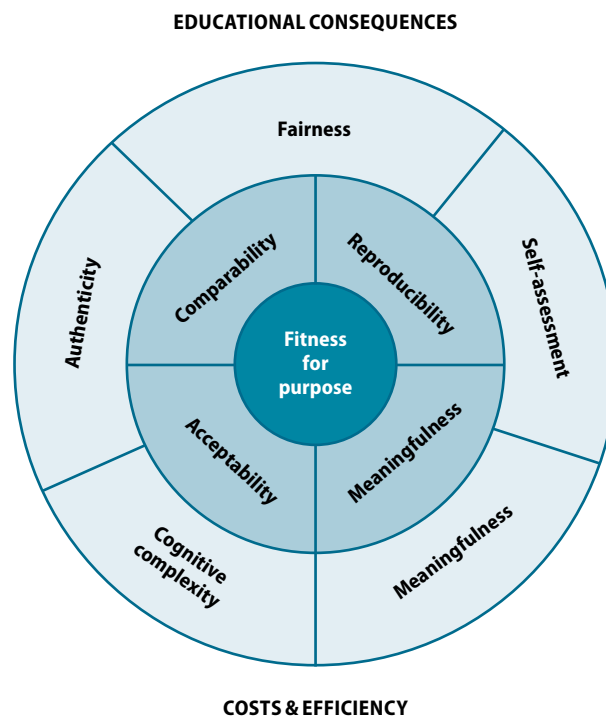
Validiteit:	De validiteit van de toets heeft te maken met de betekenis, de bruikbaarheid en de juistheid van de conclusies die uit de toetsscores worden getrokken. Voorbeelden zijn: inhoudvaliditeit, begripsvaliditeit en criteriumvaliditeit.
Betrouwbaarheid:	De betrouwbaarheid van de toets geeft aan in hoeverre er vertrouwen kan zijn in de toets als meting, ongeacht de inhoud van de toets.
Objectiviteit:	De meest gangbare betekenis van objectiviteit is intersubjectiviteit: is de toetsuitkomst onafhankelijk van storende invloeden van beoordelaars?
Transparantie:	Transparantie wil zeggen dat alle informatie voorhanden is die de best denkbare voorbereidings- en antwoordstrategie van studenten mogelijk maakt.
Normering:	Het doel van normering is het hard kunnen maken van beslissingen. Met name de relevantie of representativiteit van de toets speelt daarbij een rol.

kunnen van de student zoals de toets beoogt te meten. Om vast te stellen of een toets rechtvaardig is, onderscheiden zij vijf kwaliteitscriteria (zie box 3). Het betreft validiteit, betrouwbaarheid, objectiviteit, transparantie en normering. In box 3 zijn de kwaliteitscriteria beschreven.

#### **Kwaliteit van het toetsprogramma**

De laatste jaren is (inter)nationaal geïnvesteerd in een uitbreiding van criteria waarbij aansluiting met het leerproces en andere toetsen in het programma een prominentere plek krijgt. Baartman (2008) heeft twaalf criteria benoemd om de kwaliteit van toetsing op programmaniveau te analyseren (zie box 4; uit Gulikers en anderen, 2008).

### BOX 4: TWAALF KWALITEITSCRITERIA VOOR HET TOETSPROGRAMMA



Een afzonderlijke toets kan niet aan alle kwaliteitscriteria voldoen. Elke toets zal een compromis zijn tussen deze kwaliteiten. Als toetsen deel uitmaken van een groter geheel, een toetsprogramma, kan de kwaliteit van een individuele toets binnen het totaal van de kwaliteit van het toetsprogramma worden beschouwd. De bovenstaande compromissen binnen een toets worden anders gesloten als toetsen een plaats en functie hebben binnen een geheel van een onderwijsprogramma (Van der Vleuten & Schuwirth, 2005). Niet de enkele toets moet dan maximaal voldoen aan de toetskwaliteiten, maar het toetsprogramma als geheel. Daarbij geldt: de som is meer dan het geheel van de delen.

#### **Kwaliteit van toetsbeleid**

In de punt van de piramide staat de kwaliteit van het toetsbeleid. Hier gaat het om de mate waarin de zaken in box 5 door de opleiding zijn beschreven (voorbeeld van de Hogeschool Arnhem Nijmegen, 2012).

