

Dr. Gerard J.J.M. Straetmans
Prof.Dr. Piet F. Sanders

Beoordelen van competenties ***van docenten***

Deze brochure (oplage 250 exemplaren) wordt uitgegeven door het Landelijk Programmamanagement Educatief Partnerschap van de HBO-raad en wordt in een beperkt aantal gratis verspreid onder de lerarenopleidingen. U kunt deze brochure ook vinden op de EPS-website: www.educatiefpartnerschap.nl/publicaties

Voorwoord

Het lerarentekort en de noodzaak verkorte opleidingsroutes naar het lerarenberoep te ontwikkelen hebben de discussie over het beoordelen van leraren nieuw leven ingeblazen. Hoe kan de kwaliteit van de in-, door- en uitstroom in verschillende opleidingsroutes worden gewaarborgd? Met het oog hierop wordt er de afgelopen jaren in Nederland gewerkt aan de ontwikkeling van verschillende 'assessment'-instrumenten. Termen als 'assessment', 'competenties' en 'portfolio' zijn voor veel lerarenopleiders nieuw. Wat is voor hen het belang hiervan? Lerarenopleiders zijn gediend met antwoorden op vragen als:

- Wat is 'assessment' precies? Wat voor 'assessment'-systemen bestaan er in het buitenland? Welke ervaringen zijn daarmee opgedaan?
- Aan welke eisen moet de formulering van leraarscompetenties voldoen?
- Aan welke kwaliteitseisen moeten 'assessment'-instrumenten voldoen?
- Welke instrumenten zijn nu beschikbaar of in ontwikkeling?
- Kunnen lerarenopleiders de nu ontwikkelde 'assessment'-instrumenten zomaar gebruiken?

Wat zijn de sterke en zwakke kanten ervan?

Deze vragen zijn vanuit het EPS-project Flankerend onderzoek voorgelegd aan de auteurs van deze brochure. Dr. Gerard Straetmans en prof. dr. Piet Sanders zijn erin geslaagd voor een publiek van lerarenopleiders en beleidsmakers op heldere wijze uiteen te zetten wat er allemaal komt kijken bij het beoordelen van competenties van docenten. Met sprekende voorbeelden en schema's doen zij de complexiteit van de materie uit de doeken.

Straetmans en Sanders laten zien wat er nodig is voor kwaliteit bij het beoordelen van leraren en wat er aan kwaliteitsbevordering kan worden gedaan. Dit laatste hoeft niet per se ingewikkeld te zijn. Tegen de achtergrond van te stellen kwaliteitseisen beschrijven de auteurs een viertal instrumenten ter beoordeling van leraarscompetenties die nu in Nederland in ontwikkeling zijn. Zij evalueren verschillende kenmerken van deze instrumenten en doen suggesties voor het gebruik en de kwaliteitsverbetering ervan. Moge deze aflevering van de EPS-reeks menige professionele discussie rond het opleiden en beoordelen van leraren verhelderen!

Dr. C.N. Brouwer

Projectleiding Flankerend onderzoek EPS

Inhoud

VOORAF (binnenzijde omslag)	
VOORWOORD	pag. 2
INLEIDING	pag. 5
1 OPLEIDEN OP BASIS VAN COMPETENTIES	pag. 6
1.1 COMPETENTIES: IS ER IETS NIEUWS ONDER DE ZON?	pag. 7
1.2 COMPETENTIES: OP ZOEK NAAR EEN BRUIKBARE OMSCHRIJVING	pag. 8
1.3 COMPETENTIEGERICHTE LERARENOPLEIDING	pag. 10
2 BEOORDELEN VAN COMPETENTIES	pag. 12
2.1 FUNCTIES VAN COMPETENTIEBEOORDELING	pag. 12
2.2 METHODEN VOOR HET BEOORDELEN VAN COMPETENTIES	pag. 15
2.2.1 Methoden voor het verzamelen van bewijs voor competentie in een werksituatie	pag. 17
2.2.2 Beoordelen onder testcondities	pag. 23
2.3 VAN SCORES NAAR CONCLUSIES	pag. 32
2.3.1 Over doelprestaties en middelprestaties	pag. 32
2.3.2 Kwaliteitsaspecten van een assessment	pag. 35
2.4 DE BESTE METHODE IS EEN METHODEMIX.	pag. 41
2.4.1 Assessment Center	pag. 44
2.4.2 De portfolio methode	pag. 48
3 ASPECTEN EN KWALITEITSCRITERIA VAN COMPETENTIE-INSTRUMENTEN	pag. 54
3.1 ASPECTEN VAN COMPETENTIE-INSTRUMENTEN VOOR DOCENTEN	pag. 54
3.1.1 Wie wordt beoordeeld?	pag. 54
3.1.2 Waartoe wordt beoordeeld?	pag. 54

3.1.3	Wat wordt beoordeeld?	pag. 55
3.1.4	Hoe wordt beoordeeld?	pag. 57
3.1.5	Door wie wordt beoordeeld?	pag. 59
3.2	KWALITEITSCRITERIA VOOR COMPETENTIE-INSTRUMENTEN VOOR DOCENTEN	pag. 62
3.2.1	Psychometrische kwaliteitscriteria	pag. 63
3.2.2	Procedurele kwaliteitscriteria	pag. 64
4	BESCHRIJVING EN EVALUATIE VAN COMPETENTIE-INSTRUMENTEN	pag. 66
4.1	COMPETENTIE-INSTRUMENT 'BEWEZEN GESCHIKT'	pag. 66
4.1.1	Beschrijving van 'Bewezen Geschikt'	pag. 66
4.1.2	Evaluatie van 'Bewezen Geschikt'	pag. 72
4.2	COMPETENTIE-INSTRUMENT 'BEKWAAM BEOORDEELD'	pag. 75
4.2.1	Beschrijving van 'Bekwaam Beoordeeld'	pag. 75
4.2.2	Evaluatie van 'Bekwaam Beoordeeld'	pag. 80
4.3	COMPETENTIE-INSTRUMENT 'INTEGRATIEF BEOORDELINGSMOMENT'	pag. 82
4.3.1	Beschrijving van 'Integratief BeoordelingsMoment'	pag. 82
4.3.2	Evaluatie van 'Integratief BeoordelingsMoment'	pag. 86
4.4	COMPETENTIE-INSTRUMENT 'ALS DOCENT IN BEELD'	pag. 87
4.4.1	Beschrijving van 'Als Docent in Beeld'	pag. 87
4.4.2	Evaluatie van 'Als Docent in Beeld'	pag. 91
4.5	BESCHRIJVING VAN INSTRUMENTEN VOOR DE BEOORDELING VAN COMPETENTIES VAN PRAKTISERENDE DOCENTEN	pag. 92
	TOT SLOT	pag. 96
	INTERESSANTE WEBSITES	pag. 98
	LITERATUUR	pag. 99
	COLOFON	pag. 108

.....	pag. 55
.....	pag. 57
.....	pag. 59
.....	pag. 62
.....	pag. 63
.....	pag. 64
.....	pag. 66
.....	pag. 66
.....	pag. 66
.....	pag. 72
.....	pag. 75
.....	pag. 75
.....	pag. 80
.....	pag. 82
.....	pag. 82
.....	pag. 86
.....	pag. 87
.....	pag. 87
.....	pag. 91
.....	pag. 92
.....	pag. 96
.....	pag. 98
.....	pag. 99
.....	pag. 108

Inleiding

Competenties en competentiegericht opleiden staan momenteel in de schijnwerpers van onderwijskundig Nederland. Vaktijdschriften, studiedagen en seminars, ze staan vaak bol van de verhalen over hoe het competentiedenken de kwaliteit van het opleiden kan verbeteren. Wie zijn oor te luisteren legt, bespeurt nogal wat scepsis. 'Is dit het nu?'; 'Hebben we dit al niet eerder gehoord?'; of 'Laten we maar even rustig afwachten want straks komen ze toch weer met iets anders!'. Eerlijke reacties hierop kunnen hard aankomen: Nee, dit is het niet. Dé oplossing zal er trouwens nooit komen aangezien de maatschappij en dus het onderwijs aan voortdurende verandering onderhevig is. Ja, dit heeft u al eerder gehoord; het denken in competenties is niet nieuw. In de tachtiger jaren van de vorige eeuw is met name in het beroepsonderwijs veel te doen geweest over het formuleren van meetbare en voor alle betrokken partijen herkenbare uitkomsten van beroepsonderwijs. De huidige discussies over competentiegericht onderwijs gaan echter verder. Ze gaan over het loslaten van de gedachte dat er maar één of enkele manieren zijn om een beroeps-

kwalificatie te halen. Ze gaan over de wens om de nu nog vaak starre programmering in afzonderlijke leergebieden los te laten. En ze gaan over een opleidingsdidactiek die studenten meer verantwoordelijkheid geeft voor het eigen leren ten einde de extrinsieke motivatie (juist zoveel doen dat je de studiepunten binnenhaalt) om te buigen in een meer intrinsiek gerichte motivatie.

Deze brochure is geheel gewijd aan het beoordelen van competenties. In hoofdstuk 1 gaan we terug naar de 'roots' van het competentiedenken; we wagen een poging tot verheldering van het begrip en wijden enkele beschouwingen aan de implementatie ervan in de lerarenopleiding. In hoofdstuk 2 wordt de focus gericht op functies van competentiebeoordeling, op sterke en zwakke eigenschappen van diverse beoordelingsmethoden en op de noodzaak om te kiezen voor een methodemix. Hoofdstuk 3 is geheel gewijd aan een beschrijvingskader voor assessments. In hoofdstuk 4 tenslotte worden enkele, momenteel in de belangstelling staande assessments tegen het licht gehouden, gebruikmakend van het beschrijvingskader uit hoofdstuk 3.

I · Opleiden op basis van competenties

Een werkloze jongeman moet van Arbeidsvoorziening een baan zoeken. Op een van zijn dwaaltochten door de stad ziet hij dat de plaatselijke Mercedes dealer iemand zoekt om de auto's, die aankomen van de importeur, schoon te maken met het oog op de aflevering aan de klant. Hij meldt zich en krijgt de baan op voorwaarde dat hij een cursus 'autopoetsen' volgt. Zijn verweer dat iedereen auto's kan poetsen, wordt weggewuifd met het motief dat het hier niet om gewone auto's gaat maar om auto's van het merk Mercedes. De cursus blijkt meer om het lijf te hebben dan verwacht:

- theorie over de samenstelling van autolakken;
- theorie over de samenstelling van veelgebruikte kunststoffen in de autofabricage;
- theorie over de invloeden van schoonmaakmiddelen op lakken en kunststoffen;
- theorie over de samenstelling van autowas.

De cursus wordt afgesloten met een schriftelijke toets.

Overdreven? Ja en nee. Ja, omdat dit natuurlijk echt te gek voor woorden is. Nee, omdat er een gelijkenis is met nogal wat beroepsopleidingen waarvan het curriculum werd opgebouwd op basis van wat de zittende staf belangrijke kennis en vaardigheden vond ('Dit is niet zomaar kennis...'), in plaats van op een analyse van de taken waarmee beginnende beroepsbeoefenaren geconfronteerd worden. Competentiegericht opleiden behelst dat de opbouw van het curriculum wordt gebaseerd op de startbekwaamheden (competenties) van de beroepsbeoefenaar en ook dat beslissingen over voortgang en certificering daarop gebaseerd zijn.

Competentiegericht opleiden is niet zomaar uit de lucht komen vallen, zeker niet als thema binnen de lerarenopleidingen. De klachten over de kwaliteit van de leraren, de ernstige tekorten aan leraren op de arbeidsmarkt, de veranderde didactiek in alle vormen van onderwijs waarvoor leraren worden opgeleid, zijn enkele belangrijke redenen voor de populariteit van het onderwerp in kringen van lerarenopleiders.

Straks wijden we enkele woorden aan de vernieuwingen van de lerarenopleidingen in Nederland. Maar pas nadat we getracht hebben het lastige begrip competentie te omschrijven. We beginnen met een korte beschouwing over het ontstaan van competentiegericht opleiden, zo'n 40 jaar geleden in de Verenigde Staten.

1.1 Competenties: is er iets nieuws onder de zon?

Nogal wat mensen denken dat het competentiedenken iets is van de laatste tien jaar. Niets is minder waar. Weliswaar heeft de discussie de laatste tien jaar een enorme opleving doorgemaakt, feit is dat de wortels liggen in het begin van de jaren zestig van de vorige eeuw in de Verenigde Staten. Het was de tijd dat de Amerikanen hevig geschokt waren doordat zij op ruimtevaartgebied overvleugeld waren door wat denigrerend een 'totalitaire landbouwstaat' werd genoemd¹. Onmiddellijk werden vraagtekens gezet bij de kwaliteit van het onderwijs en maatregelen genomen. Voortaan moest in heldere en eenduidig geformuleerde doelstellingen worden vastgelegd waar onderwijs toe moest leiden. Ook de lerarenopleidingen moesten onder het mes en begin zeventiger jaren was het onderwijs in veel teacher colleges competency-based.

¹ In 1957 waren de Sovjets de eersten om een kunstmaan te lanceren en deze in een baan om de aarde te brengen.

Met technieken die ontleend waren aan behaviouristische denkbeelden op het gebied van leren en instructie werd het geven van onderwijs geanalyseerd in reeksen van goed-afgebakende taken. De ratio hiervan was dat men dacht, dat als leraren de geïdentificeerde en beschreven taken zouden beheersen, ze goede docenten zouden zijn. De teleurstelling was groot toen er geen duidelijke relatie bleek te bestaan tussen het correct kunnen uitvoeren van de honderden goed-afgebakende taken in de opleiding en het geven van goed onderwijs (Houston, 1974). De fout die men begaan heeft, is dat uitvoering van taken gelijk werd gesteld aan competentie. In die visie is iemand bijvoorbeeld een competent tv-reparateur als hij in staat is om de 25 meest voorkomende storingen bij televisietoestellen te verhelpen. Daarbij wordt voorbij gegaan aan het feit dat er storingen kunnen zijn die om een combinatie van handelingsroutines vragen of om wijzigingen daarvan. De competentie is dan een optelsom van taken die uitgevoerd kunnen worden, maar niet meer dan dat. Van deze wel erg beperkte kijk op competentie zijn we inmiddels verlost. Maar om nou te zeggen dat we nu precies weten wat we moeten verstaan onder competentie is toch wat veel gezegd, zoals uit de volgende paragraaf zal blijken.

