

Beoordelen als fundament van goed opleiden in het beroepsonderwijs: een analyse van toetsprogramma's in het mbo en hbo

L. K. J. Baartman en J. T. M. Gulikers

Samenvatting

Het bewust inzetten van geschikte (combinaties) van toetsvormen vormt een belangrijk fundament voor goed opleiden in het beroepsonderwijs. In de onderwijspraktijk en onderzoek wordt echter vaak gestart met de instrumentele hoe-vraag van de toetsvormen in plaats van de inhoudelijke wat-vraag van de goede beroepsbeoefenaar. Het doel van dit artikel is het analyseren van negen toetsprogramma's in het hbo en twee toetsprogramma's in het mbo, vanuit het perspectief van de inhoudelijke wat-vraag: de professionele bagage en taakuitvoering, de kern van het beroep in het werkveld en het ontwikkelvermogen van de startend beroepsbeoefenaar. De onderzoeksvragen zijn: (1) In hoeverre wordt de professionele bagage en professionele taakuitvoering beoordeeld, (2) In hoeverre wordt de kern van het beroep in het werkveld beoordeeld, en (3) In hoeverre stimuleert de beoordeling het ontwikkelvermogen noodzakelijk voor de startend beroepsbeoefenaar? De resultaten laten zien dat er meer aandacht nodig is voor een ontwerpgericht en programatisch perspectief op toetsing in het hbo en mbo. Startend vanuit de beroepsbeoefenaar kan duidelijker worden geoperationaliseerd wat een beoordelaar precies wil zien en horen van de student gedurende de toetsing.

1 Beoordelen als fundament van goed opleiden in het beroepsonderwijs: een analyse van toetsprogramma's in het mbo en hbo

Beoordeling in het beroepsonderwijs bepaalt de kwalificering van studenten voor de arbeidsmarkt en is een van de meest sturende factoren achter het leren van studenten (Boud, 2000). Het bewust inzetten van geschikte (combinaties) van toetsvormen vormt daarmee een belangrijk fundament voor goed

opleiden in het beroepsonderwijs (mbo en hbo). Een valide beoordeling betekent dat zowel de inhoud als de vorm van beoordeling passend zijn bij het doel. Van der Vleuten, Schuwirth, Scheele, Driessen en Hodges (2010) omschrijven dit als "fitness for purpose". Validiteit wordt echter niet bepaald door een losse toets op een bepaald moment in de opleiding, maar door alle toetsen tezamen in een (deel van) de opleiding, bijvoorbeeld het eerste jaar of een specialisatiefase. Wil beoordeling echt het fundament vormen van goed beroepsonderwijs, dan moeten de inhoud en de vorm worden bekeken op programmaniveau. Dit betekent dat toetsen niet meer kunnen worden gezien als losse entiteiten, maar als een samenhangend geheel in een toetsprogramma (Baartman, Bastiaens, Kirschner, & Van der Vleuten, 2007; Van der Vleuten et al. 2010). Een toetsprogramma is een bewust gekozen combinatie van toetsvormen die in samenhang een opleiding in staat stelt om een valide en betrouwbaar oordeel te vellen over de competenties van de student. Het denken in termen van toetsprogramma's impliceert dat toetsing een inherent onderdeel is van curriculumontwikkeling. Zo presenteren Van der Vleuten et al. (2012) een model voor toetsprogramma's dat aansluit bij het 4C/ID model voor curriculumontwerp (Van Merriënboer, 1997) en beschrijven Tillema, Kessels en Meijers (2000) een assessmentsysteem waarmee een betere connectie kan worden bereikt tussen assessment en instructie. Toetsing is hiermee niet het sluitstuk, maar juist een integraal onderdeel van curriculumontwerp (Sluijsmans, Straetmans, & van Merriënboer, 2008; Tillema et al., 2000).

De kwaliteit van toetsing en het curriculum als geheel wordt bepaald door het vaststellen van de eindkwalificaties of "objectives" van het onderwijs, het uitwerken van deze eindkwalificaties in beoogde leerop-

brenghen per onderwijseenheid en het vervolgens specificeren van de beoordelingscriteria voor de toetsen. Baartman, Kloppenburg en Prins (2013) noemen dit proces van operationalisatie als één van de onderdelen van de validiteit van toetsprogramma's. In het ontwerpen van curricula is deze specificering en uitwerking van de inhoud echter vaak onderbelicht en wordt te snel overgegaan naar het kiezen van toetsvormen. Nieuwe toetsvormen worden bijvoorbeeld ingezet "omdat een proeve van bekwaamheid past bij competentiegericht onderwijs", zonder na te denken over of dit de meest passende toetsvorm is. Een review van Van den Berg en De Bruijn (2009) laat zien dat deze gerichtheid op het methodische aspect een kenmerk is van veel onderzoek in het beroepsonderwijs. De leerinhoud en de veranderde inhoud in termen van kwalificaties die het beroep kenmerken staan minder centraal.

Het doel van dit artikel is daarom om de huidige toetsing in het beroepsonderwijs te analyseren startend vanuit het inhoudsvraagstuk. Het inhoudsvraagstuk is vertaald in drie kijkkaders die tezamen beschrijven wat het betekent om een goede startende beroepsbeoefenaar te zijn.

Dit artikel beschrijft de analyse van de kwaliteit van 11 toetsprogramma's van verschillende mbo en hbo opleidingen aan de hand van de drie kijkkaders. Hiertoe is een inhoudelijke koppeling gemaakt tussen de drie kijkkaders en het gevalideerde raamwerk van twaalf kwaliteitscriteria voor Competentie Assessment Programma's (Baartman et al., 2007) waarmee de 11 toetsprogramma's in eerste instantie geanalyseerd zijn (Baartman, Gulikers, & Dijkstra, 2013; Gulikers, Biemans, & Mulder, 2009). De onderzoeksvraag die centraal staat in dit artikel is: In hoeverre sluit de toetsing aan bij het inhoudelijke doel van de opleiding, namelijk het opleiden van goede startende beroepsbeoefenaren? Dat wil zeggen: (1) In hoeverre wordt de professionele bagage en professionele taakuitvoering beoordeeld, (2) In hoeverre wordt de kern van het beroep in het werkveld beoordeeld, en (3) In hoeverre stimuleert de beoordeling het ontwikkelvermogen noodzakelijk voor de startende beroepsbeoefenaar? Dit artikel start met een verdere uitwerking

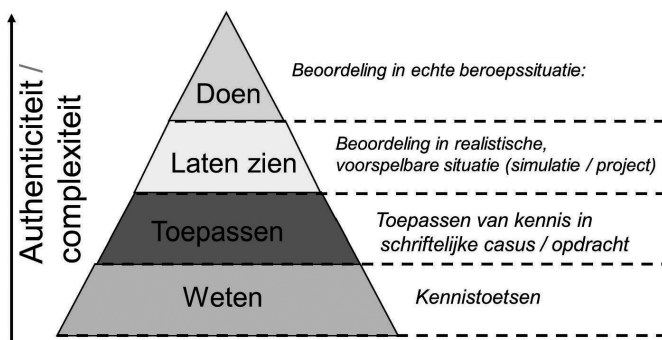
van de drie kijkkaders. In ieder kijkkader wordt de betreffende inhoudsvraag geproblematiseerd en geëxpliciteerd en worden de implicaties voor toetsing beschreven. Vervolgens wordt een koppeling gelegd tussen de drie kijkkaders en de twaalf kwaliteitscriteria waarmee de 11 toetsprogramma's zijn geanalyseerd. In de resultatensectie wordt een beschrijving gegeven van de casussen in het mbo en hbo, waarna een typering wordt gegeven van de 11 toetsprogramma's vanuit het oogpunt van de drie kijkkaders. Tot slot worden op basis van de verschillen en overeenkomsten tussen de casussen in het mbo en hbo enkele conclusies getrokken over sterke en zwakke schakels in de huidige toetsprogramma's voor de beoordeling van toekomstige beroepsbeoefenaren en worden suggesties gedaan voor verbetering.

