

Bewust en bekwaam toetsen

Wat zouden lerarenopleiders moeten weten over toetsing?

Dominique Sluijsmans, Zuyd Hogeschool & Universiteit Maastricht

Sabine van Eldik, Hogeschool van Arnhem en Nijmegen

Desirée Joosten-ten Brinke, Fontys Lerarenopleiding Tilburg & Open Universiteit Heerlen

Linda Jakobs, Hogeschool van Arnhem en Nijmegen

Samenvatting

Kwaliteit van toetsen staat prominent op de politieke en maatschappelijke agenda van het hoger onderwijs. In dit artikel wordt betoogd dat kwaliteit van toetsen in hoge mate wordt bepaald door de toetsbekwaamheid van opleiders en hun bewustzijn van het belang van goed toetsen. Tot op heden is onvoldoende helder welke kennis lerarenopleiders in huis moeten hebben om die kwaliteit te kunnen borgen. Het ontbreken van een nationale kennisbasis 'Toetsing voor lerarenopleiders' vormt het uitgangspunt van ons onderzoek Bewust en bekwaam toetsen. In dit mede door SURF gefinancierd onderzoek hebben vier lerarenopleidingen hun krachten gebundeld om de kwaliteit van toetsbekwaamheid te versterken: het ILS van de HAN, FLOT van Fontys Hogescholen, de ULO van de VU en De Nieuwste PABO van de Hogeschool Zuyd. Een eerste belangrijke vraag in dit project is: Welke toetskennis is van belang voor lerarenopleiders? Om deze vraag te beantwoorden is een literatuurstudie uitgevoerd en zijn schriftelijke interviews afgenomen bij (inter)nationale toetsexperts. Resultaten laten zien dat er geen eenduidig beeld is over wat lerarenopleiders zouden moeten beheersen en dat reeds bestaande standaarden vaak abstract geformuleerd zijn. De kwaliteitspiramide voor eigentijds toetsen en beoordelen (Joosten-ten Brinke, 2011) is gebruikt voor een eerste aanscherping van leerdoelen over toetsing. In totaal zijn 57 leerdoelen benoemd en verdeeld over de vier niveaus van de piramide: 8 leerdoelen op het niveau van toetsitems, 28 leerdoelen op het niveau van toetsen, 9 leerdoelen op het niveau van toetsprogramma's en 12 leerdoelen op het niveau van toetsbeleid. Verdere validering van de leerdoelen is nodig voordat instrumenten kunnen worden ontwikkeld waarmee lerarenopleiders zich kunnen professionaliseren. Het artikel sluit af met enkele aanbevelingen en een inlink in het vervolg van het project.

Ontwikkelingen op het gebied van toetsing

Het beoordelen van studenten is één van de meest kritische taken van een leraar (Mertler & Campbell, 2005). Onderzoek wijst uit dat leraren veel tijd besteden aan het voorbereiden, uitvoeren en nakijken van toetsen (Plake, 1993; Stiggins, 1999; Stiggins & Conklin, 1992). Als leraren niet bekwaam genoeg zijn in toetsen, kan dit nadelige gevolgen voor de kwaliteit van het onderwijs (Alkharusi, Aldhafri, Alnabhani, & Alkabani, (2012); Popham, 2009). Het belang van goed toetsen blijft ook in Nederland niet onopgemerkt. In het accreditatiestelsel betreft één van de drie standaarden voor de beperkte opleidingsaccreditatie de toetsing (NVAO, 2010): de opleiding dient te beschikken over een adequaat systeem van toetsing en men moet kunnen aantonen dat de beoogde eindkwalificaties gerealiseerd worden. Ook hebben examencommissies door de wijzigingen in de Wet op het Hoger onderwijs een nieuwe rol gekregen: naast hun aanvankelijke procedurele verantwoordelijkheid krijgen zij nu ook een inhoudelijke toetskundige verantwoordelijkheid in het borgen van de kwaliteit van toetsen en beoordelen. Recentelijk is het rapport *Vreemde ogen dwingen* gepubliceerd (Commissie Bruijn, 2012). Deze commissie heeft een aantal maatregelen geformuleerd waarmee externe validering van de kwaliteit van beoordelen kan worden vormgegeven. Voorbeelden van deze maatregelen zijn het gezamenlijk ontwikkelen van landelijke toetsen door de opleidingen zelf, het certificeren van examinatoren,

het inrichten van een basis- en seniorkwalificatie examinering (BKE/SKE) voor leraren en het borgen van toetsdeskundigheid in visitatiecommissies.

Ook wetenschappelijk gezien heeft een ware paradigmashift plaatsgevonden: toetsen worden niet langer beschouwd als een domein dat toebehoort aan de psychometrici, maar krijgt ook vanuit leertheoretisch perspectief steeds meer aandacht. Toetsing wordt benaderd als een ontwerpvragestuk, waarbij de aansluiting met het leren essentieel is (zie ook Schuwirth & Van der Vleuten, 2006). Dit betekent dat toetsing naast de gangbare summatieve (= selecterende, beslissende) functie ook steeds meer een formatieve (= ontwikkelingsgerichte) functie gaat vervullen. Bij deze vormende functie van toetsing staat de bijdrage die de toetsing levert aan het leerproces van de student centraal. Er is veel onderzoek voorhanden dat aantoont dat het inzetten van ontwikkelingsgerichte toetsing leidt tot een verbetering in prestatie (Black & Wiliam, 1998). Recentelijk wordt naast de summatieve en formatieve functie van toetsing de duurzame functie van toetsing genoemd. Duurzaam toetsen betekent dat leren voor de lange termijn wordt gestimuleerd (Boud, 2000; Sluijsmans, 2008). Toetsen staan in dit perspectief niet meer op zichzelf, maar vormt een onderdeel van een programma van toetsen dat past binnen het toetsbeleid van opleidingen dat gericht is op het voorbereiden van studenten op een leven lang leren. Daarbij wordt een belangrijke rol toegekend aan de student, door deze geleidelijk te stimuleren zelf sturing te nemen over het leren. Het besef bij lerarenopleidingen dat de kwaliteit van toetsing voor een groot deel de kwaliteit van het onderwijs representeert en dus leren en beoordelen goed op elkaar aan moet sluiten, heeft drie ingrijpende implicaties gehad voor de inrichting van het toetsbouwwerk en de rol van actoren hierin.

Implicaties voor de onderwijspraktijk

Een eerste implicatie is een sterke uitbreiding aan toetsvormen en -instrumenten. De opkomst van het competentiegerichte leren in de lerarenopleiding heeft geresulteerd in een mix van toetsvormen die doorgaans worden bestempeld als 'innovatief'. Voorbeelden zijn de generieke toepassing van (digitale) portfolio's, proeven van bekwaamheid en simulaties. Tegelijkertijd maken studenten, inspecties, maar ook experts op het gebied van toetsing zich zorgen of de toetsing van kennis wel genoeg wordt geborgd (Straetmans, 2006). Toetsservicesystemen maken het mogelijk kennis en vaardigheden via digitale toetsen te meten. Een voorbeeld in het hbo hiervan is het *10voordeleraar*-project waarin kennistoetsen ontwikkeld worden die aansluiten bij een brede kennisbasis (HBO-raad, 2012; Joosten-ten Brinke, Lansu, Van Leuven, & Sluijsmans, 2012).