1.2 Competenties: op zoek naar een bruikbare omschrijving

Een belangrijke fout van de vroege op behaviouristische leest geschoeide competentiebeweging is dat ze bij het analyseren van taken teveel bleef uitgaan van de bestaande curricula. De leerstof, ondergebracht in verschillende vakken, was het vertrekpunt in plaats van de competenties van de betreffende functie. Niet dat men eenzijdig gericht was op kennis; voor vaardigheden bestond ook toen al aandacht. Maar vaardigheid is geen synoniem voor competentie. Wij beseffen terdege dat we nu op glad ijs komen. Als we beweren dat vaardigheid en competentie andere dingen zijn, moet dit gestaafd kunnen worden aan de hand van definities. Nu zijn er voor beide begrippen net zoveel verschillende definities als deskundigen, zoals Bos (1998) constateerde in een door haar uitgevoerde literatuurstudie. Probleem met de meeste definities is dat de omschrijvingen op hun beurt begrippen bevatten die niet helder gedefinieerd zijn. Bijvoorbeeld: ‘..competentie is de bekwaamheid of het vermogen...’ (de Bie & Mostert, 2000). We zien er daarom van af nieuwe definities toe te voegen en volstaan met een beschrijving van de verschillen tussen vaardigheden en competenties.

Daarvoor gaan we uit van een gemakkelijk voorbeeld, dat waarschijnlijk onrecht doet aan de complexiteit van het vraagstuk, maar aan de hand waarvan het onderscheid tussen de twee begrippen duidelijk gemaakt kan worden. Bij het tennisspel speelt de opslag (service) een cruciale rol. Anders gezegd: serveren is een belangrijke competentie van een tennisspeler. Het kunnen uitvoeren van de complexe beweging is een vaardigheid die na langdurige oefening en voldoende feedback steeds beter beheerst zal worden maar niet zonder meer gelijk gesteld mag worden met competentie. In onze ogen is een competentie een complexe vaardigheid die:

1. wendbaar is;
2. in een relevante context wordt aangewend en
3. waarvan de uitvoering en/of het resulterende product voldoet aan een kwaliteitsstandaard.

WENDBAARHEID

Een vaardigheid is een reeks activiteiten die voor een groot deel automatisch verloopt. De toevoeging ‘voor een groot deel’ is erg belangrijk. Tenzij het om een zeer eenvoudige taak gaat, is op bepaalde momenten bewuste sturing van dit proces noodzakelijk wil de vaardigheid onder allerlei omstandigheden tot het gewenste resultaat leiden.

Om bij het voorbeeld te blijven: de wijze waarop de complexe serveerbeweging wordt uitgevoerd, dient af te hangen van omstandigheden als: eerste of tweede service; de positie van de tegenspeler; de ondergrond; de windkracht; de kwaliteit van de bal; enz. Het is belangrijk dat de tennisser de omstandigheden die moeten leiden tot aanpassing van de service herkent en daarop inspeelt met kleine aanpassingen in de beweging.

RELEVANTE CONTEXT

Een vaardigheid, opgevat als een reeks activiteiten die, achter elkaar uitgevoerd, tot een bepaald resultaat leidt, kan in principe in een 'contextarme' situatie verworven worden. Zolang de afmetingen en de ondergrond van de oefenplaats maar overeenkomen met die van een tennisbaan en er in het midden iets hangt waar de bal overheen moet gaan, is aan alle voorwaarden voldaan om de vaardigheid in te oefenen. Het is echter maar de vraag of de zo verworven vaardigheid ook in een relevante context met succes kan worden toegepast. Tijdens een tenniswedstrijd spelen nog andere factoren mee dan de puur technische, zoals: het vermogen tot concentratie, stressbestendigheid, vermoeidheid, e.d.

KWALITEITSSTANDAARD

Een vaardigheid is geen alles-of-niets situatie. Het gaat om een continue schaal waarop geen duidelijk beginpunt noch een duidelijk eindpunt, zijnde de ultieme vaardigheid, is neergelegd. Het kan immers altijd beter. Van beheersing kan dan ook niet gesproken worden zonder te refereren aan de context waarin de vaardigheid moet worden toegepast. Bij een tennisser die op C-niveau speelt worden minder eisen aan de vaardigheid gesteld dan bij een tennisser van het B-niveau.

Uit het bovenstaande moge duidelijk zijn geworden dat het bij competenties gaat om een integratie van vaardigheden, kennis, attitudes en andere persoonlijke eigenschappen. Deze opvatting vinden we ook terug bij Gonczi (1994), een Australisch onderwijskundige die zich al vele jaren sterk maakt voor competentiegericht opleiden van professionals. Gonczi is een aanhanger van de geïntegreerde (holistische) benadering van competentiegericht opleiden en beoordelen. In zijn visie is van competentie sprake als complexe combinaties van kennis, vaardigheden, attitudes en normen en waarden in een specifieke probleemsituatie worden

aangewend om het probleem te begrijpen en tot een oplossing te komen die voldoet aan de geldende kwaliteitsnormen. Met klem wijzen we op de zinsnede 'in een specifieke probleemsituatie'. Competentie wordt dus niet opgevat als een generieke bekwaamheid die in alle probleemsituaties tot succes zal leiden. Dat is geheel in lijn met de onderzoeksliteratuur over verschillen tussen beginners en experts, waaruit geconcludeerd kan worden dat expertise - zijnde een zeer hoog competentieniveau - domeinspecifiek is (Glaser, 1991).

1.3 Competentiegerichte lerarenopleiding

De geïntegreerde benadering van competenties leidt tot andere denkbeelden over opleiden, zoals moge blijken uit de volgende passage uit een artikel van Hager & Gillis (1995): '...The integrated approach acknowledges competence as a developmental process or continuum that promotes occupational growth. Furthermore, it is plausible to suggest that this approach to competency-based assessment overcomes all of the significant objections to the competency standards that have been identified in the literature. It allows for, among other things, such higher level considerations as the role of ethics and

values in competent performance, the need for reflective practice, the importance of diverse contexts and the fact that, in many cases, there is more than one way of performing competently.'

In Nederland is deze opvatting inmiddels gemeen goed geworden bij de lerarenopleiders. In Educatief Partnerschap, een vernieuwingsplan van de eerste- en tweedegraads lerarenopleidingen, wordt afstand gedaan van het modulair opgebouwde onderwijs waarin studenten bekwaam worden geacht als ze voldoende studiepunten hebben verzameld op afzonderlijke studie-onderdelen. Men heeft er geen vertrouwen meer in dat op deze manier leraren worden opgeleid die op een efficiënte en effectieve manier kennis, vaardigheden en houding geïntegreerd weten aan te wenden in complexe beroeps-situaties. Bij competentiegericht opleiden wordt het curriculum niet opgebouwd rondom de vaktheorie maar vanuit de concrete beroepspraktijk. Heel recent heeft dit geleid tot de ontwikkeling van een systematiek, het kijkkader voor competenties genoemd, om competenties te beschrijven die recht doen aan de hierboven geschetste integratieve, holistische benadering (Dietze, Jansma, & Riezebos, 2000). In het kijkkader worden 16 competenties onderscheiden als

resultaat van de kruising van vier contexten van beroepshandeling (werken met leerlingen; werken met team en in de organisatie; werken met/in de omgeving; werken aan eigen ontwikkeling) met vier beroepsrollen (pedagogisch handelen; vakdidactisch handelen; organisatorisch handelen; interpersoonlijk handelen). Het is aan de opleidingen om deze en eventueel andere, zelf geformuleerde competenties te concretiseren en te operationaliseren aan de hand van een specificatie van onderliggende aspecten van kennis, vaardigheden en houding.

De ommezwaai naar competenties als vertrekpunt voor de opbouw van het curriculum is niet de enige vernieuwing die wordt nagestreefd. Het besef dat leraren professionals zijn en dus in staat moeten zijn hun professionaliteit vorm te geven en te bewaken, moet al tijdens de opleiding ontstaan en verder worden uitgebouwd. Dat lukt niet in een curriculum dat van a tot z vastligt en van studenten slechts vraagt te consumeren. Leren leren, zelf sturing geven aan het leerproces en daar verantwoordelijkheid voor dragen, zijn elementen in een nieuwe didactiek die de basis legt voor het opleiden van gemotiveerde professionals die in staat zijn tot zelfreflectie en persoonlijke groei.

Het zal duidelijk zijn dat in een dergelijk curriculum geen plaats is voor een onderwijsconcept dat dezelfde leerstof op dezelfde manier en op hetzelfde moment aanreikt aan studenten. In de nieuwe didactiek is geen centrale rol meer weggelegd voor klassikaal onderwijs; nieuwe leerstof zal vooral door zelfstudie verworven worden. Een andere belangrijke verandering is dat het belang van het klassieke op kennis gerichte tentamen zal inboeten ten gunste van vormen van assessment die helderheid kunnen verschaffen over de verwerving van de competenties. Stagebeoordeling, zelf-assessment, vaardigheidstoetsen en portfolio zijn termen die nu regelmatig opduiken als het gaat om het beoordelen van competenties. Over de voor- en nadelen van deze methoden en de gevolgen daarvan voor het trekken van conclusies over competenties gaat het volgende hoofdstuk.

2 · Beoordelen van competenties

Competentiegericht opleiden vraagt niet alleen om een herziening van de op traditionele leest geschoeide curricula maar ook, en zeker niet in de laatste plaats, om een grondige revisie van het systeem van toetsing en afsluiting. Niet dat alle kennistoetsen in een klap in de ban gedaan moeten worden; die blijven nuttig aangezien elke competentie, zeker competenties van professionals zoals leraren, gebaseerd zijn op kennisbestanden. Wat uitgebannen moet worden, is de tot dusver tamelijk eenzijdige nadruk op kennistoetsen in veel beroepsopleidingen en daarmee de gedachte dat de studiepunten, die vaak de beloning vormen voor een voldoende resultaat op dergelijke toetsen, een garantie zouden bieden voor de kwaliteit van de latere beroepsuitoefening. Als competentiegericht opleiden geen modegril wil zijn die na enkele jaren bijgeschreven kan worden als de zoveelste mislukte onderwijsinnovatie, is meer aandacht voor het beoordelen van competenties absoluut noodzakelijk. Uit de onderwijskundige literatuur is bekend dat toetsen een sturende werking uitoefenen op het

daaraan voorafgaande onderwijs en dat het officiële leerplan soms al heel snel na de introductie van de innovatie moet wijken voor de onderwijsinhouden die gedictieerd worden door de afsluitende toetsen (zie bijvoorbeeld Frederiksen, 1984 en Madaus, 1988). Bij curriculumherzieningen dienen vraagstukken van toetsing en afsluiting van meet af aan in beschouwing te worden genomen.

2.1 Functies van competentiebeoordeling

Wij leven in een periode van snelle veranderingen. De tijd dat personen door het volgen van één enkele beroepsopleiding zich voor het leven kwalificeerden voor een beroep ligt definitief achter ons. Niemand kan zich straffeloos onttrekken aan de noodzaak van bijscholing. Dat geldt zeker voor docenten die, als het goed is, toch beschouwd dienen te worden als de kennismanagers van de natie en derhalve een cruciale rol spelen in een economie die zo afhankelijk is van kennis. Het beoordelen van competenties is daarom niet alleen van belang gedurende de initiële opleiding maar ook tijdens de periode van

actieve beroepsuitoefening, bijvoorbeeld om na te gaan of men nog voldoende is toegerust voor de eisen die het beroep stelt. Zo bezien kunnen we de volgende functies onderscheiden van het beoordelen van competenties:

- Nemen van instroombeslissingen;
- Nemen van voortgangsbepalingen;
- Nemen van certificeringbeslissingen;
- Auditering van de zittende staf met het oog op de professionele ontwikkeling van het individu en op de behoefte van de opleiding.

INSTROOMBESLISSINGEN

Instroombeslissingen worden belangrijker naarmate er, als gevolg van maatschappelijke ontwikkelingen, minder studenten zijn die de 'koninklijke weg' naar een bepaalde beroepsopleiding volgen. Zo zijn er deeltijdopleidingen voor reeds in het beroep actieve leraren die een bevoegdheid voor een ander vak willen halen of die een bevoegdheid voor een hoger onderwijstype willen. Steeds populairder worden de urgentieprogramma's voor middelbaar of hoger opgeleiden die in het onderwijs aan de slag willen. Hierover is momenteel veel te doen omdat de grote tekorten aan leraren in de verschillende onderwijssectoren de overheid gedwongen hebben om te zien

naar onconventionele maatregelen. Proberen om meer mensen te laten herintreden (vooral vrouwen die hun beroep weer willen oppakken na een periode waarin ze voor de kinderen gezorgd hebben en ouderleraren), is een tamelijk voor de hand liggende oplossing. Dat kan van de maatregel om de nood te lenigen door middel van zij-instroom in het beroep niet gezegd worden. Zij-instromers zijn mensen met een hbo- of wetenschappelijke opleiding die, na gebleken geschiktheid, aan de slag gaan in het onderwijs en daarnaast een (beperkte) opleiding volgen. Bij dergelijke opleidingen dient 'aan de poort' bepaald te worden of de aanmelder wel geschikt is voor het beroep en wat de meest effectieve en efficiënte route (denk aan vrijstellingen) is op weg naar de bevoegdheid voor elke individuele student. Het is niet moeilijk in te zien dat competentiegerichte curricula hier een meerwaarde hebben boven traditioneel ingerichte curricula. Het maakt niet uit waar je de kennis, vaardigheid en ervaring hebt opgedaan; als je kunt aantonen bepaalde competenties te bezitten, krijg je vrijstellingen. Instroombeslissingen zijn niet het exclusieve terrein van deeltijdopleidingen. Ook in voltijdopleidingen, bij studenten die de koninklijke weg bewandeld hebben, kan het nuttig zijn om kort na het

instroommoment, zoals bijvoorbeeld aan het eind van de propedeutische fase, de geschiktheid voor het beroep vast te stellen. Het kan de student helpen om frustratie en verspilling van dure opleidingstijd te voorkomen en de opleiding om korting op de financiering te voorkomen doordat het interne rendement terugloopt.

VOORTGANGSBESLISSINGEN

Onder voortgangsbeslissingen verstaan we hier alle beslissingen die genomen worden om elke deelnemer optimaal te begeleiden bij het verwerven van de benodigde competenties. Dit vereist dat regelmatig vastgesteld wordt of voorwaardelijke kennis aanwezig is en in hoeverre competenties of onderdelen daarvan al beheerst worden.