#### **Kwaliteit van de toetsorganisatie**

Vervolgens is de hele organisatie van het toetsbouwwerk belangrijk om toetskwaliteit te kunnen garanderen. Dit komt terug in de criteria als kosten en haalbaarheid

te borgen (zie bijvoorbeeld [www.toetsen-opschool.nl](http://www.toetsen-opschool.nl) of [www.toetswijzer.nl](http://www.toetswijzer.nl)). Belangrijke criteria voor toetsitems zijn weergegeven in box 2.

#### **Kwaliteit van toetsen**

Op het niveau van toetsen beschrijven Bax en van Berkel (2002) dat een toetsuitslag een eerlijke en rechtvaardige uitspraak moet doen over het kennen en

(Baartman, 2008). Toetsorganisatie gaat over zaken als het vaststellen van de rollen, taken en verantwoordelijkheden van betrokkenen bij toetsing, waaronder toets- en examencommissies, de werkprocessen rondom de toetsing, de logistieke organisatie van de toetsing en het vaststellen van toetsroosters.

#### **Kwaliteit van de toetsbekwaamheid**

Last but not least is de toetsbekwaamheid van docenten, examencommissies en andere betrokkenen van essentieel belang, aangezien kwaliteit van toetsing in hoge mate wordt bepaald door de kwaliteit van de beoordelaar.

Toetsdeskundigen in het hoger onderwijs constateren dat het met de kennis en kunde van opleiders over toetsen niet goed is gesteld (Rixsen, 2010). Deze mening is niet ongegrond: Docenten in het hoger onderwijs zijn vaak vakexperts en zijn niet of nauwelijks opgeleid in toetsing en hebben onvoldoende kennis van de basisprincipes van toetsen en beoordelen (ISO, 2009; Straetmans, 2006). Dat is zorgelijk, omdat het vaststellen van een betrouwbaar beeld van de kennis en kunde van de student en het nemen van vaak ingrijpende beslissingen over de studieloopbaan van studenten, kwalitatief goede beoordelaars vereist. Elke docent moet in staat zijn om toetsinformatie te interpreteren en te gebruiken om vast te stellen waar onze studenten staan en hoe zij verder kunnen bijdragen aan het leren van de student (Absolum, Flockton, Hattie, Hipkins, & Reid, 2009). Brookhart (2009) heeft toetscompetenties van docenten beschreven (box 6). Deze toetscompetenties van docenten passen bij de vernieuwde visie op toetsen en beoordelen waarbij de rol van de student meer centraal staat.

#### **Hoe ziet het fundament van de methodiek er uit?**

Nadat we toetskwaliteit een nadere invulling hadden gegeven (zie box 1), zijn we op zoek gegaan naar een kader dat als basis kon dienen voor de ontwikkeling van de methodiek om toetskwaliteit te

#### **BOX 5: KWALITEITSCRITERIA VOOR TOETSBELEID**

- Wat is de aanleiding, de doelstelling en de reikwijdte van het toetsbeleidsplan?
- Wie zijn de belanghebbenden van dit toetsbeleidsplan?
- Wat zijn de uitgangspunten van toetsbeleid en de borging van de kwaliteit?
- Wat is de relatie met de visie op leren en toetsen?
- Wat is de samenhang van toetsing met beroepskwalificaties en de onderwijsleeromgeving?
- Wie heeft taken, bevoegdheden en verantwoordelijkheden met betrekking tot de uitvoering van toetsing?
- Hoe ziet de toetscyclus eruit: ontwikkeling (toets(vorm), cesuur), overeenstemming over criteria en standaarden, uitvoering, evaluatie, verbetering?
- Hoe wordt validiteit en betrouwbaarheid geborgd en aangetoond?
- Hoe faciliteert het management de toetscyclus en professionalisering?
- Wie heeft taken, bevoegdheden en verantwoordelijkheden met betrekking tot het beleid en de uitvoering van kwaliteitszorg (op het secundaire niveau)?
- Hoe werkt de borging en verbetering van kwaliteit (szorgsysteem) rondom toetsing?