Een tweede implicatie is dat alle ontwikkelingen om een herijking van de kwaliteitscriteria vragen. Toetsing wordt een integraal onderdeel van een leertraject, waardoor naast de gangbare psychometrische criteria als betrouwbaarheid en validiteit een uitbreiding van criteria nodig is. 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 en anderen (2006), de zeven principes van Assessment 2020 (Boud and associates, 2010) en de tien kenmerken van duurzaam toetsen (Sluijsmans, 2010). Digitale toetsing vraagt nog eens om specifieke kwaliteitscriteria, zoals veiligheid en technische mogelijkheden (Draaijer & Hartog, 2007).

Een derde implicatie - die in dit artikel centraal staat - is dat de inzichten over de wijze van toetsing een uitbreiding van kennis en vaardigheden over toetsing vraagt die lerarenopleiders

vanuit hun eigen opleiding niet of nauwelijks 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 lerarenopleider zich verder moeten professionaliseren. Gezien de consequenties die toetsresultaten hebben voor studenten, is het essentieel dat lerarenopleiders begrijpen wat goed toetsen betekent (Alkharusi, Mahdi Kazem, & Al-Musawai, 2011; Sluijsmans, 2013). Het is belangrijk de lerarenopleider handvatten te bieden om de toetsing beheersbaar en transparant te houden. Bewustwording over wat kwaliteit van toetsing betekent voor kwaliteit van leren is een voorwaarde om tot goede toetsen te komen.

Hoewel eerder genoemde ontwikkelingen ertoe hebben geleid dat professionaliseringsactiviteiten zich steeds meer richten op toetsing, blijkt professionalisering lang niet altijd tot de gewenste effecten te leiden. Studenten oordelen nog steeds vaak negatief over de wijze van toetsing (ISO, 2009). De huidige toetsprogramma's bestaan vaak uit veel kleine toetsen, waardoor de samenhang voor de student zoek is en er sprake is van versnippering (Van der Vleuten en anderen, 2010). Formatief toetsen is nog ondervertegenwoordigd en ook de interbeoordelaarsbetrouwbaarheid laat vaak nog te wensen over. Verder zijn er opleidingen waarin onvoldoende duidelijk is wat er precies wordt getoetst, omdat de beoordelingscriteria onvoldoende duidelijk zijn. Over het algemeen hebben veel opleidingen moeite om bij leerwegonafhankelijke toetsing de validiteit, authenticiteit en betrouwbaarheid van niet voorgeschreven bewijsmateriaal te bepalen (Albers, Jakobs, Kneefel, Peters, & Sluijsmans, 2009).

Het project Bewust en bekwaam toetsen

Ondanks de druk om kwaliteit van toetsing te borgen, is er onvoldoende zicht op wat een lerarenopleider zou moeten weten over toetsing. Toetsing staat op de kaart, maar helaas zijn in Nederland de beschrijvingen over de gewenste kennis en kunde onvoldoende vertaald in leerdoelen (zie bijvoorbeeld Bekker-Holtland en anderen, 2010; Tillema, Dirkse-Hulshar, & Van den Hurk, 2002; Van Dinther, 2009). Ook in de beroepsstandaard voor lerarenopleiders is de bekwaamheidseis voor toetsing te abstract omschreven (Koster & Dengerink, 2000): *“De lerarenopleider ontwikkelt instrumenten voor (zelf)beoordeling van professionele bekwaamheden en beoordeelt deelnemers op hun geschiktheid”*. Er is echter niet uitgewerkt welke kennis hieraan ten grondslag ligt. In de kennisbasis voor lerarenopleiders (<http://www.kennisbasislerarenopleiders.nl/>) is met name binnen de domeinen 'leren en lerenden' en 'curriculum en assessment' aandacht voor toetsen en beoordelen. Het draait hier om de vragen 'wanneer heeft iemand iets geleerd' en 'hoe toets je of de doelen van het curriculum worden bereikt'. Ook in een recente reviewstudie over de rollen van leraren staat de rol van beoordelaar niet expliciet benoemd (Lunenburg, Dengerink, & Korthagen, 2013). Voor de lerarenopleider die naast het beschikken over kennis over toetsing ook een voorbeeld moet zijn voor de studenten in de uitvoering (het zogenoemde congruentieprincipe; Koster & Korthagen, 1997; Swennen, Lunenburg, & Korthagen, 2008) ontbreekt deze omschrijving volledig. Een goede lerarenopleider leert studenten expliciet hoe je als leraar en als school(team) goed toetst in een specifieke onderwijscontext (Joosten-ten Brinke, 2011).

Het toenemend belang van toetsbekwame leraren(opleiders) en het ontbreken van een nationale kennisbasis 'Toetsing voor lerarenopleiders' vormden de aanleiding voor het onderzoek

Bewust en bekwaam toetsen. In dit artikel rapporteren we over de eerste fase van het onderzoek dat zich richtte op het in kaart brengen van relevante leerdoelen over toetsen. De centrale vraag in het onderzoek is: Welke toetskennis is van belang voor lerarenopleiders?

Methode

Om een antwoord te vinden op deze vraag is gestart met een literatuuronderzoek gericht op internationale, expliciete beschrijvingen van kennis, standaarden of toetsvaardigheden voor leraren. Er is bewust gezocht op de doelgroep leraren, omdat toetskennis voor leraren in hoge mate overeenkomsten heeft met die van lerarenopleiders. Voor de zoektocht is gebruik gemaakt van zoekmachines als Google Scholar, Eric, en Omega. Gebruikte zoektermen waren 'assessment', 'testing' in combinatie met 'skills', 'competencies', 'knowledge' en 'literacy'. Er is literatuur gezocht in de periode 1990-2012. De zoektocht leverde aanvankelijk 24 publicaties op, maar in slechts acht van deze studies zijn expliciete beschrijvingen opgenomen van toetsinhouden, -kennis en -vaardigheden gericht op de doelgroep leraren.

We hebben tevens 28 (inter)nationale toetsexperts via e-mail benaderd. Deze experts hebben veelvuldig gepubliceerd op het gebied van toetsen, drie van hen zijn werkzaam in de lerarenopleiding. Alle experts zijn leraar. We hebben hen de volgende vraag voorgelegd: Welke kennis over toetsing mag van een leraar worden verwacht? Van de 28 experts die zijn benaderd hebben er 12 gereageerd op onze vraag (36%; 9 internationaal, 3 nationaal).

Resultaten

Tabel 1 geeft de genoemde kennis en vaardigheden per studie weer (in chronologische volgorde). Uit de beschrijvingen van de kennis en vaardigheden blijkt dat er verschillende manieren zijn om leerdoelen te ordenen. In bepaalde studies wordt dit beperkt tot kennis van een aantal concepten (bijvoorbeeld Schafer, 1991), terwijl in andere studies doelen zijn beschreven op basis van concrete ontwerp vragen (bijvoorbeeld Stiggins, 2009). Enkele studies beschrijven niet zozeer een lijst met kennis en vaardigheden, maar benoemen een aantal inhoudsgebieden waarin leraren onderlegd dienen te zijn (zie Popham, 2009). Het abstractieniveau van de leerdoelen is over het algemeen hoog (bijvoorbeeld Stiggins, 1999; Absolum en anderen, 2009), waardoor vertaling naar de praktijk lastig is.

