

Ontwikkeling van een zelfbeoordelingsinstrument voor docentcompetenties

E. C. Roelofs, M. Nijveldt en D. Beijaard¹

Samenvatting

Deze studie richt zich op de ontwikkeling en beproeving van een webgebaseerd instrument voor zelfbeoordeling ten behoeve van docenten in het hoger beroepsonderwijs. Het instrument is gebaseerd op wetenschappelijke inzichten over relevante taakgebieden van deze doelgroep. Het instrument bestrijkt 18 onderscheiden competenties, verdeeld over vijf taakgebieden. Gebruikers krijgen na afname ervan feedback in de vorm van scoreprofielen en een schriftelijke toelichting. In twee rondes zijn proefversies van het instrument voorgelegd aan hbo-docenten, deels schriftelijk en deels webgebaseerd. De resultaten van een pilot en een vervolgstudie wijzen uit dat de competenties betrouwbaar kunnen worden vastgesteld met behulp van 18 schalen (totaal 182 items). Voor twee competenties zijn bovendien meer verfijnde schalen ontwikkeld, die differentiëren naar deelaspecten van deze competenties en naar specifieke beroepssituaties. Ook deze schalen blijken betrouwbaar. De psychometrische meerwaarde van deze differentiatie verschilt voor de twee competenties. Geïnterviewde hbo-docenten laten weten dat het achterliggende domein van het instrument herkenbaar en representatief is. Bovendien zet de zelfbeoordeling hen aan tot reflectie op het eigen beroep, waarmee een proces van verdere professionalisering wordt ondersteund.

1 Inleiding

In diverse onderwijssectoren is sprake van een toenemende belangstelling voor de beoordeling van docentcompetenties. Naast de toegenomen aandacht voor kwaliteitsbewaking in het algemeen is dit te verklaren door het streven naar erkenning van het leraarsberoep als professie (Verloop, 2003). Centraal in deze studie staan het ontwikkelen en beoordelen van competenties van docenten in het hoger beroepsonderwijs (hbo). Gerappor-

teerd wordt over de ontwikkeling van een webgebaseerd zelfbeoordelingsinstrument voor hbo-docenten, bedoeld ter ondersteuning van hun professionele ontwikkeling.

Instrumenten om competenties van docenten gedurende hun loopbaan te volgen zijn in het hbo nauwelijks voorhanden. De ontwikkeling van beoordelingsinstrumenten past echter bij de beleidsontwikkelingen in het hbo. Via een stelsel van kwaliteitszorg worden opleidingen periodiek gevisiteerd en beoordeeld op een aantal kwaliteitskenmerken, zoals de kwaliteit van de onderwijsinhoud, de vormgeving van het onderwijsleerproces, de behaalde rendementen en de realisatie van organisatorische randvoorwaarden (Nederlands-Vlaamse Accreditatieorganisatie, 2003). Bij de beoordeling van kwaliteit van hbo-opleidingen wordt een groot belang gehecht aan de bekwaamheden of competenties van docenten, zoals ook de Onderwijsraad (2006; 2007) in recente rapporten aangeeft.

Deze studie beoogt theoretische en praktische kennis te vergaren over aspecten van validiteit, betrouwbaarheid en praktische bruikbaarheid van zelfbeoordelingsinstrumenten voor competentiebeoordeling. Dit alles in de context van het gebruik ervan voor de professionele ontwikkeling van hbo-docenten. In het vervolg van dit artikel gaan we in op de problematiek van beoordelen van docentcompetenties, het gebruik van zelfbeoordeling als instrument, inclusief het in deze studie ontwikkelde instrument (Eduprof (Roelofs & Beijaard, 2004), en de opzet en resultaten van het onderzoek naar dit instrument. Het artikel besluit met een aantal conclusies en aanbevelingen.

2 Het beoordelen van docentcompetentie

2.1 Doelen van competentiebeoordelingen

De beoordeling van docentcompetentie dient in het algemeen één of meer van de volgende

doelen:(1) het bijhouden en bevorderen van professionele groei van docenten, (2) het monitoren van de implementatie van onderwijsvernieuwing en (3) het beoordelen van de kwaliteit van docenten met het oog op certificering, werving, selectie of promotie (Boyd,1989; Sternberg & Horvath, 1995; Verloop, 2003).

Bij het bevorderen van professionele groei en het monitoren van onderwijsvernieuwingen is veelal sprake van formatieve beoordeling, terwijl bij het beoordelen van de kwaliteit van docenten meestal sprake is van summatieve beoordeling. Verschillende auteurs pleiten bij het toepassen van docentbeoordelingen in schoolorganisaties voor een systeem waarbij op continue basis sprake is van formatieve beoordeling en periodiek van summatieve beoordeling (Danielson & McGreal, 2000; Howard & McColskey, 2001). Beide vormen moeten in de tijd dus wel van elkaar gescheiden blijven. Formatieve beoordeling moet volgens genoemde auteurs docenten de vrijheid bieden om bepaalde onderdelen van hun onderwijsrepertoire te verbeteren, nadat eerst is vastgesteld welke onderdelen goed of minder goed worden beheerst. Bij summatieve beoordeling wordt een totaalbeeld verkregen op een breed scala aan competenties. Dan staat niet meer het leerproces van docenten centraal, maar de mate waarin ze beschikken over vooraf vastgestelde competenties.

2.2 Uitgangspunten bij competentiebeoordeling

Geleidelijk aan ontstaat een kennisbasis over de wijze waarop docentcompetenties beoordeeld zouden moeten worden. Samenattend wordt bij onderzoek en instrumentontwikkeling gewerkt vanuit de volgende uitgangspunten (Van der Schaaf, 2005; Tigelaar, 2005; Uhlenbeck, 2002): beoordelingen zijn integratief van aard, houden rekening met de specifieke context van de docent, komen tot stand op basis van het gebruik van een mix van methoden en dienen bij te dragen aan de verdere professionalisering van docenten (zie ook Dwyer, 1998; Frederiksen, Sipusic, Sherin, & Wolfe, 1998). Hieronder worden deze uitgangspunten nader toegelicht.

Het competentiebeprip als basis voor integratieve beoordeling van cognities, gedrag en gevolgen

Bij het beoordelen van docentcompetenties wordt recentelijk steeds minder nadruk gelegd op het 'juiste' professionele handelen en meer op het formuleren van principes en criteria voor verantwoord professioneel handelen (Dwyer, 1994, 1998). In dit verband benadrukt Dwyer dat ook onderliggende beslisprocessen in beeld gebracht moeten worden.

Meer in het algemeen kwam met de intrede van het competentiebeprip de nadruk bij beoordeling minder te liggen op geïsoleerde kenniselementen en gedragingen van docenten en meer op integratie van kennis, opvattingen en vaardigheden. Competenties worden beschouwd als het resultaat van een complexe mix van kennis, opvattingen en vaardigheden die situatiespecifiek worden omgezet in professioneel handelen (Reynolds, 1992; Shulman, 1987). Competent handelen betekent niet alleen het vertonen van door deskundigen als bekwaam of effectief erkend gedrag, maar ook het kunnen afwegen van alternatieven en verantwoord van het vertoende gedrag in uiteenlopende, specifieke en onvoorspelbare, contexten (Dwyer, 1994). Bij de beoordeling dienen naast gedragscomponenten dan ook cognities te worden betrokken (Uhlenbeck, 2002).

De toenemende integratie van theorieën over onderwijzen en theorieën over leren maakt het bovendien voor de hand liggend om bij de beoordeling het handelen en de verantwoording daarvan zinvol in verband te brengen met de gevolgen voor leerlingen (Shuell, 1993; Vermunt & Verloop, 1999), bijvoorbeeld als het gaat om de verwerving van begrippen.

Standaarden voor docentcompetenties kunnen dan ook gedefinieerd worden vanuit de gevolgen. Gestart wordt met gewenste op te roepen leerprocessen bij leerlingen, waarna wordt teruggeredeneerd naar acceptabele wijzen van handelen van leerkrachten en onderliggende beslisprocessen, die het handelen legitimeren aan de hand van verdedigbare persoonlijke kennisbestanden en overwegingen die verband houden met de specifieke onderwijssituatie. Frederiksen en collega's.

(1998) hebben deze werkwijze al eens toegepast in een studie over het beoordelen van videoportfolio's.

Om gevolgen, gedrag en onderliggende beslisprocessen in samenhang te kunnen beoordelen, is het gebruik van een interpretatief procesmodel voor competent handelen waarin al deze aspecten een plaats hebben van belang (Kane, 1992). Niet de afzonderlijke componenten van competent handelen, zoals het waargenomen gedrag, de afwegingen en de gevolgen, maar de combinatie ervan wordt beoordeeld. Ten behoeve van het beoordelen van docentcompetenties hebben Roelofs en Sanders (2003; 2007) de samenhang tussen deze componenten samengevat in een procesgericht model voor competent handelen, waarbij ook de contextspecificiteit en de complexiteit van competent handelen tot uitdrukking komen. Competentie kan volgens dit model afgelezen worden als de mate waarin personen, gebruikmakend van een persoonlijke kennisbasis, bij het uitvoeren van taken in een specifieke context verantwoorde beslissingen kunnen nemen, resulterend in werkgedrag (probleemaanpak, interventie) dat bijdraagt aan vooraf als wenselijk beschouwde resultaten. Deze resultaten kunnen individueel worden afgelezen bij cliënten (in geval van docenten bij leerlingen) of collega's, maar ook op het niveau van afdelingen (bijv. slaagcijfers). In Figuur 1 is het model weergegeven. Uit dit model kunnen diverse operationele definities van competentie worden afgeleid, zoals ook ge-

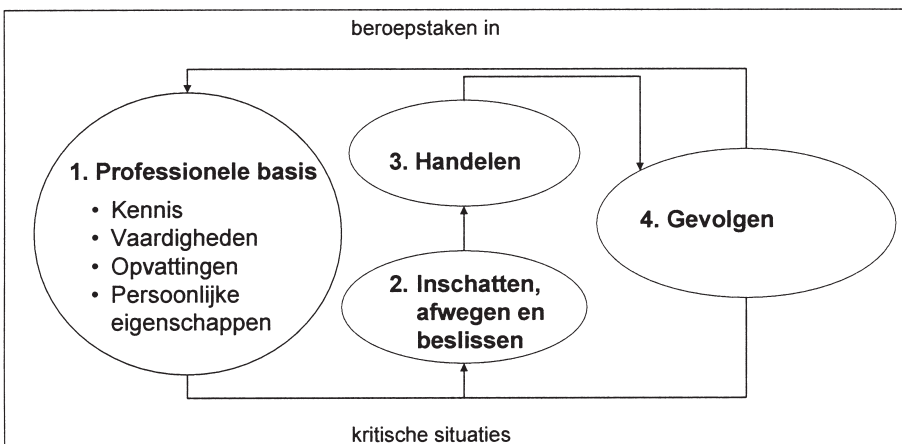
beurd is bij de ontwikkeling van het instrument Eduprof.