CERTIFICERINGSBESLISSINGEN

Anders dan in traditionele curricula is er in competentiegerichte opleidingen een minder strikte scheiding tussen voortgangs- en certificeringsbeslissingen. Het afsluitende examen als ultiem beslissingsmoment past niet in een opleidingsmodel dat gebaseerd is op competenties aangezien zo'n examen de kenmerken heeft van een momentopname en, zoals we hierna nog zullen zien, daardoor

nooit in voldoende mate representatief kan zijn voor de competenties in kwestie. Of iemand het diploma uitgereikt mag krijgen, is een beslissing waarvoor zwaarwegend en omvangrijk bewijs geleverd moet worden. Dat bewijs mag gerust gedurende de hele opleiding verzameld worden mits maar duidelijk is wat de status ervan is en er heldere procedures zijn om uit alle verzamelde bewijs een eenduidige conclusie te trekken.

AUDITERING

Tenslotte speelt het beoordelen van competenties een cruciale rol in het human resources concept 'competentiemanagement'. Net als bedrijven in de commerciële sector moeten schoolorganisaties alle zeilen bijzetten om deskundig personeel te werven en te behouden. Competentiemanagement is dat onderdeel van het personeelsbeleid waarbij gedragsgebonden hoedanigheden van arbeidskrachten als object van analyse en geïntegreerde sturing worden gebruikt (Thijssen, 2000). De resultaten worden gebruikt op de deelgebieden werving en selectie, inzetbaarheid (employability), functiebeoordeling en prestatiebeloning. Op al die deelgebieden spelen competenties een belangrijke rol, hoewel steeds vanuit een andere invalshoek.

- Bij werving en selectie gaat de aandacht vooral uit naar meer algemene, maar desondanks ontwikkelbare, eigenschappen van personen die men van belang acht voor de functie-uitoefening. Een school zal bijvoorbeeld graag willen beschikken over docenten met een hoge mate van flexibiliteit.
- Op het deelgebied inzetbaarheid is de focus gericht op het handelingsrepertoire waarover de diverse functionarissen dienen te beschikken. Zo luidt een van de startbekwaamheden voor leraar secundair onderwijs dat hij moet kunnen handelen in onverwachte situaties. Dit betekent dat hij een onderwijssituatie of begeleidingsactiviteit ter plekke improviserend moet kunnen voorbereiden, uitvoeren en evalueren.
- Bij een functiebeoordeling ligt de nadruk op concreet vertoond gedrag in relevante werksituaties. De leidinggevende van een docent zou bijvoorbeeld bij leerlingen kunnen navragen of de docent een zieke collega goed vervangen heeft.
- Bij prestatiebeloning tenslotte, wordt niet in de eerste plaats gelet op het proces maar op het resulterende product. Dat maakt vergelijking met een 'hard' criterium vaak mogelijk, wat vervolgens mogelijkheden biedt voor belonings-

differentiatie. Bijvoorbeeld: docenten die vaak en naar tevredenheid invallen voor collega's verdienen extra dagen voor een sabbatical.

Uit deze opsomming moge blijken dat het beoordelen van competenties diverse doelen kan dienen. Het doel is bepalend voor de wijze waarop competentie wordt vastgesteld. Daarvoor zijn verschillende methoden beschikbaar die in de volgende paragraaf uitvoerig aan de orde komen.

2.2 Methoden voor het beoordelen van competenties

Competenties hebben te maken met het kunnen uitvoeren van taken die in een of andere vervolgsituatie (werk, opleiding of andere maatschappelijke situatie) van belang zijn. Het gaat daarbij bijna altijd om taakstellingen in een relevante context waarin kennis, vaardigheden en attitudes gecombineerd worden tot handelingen die tot het gewenste resultaat moeten leiden. Gelet op deze formulering zal het voor iedereen duidelijk zijn dat kennistoetsen niet de eerste keuze moeten zijn als het gaat om het vaststellen van competentie. In plaats daarvan zijn we op zoek naar instrumenten die op een veel directere

wijze kunnen 'meten' waarin we geïnteresseerd zijn. In de meer populaire onderwijskundige literatuur wordt die kwaliteit vaak aangeduid met het begrip 'authenticiteit'. De authenticiteit van een instrument is gradueel en wordt bepaald door de authenticiteit van: a) de aangeboden taak of taken, b) de wijze waarop gereageerd kan worden op die taken en c) de context waarin de taakuitvoering plaatsvindt. Grant Wiggins, een Amerikaans voorvechter van grotere authenticiteit bij het vaststellen van prestaties van leerlingen en studenten, noemt onder andere de volgende kenmerken van authentieke prestatiebeoordeling (Wiggins, 1993):

- Uitdagende en relevante taken. De taken zijn replica's van of afgeleid van het soort problemen waarmee beroepsbeoefenaren in hun praktijk geconfronteerd worden.
- Natuurgetrouwe context waarbinnen de taken worden uitgevoerd. Dit betekent in feite dat een groot deel van de maatregelen die normaal getroffen worden om toetsing zo efficiënt mogelijk te laten verlopen, komen te vervallen of op zijn minst heroverwogen moeten worden. Denk daarbij aan de extreme pogingen tot geheimhouding van toetsmateriaal en aan de beperkingen die worden opgelegd aan toetstijd en -plaats en

aan de toegestane hulpmiddelen.

- Het gaat om taken die niet routinematig uitgevoerd kunnen worden. Echte problemen in de beroepspraktijk zijn zelden 'op de automatische piloot' op te lossen maar doen een groot beroep op het kennis- en vaardigheidsrepertoire en op het beoordelingsvermogen van de taakuitvoerder als het gaat om de vraag welke kennis en vaardigheden wanneer ingezet moeten worden.
- De opdrachten zelf alsmede de criteria en standaarden die gehanteerd worden bij de beoordeling van de prestaties moeten helder zijn voor alle betrokkenen. Met het oog daarop mogen taken voorbereid worden en eventueel met de assessoren worden doorgesproken.
- De beoordeling gebeurt door een getrainde assessor. Het werken met authentieke taken maakt het voorspellen van de taakuitvoering lastig. Het gevolg is dat de prestatiecriteria nooit volledig kunnen zijn afgestemd op alle mogelijke taakuitvoeringen en derhalve geïnterpreteerd zullen moeten worden. Door assessoren te trainen kan geprobeerd worden deze interpretaties in dezelfde richting te laten gaan.

Er zijn heel veel verschillende manieren denkbaar waarop informatie verzameld kan worden met het doel een uitspraak te doen over het al dan niet competent zijn van een persoon. Bijna elke auteur die over dit onderwerp schrijft, waagt een poging om de lezer te informeren over de vele keuzemogelijkheden die er zijn. De volgende methoden worden door bijna elke auteur genoemd:

- observatie in de werksituatie;
- interview;
- vaardigheidstoets;
- kennistoets;
- zelfbeoordeling;
- stageverslag;
- portfolio.

Deze lijst is gemakkelijk uit te breiden, want het aantal manieren waarop informatie verzameld kan worden die kan gelden als bewijs voor competentie wordt slechts beperkt door de creativiteit. Vaak verschillen die methoden niet echt van elkaar. In wezen is er maar één aspect waarop beoordelingsmethoden essentieel van elkaar verschillen en dat is de setting waarin het bewijs voor competentie verzameld wordt. Dat kan de werksituatie zijn of een testsituatie.

2.2.1 Methoden voor het verzamelen van bewijs voor competentie in een werksituatie

Als je wil weten of iemand in staat is lekkere pannenkoeken te bakken, dan is de beste manier om dat vast te stellen een keer bij die persoon op bezoek te gaan en bij hem pannenkoeken te gaan eten. ‘The proof of the pudding is in the eating’, zeggen de Engelsen. Wil je weten of iemand goed kan lesgeven dan ligt het voor de hand om die persoon te gaan observeren tijdens zijn lesgeven. We zullen deze methode nu verder aanduiden als observatie in de werksituatie. Observeren in de werksituatie kost veel tijd en daarom wordt vaak uitgeweken naar een methode die efficiënter is en die we hier retrospectie zullen noemen. Van beide methoden worden de voor- en nadelen besproken.

OBSERVATIE IN DE WERKSITUATIE

De auteurs van deze brochure herinneren het zich nog heel goed: eens in de zoveel tijd gebeurde er in de klas iets onverwachts. De deur ging open en daar verscheen het hoofd der school met in zijn kielzog een meneer (het was toen nooit een mevrouw) van wie je meteen wist dat die nog belangrijker moest zijn dan eerstgenoemde.

De deftige meneer beende, een aktetas stevig onder zijn arm geklemd, naar een stoel achter in de klas die daar speciaal voor dit soort gelegenheden stond. Het hoofd der school spoedde zich naar de inmiddels wat wit weg getrokken leerkracht met de boodschap dat hij/zij door moest gaan met de les en dat meneer de inspecteur gewoon even wilde kijken. De inspecteur pakte een aantekenschrift, stak een sigaartje op en schreef zo nu en dan iets op. Een enkele keer stelde hij wel eens een vraag, meestal aan de leerkracht maar ook wel eens aan een leerling. Veel langer dan een half uur bleef hij nooit.

Er is niets mis met observeren in de werksituatie; de te beoordelen competenties worden in hun volle complexiteit geëtaleerd in een natuurlijke context en de beoordelaar verzamelt precies de informatie waarnaar de interesse uitgaat. Helaas geldt dat niet voor alle beroepen want sommige beroepen, vooral die welke wij tegenwoordig aanduiden als professionele beroepen, hebben een dermate breed competentieprofiel dat het erg veel tijd gaat kosten voordat men op basis van observaties een conclusie kan trekken over de competenties van dergelijke beroepsbeoefenaren. Omdat die tijd er meestal niet is, loopt men het risico dat er verkeerde

conclusies getrokken worden op grond van een te kleine, niet-representatieve steekproef van uitgevoerde taken. Een ander probleem in dit verband heeft betrekking op de scoring. Beoordelaars van competenties zijn meestal erkend expert in het betreffende beroep en weten achteraf best te zeggen of ze, om een toch nog steeds belangrijk aspect van het leraarsberoep maar weer als voorbeeld te nemen, een goede les gezien hebben. Maar zo'n uitspraak moet natuurlijk gemotiveerd kunnen worden door datgene wat geobserveerd is te staven aan expliciet geformuleerde prestatiecriteria. En hier speelt de complexiteit van de reële werksituatie ons parten. Het blijkt erg moeilijk om een gestandaardiseerd scoreformulier te ontwikkelen dat geschikt is om vertoond gedrag in een reële werksituatie te scoren. Als de beoordelingsaspecten op het formulier te ruim geformuleerd zijn, laten ze erg veel ruimte voor interpretatieverschillen. Zijn ze minutieus geformuleerd dan worden de lijsten zo uitgebreid dat ze niet meer hanteerbaar zijn en dat men door de veelheid van beoordelingsaspecten het zicht op de competentie verliest (het geheel is meer dan de som der delen). Hier wordt de prijs betaald voor de keuze om de competentie in zijn volle complexiteit te beoordelen. Een positief geluid tussen al deze sombere beschou-

wingen is dat zowel beoordeelden als beoordelaars vaak een goed gevoel hebben bij deze methode. Beide partijen hebben het idee dat ze bezig zijn met datgene 'waar het om gaat'.

Samenvattend komen we tot de volgende belangrijkste voor- en nadelen:

Voordelen	Nadelen
Hoge acceptatiegraad bij zowel de beoordeelden als de beoordelaars	Praktisch niet of minder goed uitvoerbaar bij professionele competenties Risico van verkeerde conclusies als gevolg van een te gering aantal uitgevoerde taken Problematische scoring

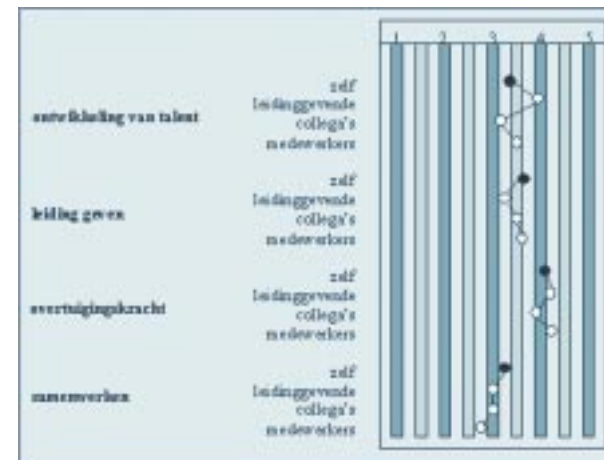
RETROSPECTIE

Een van de belangrijkste tekortkomingen van de proof-of-the-pudding-methode (observatie in de werksituatie) is dat vaak de tijd ontbreekt om personen gedurende langere tijd te observeren.

Het gevolg is meestal dat het geobserveerde niet-representatief is voor de competenties van de betreffende persoon. Soms wordt de oplossing gezocht in de ont koppeling van het observatiemoment en het tijdstip van scoring. Dan vraagt men de beoordelaar om, op een moment dat het hem of haar uitkomt, te reflecteren op het gedrag van de te beoordelen persoon en dat vervolgens te beoordelen aan de hand van een gestandaardiseerde lijst met beoordelingsaspecten. De beoordelaar bepaalt zelf op welke observaties hij zijn beoordelingen baseert. Dat roept een probleem op, namelijk dat we niet weten of de scoring op de beoordelingsaspecten wel gebaseerd is op relevant gedrag. Zo zal het bijvoorbeeld niet de bedoeling zijn dat een beoordelingsaspect als 'kan goed orde houden' gescoord wordt op grond van een zeer negatieve en daardoor in het geheugen gebrande ervaring tijdens een klassenuitstapje. Een groot voordeel van de retrospectiemethode is dat ze weinig voorbereiding vereist waardoor het organisatorisch eenvoudiger wordt om meerdere personen om een beoordeling te vragen.