1.1 Kijkkader 1: professionele bagage en professionele taakuitvoering

Kijkkader 1 start vanuit het idee dat een competente startende beroepsbeoefenaar beschikt over een complete en samenhangende set van professionele bagage, bestaande uit kennis, vaardigheden en houdingen (Schaap, De Bruijn, Van der Schaaf, & Kirschner, 2009; Schaap & Van Schaik, dit nummer), en het vermogen om deze kennis, vaardigheden en houdingen op een flexibele en adequate manier in te zetten bij het werken aan complexe beroepstaken (Roelofs, 2006): de professionele taakuitvoering. Deze startbekwaamheid, in tegenstelling tot de vakvolwassen beroepsbeoefenaar, staat beschreven in de mbo kwalificatiedossiers (www.kwalificatiesmbo.nl) en de Dublin Descriptoren voor het hbo.

In de literatuur over toetsing wordt een duidelijke onderscheid gemaakt tussen de professionele bagage (opgevat als kennis en vaardigheden) en de professionele taakuitvoering. De piramide van Miller (1990) geeft hier mooi uitdrukking aan (zie Figuur 1).

Miller analyseerde in het medische domein de vraag "wat maakt iemand (i.e., een arts) een goede beroepsbeoefenaar?". Dit resulteerde in een piramide met vier lagen die tezamen beroepsbekwaamheid beschrijven: "knows", "knows how", "shows how" en



Figuur 1. De piramide van Miller (bewerkte versie)

“does”, waarbij de onderste lagen het fundament vormen voor de hogere lagen. Het fundament van de piramide en daarmee ook het fundament van beroepsbekwaamheid ligt in het “knows” oftewel het weten en het “knows how”, oftewel toepassen van kennis/probleem oplossen en routinematige vaardigheden (b.v. bordschrijven, veiligheidsprocedures uitvoeren). De derde en vierde laag gaan over het laten zien van daadwerkelijk bewust gedrag, waarbij de student kennis, vaardigheden en houdingen moet integreren. Het hoogste niveau van de piramide is “does” waarin echt professioneel gedrag tot uiting komt in reële en volwaardig complexe beroepssituaties. Het niet alleen om het, wellicht toevalligerwijs, laten zien van professioneel gedrag, maar ook om het nemen van de juiste beslissing in een variatie aan professionele situaties, met andere woorden, om het bewust inzetten van kennis, vaardigheden en houdingen passend bij een bepaalde situatie. Roelofs (2006) onderzocht bijvoorbeeld hoe competent handelen van docenten geduid kan worden en omschrijft dit in een procesmodel. Een dergelijk procesmodel omschrijft op het niveau van een concrete beroepstaak langs welke processen van denken en handelen een beroepsbeoefenaar een taak uitvoert en hoe hierin combinaties van kennis, vaardigheden en houding worden ingezet. Hierbij vinden processen plaats van inschatten, afwegen en beslissen.

Bovenstaande omschrijving van de professionele bagage en taakuitvoering van de beroepsbeoefenaar heeft implicaties voor de keuze van toetsvormen. Miller (1990) noemt multiple choice toetsen, open vragen en

casustoetsen als geschikte toetsen op het “knows” niveau, waarin gevraagd wordt naar de reproductie van het kennisdomein. Hier gaat het hier met name om herinneren en begrijpen (uitleggen, voorbeelden noemen). Het niveau “knows how” is gericht op het toepassen van deze kennis in het oplossen van problemen of routinematige vaardigheden. Ook hier zijn multiple choice toetsen, open vragen en casustoetsen geschikt, maar moet duidelijk worden gevraagd naar toepassing. In de taxonomie van Krathwohl (2002) gaat het dan bijvoorbeeld om toepassen, analyseren, en evalueren van bijvoorbeeld een casus of beroepssituatie. Ook vaardigheidstoetsen (bijvoorbeeld het vertonen van bloembindingstechnieken door vijf bloemenboeketten te maken) beoordelen dit “knows how” niveau. Hierbij wordt nog weinig tot geen integratie van kennis, vaardigheden en houdingen gevraagd en is er ook geen sprake van een complexe, realistische beroepssituatie. Toetsvormen die recht doen aan de professionele taakuitvoering zijn te vinden op het “shows how” en het “does” niveau. Een student moet bij beoordeling op het niveau van “shows how” daadwerkelijk laten zien wat hij kan in bijvoorbeeld een beroepstaak, project of simulatie. De student moet zijn kennis en routinematige vaardigheden inzetten en combineren tot het laten zien van competent gedrag in een meer authentieke (maar vaak gesimuleerde of gestandaardiseerde) beroepstaak waarin ook een professionele houding tot uiting komt. De beoordeling door de docent gebeurt bijvoorbeeld door middel van observatie en het gebruik van rubrics of observatie-

formulieren. Beoordeling van het “does” niveau is het meest complex en ongestandaardiseerd. Hier laten toekomstig beroepsbeoefenaren zien dat zij kunnen functioneren in de complexe en soms onvoorspelbare authentieke beroepspraktijk, bijvoorbeeld door beoordeling op de werkplek.

1.2 Kijkkader 2: de kern van het beroep in het werkveld

Beroepsbekwaamheid is sterk afhankelijk van het specifieke beroep waarvoor wordt opgeleid (Mulder, in press; Mulder, Wesseling, & Bruijstens, 2005): de kerntaken van de verpleegkundige verschillen van de kerntaken van een bloemist en als gevolg daarvan krijgen competenties, bijvoorbeeld “vakdeskundigheid toepassen”, een andere invulling. Competenties krijgen dus pas betekenis in de context waarin ze worden ingezet: “competence as situated professionalism” (Mulder, in press). Dit tweede kijkkader specificeert wat beroepsbekwaam gedrag in een bepaald beroep betekent. Dit vraagt om een analyse van de kerntaken van het beroep, ook wel ‘job analysis’ genoemd (Robinson, 2012):

Job analysis is the process through which the component parts of a job are identified [.....] The process can be either task-oriented or person-oriented. Task-oriented approaches seek to divide a job into its various tasks, activities, and responsibilities, while person-oriented approaches focus on the underlying skills and attributes required by successful performers. A combination of both approaches is desirable in order to elicit the most complete picture of a job.

Een dergelijke analyse biedt inzicht in de kerntaken van het beroep en maakt concreet wat dit betekent voor de kennis en vaardigheden van de student. Deze kerntaken (of een selectie daarvan) dienen vervolgens als uitgangspunt bij het vormgeven van toetsen die representatief en relevant zijn voor de toekomstige beroepsbeoefenaar, ook wel authentiek toetsen genoemd (Gulikers, Bastiaens, Kirschner, 2004). Authentiek toetsen kan gedurende alle jaren en alle niveaus van een beroepsopleiding en dus niet alleen om startbekwaamheid te beoordelen. Hiervoor is het noodzakelijk de inhoudsvraag zo concreet mogelijk te beantwoord: “wat moet een stu-

dent van deze opleiding, op dit niveau, in dit leerjaar, voor beroepstaken kunnen uitvoeren”?. Dit is de referentiesituatie waarop een authentieke toetsituatie kan worden gebouwd.