Situatiespecifiek beoordelen

Aansluitend bij de opvatting dat competenties worden getoond in specifieke contexten en ook binnen zekere grenzen contextspecifiek van aard zijn, wint het idee van situatiespecifiek beoordelen thans terrein. Tot het begin van de jaren negentig werd bij de beoordeling van docenten veel gebruik gemaakt van methodes die weinig rekening hielden met de specifieke beroepssituatie van de docent. Gegevens uit lesobservaties, checklists en beoordelingsschalen werden niet verzameld in relatie tot een specifiek vak, het type leerling waaraan werd onderwezen, en de daarbij centraal staande doelen (Stodolsky, 1990). Hedendaagse auteurs pleiten voor beoordeling in specifieke contexten (Danielson & McGreal, 2000; Darling-Hammond, & Snyder, 2000; Scriven, 1995). Door bij de beoordeling rekening te houden met de specifieke beroepscontext van de leraar worden beoordelingen specifiek en daarmee meer valide.

Gebruik van een methodenmix

De tendens van situatiespecificiteit sluit aan bij het streven om authentieke vormen van toetsing te ontwikkelen (Wiggins, 1989). Daarbij wordt ernaar gestreefd de beoordelingssituaties zoveel mogelijk te laten lijken op de werkelijkheid. Deze zogenaamde *hands-on*-methoden van beoordelen bieden



Figuur 1. Procesmodel voor professioneel handelen in een specifieke taakomgeving.

de docent de gelegenheid om bewijs voor competentie aan te dragen door het uitvoeren van authentieke beroepstaken in complexe beroepssituaties, zoals het verzorgen van een proefles en het samenstellen van een videoportfolio op basis van meerdere lessen.

Hands-on-methoden hebben als nadelen dat ze duur zijn, moeilijk te standaardiseren en maar een beperkt deel van het in de praktijk verlangde onderwijsrepertoire bestrijken (Straetmans & Sanders, 2001). Anders gezegd, de methoden hebben een beperkte generaliseerbaarheid naar het universum van onderwijssituaties (Kane, 1992, 2004). Om deze reden kan gekozen worden voor de inzet van minder authentieke, zogenaamde *hands-off*-instrumenten, die weliswaar minder dicht bij de praktijk van de docent staan, maar wel een groter gebied bestrijken, bijvoorbeeld door de beslisvaardigheid in een groot aantal denkbare onderwijssituaties te beoordelen. Meer in het algemeen wordt een mix van methoden aanbevolen, zodat de zwakten van de afzonderlijke methoden kunnen worden gecompenseerd (Dwyer, 1994, 1998; Uhlenbeck, 2002).

Bijdrage aan professionalisering

Wil beoordeling een bijdrage leveren aan professionalisering van docenten dan zijn systematische feedback op de beroepsuitoefening en een proces van reflectie belangrijk. Duke en Stiggins (1990) noemen een aantal voorwaarden die de kans daarop vergroten. Deze hebben betrekking op kenmerken van het beoordelingssysteem, de beoordelaars en de door hen te geven feedback, en de omgeving waarin beoordeeld wordt.

Het beoordelingssysteem kan onder de volgende voorwaarden professionele ontwikkeling bevorderen: de docent heeft invloed op de totstandkoming van beoordelingscriteria, de *performance*-standaarden zijn helder voor docenten, de docent is op de hoogte van deze standaarden, de docent acht de standaarden geschikt voor de eigen situatie en bij de beoordeling wordt frequent gebruik gemaakt van lesobservaties en van gegevens over leerprestaties.

De volgende kenmerken van beoordelaars hebben een positief effect op professionele ontwikkeling: beoordelaars moeten geloofwaardig zijn als bron voor feedback, een hel-

pende relatie tot de docent hebben, betrouwbaar zijn, een niet-bedreigende werkwijze hanteren, geduldig en flexibel zijn, kennis hebben van de ambachtelijke aspecten van het onderwijzen, beschikken over het vermogen om suggesties te modelleren, bekend zijn met de leerlingen van de betrokken leerkracht, onderwijservaring hebben en de tijd nemen voor zowel de dataverzameling als de feedback. Uhlenbeck, Verloop en Beijaard (2002) onderstrepen in dit verband het belang van een goede training van beoordelaars.

Meer specifiek hebben kenmerken van zinvolle feedback betrekking op: de kwaliteit van de geuite ideeën, de diepgang van de informatie, de specificiteit van de informatie, de verantwoording van de informatie, de formaliteit van de feedback, de beschrijvende aard van de feedback, de timing van de feedback, de verbondenheid van de feedback met standaarden en de frequentie van de feedback (Duke & Stiggins, 1990).

Tot slot zijn de volgende kenmerken van de omgeving waarin beoordeeld wordt van invloed op professionalisering: helderheid van het beoordelingsbeleid, beschikbare tijd voor professionalisering, rustige werkverhoudingen zonder wrijvingen, mogelijkheid om onder werktijd te professionaliseren, de mate waarin professionele groei als voornaamste doel van evaluatie wordt gezien en de mate waarin de resultaten van beoordeling gebruikt worden voor verdere professionele ontwikkeling. Beijaard en Uhlenbeck (2001) schrijven over dit laatste het volgende: “bepaal samen met docenten specifieke, bereikbare doelen en maak gebruik van ervaren docenten om de prestaties van minder ervaren docenten te verbeteren”.

2.3 Ontwikkeling van standaarden voor competent handelen

Beoordelingsinstrumenten moeten zijn gebaseerd op een relevante en representatieve set standaarden. Om tot die standaarden te komen, worden in de literatuur de volgende werkwijzen beschreven (Boyd, 1989; Danielson, 1996; Danielson & McGreal, 2000; Dwyer, 1998; Howard & McColskey, 2001; Uhlenbeck, 2002; Verloop, Beijaard, & Van Driel, 1998):

- op basis van onderzoeksliteratuur samen-

stellen van een gefundeerd raamwerk voor competent handelen;

- uitwerken van de rollen die de docent in de beroepssituatie moet vervullen en de vaardigheden en deelvaardigheden die hiervoor nodig zijn; op grond daarvan de vereisten bepalen voor het beroep;
- gedetailleerd onderzoeken van het functioneren van de docent in praktijksituaties; het gaat hier, in tegenstelling tot de vorige procedure, om een empirische analyse van beroepsmatig functioneren;
- raadplegen van als excellent bekend staande docenten;
- in kaart brengen van docentvariabelen waarvan uit onderzoek is gebleken dat deze samenhangen met goede leerlingprestaties, en
- samenstellen van een brede commissie waarin praktiserende professionals een belangrijke stem hebben, waardoor aan zowel theoretische als praktische perspectieven recht wordt gedaan, hetgeen bovendien de acceptatie van te ontwikkelen standaarden bevordert.

Door deze combinatie van perspectieven en bronnen kan een voor het beroep dekkend overzicht van standaarden worden verkregen dat bovendien kan rekenen op een brede steun in het onderwijsveld.

In Nederland is een begin gemaakt met de ontwikkeling van standaarden voor het leraarberoep. Ten behoeve van de wet Beroepen in het Onderwijs is in 2006 een beschrijving gemaakt van startbekwaamheidseisen voor docenten in het basis en voortgezet onderwijs door de Stichting Beroepskwaliteit Leraren (2003). Het is aan lerarenopleidingen om deze eisen verder uit te werken tot concrete standaarden.

3 Zelfbeoordeling als instrument voor competentiebeoordeling

3.1 Zelfbeoordeling: kenmerken en uitgangspunten

In deze studie staat zelfbeoordeling met behulp van vragenlijsten centraal. Docenten geven aan de hand van een aantal stellingen een retrospectieve beoordeling van het eigen werkgedrag. Door op een groot aantal aspecten

van competentie te vragen naar de mate waarin een voorgestelde handelwijze van toepassing is op de eigen praktijk, kan de docent zichzelf bewust worden van het eigen handelen. Een voordeel van het stellen van een groot aantal vragen is dat een grote dekking kan worden bereikt voor verschillende domeinen van competent handelen. Nadelen van deze werkwijze zijn dat inschattingen van docenten niet vrij zijn van sociale wenselijkheid en dat de complexiteit van het competent handelen noodzakelijkerwijs wordt gereduceerd tot uitspraken over gepercipieerd werkgedrag. Uit verschillende studies is gebleken dat zelfbeoordelingen door de docent een positiever beeld van de docent opleveren dan wanneer leerlingen dezelfde docent beoordelen (Ostrander, 1996; Roelofs, Franssen, Houtveen, & Lagerweij, 1999; Zill & Nolin, 1994). Uit diezelfde studies bleek ook dat er een duidelijke samenhang bestaat tussen docent- en leerlingpercepties op het niveau van de uit meerdere items samengestelde schalen. Het scoreprofiel op de verschillende schalen van zelfbeoordelingen kon als betrouwbare maatstaf worden genomen om uitspraken te doen over vaak en minder vaak gepraktiseerde aspecten van onderwijzen, zoals procesgerichte instructie, aansluiting bij de belevingswereld van leerlingen, en samenwerkend leren.

De genoemde nadelen van zelfbeoordeling zouden onoverkomelijk zijn als deze zou worden ingezet voor *high-stakes*-doeleinden, zoals certificering en selectie. Zelfbeoordelingen kunnen echter zinvol ingezet worden voor professionaliseringsdoeleinden. Zo blijken zelfbeoordelingen (praktiserende) docenten meer bewust te maken van hun beroepstaken, omdat ze aanzetten tot reflectie (Boud, 1995; Duke & Stiggins, 1990). Wil zelfbeoordeling zinvol worden ingezet voor professionaliseringsdoeleinden, dan dient volgens Barber (1990) aan de volgende voorwaarden te worden voldaan:

- het docentgedrag dient nauwkeurig te worden geoperationaliseerd;
- uit resultaten van zelfbeoordelingen moet duidelijk worden op welke aspecten van competentie verbeteringen mogelijk zijn; die aspecten moeten bovendien veranderbaar zijn, en

- de feedback op grond van zelfbeoordelingen dient concrete suggesties voor verbetering te bevatten.

