

Wat is de overeenstemming tussen competentiebeoordeling door middel van portfolio en het oordeel over competent handelen in de praktijk?

Drs. S. Weijzen
(s.weijzen@han.nl) is werkzaam bij NDO, faculteit Educatie, Hogeschool van Arnhem en Nijmegen. Dr. M. van den Hurk is werkzaam bij Onderwijskunde, Radboud Universiteit Nijmegen.

Inzet van portfolio's om competenties te beoordelen betekent onherroepelijk discussie over validiteit en betrouwbaarheid. Ten behoeve van dit onderzoek is een beoordelingsprocedure ontwikkeld voor derdejaarsstudenten ergotherapie aan de Hogeschool van Arnhem en Nijmegen. Centraal in dit onderzoek staat de vraag in welke mate het oordeel van docenten over competenties van studenten, aangetoond middels portfolio, overeenkomt met het oordeel van praktijkbegeleiders over competenties van studenten, aangetoond middels praktisch handelen in de stage, met behulp van eenzelfde beoordelingsprocedure.

Op een aantal competenties bleek sprake te zijn van een hoge mate van overeenkomst tussen de scores die de studenten op de opleiding haalden en de indruk die de praktijkbegeleiders gaven. Dit gegeven kan duiden op een valide beoordelingsprocedure. Echter, op een aantal criteria vertoonden de scores weinig overeenkomst.

INLEIDING

De verwerving van competenties staat steeds meer centraal in het hoger onderwijs. Het competentiebeprij verwijst naar een integratie van kennis, vaardigheden en attitudes (Westera, 2001) die kan worden toegepast in een specifieke arbeidscontext. De focus op competenties in het onderwijs vraagt om beoordelingsvormen waarbij kennis, vaardigheden en attitudes geïntegreerd beoordeeld kunnen worden. Daarnaast dient de beroepssituatie zo centraal mogelijk te staan en moet dat uitgangspunt ook zichtbaar terugkomen in de wijze van beoordeling. Niet elke beoordelingsvorm is geschikt voor het beoordelen van competenties. Van Diggele en Straetmans (2002) noemen als eisen aan beoordelingsvormen die passen bij competentiegericht onderwijs, dat deze ruimte moeten bieden voor persoonlijke invulling en prestaties moeten uitlokken die laten zien dat studenten hun eigen doelen kunnen en willen bereiken. Daarnaast moe-

ten de beoordelingen toelaten dat studenten een eigen vorm kiezen waarin het best zichtbaar wordt wat ze kunnen presteren en dat toetsen verschillende condities (omstandigheden) moeten toestaan en daarmee flexibel zijn. Het portfolio is een instrument dat in toenemende mate gebruikt wordt om competenties aan te tonen en te beoordelen.

Competentiebeoordeling en portfolio

Een portfolio kan worden omschreven als een selectie van illustratief materiaal dat een beeld geeft van de wijze waarop studenten taken hebben uitgevoerd. Een portfolio kan meerdere doelen hebben. Het kan ontwikkelingsgericht worden ingezet om het leerproces te ondersteunen of om verzamelde competentiebewijzen te beoordelen. Met dit laatste verschaft het portfolio inzicht in bereikte competentieniveaus: niveaus van vaardigheden, kennis en attitudes die nodig zijn om relevante taken succesvol te kunnen uitvoeren (Driessen, Beijaard, Van Tartwijk & Van der Vleuten, 2002). Een portfolio wordt in een authentieke context ontwikkeld en de bewijsstukken die erin opgenomen worden zijn levensecht. Het levensechte karakter van een portfolio neemt nog toe wanneer de student zelf kan beslissen wat wordt opgenomen en waarin deze keuze expliciet wordt toegelicht en verantwoord (Dochy, Schelfhout & Janssens, 2003).

Daar waar portfolio's in eerste instantie met name formatief ingezet werden om het praktisch handelen in authentieke contexten vast te leggen en om studenten aan te moedigen om op hun handelen te reflecteren, worden portfolio's nu ook steeds meer voor summatieve doeleinden ingezet, dat wil zeggen dat ze gebruikt worden om een (eind)oordeel uit te spreken over de mate van competentieverwerving door de student. Echter, Pitts, Coles en Thomas (2001) concludeerden dat portfolio's vooral waardevol zijn voor formatieve toetsing en dat het summatief toetsen van portfolio's problematisch kan zijn. Om die reden dient nauw te worden gekeken naar psychometrische kwaliteiten van portfoliobeoordelingen, met name de validiteit en betrouwbaarheid (Driessen e.a., 2005).

De validiteit is de mate waarin de test aan zijn doel beantwoordt. Validiteit is belangrijk omdat ze een voorwaarde vormt voor de rechtvaardiging van de beslissingen over het wel of niet verstrekken van professionele kwalificaties (diploma's en certificaten) (Van Zutven, Polderdijk & De Volder, 2004). Voorwaarde voor een valide toetsing bij een portfoliobeoordeling is dat het bewijsmateriaal dat de student opneemt in het portfolio een valide beeld geeft van zijn/haar kunnen. Eisen die om die reden aan het bewijsmateriaal moeten worden gesteld, zijn: *authenticiteit* (heeft de student dat bewijsmateriaal opgenomen dat geschikt is om een uitspraak te doen over zijn/haar competentieniveau?), *actualiteit* (weerspiegelt het bewijsmateriaal het huidige competentieniveau?) en *relevantie* (in hoeverre dekt het bewijsmateriaal de belangrijkste elementen van de te beoordelen competenties?) (Driessen e.a., 2002).