In het beroeps onderwijs zien we een sterke toename van beoordeling op de werkplek als belangrijk onderdeel van een valide beoordeling van beroepsbekwaamheid (Gulikers, Biemans & Mulder, 2009). Echter, ook hierbij wordt vaak niet gestart vanuit de kerntaken van het beroep. Onderzoek laat echter zien dat de werkplek niet automatisch een goede en authentieke leer- en toetsomgeving is (Poortman, Illeris en Nieuwenhuis, 2011; Tynjälä, 2008). Studenten draaien op de werkplek mee in het reguliere werkproces in plaats van dat de stageplek expliciet wordt ingezet om bepaalde bekwaamheden te laten zien en beoordelen (Gulikers et al., 2009; Poortman et al., 2011; Toohey, Ryan, & Hughes, 1996). Ook beoordelingssituaties waarin docenten proberen de beroepswerkelijkheid na te bootsen in meer schoolse settings leiden niet altijd tot de gewenste resultaten. Dit komt onder andere doordat docenten niet altijd een goed en up-to-date beeld hebben van de huidige en veranderende beroepspraktijk. Cennamo en collega's (2011) geven een mooi voorbeeld van het belang van de betrokkenheid van echte architecten, naast docenten en medestudenten. Alleen zij kunnen, ook in de schoolse context, studenten de normen van de beroepspraktijk bijbrengen, feedback geven vanuit dit blikveld en beoordelen of de eindproducten van de studenten aan deze professionele normen voldoen. Met andere woorden: een authentieke beoordeling vraagt niet alleen om het meer realistisch laten lijken van de toetsituatie (Cummings & Maxwell, 1999). Om de kern van het beroep in het werkveld te kunnen beoordelen moet de toets aan bepaalde randvoorwaarden moeten voldoen. Het vijf-dimensionale model van Gulikers en collega's (2004) benoemt vijf toetskenmerken die meer of minder op de echte beroepspraktijk kunnen lijken. Deze vijf dimensies zijn (1) de toetstaak (2) de fysieke context en daarin beschikbare bronnen en materialen (3) de sociale context bestaande uit samenwerking-/hulpmogelijkheden (4) de toetsvorm (5) het toetsresultaat

en – criteria waarop de student beoordeeld wordt. Hoe meer deze vijf kenmerken in een toets worden vormgegeven overeenkomstig de professionele referentiesituatie op het niveau van de student, hoe groter de kans dat de kern van het beroep in het werkveld wordt beoordeeld.

1.3 Kijkkader 3: het ontwikkelvermogen van de beroepsbeoefenaar

De beroepspraktijk is veranderlijk en de precieze invulling van de kerntaken verandert voortdurend. Studenten moeten niet worden opgeleid voor een stabiel beroep, maar juist kunnen omgaan met deze veranderlijkheid. Bovendien komen steeds meer afgestudeerden niet terecht in het beroep waarvoor ze zijn opgeleid, maar in een aanpalend beroep. Dit betekent dat beroepsbeoefenaren in staat moeten zijn om hun kennis, vaardigheden en houdingen in verschillende samenstellingen te integreren tot beroepsbekwaam gedrag (Baartman & De Bruijn, 2011). Ze moeten voortdurend kritisch blijven kijken naar hun eigen kunnen in relatie tot wat ze graag willen/moeten bereiken en zich vervolgens in de gewenste richting ontwikkelen (Boud, 2000; Poortman et al., 2011). Tevens moeten ze in staat zijn transfer te maken naar nieuwe beroepssituaties (Tuomi-Gröhn & Engeström, 2003). Het derde en laatste kijkkader is daarom gericht op het ontwikkelvermogen van de beroepsbeoefenaar, de capaciteit van een professional om zijn/haar competenties verder te ontwikkelen op basis van kritische analyse van (werk)ervaringen (Kloppenburger, 2011).

In de toetsliteratuur is het belang van dit ontwikkelvermogen en de sturende rol van toetsing in het stimuleren en/of waarderen van dit ontwikkelvermogen de laatste 15 jaar sterk gegroeid (Boud, 2000; Nicol & Macfarlane-Dick, 2006; Wiliam, 2011). In dit verband wordt steeds meer gesproken over “sustainable assessment” of “learning-oriented assessment” (Boud, 2000; Carless, 2007; Davies & Ecclestone, 2008). Hierin zijn toetsen niet alleen sluitstukken, maar ook juist leermomenten. Hiermee hebben we het dan expliciet over de formatieve functie van toetsing, waarvan de kern is te beschrijven in drie processen: (1) waar wil/moet ik naartoe (feedforward), (2) waar sta ik nu/wat laat mijn huidige

prestatie zien (feedback), en (3) wat kan ik doen om het gat te dichten (feedforward)? (Hattie & Timperley, 2007; Wiliam, 2011). Deze drie processen beschrijven de kern van het ontwikkelvermogen van de competente beroepsbeoefenaar.

Ook het derde kijkkader heeft implicaties voor de toetsing. Formatieve toetsing moet studenten stimuleren hun eigen prestaties kritisch te analyseren en moet rijke feedback opleveren. Daarnaast moet een toetsprogramma studenten de gelegenheid bieden om de verkregen feedback te verwerken, resulterend in een verbeterde prestatie in een volgende toetssituatie (Wiliam, 2011). Ook voor het laten zien van transfer zijn meerdere toetssituaties in een toetsprogramma noodzakelijk. Toetsvormen met een heel duidelijke formatieve functie zijn peer-assessment en zelf-assessment. Onderzoek heeft aangetoond dat deze toetsvormen een positieve bijdrage leveren aan de professionele houding, het kritisch kijken naar het eigen functioneren, en het geven van feedback op elkaars prestaties (Boud, 2000; Sluijsmans, Brand-Gruwel, & Van Merriënboer, 2002). Een belangrijke voorwaarde om zelf- en peer-assessment te laten slagen is transparantie van de beoordelingscriteria en standaarden. Studenten moeten hun eigen werk kunnen vergelijken met een bepaalde standaard, en moeten beschikken over inzicht in de mogelijkheden om hun werk vervolgens te verbeteren (Nicol & Macfarlane-Dick, 2006; Sadler, 1989). Ook een portfolio biedt veel mogelijkheden voor het stimuleren en/of beoordelen van ontwikkelvermogen. Een portfolio bevat vaak geschreven reflecties en zelfevaluaties (Van Tartwijk, Driessen, Van der Vleuten, & Stokking, 2007; Wade & Yarbrough, 1996), en ontwikkelingsplannen en leerdoelen die kunnen bijdragen aan het monitoren van de eigen professionele ontwikkeling over verschillende professionele situaties heen (Mathers, Challis, Howe, & Field, 1999). Het gaat hierbij dus om longitudinaal beoordelen op verschillende momenten door de leerloopbaan van de student om het ontwikkelvermogen en transfer zichtbaar te maken.

2 Methode

2.1 Analyse van de casussen op basis van de drie kijkkaders

De 11 casussen in dit artikel, bestaande uit 2 mbo en 9 hbo toetsprogramma's, zijn in eerdere studies (Baartman, Gulikers, & Dijkstra, 2013; Gulikers, Biemans, Mulder, 2009) beschreven en geanalyseerd vanuit twaalf kwaliteitscriteria voor Competentie Assessment Programma's (Baartman et al., 2007; Baartman, Prins, Kirschner, & Van der Vleuten, 2011). Voor dit artikel is een secundaire analyse (Hinds, Vogel, & Clarke-Steffen, 1997) uitgevoerd vanuit de drie kijkkaders zoals hierboven beschreven. De 12 kwaliteitscriteria staan gepresenteerd in Tabel 1 en vormen een gevalideerd en veel gebruikt analysekader om de kwaliteit van toetsprogramma's in het beroepsonderwijs te evalueren (zie Baartman et al., 2007).

Voor de analyse op de drie kijkkaders is een inhoudelijke koppeling gelegd tussen de drie kijkkaders en de 12 kwaliteitscriteria, op basis waarvan analysevragen zijn geformuleerd (zie Tabel 2). Bijvoorbeeld kijkkader 1

(professionele bagage en taakuitvoering) sluit aan bij het kwaliteitscriterium "fitness for purpose", de vraag of alle relevante kennis, vaardigheden en houdingen op een geïntegreerde manier worden getoetst. Een ander kwaliteitscriterium dat aansluit bij kijkkader 1 is cognitieve complexiteit, wat gaat over het beoordelen van bewust en doordacht handelen (beslisprocessen) in de professionele taakuitvoering.