#### **BOX 6: TOETSCOMPETENTIES VAN DOCENTEN**

- Docenten moeten begrijpen wat leren is in de context waarin ze les geven.
- Docenten moeten in staat zijn om toegankelijke en meetbare leerdoelen te formuleren die aansluiten bij de gewenste inhoud en beheersingsniveau volgens de curriculum doelen en visies.
- Docenten moeten beschikken over verschillende strategieën om met studenten te communiceren over de leerdoelen en hoe deze bereikt kunnen worden.
- Docenten moeten de bedoelingen en het gebruik van verschillende beoordelingsmethodes kennen en vaardig zijn om deze te gebruiken.
- Docenten moeten kunnen vaststellen dat de toetsvormen passen bij de te meten kennis en vaardigheden van hun studenten.
- Docenten moeten effectieve en bruikbare feedback op studentenwerk kunnen geven.
- Docenten moeten in staat zijn om beoordelingsmodellen te construeren die leiden tot bruikbare informatie om beslissingen te kunnen nemen over studenten en/of klassen. Deze beslissingen moeten leiden tot verbeteringen van het leren, groeien of ontwikkelen van de student.
- Docenten moeten externe toetsresultaten kunnen gebruiken en op de juiste manier interpreteren ten behoeve van de juiste beslissingen.
- Docenten moeten hun interpretaties van resultaten en beslissingen kunnen communiceren naar de studenten, schoolleiding en ouders.
- Docenten moeten hun studenten kunnen helpen om op basis van toetsinformatie beslissingen te nemen.
- Docenten zijn zich bewust van hun wettelijke en ethische verantwoordelijkheden bij het beoordelen van studenten.

analyseren. Na vergelijking van een aantal kwaliteitsmodellen bleek het auditinstrument van de Stichting Duurzaam Onderwijs (SDO; 2010) het beste aan te sluiten bij onze visie en doelen. Met dit instrument is het mogelijk te meten in hoeverre opleidingen er in slagen het thema duurzame ontwikkeling in te bedden in hun

onderwijs. Belangrijke uitgangspunten voor de methodiek zijn eigenaarschap over de uitkomsten van de analyse en het stimuleren van de motivatie verbeteringen te willen implementeren. De basis voor het auditinstrument is een model voor kwaliteitsmanagement, ontwikkeld door de European Foundation for Quality

## BOX 7: HET ONTWERPKADER VAN DE METHODIEK

		VERBETERSTAPPEN				
		Minimale kwaliteit 1	2	3	4	Ideale kwaliteit 5
PDCA	Kwaliteit toetsbeleid					
PDCA	Kwaliteit toetsprogramma					
PDCA	Kwaliteit toetsen					
PDCA	Kwaliteit Toetsitems					
PDCA	Kwaliteit Toetsorganisatie					
PDCA	Toets-bekwaamheid					

Management (EFQM), en uitgebreid door het Instituut Nederlandse Kwaliteit (INK), het EFQM-INK-model (Oakland, 1993). Uitgangspunt voor dat model is dat een organisatie zich ten aanzien van een aantal criteria kan bevinden in een bepaalde ontwikkelingsfase. Bij het auditinstrument van SDO gaat het om twintig criteria, gegroepeerd in vijf aandachtsgebieden en gerelateerd aan de Deming-cirkel voor kwaliteitsmanagement: PLAN-DO-CHECK-ACT. Zowel het EFQM-model als het INK-model zijn erkende methoden die breed worden toegepast binnen de kwaliteitszorg in Europa (zie bijvoorbeeld Hardjono & Hes, 1999). Het auditinstrument van SDO heeft de volgende kenmerken die ook voor onze methodiek zinvol zijn: (1) Door zelfevaluatie komen opleiding tot een positie- en ambitiebepaling, (2) Er wordt uitgegaan van fasen van ontwikkeling, (3) Er wordt gebruik gemaakt van één set van criteria die alle betrokkenen hanteren en (4) Er is een relatie met de PDCA-cyclus van Deming.

Voor de positie- en ambitiebepaling van onze methodiek gebruiken we vijf fasen waarbij fase 1 uitgaat van een minimale kwaliteit van toetsing en fase 5 van een ideale kwaliteit van toetsing. Deze fasen worden ingevuld op de zes bouwstenen (zie box 1). Dit leidt tot het volgende ontwerp kader van de methodiek, waarbij de gele cellen fictieve invullingen zijn van een positiebepaling (box 7).