In het hoger onderwijs wordt deze methode sinds jaar en dag gebruikt om de mening van studenten te peilen over de kwaliteit van verzorgde cursussen. Een steeds weer terugkomende klacht van docenten

is dat dergelijke meningen niet objectief zijn en verband zouden houden met behaalde tentamenresultaten. Van Os (1999) toonde aan dat hiervan geen sprake is en dat studenten in het hoger onderwijs een betrouwbaar, eerlijk en valide oordeel geven over de kwaliteit van het onderwijs in het algemeen en de docent in het bijzonder. Een belangrijk voordeel van het inschakelen van studenten is dat veel personen allemaal vanuit hun eigen beleving een oordeel geven over de competenties van een docent. Met name wanneer de studentoordelen worden aangevuld met de beoordelingen van relevante anderen zoals collega's, direct leidinggevenden en -niet te vergeten - de betreffende docent zelf, ontstaat een tamelijk compleet beeld van het functioneren dat, mits daarover van tevoren heldere afspraken zijn gemaakt, gebruikt kan worden voor doelen als employability en personeelsbeoordeling. Deze methode wordt in de literatuur aangeduid met de term 360-graden beoordeling. Figuur 1 laat een voorbeeld zien van een rapport van een 360-graden beoordeling op het gebied van managementvaardigheden.



Figuur 1. Fragment uit een rapportage naar aanleiding van een 360-graden feedback onderzoek. Verklaring schaal: 1 = onvoldoende; 2 = zwak; 3 = matig; 4 = voldoende; 5 = goed.
bron: PiMedia BV, Arnhem 1993.

Zelfbeoordeling geniet steeds meer belangstelling en daar zijn goede redenen voor. Het dwingt tot introspectie, tot het kritisch beschouwen van waar je staat en hoe je functioneert. Het is niet onredelijk om te veronderstellen dat de motivatie om gedrag te veranderen groter zal zijn als de persoon in kwestie zelf daartoe de noodzaak onderkent. Maar zijn mensen wel in staat om kritisch naar hun functioneren te kijken? Straetmans (2000) onderzocht dit bij cursisten uit het middelbaar beroepsonderwijs. Hij vroeg 31 cursisten om hun vaardigheid op het gebied van basishandelingen met het pakket Windows 95 te beoordelen. Elke cursist repondeerde daartoe op 28 zogeheten can-do statements. Bijvoorbeeld:

Kunt u de map 'Auto's' met alle daarin opgeslagen bestanden kopiëren van station A naar station C?

Ja, vlot en foutloos	Met wat proberen lukt dat wel	Nee, dat lukt me niet
----------------------	-------------------------------	-----------------------

Een week daarna moesten alle proefpersonen een vaardigheidstoets afleggen waarin dezelfde handelingen uitgevoerd moesten worden als die in de can-do statements aan de orde waren gesteld. De helft van de proefpersonen was bij de zelfbeoordeling al op de hoogte van de komende vaardigheidstoets, de andere helft niet.

Constateringen:

- Slechts 6 van de 31 proefpersonen bleken in staat tot een redelijk accurate zelfbeoordeling (dat wil zeggen een overeenstemming van 60% of hoger tussen de scores op zelfbeoordeling en de vaardigheidstoets);
- 17 proefpersonen overschatten hun vaardigheid (soms ernstig) en 1 onderschatte de eigen vaardigheid;
- Proefpersonen die van tevoren op de hoogte waren van de komende vaardigheidstoets waren minder geneigd hun vaardigheid ernstig te overschatten;
- Er was een positief en significant verband tussen de totaalscores op de zelfbeoordeling en die op de vaardigheidstoets. In combinatie met een bepaalde censuur leidde dit tot redelijk nauwkeurige beslissingen over beheersing op basis van zelfbeoordeling.

De conclusie luidde dat zelfbeoordelingen op itemniveau onvoldoende betrouwbaar zijn, maar dat op basis van de totaalscore op het zelfbeoordelingsinstrument accurate beslissingen over beheersing mogelijk waren.

Uit een onlangs uitgevoerd onderzoek is gebleken dat in Nederland ruim een kwart van 350 aangeschreven bedrijven (industrie, zakelijke dienstverlening, non-profit en handel) deze methode toepast (Jellema, 2000). De overgrote meerderheid daarvan (95 procent) was van mening dat 360-graden beoordeling een geschikt instrument vormt voor de personeelsontwikkeling. Ruim 80 procent vond de acceptatie door beoordelaars en beoordeelde goed tot zeer goed. Van Hooft (1999) deed onderzoek naar de kwaliteit van (een specifieke) 360-graden beoordelingsmethode. Hij concludeerde dat de beoordelingen over het algemeen betrouwbaar waren maar dat sommige beoordelingsaspecten zich beter lieten beoordelen dan andere. De betrouwbaarheid van de beoordelingen, gemeten als de overeenstemming tussen beoordelaars liet te wensen over als die beoordelaars uit verschillende groepen kwamen. Doordat elke groep vanuit zijn eigen plaats in en visie op de organisatie naar het functioneren kijkt, ontstaan er tussen deze groepen beoordelingsverschillen. Zelfbeoordelingen vielen relatief hoog uit (zie ook schema op pag. 21) en beoordelingen door leidinggevenden vertoonden veel spreiding. Al met al lijkt retrospectie als methode eenvoudiger

in te zetten dan observatie in de reële werksituatie, tenzij men zeer veel tijd en geld ter beschikking heeft. Een nadeel dat niet onbesproken mag blijven, heeft betrekking op het gedrag dat beoordeeld wordt met de retrospectie methode. De beoordelingen komen niet exclusief tot stand op basis van doelbewuste observaties maar vooral op basis van herinneringen aan toevallige waarnemingen van gedrag. Wat wordt vastgelegd in de beoordelingen is niet zozeer het maximale kunnen van iemand maar zijn typische gedrag. Het typische gedrag kan om allerlei redenen afwijken van het gedrag waartoe iemand echt in staat is.

Voordelen	Nadelen
Relatief eenvoudig uitvoerbaar	Onzekerheid of beoordelingen op het gewenste gedrag gebaseerd zijn
Grote aantallen beoordelaars vergroten de betrouwbaarheid van het gemiddelde oordeel.	Acceptatie door te beoordelen personen is soms een probleem
	Meet vooral typisch gedrag

2.2.2 Beoordelen onder testcondities

Wanneer competenties niet gedemonstreerd kunnen worden in de reële werksituatie blijft er weinig anders over dan de te beoordelen personen hun competenties te laten demonstreren in een nagebootste werksituatie. Dat kan overigens heel realistisch zijn. De meesten van ons staan er niet bij stil, maar tijdens het rijexamen, zoals afgenomen door het CBR, demonstreert een kandidaat zijn rijvaardigheid in een nagebootste werkelijkheid. Het verschil met de reële verkeerssituatie schuilt voornamelijk in de aanwezigheid van een examiner die de weg wijst en ingrijpt als het gevaarlijk wordt. Nabootsing kan verschillen in de mate waarin ze representatief zijn voor de reële werksituatie (Fitzpatrick & Morrison, 1971). Straetmans (1993) onderscheidde drie niveaus van representativiteit en koppelde daaraan instrumenten voor het beoordelen van competenties.

HOGE REPRESENTATIVITEIT: DE WORK SAMPLE TEST

Indien haalbaar gaat de voorkeur uit naar een instrument met een hoge representativiteit. Dergelijke instrumenten worden wel aangeduid met de

Engelse term 'hands-on'. Hands-on wil zeggen dat de te beoordelen personen in een realistische setting realistische taken uitvoeren en daarbij gebruik (kunnen) maken van alle hulpmiddelen die in een reële werksituatie ook gebruikt zouden kunnen worden. De grootste bekendheid geniet de work sample test, in Nederland ook wel arbeidsproef genoemd. Het onvertaald laten van de term heeft echter onze voorkeur omdat 'arbeidsproef' ten onrechte de indruk wekt niet ingezet te kunnen worden voor professionele beroepen.

Bij een work sample test moet een te beoordelen persoon, onder omstandigheden die de reële werksituatie zeer dicht benaderen, taken uitvoeren die ontleend zijn aan de toekomstige functie. Het bekendste voorbeeld is het rijexamen. Eveneens tot de verbeelding sprekend is de toets die artsen in opleiding moeten afleggen om te bewijzen dat ze competent zijn in het voeren van een anamnese-gesprek, het verrichten van klinisch onderzoek, het stellen van een diagnose en het voorstellen van een behandeling. Dichter bij huis voor de meeste lezers zijn de lessen die studenten van lerarenopleidingen tijdens hun stageperiodes moeten geven. Ook hier wordt de reële werksituatie dicht benaderd. Het grote verschil met de reële werksituatie is dat handhaving

van de orde vaak primair berust op de autoriteit van de echte leraar. In de tijd dat de auteurs van deze brochure zelf werden opgeleid tot onderwijzer, verliet de mentor (de onderwijzer die belast was met de begeleiding) soms het klaslokaal om te zien of de stagiar in staat was de orde te handhaven gedurende zijn afwezigheid. Tijdens de huidige lio-stage is deze situatie standaard: Het merendeel van de lessen wordt daarin gegeven zonder dat er direct toezicht is van een ervaren docent. Hoewel de verantwoordelijkheid van de effectiviteit van het gegeven onderwijs uiteindelijk bij de stagebiedende school ligt, zien we de lio-stage, als daarin tenminste een beoordelingsmoment is opgenomen, toch meer als een voorbeeld van beoordelen in de reële werksituatie dan als een work sample test.

Personen die een work sample test moeten maken zijn er over het algemeen van overtuigd dat ze daarmee beoordeeld worden op aspecten die er werkelijk toe doen. De uitslag van dergelijke tests wordt dan ook meestal gerespecteerd en geaccepteerd, zowel door de beoordeelde als door degenen die geacht worden beslissingen te nemen op grond van de testuitslag.

In een work sample test voert een persoon realistische taken uit onder realistische omstandigheden.

Bij de uitvoering van dergelijke taken wordt vaak een beroep gedaan op het hele scala van menselijke bekwaamheden: kennis, cognitieve, perceptuele, motorische en affectieve vaardigheden. Omdat het verloop van de taakuitvoering zich maar in beperkte zin laat voorspellen is de scoring een probleem. Noodgedwongen wordt meestal met vrij globale scoringsrubrieken gewerkt en wordt het aan de beoordelaar overgelaten in welke rubriek een gedragsaspect gescoord dient te worden. Bij het rijexamen bijvoorbeeld wordt de kandidaat op tien aspecten beoordeeld, zoals het in- en uitvoegen, het gedrag bij kruispunten en de uitvoering van bijzondere verrichtingen. Maar het is onmogelijk om scoringsvoorschriften op te stellen die zo concreet geformuleerd zijn dat het in elke verkeerssituatie duidelijk is hoe het gedrag van de kandidaat gescoord moet worden. Bij de proeflessen van studenten van lerarenopleidingen zal het niet anders zijn. Elke les is uniek en kan zich in onverwachte zin ontwikkelen en kan daarom slechts aan de hand van globale scoringsvoorschriften beoordeeld worden. Een ander nadeel van work sample tests is dat de uitvoering en soms ook de voorbereiding veel tijd kosten, zowel van de te beoordelen persoon als van de assessor. Men zal daarom geneigd zijn het aantal

work sample tests zo beperkt mogelijk te houden. Dat kan gemakkelijk tot verkeerde conclusies over competentie leiden aangezien succesvolle uitvoering van de ene taak geen garantie is voor de succesvolle uitvoering van andere, soortgelijke taken. Als een student van een lerarenopleiding in een les aardrijkskunde heeft gedemonstreerd op zinnige wijze gebruik te kunnen maken van een nieuw medium als Internet, wil dat nog niet zeggen dat hij dit in een les rekenen ook zou kunnen.

Resumerend kunnen we de volgende voor- en nadelen van de work sample test opnoemen:

Voordelen	Nadelen
<p>Zeer realistisch</p> <p>Hoge acceptatiegraad voor zowel de te beoordelen personen als voor de beslissingnemers</p>	<p>Moelijk om tot betrouwbare scoring te komen</p> <p>Afname kost veel tijd waardoor men het aantal uit te voeren taken beperkt wil houden. Dit kan tot verkeerde conclusies leiden</p>

GEMIDDELDE REPRESENTATIVITEIT: SKILL SAMPLE TEST

Bij sommige competenties kan het realiteitsgehalte van een work sample test als te hoog ervaren worden. Daar kunnen uiteenlopende redenen voor zijn. Soms gaat het om competenties die voor hun bewijsuitvoering van taken vereisen die erg gevaarlijk of duur zijn. In het onderwijs is daar meestal geen sprake van maar bij beroepen als die van brandweerman, arts, vlieger, enz. zijn er genoeg voorbeelden van te noemen. Soms ook wil men meer greep krijgen op het gedrag van de te beoordelen persoon. Bijvoorbeeld als men per se wil zien dat een persoon een bepaalde complexe vaardigheid, die deel uitmaakt van de competentie, onder de knie heeft. Bij een work sample test zijn er vaak maar weinig mogelijkheden om invloed uit te oefenen op het te vertonen gedrag. Wanneer de opdracht eenmaal in uitvoering is genomen, verloopt die voor een groot deel autonoom. Bij het rijexamen, bijvoorbeeld, kunnen noodsituaties niet afgedwongen worden en evenmin is het mogelijk om de verkeersintensiteit of de weersgesteldheid te beïnvloeden. Bij een proefles kan men bijna niet anders dan afwachten hoe de klas reageert op de activiteiten van de lesgever. Een andere reden kan zijn dat men het realiteitsgehalte

van een work sample test te hoog vindt met het oog op de gevorderdheid van de te beoordelen persoon terwijl men toch graag wil weten hoe het staat met de ontwikkeling van de betreffende competentie. In dergelijke gevallen neemt men vaak zijn toevlucht tot de skill sample test. De term geeft al aan dat hier niet meer de competentie als geheel wordt getoetst maar één of enkele geïsoleerde vaardigheden. Dit kan bereikt worden door de volgende maatregelen:

- De opdracht heeft minder vrijheidsgraden. Er wordt bijvoorbeeld concreter aangegeven welk gedrag vertoond moet worden of aan welke eisen het resulterende product moet voldoen;
- Tijdens de taakuitvoering worden stimuli ingevoerd die het verlangde gedrag uitlokken.

De skill sample test is in allerlei domeinen inzetbaar. Met name als de te beoordelen vaardigheid technisch georiënteerd is, kan het verlangde gedrag uitgelokt worden door machines, gereedschappen of andere technische ontwerpen op een bepaalde, voor-geprogrammeerde manier te laten reageren op acties van de te beoordelen persoon. Dergelijke ontwerpen worden simulatoren genoemd (denk aan de vliegsimulator, de gesimuleerde controlekamer van een chemische fabriek, en dergelijke). We zullen hier niet nader ingaan op de simulator omdat er op dit

moment geen simulatoren zijn waarmee bepaalde (sub)competenties van docenten getraind en vastgesteld kunnen worden. Dat ze er zullen komen staat echter vast. Zodra de spraakherkenningstechnologie wat verder ontwikkeld is, komen er mogelijkheden om virtuele leerlingen te laten reageren op wat er gezegd wordt. Voorlopers zijn simulatoren die gebruikt worden om mensen te trainen in het optreden voor een zaal en, meer in de entertainment-sfeer, de simulator waarbij je als dirigent een orkest mag leiden.