3.2 De ontwikkeling van EduProf, een instrument voor zelfbeoordeling

EduProf is, zoals gezegd, een instrument waarmee docenten in het hbo op betrouwbare en efficiënte wijze hun competenties kunnen beoordelen. Bij de ontwikkeling van het instrument zijn de volgende fasen doorlopen (Dwyer, 1998):

1. De formulering van een visie op onderwijs in het hoger onderwijs waarin de vraag centraal staat: wat is professioneel gedrag en wat niet?
2. De beschrijving van de taakgebieden en daarbij behorende *performance*-criteria, aansluitend bij wetenschappelijke inzichten en bestaande indelingen betreffende docentcompetenties voor docenten in het basis en voortgezet onderwijs. Aan de basis van het instrument ligt een beschrijving van de competenties op vijf onderscheiden taakgebieden: voorbereiden van onderwijs, uitvoeren van onderwijs, toetsen en evalueren van onderwijs, functioneren in de opleidingsorganisatie, en werken aan de eigen professionalisering. Deze beschrijving is tot stand gekomen na een uitvoerige literatuurstudie op een aantal relevante deelterreinen om een beeld van de *state of the art* te krijgen: leren en onderwijzen, krachtige leeromgeving, leren van docenten, (stadia in) professionele ontwikkeling van docenten, kennisbasis van docenten, schoolontwikkeling en schoolverbetering. Naast deze *'state of the art'*-inzichten zijn bestaande competentiebeschrijvingen bestudeerd en betrokken bij de competentiebeschrijving ten behoeve van EduProf. De resultaten ervan zijn opgenomen in een basisdocument (Roelofs, Van Diggele, Sanders, Groot, Nedermeijer, & Beijaard, 2002).
3. De ontwikkeling van itemsets voor zelfbeoordeling, competentieschalen, en de afleiding van beheersingsstandaarden op de afzonderlijke schalen. Om vast te stellen of docenten een bepaalde competentie voldoende beheersen, is in samenwerking met een panel deskundigen een grens-

score bepaald voor elke competentie. Het panel bestond uit twee ervaren docenten hoger onderwijs, twee onderwijsadviseurs hoger onderwijs, twee ervaren toetsontwikkelaars en twee ervaren lerarenopleiders. De grenswaarde vertegenwoordigt de score die deskundigen van een gemiddelde startende docent verwachten. Het panel kreeg hiertoe de vraag voorgelegd welk percentage beheersing een grenskandidaat zou moeten behalen, dus net bekwaam is om onderwijs te verzorgen in het hoger beroepsonderwijs. Deze methode wordt wel aangeduid als de Angoff-methode voor het bepalen van standaarden (Dwyer, 1998; Vos & Knuver, 2000).

4. Een eerste beproeving van het instrument bij vertegenwoordigers van toekomstige gebruikers, waarbij gelet is op de inhoudelijke relevantie van het itemdomein, de representativiteit van de itemverzameling voor dat domein, de scorbaarheid van de ontwikkelde items en de betrouwbaarheid van de ontwikkelde schalen voor de verschillende competenties.
5. Het verrichten van een vervolgstudie bij een grotere groep docenten.
6. De ontwikkeling van een definitief webgebaseerd programma aan de hand waarvan docenten uit het hbo hun competenties op vijf taakgebieden kunnen onderzoeken.

EduProf stelt de gebruikers in staat om op verschillende niveaus van nauwkeurigheid en specificiteit gegevens te verzamelen over hun competenties. Daartoe zijn een Quick scan en een zogenaamde Dieptescan ontwikkeld. Door de afname van de Quick scan, welke 53 vragen omvat, kunnen docenten een eerste beeld verkrijgen van de beheersing van de docentcompetenties. Resultaten uit de Quick scan kunnen een reden zijn om een competentie nauwkeuriger te onderzoeken met behulp van de zogenaamde Dieptescan, die in totaal meer dan 200 vragen bevat. EduProf berekent direct na afloop van de *scans* per competentie zogeheten beheersingsscores, die worden uitgedrukt op een procentuele schaal (van 0 tot 100). Bij de Dieptescan wordt als grens voor het beheersen van een competentie een score aangehouden van 65%. Afname van de EduProf zelfscans le-

vert een rapportage op in de vorm van een grafiek met een competentieprofiel waarbij scores in een figuur worden afgezet tegen de beheersingsstandaard. De figuren worden gevolgd door een toelichtend rapport. Daarin wordt per competentie steeds beschreven wat de gevolgen zijn van wel of niet beheersen ervan voor de direct betrokkenen: studenten, collega's of de organisatie als geheel. Een voorbeeld van commentaar bij niet beheersing van competentie 8 (begeleiden van leerprocessen) luidt: "Sluit uw begeleiding niet aan op hoe studenten hun studietaken (literatuur, opdrachten, werkstukken) aanpakken, dan bestaat het risico dat ze herhaaldelijk in dezelfde valkuilen stappen, dat misverstanden of foutieve opvattingen hardnekkig blijven bestaan, kortom dat een professionele taakuitvoering gevaar loopt."

De eerste drie fasen van de instrumentontwikkeling zijn beschreven in Nijveldt en Roelofs (2003). Deze studie betreft een rapportage van de vierde en vijfde fase van instrumentontwikkeling, een pilot en vervolgstudie bij vertegenwoordigers van toekomstige gebruikers, en betreft alleen de bovengenoemde Dieptescan.

De centrale vraagstelling van de studie is tweeledig:

- Is het mogelijk om een instrument voor zelfbeoordeling te ontwikkelen voor hbo-docenten dat voldoet aan hedendaagse eisen op het terrein van docentbeoordeling betreffende betrouwbaarheid, validiteit en praktische inzetbaarheid bij professionele ontwikkeling?
- In welke mate is het vanuit psychometrisch en onderwijskundig opzicht wenselijk het instrument te differentiëren naar specifieke beroepscontexten?

4 Onderzoeksozet

In de periode tot aan de lancering van EduProf is allereerst een pilotstudie verricht naar de totale itemset. Afname gebeurde bij een deel van de respondenten op papier en bij een ander deel webgebaseerd. In de pilotstudie werden psychometrische gegevens over het instrument verzameld. Daarnaast werden bij

stakeholders gegevens verzameld over de door hen gepercipieerde kwaliteit van de verschillende onderdelen van EduProf. Na de pilotstudie vond op basis van de bijgestelde itemset een vervolgstudie plaats onder een grotere groep docenten, deels op papier en deels *on line* via het inmiddels ontwikkelde programma. Doel van de vervolgstudie was het onderzoeken van de stabiliteit van de betrouwbaarheidscoëfficiënten. Immers, de ware betrouwbaarheid is afhankelijk van de steekproefgrootte. Een steekproef dient voldoende omvang te hebben willen resultaten over betrouwbaarheid met enige zekerheid kunnen worden gehanteerd (Feldt, 1965). Onderzocht zijn de betrouwbaarheden voor de te onderscheiden subschalen en de stabiliteit van de betreffende coëfficiënten. Daarnaast is nagegaan of er empirische aanwijzingen zijn die vereisen dat scores op een competentie worden gedifferentieerd naar specifieke beroepscontexten.

4.1 Instrumenten en dataverzameling

Instrument EduProf

- Algemene docentkenmerken

Als introductie op de Dieptescan is gevraagd naar de volgende persoonlijke gegevens: leeftijd, sekse, didactische vooropleiding, aantal jaren onderwijservaring en de sector waarbinnen onderwijs verzorgd wordt.

- Dieptescan: de scan Algemene beheersing en Deelaspectenscan

In onderhavige studie zijn twee onderdelen van de Dieptescan afgenomen: de scan Algemene beheersing en de Deelaspectenscan. De scan Algemene beheersing heeft betrekking op de eerder genoemde vijf taakgebieden, die nader zijn uitgesplitst in achttien competenties (zie Tabel 3). De vragenlijst bevat 182 items en verschaft per competentie een algemeen beeld over de eigen bekwaamheid.

Het tweede onderdeel betreft de Deelaspectenscan, waarbij twee competenties zijn onderverdeeld in deelcompetenties die bovendien worden bevraagd binnen specifieke contexten (zie Tabel 4). Het betreft competentie 7, gesprekken en opdrachtenbesprekingen begeleiden (totaal 32 items), en com-

petentie 8, begeleiding van leerprocessen (45 items). Binnen competentie 7 worden drie deelcompetenties onderscheiden. Allereerst betreft dit de deelcompetentie *het bewaken van de inhoud van gesprekken*. Deze deelcompetentie wordt bevraagd binnen twee afzonderlijke beroepssituaties: *groeps gesprekken voeren* ($k = 8$) en *discussies leiden* ($k = 10$). De tweede deelcompetentie betreft *het bewaken van vormtechnische aspecten van gesprekken* ($k = 5$) en de derde *het begeleiden van de gespreksdynamiek* ($k = 9$). Beide laatstgenoemde deelcompetenties worden generiek bevraagd, dus niet in afzonderlijke beroepssituaties. Binnen competentie 8, *het begeleiden van leerprocessen*, worden twee deelcompetenties onderscheiden: *beeldvorming van het leerproces* en *ondersteuning van het leerproces*. Deze beide deelcompetenties worden elk bevraagd binnen drie beroepssituaties: *het voeren van gesprekken in groepsonderwijs*, *het bespreken van opdrachten in groepsverband* en *het begeleiden van afstudeeropdrachten*. Door combinatie van deelcompetenties en beroepssituaties ontstaan voor competentie 8 zes setjes items (zie Tabel 4).

De items van beide scans hebben de vorm van stellingen over concreet onderwijsgedrag waarbij docenten op een vijfpuntsschaal kunnen aangeven in welke mate de betreffende stelling op hen van toepassing is. In de stellingen wordt in veel gevallen docentgedrag beschreven dat in verband gebracht wordt met het leren van de leerlingen. Daarmee volgen we het gehanteerde model van competent handelen. Twee voorbeelden: “Ik bespreek met studenten welke strategieën ze hanteren om relevante literatuur te bestuderen” en “Als studenten problemen hebben bij het bestuderen van de leerstof, geef ik hen zodanige handreikingen dat ze blijkens hun reacties adequater verder kunnen.” De antwoorden kunnen variëren van *in het geheel niet op mij van toepassing* tot *volledig op mij van toepassing*. Deze manier van beantwoorden kwam uit een eerste beproefing van de vragenlijsten onder enkele proefpersonen als meest duidelijk naar voren. Alternatieve antwoordcategorieën met globale frequentieaanduidingen als *vaak*, *soms* en met meer precieze aanduidingen als *1 keer per maand*

en *elke dag*, vielen af; de eerste omdat *vaak* voor iedereen iets anders kan betekenen, de laatste omdat deze niet eenduidig te interpreteren zijn in relatie tot het aantal door de docent te verzorgen cursussen of bijeenkomsten. Om positieve of negatieve antwoordtendenties tegen te gaan is een deel van de items negatief gesteld. Dit betrof ongeveer 10% van de items. Deze items werden achteraf gespiegeld.

De scores van de docenten op de afzonderlijke competenties werden berekend door alle items een score toe te kennen van 0 (*in het geheel niet op mij van toepassing*) tot 4 (*volledig op mij van toepassing*). In Bijlagen 1 en 2 zijn ter illustratie twee itemsets opgenomen van onderdelen uit respectievelijk de scan Algemene Beheersing en de Deelaspectenscan. Voor de berekening van de scores per competentie en deelcompetentie zijn de scores van alle bijbehorende items gesommeerd. De somscores werden daarna omgezet in een percentagescore door de behaalde score te delen door de maximaal te behalen score.

• Vragenblok kwaliteit van de items

In de pilot is bij de scan Algemene beheersing en Deelaspectenscan een vragenblok ingebouwd dat betrekking had op de kwaliteit van de items. Na elke vraag konden docenten aankruisen of de vraag duidelijk was, relevant voor het beroep en of het gevraagde gedrag redelijkerwijs verwacht mocht worden van een hbo-docent.