### Hoe ziet een eerste ontwerp van een methodiek er uit?

Op basis van het ontwerp kader in box 7 kan een kwalitatieve invulling worden gegeven aan elke cel. Om tot een uitspraak te komen over de kwaliteit van elk van de zes bouwstenen is op basis van de ingebrachte vraagstukken vanuit opleidingen door de onderzoekers een verfijning aangebracht. Voor elke bouwsteen gelden drie dimensies. Deze dimensies zijn ontwerp (hoe komt de betreffende bouwsteen tot stand?), criteria (welke criteria worden gebruikt om de kwaliteit van de betreffende bouwsteen vast te stellen?)

en kwaliteitsborging (hoe vindt de kwaliteitsborging van de betreffende bouwsteen plaats?). Deze verfijning heeft geleid tot het eerste prototype van de methodiek. In box 8 wordt deze uitwerking gepresenteerd voor de bouwsteen toetsbekwaamheid, waarbij steeds de daaropvolgende fase ook borging van de voorafgaande fase betekent. Met andere woorden: wanneer de opleiding zich positioneert in fase 3 is ook voldaan aan fase 1 en 2.

Op dezelfde wijze als box 8 zijn ook de andere bouwstenen beschreven.

### Welke stappen zijn nodig voor een onderbouwde analyse van toetskwaliteit?

Met het ontwerp kader uit box 7 en de ingevulde cellen waarvan een uitgewerkt voorbeeld is gegeven in box 8, kan een opleiding vervolgens aan de slag. Dit vraagt een goede voorbereiding. Het is van belang vast te stellen wie het proces van het analyseren van de kwaliteit van toetsing, het vaststellen van de ambitie en verbeterstappen en het uitvoeren en monitoren van de verbeterstappen binnen de opleiding coördineert (de moderator). Vervolgens moet worden vastgesteld of er een vaste groep van betrokkenen alle bouwstenen analyseert of dat er onderscheid gemaakt wordt per bouwsteen. Een voordeel van het analyseren door één groep is dat de beoordeling op eenzelfde manier plaatsvindt. Een nadeel zou kunnen zijn dat de deelnemers mogelijk meer specifiek inzicht hebben in de kwaliteit van specifieke lagen en sommige lagen onderbelicht zijn in expertise. Wanneer duidelijk is wie de analyse gaan uitvoeren, voeren de deelnemers individueel de eerste stap uit. Ieder stelt vast per dimensie per bouwsteen wat de huidige kwaliteit van toetsing is, aangevuld met een verantwoording van het gegeven oordeel. Dit gebeurt door te verwijzen naar bewijsstukken om het ingenomen standpunt te illustreren. Mogelijke bewijsstukken zijn het beschreven toetsbeleid van opleiding, werkplannen/jaarverslagen van examencommissie, interne

**BOX 8: UITWERKING VAN DE BOUWSTEEN TOETSBEKWAAMHEID****KWALITEIT VAN DE TOETSBEKWAAMHEID VAN BETROKKENEN (beoordelaars, ontwikkelaars, examencommissies)**