Melville, Rees, Brookfield en Anderson (2004) stellen vast dat, vanwege de open structuur van portfolio's, de validiteit van portfoliobeoordelingen vaak hoog is. Ook Gonczy (1994) concludeert dat portfoliobeoordelingen dichtbij staan bij hetgeen dient te worden gemeten, namelijk de integratie van de kennis, vaardigheden en attitudes die nodig zijn om competent te functioneren in de praktijk.

Bij summatieve toetsing van competenties middels het portfolio is een belangrijke validiteitsvraag of de toetsing een adequate reflectie is van de competenties die nodig zijn voor het latere functioneren in de beroepspraktijk. Dus als de beoordeling van de competenties door middel van het portfolio adequaat is, dient het een juiste voorspelling te zijn van het later functioneren in de beroepspraktijk. Hierbij ligt de nadruk vooral op predictieve validiteit van de beoordeling (Van Zutven e.a., 2004).

Een adequate beoordeling hangt niet alleen af van de kwaliteit van de bewijslast in het portfolio, maar tevens van de betrouwbaarheid van de beoordelingsprocedure. Dierick, Dochy en Van de Watering (2001) pleiten voor een verruiming van de traditioneel gebruikte kwaliteitscriteria als validiteit en betrouwbaarheid om recht te doen aan de eigenheid van bijvoorbeeld portfoliobeoordeling. Als kwaliteitscriteria van assessment bevelen zij aan: transparantie van de assessmentprocedure, eerlijkheid, cognitieve complexiteit, authenticiteit van taken en directheid van assessment, waarbij globaal beoordelen van het assessment dezelfde resultaten geeft als beoordeling volgens standaarden op deelaspecten.

Bij de betrouwbaarheid van portfoliobeoordeling gaat het vooral om de vraag in welke mate de beslissing 'of iemand wel of niet competent is', betrouwbaar is. Enerzijds gaat het dus om de daadwerkelijke competentie vast te stellen, anderzijds gaat het om te onderzoeken in hoeverre het vertoonde gedrag te generaliseren is naar bijvoorbeeld de beroepswerkelijkheid. Het aspect van de betrouwbaarheid komt met name aan de orde bij:

- het formuleren van de vragen, opdrachten of werkwijzen;
- de objectiviteit of intersubjectiviteit bij de beoordeling;
- het formuleren van de beoordelingscriteria en de beoordelingsmethode (Van Zutven e.a., 2004).

De laatste twee aspecten, intersubjectiviteit en beoordelingsmethode, zijn eisen op procedureel niveau van de beoordeling. Elshout-Mohr, Oostdam, Dietze en Snoek (2000) zien op dit niveau mogelijkheden de betrouwbaarheid bij portfoliobeoordelingen te verhogen: een zorgvuldig geplande en gecontroleerde beoordelingsprocedure die voor iedereen identiek is. Straetmans, Sluijsmans, Bolhuis en Van Merriënboer (2003) sluiten hierop aan door aan te geven dat het bij betrouwbaarheid vooral gaat om een gedetailleerde beoordelingsprocedure, met meerdere, getrainde beoordelaars.

Naast een zorgvuldig geplande beoordelingsprocedure is het ten behoeve van een betrouwbare beoordeling, evenals bij het aspect van validiteit, van belang dat het bewijsmateriaal dat de student in het portfolio opneemt 'te vertrouwen is' als zijnde een adequate afspiegeling van zijn/haar competentieniveau. Vandaar dat eisen aan het bewijsmateriaal toegevoegd kunnen worden als *variatie* (in hoeverre gaat het in het bewijsmateriaal om verschillende contexten en rollen waarin competent gedrag vertoond moet worden?) en *kwantiteit* (toont de student in voldoende mate aan dat het niet gaat om een toevalstreffer, maar om echte beheersing?).

Er bestaat veel discussie over de vraag of er met meer dan wel met minder gedetailleerde kaders moet worden gewerkt bij portfolio-beoordeling. Gedetailleerde kaders verhogen de betrouwbaarheid, maar schieten tekort als het gaat om validiteit. In het algemeen wordt het als niet wenselijk gezien om vergaand gedetailleerde beoordelingskaders te ontwerpen; op dit punt verkiest men validiteit boven betrouwbaarheid (Inspectie van Onderwijs, 2003). Voorgesteld wordt een middenweg tussen analytisch en globaal beoordelen te hanteren (Van Tartwijk & Wubbels, 2002; Driessen e.a., 2002). Men geeft dan één totaalscore, waarbij de beoordeling van een aantal afzonderlijke criteria wordt gebruikt om het eindoordeel te beargumenteren. Op die wijze beoordeelt men de volledige competentie, als integratie van kennis, vaardigheden en attitudes, maar kan er in de argumentatie wel aandacht zijn voor afzonderlijke delen.

In onderzoeken wordt steeds meer bewijs geleverd voor het feit dat globale oordelen ten minste zo betrouwbaar zijn als lijstjes die kunnen worden afgevinkt op specifieke criteria (Regehr, MacRae, Reznick & Szalay, 1998; Schwatz, Colliver, Bardes, Charon, Fried & Moroff, 1999). Zo vonden Regehr e.a. (1998) in een onderzoek naar de psychometrische eigenschappen van respectievelijk analytische 'afvinklijstjes', globale beoordelingsschalen voorafgegaan door een 'afvinklijstje' en globale beoordelingsschalen op zichzelf, dat globale beoordelingsschalen die door ervaren beoordelaars ingevuld waren, een hogere betrouwbaarheid en een betere validiteit vertoonden dan de analytische 'afvinklijstjes'. Bovendien verbeterde de aanwezigheid van de afvinklijsten voorafgaand aan de globale beoordelingsschalen noch de betrouwbaarheid noch de validiteit van de beoordelingsschalen ten opzichte van de globale beoordelingsschalen op zichzelf. De resultaten veronderstellen dat globale beoordelingsschalen, die door ervaren beoordelaars worden ingevuld, minstens zo geschikt zijn om summatief te beoordelen, als analytische lijstjes.