Om de kwalitatieve data uit de casussen te analyseren zijn alle data samengevoegd in verschillende matrices (Miles & Huberman, 1994). Eerst zijn de data geordend op basis van de 12 kwaliteitscriteria, omdat dit het gezamenlijke uitgangspunt was bij de dataverzameling in de 2 mbo-opleidingen en de 9 hbo-opleidingen. Vervolgens zijn de resultaten per casus onder de kijkkaders geplaatst, op basis van de koppeling tussen de 12 kwaliteitscriteria en de drie kijkkaders (Tabel 2). Per kijkkader ontstond zo een matrix, apart voor de mbo-casussen en de hbo-casussen, waarin de kwaliteitscriteria in de rijen stonden en de casussen in de kolommen. Deze matrices zijn vervolgens door beide auteurs

Tabel 1

Twaalf Kwaliteitscriteria voor Competentie Assessment Programma's (gebaseerd op Baartman et al., 2007)

Kwaliteitscriteria	Omschrijving
Fitness for purpose	Inhoudelijke afstemming tussen het CAP en de opleidingsdoelen. Het CAP meet de beoogde competenties.
Cognitieve complexiteit	Beoordeling van denkprocessen tijdens de professionele taakuitvoering.
Authenticiteit	De overeenkomst tussen de beoordelingsvormen en de toekomstige beroepssituatie.
Ontwikkeling van zelfsturend leren	De toetsen stimuleren studenten om te reflecteren, zichzelf en elkaar te beoordelen en eigen leerdoelen te formuleren.
Transparantie	Het CAP is bekend bij en begrijpelijk voor studenten, docenten en beoordelaars uit het werkveld.
Herhaalbaarheid	Beslissingen over studenten zijn gebaseerd op meerdere bewijzen, meerdere assessmentsituaties en meerdere beoordelaars.
Vergelijkbaarheid	De opdrachten, beoordelingsprocedure en beoordelingscriteria zijn vergelijkbaar voor alle studenten.
Eerlijkheid	Alle studenten krijgen een eerlijke kans hun competenties te demonstreren, bijvoorbeeld door veel verschillende soorten opdrachten en bewaarde beoordelaars te gebruiken. Studenten ervaren de beoordeling ook als eerlijk.
Acceptatie	Alle betrokkenen bij een CAP kunnen zich vinden in de beoordelingscriteria, toetsvormen en procedures.
Betekenisvolheid	De beoordeling levert waardevolle feedback op voor studenten en is een leermoment.
Onderwijsgevolgen	De effecten van de beoordeling op het leren van studenten, en op de instructie of begeleiding die wordt gegeven door docenten.
Tijd en kosten	De haalbaarheid van de uitvoering van CAP voor studenten en beoordelaars.

Tabel 2
De drie kijkkaders met bijbehorende analysevragen en kwaliteitscriteria

	Kijkkader	Analysevragen	Kwaliteitscriteria
1.	Kenmerken van startende beroepsbeoefenaren	Worden alle kenmerken van een goede beroepsbeoefenaar beoordeeld? Is de combinatie van gebruikte toetsvormen geschikt voor het beoordelen van alle kenmerken van de beroepsbeoefenaar?	Fitness for purpose Authenticiteit Cognitieve complexiteit Eerlijkheid Transparantie
2.	Aansluiting tussen school en werk	In hoeverre is de toetsing representatief voor de toekomstige beroepspraktijk? In hoeverre is het werkveld betrokken bij de toetsing? Hoe ziet hun rol eruit?	Authenticiteit Transparantie Acceptatie Herhaalbaarheid
3.	Beoogde leerprocessen	In hoeverre wordt toetsing bewust ingezet om het leren te stimuleren? In hoeverre levert de toetsing feedback op? In hoeverre ervaren studenten de toetsing als een leermoment?	Ontwikkeling van zelfsturend leren Onderwijsgevolgen Betekenisvolheid Alignment onderwijs-toetsing Alignment stage-toetsing

onafhankelijk van elkaar geanalyseerd, resulterend in een samenvatting per kijkkader met daarin steeds een expliciete verwijzing naar de kwaliteitscriteria behorende bij het kijkkader. Vergelijking en overleg hebben geresulteerd in één samenvatting voor de hbo-casussen en één samenvatting voor de mbo-casussen.

2.2 De hbo-casussen

In dit onderzoek is de kwaliteit van het toetsprogramma van negen opleidingen in een grote hbo-instelling in Nederland onderzocht (voor de volledige studie, zie Baartman, Gulikers & Dijkstra, 2013). Bij elk toetsprogramma gaat het om een combinatie van verschillende toetsvormen, bijvoorbeeld alle toetsen in het eerste leerjaar van de opleiding of in de hoofdfase van de opleiding. De opleidingen voerden een zelfevaluatie uit met 4-11 docenten en 3-10 studenten. De zelfevaluatie bestond uit drie fasen. In de eerste fase werd een training verzorgd, waarin een uitleg werd gegeven over de 12 kwaliteitscriteria en het verzamelen van bewijzen voor kwaliteit. Ook werd het precieze toetsprogramma bepaald om ervoor te zorgen dat alle evaluatoren hetzelfde uitgangspunt voor de evaluatie hanteerden. In de tweede fase vulden alle evaluatoren individueel een elektronische vragenlijst

in waarmee ze de kwaliteit van hun toetsprogramma beoordeelden. In deze vragenlijst zijn de 12 kwaliteitscriteria verder uitgewerkt in 4-6 indicatoren per criterium te scoren op “in hoeverre voldoet uw toetsprogramma aan deze indicator” op een schaal van 1-100% en te onderbouwen met concreet bewijs. De derde fase bestond uit een groepsinterview waarin op een gestructureerde wijze de resultaten van de individuele evaluaties met elkaar werden geconfronteerd en besproken, waardoor de evaluatie van het toetsprogramma vanuit verschillende invalshoeken werd beraamd. Gezamenlijk werden sterke en zwakke kwaliteitsaspecten benoemd.

2.3 De mbo-casussen

In dit onderzoek is de kwaliteit van de summatieve toetsing onderzocht, zoals die wordt gebruikt binnen de Agrarische OpleidingsCentra (AOC's) (Gulikers, Biemans, & Mulder, 2009). De casussen zijn representatief voor een groot deel van de opleidingen zoals die worden verzorgd binnen de AOC's. Twaalf AOC's hebben in samenwerking met kenniscentrum Aequor en in nauwe afstemming met brancheorganisaties voor alle opleidingen summatieve Proeven van Bekwaamheid (PvB's) ontwikkeld (zie: www.degroenestandaard.nl). Iedere PvB bestaat uit

een combinatie van twee toetsvormen: (1) een observatie op het bedrijf waar een leerling laat zien dat hij de kritische beroepssituatie kan uitvoeren (“does”) en (2) een criteriumgericht interview waarin de leerling zijn keuzes moet beargumenteren (“knows how”). Steeds zijn twee assessoren betrokken: een docent en een praktijkbeoordelaar, meestal stagebegeleider. Iedere PvB is dus een klein toetsprogramma an sich. Een combinatie van een aantal van deze PvB’s (variërend van 2-4 per opleiding) vormt het toetsprogramma van de hele opleiding.

Deze casus beschrijft de implementatie van twee van deze PvB’s, beide in een diergerelateerde mbo-opleiding (niveau 2 en 3). De kwaliteit van de geïmplementeerde PvB’s is geanalyseerd aan de hand van bovengenoemde 12 kwaliteitscriteria, met enkele aanpassingen voor het specifieke doel van het originele onderzoek (Gulikers, Biemans, & Mulder, 2009). Het criterium “onderwijsgevolgen” is verder opgesplitst in het stimuleren van (1) reflectie en persoonlijke ontwikkeling, (2) generieke competentieontwikkeling, en (3) motivatie. Het criterium “alignment” (Biggs, 1996) is toegevoegd als extra kwaliteitscriterium, omdat uit eerder onderzoek is gebleken dat dit een cruciale voorwaarde is om te komen tot een goede toetskwaliteit (Gulikers, Kester, Kirschner, & Bastiaens, 2008). Het criterium alignment vraagt expliciet naar de relatie tussen de toetsactiviteiten en de leeractiviteiten op (1) school en (2) stage. Voor de evaluatie van de twee mbo-casussen zijn individuele evaluatielijsten uitgezet bij studenten, docent-assessoren, praktijkassessoren en toetsontwikkelaars. Met deze vier doelgroepen zijn vervolgens semi-structureerde individuele of groepsinterviews afgenomen gestructureerd langs de kwaliteitscriteria.

3 Resultaten

De resultaten beschrijven de belangrijkste bevindingen van de hbo en mbo toetsprogramma’s op de drie kijkkaders met tussen haakjes steeds de bijbehorende kwaliteitscriteria.