	1	2	3	4	5
<p><b>Ontwerp</b></p> <p>Hoe komt de ontwikkeling en inzet van toetsbekwaamheid tot stand?</p>	Toetsbekwaamheid wordt informeel en non-formeel door betrokkenen zelf ontwikkeld. Elke betrokkene wordt ingezet bij toetsing, ongeacht de toetsbekwaamheid.	Toetsbekwaamheid wordt formeel op individueel niveau ontwikkeld op basis van afspraken met de leidinggevende. Elke betrokkene wordt ingezet op toetsen waarvoor aantoonbaar scholing is gevolgd.	Toetsbekwaamheid wordt ontwikkeld op basis van geconstateerde hiaten binnen een team en op teamniveau gemaakte afspraken. Elke betrokkene die wordt ingezet bij toetsing krijgt taken die passen bij zijn of haar toetsbekwaamheid en motivatie.	Toetsbekwaamheid wordt collectief ontwikkeld op basis van de visie die er op organisatieniveau is over goed toetsen. Toetsbekwaamheid wordt ingezet op basis van een middellange termijn visie op kwaliteit van toetsen (Hoe zetten we toetsbekwaamheid zo in dat de kwaliteit van toetsing binnen de opleiding is geborgd?).	Toetsbekwaamheid wordt collectief ontwikkeld op basis van een visie die er maatschappelijk en in de wetenschap is over goed toetsen ten behoeve van leven lang leren. Toetsbekwaamheid wordt ingezet op basis van deze onderbouwde visie op kwaliteit van toetsen.
<p><b>Criteria</b></p> <p>Welke informatie wordt gebruikt om de kwaliteit van de toetsbekwaamheid vast te stellen?</p>	Toetsbekwaamheid wordt informeel en non-formeel door het individu zelf vastgesteld.	Toetsbekwaamheid wordt formeel op individueel niveau vastgesteld door de leidinggevende op basis van duidelijke criteria.	Toetsbekwaamheid wordt vastgesteld door feedback van en intervisie met collega's op basis van duidelijke criteria.	Toetsbekwaamheid wordt op collectief niveau vastgesteld ontwikkeld op basis van de visie die er op organisatieniveau is over goed toetsen.	Toetsbekwaamheid wordt op collectief niveau vastgesteld op basis van een visie die past bij de ontwikkelingen vanuit de maatschappij en de wetenschap over (lange termijn) leren.
<p><b>Kwaliteitsborging</b></p> <p>Hoe wordt de kwaliteit van de toetsbekwaamheid geborgd?</p>	Toetsbekwaamheid wordt intern ad hoc op individueel niveau vastgesteld wanneer daar vraag naar is (bijv. bij accreditatie). Professionalisering vindt ad hoc op basis van individuele behoefte plaats met incidentele scholing (bijv. cursus, studiedagen). De keuze van scholing wordt bepaald op basis van de hiaten in toetsbekwaamheid van het individu. Toetsbekwaamheid wordt door de betrokkene zelf vastgesteld op basis van bijv. studentevaluaties.	Toetsbekwaamheid wordt intern één keer per jaar in het R&O gesprek vastgesteld door de leidinggevende op individueel niveau op basis van voorgeschreven eisen uit de wet. Professionalisering vindt ad hoc op basis van de behoefte van een team plaats met incidentele scholing (bijv. cursus, studiedagen). Toetsbekwaamheid wordt door de betrokkene zelf vastgesteld en met de leidinggevende besproken in bijv. een R&O gesprek.	Toetsbekwaamheid wordt intern één keer per jaar op collectief niveau vastgesteld op basis van voorgeschreven eisen uit de wet waarbij het resultaat een nauwkeurig overzicht is van de mate waarin toetsbekwaamheid op organisatieniveau is geborgd. Professionalisering vindt plaats op basis van een middellange termijn met structurele scholing op individueel en collectief niveau (bijv. opleiding toetsdeskundige). De keuze van scholing wordt bepaald op basis van de hiaten in toetsbekwaamheid van het team. Toetsbekwaamheid wordt door de betrokkene zelf vastgesteld door het vergaren van 360 graden feedback (studenten, collega's, werkveld) en met de leidinggevende besproken in bijv. een R&O gesprek.	Toetsbekwaamheid wordt één keer per jaar op collectief niveau vastgesteld door een externe partij op basis van voorgeschreven eisen uit de wet waarbij het resultaat een nauwkeurig overzicht is van de mate waarin toetsbekwaamheid op organisatieniveau is geborgd. Professionalisering vindt plaats met het oog op de middellange termijn met structurele scholing op collectief niveau (bijv. opleiding toetsdeskundige). De keuze van scholing wordt bepaald op basis van de visie van de opleiding op goed toetsen. Toetsbekwaamheid wordt door de betrokkene zelf vastgesteld door het vergaren van 360 graden feedback (studenten, collega's, werkveld) en met de leidinggevende besproken in bijv. een R&O gesprek.	Toetsbekwaamheid wordt één keer per jaar op collectief niveau vastgesteld door een externe partij op basis van voorgeschreven eisen uit de wet waarbij het resultaat een nauwkeurig overzicht is van de mate waarin toetsbekwaamheid is geborgd in relatie tot maatschappelijke en wetenschappelijke ontwikkelingen. Professionalisering vindt plaats met het oog op de lange termijn met structurele scholing op collectief niveau (bijv. opleiding toetsdeskundige). De keuze van scholing wordt bepaald op basis van de wensen vanuit de maatschappij en de wetenschap over (lange termijn) leren. Toetsbekwaamheid wordt door de betrokkene zelf vastgesteld door het vergaren van 360 graden feedback (studenten, collega's, werkveld). Daarnaast wordt de toetsbekwaamheid vastgesteld door een externe, gekwalificeerde partij (bijv. een opleiding).