De per email benaderde experts die op onze vragen reageerden (E1 t/m E12) benoemden relevante toetskennis op verschillende manieren. Een enkele keer werd deze kennis gerelateerd aan een specifieke toetsvorm. "*Lerarenopleiders moeten proeven van bekwaamheid kunnen ontwerpen en uitvoeren*" (E6). Veel vaker (10x) hadden de reacties betrekking op visie en toetsontwerp, zoals "*Het kunnen vertalen van het toetsbeleid naar toetspraktijk*" (E3), "*Zicht hebben op de doelen van toetsing en weten hoe deze te bereiken zijn*" (E7), "*Bewustzijn dat een menselijk oordeel leidt tot veel ruis*" (E8), "*Zicht op eigen opvattingen op toetsing*" (E2). Een aantal reacties richtten zich op onderdelen van het toetsproces zoals over analyse: "*Beheersen van statistische maten (gemiddelde, standaarddeviatie, modus, mediaan)*" (E1), "*Het maken van diagnostische analyses van totaal- en subscores op toetsen*" (E6). Over item- en toetsconstructie werden de volgende doelen benoemd: "*Construeren van toetsitems*" (E1), en "*Antwoordmodellen opstellen*" (E4). Bij interpretatie en communicatie werden specifiek genoemd: "*Koppelen van toetsresultaten aan toets- en onderwijsdoelen*" (E5), "*Weten hoe je waardevolle feedback geeft die het leren daadwerkelijk versterkt*" (E1) en "*Communiceren van toetsresultaten naar de*

Tabel 1

Overzicht van studies met beschrijvingen van toetskennis en -vaardigheden voor leraren

Auteur	Kennis en vaardigheden
American Federation of Teachers, National Council on Measurement in Education, & National Education Association (1990)	Leraren zijn in staat/vaardig: <ul style="list-style-type: none"> ▶ om de juiste toetsvormen te kiezen in relatie tot te nemen beslissingen; ▶ in het ontwerpen van toetsvormen in relatie tot te nemen beslissingen; ▶ in het verzamelen, scoren en interpreteren van de resultaten van externe en zelfontworpen toetsen; ▶ in het gebruiken van toetsresultaten bij het maken van beslissingen over individuele studenten, het plannen van instructie, het ontwikkelen van het curriculum en schoolverbetering; ▶ in het ontwikkelen van valide scoringsmethodes. ▶ in het communiceren van toetsresultaten naar studenten, ouders, andere derden; ▶ in het herkennen van onethische, illegale en andere onjuiste toetsvormen en gebruik van toetsinformatie;.
Schafer (1991)	Kennis van: <ul style="list-style-type: none"> ▶ basisconcepten; ▶ gebruik van toetsen; ▶ plannen en ontwikkelen van toetsen; ▶ interpreteren van toetsresultaten; ▶ beschrijven van resultaten; ▶ evalueren en beschrijven van toetsen; ▶ feedback en cijfers geven; ▶ ethische zaken.
Stiggins (1999)	<ul style="list-style-type: none"> ▶ Verbinden van toetsing aan duidelijke leerdoelen; ▶ verhelderen van verwachtingen over leerdoelen; ▶ toepassen van passende toetsvormen; ▶ ontwikkelen van goede toetsvragen and beoordelingscriteria met een passende sampling; ▶ voorkomen van 'bias' in toetsing; ▶ effectief communiceren over toetsresultaten; ▶ gebruiken van toetsen als onderwijs/feedback.
NSW-IoT (2005)	Zie Bijlage 1. Bijzonder is dat standaarden zijn geformuleerd op vier niveaus.
Absolum en anderen (2009)	Om toetsbekwaam te zijn moeten leraren de kennis, vaardigheid en motivatie hebben om verantwoord de ontwikkeling van studenten te beoordelen. Elke leraar 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.
Popham (2009)	Leraren hebben kennis en vaardigheden op de volgende terreinen: <ol style="list-style-type: none"> 1 de basisfunctie van toetsing, namelijk het verzamelen van bewijs op basis waarvan beslissingen kunnen worden genomen over de studenten; 2 betrouwbaarheid van toetsen (bijvoorbeeld het borgen van interne consistentie); 3 het belang van de drie typen validiteit: inhouds-, criterium en construct-validiteit; 4 hoe bias in toetsing kan worden herkend en voorkomen; 5 het ontwerp en het verbeteren van toetsitems; 6 het scoren van antwoorden van studenten, in het bijzonder het gebruik van rubrieken; 7 het ontwikkelen en toetsen van proeven van bekwaamheid, portfolio's, tentoonstellingen, peer assessments en self assessments; 8 ontwerpen en implementeren van formatief toetsen; 9 het verzamelen en interpreteren van bewijs over de houdingen, interesses en waarden van studenten; 10 het interpreteren van toetsresultaten op landelijke, gestandaardiseerd toetsen; 11 het toetsen van anderstaligen of studenten met handicaps;