Met dit onderzoek wordt nagegaan of het inderdaad mogelijk is om vanuit de procedurele eisen die gesteld moeten worden aan betrouwbaarheid, een valide beoordelingsprocedure voor hbo-studenten te ontwikkelen om competenties in portfolio's te beoordelen. Volgens Van Zutven e.a. (2004) is een procedure geschikt voor competentiegericht beoordelen, wanneer sprake is van predictieve validiteit: het voorspellend vermogen ten opzichte van toekomstige situaties, studiesucces of succes in de beroepspraktijk.

Om hier meer inzicht in te verkrijgen staat derhalve de volgende onderzoeksvraag centraal:

In welke mate komt het oordeel van de docent over competenties aangetoond middels portfolio overeen met het oordeel van praktijkbegeleiders over competenties aangetoond middels praktisch handelen met behulp van eenzelfde beoordelingsprocedure?

METHODE

Deelnemers

Het onderzoek heeft plaatsgevonden bij de opleiding ergotherapie van de Hogeschool van Arnhem en Nijmegen (HAN).

- Er zijn 28 portfolio's beoordeeld door 14 beoordelaars. De beoordelaars vulden voor elke student het beoordelingsinstrument in, naar aanleiding van de gegevens uit het portfolio van de student.
- Er zijn 27 derdejaarsstudenten beoordeeld door 27 praktijkbegeleiders, die het beoordelingsinstrument invulden naar aanleiding van een indruk die ze hebben van het praktisch handelen van de studenten tijdens de stage. In totaal werd aan 33 praktijkbegeleiders de vraag gesteld om het beoordelingsinstrument in te vullen (respons = 82%).

Uiteindelijk zijn 21 gepaarde waarnemingen (waarnemingen van docenten op de opleiding en praktijkbegeleiders bij dezelfde student) in het onderzoek opgenomen.

Instrumenten

Portfolio

Het portfolio van studenten kan worden getypeerd als een beoordelingsportfolio. Dit portfolio betreft een middel waarmee de student zijn/haar ontwikkeling, prestaties of competenties kan laten beoordelen. Het portfolio van de studenten bestond zowel uit een procesgedeelte, waarin de ontwikkeling ten aanzien van de vijf competenties beschreven was, als uit een presentatiegedeelte, waarin het bewijsmateriaal ten behoeve van de beoordeling opgenomen was. Alleen het presentatiedeel werd beoordeeld.

Beoordelingsinstrument

Het beoordelingsinstrument voor het portfolio bestaat uit drie stappen.

Stap 1: Bepalen of het portfolio kan worden gebruikt om de competentieontwikkeling van de student te beoordelen

De factoren genoemd bij deze stap zeggen iets over de kwaliteit van het materiaal dat is opgenomen in het portfolio, zonder het materiaal inhoudelijk te waarderen. Wanneer deze stap in zijn geheel met een voldoende wordt gewaardeerd, wordt de bewijslast van de student geschikt verondersteld om summatief te beoordelen.

Aspecten	Waardering	
Authenticiteit:		
Is het product daadwerkelijk afkomstig van de stagiair? Is aan het product feedback gekoppeld van een andere persoon dan de student zelf?	+++	---
Actualiteit:		
Weerspiegelt het bewijs het huidige competentieniveau van de stagiair?	+++	---
Relevantie:		
Hoe specifiek is het bewijs voor de te beoordelen competentie?	+++	---
Kwantiteit:		
Is het aantal bewijzen, of de duur van een ervaring, voldoende om uitspraken te doen over de competentie?	+++	---
Variatie:		
Is de competentie in verschillende contexten bewezen? NB Verschillende contexten pas na twee stageperiodes beoordelen	+++	---
Conclusie:		
Advies n.a.v. de hiervoor genoemde aspecten:	+++	---

Beoordelingsmodel voor het waarderen van producten/bewijzen in een portfolio ten behoeve van deelname aan een summatieve competentiebeoordeling (Van Diggele & Straetmans, 2002).

Stap 2: *Beoordeling van het algemene niveau van beheersing van de competenties* → *adviesering*

NB Deze stap moet aan het eind van het jaar met ++ worden beoordeeld om studiepunten te verdienen voor de afzonderlijke competenties.

In deze fase van de opleiding wordt verwacht dat studenten zich kunnen herkennen in de Europese Dublin-descriptoren die passen bij de hbo-kwalificaties. Door het toevoegen van deze stap aan het instrument wordt gewaarborgd dat het handelen van studenten niet alleen wijst in de richting van competentie, maar dat studenten daarnaast ook op het juiste niveau blijf geven van competentieverwerving.

In de Dublin-descriptoren komen criteria terug als professionalisering, transfervermogen, sociaal-communicatieve vaardigheden, methodisch en reflectief handelen en denken, probleemgerichte creativiteit en een brede, multidisciplinaire basis. Criteria die een student in staat stellen om zichzelf te sturen in de aanpak van steeds complexere taken, waarvan niet altijd op voorhand aan te geven is wat standaardprocedures zijn.