Afname van de competentiescans

Tabel 1 geeft weer welke groepen uit de pilotstudie en de vervolgstudie welke onderdelen van de beide scans hebben doorlopen. Ten behoeve van de pilot zijn 161 docenten verzocht om de vragenlijst schriftelijk in te vullen voor de eerste twee groepen competenties van de Scan Algemene beheersing en de Deelaspectenscan; 36 docenten (20%) afkomstig uit zes hogescholen hebben uiteindelijk deelgenomen.

Tevens is in de pilotstudie een groep van 30 hbo-docenten afkomstig uit vier hogescholen aangeschreven om deel te nemen aan een experimentele *on line* versie van EduProf. Deze had betrekking op de compe-

Tabel 1

Afnamen in de verschillende onderzochte groepen

Scan Algemene beheersing	Pilot	Vervolgstudie
Competentiegroep 1: voorbereiden van onderwijs	Postenqu�ete hbo-docenten ($n=36$) van 6 hogescholen	Web-based afname ($n=50$) onder hbo-docenten van 16 hogescholen Schriftelijke ($n=56$) afname onder hbo-docenten van een hogeschool en een lerarenopleiding
Competentiegroep 2: uitvoeren van onderwijs		
Competentiegroep 3: evalueren en bijstellen van onderwijs		
Competentiegroep 4: functioneren in de organisatie		
Competentiegroep 5: professionaliseringstaken		
Deelaspectenscan		
Competentie 7: gesprekken	Postenqu�ete hbo-docenten ($n=36$)	
Competentie 8: begeleiden van leerprocessen		

tenties 3 tot en met 5 van de scan Algemene beheersing; 22 docenten (73%) hebben hun medewerking verleend.

Voor de vervolgstudie zijn innovatie- en opleidingsmanagers van in totaal 16 hogescholen benaderd met de vraag om hun docenten deel te laten nemen aan de webgebaseerde versie van EduProf. Deelname aan het onderzoek zou geschieden op basis van vrijwilligheid. In totaal werden 113 toezeggingen verkregen van individuele docenten. Mede door onwennigheid van respondenten met webgebaseerde instrumenten – een aantal bewaarde hun gegevens niet – is een deel van de data verloren gegaan. Weliswaar doorliepen alle docenten verschillende onderdelen van EduProf, zoals de Quick scan, maar van slechts 44% ($n = 50$) van hen kwamen data beschikbaar over de scan Algemene beheersing. Een tweede groep uit de vervolgstudie betrof 90 docenten van twee grote hogescholen bij wie de vragenlijsten schriftelijk zijn afgenomen in het kader van een kennismakingsochtend over EduProf, waarbij docenten tevens de gelegenheid werd geboden om de competentiescans individueel te doorlopen. Tijdens deze kennismaking met EduProf is geen enkele inhoudelijke informatie gegeven over de te beoordelen competenties, noch over de items. Wel werden de resultaten na afloop van de afname nabesproken. Van 56 docenten zijn bruikbare data verkregen (respons 70%).

In verband met het vrijwillige karakter

van deelname was de vorming van een landelijk representatieve steekproef niet mogelijk, hetgeen generaliserende uitspraken over competenties naar de totale populatie hbo-docenten bemoeilijkt. Dit was echter ook niet de opzet van deze studie. Onderzoek naar de stabiliteit van de betrouwbaarheidsco effici nten was op basis van de steekproef wel mogelijk.

Interviews

Uit de pilotgroep is ten behoeve van een interview een zo gevarieerd mogelijke selectie gemaakt op basis van onderrwijservaring, het vakgebied waarin men lesgeeft en scores op de scans van tien docenten en drie kaderfunctionarissen. Alle respondenten werd verzocht om ter voorbereiding op het gesprek de terugrapportage en de toelichtende teksten op de competenties zorgvuldig te bestuderen. De interviews duurden ongeveer een uur en waren halfgestructureerd van opzet. De volgende onderwerpen kwamen aan bod: inhoudsvaliditeit en ecologische validiteit, legitimering van het competentiedomein, bruikbaarheid in de beroepscontext, kwaliteit van de terugrapportage, kwaliteit van de toelichtende teksten bij de competenties en de mogelijke bijdrage van het instrument aan professionele ontwikkeling.

4.2 Respondenten

Tabel 2 geeft weer hoe de respondenten van de twee onderzoeksgroepen zijn verdeeld

over de in het onderzoek betrokken onderwijssectoren. Deze tabel laat tevens zien welke didactische vooropleiding docenten hebben genoten. De onderzoeksgroep ($n = 58$) van de pilot omvatte 56% mannen en 44% vrouwen. In de vervolgstudie ($n = 106$) was de verdeling respectievelijk 58,8 en 41,1%. De gemiddelde leeftijd van de respondenten was in de pilot 45 jaar ($sd = 9,2$) en in de vervolggroep 47,4 ($sd = 9,7$). Uit een vergelijking met recente populatiegegevens van het Ministerie van Onderwijs, Cultuur en Wetenschap (2007) en de HBO-raad (2006) kan worden opgemaakt dat de steekproef van de vervolgstudie representatief is naar leeftijd en geslacht.

Het aantal jaren onderriservaring varieerde van een half jaar tot 29 jaar in de pilot en van 4 tot 40 jaar in de vervolggroep. Het gemiddeld aantal jaren onderriservaring bedroeg in de pilot 14 jaar ($sd = 9,9$) en in de vervolggroep 19,2 jaar ($sd = 9,7$). De vervolggroep heeft dus langere onderriservaring dan de pilotgroep.

De respondenten van de vervolggroep zijn

afkomstig uit 17 van de 45 hogescholen in Nederland. Van hen komt 48% uit de pedagogische sector, waarbij het (leren) verzorgen van onderris het object van de opleiding is. Van de overige docenten komt 19% uit de economische sector, 14,5% uit de sociaal-agogische sector, 6% uit de kunstsector, 4,8% uit de sector gezondheidszorg, en 1% uit de sector techniek.

Bijna de helft van de ondervraagden (49,4%) uit de vervolggroep heeft naast de functie van docent een of meer andere rollen in hun organisatie. Het gaat daarbij om een van de volgende rollen: stagebegeleider (30,8%), leerplanontwikkelaar (29,1%), mentor (27,9%), opleidingscoördinator (18,9%), managementfunctionaris (10,3%), projectleider van een onderrisvernieuwing (7,1%), supervisor (4,6%), en onderrisontwikkelaar (27,3%).

Al met al vormt de vervolggroep een steekproef die veel spreiding over relevante variabelen laat zien, maar die niet representatief is voor de populatie docenten hoger onderris. Wel heeft deze groep een vol-

Tabel 2

Verdeling van respondenten over sector en didactische vooropleiding

Sector waarin werkzaam:	Didactische vooropleiding					Totaal
	Onbekend	Geen didactische vooropleiding	Didactische cursus(sen)	Eerstegraads lerarenopleiding	Tweedegraads lerarenopleiding	
Pilot						
Technisch (informatica)	0	5	3	3	5	16
Economisch	0	5	3	2	2	12
Gezondheidszorg	0	3		3	2	8
Sociaal Agogisch	0	10	8	2		20
Totaal	2	23	14	10	9	58
Vervolggroep						
Technisch (informatica)	0	0	0	1	0	1
Economisch	0	8	11	2	1	22
Pedagogisch	6	0	11	22	12	51
Gezondheidszorg	0	2	2	0	0	4
Sociaal-agogisch	0	14	4	2	0	20
Kunstonderris		2	0	0	0	2
Anders	0		6	0	0	6
Totaal	6	26	34	27	13	106

doende omvang om stabiele schatters van betrouwbaarheidsindices te verkrijgen.

Van de 10 hbo-docenten die deelnamen aan het interview waren vier afkomstig uit de sociaal agogische sector, drie docenten uit de sector economie en drie docenten uit de sector techniek. Vier van de docenten hadden minder dan vijf jaar ervaring; zes hadden gemiddeld 17 jaar ervaring. Bovendien namen drie kaderfunctionarissen deel aan de interviews, waaronder een onderwijskundige binnen het hbo, een opleidingsmanager en een adviseur bij een bureau voor onderwijsmanagement.

4.3 Analyses

Om de betrouwbaarheid van de verschillende subschalen van de competentieascans te onderzoeken zijn coëfficiënten Cronbach's α berekend. Op grond van de itemanalyses zijn items verwijderd uit de competentieschalen waarvan de itemrestcorrelatie lager was dan 0,20. Voor de gegevens van de vervolggroep zijn ook de 90% betrouwbaarheidsintervallen rondom de α -coëfficiënten berekend. Hiermee wordt een beeld verkregen van de stabiliteit van de α -coëfficiënten. Aangezien α een onderschatting is van de betrouwbaarheid, is bovendien de hoogste ondergrens van de betrouwbaarheid, de zogenaamde Greatest Lower Bound (GLB, Verhelst, 2000) berekend. Deze is interpreteerbaar als een α -coëfficiënt.

Om na te gaan of de gemaakte specifiekere onderscheidingen naar deelcompetenties en beroepssituaties empirisch te rechtvaardigen zijn, werden intercorrelaties berekend. Hoge intercorrelaties wijzen op een beperkte discriminante validiteit. Bijvoorbeeld, het onderscheid tussen bekwaamheid in 'beeldvorming' en in 'ondersteuning' bij het begeleiden van leerprocessen is moeilijk staande te houden als de samenhang tussen scores op die deelcompetenties hoog is. Een lage samenhang rechtvaardigt daarentegen een differentiatie naar deze deelcompetenties.

De interviews zijn eerst letterlijk uitgeschreven, waarna uitspraken van de respondenten zijn geïdentificeerd en geordend onder de rubrieken die zijn genoemd bij de interviews als databron.

5 Resultaten

Om inzicht te krijgen in de kwaliteit van de verschillende competentieschalen van de scan Algemene beheersing en de Deelaspectenscan wordt in deze paragraaf allereerst een overzicht gegeven van de psychometrische gegevens van de schalen. Daarbij wordt tevens gerapporteerd over de kwaliteit van de vragen van de beide scans. Tot slot wordt aandacht besteed aan de mate van samenhang tussen de scores op de scan Algemene beheersing en de Deelaspectenscan.

5.1 Betrouwbaarheid en validiteit van de competentiescans

De scan Algemene beheersing

In Tabel 3 staan de psychometrische gegevens van de competentieschalen voor de scan Algemene Beheersing vermeld. De gegevens zijn uitgesplitst naar de pilotgroep en de vervolggroep. In de tabel zijn tevens de beheersingsscores per competentie voor beide groepen opgenomen.

Tabel 3 laat zien dat de betrouwbaarheden van de schalen in de pilot liggen tussen 0,52 en 0,88. De α -coëfficiënten liggen voor het merendeel van de schalen boven 0,70. De schalen van competentie 12 (*maken van toetsen*) en 13 (*hanteren van toetsinstrumenten*) zijn respectievelijk matig en laag betrouwbaar. Daarnaast blijkt dat van competentie 5 slechts twee items resteren.