Auteur	Kennis en vaardigheden
Vervolg Popham	12 het voorbereiden van studenten op landelijke toetsen; 13 het bepalen van de waarde van een landelijke, gestandaardiseerd toets voor het evalueren van de kwaliteit van het onderwijs.
Stiggins (2009)	1 Waarom beoordelen? (proces van toetsen en resultaten zijn in lijn met duidelijke doelen): <ul style="list-style-type: none"> ▶ de leraar weet voor wie de toets is bedoeld. ▶ de leraar begrijpt de motivatie van studenten voor toetsen en kan toets-ervaringen gebruiken om motivatie te verhogen; ▶ de leraar kan toetsprocessen en -resultaten inzetten voor formatieve toets-activiteiten; ▶ de leraar gebruikt toetsresultaten voor summatief toetsen; ▶ de leraar kan formatief en summatief toetsen goed balanceren. 2 Wat beoordelen? (toetsen reflecteren duidelijke en relevante leerdoelen): <ul style="list-style-type: none"> ▶ de leraar heeft duidelijke leerdoelen geformuleerd gekoppeld aan standaarden; ▶ de leraar begrijpt de verschillen tussen type leerdoelen; ▶ de leraar heeft een plan voor het toetsen van leerdoelen over langere tijd. 3 Hoe beoordelen? (leerdoelen zijn vertaald in toetsen die accurate resultaten opleveren): <ul style="list-style-type: none"> ▶ de leraar begrijpt de verschillende toetsvormen; ▶ de leraar kan toetsvormen kiezen die overeenkomen met beoogde leerdoelen; 4 Hoe communiceren? (toetsresultaten worden effectief gecommuniceerd): <ul style="list-style-type: none"> ▶ de leraar rapporteert en vat toetsinformatie samen op een manier dat deze leren representeert; ▶ de leraar kiest de beste manier van rapporteren (cijfers, narratieven, portefeuilles) die past bij de leerdoelen en de studenten; ▶ de leraar interpreteert en gebruikt toetsen correct; ▶ de leraar communiceert toetsresultaten effectief naar studenten; ▶ de leraar communiceert toetsresultaten effectief naar derden (ouders, collega's, andere stakeholders); 5 Hoe studenten betrekken? <ul style="list-style-type: none"> ▶ de leraren maakt leerdoelen duidelijk voor studenten; ▶ de leraar betreft studenten bij het ontwerp van de toets; ▶ de leraar betreft studenten bij het beoordelen, bijhouden en bepalen van de leerdoelen; ▶ de leraar betreft studenten bij het communiceren over hun leren.
Brookhart (2011)	<ul style="list-style-type: none"> ▶ leraren begrijpen wat 'leren' is in de context waarin ze les geven; ▶ leraren zijn in staat om toegankelijke en meetbare leerdoelen te formuleren die aansluiten bij de gewenste inhoud en beheersingsniveau volgens de curriculumdoelen en visies; ▶ leraren beschikken over verschillende strategieën om met studenten te communiceren over de leerdoelen en hoe deze bereikt kunnen worden; ▶ leraren kennen de bedoelingen en het gebruik van verschillende beoordelingsmethodes en zijn vaardig om deze te gebruiken; ▶ leraren kunnen vragen in klassensituaties, toetsitems en performance assessmenttaken analyseren om er zeker van te zijn dat ze passen bij het meten van de kennis en vaardigheden van hun studenten; ▶ leraren kunnen effectieve en bruikbare feedback op studentenwerk geven; ▶ Leraren zijn in staat 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; ▶ leraren kunnen externe toetsresultaten gebruiken en op de juiste manier interpreteren ten behoeve van de juiste beslissingen; ▶ leraren kunnen hun interpretaties van resultaten en beslissingen communiceren naar de studenten, schoolleiding en ouders; ▶ leraren kunnen hun studenten helpen om op basis van toetsinformatie beslissingen te nemen; ▶ leraren zijn zich bewust van hun wettelijke en ethische verantwoordelijkheden bij het beoordelen van studenten.

school, ouders en studenten" (E3). Formatief beoordelen wordt door de experts ook nadrukkelijk genoemd: "De definitie en onderliggende principes van formatief toetsen kennen en dit ook opnemen in de visie op leren (in de 'spirit' van onderwijzen)" (E11), of "Zicht op hoe formatief toetsen kan bijdragen aan zelfregulerend leren" (E11). Eén expert noemde dat lerarenopleiders "onderzoek moeten kunnen doen naar toetsing" (E8).

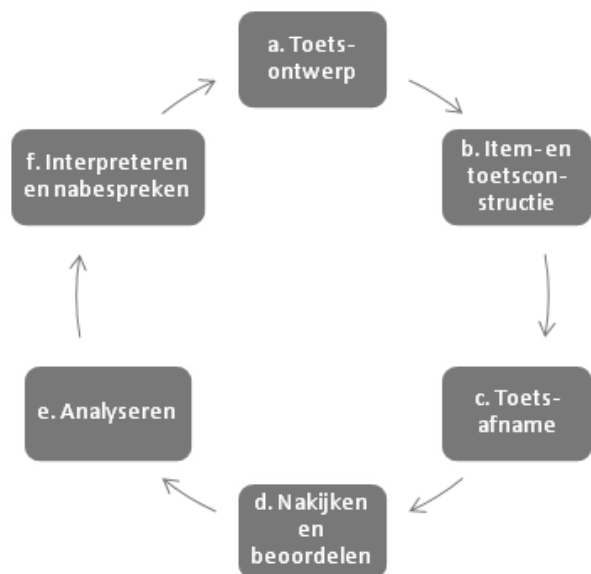
Vier experts (E1, E7, E8, E10) geven aan dat "het niet zo moet zijn dat lerarenopleiders en leraren psychometrici worden. Gebrek aan basiskennis over de kernbegrippen uit de psychometrie maakt echter het ontwerpen en analyseren van toetsen moeilijk". Een advies (E8) is dat elke opleider een basiscursus heeft gevolgd over toetsing met bijvoorbeeld uitgangspunten van de klassieke testtheorie, item-response theorie en generaliseerbaarheidstheorie. Zeven van de tien experts benaderen het gevaar van een te instrumentele focus op toetsing en wat ze op dit gebied moeten beheersen. Ze benadrukken het belang van het kunnen formuleren van opvattingen over toetsing, het vormen van een toetsvisie en het doorzien van de toetsfactoren die het leren zowel negatief als positief beïnvloeden.

De kwaliteitspiramide voor eigentijds toetsen en beoordelen

De resultaten van de literatuurstudie en de expertinterviews gaven zicht op welke kennis over toetsen voor (leraren)opleiders van belang is. Daarbij kunnen de volgende zaken worden geconstateerd:

- ▶ er bestaat geen eenduidig beeld over welke kennis (leraren)opleiders moeten beheersen, maar er zijn wel duidelijke patronen te herkennen;
- ▶ de kennis richt zich op verschillende toets-entiteiten (toetsitem, toets, toetsprogramma en toetsbeleid);
- ▶ de kennis binnen de entiteit 'toets' richt zich op specifieke fasen in een toetsproces.

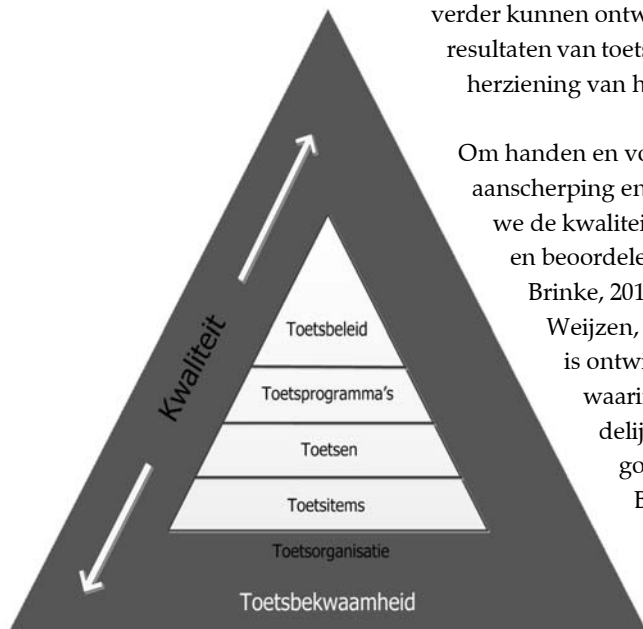
In Figuur 1 is dit toetsproces gevisualiseerd. In de ontwerpfase is kennis over de functie van toetsen, de keuze voor toetsvormen, de beoordelaars en de toetsvorm belangrijk. Bij de fase van item- en toetsconstructie is deskundigheid nodig in het beoordelen van de kwaliteit van de geconstrueerde vragen en samengestelde toetsen. Na een toetsafname is kennis nodig over de analyse van de toetsvragen en toets als geheel en hoe resultaten worden geïnterpreteerd en nabesproken met studenten. Kennis over feedback en feedforward geven is nodig om studenten handvatten te bieden over hun prestatie en de wijze hoe ze die in de toekomst



Figuur 1. Het Toetsproces.