3.1 Hbo-casussen

Professionele bagage en professionele taakuitvoering

In de hbo-casussen wordt de kennisbasis van de startend beroepsbeoefenaar breed getoetst door middel van “knows” en “knows how” toetsen. Dit gebeurt bij de afzonderlijke vakken in het curriculum, maar ook in thematoetsen. Knows-how, met name methodologische en/of theoretische onderbouwing, krijgt veel aandacht in de vorm van verantwoordingsgesprekken (cognitieve complexiteit), vaak gekoppeld aan projecten of performance assessments. Studenten en docenten vinden ook dat de toetsen zijn gericht op het denkproces, inzicht en toepassen van kennis (cognitieve complexiteit) waar zij ook veel waarde aan hechten (fitness for purpose). Docent: “kennis is de sleutel tot het nemen van goede beslissingen en handelen”. Een andere docent merkt op “zonder kennis is het niet mogelijk een competentie te laten zien. De student kan een trucje laten zien, maar bij de assessments zit aan het eind een verantwoordingsgesprek waaruit moet blijken dat de student ook kennis van zaken heeft”. Studenten vinden dat ze veel dingen moeten beargumenteren: “bij toepassingsvragen wordt bijvoorbeeld gevraagd waarom of licht toe”. Voor het toetsen van de professionele taakuitvoering maken de meeste hbo-opleidingen gebruik van projecten, performance assessments en/of stagebeoordeling (“shows how” of “does” toetsing). Echter, bij projecten merken de studenten op dat de beoordelingscriteria meer zijn gericht op kennis dan op probleemoplossend handelen en de daadwerkelijke performance in de praktijk (fitness for purpose). Daarnaast benadrukken docenten en studenten dat het in de toetsen altijd gaat om een selectie uit de totale kennisbasis: niet alle kennis kan aan de orde komen in de toetsing (fitness for purpose). De beoordeling van attitudes wordt over het algemeen lager gescoord, al zijn hier grote verschillen te zien tussen de verschillende opleidingen. In sommige opleidingen is weinig aandacht voor het beoordelen van attitudes: “studenten doen hier lacherig over”, “je kunt toetsen halen zonder aanwezig te zijn” en “dit is niet zichtbaar voor de docent, want die ziet studenten niet aan het werk”. Kortom, bij de hbo-opleidingen ligt de nadruk op het toetsen

van de kennisbasis van de startend beroepsbeoefenaar en minder op het uitvoeren van professionele taken in de praktijk. Een aandachtspunt voor kijkkader 1 is dat veel van de hbo-opleidingen werken met beroepscompetentieprofielen voor de opleiding, maar dat deze meestal niet zijn terug te vinden in de leerdoelen per cursus en de uiteindelijke toetsdoelen en beoordelingscriteria (transparantie). Vaak wordt opgemerkt dat deze opleidingscompetenties wel in de studiehandleiding staan genoemd, maar studenten hier niet naar kijken en de competenties ook niet worden benoemd of gebruikt in de dagelijkse lespraktijk.

De kern van het beroep in het werkveld

De betrokkenheid van het werkveld bij de toetsing is in hbo-casussen over het algemeen laag. In hbo vindt de meeste toetsing op school plaats. Dit geldt voor de kennisgerichte toetsing, maar ook voor de toetsing van de professionele taakuitvoering, die wordt getoetst via bijvoorbeeld projecten en simulaties. Studenten beoordelen hun toetsprogramma in de meeste gevallen echter wel als representatief voor hun toekomstige beroepspraktijk en als goed voorbereidend op die beroepspraktijk (authentiek), met de kanttekening dat “de sfeer in de klas relaxter” en “informeler dan met een echte patiënt en behandelaar” is. Het werkveld is met name betrokken bij de beoordeling op de stage, waar de stagebegeleider optreedt als mede-beoordelaar. Het eindoordeel ligt bij de docent van de opleiding (herhaalbaarheid). Het werkveld heeft vaak geen duidelijk beeld van de competenties en de verschillende beoordelingsvormen in de totale opleiding (transparantie). Veel docenten/studenten zeggen niet te weten hoe het bedrijfsleven denkt over de beoordeling: “ze hebben niet echt een mening en doen weinig moeilijk over de invulling van de opleiding. Ze nemen aan dat het goed zit” (acceptatie). Studenten vinden dat hun docenten een goed beeld hebben van de beroepspraktijk, omdat ze vaak zelf werkzaam zijn in de praktijk (authenticiteit). Eén opleiding beschrijft dat de school het stagebedrijf wel op de hoogte brengt en een pakket aan informatie meegEEft, maar noemt de communicatie met het werkveld wel “een ondergeschoven kindje”

(transparantie). Een andere opleiding geeft aan graag meer te willen samenwerken met de beroepspraktijk om de beroepscompetenties om te zetten naar les/toetsmaterialen.

Het ontwikkelvermogen van de startend beroepsbeoefenaar

In de hbo-casussen wordt toetsing meestal niet bewust ingezet om het leren te stimuleren. Studenten en docenten zijn zich weinig bewust van de invloed van verschillende toetsvormen op het leerproces (onderwijsgevolgen). Studenten noemen dat zij vooral leren om een toets te halen: “als je goed leert, krijg je hoge punten” en “je wilt het halen, dus daarom bereid je je goed voor”. Studenten zien beoordeling over het algemeen als een “docenten-ding”. Ze zien de toetsen ook niet als een leermoment, maar als een verplicht nummer of iets dat ze moeten inleveren. Soms zien ze achteraf wel dat ze veel hebben geleerd van een toetsmoment. Op het moment van de toetsing zelf zien ze dit soms niet, wat bij performance assessments ook komt door de stress van het toetsmoment zelf (betekenisvolheid). Studenten vinden de opdrachten leerzaam, willen opdrachten graag nabespreken en zijn zich bij bijvoorbeeld tussentijdse presentaties duidelijk bewust van het feit dat ze wat leren. Studenten waarderen feedback op basis van de beoordelingscriteria, want “zo kun je precies zien wat je goed doet en waar je aan moet werken” (betekenisvolheid).

Bij de toetsing in de hbo-casussen is over het algemeen weinig aandacht voor leren leren, er wordt weinig gebruik gemaakt van peer assessment en de studenten hebben weinig inbreng in de toetsing (ontwikkeling van zelfsturend leren). Studenten zijn wel positief over peer assessment als dit wordt gebruikt “dit werkt beter dan een docent die er maar omheen praat” of “ik vond het in het begin heel moeilijk om feedback te geven tijdens het oefenen, maar nu gaat dat veel beter” (betekenisvolheid). Het formuleren van eigen leerdoelen gebeurt niet veel. Studenten moeten voldoen aan de vastgestelde leerdoelen en in een volgende beoordeling wordt niet meer teruggegrepen op eerder gegeven feedback en genoemde verbeterpunten (betekenisvolheid; ontwikkeling van zelfsturend leren). Docen-

ten geven aan vaak niet voldoende tijd te hebben om goede feedback te geven, een beoordeling na te bespreken met studenten of leerdoelen op te stellen met een student.

3.2 Mbo-casussen

Professionele bagage en professionele taakuitvoering

De combinatie van toetsvormen in de mbo-casussen doelt in eerste instantie op professionele taakuitvoering en vervolgens onderbouwing op basis van de professionele kennisbasis. Betrokkenen zijn unaniem in de waardering van de toets voor professionele taakuitvoering (fitness for purpose). Echter, het doel van het CGI is voor leerlingen niet zo helder; zij zien dit niet als een “knows how”-toets gekoppeld aan eerder vertoond gedrag (fitness for purpose; transparantie). De meeste leerlingen verwachtten meer kennisvragen in het CGI. Met andere woorden, de kennisbasis van de beroepsbeoefenaar wordt in de ogen van de leerlingen niet beoordeeld met deze manier van toetsen (fitness for purpose). Praktijkbeoordelaars mbo-3 vinden juist dat deze combinatie van toetsvormen de kern van beroepsbekwaamheid beoordeelt (authenticiteit): “je kijkt echt naar het dagelijkse functioneren op de werkplek in zijn geheel” en “in het CGI worden precies die vragen gesteld die ik mezelf eigenlijk ook iedere dag zou moeten stellen” (praktijkbeoordelaar). Volgens hen wordt de kennisbasis dus op een juiste manier beoordeeld door deze manier van toetsen (fitness for purpose; authenticiteit). Bij mbo-2 zien alle deelnemers de PvB voornamelijk als een beroepsrelevante handelingstoets (fitness for purpose). Het toetsen van een kennisbasis of toepassing van een kennisbasis vinden zij minder relevant voor mbo-2 leerlingen (cognitieve complexiteit). Het nut en de bedoeling van het CGI blijkt dan ook voor de mbo-2 docenten en vooral praktijkbeoordelaars niet helder (fitness for purpose; transparantie).