audit rapportages of accreditatieverslagen (toetsbeleid), overzicht toetsprogramma, toetsmatrijzen, toetsanalyses (toetsprogramma/toetsen), overzicht trainingen, R&O gesprekken (toetsbekwaamheid), toetsroosters, werkprocessen toetscommissie en studentevaluaties (toetsorganisatie). In een consensusbijeenkomst vindt vervolgens voor iedere dimensie per bouwsteen overleg plaats om tot consensus te komen. Waar die niet kan worden bereikt, wordt die fase gekozen waarover alle deelnemers overeenstemming hadden. De bijeenkomst wordt geleid door de moderator die niet deelneemt aan de besluitvorming.

In de tweede stap van de methodiek formuleren de betrokkenen bij de opleiding in een aantal groepsessies een ambitie en bijbehorende concrete verbeterstappen ten aanzien van de kwaliteit van toetsing die richting geven aan de koers voor de overeengekomen beleidsperiode. De deelnemers bepalen de gewenste fase. Ook formuleren ze concrete doelen en daarbij horende activiteiten die het behalen van de gewenste fase mogelijk maken. De bijeenkomst eindigt met het vaststellen van die verbeterstappen die de meeste prioriteit hebben. Het resultaat van de meting is: een beschrijving van de huidige situatie (in de vorm van een getal per dimensie per bouwsteen) en een verantwoording, een beschrijving van de gewenste situatie, een datum waarop die gewenste situatie bereikt moet zijn en een lijst van eerste prioriteiten in het bereiken van een gewenste situatie. Als derde stap worden de verbeterstappen door de betrokkenen van de opleiding uitgevoerd. De examencommissies, verantwoordelijk voor de kwaliteit van toetsing, monitoren de uitvoering van de verbeterstappen en sturen waar nodig bij. De opleiding bepaalt zelf de frequentie waarmee ze de methodiek ten uitvoer brengt.

### **Hoe kunnen we de validiteit, betrouwbaarheid, bruikbaarheid en transparantie van de methodiek vaststellen?**

Omdat de methodiek in wezen een toetsinstrument is, gelden dezelfde kwaliteitscriteria die ook voor andere toetsinstrumenten gelden. De drie belangrijkste criteria zijn betrouwbaarheid (is de methodiek nauwkeurig genoeg, geeft de uitkomst een betrouwbare afspiegeling van de daadwerkelijke toetskwaliteit?), validiteit (dekt de methodiek het domein van toetskwaliteit, worden er geen zaken over het hoofd gezien?) en transparantie (is de methodiek inzichtelijk genoeg voor de gebruiker, worden de instructies en bronnen zo begrepen als ze zijn bedoeld?). Om te onderzoeken of de methodiek valide, betrouwbaar en transparant is, wordt het prototype in de komende maanden voorgelegd aan betrokkenen van een aantal opleidingen (leden toets- en examencommissies, management). In focusgroep interviews staan vragen centraal als:

- Past deze opzet bij de vraag hoe kwaliteit van toetsing en de verbeterstappen kan worden bepaald?
- Welke begrippen zijn niet duidelijk?
- Welke aanvullende informatie heb je nodig om de methodiek op een goede manier te begrijpen en te benutten?
- Wie kunnen worden betrokken bij de positie- en ambitiebepaling?

Op basis van de input van de opleidingen wordt de methodiek waar nodig aangepast en wordt een handleiding ontwikkeld om de methodiek te gebruiken met als resultaat een try-out versie. Tijdens de try-outfase testen een aantal opleidingen (delen van) de methodiek (try-out versie) en proberen ze het bijbehorende ondersteunende instructiemateriaal uit. De projectleden begeleiden de try-outs. Vragen die in de try-out fase centraal staan zijn onder andere: In hoeverre lukt het opleidingen om via individuele oordelen van

betrokkenen over de huidige kwaliteit van toetsing te komen tot een gemeenschappelijk oordeel over de kwaliteit van toetsing? In hoeverre herkennen opleidingen zich in het resultaat van de analyse, de vastgestelde kwaliteit van toetsing? In hoeverre lukt het opleidingen om op basis van de analyse van de huidige kwaliteit een ambitie te bepalen? Welke hulpmiddelen hebben opleidingen nodig om daadwerkelijk aan de slag te gaan om de vastgestelde ambitie te realiseren?