Bij de formulering van deze criteria is een keuze gemaakt voor criteria die het meest kenmerkend zijn voor de reflectieve en zelfsturende kwaliteiten die in deze fase van de opleiding verwacht worden.

Dergelijke criteria zijn niet zozeer competentiegebonden, oftewel specifiek verbonden met één of meer beroepscompetenties, maar deze liggen generiek besloten in het handelen van een bekwame hbo-professional. Verwacht mag worden dat deze criteria in de reflecties van studenten terugkomen.

Aspect	Oordeel	Toelichting
Is de stagiair in staat om zelfstandig te handelen m.b.t. de te beoordelen competenties of heeft hij veel tijd of hulp nodig?	++ + +/- - --	
Is de kwaliteit van de reflectie dusdanig dat de stagiair aantoont in staat te zijn verbinding te leggen tussen zijn ervaringen en de theorie (t-tjes en T's). Wekt de stagiair het vertrouwen dat hij zijn inzichten kan generaliseren?	++ + +/- - --	
Geeft de stagiair blijk van een onderbouwde visie en handelt hij volgens die visie?	++ + +/- - --	
Weegt de student op een professionele wijze alternatieven af?	++ + +/- - --	
Herkent de stagiair morele dilemma's in de praktijk en gaat hij hier zorgvuldig mee om?	++ + +/- - --	
Samenvattend oordeel:	++ + +/- - --	

Stap 3: *Formuleren van feedback over de mate waarin de student de verschillende competenties beheerst → advisering*

Van de studenten werd verwacht dat ze in staat zouden zijn om middels de bewijslast die ze opgenomen hadden in het portfolio, vijf competenties aan te tonen: diagnosticeren, interveniëren 1, interveniëren 2, het leveren van een bijdrage aan organisatieprocessen en deskundigheidsbevordering. Bij het beoordelen van deze criteria werd gelet op de kwaliteit van de bewijslast (beroepsproducten) en de feedback van derden.

Competentie	Materiaal genoeg?	Wijst het materiaal in de richting van competentie?	Toelichting oordeel en advies
1. Diagnosticeren	++ +/- --	++ + +/- - --	
<ul style="list-style-type: none"> - vraag verhelderen - gegevens verzamelen door gericht onderzoek - handelingsprobleem definiëren 			
2. Interviëren 1	++ +/- --	++ + +/- - --	
<ul style="list-style-type: none"> - ergotherapeutisch interventieplan opstellen - interventieplan uitvoeren - proces en productevaluatie - dossier voeren en rapporteren - evalueren 			
3. Interviëren 2	++ +/- --	++ + +/- - --	
<ul style="list-style-type: none"> - adviesvraag - probleemanalyse - probleemdefinitie - verkennen oplossingsrichtingen - typeren en selecteren oplossing - advies(rapportage) - implementeren oplossingen - evaluatie 			

Competentie	Materiaal genoeg?	Wijst het materiaal in de richting van competentie?	Toelichting oordeel en advies
4. Bijdrage leveren aan organisatieprocessen	++ +/- --	++ + +/- - --	
<ul style="list-style-type: none"> - contacten leggen en onderhouden - signaleren ontwikkelingen en knelpunten - waarderen van werk- en beleidsprocessen - verbeterplannen - implementatie 			
5. Bevorderen eigen deskundigheid	++ +/- --	++ + +/- - --	
<ul style="list-style-type: none"> - Gaat actief om met feedback - Maakt S/Z-analyse - Geeft sturing aan eigen leerproces (stageplan) - Kan zelfstandig theorie gebruiken - Gebruikt inzichten uit gevolgde scholingen 			

De criteria die in het beoordelingsinstrument zijn opgenomen, zijn dusdanig geformuleerd dat een middenweg tussen analytisch en globaal oordelen mogelijk is. Dat wil zeggen: de criteria zijn niet heel specifiek uitgewerkt naar een lijstje van kenmerken die afgevinkt kunnen worden; er worden wel kenmerken genoemd, maar de kenmerken zijn slechts een hulpmiddel om tot een totaal (globaal) oordeel te komen.

Procedure

Studenten werden op de opleiding in januari 2005 beoordeeld door beoordelaars die aan de hand van het beoordelingsinstrument de portfolio-inhoud van de student op waarde schatten.

Ten behoeve van de betrouwbaarheid is in de ontwikkelde procedure een aantal maatregelen genomen, zoals het trainen van de beoordelaars, het beoordelen door meerdere beoordelaars, het ontwerpen van (objectieve) beoordelingscriteria die voor elke student gelden, het zorgen voor tussentijdse terugkoppeling op het portfoliomateriaal in de vorm van studieloopbaanbegeleiding, het scheiden van begeleiding en beoordeling, het zoeken van een middenweg tussen een globale en analytische beoordeling, het verhogen van de authenticiteit door studenten te vragen om al het bewijsmateriaal van feedback door derden te laten voorzien, en het bieden van een tweede kans aan studenten om door middel van een gesprek hun reflectieniveau alsnog aan te tonen. Over alle competenties afzonderlijk wordt een oordeel uitgesproken, aangezien studiepunten per competentie worden toegekend.

Praktijkbegeleiders van studenten die op de opleiding in januari op basis van het portfolio reeds beoordeeld waren, zijn in april 2005 telefonisch benaderd met de vraag of ze voor de studenten, die ze op dat moment begeleidden, eenzelfde instrument wilden invullen als eerder op de opleiding gedaan was, op basis van een indruk van het competentieniveau van de studenten.