De items die in de pilot zijn verwijderd vanwege lage R_{ii} -waarden zijn voorafgaand aan de vervolgstudie anders verwoord. Tevens is bij de herziening van de items gekeken naar de resultaten op de beoordelingen van de afzonderlijke items. Deze beoordelingen vielen in het algemeen positief uit. Als er negatieve oordelen waren over de items, dan hadden deze vooral betrekking op onduidelijke formuleringen. Het aantal items per competentie dat beoordeeld werd als *onduidelijk geformuleerd, irrelevant of te vergaand* was relatief klein. Als grens voor verwijdering werd genomen dat 20% van de docenten een negatief oordeel uitsprak over een item; 10 items kwamen in aanmerking voor deze kwalificatie. Vijf items waren volgens dit criterium niet voldoende duidelijk

Tabel 3

Psychometrische gegevens van de schalen in de pilot en de vervolggroep

Taakgebied en competentie	Pilotgroep					Vervolggroep						
	K	n	α	Gem.	Sd.	90% betr.						
						K	n	α	interv.	GLB	Gem.	Sd.
Vorbereiden van onderwijs												
1 Presenteren van studieonderdeel	7	36	,74	75,7	15,9	7	99	,75	,68-.81	,84	75,8	17,6
2 Vorbereiden van bestaand studieonderdeel	7	36	,71	76,2	15,1	10	96	,74	,67-.80	,85	83,6	11,7
3 Vorbereiden van zelfstandig leren	11	36	,80	63,2	15,4	11	101	,81	,76-.85	,90	73,0	13,5
4 Vorbereiden beroepsgerichte onderwijs-activiteiten	6	36	,69	57,1	16,8	9	100	,76	,69-.81	,86	74,2	14,0
5 Gebruik van ict in het onderwijs	2	36	,75	54,5	28,9	9	102	,89	,86-.92	,95	60,8	25,1
Uitvoeren van onderwijs												
6 Uitleg geven in uiteenlopende onderwijs-regimes	12	36	,77	69,6	12,2	12	98	,79	,73-.84	,90	78,3	11,5
7 Gesprekken en opdracht-besprekingen begeleiden	14	36	,85	68,7	14,7	14	103	,83	,79-.87	,93	77,4	12,2
8 Begeleiding van leerprocessen	12	36	,84	66,3	14,4	13	103	,83	,79-.87	,92	75,4	12,7
9 Organiseren van onderwijs-situaties	4	36	,71	63,2	17,8	7	95	,69	,60-.76	,81	77,3	15,0
10 Creëren van positief leerklimaat	9	36	,84	64,1	15,8	9	95	,80	,74-.84	,90	76,2	13,8
11 Stimuleren van positieve studiehouding studenten	8	36	,83	72,7	15,9	8	97	,75	,68-.81	,83	77,8	13,0
Toetsen / evalueren onderwijs												
12 Maken van toetsen	3	36	,62	68,7	19,4	8	94	,83	,78-.87	,90	78,4	15,7
13 Hanteren van toets-instrumenten	5	36	,52	72,7	13,9	4	95	,63	,51-.72	,65	80,9	12,0
14 Evalueren en bijstellen van studieonderdeel	12	22	,88	69,5	18,3	12	97	,88	,85-.91	,94	81,4	14,6
Functioneren in opleidingsorganisatie												
15 Organisatie-conform werken	9	22	,72	71,1	15,5	10	99	,78	,73-.83	,88	80,2	13,1
16 Overleggen	7	22	,74	69,2	12,4	9	98	,83	,78-.87	,91	76,9	15,3
Eigen professionalisering												
17 Individueel professionaliseren	11	22	,88	64,5	16,8	10	96	,76	,70-.82	,90	76,4	18,4
18 Leren in teamverband	12	22	,84	67,1	15,8	12	96	,84	,80-.88	,94	76,9	13,4

geformuleerd. Drie items werden als niet voldoende relevant beschouwd en bij twee items vonden ondervraagden dat het bevraagde onderwijsgedrag redelijkerwijs niet mocht worden verwacht van hbo-docenten. De betreffende items werden geherformuleerd. De schaal voor competentie 5 (*gebruik van ict in het onderwijs*) is opnieuw geconstrueerd aan de hand van een meer uitgebreide operationalisatie. Het resultaat is een aanzienlijke uitbreiding van het aantal items en een hoge betrouwbaarheid.

Over het algemeen blijven de α -waarden voor de schalen in de vervolgstudie zeer bevredigend. Ze variëren tussen 0,63 en 0,89 (zie Tabel 3). De GLB-coëfficiënten volgen hetzelfde patroon, maar liggen alle wat hoger. Twee van de 18 competentieschalen bezitten α 's tussen 0,60 en 0,70 en zijn daarmee matig betrouwbaar: 9 (*onderwijs organiseren*), en 13 (*hanteren van toetsinstrumenten*). Met in achtname van de brede intervallen rondom de α -coëfficiënten moeten resultaten op vooral de laatste schaal met enige voorzichtigheid worden geïnterpreteerd.

Van de scan Algemene beheersing zijn de schalen voor acht competenties ongewijzigd gebleven ten opzichte van de pilot. Verder valt op dat de behaalde scores op de verschillende competenties in de vervolgstudie duidelijk hoger uitvallen dan in de pilot. Het is zelfs zo dat de gemiddelde beheersingsscores in de vervolgstudie ruim boven de vooraf via experts vastgestelde beheersingsniveaus (65%) liggen. *T*-toetsen op de gemiddelde scores van zeven van de acht identieke schalen van de twee onderzochte groepen (pilot-versus vervolgstudie) laten significante verschillen zien. Voor de competenties 3, 6, 7, 10, 11, 14 en 18 werden respectievelijk significante *t*-waarden (allen getoetst met $\alpha = 0,01$) gevonden van $-3,5$ ($df = 1, 134$), $-5,1$ ($df = 1, 130$), $-3,1$ ($df = 1, 136$), $-3,8$ ($df = 1, 129$), $-4,8$ ($df = 1, 130$), $-2,9$ ($df = 1, 118$) en $-2,7$ ($df = 1, 117$). De verschillen in scores kunnen deels worden toegeschreven aan de niet gelijkwaardige steekproeven in beide studies. Zo bestaat een aanzienlijk deel (48%) van de vervolgstudie uit docenten die zelf leraren opleiden. In de pilotgroep werkt geen van de docenten aan een lerarenoplei-

ding, ofschoon wel een deel van die groep docenten die opleiding zelf als achtergrond heeft. Op grond van deze kenmerken zouden dus hogere competentiescores van de vervolgstudie mogen worden verwacht.

De Deelaspectenscan

Tabel 4 geeft een overzicht van de psychometrische gegevens van de schalen behorend bij *gesprekken voeren en begeleiden van leerprocessen* van de Deelaspectenscan.

De tabel laat zien dat de psychometrische kwaliteit van de schalen in de vervolgstudie over het algemeen bevredigend is, waar in de pilotgroep de schaal 8B1 (*geven van ondersteuning tijdens begeleiding van afstudeeropdrachten*) nog een lage betrouwbaarheid heeft. Bij drie schalen zijn voorts de betrouwbaarheidsintervallen rondom de α -coëfficiënten tamelijk breed: schaal 7B, 8B1 en 8B2. Met andere woorden, er is nog enige onzuiverheid in de geschatte α -coëfficiënten voor die schalen.

Geen van de items van de Deelaspectenscans werd door meer dan acht procent van de respondenten negatief beoordeeld. Dit wijst erop dat de kwaliteit van de items van de Deelaspectenscan nog iets beter werd beoordeeld dan die van de scan Algemene beheersing.

Opnieuw valt op dat docenten van de vervolgstudie hogere scores op de corresponderende schalen dan docenten van de pilotgroep. Op twee van de zes identieke schalen is dit verschil significant: het begeleiden van de gespreksdynamiek als onderdeel van gesprekken begeleiden ($t = -4,4$, $df = 43,7$, $p \leq 0,01$) en beeldvorming van het leerproces tijdens opdrachtbesprekingen ($t = -2,9$, $df = 88$, $p \leq 0,01$).

5.2 Differentiatie van competenties naar deelcompetenties en beroeps-situaties

De scan Algemene beheersing is bedoeld om een generiek beeld te geven van het behaalde niveau op de onderscheiden competenties. De Deelaspectenscan biedt inzicht in de beheersing van deelcompetenties, bijvoorbeeld 'de inhoud van gesprekken bewaken' en deelcompetenties in specifieke beroepsituaties, bijvoorbeeld *de inhoud van discussies bewa-*

Tabel 4

Psychometrische gegevens van de schalen gesprekken voeren en begeleiden van leerprocessen van de Deelaspectenscan

Deelcompetentie	Pilotgroep					Vervolggroep						
	K	n	α	Gem.	Sd.	K	n	α	90% betr. interv.	GLB	Gem.	Sd.
7 Gesprekken en opdrachtbesprekingen begeleiden												
7A1: De inhoud van gesprekken bewaken	7	19	,81	70,2	13,1	8	102	,83	,73-,90	,94	77,8	13,7
7A2: De inhoud van discussies bewaken	8	19	,86	59,4	13,5	10	27	,87	,79-,93	,97	72,7	14,5
7B: Bewaken van vormtechnische aspecten van gesprekken	5	19	,74	64,6	13,8	5	30	,73	,55-,86	,87	74,9	16,2
7C: De gespreksdynamiek begeleiden	9	19	,70	63,5	12,1	9	102	,86	,78-,92	,96	75,2	15,0
8. Begeleiden leerprocessen												
A. Beeldvorming van het leerproces												
8A1: Afstudeeropdracht	3	16	,73	80,4	14,5	3	22	,83	,67-,92	,85	84,1	13,6
8A2: Groeps-onderwijs-gesprekken	6	16	,73	61,0	14,4	8	74	,74	,66-,82	,87	74,2	16,6
8A3: Opdracht-besprekingen	9	16	,90	60,2	17,1	9	74	,80	,73-,85	,89	72,7	15,4
B. Ondersteuning												
8B1: Begeleiden afstudeer-opdrachten	3	16	,59	66,1	15,5	4	23	,71	,46-,84	,75	83,4	12,0
8B2: Groeps-gesprekken	6	16	,70	65,9	15,0	6	74	,69	,58-,78	,79	76,8	13,0
8B3: Opdracht-besprekingen in groepsverband	14	16	,95	71,1	15,0	15	74	,88	,84-,91	,96	78,6	12,4

ken. In de Tabellen 5 en 6 zijn de correlaties vermeld tussen de schaalscores op de scan Algemene beheersing en de schaalscores op de Deelaspectenscan. Ook zijn correlaties vermeld tussen de verschillende subschalen van de Deelaspectenscan onderling.