### **Wat is ons beoogde ideaal?**

We beseffen dat voorgestelde methodiek veel tijd en energie zal vragen deze te structureel in te zetten. Toch hopen we na afronding van het project de volgende opbrengsten te hebben behaald:

- Een verantwoording van het validatieproces van de beschrijving van de cellen (waarom dekken deze alle aspecten van kwaliteit van toetsing en alle fasen?);
- Een overzicht van bronnen c.q. bewijzen die opleidingen moeten voorleggen om de positie te bepalen (welke bronnen c.q. bewijzen ondersteunen de positiebepaling?);
- Een advies met betrekking tot een representatieve gebruikersgroep (welke personen moeten zijn betrokken om tot een betrouwbare positiebepaling te komen?);
- Voorbeelden van uitgewerkte werkprocessen op elke bouwsteen van de piramide (welke werkprocessen zijn illustratief om opleidingen te helpen bij hun positiebepaling?);
- Een format van een verantwoordingsdocument waarin wordt toegelicht op basis waarvan een positie en ambitie is bepaald (hoe wordt aangetoond dat de positiebepaling klopt?).



Vragen voor verder onderzoek kunnen zijn:

- Wat is het effect van het gebruik van de methodiek op de know-how over toetskwaliteit?
- Wat is het effect van de methodiek op het kunnen formuleren van een collectieve visie op de kwaliteit van toetsing?
- Wat is het effect van de methodiek op het kunnen vaststellen van huidige en wenselijke toetskwaliteit?
- Wat is het effect van de methodiek op het kunnen formuleren van een duurzaam systeem van kwaliteitszorgsysteem voor toetsing?

### Tot slot

In dit praktisch artikel hebben we een eerste prototype van een methodiek beschreven waarmee opleidingen de kwaliteit van toetsing kunnen analyseren om zo te komen tot positie- en ambitiebepaling. Kritische reflecties met collega's bieden houvast voor een herzien ontwerp van de methodiek. Belangrijke zaken die worden genoemd zijn: het verder expliciteren van het doel van de methodiek, het ontwikkelen van begeleidend materiaal, zoals een handleiding voor begeleiders, zicht krijgen op de doorlooptijd van het analyseren en het komen tot verbeterstappen en op de tijdsinspanning van betrokkenen. Om opleidingen zelf de kwaliteit van de toetsing te laten vaststellen op basis van door hen opgestelde criteria, moeten we voorzichtig zijn met het beschikbaar stellen van good-practices, zoals voorbeelden van toetsprogramma's en toetsbeleid. Door het zelf te ontwikkelen, is de opleiding eigenaar en heeft men binnen de opleiding de neuzen dezelfde kant op. Ook zou in de methodiek de rol van het management explicieter zichtbaar moeten zijn. Immers managers zorgen voor facilitering en goede werkcondities van de verschillende commissies die met kwaliteit van toetsing bezig zijn, maar monitoren ook of docenten/teams hun werk uitvoeren.

Het blijft van belang er voor te waken dat een goede schriftelijke vastlegging niet

voldoende is voor kwaliteit. Als de docenten/uitvoerders er niet van op de hoogte zijn, niet weten hoe ze het moeten uitvoeren of het gewoonweg niet doen, heb je nog geen kwaliteit. Het ontwikkelingsgericht bezig willen zijn met de kwaliteit van toetsing zou ook een onderdeel van de piramide kunnen zijn. Het willen leren van elkaar, het toetsen van kwaliteit met elkaar, het elkaar aanspreken op fouten.

Als er binnen de opleiding geen goede cultuur is op dit gebied, dan is het realiseren van kwaliteit van toetsen en hier ontwikkelingsgericht mee omgaan een te grote stap.