Aan de praktijkbegeleiders die respons gaven aan dit verzoek (82%) is in mei 2005 hetzelfde instrument voorgelegd als het beschreven instrument dat is ingevuld door de beoordelaars op de opleiding. Enige verschil is dat de praktijkbegeleiders stap 1 niet

hoefden te doorlopen, omdat deze stap slechts betrekking heeft op de inhoud van het portfolio (kwaliteit van de bewijslast).

Stap 2 en 3 hebben de begeleiders helemaal ingevuld; het verschil zat in de instructie die voorafgaand aan de stappen gegeven werd, aangezien het functioneren in de praktijk gedefinieerd werd als de indruk die de student maakt op de praktijkbegeleider(s) als startbekwame ergotherapeut. Daar waar voor de beoordelaars op de opleiding de volgende instructie genoemd werd: 'Spreek ten aanzien van onderstaande aspecten een oordeel uit of uit het portfolio blijkt dat de student in staat is dit te doen en formuleer eventueel advies', was de instructie voor de praktijkbegeleiders als volgt: 'Spreek ten aanzien van onderstaande aspecten een oordeel uit of uit het praktisch handelen blijkt dat de stagiair in staat is dit te doen en formuleer eventueel advies'. In tegenstelling tot de beoordelaars op de opleiding waren de praktijkbegeleiders niet getraind en was in dit geval geen sprake van meerdere beoordelaars, maar praktijkbegeleiders zien studenten over een langere periode en ze hanteerden dezelfde criteria als de beoordelaars op de opleiding.

Analyse

Allereerst zijn de gemiddelde scores van de beoordelaars op de opleiding en van de praktijkbegeleiders berekend. Vervolgens zijn correlaties berekend tussen de scores van de beoordelaars over de inhoud van het portfolio en de scores van de praktijkbegeleiders over het praktisch handelen.

Zowel per competentie criterium werd een correlatie (Pearsons ρ) berekend, als voor de criteria die verwijzen naar het algemene niveau van beheersing van de competenties. Daarnaast is middels t -toetsing gekeken of de gemiddelde scores tussen beoordelaars op de opleiding en praktijkbegeleiders significant verschillen. Om het gestandaardiseerde verschil van de gemiddelden te bepalen is de effectgrootte (Cohens d) berekend, waarbij een effectgrootte van <0.20 geïnterpreteerd kan worden als een klein of minimaal effect, $0.20-0.80$ als een matig effect en >0.80 als een groot effect.

RESULTATEN

In tabel 1 zijn de gemiddelde scores te zien van de beoordelaars op de opleiding en de praktijkbegeleiders op de competenties. Bovendien worden de correlaties en het verschil tussen de gemiddelde scores per competentie zichtbaar.

Op sommige criteria komt het oordeel van de docenten op de opleiding en de praktijkbegeleiders in hoge mate overeen: significante correlaties op een significantieniveau van $p=.01$ zijn te zien met betrekking tot de competenties 'diagnosticeren' ($\rho=.80$), 'interveniëren 1' ($\rho=.63$), 'deskundigheidsbevordering' ($\rho=.65$) en het algemene beheersingsaspect 'zelfstandigheid' ($\rho=.66$). Wat betreft het 'samenvattend oordeel met betrekking tot stap 1' ($p=.46$) is een significante correlatie gevonden op het niveau van $p=.05$.

Bij de overige criteria is de mate van overeenkomst tussen docenten en praktijkbegeleiders laag. De correlaties variëren van $\rho=.08$ ('bijdrage leveren aan organisatieprocessen'), tot $\rho=.38$ ('afwegen alternatieven').

De gemiddelden van een drietal criteria verschillen significant van elkaar. Praktijkbegeleiders beoordelen significant positiever dan docenten op de opleiding bij de criteria 'afwegen van alternatieven' ($\rho=.38$, $t=2.65$, $p<.01$), 'herkennen van morele dilemma's' ($\rho=.35$, $t=2.99$, $p<.00$) en het samenvattend oordeel dat gegeven wordt ($\rho=.46$, $t=2.65$, $p<.00$). De effectgrootte is bij deze criteria groter dan .58, wat een matig effect betekent.

De tendens die uit deze drie criteria is af te leiden, is dat praktijkbegeleiders over het algemeen positiever zijn ten aanzien van de studenten dan uit de portfoliobeoordeling op de opleiding naar voren komt.

Opmerkelijk is dat bij de criteria 'interveniëren 2' en 'bijdrage leveren aan organisatieprocessen' geen verschil te zien is in gemiddelde score die door docenten op de opleiding en praktijkbegeleiders gegeven wordt. En dat, terwijl de twee scores (van docenten en van de praktijkbegeleiders) op het laatste criterium, 'bijdrage leveren aan organisatieprocessen', nauwelijks correlatie met elkaar vertoonden ($\rho=.08$). Individuele scores verschillen blijkbaar behoorlijk.

Tabel 1 Resultaten van aantal paren (n), correlatie (R), gemiddelde en standaarddeviatie van praktijkbegeleiders (X_1 en S_1), gemiddelde en standaarddeviatie van docenten op de opleiding (X_2 en S_2), verschil tussen twee gemiddelden ($M_1 - M_2$), T-toets voor verschil tussen gemiddelden (t), aantal vrijheidsgraden (df), de overschrijdingskans (Sig 2-tailed) en effectgrootte (d)