Is deze PvB dan een volwaardige combinatie van toetsmethoden om de relevante aspecten van beroepsbekwaamheid (professionele bagage en taakuitvoering) te beoordelen? Beoordelaars en leerlingen in beide casussen geven aan dat zij deze PvB niet volledig vinden voor het tonen én beoordelen van de benoemde werkprocessen en beoorde-

lingscriteria (eerlijkheid). De twee belangrijkste redenen hiervoor zijn (a) niet alle werkprocessen zijn zichtbaar op ieder bedrijf of op het moment van observatie, en (b) de leerling heeft al zijn werkprocessen eerder laten zien in een voorbereidend portfolio of gedurende zijn stage. Het expliciet benoemen van het portfolio als onderdeel van het toetsprogramma, tezamen met de PvB's, zou hier een oplossing kunnen bieden.

De kern van het beroep in het werkveld

Leerlingen, beoordelaars en ontwikkelaars ervaren de PvB's als representatief voor de toekomstige beroepspraktijk (authenticiteit). Authenticiteit was ook het belangrijkste uitgangspunt bij de ontwikkeling van deze mbo: het werkveld is nauw betrokken geweest bij de ontwikkeling van de PvB's. De kerntaken en de daaruit volgende beschrijving van kritische beroepssituaties voor de PvB's zijn gezamenlijk beschreven door docenten en beroepsbeoefenaars (of branchevertegenwoordigers) en worden door praktijkbeoordelaars duidelijk herkend als relevant en representatief voor het beroep (acceptatie; authenticiteit). De beoordeling wordt uitgevoerd door een docent en een praktijkbeoordelaar (herhaalbaarheid). Leerlingen waarderen met name de nauwe betrokkenheid van hun stagebegeleider als mede-beoordelaar bij deze toetsing en de uitvoering van de toetsing op hun stageplek (authenticiteit; acceptatie). Leerlingen en praktijkbeoordelaars geven aan dat voor hen helder is waarop wordt beoordeeld (transparantie). Uit de interviews blijkt echter dat zij vooral uitgaan van het holistische totaaloordeel “kan de leerling de kritische beroepssituatie aan?” en geen notie hebben van onderliggende kerntaken, werkprocessen, competenties of andere beoordelingscriteria.

Het ontwikkelvermogen van de startend beroepsbeoefenaar

Wat betreft het laatste kenmerk van de goede beroepsbeoefenaar (het ontwikkelvermogen), kunnen we kort zijn: hier is geen aandacht voor in de toetsing van de mbo-opleidingen. Belangrijk om te vermelden is dat het beoordelen van ontwikkelvermogen ook niet het doel was van deze toetsing. In de mbo-casus-

sen wordt de PvB expliciet niet ingezet om het leren van leerlingen, op wat voor manier dan ook, te stimuleren. Leerlingen, beoordelaars en toetsontwikkelaars ervaren de PvB dan ook niet als een leermoment (betekenisvolheid). Toetsontwikkelaars geven aan dat deze toetsing hier niet voor is bedoeld, omdat het een summatieve toets is. De summatieve en formatieve functies van toetsing zijn strikt gescheiden en feedback geven vinden zij daarom geen onderdeel van deze summatieve toetsing (ontwikkeling van zelfgestuurd leren). Zelf- of peer-beoordeling vindt in zijn geheel niet plaats. Opvallend is wel dat docenten en praktijkopleiders verwachten dat de PvB automatisch aanzet tot reflectie, terwijl leerlingen op zowel mbo-2 als mbo-3 niveau dit kenmerk laag scoren (onderwijsgevolgen). De leerlingen van beide opleidingen ervaren eerder een negatieve relatie tussen toetsing en leren omdat zij geen alignment zien tussen datgene wat ze op school leren en wat ze in de PvB moeten laten zien (alignment onderwijs-toetsing). Daarentegen scoren zij de alignment tussen de PvB en hun stage unaniem hoog. Docenten beoordelen de alignment tussen leeractiviteiten op school en de PvB echter wel goed. Met andere woorden, zij verwachten dat schoolse leeractiviteiten leerlingen goed voorbereiden op de PvB.

4 Conclusies

Het doel van dit artikel was om de huidige toetsing in het beroeps onderwijs te analyseren vanuit het inhoudsvraagstuk, uitgewerkt in een drietal kijkkaders die tezamen beschrijven wat het betekent om een goede startende beroepsbeoefenaar te zijn. Dit hebben we gedaan door een kritische (secundaire) analyse uit te voeren van 9 toetsprogramma's uit het hbo en 2 uit het mbo. De drie onderzoeksvragen behorende bij de drie kijkkaders waren in hoeverre (1) de professionele bagage en professionele taakuitvoering en (2) de kern van het beroep in het werkveld worden beoordeeld, en (3) de beoordeling het ontwikkelvermogen van de startend beroepsbeoefenaar stimuleert. Op basis van de verschillen en overeenkomsten uit de mbo en hbo-casussen vallen enkele belangrijke conclusies te trek-

ken als het gaat om het zo goed mogelijk beoordelen van toekomstige beroepsbeoefenaren.

Wat betreft de eerste onderzoeksvraag laten de hbo-casussen een nadruk zien op toetsing van de kennisbasis door gebruik van "knows" en "knows how" toetsvormen, terwijl in het mbo vooral de toepassing in de professionele taakuitvoering wordt beoordeeld door middel van "does" toetsen op de stageplek. De "shows-how" en "does" toetsen in het hbo worden voornamelijk op kennisgerichte beoordelingscriteria beoordeeld, terwijl in het mbo de koppeling van de "does" toets en de "knows-how" toets (CGI) vanuit theoretisch oogpunt sterk is, maar de leerlingen dit niet als een "knows-how" toets herkennen en ook beoordelaars het nut en gebruik van het CGI niet helder voor ogen hebben. Beide casussen laten zien dat het bewust nadenken over de inhoudsvraag en het hierbij bewust kiezen en inzetten van een combinatie van toetsvormen (een toetsprogramma) om alle onderdelen van een startend beroepsbeoefenaar te kunnen beoordelen, niet kenmerkend is voor mbo en hbo toetspraktijken. De kern van toetskwaliteit ligt echter in deze geschiktheid voor onderwijsdoelen (Baartman et al., 2007).

Wat betreft de tweede onderzoeksvraag zijn zowel het hbo als het mbo in eerste instantie positief over de mate waarin het toetsprogramma de kern van het beroep beoordeelt. Echter de betrokkenheid en acceptatie van het werkveld is in het hbo nog gering, terwijl eerder onderzoek (Tynjälä, 2008) het belang van een goede communicatie tussen school en het bedrijfsleven benadrukt. In het mbo is er veel werkveldbetrokkenheid, maar deze casussen laten wederom zien (zie ook Poortman et al., 2011) dat het 'simpelweg' toetsen op de werkplek niet garandeert dat de kern van het beroep wordt beoordeeld. Een toets in de authentieke beroepssituatie is niet automatisch een volledig valide toets (Van der Vleuten et al., 2010). Het vijf-dimensionale model van Gulikers en collega's (2004) laat ook zien dat de fysieke context (de werkplek) slechts één van de vijf toetskenmerken is die authentieker gemaakt kunnen worden. Bovendien blijken een authentieke toetstaak en authentieke toetscri-

teria van sterkere invloed te zijn op de prestatie van studenten dan de authentieke fysieke context (Gulikers et al., 2008). Dit sluit aan bij de argumenten van Van der Vleuten en collega's (2010) dat de stimulus van een toets (= de toetstaak) de validiteit van de toets bepaalt. Concluderend, ook op de werkplek moet worden nagedacht over het inrichten van een specifieke toetsituatie, waarin de te beoordelen competenties op het juiste niveau worden aangesproken.