### Referenties

- Absolum, M., Flockton, L., Hattie, J., Hipkins R., & Reid, I. (2009) *Directions for Assessment in New Zealand (DANZ)*. New Zealand: Ministry of Education.
- Baartman, L.K.J. (2008). *Assessing the assessment. Development and use of quality criteria for Competence Assessment Programmes*. Academisch proefschrift. Utrecht: Universiteit van Utrecht.
- Boud, D. and Associates (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Sydney: Australian Learning and Teaching Council.
- Commissie Bruijn (2012). *Vreemde ogen dwingen. Eindrapport Commissie externe validering examenkwaliteit hoger beroepsonderwijs*. Den Haag: HBO-raad.
- Gerritsen-Van Leeuwenkamp, K. (2012). *Het Relatieve Belang van Vijftig Kwaliteitskenmerken van Toetsing voor Studententevredenheid in het Hoger Beroepsonderwijs*. Masterscriptie. Heerlen: Open Universiteit.
- Gulikers, J., Sluijsmans, D.M.A., & Baartman, L. (2008). The power of assessment in teacher education. In A. Swennen & M. van der Klink (Eds.), *Becoming a Teacher Educator. Theory and Practice for Novice Teacher Educators* (pp. 175- 191). Dordrecht, Netherlands: Kluwer Academic Publishers Group.
- Hogeschool Arnhem Nijmegen (2012). *Inhouden toetsbeleid*. Nijmegen: Hogeschool Arnhem Nijmegen.
- Hardjono, T.W., & Hes, F.W. (1999). *De Nederlandse Kwaliteitsprijs en Onderscheiding* (5e oplage). Deventer: Kluwer Bedrijfswetenschappen.
- 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).

[www.onderwijsinspectie.nl/actueel/publicaties/Boekhouder+of+wakend+oog.html](http://www.onderwijsinspectie.nl/actueel/publicaties/Boekhouder+of+wakend+oog.html).

- ISO (2009). *Toetsing een vak apart*. Utrecht: Interstedelijk studenten overleg.
- Joosten-ten Brinke, D., Sluijsmans, D.M.A., & Peeters, A. (2012). *De toetsing getoetst. Rapportage van de eerste projectfase gericht op het ontwerp van een methodiek om toetskwaliteit te analyseren, verbeteren en borgen*. Nijmegen: Hogeschool van Arnhem en Nijmegen.
- Ministerie van OC&W. (2012). Kamerbrief hoger beroepsonderwijs, 17 jan 2012. Den Haag: OCW.
- Mager, U., & Nowak, P. (2012). Effects of student participation in decision making at school. A systematic review and synthesis of empirical research. *Educational Research Review*, 7, 38-61. Doi: 10.1016/j.edurev.2011.11.001.
- Nederlands-Vlaamse Accreditatieorganisatie (2010). *Beoordelingskaders ten behoeve van het accreditatiestelsel*. NVAO: Den Haag.
- Oakland, J.S. (1993). *Total Quality Management: the route to improving performance*. Butterworth-Heinemann.
- Onderwijsraad (2006). *Advies Examinering: draagvlakken toegankelijkheid, uitgebracht aan de staatssecretaris van Onderwijs, Cultuur en Wetenschap*. Nr. 20060320/865. Den Haag (november). [www.onderwijsraad.nl/upload/publicaties/316/documenten/examinering\\_\\_\\_draagvlak\\_en\\_toegankelijkheid.pdf](http://www.onderwijsraad.nl/upload/publicaties/316/documenten/examinering___draagvlak_en_toegankelijkheid.pdf).
- Sluijsmans, D.M.A. (2008). *Duurzaam beoordelen in vraaggestuurd leren* (lectorale rede). Nijmegen: Hogeschool van Arnhem en Nijmegen.
- Sluijsmans, D.M.A. (2010). Beoordelen met het oog op de toekomst: Tien kenmerken van duurzaam beoordelen. *Examens*, 7, 5-9.
- Smits, R. (2011). Examencommissies HBO worden belangrijker! *Examens, tijdschrift voor de toetspraktijk*, 2, 16-19.
- Stichting Duurzaam Hoger Onderwijs (2010). *ALSHE: Auditing instrument for sustainability in higher education*. Duurzaam Hoger Onderwijs.
- Straetmans, G.J.J.M. (2006). *Bekwaam beoordelen en beslissen* (Lectorale rede). Deventer: Saxion.
- Van Berkel, H., & Bax, A. (2002). *Toetsen in het hoger onderwijs*. Houten: BohnStafleu van Loghum.
- Van der Vleuten, C.P.M., & Schuwirth, L.W.T. (2005). Assessment of professional competence: from methods to programmes. *Medical Education*, 39, 309 -317.
- Van der Vleuten, C.P.M., & Schuwirth, L.W.T. (2011). Een model voor programmatische toetsing. *Examens*, 3, 5-9.