Criterion	n Paren	R	X_1	S_1	X_2	S_2	$M_1 - M_2$	t	df	Sig (2-tailed)	d
Diagnosticeren	19	.80**	4.4	.61	4.6	.56	.16	1.83	18	.08	.43
Interveniëren 1	19	.63**	4.2	.80	4.3	.58	-.16	-1.14	18	.26	-.27
Interveniëren 2	13	1.00	4.2	.56	4.2	.56	.00		12		
Bijdrage leveren aan organisatieprocessen	19	.08	4.3	.82	4.3	1.0	.00	.00	18	1.00	
Bevorderen eigen deskundigheid	19	.65**	4.3	1.0	4.5	.70	-.16	-.90	18	.38	-.22
Zelfstandigheid	21	.66**	4.4	.87	4.6	.51	-.14	-1.00	20	.33	.43
Kwaliteit reflectie	21	.20	4.5	1.1	3.8	1.4	.67	1.96	20	.06	.44
Visie	21	.29	4.5	.87	3.8	1.4	.62	2.03	20	.06	.44
Afwegen alternatieven	21	.38	4.4	1.1	3.6	1.4	.81	2.65	20	.01	.58
Herkennen morele dilemma's	21	.35	4.6	.74	3.8	1.4	.86	2.99	20	.00	.66
Samenvattend oordeel m.b.t. stap 2	21	.46*	4.3	.86	3.7	1.1	.62	2.65	20	.01	.58

* $p<.05$.

** $p<.01$.

CONCLUSIE EN DISCUSSIE

Doel van dit onderzoek was om na te gaan of het mogelijk was een beoordelingsprocedure te ontwikkelen, die het beoordelen van portfolio's mogelijk maakt op zodanige wijze dat die predictief valide is en een betrouwbaar beeld geeft van de mate van competentieverwerving door de student. Dit werd gemeten door na te gaan in welke mate studenten die op de opleiding een bepaalde beoordeling krijgen, ook in de praktijk overeenkomstig hun beoordeling presteren. Concreet werd de vraag gesteld: *In welke mate komt het oordeel van de docent over competenties aangetoond middels portfolio overeen met het oordeel van praktijkbegeleiders over competenties aangetoond middels praktisch handelen?*

Het oordeel van praktijkbegeleiders werd gebruikt als referentiepunt ter controle van de betrouwbaarheid en predictieve validiteit van de procedure op de opleiding.

Aan de hand van dit onderzoek kan nog niet worden gesteld dat op alle criteria die zijn beoordeeld, voldoende overeenkomst is tussen beoordeling op de opleiding en prestatie in de praktijk. Toch is over vijf van de elf criteria een hoge mate van overeenkomst gevonden tussen het oordeel van docenten en van praktijkbegeleiders. Van die vijf criteria worden drie criteria gemeten in stap 3 van de procedure. Deze stap verwijst rechtstreeks naar de competenties die van studenten worden verwacht. Dat geeft aan dat over deze competenties een valide oordeel uitgesproken wordt middels beoordeling van het portfolio. De portfoliobeoordeling staat hier dicht bij hetgeen gemeten dient te worden, namelijk de integratie van de kennis, vaardigheden en attitudes die nodig zijn om competent te functioneren in de praktijk.

Gezien de tendens dat praktijkbegeleiders positiever oordelen over de studenten kan het zijn dat de ontwikkelde beoordelingsprocedure bij portfoliobeoordeling te streng selecteert. Temeer, aangezien de mate van overeenstemming tussen docenten op de opleiding en praktijkbegeleiders bij de criteria in stap 2 (algemene aspecten van competentiebeheersing) vaak laag was. Deze stap wordt in de procedure als voorwaardelijk beschouwd om tot een verdere, inhoudelijke beoordeling van de competenties over te gaan. Bekeken moet worden of deze algemene aspecten een geschikte afspiegeling vormen van de Dublin-descriptoren, uitgewerkt naar hbo-criteria in deze fase van de opleiding. Wellicht dat het aspiratieniveau dat met deze aspecten wordt beoogd, te hoog is voor derdejaarsstudenten.

Ondanks dat de resultaten van dit onderzoek niet zonder meer aanleiding geven om te kunnen spreken van een geheel valide en betrouwbare beoordelingsprocedure, kunnen verklaringen worden gevonden voor het feit dat het oordeel van de docenten en de praktijkbegeleiders op een zestal criteria van elkaar verschilt.

Eén verklaring is dat de praktijkbegeleiders niet bekend waren met de criteria op het moment van invullen. Praktijkbegeleiders gaven regelmatig aan dat ze het lastig vonden om aan de hand van zo weinig criteria, die naar hun idee verschillende 'subcriteria' kenden die ze verschillend zouden beoordelen, een totaal oordeel te geven over de student. Praktijkbegeleiders wilden graag een sterker analytisch instrument om hun indruk

van de student te geven. Een globaal oordeel vonden ze, op basis van het instrument, lastig te geven, wat ook niet vreemd is, gezien de conclusie van Regehr e.a. (1998) dat globale beoordelingsschalen betrouwbaar zijn, mits ingevuld door ervaren beoordelaars.

Een andere verklaring voor het verschil in oordeel tussen docenten op de opleiding en praktijkbegeleiders is dat praktijkbegeleiders vanuit een begeleidingsrelatie met de student gevraagd werd een oordeel te geven, terwijl op de opleiding begeleiding en beoordeling gescheiden zijn, waardoor een objectief oordeel eenvoudiger te geven is. Praktijkbegeleiders maken ontwikkelingen van studenten mee en hebben een relatie met de studenten, hetgeen van invloed kan zijn op het oordeel.