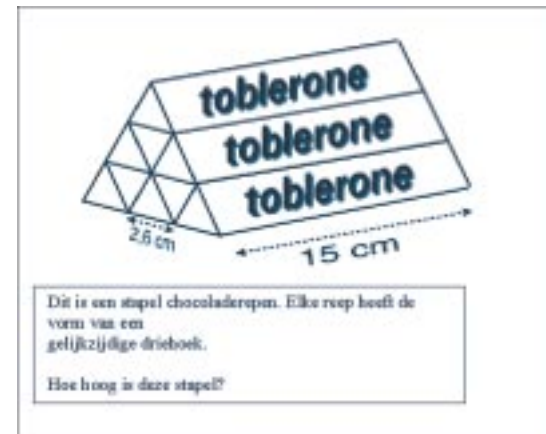
Als het om interpersoonlijke vaardigheden gaat, worden vaak acteurs gebruikt. In het medisch onderwijs bijvoorbeeld wordt de skill sample test gebruikt om onderzoeks- en behandelingsvaardigheden te trainen en vast te stellen. Regelmatig wordt daarbij gebruik gemaakt van acteurs die getraind zijn in het nabootsen van een patiënt met een bepaalde ziekte of kwetsuur en die volgens een script reageren op vragen en handelingen van de te beoordelen persoon.

Het uitlokken van gedrag heeft twee belangrijke voordelen. Anders dan in een work sample test hoeft men niet passief af te wachten of en wanneer het verlangde gedrag zich zal voordoen.

In dezelfde tijd kan met de skill sample test meer relevant gedrag beoordeeld worden. Bijkomend voordeel is dat er met meer gedetailleerde scoringsrubrieken gewerkt kan worden. Immers, men weet welk gedrag er te verwachten is. Beoordelaars hoeven daardoor minder interpreterend te werk te gaan bij de beoordeling, wat leidt tot betrouwbaarder scores. Nadeel van het expliciet uitlokken van gedrag is dat er meer tijd in de voorbereiding gaat zitten (vooral de meer complexe gedragingen vereisen een gedegen training van degenen die op 'voor-geprogrammeerde' wijze gedrag moeten uitlokken bij de te beoordelen persoon). En natuurlijk wordt er ingeleverd op het gebied van de representativiteit. Het verlangde gedrag wordt als het ware uit de te beoordelen persoon 'geperst' en het is maar de vraag of de betreffende persoon in de reële werksituatie in staat zou zijn om het verlangde gedrag te vertonen als respons op minder nadrukkelijk aanwezige stimuli. In het schema hiernaast en op pag. 28 is een (niet volledig uitgewerkt) voorbeeld opgenomen van een skill sample test voor het beoordelen van docent-competenties.

Opdracht: Begeleid een groepje leerlingen die in de wiskundeles samen onderstaande opgave moeten zien op te lossen door de stelling van Pythagoras te gebruiken. Het is de bedoeling dat de leerlingen leren van elkaar door de verschillende inzichten die er bij hen leven te laten uitwisselen. De groep moet unaniem achter de voorgestelde oplossing staan.

Tijd: 20 minuten



Vorbereiding door testleider

Een van de leerlingen (leerling A) krijgt als (geheime) instructie mee dat hij zijn eigen foutieve benadering niet wil opgeven voor een door de rest van de groep overeengekomen aanpak. Een andere leerling (B) krijgt als (geheime) instructie mee om de suggestie van leerling A op botte wijze af te wijzen. Leerling A wordt geïnstrueerd om zich als volgt te gedragen: Hij komt als eerste van de groep met zijn (foute) visie van een oplossing: opmeten met een liniaal. Als leerling B deze visie bot afwijst, trekt leerling A zich terug uit het groepsproces. Hij is pas weer bereid om mee te doen als er opnieuw aandacht aan zijn voorstel wordt besteed door de groep en er overtuigend wordt aangetoond wat er mis mee is.

Beoordeling

De kandidaat wordt primair beoordeeld op de sturing die hij geeft aan het groepsproces. Hij moet ervoor zorgen dat elk groepslid zijn inbreng heeft, dat de sfeer goed blijft, dat de oplossing binnen het bereik van de groep blijft zonder zelf te gerichte hints te geven. Voor elk van deze aspecten kunnen enkele prestatiecriteria geformuleerd worden die te scoren zijn op een beoordelingsschaal zoals hieronder afgebeeld:

Stimuleren van het groepsproces

De kandidaat:

4	3	2	1
weet door zijn houding en vragen elke leerling te prikkelen tot het leveren van een bijdrage aan het groepsproces.	weet de meeste groepsleden actief bij het oplosproces te betrekken	laat zich te veel leiden door enkele leerlingen met veel inbreng	heeft vooral oog voor zijn eigen opvattingen en legt die op aan de groep.

Resumerend kunnen we de volgende voor- en nadelen van de skill sample test opnoemen:

Voordelen	Nadelen
Efficiënt omdat gedrag meer gericht kan worden uitgelokt	Vereist meer voorbereidings tijd (schrijven en inoefenen van 'script' om gedrag uit te lokken)
Scoring betrouwbaarder dan bij work sample test	Gereduceerde representativiteit roept vragen op over de betekenis van de scores

LAGE REPRESENTATIVITEIT: COGNITIVE SKILL TEST

Er zijn situaties denkbaar waarin noch voor de work sample test noch voor de skill sample test gekozen kan worden. Steeds vaker vormen de competenties de bouwstenen voor beroepsopleidingen en worden studenten vanaf de start van de opleiding geconfronteerd met realistische praktijkproblemen. Meestal is de problematiek die in een work sample test wordt aangeboden nog te complex voor de beginnende student. Bovendien is het organisatorisch en financieel vaak niet haalbaar om elke student frequent

een toets van hoge of gemiddelde representativiteit af te nemen. Dan kan de cognitive skill test uitkomst bieden. Met een cognitive skill test kan worden vastgesteld of een student weet hoe hij/zij zou moeten handelen om, gegeven een bepaalde probleemsituatie, tot een oplossing te komen. De complex skill test meet probleemoplosvaardigheid; een belangrijk maar op zichzelf niet toereikend aspect van een beroepscompetentie. Echt harde bewijzen voor competentie zijn niet te verkrijgen zonder uitvoering van complexe taken in de beroepspraktijk of in situaties die daar sterk op lijken. Daarbij gaat het om taken waarin denken en doen hand in hand gaan en behalve cognitieve vaardigheden (zoals probleemoplosvaardigheid) ook motorische, perceptuele en affectieve vaardigheden een belangrijke rol spelen. Desalniettemin kunnen cognitive skill tests een belangrijke rol spelen tijdens een beroepsopleiding. Uit de literatuur is bekend dat experts op een vakgebied zich onderscheiden van beginners doordat ze in hun geheugen veel meer probleemsituaties en daarbij behorende aanpakken hebben opgeslagen. Hoe meer probleemsituaties worden aangeboden des te groter de kans dat die later herkend worden en de persoon op passende wijze zijn vaardigheden weet te mobiliseren.

De cognitive skill test vergt weinig voorbereiding en kan snel worden afgenomen, wat betekent dat er tijd is om studenten meerdere problemen voor te leggen. Bovendien kunnen de aan te bieden problemen qua complexiteit afgestemd worden op de gevorderdheid van de studenten; iets wat met work sample tests minder goed mogelijk is omdat de complexiteit daarvan maar beperkt onder controle te houden is.

De cognitive skill test kent vele gedaanten, van de oude vertrouwde casustoets die op papier wordt aangeboden en uitgewerkt en door een docent wordt nagekeken tot de moderne multimediatest die op een beeldscherm wordt gepresenteerd en gemaakt en door de computer wordt nagekeken. In laatstgenoemde varianten wordt de probleemsituatie vaak verrassend realistisch gepresenteerd, met name wanneer daarvoor videobeelden worden gebruikt. Zo wordt er bij de Koninklijke Landmacht een multimediatest gebruikt om de triage-competentie te trainen van militaire artsen (een triage is het maken van een verdeling in oorlogsgewonden met het oog op hun latere behandeling). Het programma is primair ontwikkeld voor instructiedoeleinden maar zou met enige aanpassingen ook heel goed voor toetsing kunnen worden ingezet (Krol, 1992).

In dit programma wordt de militaire arts geconfronteerd met een virtueel slagveld met een grote diversiteit aan gewonde soldaten. De opdracht is om binnen een bepaald tijdsbestek een schifting in de gewonden aan te brengen die zal leiden tot een zo groot mogelijk aantal overlevenden. De militaire arts kan daarbij allerlei onderzoek doen en eventueel snelle medische handelingen verrichten. De simulatie is dynamisch, wat inhoudt dat de situatie van de gewonden kan veranderen onder invloed van tijd en/of medische handelen.

Een ander voorbeeld is de SQ-test, een multimediatest voor interpersoonlijke effectiviteit (van der Maessen de Sombreff, Westen & de Veer, 2000). Met de SQ-test kan worden vastgesteld hoe goed personen inzicht hebben in de kwaliteit van gedrag in sociale situaties. Tijdens een testafname wordt een persoon geconfronteerd met een groot aantal sociale interacties in werksituaties. Deze minicassussen verlopen allemaal conform de volgende procedure:

- Eerst wordt er achtergrondinformatie gegeven over de sociale interactie.
- Vervolgens krijgt men een video-opname te zien van een cruciale fase in de sociale interactie.

- Daarna moet de kandidaat van vier mogelijke reacties, alle te starten als videofilm, op een vierpuntsschaal aangeven hoe effectief die zijn. De antwoorden worden beoordeeld door ze te vergelijken met het consensuoordeel van experts. Uit onderzoek is gebleken dat de betrouwbaarheid van dit consensuoordeel hoog is.
- Er wordt gerapporteerd in de vorm van een profielscore bestaande uit een totaalscore en deelscores voor onderliggende relevante gedrags aspecten.

Sedert september 1999 werkt een aantal onderwijsorganisaties aan de ontwikkeling van een docenten development center (over wat dat is komen we verderop nog te spreken) waarvan het belangrijkste onderdeel betrekking heeft op het vaststellen van sociale competenties. Daarvoor wordt de procedure gebruikt die ten grondslag ligt aan de SQ-test. Middels literatuuronderzoek en gesprekken met ervaringsdeskundigen (docenten uit verschillende onderwijssectoren) is een overzicht verkregen van de sociale competenties die van belang zijn voor het lerarenberoep. Momenteel wordt gewerkt aan schaalconstructie en uitwerken en verfilming van scripts van kritieke situaties (Instituut voor Leraar en

School, 2000). Een aardige ontwikkeling waarvan wij met grote belangstelling het valideringsonderzoek zullen volgen.

Met een cognitive skill test wordt vastgesteld of iemand weet hoe er in bepaalde situaties gehandeld moet worden, niet of hij het ook werkelijk kan en dat is natuurlijk een zwak punt. Wie zou docenten een bevoegdheid willen geven die alleen maar in een computersimulatie hebben laten zien om te kunnen gaan met complexe klassensituaties? Van de andere kant ben je met de cognitive skill test in staat om grote aantallen situaties in korte tijd te presenteren, waaronder situaties die tamelijk zeldzaam maar toch kritisch zijn. De scoring kan heel betrouwbaar zijn omdat van tevoren alle handelingsalternatieven uitgewerkt zijn en deskundigen een unaniem oordeel hebben uitgesproken over de beste oplossing. Overigens geldt dit laatste alleen voor cognitive skill tests waarin een groot aantal korte taaksituaties wordt aangeboden. Bij cognitive skill tests, zoals de hiervoor besproken triage-simulatie, waarin de nadruk meer ligt op het probleemoplosproces, is de scoring toch vaak problematisch omdat experts onderling niet tot overeenstemming kunnen komen over de kwaliteits-

verschillen tussen de vele, mogelijke oplossroutes. Resumerend zijn de voor- en nadelen van de cognitive skill test:

Voordelen	Nadelen
In korte tijd kan een groot aantal verschillende taaksituaties worden voorgelegd	Hoge ontwikkelingskosten
Afname kost weinig voorbereiding en begeleiding	Kennis is een noodzakelijke maar onvoldoende voorwaarde voor competentie
Betrouwbare scoring (niet voor processen)	

Verderop zullen we de hierboven besproken methoden nog evalueren aan de hand van drie kwaliteitskenmerken die van cruciaal belang zijn voor de kwaliteit van beslissingen. Voor een goed begrip daarvan is het nodig eerst stil te staan bij het probleem van het trekken van conclusies over competentie op basis van assessmentresultaten.

2.3 Van scores naar conclusies

2.3.1 Over doelprestaties en middelprestaties

Prestatiebeoordeling is niet het exclusieve domein van onderwijsgevendenden. Bij alle menselijke activiteiten komt op een zeker moment de behoefte op om prestaties te beoordelen. In de sport bijvoorbeeld is die behoefte heel nadrukkelijk aanwezig. Toch is er een groot verschil tussen prestatiebeoordeling in de sport en prestatiebeoordeling in het onderwijs. Bij een sportwedstrijd gaat de belangstelling van de assessoren uit naar de prestatie zelf. Zo gaat het bij een hoogspringwedstrijd om die ene of die enkele sprongen die de atleet uitvoert; die zijn het doel van het assessment. Als we in het onderwijs prestaties van deelnemers beoordelen, zijn we niet in de eerste plaats geïnteresseerd in de prestaties op de uitgevoerde taken. Wat we eigenlijk willen weten is of de betreffende deelnemers hetzelfde prestatieniveau zouden bereiken op andere, soortgelijke taken uitgevoerd onder wisselende omstandigheden. De prestaties zijn hier niet het doel maar een middel om te komen tot een uitspraak over een competentie of andere persoonsgebonden eigenschap van de beoordeelde. Ga maar na: als een leraar bij zijn leerlingen

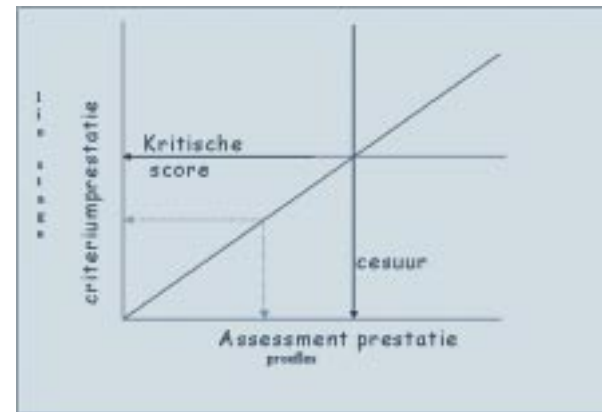
een toets afneemt met twintig opgaven over het optellen en aftrekken van ongelijknamige breuken, dan gaat zijn interesse niet primair uit naar de prestatie op deze twintig opgaven maar naar de prestatie op alle mogelijke, soortgelijke breukopgaven zodat hij een uitspraak kan doen over de beheersing van de vaardigheid 'optellen en aftrekken van ongelijknamige breuken'.