Tabel 5 wijst uit dat er significant positieve correlaties bestaan tussen de behaalde score op competentie 7 van de scan Algemene beheersing en de deelcompetenties van de

Deelaspectenscan op deze competentie. Vooral de deelcompetenties *de inhoud van de gesprekken bewaken* ($r = 0,81$) en *het begeleiden van de gespreksdynamiek* ($r = 0,83$) hangen sterk samen met de algemene competentie *gesprekken en opdrachtbesprekingen begeleiden*. De samenhang tussen de score op de algemene competentie en de scores op respectievelijk *het begeleiden van vormtechnische aspecten van gesprekken* en *het bewa-*

Tabel 5

Pearson product-momentcorrelaties tussen de scores van de scan Algemene beheersing en Deelaspectenscan op competentie 7. Gesprekken en opdrachtbesprekingen begeleiden (vervolgstudie)

Score op (deel) competentie	Score op (deel)competentie			
	scan Algemene beheersing competentie 7	7A1	7A2	7B
7A1.				
De inhoud van de gesprekken bewaken	0,81**			
7A2.				
De inhoud van de discussies bewaken	0,42*	0,79**		
7B.				
Bewaken van vormtechnische aspecten van gesprekken	0,61**	0,39*	0,45*	
7C.				
De gespreksdynamiek begeleiden	0,83**	0,50**	0,25	0,39*

*= $p \leq 0,05$, **= $p \leq 0,01$

ken van discussie-inhoud is duidelijk lager (respectievelijk 0,61 en 0,42).

In Tabel 5 worden ook de correlaties tussen de scores op de verschillende deelcompetenties weergegeven. Hoge correlaties wijzen op een beperkte discriminante validiteit. Een lage samenhang rechtvaardigt een onderscheid in deelcompetenties. De correlaties tussen de score op *het bewaken van het vormtechnische aspect van gesprekken* en de scores op de andere deelcompetenties liggen alle onder de 0,50 (resp. 0,39, 0,45, 0,39). Ook valt op dat de score op *het bewaken van de discussie-inhoud* niet significant samenhangt met de score op *het begeleiden van de gespreksdynamiek* ($r = 0,25$). De score op de deelcompetentie *het bewaken van de gespreksinhoud* en score op *het bewaken van de discussie-inhoud* hangen sterk samen ($r = 0,79$).

Tabel 6 laat zien dat er significant positieve correlaties bestaan tussen de behaalde score op competentie 8 van de scan Algemene beheersing (*begeleiden van leerprocessen*) en de scores op vijf van de zes onderdelen van de deelaspectenscan. De correlaties tussen *begeleiden van leerprocessen* in het algemeen enerzijds en de scores voor respectievelijk *beeldvorming* en *ondersteuning* bij *groepsonderwijs* en *opdrachtbesprekingen* anderzijds zijn hoog (waarden tussen 0,71 en

0,87). De scores op *beeldvorming* en *ondersteuning* bij *afstudeeropdrachten* hangen minder sterk samen met de algemene score voor begeleiding van leerprocessen (respectievelijk 0,45 en 0,51). De correlaties tussen de deelcompetenties in specifieke beroepssituaties, zowel binnen dezelfde beroepssituatie als tussen verschillende beroepssituaties variëren van aanzienlijk tot hoog (tussen 0,52 en 0,85).

Opmerkelijk is de hoge samenhang tussen de scores op de twee deelcompetenties *beeldvorming* en *ondersteuning* binnen de beroepssituatie *bespreken van opdrachten* (respectievelijk $r = 0,87$ en 0,80). Vier van de vijf correlaties onder de grens van 0,70 komen voor rekening van de beroepssituatie *begeleiden van afstudeeropdrachten*. Daarnaast valt *beeldvorming* bij begeleiding tijdens groeps- en groeps- en groeps- en groeps- onderwijs niet samen met *ondersteuning* tijdens groeps- en groeps- onderwijs, getuige de correlatie tussen de betreffende scores ($r = 0,59$).

5.3 Praktische inzetbaarheid van het instrument

Inhoudsvaliditeit en ecologische validiteit
Zeven respondenten vonden de niet-situatiegebonden scan Algemene beheersing al concreet. Een iets groter deel ($n = 9$) vond dat de Deelaspectenscan in hoge mate is toegespitst

Tabel 6

Pearson product-momentcorrelaties tussen de scores van de scan Algemene beheersing en Deelaspectenscan op competentie 8. Begeleiden van leerprocessen (vervolgstudie)

Score op (deel) competentie	Scan Algemene Beheersing competentie 8	Score op (deel)competentie				
		8a1	8a2	8a3	8b1	8b2
A. Beeldvorming van het leerproces						
8A1. Afstudeeropdracht	0,45					
8A2. Groepsonderwijs, gesprekken	0,71**	0,56**				
8A3. Opdrachtbesprekingen	0,87**	0,56**	0,80**			
B. Ondersteuning						
8B1. Begeleiden van afstudeeropdrachten	0,51*	0,85**	0,52*	0,55**		
8B2. Groepsgesprekken	0,80**	0,68**	0,59**	0,71**	0,73**	
8B3. Opdrachtbesprekingen in groepsverband	0,86**	0,77**	0,66**	0,82**	0,76**	0,78**

*= $p \leq 0,05$, **= $p \leq 0,01$

op concrete situaties. De kaderfunctionarissen, aan wie gevraagd werd of het instrument erin slaagt om diepgaand op competenties in te gaan, vonden dit allen ($n = 3$) het geval. Alle docenten ($n = 10$) vonden dat het instrument een werkelijkheidsgetrouw inzicht geeft in situaties. De kaderfunctionarissen waren unaniem van mening dat het instrument een nog meer werkelijkheidsgetrouw inzicht zou geven wanneer dezelfde vragen over de docent in kwestie ook zouden worden voorgelegd aan collega's en studenten.

Legitimering van het competentiedomein

De relevantie van de onderscheiden didactische competenties werd door alle respondenten onderschreven. Van de dertien respondenten konden er zes met zekerheid zeggen dat zij de competenties representatief vinden voor het taakgebied. Drie respondenten zeiden te weinig overzicht te hebben om hier een uitspraak over te kunnen doen. Een van

de respondenten zou als aanvulling op de onderscheiden competenties de 'modelfunctie' van docenten als een competentie zien. Drie respondenten hadden een weinig uitgesproken mening.

Kwaliteit van de terugrapportages van het instrument

Alle 13 respondenten vonden de feedback uitgebreid en specifiek genoeg. Volgens een vijftal docenten zou de feedback eventueel uitgebreid kunnen worden met tips over mogelijke manieren van aanpak bij het ontwikkelen van de afzonderlijke competenties. Andere docenten ($n = 5$) zouden dit te ver vinden gaan. Het zou ten koste kunnen gaan van de prettige toon die de feedback nu weet te raken. De docent wordt nu benaderd als een professional en daar past volgens deze docenten bij dat ze zelf de verantwoordelijkheid kunnen nemen om een vervolg te geven aan de feedback. Onder deze verantwoorde-

lijkheid valt ook dat docenten waar nodig de hulp van collega's of leidinggevendenden inroepen. Het instrument is daarmee slechts een middel om het proces van professionalisering op gang te brengen.

Kwaliteit van de toelichtende teksten bij de competenties

De toelichtende teksten bij de competenties waarvan docenten indien gewenst gebruik kunnen maken, werden door docenten niet unaniem gezien als een waardevol onderdeel van het instrument, hoewel de docenten de kwaliteit ervan positief beoordeelden. Zes docenten die de toelichtende teksten als essentieel beschouwden, vonden dat de teksten inzicht bieden in de theorie die ten grondslag ligt aan het instrument. Zij vonden het belangrijk om voorafgaand aan de zelfbeoordeling te weten waarop zij zichzelf zouden beoordelen en waarom. Docenten die de toelichtende teksten overbodig vonden ($n = 4$), gaven aan dat de inhoud en het belang van de competenties al voor zich spreken.

Bruikbaarheid van het instrument in de beroepscontext

Aan de bruikbaarheids van een aanvaardbare tijdsinvestering wordt volgens kaderfunctionarissen tegemoet gekomen doordat docenten niet alle vragen van het instrument hoeven in te vullen. Afhankelijk van hun interesses en de uitkomsten van de Quick scan kunnen zij ervoor kiezen de vragen van één of meer competenties van de scan Algemene beheersing te beantwoorden. De ondervraagde docenten achtten het niet reëel om aan alle competenties tegelijk te werken. Volgens hen is het van belang docenten ervan te doordringen dat het doordenken van de consequenties van de resultaten veel meer tijd vergt dan het doorlopen van de instrumenten. Daarover zou van tevoren duidelijkheid moeten worden geschapen.

Mogelijke effecten van de zelfbeoordeling

Docenten waren van mening dat het ontwikkelde instrument kan stimuleren tot bewustwording van de docentrol, het bedenken en uitproberen van handelingsalternatieven en het ondernemen van professionele activiteiten. De vier beginnende docenten (geopera-

tionaliseerd als 0-5 jaar ervaring) verschillen in hun mening van de ervaren docenten ($n = 6$). Zij gaven aan dat de zelfbeoordeling heeft bijgedragen aan bewustwording, terwijl de ervaren docenten dit niet zo hebben ervaren. Twee van hen waren van mening dat het instrument meer geschikt zou zijn voor beginnende dan voor ervaren docenten. Bovendien gaven zij aan persoonlijk vooral geïnteresseerd te zijn in zelfbeoordeling op het gebied van organisatie-, onderzoeks- en ontwikkeltaken, taken die doorgaans door meer ervaren docenten worden vervuld. Ook werd de mogelijkheid geopperd om *scans* te ontwikkelen voor specifieke rollen zoals die van studieloopbaanbegeleider of die van coach van beginnende docenten.

De activiteiten die de ondervraagden zouden willen ondernemen naar aanleiding van de feedback zijn: het bespreken van de resultaten met collega's, observatie door collega's, het zoeken van draagvlak voor professionalisering binnen het team, het bedenken van vervolgstappen op de zelfbeoordeling, begeleidde intervisie, het aanvragen van een functioneringsgesprek, het volgen van een cursus en het bestuderen van literatuur. Docenten zouden bij het ondernemen van activiteiten idealiter de hulp van collega's inroepen.

6 Conclusies en discussie

6.1. Betrouwbaarheid, validiteit en praktische inzetbaarheid

In deze studie zijn vijf taakgebieden van hbo-docenten achtereenvolgens geoperationaliseerd in 18 competenties en uitspraken bij deze competenties. Op basis van de pilotstudie en de vervolgstudie bleek het mogelijk een instrument voor zelfevaluatie te ontwikkelen met voor iedere competentie (redelijk) betrouwbare schalen. Aan deze operationalisatie ligt een uitgebreide literatuurstudie ten grondslag, beschreven in een basisdocument (Roelofs et al., 2002). De respondenten gaven aan dat de competenties representatief zijn voor de taakgebieden waarop hun werk als docent betrekking heeft en relevant zijn voor hun onderwijspraktijk. Omdat een beperkt aantal docenten is geïnterviewd, kunnen we alleen met de nodige voorzichtigheid

concluderen dat het instrument inhoudelijk en ecologisch valide is. Aanwijzingen voor de validiteit van het instrument zijn voorts verkregen door de bevinding dat docenten in de vervolggroep, waarvan het merendeel een lerarenopleiding had genoten, een hogere score behaalden dan docenten uit de pilot-groep, waarvan het merendeel geen lerarenopleiding had genoten.