Met betrekking tot het derde kijkkader blijkt de relatie tussen leren, ontwikkelvermogen, en toetsen zowel in de mbo als hbo-casussen nog voor verbetering vatbaar. Kijkkader 3 geeft aan dat toetsing het leren stuurt via feedup, feedback en feedforward (Hattie & Timperley, 2007). Mbo-leerlingen ervaren een mismatch tussen wat en hoe ze leren op school en wat ze moeten doen in de praktijktoetsing, terwijl docenten deze mismatch niet zien. Bovendien was het voor leerlingen niet transparant wat er in het CGI van hen werd verwacht en zijn zij op school niet voorbereid op het soort vragen (i.e., waarom/onderbouwingsvragen) dat hierin werd gesteld. Ook laten de casussen zien er weinig feedback op prestaties wordt gegeven, omdat (1) docenten aangeven hier geen tijd voor te hebben of (2) men vindt of denkt dat een summatieve toets niet bedoeld is om feedback te geven. Wanneer er feedback mogelijkheden zijn, of gebruik wordt gemaakt van bijvoorbeeld peer-assessments, blijken latere toetsen hier niet op terug te grijpen. Het nut van formatieve peer of self-assessments en feedback wordt uiteraard te niet gedaan als er geen mogelijkheid is om iets met deze feedback te doen in een nieuwe (toets)situatie (Nicol & Macfarlane-Dick, 2006; Wiliam, 2011).

Tot slot kan een aantal algemene conclusies worden getrokken, over de verschillende kijkkaders heen. Ten eerste blijkt uit alle kijkkaders dat er bij het vormgeven van toetsing duidelijker vanuit het inhoudsvraagstuk moet worden gestart. Dit betekent dat toetsing, en de opbouw van toetsprogramma's, meer moet worden benaderd als een ontwerp-vraagstuk en als integraal onderdeel van curriculumontwerp (Sluijsmans et al., 2008; Sluijsmans, Peeters, Jakobs, & Weijzen 2012). De startvraag is steeds de operationalisatie van de

inhoudsvraag: wat zijn de opleidingskwalificaties, hoe kunnen deze kwalificaties worden uitgewerkt in niveaus en beoogde leeropbrengsten voor verschillende onderdelen van het curriculum en uiteindelijk worden vertaald in beoordelingscriteria of scoringsmodellen voor de toetsen (Baartman, Kloppenburg & Prins, 2013).

Ten tweede speelt dit starten bij het inhoudsvraagstuk ook een rol bij losse toetsen. De toetspraktijk in het Nederlandse beroeps- onderwijs laat bijvoorbeeld zien dat er steeds meer wordt beoordeeld in de professionele beroepspraktijk, maar dat beoordelaars vaak niet goed weten wat ze hierin precies willen beoordelen (Gulikers, 2010). Bij het beoordelen in de professionele taakuitvoering ("shows how" en "does") moeten beoordelaars zich buigen over vragen als: wat wil ik precies zien of horen om te beoordelen of een student competent is in "beslissingen en activiteiten initiëren"? Bij een CGI spelen vragen als: wat wil ik van de student horen om te achterhalen of hij/zij adequate beslissingen kan nemen en bewust heeft gehandeld? Dit vraagt gedegen aandacht in het toetsontwerpproces waarin beoordelaars, bij voorkeur gezamenlijk, operationaliseren wat een competentie precies betekent in een specifieke toetsituatie. Sluijsmans en collega's (2012) noemen de professionalisering van docenten in het operationaliseren van deze leer-/toetsdoelen als een kernvoorwaarde voor kwalitatief goede toetsing in het beroepsonderwijs. Ook Gulikers en Pat-el (2013) laten zien dat de kritische succesfactor voor de invoering van competentiegericht beoordelen in het vmbo ligt in het begrip van wat competenties zijn en hoe je die in onderwijs en toetsing kunt operationaliseren.

Ten derde worden de mogelijkheden van toetsing om het ontwikkelvermogen te stimuleren nog niet volledig benut. Zo wordt in de meeste casussen een strikte scheiding gehanteerd tussen formatief en summatief beoordelen en worden summatieve toetsen niet gezien als leermoment, terwijl een summatief toetsmoment veel kansen biedt voor verdere ontwikkeling (Carless, 2007). Hiervoor verdient het aanbeveling, zeker ook in het kader van effectief én efficiënter toetsen, om toetsprogramma's te ontwikkelen die de formatie-

ve en summatieve doelen verenigen (Van der Vleuten et al., 2012) en te experimenteren met de formatieve functie van toetsen die in eerste instantie een summatief doel hebben (Carless, 2007; Schuurs & Verhoeven, 2010). Ook voor het stimuleren en kunnen beoordelen van transfer is het denken in termen van een toetsprogramma vereist. Dit noodzaakt een diversiteit aan beroepssituaties waarin de student transfer kan laten zien en beargumenteren (“geldt dit alleen in deze specifieke situatie? Zou je het in een andere situatie ook zo aanpakken?”) en waarin de student feedback en leerdoelen kan meenemen naar een volgende toetsituatie.

Tot slot legt de huidige literatuur over beoordeling over het algemeen geen expliciete koppeling tussen beoordeling en de daadwerkelijke leerprocessen zoals we die beogen in het beroepsonderwijs. Literatuur over het effect van verschillende vormen van toetsing op het leren laat wisselende effecten zien en leren wordt in deze studies vaak geoperationaliseerd als bijvoorbeeld diep leren of oppervlakkig leren (Gijbels & Dochy, 2006; Gulikers et al., 2008), wat nog weinig concrete informatie geeft over de daadwerkelijke leerprocessen van studenten. Carless (2007, p.57) benadrukt dat “All assessments lead to some kind of student learning, but a fundamental challenge is to stimulate the right kind of learning”. In vervolgonderzoek zou daarom een link gelegd moeten worden tussen literatuur over bijvoorbeeld transparantie van beoordelingscriteria, feedback en effecten van verschillende beoordelingsvormen op het leren, en literatuur over leren in beroepsonderwijs, bijvoorbeeld over integratie (Baartman & De Bruijn, 2011), boundary crossing (Akkerman & Bakker, 2011) en recontextualiseren (Van Oers, 1998).

Literatuur

- Akkerman, S. F., & Bakker, A. (2011). Boundary crossing and boundary objects. *Review of Educational Research, 81*, 132-169.
- Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2007). Evaluating assessment quality in competence-based education: A qualitative comparison of two frameworks. *Educational Research Review, 2*, 114-129.
- Baartman, L. K. J., & De Bruijn, E. (2011). Integrating knowledge, skills and attitudes: Conceptualising learning processes towards vocational competence. *Educational Research Review, 6*(2), 125-134.
- Baartman, L. K. J., Gulikers, J. T. M., & Dijkstra, A. (2013). Factors influencing assessment quality in higher vocational education. *Assessment & Evaluation in Higher Education*. DOI:10.1080/02602938.2013.771133.
- Baartman, L. K. J., Kloppenburg, R., & Prins, F. J. (2013). Kwaliteit van toetsprogramma's. In A. Bax, H. Van Berkel & D. Joosten-Ten Brinke, (Eds.), *Toetsen in het Hoger Onderwijs* (3e druk, pp. 51-62). Houten: Bohn Stafleu van Loghum.
- Baartman, L. K. J., Prins, F. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2011). Self-evaluation of assessment programs: A cross-case analysis. *Evaluation and Program Planning, 34*, 206-216.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education, 32*(3), 347.
- Boud, D. (2000). Sustainable Assessment: rethinking assessment for the learning society. *Studies in Continuing Education, 22*, 151-167.
- Carless, D. (2007). Learning-oriented assessments: conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*(1), 57 - 66.
- Cennamo, K., Brandt, C., Scott, B., Douglas, S., McGrath, M., Reimer, Y., & Vernon, M. (2011). Managing the complexity of design problems through studio-based learning. *The Interdisciplinary Journal of Problem-based Learning, 5*(2), 12-36.
- Cumming, J. J., & Maxwell, G. S. (1999). Contextualising authentic assessment. *Assessment in Education: Principles, Policy & Practice, 6*, 177-194.