Als de prestatie zelf het doel van het assessment is, zoals bijvoorbeeld bij een hoogspringwedstrijd, zijn er geen problemen met de te trekken conclusies. Die komen heel eenduidig voort uit de prestaties en een vastgesteld criterium. Dat criterium kan absoluut of relatief georiënteerd zijn. In het eerste geval krijgt bijvoorbeeld iedereen een prijs die hoger dan 180 cm springt. In het tweede - meer gebruikelijke - geval wordt de eerste prijs uitgereikt aan degene die over een hogere lat is gesprongen dan de overige deelnemers. Als daarentegen niet de prestatie zelf het doel is maar een achterliggende vaardigheid of competentie, wordt het trekken van conclusies een stuk lastiger. Er zal beargumenteerd moeten worden waarom de conclusie gerechtvaardigd is. Dat een dergelijke argumentatie belangrijk is, weet iedereen die te maken heeft met de certificering

van belangrijke competenties. In de VS wordt al jarenlang intensief nagedacht over deze problematiek omdat gedupeerden (mensen die menen dat het verlangde certificaat hen ten onrechte werd onthouden of mensen die zich bedreigd voelen door incompetent beroepsbeoefenaars) daar sneller de gang naar de rechter maken om hun gelijk te halen.

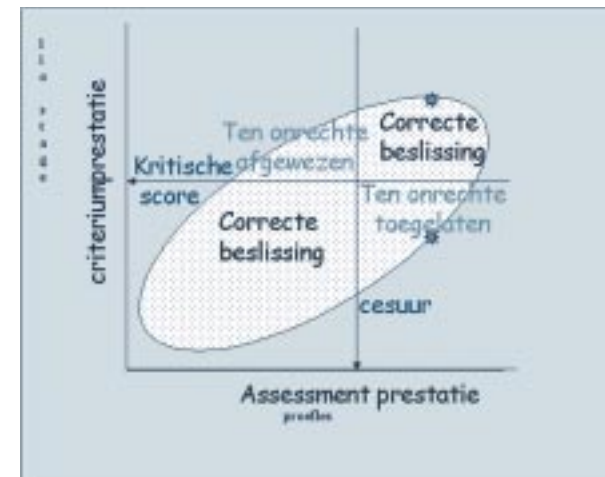
Wat is precies het probleem? Elk assessment heeft tot doel conclusies te trekken over vaardigheden of competenties van mensen. Daarbij wordt ervan uitgegaan dat het gedrag vertoond tijdens het assessment in hoge mate soortgelijk gedrag kan voorspellen in een relevante vervolg-situatie. Laat u even meevoeren in de volgende hypothetische situatie. Veronderstel eens dat studenten van een lerarenopleiding pas tot de lio-stage worden toegelaten als ze een voldoende resultaat hebben behaald op een proefles. Veronderstel verder dat, in het kader van een experiment, de schoolleiding bepaald heeft dat alle kandidaten tot de lio-stage worden toegelaten, ook die met een onvoldoende resultaat op de proefles. Veronderstel vervolgens dat van alle kandidaten tijdens de lio-stage op uitgebreide schaal video-opnames zijn gemaakt van allerlei lessituaties en dat die beoor-

deeld zijn aan de hand van dezelfde prestatiecriteria als die gebruikt werden bij de beoordeling van de proefles. Van elke kandidaat zijn dan twee totaal-scores beschikbaar, één gebaseerd op de proefles en één gebaseerd op de videofragmenten tijdens de lio-stage. In figuur 2 wordt deze situatie grafisch weergegeven. Op de horizontale as wordt de prestatie afgebeeld zoals vertoond tijdens de proefles. Op de verticale as is de prestatie afgebeeld zoals vastgesteld in de lio-stage. In figuur 2 hiernaast liggen alle coördinaten op de diagonaal, wat betekent dat bij een bepaalde score op de proefles (de horizontale schaal) slechts één score hoort op het criterium dat voorspeld moet worden (de prestatie tijdens de lio-stage, afgezet op de verticale schaal). In deze situatie is het assessment een perfecte voorspelling van het criterium en worden over alle kandidaten correcte beslissingen genomen.



Figuur 2. Een perfect verband tussen assessment- en criteriumprestatie.

Figuur 3 geeft een wat realistischer - maar ook nog heel positief - beeld van de voorspellende kracht van het assessment. Over het algemeen komt een lage score op het assessment overeen met een lage score op het criterium en geldt dezelfde tendens voor middelmatige en hoge scores. Maar bij een bepaalde score op het assessment blijken nu verschillende scores op het criterium te kunnen horen. De sterretjes in figuur 3 geven aan hoe ver scores op het criterium uit elkaar kunnen liggen, uitgaande van dezelfde testprestatie. Die scores op het criterium kunnen zelfs zo uiteenlopend zijn, zoals hier het geval is, dat er verschillende kwalificaties bij horen. Duidelijk is ook te zien dat niet over alle 'kandidaten' een correcte beslissing wordt genomen. Er zijn bijvoorbeeld kandidaten die ten onrechte worden aangemerkt als niet-beheersers. Dat zijn personen die een onvoldoende haalden voor de proefles maar die in de criteriumsituatie toch voldoende bleken te presteren. Maar er zijn ook personen die ten onrechte zijn aangemerkt als beheersers. Zij haalden weliswaar een voldoende voor de proefles maar gaven in de criteriumsituatie blijk van een onvoldoende beheersingsniveau.



Figuur 3. Verkeerde beslissingen als gevolg van verschillen tussen assessment- en criteriumtaken.

2.3.2 Kwaliteitsaspecten van een assessment

Als we verkeerde beslissingen over personen zoveel mogelijk willen vermijden, zullen we moeten proberen het verband tussen de assessment- en criteriumprestaties zo groot mogelijk te doen zijn. Alleen dan zullen de coördinaten weer in een smaller gebied rondom de ideale diagonaal komen te liggen.

Om dat te bewerkstelligen is het noodzakelijk dat toetsconstructeurs zich bewust zijn van de obstakels die het trekken van conclusies over competentie op basis van testprestaties in de weg staan. Kane (1992) verwoordt het als volgt: ‘...in evaluating the validity of assessments of professional competence, it is important to look for possible flaws in the chain of inferences from the results of the measurement procedure to conclusions about competence.’

‘Chain of inference’ zullen we in het vervolg aanduiden met de term ‘kwaliteitsketen’ omdat het gaat om een reeks van achtereenvolgens te evalueren kwaliteitsaspecten (de schakels in de keten). Elke schakel in deze keten moet solide zijn om verantwoorde conclusies over competentie te kunnen trekken. De belangrijkste schakels hebben te maken met de accuraatheid, generaliseerbaarheid en extrapolatieerbaarheid van de assessmentscores. Deze schakels zijn alle aspecten van de meer bekende psychometrische kwaliteitsbegrippen ‘betrouwbaarheid’ en ‘validiteit’, waarover meer in paragraaf 3.2.1. In figuur 4 wordt de kwaliteitsketen grafisch weergegeven alsmede de gevolgen van het verbreken van de keten voor de te trekken conclusies.



Figuur 4. De ‘kwaliteitsketen’ en de gevolgen van niet solide schakels.

ACCURAAUWHEID VAN DE SCORE

De geldigheid van de conclusie over het prestatieniveau is afhankelijk van de accuraatheid

van de score. Als de score niet accuraat is dan geeft die geen betrouwbare indicatie van het prestatie-niveau; dit had dan even goed hoger of lager kunnen uitvallen. De accuraatheid van de score wordt voor een belangrijk deel bepaald door de duidelijkheid en specificiteit van de geformuleerde prestatiecriteria waarmee het waargenomen gedrag beschreven kan worden. Soms is dat goed te doen. Denk bijvoorbeeld aan de keuzeopties van een meerkeuzevraag die in feite niets anders zijn dan beknopt geformuleerde beschrijvingen van (de uitkomst van) gedrag dat men op voorhand verwacht van een kandidaat aan wie een probleem is voorgelegd. Maar naarmate de voorgelegde problemen complexer worden en meer een 'open-einde karakter' hebben, wordt het moeilijker om te anticiperen op alle mogelijke te verwachten gedragingen of uitkomsten daarvan en daarvoor heldere en voldoende specifieke prestatiecriteria te formuleren.

Een score is accuraat als die het ware vaardigheidsniveau van de kandidaat dicht benadert. Dat is de gemiddelde score die de kandidaat zou halen wanneer die de voorgelegde taken uit het assessment een oneindig aantal keren zou uitvoeren, aangenomen dat de prestaties niet beïnvloed zouden

worden door het vermogen van de kandidaat om lering te trekken uit de herhaling of door het optreden van vermoeidheidsverschijnselen (vrij naar De Groot & Van Naerssen, 1977). De accuraatheid wordt vooral bedreigd door onvolkomenheden die te maken hebben met de omstandigheden bij de afname en het gebruikte scoringssysteem. Crooks, Kane en Cohen (1996) noemen in dat verband de volgende bedreigingen:

- *Geringe motivatie.* Kandidaten met een geringe motivatie laten niet het 'achterste van de tong' zien. Hun score is dan geen betrouwbare indicatie voor de werkelijke capaciteiten. De motivatie om te presteren komt onder druk te staan als kandidaten geen vertrouwen hebben in de gebruikte assessmentmethode of in het ongewisse zijn over hoe, waarvoor en door wie de scores gebruikt worden.
- *Testangst.* Als de motivatie om goed te presteren zeer hoog is, ontstaat soms testangst. Dit verschijnsel is vrij algemeen als er zware beslissingen genomen worden op grond van de uitslag. Het probleem is echter dat de uitwerking niet identiek is bij elk individu: sommigen presteren ver onder hun kunnen terwijl anderen onder druk van de situatie extra goed presteren.

- *Niet opgevolgde testcondities.* Dit kan de score ten onrechte verlagen of verhogen. Het eerste kan bijvoorbeeld het gevolg zijn van onvoldoende toebemeten tijd, van ongunstige omstandigheden (lawaai, onvoldoende verlichting) of van onvoldoende duidelijke instructies. Van het tweede kan sprake zijn als de kandidaten meer voorbereiding op het assessment krijgen dan is toegestaan of tijdens het assessment meer tijd of hulp krijgen dan waar ze volgens de beschreven testcondities recht op hebben.
- *Gebrek aan scoreconsistentie.* Riskant wordt de interpretatie van scores als blijkt dat de hoogte van een score bepaald wordt door andere zaken dan de prestatie zelf. Deze situatie wordt vaak pas zichtbaar als de intra- en/of interbeoordelaars overeenstemming zijn vastgesteld. Eerstgenoemde overeenstemmingsmaat betreft de overeenstemming tussen scores op een meting en een herhaalmeting toegekend door een bepaalde beoordelaar. De tweede overeenstemmingsmaat heeft betrekking op de overeenstemming tussen de scores van een groep beoordelaars. Het is verleidelijk om te trachten deze overeenstemmingsmaten te verhogen door bepaalde aspecten maar niet meer te beoordelen of minder zwaar te

laten meetellen voor de bepaling van de totaal-score, maar dit kan ernstige gevolgen hebben voor de validiteit van de score.

- *Te gedetailleerde of te globale scoring.* Met een te gedetailleerde scoring loopt men het gevaar dat de score geen goede afspiegeling is van de geleverde prestatie. De kwaliteit van een diapresentatie per computer, bijvoorbeeld, moet niet primair beoordeeld worden op de technische uitvoering van elke afzonderlijke dia maar op de indruk die de presentatie als geheel achterlaat bij het publiek. Het probleem is om een tamelijk vage kwaliteit als 'indruk' te vertalen in een reeks prestatiecriteria die objectief kunnen worden vastgesteld. Als dat niet lukt wordt vaak besloten om de prestatie als geheel te beoordelen. Maar ook dat is de verkeerde weg aangezien de literatuur over het beoordelen van producten en processen eenduidig is over de onbetrouwbaarheid van globale beoordelingen.

GENERALISEERBAARHEID VAN DE SCORE

De score, zijnde een kwantificering van de waargenomen prestatie, is op zichzelf niet interessant. Ga maar na. Een docent die een toets afneemt met 20 opgaven over het optellen en aftrekken van

ongelijknamige breuken is niet primair geïnteresseerd in de uitkomst van een opgave als: $\frac{3}{4} - \frac{1}{2}$; hij wil gewoon weten of zijn leerlingen de algoritmes beheersen voor het optellen en aftrekken van eenvoudige, ongelijknamige breuken. Een score ontleent zijn betekenis dus aan het feit dat hij gegeneraliseerd kan worden naar een score die de kandidaat naar verwachting zal halen op het toetsdomein; dat bevat alle mogelijke toetsopgaven die voor dat stuk leerstof geconstrueerd kunnen worden. Eenvoudig gezegd komt het erop neer dat de betekenis van de score zich niet mag beperken tot de aangeboden taken in het assessment maar zich moet uitstrekken naar alle soortgelijke taken die aangeboden hadden kunnen worden, rekening houdend met variaties in taken, omstandigheden en beoordelaars. Laten we ter verduidelijking nog een ander voorbeeld geven. Volgens Robert Gagné, een bekend leertheoreticus, moet om het leerproces bij een lerende succesvol te laten verlopen, externe ondersteuning geboden worden middels de volgende interventies (Gagné, 1985):

- aandacht trekken;
- op de hoogte stellen van het leerdoel;
- stimuleren tot herinnering van eerder geleerde, voorwaardelijke kennis;

- aanbieden van de nieuwe leerstof met veel nadruk op onderscheidende kenmerken;
- ondersteunen van de verwerving door de leerstof betekenisvol te maken voor elke individuele lerende;
- uitlokken van gedrag waaruit kan blijken welke onderdelen al beheerst worden en welke nog niet;
- bijsturen van de instructie;
- definitief vaststellen of het leerdoel bereikt is;
- bieden van gelegenheid tot oefening zodat het geleerde beklijft en er makkelijker transfer optreedt.