Een derde verklaring voor het verschil is dat docenten op de opleiding en praktijkbegeleiders andere eisen hanteren ten aanzien van de criteria. Docenten op de opleiding lijken hogere eisen te stellen aan de algemene aspecten van competentiebeheersing, zoals bijvoorbeeld 'de kwaliteit van reflectie', 'het herkennen van morele dilemma's'. De docenten die het portfolio beoordelen, beoordelen per criterium, aan de hand van schriftelijk bewijs. Het niveau van schriftelijk formuleren, redeneren en onderbouwen wordt meegenomen. Een veel gegeven toelichting door docenten op de portfolio's van de studenten is dat veel zaken impliciet blijken maar nog te weinig geëxpliciteerd worden in het portfolio. Bovendien wordt nog vaak de integratie tussen handelen en theorie gemist. Aan dat laatste criterium zal de opleiding waarschijnlijk meer eisen stellen dan de praktijk.

De competentie bij welke de overeenkomst tussen docenten op de opleiding en praktijkbegeleiders opvallend laag is, is 'bijdrage leveren aan organisatieprocessen'. Als wordt gekeken naar de indicatoren die horen bij dit criterium, zoals bijvoorbeeld het leggen en onderhouden van contacten of het waarderen van werk- en beleidsprocessen, lijkt het aannemelijk dat deze competentie vrij lastig in een portfolio aan te tonen is. Concreet bewijsmateriaal dat deze indicatoren kan ondersteunen is niet zomaar voor handen. Het zijn indicatoren die met name incidenteel in de praktijk voorkomen en die lastig zijn vast te leggen in een portfolio. Aanwijzingen voor dit idee zitten in de toelichting die docenten, respectievelijk praktijkbegeleiders geven: opvallend is dat docenten met name bij deze competentie vaak aangeven bewijsmateriaal te missen. Bovendien, als ze praten over de kwaliteit van de competentie, dan praten ze in de meeste gevallen over de kwaliteit van rapporten, verslagen, projecten: kortom, concrete producten (bewijsmateriaal). Ook wijzen ze op de relatie van het bewijsmateriaal met de literatuur. Daar waar studenten dilemma's uit de praktijk, plannen en discussies beschrijven, wordt dit in het bijzonder door docenten op de opleiding gewaardeerd.

Praktijkbegeleiders hebben het bij de competentie 'bijdrage leveren aan organisatieprocessen' vooral over luistervaardigheden, contacten in en om het werkveld en persoonskenmerken zoals flexibiliteit, sociale invoelendheid, lef en accuratie. Het is zinvol om te overwegen of het meten van een dergelijke competentie wel in de portfoliobeoordelingsprocedure moet worden opgenomen of wellicht anders kan worden gedefinieerd, zodat deze beter aansluit bij concreet bewijsmateriaal en daarmee betrouwbaar kan worden beoordeeld.

Duidelijk wordt dat het portfolio niet automatisch een reflectie is van het praktisch handelen van de student. In het portfolio gaat het om concrete gegevens of concrete beschrijvingen. Het aanleggen van een portfolio is een vaardigheid op zichzelf, waar de opleiding aandacht voor moet hebben zodat uiteindelijk een meer valide beoordeling van wel of niet competent functioneren mogelijk is. Bovendien moet men alert zijn op het feit dat sommige competenties en bijbehorende indicatoren wellicht niet gemakkelijk middels portfolio kunnen worden aangetoond. Daarnaast is een student die in staat is een goed geschreven weerslag te geven van het praktisch handelen, nu eenmaal in het voordeel als portfolio het instrument is aan de hand waarvan wordt getoetst, ondanks de hiervoor genoemde oplossingen, die weliswaar aannemelijk lijken maar in de praktijk, uit het oogpunt van arbeidsintensiteit, niet echt wenselijk zijn. De vraag is of opleidingen zich dit altijd realiseren.

Het is aan te bevelen om naast portfolio-beoordeling ook andere competentiegerichte toetsvormen te blijven hanteren, zoals performance-assessments en interviews. Daarnaast moet op opleidingen aandacht zijn voor het bewijzen met portfolio als instrument.

Hoewel aan de hand van dit onderzoek niet geconcludeerd kan worden dat met de ontwikkelde beoordelingsprocedure een hoge mate van overeenkomst gevonden werd tussen beoordeling op de opleiding en prestatie in de praktijk, kende dit onderzoek beperkingen, waardoor conclusies lastig generaliseerbaar zijn.

De procedure was slechts voor een heel beperkte groep ontwikkeld: derdejaars hbo-studenten ergotherapie, waardoor het aantal deelnemers aan het onderzoek laag was.

Daarnaast is het oordeel van praktijkbegeleiders als referentiepunt genomen, een oordeel dat om genoemde redenen zijn beperkingen kent en beïnvloed kan worden door aspecten als relatie met de student en persoonlijkheid van de begeleider.

Bovendien is in dit onderzoek de predictieve validiteit gemeten, aangezien deze vorm van validiteit volgens Van Zutven e.a. (2004) het meest aan de orde is bij het meten van professionele (beroeps)competenties. Onderzoek naar de inhoudsvaliditeit, de kwaliteit van de definiëring van de beoogde competenties in een aantal indicatoren, zou een toegevoegde waarde zijn.

Resultaten van dit onderzoek riepen bovendien de vraag op of het portfolio daadwerkelijk een wenselijk middel is om blijk te geven van competentie. Om die reden zou onderzoek naar de consequentiële validiteit van portfolio-beoordeling waardevol zijn. Op dat moment kan het onderzoek in een breder perspectief worden geplaatst in termen van gevolgen van de interpretatie over assessment en de effecten van een dergelijke portfolio-beoordelingsprocedure op het leergedrag van studenten (Dierick e.a., 2001).