- Davies, J., & Ecclestone, K. (2008). 'Straitjacket' or 'springboard' for sustainable learning? The implications of formative assessment practices in vocational learning cultures. *Curriculum Journal*, 19(2), 71 - 86.
- Gijbels, D., & Dochy, F. J. R. C. (2006). Students' assessment preferences and approaches to learning: Can formative assessment make a difference? *Educational Studies*, 32, 399-409.
- Gulikers (2010). Groen Proeven, verslag van monitoren groen proeven pilots 2009-2010. Ede: AOC raad. Retrieved from www.groenproeven.nl
- Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Design*, 52(3), 67-85.
- Gulikers, J., Biemans, H., & Mulder, M. (2009). Developer, teacher, student and employer evaluations of competence-based assessment quality. *Studies in Educational Evaluation*, 35(2-3), 110-119.
- Gulikers, J. T. M., Kester, L., Kirschner, P. A., & Bastiaens, T. J. (2008). The effect of practical experience on perceptions of assessment authenticity, study approach, and learning outcomes. *Learning and Instruction*, 18, 172-186.
- Gulikers, J. & Pat-el, R. (2013, August). A comprehensive approach to assessment innovations. Paper presented at the EARLI conference, München, Germany.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77, 81-112.
- Hinds, P. S., Vogel, R. J., & Clarke-Steffen, L. (1997). The Possibilities and Pitfalls of Doing a Secondary Analysis of a Qualitative Data Set. *Qualitative Health Research*, 7, 408-424.
- Kloppenburg, R. (2011). Bekwaam beoordeeld. Inhoud, functie en kwaliteit van competentiegerichte assessments in social work opleidingen. Unpublished doctoral dissertation. Universiteit Utrecht, the Netherlands.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41, 212-218.
- Mathers, N. J., Challis, M. C., Howe, A. C., & Field, N. J. (1999). Portfolios in continuing medical education – effective and efficient? *Medical Education*, 33, 521-530.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook*. Thousand Oaks, CA: Sage Publications.
- Miller, G. E. (1990). Assessment of clinical skills/competence/performance. *Academic Medicine*, 65(9), 63-67.
- Mulder (in press). Conceptions of professional competence. In S. Billett, C. Harteis, H. Gruber (Eds), *International Handbook on Research into professional and practice-based learning*. Springer.
- Mulder, M., Wesselink, R. & Bruijstens, H. (2005). Job profile research for the purchasing profession. *International Journal of Training and Development*, 9, 185-204
- Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31, 199-218.
- Poortman, C.L., Illeris, K. & Nieuwenhuis, L. (2011). Apprenticeship: from learning theory to practice. *Journal of Vocational Education and Training*, 63, 267-287.
- Robinson, M. (2012). *What is Job Analysis?* Retrieved April 18, 2013, from http://esrccoi.group.shef.ac.uk/pdf/whatis/job_analysis.pdf.
- Roelofs, E. (2006). Een procesmodel voor de beoordeling van competent handelen. *Tijdschrift voor Hoger Onderwijs*, 24, 152-167.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119.
- Schaap, H., De Bruijn, E., Van der Schaaf, M. F., & Kirschner, P. A. (2009). Students' personal professional theories in competence-based vocational education: The construction of personal knowledge through internalisation and socialisation. *Journal of Vocational Education & Training*, 61, 81-494.
- Schuurs, U., & Verhoeven, L. (2010). Meten van leerprestaties in het (V)MBO: *Assessment for learning en assessment of learning*. NWO-rapport. Retrieved from [http://www.nwo.nl/files.nsf/pages/NWOP_89GEN3/\\$file/2010-Verhoeven_Schuurs.pdf](http://www.nwo.nl/files.nsf/pages/NWOP_89GEN3/$file/2010-Verhoeven_Schuurs.pdf)
- Sluijsmans, D. M. A., Brand-Gruwel, S., & Van Merriënboer, J. J. G. (2002). Peer assessment training in teacher education. Effects on performance and perceptions. *Assessment and Evaluation in Higher Education*, 27, 443-454.
- Sluijsmans, D., Peeters, A., Jakobs, L., & Weijzen, S. (2012). De kwaliteit van toetsing onder de loep. *Onderwijsinnovatie*, 4, 17-25.

- Sluijsmans, D. M. A., Straetmans, G. J. J. M., & van Merriënboer, J. J. G. (2008). Integrating authentic assessment with competence-based learning in vocational education: the Protocol Portfolio Scoring. *Journal of Vocational Education & Training*, 60, 159-172.
- Tillema, H. H., Kessels, J. W. M., & Meijers, F. (2000). Competencies as building blocks for integrating assessment with instruction in vocational education: a case from the Netherlands. *Assessment and Evaluation in Higher Education*, 25, 265-278.
- Toohy, S., Ryan, G., & Hughes, C. (1996). Assessing the practicum. *Assessment & Evaluation in Higher Education*, 21, 215 - 227.
- Tuomi-Gröhn, T., Engeström, Y. (2003). *Between school and work: New perspectives on transfer and boundary-crossing*. Amsterdam, the Netherlands: Pergamon.
- Tynjälä, P. (2008). Perspectives into learning at the workplace. *Educational Research Review*, 3, 130-154.
- Van den Berg, N., & De Bruijn, E. (2009). *Het glas vult zich. Kennis over vormgeving en effecten van competentiegericht beroepsonderwijs*. Retrieved from <http://www.ecbo.nl/ECBO/downloads/publicaties/Eindrapportage%20review%20cgo%20uitgebreide%20versie.pdf>
- Van der Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & Van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, 34, 205-214.
- Van der Vleuten, C. P. M., Schuwirth, L. W. T., Scheele, F., Driessen, E. W., & Hodges, B. (2010). The assessment of professional competence: building blocks for theory development. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 24, 703-719.
- Van Merriënboer, J.J.G. (1997). *Training complex cognitive skills*. Englewood Cliffs, NJ: Educational Technology Publications.
- Van Oers, B. (1998). From context to contextualising. *Learning and Instruction*, 8, 473-488.
- Van Tartwijk, J., Driessen, E., Van der Vleuten, C., Stokking, K. (2007). Factors influencing the successful introduction of portfolios. *Quality in Higher Education*, 13, 69-79.
- Wade, R. C., & Yarbrough, D. B. (1996). Portfolios: a tool for reflective thinking in teacher education. *Teaching and Teacher Education*, 12, 63-79.
- William, D. (2011). What is assessment for learning? *Studies in Educational Evaluation*, 37, 3-14.

Auteurs

Liesbeth Baartman is werkzaam bij de Hogeschool Utrecht, Lectoraat Beroepsonderwijs. **Judith Gulikers** werkt bij Wageningen Universiteit, Educatie- en competentiestudies.

Correspondentieadres: liesbeth.baartman@hu.nl

Abstract

Assessment as the cornerstone of high quality vocational education: an analysis of 11 assessment programmes in senior secondary and higher vocational education

The choice and implementation of (combinations) of assessment methods is an important prerequisite for high quality vocational education. Research and practical implementation often start with the question how to assess students instead of the question what needs to be assessed. Therefore, the goal of this article is to analyse current assessment methods in vocational education, starting from the content perspective of what needs to be assessed. Three themes were formulated that together describe what it means to be a good starting professional in the workplace. Next, 11 assessment programmes in senior secondary and higher vocational education were analysed. Research questions were: (1) to what extent do the assessments capture professional knowledge and performance, (2) to what extent do the assessment methods capture the core of the vocation in the workplace, and (3) to what extent do the assessments stimulate the capacity for further lifelong learning. Similarities and differences between the cases in senior secondary and higher vocational education show the need to approach assessment as a design question and to think in term of assessment programmes instead of single assessments alone. Starting from the content of what needs to be assessed, the assessments can be adequately operationalized on the level of assessment programmes and curricula and single assessments.