Als wij zouden willen vaststellen of studenten van een lerarenopleiding deze interventies inderdaad toepassen in hun onderwijs, dan is het riskant om te willen volstaan met één of enkele waarnemingen. Immers, het kan best zijn dat de betreffende persoon de interventies uitstekend weet te verwerken in lessen waarin de nadruk ligt op het verwerven van begrippen of principes maar erg veel moeite heeft om ze te gebruiken in lessen waarin het gaat om het toepassen van eerder verworven kennis in nieuwe probleemsituaties. Evenmin is het zeker dat het kunnen toepassen van de principes in een groepsinstructie garandeert dat ze ook succesvol kunnen worden toegepast in geïndividualiseerde instructies,

bijvoorbeeld in de ontwerpen voor computer-gestuurde instructies. En ook kan niet voetstoots worden aangenomen dat de vaardigheid zich uitstrekt over vakgebieden: structuur aanbrengen in een instructie blijkt vaak moeilijker bij inhouden waarin de nadruk ligt op een ander dan cognitief aspect, zoals bijvoorbeeld bij het vak tekenen, gymnastiek of muziek.

Het probleem bij het vaststellen van professionele competenties is dat die vaak dermate veelomvattend zijn dat het aantal potentiële taken te groot is om ze allemaal te laten uitvoeren. Schaarste aan tijd en middelen heeft tot gevolg dat het aantal voorgelegde taken meestal niet in verhouding staat tot het aantal potentiële taken, als gevolg waarvan het assessment niet opgevat mag worden als een representatieve steekproef uit het 'toetsdomein'.

Of generalisaties toelaatbaar zijn, wordt uitgezocht in zogeheten generaliseerbaarheidsstudies. Daarin wordt onderzocht of scores stabiel blijven als er andere taken worden aangeboden, als de prestaties onder andere omstandigheden tot stand komen, als de prestaties door andere

assessoren gescoord worden, enz. Uit dergelijke studies is gebleken dat de generaliseerbaarheid van scores op gestandaardiseerde kennistoetsen over het algemeen goed is. Jammer genoeg blijkt dit niet te gelden voor scores op assessments waarmee complexe vaardigheden worden vastgesteld. Met name de stabiliteit over taken laat te wensen over (Kane, Crooks & Cohen, 1999). Te kleine en niet representatieve steekproeven zijn doorgaans de oorzaak.

Als de score niet generaliseerbaar is maar wel accuraat, dan zegt de score alleen iets over het prestatieniveau op de aangeboden taken in het assessment maar geeft die geen betrouwbare indicatie over het prestatieniveau op alle andere taken uit het toetsdomein, die ook aangeboden hadden kunnen worden.

EXTRAPOLEERBAARHEID VAN DE SCORE

Ook de verwachte score op het toetsdomein heeft niet onze bijzondere belangstelling. Uiteindelijk willen we niet weten hoe goed iemand presteert op de assessment taken maar wat zijn score is als soortgelijke taken worden uitgevoerd onder realistische omstandigheden. Dit wordt wel de taakdomeinscore genoemd. Deze laatste kwaliteitsschakel is verreweg

het lastigste om te evalueren. De geldigheid van het afleiden van de taakdomeinscore uit de toetsdomeinscore kan in principe empirisch geverifieerd worden door assessmentcores te vergelijken met criteriumcores. Met die laatste bedoelen we scores die tot stand worden gebracht door toepassing van de best mogelijke evaluatiemethoden zonder daarbij last te hebben van gebruikelijke hinderpalen als tekort aan tijd en geld. Het moge echter duidelijk zijn dat empirische verificatie van de extrapoleerbaarheid in de meeste gevallen onhaalbaar is en eigenlijk alleen theoretische waarde heeft. Meestal is de enig begaanbare weg de extrapoleerbaarheid aan te tonen door logische argumentatie. Als overtuigend kan worden beargumenteerd dat de assessment taken representatief zijn voor het toetsdomein en laatstgenoemde een natuurgetrouwe afspiegeling is van de taken in de werksituatie, dan is het trekken van conclusies over competentie op grond van een assessmentscore tamelijk veilig.

In tabel 1 (op pag. 42 en 43) worden de in paragraaf 2.2 aan de orde gestelde methoden voor het beoordelen van competentie geëvalueerd op grond van de zojuist besproken kwaliteitsaspecten. De tabel geeft weinig reden tot optimisme over de

kwaliteit van conclusies over competentie. Alle methoden blijken één of meer zwakke schakels te hebben, althans wanneer ervan uitgegaan mag worden dat er beperkingen zijn opgelegd aan het gebruik van tijd en middelen. Als dat laatste niet het geval is, ligt het voor de hand dat observatie in de reële werksituatie tot de beste conclusies over competentie leidt.

2.4 De beste methode is een methodemix.

Als een ding duidelijk is geworden uit de vorige paragraaf dan is het wel dat er geen beste methode is. Allemaal hebben ze ergens in de 'kwaliteitsketen' een zwakke schakel die de geldigheid van conclusies over competentie aantast. Gelukkig heeft niet elke methode de zwakke schakel op dezelfde plaats zodat een methodemix een mogelijke uitweg uit het dilemma kan bieden. Een methodemix heeft bovendien als voordeel dat er meer informatie wordt verzameld. Veel informatie verzamelen is erg belangrijk; hoe meer informatie des te geringer de invloed van meetfouten op het gemiddelde resultaat. We zullen nu twee vormen van een methodemix bespreken. Allebei staan ze momenteel in de belangstelling van lerarenopleiders.

methoden	accuraatheid		generaliseerbaarheid		extrapolatie	
observatie in reële werksituatie	□	Redelijk. Mits er gewerkt wordt met meerdere getrainde beoordelaars. Die zijn nodig omdat de onvoorspelbaarheid van de reële werksituatie het noodzakelijk maakt met globaal geformuleerde beoordelingsaspecten te werken.	-	Slecht. Meestal is het waargenomen gedrag niet representatief voor het gedrag waartoe men in staat moet zijn, gegeven de competentie die in het geding is.	□	Redelijk. Weliswaar is het waargenomen gedrag het meest authentieke dat men te zien kan krijgen maar, als gevolg van de beperkte generaliseerbaarheid, toch onvoldoende voor het trekken van conclusies over competentie.
retrospectie	□	Redelijk. Zwak punt is het beroep dat gedaan wordt op het geheugen. Maar dit wordt deels gecompenseerd door de grote aantallen waarnemingen die haalbaar zijn.	-	Slecht. De beoordelingsaspecten van het instrument kunnen heel representatief zijn voor de betreffende competentie. Maar het valt niet te verifiëren of de te scoren beoordelingsaspecten in verband zijn gebracht met relevant gedrag.	-/□	Slecht tot redelijk. De scores zijn gebaseerd op authentiek gedrag maar aangezien er nauwelijks zekerheid te krijgen is over de generaliseerbaarheid, is extrapolatie naar het doeldomein zeer beperkt.
work sample test	□	Redelijk. De taakstelling lokt het gewenste gedrag uit. Daarop kan geanticipeerd worden bij het formuleren van beoordelingsaspecten. Meerdere getrainde beoordelaars zijn een vereiste om tot accurate scores te komen.	-/□	Slecht tot redelijk. Wel wat beter dan bij observatie in een reële werksituatie. Maar de bewerkelijkheid van de methode maakt dat ook hier meestal geen representatief beeld ontstaat.	□	Redelijk. Het gedrag dat vertoond wordt, wijkt niet veel af van het gedrag in een reële werksituatie. Helaas kan deze hoge natuurgetrouwheid niet voldoende compensatie bieden voor de te beperkte steekproef uit het toetsdomein.

skill sample test	□ /+	Redelijk tot goed. Nauwkeurig afgebakende en beknopte taken maken de formulering van heldere en eenduidig te interpreteren beoordelingsaspecten mogelijk.	□ /+	Redelijk tot goed. De beknopte taken bieden in principe mogelijkheden voor het aanbieden van veel taken die samen een representatieve steekproef vormen uit het toetsdomein.	□	Redelijk. Het gedrag dat vertoond wordt, wordt ten dele uitgelokt door voorgeprogrammeerde interventies. De natuurgetrouwheid heeft daaronder te lijden.
cognitive skill test	+	Goed. Nauwkeurig afgebakende en beknopte taken maken het mogelijk om het respons-gedrag in hoge mate te voorspellen en dit objectief te scoren.	+	Goed. De verregaande reductie van de complexe werkelijkheid maakt het mogelijk heel veel taken aan te bieden. Er is een grote kans dat die taken representatief zijn voor het toetsdomein.	-	Slecht. Deze methode stelt vast of de kennis die aan een competentie ten grondslag ligt, in voldoende mate beheerst en toegepast kan worden. Er wordt geen gedrag beoordeeld.

Tabel 1. Hoe solide is de kwaliteitsketen bij toepassing van verschillende beoordelingsmethoden?

2.4.1 Assessment Center

Het assessment center (AC) is een verwarrende naam. De term roept bij sommige mensen het beeld op van een locatie waar psychologen op allerlei manieren proberen iemand te doorgronden. Dat beeld is niet helemaal juist. Het woordje center verwijst niet naar een plaats maar naar de methode die gehanteerd wordt om uitspraken te kunnen doen over de competentie van een kandidaat. 'Center' staat voor het combineren (centreren) van meerdere beoordelingstechnieken in een programma (de Rijke & Dochy, 1993).

Wanneer is er sprake van een AC? Hierover wordt duidelijk verschaft in een publicatie van een internationale 'task force' die zich bezig gehouden heeft met het opstellen van richtlijnen voor het gebruik van AC's (International Task Force on Assessment Center Guidelines, 2000). Volgens deze werkgroep mag er pas van een AC gesproken worden als voldaan is aan de volgende voorwaarden:

- *Er is een functieanalyse uitgevoerd.*
Om boven tafel te krijgen waarop beoordeeld moet worden, is uitvoering van een functieanalyse noodzakelijk om dimensies, competenties en prestatiecriteria te kunnen identificeren.
- *Er is een gedragsclassificatiesysteem ontwikkeld.*
Gedragingen die de deelnemers vertonen moeten geassocieerd worden (op welke competenties heeft dit gedragsaspect betrekking) en vervolgens beoordeeld (moet het betreffende gedragsaspect met een hoge of lage score gewaardeerd worden?).
- *Er wordt van verschillende assessmenttechnieken gebruik gemaakt.* Zoals in vorige paragrafen is duidelijk gemaakt, hebben alle technieken sterke en zwakke kanten waar het gaat om de geldigheid van de scores. Door technieken te selecteren die van elkaar verschillen in hun sterke en zwakke eigenschappen wordt geprobeerd de kwaliteit van de conclusies op basis van het AC te verhogen.
- *Er wordt gebruik gemaakt van simulaties.*
Eén van de gebruikte assessmenttechnieken moet een simulatie zijn waarin de kandidaat reageert met gedrag dat relevant is voor de betreffende functie.

- *Er zijn meerdere onafhankelijk van elkaar werkende assessoren.* Om te voorkomen dat de geldigheid van de scores wordt aangetast door de subjectieve inkleuring van één beoordelaar, wordt alle gedrag door minimaal twee assessoren geobserveerd en beoordeeld. Bovendien krijgen assessoren een gedegen training waarin in ieder geval grote nadruk wordt gelegd op het onderscheid tussen objectief registreren en beoordelen van gedrag.
- *Er is een eindberaad waarin de assessoren hun bevindingen integreren in een eindoordeel.* Het is niet erg waarschijnlijk dat elke assessor tot een gelijklopende beoordeling komt van de kandidaat. In een zogeheten eindberaad gaan de assessoren met elkaar in discussie met het doel consensus te bereiken. Als niet aan al deze voorwaarden is voldaan, kan hooguit gesproken worden van een toepassing van de assessment center methode.

Het AC is dus een combinatie van verschillende instrumenten om in korte tijd veel en betrouwbare informatie te verzamelen met het oog op het nemen van beslissingen over personen. De methode bestaat al langer en is vooral na de tweede wereldoorlog populair geworden in het bedrijfsleven voor het aannemen van hogere managers. Ook nu wordt het

AC nog veelvuldig ingezet voor de werving van hogere managers, ook in het onderwijs (denk bijvoorbeeld aan het AC dat Interstudie Select gebruikt voor de selectie van schoolleiders). Langzaam maar zeker echter zien we een verschuiving optreden in de functie van 'selectie' naar 'ontwikkeling van potentieel'. In de literatuur worden AC's met laatstgenoemde doelstelling vaak aangeduid met de term Development Center (DC).

De werking van een AC wordt helder beschreven in Jansen (1993) waaruit de volgende tekst grotendeels letterlijk is overgenomen. Figuur 5 op pag. 46 geeft een schematische weergave van een AC. In een AC wordt gedrag uitgelokt dat een nabootsing is van reëel functiegedrag. Dat werkgedrag kan enerzijds beschreven worden als een verzameling van situatiespecifieke vaardigheden (competenties) die bijdraagt aan de realisering van de functie, anderzijds als de manifestatie van een specifieke set van persoonsgebonden eigenschappen, trekken of dimensies. In de situatietests die het gedrag uitlokken komen vaak meerdere, maar niet noodzakelijk alle, dimensies aan de orde. In het eindberaad worden dimensiescores bepaald door horizontale generalisatie over situatiescores en worden situatie-

scores bepaald door verticale generalisatie over dimensiescores. In een traditioneel AC, waarin de nadruk ligt op selectie van de meest geschikte persoon voor een functie, wordt meestal voor het eerste gekozen. Het AC levert dan een algemene persoonsbeschrijving op; een psychologisch rapport.