Bij de invoering van competentiegericht onderwijs is het essentieel dat de student wordt meegenomen in het denken over het onderwijs. Ondanks dat het gebruik van bepaalde toetsvormen allerlei consequenties heeft voor het leren van de student, is de opvatting van de student over wat goed toetsen is niet automatisch anders bij de invoer van een competentiegericht beoordelingsinstrument. In plaats daarvan stuit de student op onbegrip. Het is duidelijk dat de student niet zomaar meegaat in de nieuwe gedach-

tegang; dit vergt actieve betrokkenheid, goede informatie en inbedding van het portfolio als rode draad door alle geleidingen van het onderwijs. Oude constructen moeten worden losgelaten, voor welke nieuwe in de plaats komen.

Illustratief voor het hiervoor genoemde discussiepunt is een citaat van Snadden (1999): 'Until we can make a mental shift that allows us to include a more holistic approach to assessment, one which values the development of individuals over a period of time, we will continue to struggle to measure the unmeasurable, and may end up measuring the irrelevant, because it is easier': een mentale omslag in het denken over onderwijs is nodig, een omslag die een meer holistische benadering richting assessment mogelijk maakt, die zich richt op de ontwikkeling van studenten in een bepaald tijdsbestek. Vermeden moet worden dat men uit 'veiligheidsoverwegingen' toch weer kiest voor een sterk gestructureerd portfolio waarin de eindtermen richting geven aan een zelfevaluatie die wordt onderbouwd met materiaal dat in de bijlagen is opgenomen. Deze nadruk op betrouwbaarheid past binnen traditionele onderwijsmodellen, maar is niet geschikt om het onderwijs echt competentiegericht te maken.

De omslag naar competentiegericht denken en daarbinnen het portfoliogebruik is een jarenlang proces waaraan alle eenheden die samen onderwijs maken, onderhevig zijn. Dat het proces een keer gaat landen, is iets wat zeker is, maar hoe, dat is afhankelijk van de aansluiting die de afzonderlijke eenheden met elkaar vinden. En toetsing, dat is daar maar één van.

REFERENTIES

- Dierick, S., Dochy, F. & Watering, G. van de (2001) Assessment in het hoger onderwijs: over de implicaties van nieuwe toetsvormen voor de edumetrie. *Tijdschrift voor Hoger Onderwijs*, 1, 2-18.
- Diggele, J. van & Straetmans, G. (2002) Competentiegericht beoordelen: een samenvatting van de brochure 'anders opleiden, anders toetsen' (2001), deel 1 in de BVE-brochurereeks 'Perspectief op Assessment'. Citogroep.
- Dochy, F., Schelfhout, W. & Janssens, S. (2003) *Anders evalueren: Assessment in de onderwijspraktijk*. Centrum voor Opleidingsdidactiek K.U.Leuven & Tiel: Uitgeverij Lannoo.
- Driessen, E., Beijaard, D., Tartwijk, J. van & Vleuten, C. van der (2002) *Portfolio's*. Groningen: Wolters-Noordhoff.
- Driessen, E., Vleuten, C. van der, Schuwirth, L., Tartwijk, J. van & Vermunt, J. (2005) The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Medical Education*, 39, 214-220.
- Elshout-Mohr, M., Oostdam, R., Dietze, A. & Snoek, M. (2001) Assessment van competenties. *VELON, Tijdschrift voor Lerarenopleiders*, 22 (2), 48-54.
- Gonczi, A. (1994) Competence based assessment in the professions in Australia. *Assessment in Education*, 1 (1), 27-44.
- Inspectie van Onderwijs (2003) *Het gebruik van het portfolio als nieuwe toetsvorm in het hoger onderwijs: eindverslag portfolio-onderzoek*. Project Portfoliotoetsing in het hoger onderwijs.

- Melville, C., Rees, M., Brookfield, D. & Anderson, J. (2004) Portfolios for assessment of paediatric registrars. *Medical Education*, 38, 1117-1125.
- Pitts, J., Coles, C. & Thomas, P. (2001) Enhancing reliability in portfolio assessment: 'shaping' the portfolio. *Medical Teacher*, 23 (4), 351-355.
- Regehr, G., MacRae, H., Reznick, R. & Szalay, D. (1998) Comparing the psychometric properties of checklists and global rating scales for assessing performance on an OSCE-format examination. *Academic Medicine*, 73, 993-997.
- Snadden, D. (1999) Portfolios: attempting to measure the unmeasurable. *Medical Education*, 33, 478-479.
- Straetmans, G., Sluijsmans, D., Bolhuis, B. & Merriënboer, J. van (2003) Integratie van instructie en assessment in competentiegericht onderwijs. *Tijdschrift voor Hoger Onderwijs*, 3.
- Schwartz, M.H., Colliver, J.A., Bardes, C.L., Charon, R., Fried E.D. & Moroff, S. (1999) Global ratings of videotaped performance versus global ratings of actions recorded on checklists: a criterion for performance assessment with standardized patients. *Academic Medicine*, 74, 1028-32.
- Tartwijk, J. van & Wubbels, Th. (2002) Evalueren van leervorderingen met portfolio's. In: *Evalueren in het onderwijs*, Onderwijskundig lexicon, Editie III. Alphen aan de Rijn: Samson.
- Westera, W. (2001) Competences in Education: a confusion of tongues. *Journal of Curriculum Studies*, 33 (1), 75-88.
- Zutven, G. van, Polderdijk, M. & Volder, M. de (2004) *Handboek toetsplanontwikkeling in competentiegericht onderwijs: beleid voor verantwoord plannen van toetsing en examinering in het hoger onderwijs*. Stichting Digitale universiteit.