Bijlage 1: Leerdoelen op het gebied van toetsen en rapporteren van effectief leren per expertiseniveau in Australië (bron: NSW Institute of Teachers, 2005, pagina 9 en 10)

ASPECT	EXPERTISENIVEAU			
	Startbekwame docent (graduate teacher)	Professionele, competente docenten (professional competent)	Professionele, competente docenten die bewezen hebben de school op een hoger plan te brengen (professional accomplishment)	Professionele docenten met leiderschapskwaliteiten (professional leadership)
TOETSEN Verbinden van toetsen en leren	Demonstreert kennis en zet strategieën in om het leren van leerlingen te toetsen.	Gebruikt een variëteit aan effectieve strategieën om het leren van leerlingen te toetsen.	Ontwerpt en gebruikt zeer effectieve toetsstrategieën die direct relatoren aan geformuleerde leeruitkomsten.	Gebruikt specialistische kennis over toetsing voor het leiden van plannen en programma-ontwikkeling.
	Demonstreert kennis over de relatie tussen toetsing en leeruitkomsten.	Communiceert naar leerlingen de relatie tussen hun prestaties en de geformuleerde leeruitkomsten.	Adviseert en assisteert collega's in het plannen en gebruiken van effectieve toets strategieën om leeruitkomsten te bereiken.	Leidt de evaluatie van toetsbeleid en –strategieën teneinde consistentie te bewaken.
Het geven van feedback aan leerlingen	Geeft zinvolle en tijdige mondelinge en schriftelijke feedback aan leerlingen.	Geeft tijdige, effectieve en consistente mondelinge en schriftelijke feedback aan leerlingen die hen helpt te reflecteren en hun voortgang in leren te bewaken.	Staat model voor het geven van effectieve en consistente mondelinge en schriftelijke feedback aan leerlingen om te verzekeren dat deze feedback hen helpt te reflecteren en aanmoedigt.	Evalueert en bewaakt in de school de effectiviteit van mondelinge en schriftelijke feedback.
Monitoring van de voortgang van leerlingen en dossieropbouw	Demonstreert kennis van en heeft een rationale voor het bijhouden van accurate rapporteren om de voortgang in leren van leerlingen te bewaken.	Gebruikt en handhaaft effectieve en efficiënte systemen voor dossieropbouw om de voortgang in leren van leerlingen te bewaken.	Adviseert en assisteert collega's in het ontwikkelen en gebruiken van effectieve en efficiënte systemen voor dossieropbouw om de voortgang in leren van leerlingen te bewaken te ontwikkelen, te gebruiken en te handhaven.	Leidt een diepgaande analyse van de gegevens over leerlingvoortgang door het toepassen van systemen voor dossieropbouw.
RAPPORTEREN	Demonstreert begrip van de principes and praktijken over het rapporteren van resultaten naar leerlingen, ouders en verzorgers.	Rapporteert effectief resultaten over voortgang naar leerlingen, ouders en verzorgers.	Adviseert en assisteert collega's om beleid, richtlijnen and rapportagesystemen te ontwikkelen.	Evalueert en bewaakt de mate waarin de wijze van rapporteren en verantwoordwoorden in lijn is met behoeften van leerlingen, ouders en verzorgers en/of andere relevante beleidslijnen.
PROGRAMMA-EVALUATIE	Demonstreert begrip van de principes en praktijken over de wijze waarop toetsresultaten kunnen worden gebruikt voor de evaluatie van lessen en lessen-series om te komen tot verdere planning van doceer- en leeractiviteiten waarop toetsresultaten kunnen worden gebruikt voor de evaluatie van lessen en lessen-series om te komen tot verdere planning van doceer- en leeractiviteiten.	Gebruikt toetsresultaten om doceer- en leerprogramma's te evalueren om te komen tot verdere planning.	Adviseert en assisteert collega's zowel de toetsresultaten als theoretische en praktische kennis te gebruiken om doceer- en leerprogramma's te evalueren.	



Figuur 2. De kwaliteitspiramide voor eigentijds toetsen en beoordelen binnen de context van een goede toetsorganisatie en toetsbekwame leraren.

verder kunnen ontwikkelen. Tenslotte is het van belang resultaten van toetsen te gebruiken voor een mogelijke herziening van het toetsontwerp.

Om handen en voeten te geven aan een verdere aanscherping en inkadering van leerdoelen hebben we de kwaliteitspiramide van eigentijds toetsen en beoordelen als basis gebruikt (Joosten-ten Brinke, 2011; Sluijsmans, Peeters, Jakobs & Weijzen, 2012; zie Figuur 2). Deze piramide is ontwikkeld op basis van literatuur waarin beschreven wordt wat voorwaardelijk is om te komen tot kwalitatief goede toetsing (zie o.a. Baartman, Bastiaens, Kirschner, & Van der Vleuten, 2006; Boud, 2000; Boud & associates, 2010; Brown & Hattie, 2009; Draaijer & Hartog, 2007; Schuwirth & Van der Vleuten, 2006; Sluijsmans, 2008).

In de kwaliteitspiramide wordt de expertise ten aanzien van toetsen en beoordelen ingedeeld in vier categorieën waarop de kwaliteit van toetsing binnen een instituut beoordeeld en geborgd kan worden:

- ▶ Op het hoogste niveau (kwaliteit van toetsbeleid) gaat het om de kwaliteit van toetsing in termen van de effecten voor de lange termijn. Wat is de visie van een opleiding ten aanzien van toetsen en beoordelen? In hoeverre draagt toetsing bij aan duurzame leerresultaten en het vormen van een blijvende lerende houding bij leerlingen of studenten?
- ▶ Op het tweede niveau gaat het om de kwaliteit van toetsprogramma's voor een afgebakend geheel van onderwijs (bijvoorbeeld de bachelor of de propedeuse). Er moet een heldere samenhang zijn tussen toetsen en de borging van de verschillende functies van toetsing. Kwaliteitscriteria zijn gericht op toetsprogramma's als geheel (Baartman en anderen, 2006). Binnen deze piramidelaag gaat het om het kunnen ontwikkelen van een effectief toetsprogramma dat afgestemd is op de curriculumdoelen.
- ▶ Het derde niveau betreft de kwaliteit van afzonderlijke toets- en beoordelingsinstrumenten en het gebruik daarvan. De stappen uit de toetscyclus (zie Figuur 2) moeten bekend zijn en kwalitatief goed uitgevoerd worden. Voorbeelden van kwaliteitscriteria zijn validiteit, betrouwbaarheid en bruikbaarheid;
- ▶ Op het laagste niveau gaat het om de kwaliteit van taken, opdrachten of items (stap b in de toetscyclus). Hieronder wordt dan elk onderdeel van een toets verstaan waarop een student een reactie kan geven. Voorbeelden van kwaliteitscriteria zijn objectiviteit, relevantie, en moeilijkheid.