Om echter generaliserende uitspraken te kunnen doen over bekwaamheidsniveaus van verschillende subgroepen docenten, zal een valideringsstudie moeten worden verricht met een representatieve steekproef, eventueel te onderscheiden naar bijvoorbeeld sector en vooropleiding. Een methode ter validering van het instrument die dan kan worden toegepast betreft de 'contrasterende groepen methode' (Newble, Hoare, & Baxter, 1982). Daarbij wordt een instrument afgenomen bij kandidaten waarvan men mag verwachten dat ze van elkaar verschillen op grond van de oefendheid of stadium in de beroepsloopbaan. Het vormen van een representatieve steekproef bleek in ons geval, waarbij docenten vrijwillig deelnamen, niet mogelijk, ondanks de grote inspanningen die daarvoor zijn geleverd. Redenen om niet deel te nemen waren onder meer: geen tijd, het gebruik van andere (eigen) competentiescans, en twijfel aan het nut van deelname zonder flankerende professionaliseringsactiviteiten.

Met betrekking tot de bruikbaarheid van het zelfbeoordelingsinstrument binnen de beroepscontext waren de respondenten van mening dat: 1) het instrument veelzijdige keuzemogelijkheden biedt bij het bepalen van situaties en competenties die beoordeeld worden en 2) het instrument goed is in te passen binnen bestaande professionaliseringstrajecten en -activiteiten. Wil het instrument praktisch inzetbaar zijn, dan is het van belang om te beschikken over de benodigde ict-faciliteiten, in ieder geval op de werkplek, wat momenteel bijna overal het geval is.

6.2. Differentiatie van competentie naar deelcompetenties en beroepscontexten

In recente literatuur wordt een voorkeur uitgesproken voor contextgebonden evaluatie van competenties. Een argument hiervoor is

dat bevindingen over competent handelen in de ene situatie vaak niet extrapoleerbaar zijn naar andere situaties vanwege allerhande contextfactoren, zoals leerjaar, vak, en samenstelling van de klas (Andrews & Barnes, 1990; Dwyer, 1994, 1998). Een consequentie hiervan is dat competenties uitgesplitst zouden moeten worden tot op het niveau van deelcompetenties in bepaalde beroepsspecifieke situaties. Dit laatste is gedaan in de Deelaspectenscan, terwijl bij de scan Algemene Beheersing competenties meer situatie-overstijgend worden vastgesteld. De resultaten ondersteunen voor een deel de contextspecifieke stellingname.

Niet overal waar deelcompetenties vooraf werden onderscheiden, konden deze ook empirisch worden onderscheiden. Zo bleken *beeldvorming* en *ondersteuning bij begeleiden van leerprocessen* empirisch niet sterk te onderscheiden. Mogelijk lopen de vermeende deelprocessen door elkaar heen en zijn ze nauwelijks te scheiden. De deelcompetenties van gespreksvoering waren empirisch wel te onderscheiden. Zo bleken het *begeleiden van discussie-inhoud* en het *begeleiden van de gespreksdynamiek* duidelijk onderscheidbaar. Ook het begeleiden van vormtechnische aspecten van gesprekken kon als afzonderlijke deelcompetentie van gespreksbegeleiding worden gezien.

Verder waren er aanwijzingen dat (zelf gerapporteerde) bekwaamheid in de ene beroepssituatie niet automatisch voorspellend is voor bekwaamheid in een andere context, ook al wordt er gevraagd naar een op het oog vergelijkbaar handelingsrepertoire. Dit bleek het geval te zijn voor het begeleiden van leerprocessen in de context van afstudeeropdrachten dat empirisch wordt onderscheiden van het begeleiden in de context van groepsonderwijs. Mogelijk stelt afstudeerbegeleiding andere eisen aan de docent dan groepsbegeleiding.

Bij al deze gerapporteerde samenhangen moet opgemerkt worden dat het gaat om scores die uitsluitend zelfpercepties van docenten betreffen.

6.3. Perspectieven op het gebruik van het instrument

In het algemeen passen vormen van zelf-evaluatie bij de huidige opvattingen over

leren en onderwijzen, waarin de nadruk steeds meer wordt gelegd op actief en zelfge-stuurd leren in (sociaal-) constructivistische leeromgevingen (Segers, Dochy, & Cascallar, 2003). In dergelijke leeromgevingen maakt zelfevaluatie integraal deel uit van het leerproces en is daarmee formatief van aard. Zelfbeoordeling vergroot de verantwoordelijkheid voor het eigen leren, draagt bij aan *ownership* van het leerproces, vertrouwen in de kwaliteit van de beoordeling en instemming met de standaarden (Gipps, 1999). Deze opvatting geldt niet alleen voor het leren van studenten, maar ook voor de professionele ontwikkeling van docenten. Het benadrukt dat docenten verantwoordelijk zijn voor het eigen leren. Van deze professionals wordt verwacht dat zij blijvend leren, zowel uit eigen behoefte als onder invloed van ontwikkelingen uit de omgeving (vergelijk Verloop, 2003). De verwachte bijdrage van een instrument als EduProf aan de professionele ontwikkeling van docenten ligt in het op gang brengen van reflectie en het stimuleren tot het ondernemen van leeractiviteiten. In de literatuur wordt dit ook wel aangeduid met de term consequentiële validiteit: in hoeverre roept een assessment gewenste gevolgen op (Darling-Hammond & Snyder, 2000; Messick, 1995)?

Uit onderzoek blijkt dat docenten leren van het ondergaan van een beoordeling zelf: men wordt aangezet tot nadenken over het eigen functioneren, ook over aspecten waarover men nooit heeft nagedacht of die meestal impliciet blijven, waardoor de beoordeling voor de docent een leerzame ervaring kan zijn (Boud, 1995). De meerwaarde van deze leerervaring boven andere vormen van professionalisering ligt in het systematische karakter van de feedback op het functioneren (Duke & Stiggins, 1990). Met name zelfevaluatie is echter een eenzame aangelegenheid wanneer niet de stap wordt gezet naar interactie met bijvoorbeeld collega's of een leidinggevende over het resultaat van de zelfevaluatie. Deze stap wordt niet automatisch gezet. Daarvoor moeten condities gecreëerd worden door de leiding van een onderwijsinstelling, zoals het beschikbaar stellen van faciliteiten. Van de zijde van de docenten is een zekere openheid gewenst om die inter-

actie aan te gaan en om te kunnen gaan met feedback. Bij vervulling van deze condities kan zelfevaluatie een krachtig middel zijn om professionele ontwikkeling te bevorderen (Boud, 1999). Een belangrijk neveneffect daarvan is dat collega's aan allerlei aspecten van hun beroep een zelfde betekenis gaan toekennen en daarvoor een zelfde taal gaan gebruiken. Zelfevaluatie en interactie over de resultaten ervan kunnen bovendien een positief effect hebben op de onderwijs- en werkcultuur van de onderwijsinstelling. Het lijkt belangrijk om verder onderzoek te doen naar de effecten van zelfevaluatie op de professionele ontwikkeling van docenten en de cultuur van de school. Op termijn hopen we dit onderzoek te kunnen doen met onder andere het in dit artikel besproken instrument EduProf.

Noot

- 1 Met dank aan Koosje Brants, Cees Groot, Dato de Gruijter en Jan Nedermeijer (allen ICLON), en Jonne van Diggele en Piet Sanders (beiden Cito) die bij de beginfase van de instrumentontwikkeling betrokken waren. Ook dank aan Brenda Zandsteeg, voor haar inzet bij de dataverzameling ten behoeve van de vervolgstudie.

Literatuur

- Andrews, T. E., & Barnes, S. (1990). Assessment of teaching. In W.R.Houston (Ed.), *Handbook of research on teacher education* (pp. 569-598). New York: Macmillan.
- Barber, L. W. (1990). Self-assessment. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary teachers* (pp. 216-229). Newbury Park, CA: Sage Publications.
- Beijaard, D., & Uhlenbeck, A. M. (2001). Het formuleren van docentcompetenties: Werkwijze en resultaten. *VELON Tijdschrift voor Lerarenopleiders*, 22(2), 26-33.
- Boud, D. (1995). *Enhancing learning through self-assessment*. London: Kogan Page.
- Boud, D. (1999). Avoiding the traps: Seeking good practice in the use of self assessment and reflection in professional courses. *Social*