De niveaus staan niet op zichzelf. Kwaliteit van toetsing wordt geborgd door een sterke samenhang van de vier niveaus (aangegeven door de zijdelingse pijl). Is de kwaliteit op het onderste niveau niet goed, dan heeft dit negatieve invloed op de kwaliteit van de bovenliggende niveaus. Andersom, idem dito. Als onduidelijk is wat de visie op toetsing is, zal dit zijn uitwerking naar beneden hebben op de kwaliteit van toetsprogramma's, toetsen en de onderliggende items. De kwaliteit van de vier lagen wordt bepaald door de kwaliteit van de toetsorganisatie en de toetsbekwaamheid van leraren (zie de buitenste twee lagen in de piramide).

Een eerste invulling van een kennisbasis 'Toetsen voor lerarenopleiders'

Het literatuuronderzoek en de expertinterviews hebben geleid tot een overzicht van feitenkennis (terminologie), conceptuele kennis (classificaties, theorieën, principes) en procedurele kennis (technieken, methoden). Dit overzicht is vertaald in een lijst van leerdoelen: 'De lerarenopleider is in staat om ...'. Hierbij zijn de leerdoelen die specifiek gericht waren op een specifieke situatie (bijvoorbeeld gericht op anderstaligen, studenten met handicaps) buiten beschouwing gelaten. Vervolgens zijn deze leerdoelen tijdens workshops met leraren en lerarenopleiders geordend naar de vier niveaus van de kwaliteitspiramide. Daarbij zijn de leerdoelen uit Tabel 1 en de expertinterviews geconcretiseerd naar specifiekere leerdoelen en zijn leerdoelen geclusterd als ze tijdens bijeenkomsten als cluster werden benoemd. In deze workshops is geconstateerd dat de genoemde doelen over onderzoek niet verder zijn meegenomen omdat deze veel overeenkomsten vertoonden met andere leerdoelen.

Uiteindelijk zijn er per piramidelaag respectievelijk de volgende acht (toetsvragen), 28 (toetsen) negen (toetsprogramma) en 12 (toetsbeleid) geordend en waar nodig aangescherpt. Bij de leerdoelen die uit de interviews zijn afgeleid, is dit als zodanig in onderstaande lijst aangegeven.

Leerdoelen op het niveau van toetsvragen

- 1 *De lerarenopleider kent verschillende typen toetsvragen.*
- 2 *De lerarenopleider weet op welke beheersingsniveaus een toetsvraag zich richt.*
- 3 *De lerarenopleider kan op basis van leerdoelen toetsvragen construeren die voldoen aan de kwaliteitscriteria relevantie, objectiviteit, efficiëntie en juiste moeilijkheidsniveau (experts).*
- 4 *De lerarenopleider kan toetsvragen op de kwaliteitscriteria relevantie, objectiviteit, efficiëntie en juiste moeilijkheidsniveau beoordelen.*
- 5 *De lerarenopleider kan realistische, concrete situaties en authentieke casussen bedenken.*
- 6 *De lerarenopleider kan op basis van beoordelingscriteria gegeven antwoorden beoordelen.*
- 7 *De lerarenopleider kan studenten scoren op individuele toetsvragen.*
- 8 *De lerarenopleider kan toetsvragen na afname analyseren op psychometrische kwaliteit.*

Leerdoelen op het niveau van toetsen

- 1 *De lerarenopleider kan een toets ontwerpen.*
- 2 *De lerarenopleider weet wat een toetsmatrijs is en hoe deze opgesteld moet worden.*
- 3 *De lerarenopleider kent de kwaliteitscriteria van toetsinstrumenten en weet deze toe te passen en te analyseren.*
- 4 *De lerarenopleider kan toetsbare leerdoelen (gebaseerd op inhoud en beheersingsniveau) voor studenten formuleren.*
- 5 *De lerarenopleider weet hoe fraude en plagiaat voorkomen kan worden.*
- 6 *De lerarenopleider kent verschillende toetsvormen en kan een toetsvorm kiezen en verantwoorden die aansluit bij de doelstellingen.*
- 7 *De lerarenopleider kent de voor- en nadelen van verschillende toetsvormen.*
- 8 *De lerarenopleider kent het verschil tussen proces- en productbeoordeling (experts).*

- 9 De lerarenopleider kent de functies van toetsing (experts).
- 10 De lerarenopleider weet wat pre- true- en post-effecten van toetsen zijn.
- 11 De lerarenopleider kan studenten betrekken in het toetsproces en hoe zij zichzelf en elkaar kunnen beoordelen.
- 12 De lerarenopleider kan een selecterende toets construeren die de 'goede' studenten onderscheid van de 'zwakkere' studenten.
- 13 De lerarenopleider kan toetsen aanpassen aan het niveau van de studenten.
- 14 De lerarenopleider kan heldere beoordelingscriteria construeren, en een rubric ontwikkelen.
- 15 De lerarenopleider kent de beoordelaarseffecten (zoals het halo-effect, het volgorde-effect).
- 16 De lerarenopleider kan de juiste toetsomgeving voor afname inrichten.
- 17 Een lerarenopleider kent het begrip cesuur en kent methoden voor het vaststellen van de cesuur en kan de cesuur bepalen.
- 18 De lerarenopleider kan een scoremodel opstellen en cijfers toekennen.
- 19 De lerarenopleider kan toetsafnames organiseren en uitvoeren.
- 20 De lerarenopleider weet wat intersubjectief beoordelen is.
- 21 De lerarenopleider kan gedrag in authentieke situaties beoordelen.
- 22 De lerarenopleider kan een antwoordmodel construeren (experts).
- 23 De lerarenopleider kent het verschil tussen analytisch en holistisch beoordelen.
- 24 De lerarenopleider kan na afname van een toets de kwaliteit van de toets vaststellen door een kwantitatieve en kwalitatieve toetsanalyse (experts).
- 25 De lerarenopleider kan toetsresultaten gebruiken voor verbetering van het onderwijs.
- 26 De lerarenopleider kan een gegeven beoordeling beargumenteren en rechtvaardigen en uitleggen.
- 27 De lerarenopleider kan effectieve, constructieve en bruikbare feedback geven (experts).
- 28 De lerarenopleider kan kwantitatieve toetsresultaten omzetten naar betekenisvolle informatie (experts).

Leerdoelen op het niveau van toetsprogramma's

- 1 De lerarenopleider weet wat een toetsprogramma is.
- 2 De lerarenopleider weet hoe toetsen in een toetsprogramma samenhangen in relatie tot de leerdoelen.
- 3 De lerarenopleider kan verantwoorden hoe de resultaten op afzonderlijke toetsen bijdragen aan de eindkwalificaties van de opleiding.
- 4 De lerarenopleider kent kwaliteitscriteria voor toetsprogramma's en kan ze toepassen.
- 5 De lerarenopleider weet wat methodemix betekent.
- 6 De lerarenopleider kan een blauwdruk van een toetsprogramma opstellen.
- 7 De lerarenopleider kan informatie (resultaten) van verschillende toetsinstrumenten aan elkaar verbinden.
- 8 De lerarenopleider kan studenten het toetsprogramma uitleggen.
- 9 De lerarenopleider weet wat 'constructive alignment' is (experts).

Leerdoelen op het niveau van toetsbeleid in context van leven lang leren

- 1 De lerarenopleider weet wat externe validering inhoudt.
- 2 De lerarenopleider weet de visie op toetsing te vertalen in beleid en zijn ontwerp van toetsing (experts).
- 3 De lerarenopleider weet hoe hij studenten kan betrekken bij het nadenken over toetsvisie.
- 4 De lerarenopleider kent het onderscheid tussen formeel, non-formeel en informeel beoordelen.
- 5 De lerarenopleider kent het onderscheid tussen normgericht, criteriumgericht en ipsatief beoordelen.
- 6 De lerarenopleider weet wat het verschil is tussen toetsen en beoordelen.
- 7 De lerarenopleider weet hoe hij zelfsturing kan bevorderen in toetsing.
- 8 De lerarenopleider weet wat duurzaam toetsen betekent.
- 9 De lerarenopleider kan bij toetsing een coachende rol aannemen om de betrokkenheid, zelfsturing en zelfverantwoordelijkheid van de student te bevorderen (experts).
- 10 De lerarenopleider weet hoe onethische, illegale en oneerlijke toetspraktijken voorkomen moeten worden.
- 11 De lerarenopleider weet hoe de kwaliteitsborging vanuit de examencommissies geregeld is.
- 12 De lerarenopleider kent de achtergronden van toetsen en beoordelen.

De lijst met in totaal 57 leerdoelen representeert een basisniveau in toetsbekwaamheid. Dit betekent dat elke lerarenopleider deskundig zou moeten zijn op deze inhoud. De doelen op toetsitem en toetsniveau zijn voorwaardelijk om de doelen op het niveau van toetsprogramma en toetsbeleid te behalen. Dit zou betekenen dat eerst wordt geïnvesteerd in het bekwaam in leerdoelen van de onderste twee piramidelagen.

Conclusie en discussie

In dit artikel staat de vraag centraal welke toetskennis en -vaardigheden voor lerarenopleiders van belang is om kwaliteit van toetsing te kunnen versterken en te borgen. Om deze vraag te beantwoorden is een literatuuronderzoek uitgevoerd en zijn (inter)nationale experts onderzocht. De literatuuranalyse en expertantwoorden maken duidelijk dat er geen eenduidige, concreet geformuleerde set aan leerdoelen voorhanden is op basis waarvan lerarenopleiders zich kunnen bekwaam op het gebied van toetsen. De kwaliteitspiramide voor eigentijds toetsen en beoordelen (zie Figuur 2) en een overzicht van de belangrijkste stappen in een toetsproces (zie Figuur 1) zijn gebruikt om een eerste invulling te geven aan een kennisbasis 'Toetsen voor lerarenopleiders' in termen van leerdoelen. Met de uitkomsten van dit onderzoek hopen we een eerste bijdrage te hebben geleverd aan de bewustwording van lerarenopleiders dat toetsen niet alleen van belang is om goed onderwijs te borgen, maar dat het ook inzicht biedt in de eigen professionele ontwikkeling als beoordelaar. In de volgende onderzoeksfase wordt een zelfbeoordelingsinstrument ontwikkeld waarmee leraren(opleiders) hun kennis over toetsen in kaart kunnen brengen.

Een kritische reflectie ten aanzien van deze studie is op zijn plaats. Het komen tot de juiste doelen vereiste veel discussie tussen de onderzoekers en met de betrokken leraren en lerarenopleiders in de workshops. Hierdoor zijn gemaakte keuzes niet altijd transparant. Daarnaast is door de veelheid en verscheidenheid aan leerdoelen nog onvoldoende sprake van focus en vertaling naar de praktijk van de opleider. Ook is het domein van toetsing zo complex dat het lastig is de grenzen van toetsing te benoemen. Het geven van feedback is bijvoorbeeld een leerdoel dat van een andere orde is dan het analyseren van de betrouwbaarheid van een toetsitem. Verdere validering van de leerdoelen door dialoog met lerarenopleiders is nodig, met vragen als: Wat vinden zij het meest belangrijk om goed te kunnen toetsen en beoordelen? Waar liggen prioriteiten? Welke leerdoelen zien we nog over het hoofd? Welke kennis en kunde vereisen specifieke rollen, bijvoorbeeld lidmaatschap van een examencommissie en toetscommissie? Tot slot is meer houvast nodig hoe de lerarenopleider kan aantonen dat leerdoelen zijn behaald.

Aanbevelingen voor de praktijk

Het toetsen van studenten is mogelijk de belangrijkste taak van de lerarenopleiders, zeker als het gaat om toetsen die in het teken staan van leren. Het versterken van de toetsbekwaamheid van opleiders zal, naast kwaliteit van leren, ook bijdragen aan het nemen van goede beslissingen over studenten. In dit artikel zijn 57 leerdoelen beschreven die samen een eerste invulling kunnen geven aan een kennisbasis 'Toetsen voor lerarenopleiders'. Het is goed voorstelbaar dat dit een druk geeft voor lerarenopleiders. Daarom wordt in het project een zelfbeoordelingsinstrument ontwikkeld waarmee lerarenopleiders in kaart kunnen brengen wat ze nog niet weten en hoe ze zich daarin verder kunnen professionaliseren. Dat daar tot nu toe nog te weinig

aandacht voor is en er mogelijk een achterstand weggewerkt moet worden kan inderdaad lastig zijn.

Om het belang van kennis over toetsing te stimuleren zijn drie aanbevelingen relevant. Een eerste aanbeveling is dat opleiders de leerdoelen met elkaar bespreken met vragen als: Vinden we deze doelen belangrijk? Welke leerdoelen zijn prioriteit? Waar liggen nog kansen? Een tweede aanbeveling is te analyseren welke kennis en kunde er al is over toetsing en hierover via bijvoorbeeld intervisie kennis te delen. Toetsing wint aan kwaliteit wanneer opleiders als 'lerend gezelschap' samenwerken (Van der Klink, 2012). Een derde aanbeveling is te onderzoeken welke condities er nodig zijn om leerdoelen te kunnen behalen in termen van tijd en ondersteuning. Belangrijke condities zijn het organiseren van effectieve vormen van professionalisering, het stimuleren van een onderzoekende houding naar de eigen toetspraktijk en het creëren van een leergemeenschap (Sluijsmans, Joosten, & Van der Vleuten, 2013).

Alle leerdoelen die dit onderzoek heeft opgeleverd zijn in hoge mate generiek en representeren een basisniveau: elke leraar of opleider zal al deze leerdoelen in bepaalde mate moeten beheersen om op een verantwoorde manier studenten te beoordelen. De waarde zit echter vooral in de wijze waarop de lerarenopleider in een specifieke opleidingscontext de leerdoelen in het ontwerp van de toetsen vertaalt. En daarin is het creëren van voldoende professionele ruimte voorwaardelijk. Studenten bekwamen in toetsen is nog teveel een ondergeschoven kind in de lerarenopleiding, maar als lerarenopleiders hierin voorbeeldig zijn, komt dit ook ten goede aan de toetsbekwaamheid van aanstaande leraren.