- Work Education*, 18, 121-132.
- Boyd, R. T. C. (1989). Improving teacher evaluations. *ERIC Digest No. 111. ERIC Identifier ED315431*. ERIC Document Reproduction Service.
- Danielson, C. (1996). *Enhancing professional practice. A framework for teaching*. Alexandria, VA: ASCD.
- Danielson, C., & McGreal, T.L. (2000). *Teacher evaluation to enhance professional practice*. Alexandria, VA,: ASCD.
- Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teachers in context. *Teaching and Teacher Education*, 16, 523-545.
- Duke, D. L., & Stiggins, R. J. (1990). Beyond minimum competence: Evaluation for professional development. In J. Millman & L. Darling-Hammond (Eds.), *The new handbook of teacher evaluation: Assessing elementary and secondary teachers* (pp. 116-132). Newbury Park, CA: Sage Publications.
- Dwyer, C. A. (1994). Criteria for performance-based teacher assessments: Validity, standards, and issues. *Journal of Personnel Evaluation in Education*, 8, 135-150.
- Dwyer, C. A. (1998). Psychometrics of Praxis III: Classroom performance assessments. *Journal of Personnel Evaluation in Education*, 12, 163-187.
- Feldt, L. S. (1965). The approximate sampling distribution of Kuder-Richardson reliability coefficient twenty. *Psychometrika*, 30, 357-370.
- Frederiksen, J. R., Sipusic, M., Sherin, M., & Wolfe, E. W. (1998). Video portfolio assessment of teaching. *Educational Assessment*, 5, 225-298.
- Gipps, C. V. (1999). Socio-cultural aspects of assessment. In P.D. Pearson & A. Iran-Nejad (Eds.), *Review of Research in Education 24* (pp. 355-392). Washington, DC: American Educational Research Association.
- HBO-raad. (2006). *Percentage personeelsleden per hogeschool en leeftijdscategorie*. Geraadpleegd op <http://www.hbo-raad.nl/?id=137&t=kenget>.
- Howard, B. B., & McColskey, W. H. (2001). Evaluating experienced teachers. *Educational Leadership*, 58(5), 48-51.
- Kane, M. T. (1992). An argument-based approach to validity. *Psychological Bulletin*, 112, 527-535.
- Kane, M. T. (2004). Certification testing as an illustration of argument-based validation. *Measurement: Interdisciplinary Research and Perspectives*, 2, 135-170.
- Messick, S. (1995). Standards of validity and the validity of standard in performance assessment. *Educational Measurement: Issues and Practice*, 14(4), 5-8.
- Ministerie van Onderwijs, Cultuur en Wetenschappen. (2007). *Nota werken in het onderwijs 2007*. Den Haag, Nederland: Delta Hage.
- Nederlands-Vlaamse Accreditatieorganisatie (2003). *Accreditatiekader bestaande opleidingen hoger onderwijs*. Opgehaald op 1 oktober 2007, van <http://nvao.net/accreditatiekaders-nederland>.
- Newble, D. I., Hoare, J., & Baxter, A. (1982). Patient management problems: Issues of validity. *Medical Education*, 16, 137-142.
- Nijveldt, M.J. , & Roelofs, E.C. (2003). *Het beoordelen van docentcompetenties in het hoger beroepsonderwijs. Evaluatie van een instrument voor zelfbeoordeling*. OPD Memorandum. Arnhem, Nederland: Cito.
- Onderwijsraad. (2006). *Waardering voor het leerraarschap*. Den Haag, Nederland: Onderwijsraad.
- Onderwijsraad. (2007). *Kwaliteit belonen in het hoger onderwijs?* Den Haag, Nederland: Onderwijsraad.
- Ostrander, L. (1996, april). *Multiple judges of teacher effectiveness: Comparing teacher self-assessments with the perceptions of principals, students, and parents*. Paper gepresenteerd op de jaarlijkse bijeenkomst van de American Educational Research Association , New York.
- Reynolds, M. (1992). What is competent beginning teaching? A review of the literature. *Review of Educational Research*, 62, 1-35.
- Roelofs, E. C., & Beijaard, D. (2004). *EduProf, competentiescan voor HBO docenten: Handleiding*. Amsterdam: Humatix.
- Roelofs, E.C. , & Sanders, P. F. (2003). Beoordeling van docentcompetenties. In M. Mulder, R. Wesseling, H. Biemans, L. Nieuwenhuis, & R. Poell (red.), *Competentiegericht onderwijs. Gediplomeerd, maar ook bekwaam?* (pp. 277-299). Groningen, Nederland: Wolters-Noordhoff.
- Roelofs, E. C., & Sanders, P. F. (2007). Towards a framework for assessing teacher competence.

European Journal of Vocational Training, 40, 123-139

- Roelofs, E. C., Diggele, J. van, Sanders, P. F., Groot, K., Nedermeijer, J., & Beijaard, D. (2002). *Stuurdokument 'De ontwikkeling van beoordelingsinstrumenten voor docentcompetenties'*. Intern rapport. Arnhem, Nederland: Cito.
- Roelofs, E. C., Franssen, H. A. M., Houtveen, A. A. M. & Lagerweij, N. A. J. (1999) Een dieptestudie naar authentiek leren in de basisvorming: docentgedrag, methodegebruik, en leerlingpercepties. *Pedagogische Studiën*, 76, 258-272.
- Schaaf, M. van der. (2005). *Construct validation of teacher portfolio assessment*. Dissertatie. Universiteit Utrecht, Utrecht, Nederland.
- Scriven, M., (1995). A unified theory approach to teacher evaluation. *Studies in Educational Evaluation*, 21, 111-129.
- Segers, M., Dochy, F., & Cascallar, E. (Eds.) (2003). *Optimising new modes of assessment: In search of qualities and standards*. Dordrecht, Nederland: Kluwer Academic Publishers.
- Shuell, T. J. (1993). Toward an integrated theory of teaching and learning. *Educational Psychologist*, 28, 291-311.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations for the new reform. *Harvard Educational Review*, 57, 1-22.
- Sternberg, R. J., & Horvath, J. A. (1995). A prototype view of expert teaching. *Educational Researcher*, 24(6), 9-17.
- Stichting Beroepskwaliteit leraren en ander personeel. (2003). *Bekwaamheidseisen leraren*. Den Haag, Nederland: SBL.
- Stodolsky, S. S. (1990). Classroom observation. In J., Darling-Hammond & L. Millmann (Eds.), *The new handbook of teacher evaluation. Assessing elementary and secondary school teachers*. (pp. 175-190). Newbury Park, CA: California: Corwin Press.
- Straetmans, J. J. M., & Sanders, P. F. (2001). *Beoordelen van competenties van docenten*. Utrecht, Nederland: Programmamanagement Educatief Partnerschap.
- Tigelaar, D. (2005). *Design and evaluation of a teaching portfolio*. Dissertatie. Universiteit Maastricht, Maastricht, Nederland.
- Uhlenbeck, A. (2002). *The development of an assessment procedure for beginning teachers of English as a foreign language*. Dissertatie. Universiteit Leiden, Lediën, Nederland.
- Uhlenbeck, A. M., Verloop, N., & Beijaard, D. (2002). Requirements for an assessment procedure for beginning teachers: Implications from recent theories on teaching and assessment. *Teachers College Record*, 104, 242-272.
- Verhelst, N. D. (2000). *Estimating the reliability of a test from a single test administration*. Arnhem, Nederland: National Institute for Educational Measurement.
- Verloop, N. (2003). De leraar. In N. Verloop & J. Lowyck (red.), *Onderwijskunde* (pp. 195-248). Groningen, Nederland: Wolters-Noordhoff.
- Verloop, N., Beijaard, D., & Driel, J. H. van. (1998). Beoordeling van docenten. *Pedagogische Studiën*, 75, 171-187.
- Vermunt, J. D., & Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and Instruction*, 9, 257-280.
- Vos, H. J., & Knuver, J. W. M. (2000). Standaarden in onderwijskunde. In R. J. Bosker (red.), *Onderwijskundig Lexicon (Editie III). Evalueren in het onderwijs*. (pp. 59-76). Alphen aan den Rijn, Nederland: Samson.
- Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan*, 70, 703-713.
- Zill, N. & Nolin, M. J. (1994, oktober). *School learning environments as reported by teachers, parents, and students*. Paper gepresenteerd op de bijeenkomst van Safe Schools, Safe Students: A Collaborative Approach to Achieving Safe, Disciplined and Drug-Free Schools Conducive to Learning, Washington, DC.

Manuscript aanvaard: 22 december 2007

Auteurs

Erik Roelofs is als senior toetsdeskundige werkzaam bij Cito in Arnhem.

Mirjam Nijveldt was als aio werkzaam en is in 2007 gepromoveerd bij het Interfacultair Centrum voor Lerarenopleiding, Onderwijsontwikkeling en Nascholing (ICLON) van de Universiteit Leiden.

Douwe Beijaard is als hoogleraar werkzaam bij de Eindhoven School of Education (ESoE) van de TU Eindhoven.

Correspondentieadres: E. Roelofs, Cito, Postbus 1034, 6801 MG Arnhem. e-mail: Erik.roelofs@cito.nl.

Abstract

Development of an instrument for self-assessment of teacher competences

This study focuses on the development and testing of a web-based instrument for self-assessment by teachers in higher education. Scientific information about these teachers' work was used as a basis for the development of the instrument. It consists of 18 domains of competence which together represent five relevant working domains of teachers. Teachers who use the instrument, receive feedback in the form of a score profile accompanied by a written explanation. In a pilot study and a follow-up study, teachers responded to the instrument, either in written form or web-based. For this, the 18 domains of competences were elaborated into 18 scales (in total, there were 182 items). The results show that the domains of competences could be assessed reliably. This applies also to two, additional, more refined scales that differentiate according to specific situations of the profession. Teachers, who were interviewed about the instrument, recognized the content of the instrument as representative of their work as teachers in higher education. The self-assessment instrument encourages also reflection by the teachers on their work, which may support their further professional development.

Bijlage 1

Alle items voor de competentie begeleiden van leerprocessen van de scan Algemene Beheersing

1. Ik bespreek met studenten op welke wijze ze hun opdrachten plannen.
 2. Ik bespreek met studenten welke strategieën ze hanteren om relevante literatuur te bestuderen.
 3. Tijdens besprekingen ga ik na op welke punten studenten problemen hebben met het bestuderen van de literatuur.
 4. Bij de bespreking van opdrachten ga ik actief na welke problemen studenten hebben met het adequaat uitvoeren van de opdrachten.
 5. Tijdens besprekingen van opdrachten maak ik studenten bewust van aanwezige leemtes in hun kennisbestand.
 6. Ik maak samen met de student afspraken over het inleveren van producten en tussenproducten behorend bij langlopende leertaken.
 7. Ik stem de omvang van mijn hulp aan studenten af op de voortgang van een langlopende leertaak.
 8. Ik maak studenten duidelijk op welke punten de gevolgde aanpak om tot een product te komen voldoet of tekort schiet.
 9. Ik geef zodanig commentaar op opdrachten die studenten uitvoeren, dat ze zichtbaar verder kunnen met hun opdrachten.
 10. Als studenten problemen hebben bij het bestuderen van de leerstof, geef ik hen zodanige handreikingen dat ze blijkens hun reacties adequater verder kunnen.
 11. Tijdens een discussie of uitwisseling herformuleer ik de standpunten van studenten met behulp van vaktermen die gehanteerd worden in de beroepspraktijk.
 12. Bij het geven van commentaar op opdrachten waarin studenten beroepsvaardigheden oefenen, trek ik conclusies aan de hand van een professioneel (handelings)model (bijv. toonaangevende behandelmethoden of technieken voor metaalbewerking).
 13. Tijdens besprekingen van taken in groepsverband draag ik er zorg voor dat de gevolgde aanpak door cursisten wordt vergeleken met oplossingswijzen zoals die door professionele experts worden toegepast.
-

Noot. De items 1 tot en met 13 vormen tezamen één schaal.

Bijlage 1

Alle items voor de competentie begeleiden van leerprocessen van de scan Algemene Beheersing

Items voor de deelcompetentie beeldvorming

1. In gesprek met de student luister ik nauwkeurig op welke wijze hij een afstudeeropdracht aanpakt
2. Ik stimuleer de student zelf de problemen te verwoorden die hij of zij tegenkomt bij de uitvoering van een afstudeeropdracht.
3. Ik bewaak dat de gekozen probleemstelling van een afstudeeropdracht relevant is voor de toekomstige beroepspraktijk.

Items voor de deelcompetentie ondersteuning

4. Ik laat de student zelf de gemaakte planning bewaken bij een afstudeeropdracht.
 5. Ik bespreek met de student de wijze waarop hij een afstudeeropdracht zelfstandig aanpakt.
 6. Ik maak de student duidelijk op welke punten het ingeleverde product voldoet of tekort schiet met betrekking tot vooraf vastgestelde inhoudelijke kwaliteitscriteria.
 7. Ik laat de student zelf nagaan in hoeverre hij de doelen van een afstudeeropdracht heeft bereikt.
-

Noot. De itemsets 1 tot en met 3 en 4 tot en met 7 vormen elk een afzonderlijke schaal.