Referenties

- Alkharusi, H., Mahdi Kazem, A., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal of Teacher Education, 39*(2), 113-123.
- Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkabani, M. (2012). Educational assessment attitudes, competence, knowledge, and practices: An exploratory study of Muscat teachers in the Sultanate of Oman. *Journal of Education and Learning, 1*(2), 217-232.
- Absolum, M., Flockton, L., Hattie, J., Hipkins R., & Reid, I. (2009). *Directions for Assessment in New Zealand*. Wellington: Ministry of Education.
- AFT, NCME, & NEA (1990). *Standards for Teacher Competence in Educational Assessment of Students*. American Federation of Teachers, National Council on Measurement in Education, & National Education Association.
- Albers, G., Jakobs, L., Kneefel, I., Peters, P., & Sluijsmans, D. (2009). *Het verbeteren van de kwaliteit van toetsing en beoordeling*. Nijmegen: Hogeschool van Arnhem en Nijmegen.
- Baartman, L. K. J., Bastiaens, Th. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2006). The wheel of competency assessment: Presenting quality criteria for competency assessment programmes. *Studies in Educational Evaluation, 32*, 153-177.
- Bekker-Holtland, A., Brandenburg, C., de Bruin, S., Kösters, J., Reinink, M., Strijbis, J., & van der Worp, K. (2010). *Assessments in de lerarenopleiding*. Algemeen Directeurenoverleg Educatieve Faculteiten (ADEF).
- Black, P., & William, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*, 7-73.
- Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. *Studies in Continuing Education, 22*, 151-167.
- Boud, D. and associates (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Sydney: Australian Learning and Teaching Council.

- Brookhart, S. (2011). Educational Assessment Knowledge and Skills for Teachers. *Educational Measurement: Issues and Practice*, 30(1), 3-12,
- Brown, G., & Hattie, J. (2009). *Understanding teachers' thinking about assesment: Insights for developing better educational assessments*. Paper presented at the Annual Meeting of the National Council of Measurement in Education. VS: San Diego.
- Commissie Brijn (2012). *Vreemde ogen dwingen. Eindrapport Commissie externe validering examenkwaliteit hoger beroepsonderwijs*. Den Haag: HBO-raad.
- Draaijer, S., & Hartog, R. (2007). Guidelines for the design of digital closed questions for assessment and learning in higher education. *e-Journal of Instructional Science and Technology*, 10(1).
- HBO-raad (2012). www.10voordeleraar.nl. Den Haag: HBO-raad.
- ISO (2009). *Toetsing: een vak apart*. Utrecht: Interstedelijk studenten overleg.
- Joosten-ten Brinke, D. (2011). *Eigentijds toetsen en beoordelen*. Lectorale rede. Tilburg: Fontys Lerarenopleiding Tilburg.
- Joosten-ten Brinke, D., Lansu, W., Van Leuven, A., & Sluijsmans, D. (2012). Gewogen en ...? De landelijke kennistoets geschiedenis. *Kleio*, 53(1), 32-34.
- Koster, B., & Dengerink, J. (2000). *Towards a professional standard for Dutch teacher educators*. Paper presented at ATE-conference: USA: Orlando.
- Koster, B., & Korthagen, F. (1997). Opleiden in een snel veranderende samenleving. Enkele aspecten van de professionalisering van lerarenopleiders. *PML-nieuwsbrief*, 2, 2-4.
- Lunenberg, M., Dengerink, J., & Korthagen, F. (2013). *Het beroep van lerarenopleider: Professionele rollen, professioneel handelen en professionele ontwikkeling van lerarenopleiders*. Amsterdam: Vrije Universiteit.
- Mertler, C.A., & Campbell, C. (2005). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the Assessment Literacy Inventory*. Paper presented at the annual meeting of the American Educational Research Association, Montréal, Quebec, Canada.
- NSW Institute of Teachers (2005). *Professional teaching standards*. Sydney: NSW Institute of Teachers.
- Nederlands-Vlaamse Accreditatieorganisatie (2010). *Beoordelingskaders ten behoeve van het accreditatiestelsel*. NVAO: Den Haag.
- Plake, B.S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, 6(1), 21-27.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or Fundamental? *Theory into Practice*, 48, 1-11.
- Schafer, W.D. (1991). Essential assessment skills in professional education of teachers. *Educational Measurement: Issues and Practice*, 10(1), 3-6.
- Schuwirth, L.W.T., & Van der Vleuten, C.P.M. (2006). A plea for new psychometric models in educational assessment. *Medical Education*, 40, 296-300
- Sluijsmans, D.M.A. (2010). Beoordelen met het oog op de toekomst: Tien kenmerken van duurzaam beoordelen. *Examens*, 7, 5-9.
- Sluijsmans, D.M.A. (2008). *Betrokken bij beoordelen* (lectorale rede). Nijmegen: Hogeschool Arnhem en Nijmegen.
- Sluijsmans, D.M.A. (2013). *Verankerd in leren: Vijf bouwstenen voor professioneel beoordelen in het hoger onderwijs* (lectorale rede). Heerlen: Zuyd Hogeschool.
- Sluijsmans, D. M. A., Joosten-ten Brinke, D., Van der Vleuten, C.P.M. (2013). *Toetsen met leerwaarde. Een reviewstudie naar de effectieve kenmerken van formatief toetsen*. Den Haag: NWO.
- Sluijsmans, D.M.A., Peeters, A., Jakobs, L., & Weijzen, S. (2012). Kwaliteit van toetsing onder de loep. *Onderwijsinnovatie*, 4, 17-25.
- Straetmans, G.J.J.M. (2006). *Bekwaam boordelen en beslissen* (Lectorale rede). Deventer: Saxion Hogescholen.
- Stiggins, R.J. (1999). Are you assessment literate? *The High School Journal*, 6(5), 20-23.

- Stiggins, R.J. (2009). Essential formative assessment competencies for teachers and school leaders. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment*. New York: Routledge.
- Stiggins, R.J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany: State University of New York Press.
- Swennen, A., Lunenberg, M., & Korthagen, F. (2008). Preach what you teach! Teacher educators and congruent teaching. *Teachers and teaching*, 14(5), 531-542.
- Tillema, H., Dirkse-Hulshar, S. & van den Hurk, V. (2002). Naar een audit van assessmentpraktijken in de lerarenopleiding. *VELON Tijdschrift voor Lerarenopleiders*, 23(3), p. 20-26.
- Van Dinther, M. (2009). Assessor in het hbo: een vak apart. *Examens*, 1, 10-14.
- Van der Klink, M. (2011). *Bekwaam innoveren voor een toekomstbestendig hoger beroepsonderwijs*. Lectorale rede. Heerlen: Zuyd Hogeschool.
- Van der Vleuten, C. P. M, Schuwirth, L. W., Scheele, F., Driessen, E. W., & Hodges, B. (2010). The assessment of professional competence: building blocks for theory development. *Best Practice & Research Clinical Obstetrics and Gynaecology*, 24, 703-719.

Dit onderzoek is uitgevoerd met steun van Stichting SURF, de organisatie die ICT vernieuwingen in het hoger onderwijs en onderzoek initieert, regisseert en stimuleert door onder meer het financieren van projecten. Meer informatie over SURF is te vinden op de website (www.surf.nl).