		SITUATIES			
		prioriteiten stellen	vergadering leiden	onderhandelen	ondernemingsplan opstellen
D I M E N S I E S	denkkracht				
	modelinge taalvaardigheid				
	organisatievermogen				
	mentale stevigheid				
	flexibiliteit				

Figuur 5. Schematische weergave van een AC

Dat het AC een succes is geworden heeft voor een belangrijk deel te maken met de grote voorspellingskracht. De vraag is dan ook niet of het werkt maar hoe. Daar is het laatste woord nog niet over gezegd. Wel wordt steeds weer opnieuw gevonden dat scores op dezelfde dimensie in verschillende situaties lager correleren dan scores in dezelfde situatie op verschillende dimensies (Jansen, 1991). Anders gezegd: het AC is verticaal sterker dan horizontaal. Kennelijk is het heel erg moeilijk om de persoonlijke eigenschappen (dimensies) zodanig te formuleren en te operationaliseren dat ze in elke situatie op dezelfde wijze geïnterpreteerd worden. Het AC heeft dus geen constructvaliditeit maar wel steekproefvaliditeit. Aan deze bevindingen kan men de conclusie verbinden dat hantering van dimensies om de prestaties in situatietests te scoren een omweg is. Je kunt beter direct de effectiviteit van gedrag vertoond in situatietests beoordelen. De constatering dat het AC alleen steekproefvaliditeit bezit, impliceert dat het persoonsplaatje dat eruit komt er een is in termen van situatiegebonden competenties; het AC levert een profiel van sterke en zwakke competenties niet van sterke en zwakke persoonlijke eigenschappen. De laatste tijd bespeuren we een verschuiving van de dimensionele naar de situationele benadering. Dat

wordt niet in de laatste plaats veroorzaakt door schaarste aan hoger opgeleiden in de arbeidsmarkt, waardoor organisaties gedwongen worden hun talent zelf op te leiden en doordat 'opleiding en ontwikkeling' meer en meer een management-gereedschap is geworden. De uitkomsten van een situationeel gericht AC zijn geformuleerd in termen van sterke en zwakke punten op het gebied van de verlangde competenties en geven dus richting aan activiteiten in het kader van HRM (human resources management). Om misverstanden te voorkomen: een AC dat bij de scoring de situationele benadering volgt, kan zowel voor selectie- als voor ontwikkelingsdoeleinden worden gebruikt. In het eerste geval worden uit een groep geteste kandidaten die personen geselecteerd waarvan het huidige functioneren als 'behorend tot de besten' gerekend kan worden. In het tweede geval is het competentieprofiel (als resultaat van de verticale generalisatie over de dimensies) de basis voor beslissingen over welke competenties (beter) ontwikkeld dienen te worden met het oog op de behoeften van de organisatie en de carrièreperspectieven van de kandidaat. AC's die voor dergelijke doelen worden ingezet, worden vaak een Development Center (DC) genoemd.

2.4.2 De portfolio methode

Assessment centers en development centers hebben vele tot de verbeelding sprekende voordelen. Ze worden goed ontvangen door de kandidaten, de einduitslag is een goede voorspeller voor het toekomstig functioneren en de profielscore geeft aanwijzingen voor de manier waarop het onderwijs in een competentiegericht curriculum vervolgd moet worden. Tegenover deze voordelen staan ook gevoelige nadelen:

- Logistiek gezien is het uitvoeren van een assessment center een enorme krachtsinspanning voor een opleiding.
- De ontwikkeling, exploitatie en onderhoud van het assessment center vergen grote sommen geld.
- De afname is geconcentreerd op één of enkele momenten in de opleiding en kan, wanneer een student bijvoorbeeld een slechte dag heeft, een vertekend beeld geven.
- De inhoud en procedures van het assessment center zijn gestandaardiseerd. Dit betekent dat kandidaten een tamelijk passieve rol hebben; het assessment center is iets dat je ondergaat; er zijn geen mogelijkheden voor kandidaten om de focus te richten op de sterke punten. Het portfolio komt

tegenoveraan genoemde bezwaren. De portfolio methodiek is een systematische procedure voor het verzamelen van documenten en werkstukken die tezamen het bewijs moeten vormen dat de eigenaar voldoet aan de minimumeisen die voor een bepaalde competentie zijn vastgesteld. Het portfolio is meer dan een beoordelingsmethode. Het vertegenwoordigt een visie op opleiden waarin studenten geen passieve consumenten zijn van hapklare brokken bereid door vakinhoudelijke experts maar actieve constructeurs van hun eigen opleidingstraject. Deze studenten worden niet alleen verantwoordelijk gehouden voor de verwerving van de competenties maar ook voor de wijze waarop het verwervingsproces het beste gestalte kan krijgen en voor het kritisch volgen van hun vorderingen (van Kammen, Melisse, & Oudkerk-Pool, 2000).

Portfolio's duiken inmiddels overal op waar wordt nagedacht over flexibele en op competenties gebaseerde opleidingstrajecten. Uiteraard wordt het concept steeds op een andere wijze geïnterpreteerd en toegepast, afhankelijk van de specifieke randvoorwaarden die bij concrete toepassingen gelden. Maar de volgende aspecten keren steeds terug en kunnen daarom beschouwd worden als definiërende kenmerken:

- Niet het curriculum maar de competenties zijn sturend voor het te verzamelen bewijs. Bewijsstukken mogen daarom gerust afkomstig zijn van 'buiten de opleiding'.
- Het portfolio kan meerdere doelen tegelijk dienen:
 - *Reflectie.* Voor de student is het primair een instrument op basis waarvan gereflecteerd kan worden op de eigen ontwikkeling. Het kan zeer motiverend werken om de opgaande lijn af te lezen uit de verzamelde bewijsstukken. Van de andere kant kunnen dezelfde bewijsstukken de noodzaak aantonen om aan bepaalde onderdelen meer energie te besteden of een andere koers te varen om ze onder de knie te krijgen.
 - *Formatieve en summatieve evaluatie.* De opleider kan de bewijsstukken in het portfolio zowel voor formatieve als voor summatieve beoordelingsdoelen gebruiken. Maar anders dan in traditionele curricula vervalt het strikte onderscheid hier-tussen. In de traditionele visie vinden formatieve beoordelingen tijdens de opleiding plaats met als doel sturing te geven aan het onderwijs en summatieve beoordelingen aan het eind van de opleiding ten einde beslissingen te nemen over zakken of slagen. Laatstgenoemde beoordelingen zouden aan zwaardere kwaliteitseisen moeten

voldoen, volgens een breed gedragen opvatting. In competentiegerichte curricula is het, zoals we hiervoor betoogd hebben, erg onverstandig beslissingen over zakken of slagen uitsluitend te baseren op evaluaties aan het eind van het opleidingstraject. Het ontbreekt dan meestal aan tijd om een voldoende brede steekproef aan taken te laten uitvoeren die een beslissing over competentie kan rechtvaardigen. Wordt daarentegen tijdens de gehele opleiding op gezette tijden informatie verzameld over het functioneren van de student in relatie tot de competenties, dan is het risico van een te beperkte steekproef aanzienlijk kleiner. Natuurlijk hebben de taken die in het begin van de opleiding worden uitgevoerd een gereduceerde representativiteit en zijn ze minder complex en moeilijk dan taken die aan het eind aangeboden worden, maar dit valt eenvoudig te compenseren door een systeem van gewogen scoring toe te passen bij de finale beoordeling. Een mooi voorbeeld van portfolio gebruik waarbij het harde onderscheid tussen formatieve en summatieve doeleinden is komen te vervallen is de invulling die de Educatieve Faculteit Amsterdam eraan gegeven heeft (van Kammen, Melisse, & Oudkerk-Pool, 2000).

In deze faculteit is men hard op weg competentie-gerichte curricula in te richten met daarbinnen mogelijkheden tot zelfstandig en zelfverantwoordelijk leren. Het portfolio is in de eerste plaats een hulpmiddel voor de student voor reflectie en zelfsturing. Maar hetzelfde bewijsmateriaal wordt ook gebruikt voor harde beslissingen op gezette tijden in het curriculum. Zie paragraaf 4.3 voor een beschrijving en beoordeling van deze aanpak.

- *Intake en EVC.* Portfolio's kunnen ook gebruikt worden om intakebeslissingen te nemen. Het stramen van het portfolio kan potentiële studenten behulpzaam zijn bij het inventariseren en beschrijven van bewijs op basis waarvan de geschiktheid voor het beroep beoordeeld kan worden en eventueel vrijstellingen kunnen worden verleend. Heel nadrukkelijk moet hierbij gewezen worden op het nut van de veelal biografische gegevens, ook wel aangeduid met de term 'biodata', waarmee portfolio's gevuld kunnen worden tijdens de intakefase. Ackley & Arwood (1999) vonden bewijs dat ondersteuning verleent aan de opvatting van Smith & Pratt (1996) dat de toelating tot een lerarenopleiding niet alleen zou moeten berusten op de vraag of voldaan wordt

aan de vooropleidingseis maar ook op de vraag of men ervaring heeft met het geven van lessen. Eerstgenoemde auteurs concludeerden het volgende uit hun onderzoek:

'...If applicants have a long held occupational vision of themselves as teachers and they have had one or more previously succesful experiences teaching groups of young people in classroom like settings, teacher education admission officers should give serious consideration to such students for admission to their programs.'

- Het werken aan het portfolio maakt de student bewust van het feit dat hij zelf verantwoordelijk is voor zijn leerproces. Daartoe reflecteert de student op gezette tijden op het verzamelde bewijs en schrijft hij op basis daarvan een zelfevaluatie. Reflectie is een belangrijke activiteit in een onderwijsproces waarin studenten zelf verantwoording dragen voor hun leren. Richert (1990) vond dat reflectie de beste resultaten opleverde op momenten dat de betreffende personen bezig waren om materiaal te verzamelen of te bewerken voor opname in hun portfolio. Bij het maken van keuzes daarbij dachten ze terug aan wat ze deden tijdens het lesgeven, aan de materialen die ze gebruikten en aan hoe

effectief die materialen waren voor de te bereiken doelen. In het tekstkader hieronder is een voorbeeld opgenomen van een reflectief verslag van een pabostudent naar aanleiding van de eerste ervaringen binnen de oriënterende module 'didactiek van rekenen/wiskunde'. De lezer dient zich bij het lezen van de tekst te realiseren dat het reflectief verslag uitgangspunt is voor het beoordelen van de competenties door zowel de bezitter van het portfolio als diens begeleider. Als zodanig heeft het een duidelijker functie dan het meer bekende stageverslag, dat vaak als een verplicht nummer zonder consequenties wordt beschouwd.

“Aan het begin van deze module was ik vrij sceptisch wat het vak rekenen/wiskunde betreft. Ik dacht, "dat doe je wel even". Ik wist dat ik moeite met wiskunde had op de middelbare school, maar ik dacht dat rekenen nu echt op basisschoolniveau was. Hoofdrekenen en tafels leren ging me vroeger altijd al goed af. Alleen op de middelbare school werd het steeds moeilijker. Vooral de laatste jaren was wiskunde mijn slechtste vak. Mijn kennis van basisschool rekenen was dus vrij goed dacht ik zo. Ik verwachtte dus geen problemen bij dit modul. Ik wist wel dat het ook in dit vak moeilijk was om uitleg te geven, dus daar zag ik wel

tegenop. Na dit modul ben ik erachter gekomen, dat het rekenen nu ook nog een groot probleem is. Er komt zoveel bij kijken, dat weet ik zelf vaak niet eens meer. Het is al een tijd geleden, dus dat maakt het extra moeilijk. Ook het lesgeven vind ik vrij pittig. Leg maar eens uit dat 1 en 1 twee is en niet drie. Het lijkt allemaal zo vanzelfsprekend. Maar voor de kinderen en vooral die, waar rekenen net nieuw is, die snappen dat op het begin gewoon niet. Ik liep eerst stage in groep 6. Nou, daar konden ze al wat van rekenen. Toch vroeg ik mij nog vaak genoeg af, "waarom weten ze niet dat 6 keer 6 36 is, dat is toch logisch". Dat moet je bij jezelf gaan kijken en je dan afvragen, is het wel zo logisch. En dan kom je tot de conclusie: ja, voor mij NU wel, dankzij een goede uitleg en een goed geduld van mijn leraar vroeger. En datzelfde moet ik nu ook hebben. Je moet je steeds verplaatsen in het kind en daar had ik van tevoren nooit aan gedacht.

Nu ik in groep 1 stage loop zie ik dat ook en gelukkig weet ik me al een beetje in te leven in de kinderen. En wat ik ook nooit gedacht had; er wordt hier meer met rekenen gedaan dan ik had verwacht. Daar sta je nooit bij stil. Over kleuters had ik eerst niet zo'n hoge hoed op, "daar kun je toch niets mee doen", dacht ik. Maar ook daarin heb ik me vergist.

Wat me ook opgevallen is, is dat je veel beter in kleine groepen les kunt geven. Dat gaat me ook veel beter af. In groep 6 had ik vier kinderen, die wat achter liepen op de rest van de klas. Ik mocht deze kinderen iets nieuws aanleren, in een ander lokaal. Na een goede uitleg en ook veel helpen bij de eerste sommen ging het al vrij aardig. De kinderen krijgen zo veel meer aandacht en ze kunnen ook beter vragen stellen. En vaak stellen ze in zo'n groepje sneller vragen, dan in de klas. Want daar worden ze nog wel eens uitgelachen. Toen ik een les gaf bij de grote groep, was het veel onduidelijker. De klas was rumoerig, omdat sommige mensen dat al wisten en ik moet dus ook vaker uitleg geven, voor diegene, die het moeilijker ermee hadden. Mijn voorkeur gaat dus absoluut uit naar de kleinere groepen. Dat is ook iets wat ik van tevoren nooit had verwacht. Ik dacht altijd, dat lesgeven in de grote klas beter was. Waarom? Weet ik eigenlijk niet. Het is misschien zo, omdat ik het zo gewend was. Op mijn basisschool werd nooit in groepjes gewerkt. En in principe denk je toch altijd, dat 'jouw' basisschool toch wel de goede is.

Na dit moduul en beide stageperiodes, zie ik mijzelf nog niet als een goede wiskunde leraar. Ik denk, dat ik zelf de theorie nog niet 100% beheers om alles goed te doen. Het ligt vooral aan mijzelf om dit te veranderen.

Mijn leraar kan mij daar natuurlijk wel bij helpen, d.m.v. literatuur aan te bieden en mij te helpen bij al mijn vragen. Tot nu toe kan ik me redelijk redden, maar echt een goede rekenles te geven, dat zie ik nog niet zo zitten. Maar ik denk dat dat nog wel komt.

De kern van deze module is. Het zien van goede mogelijkheden van leertechnieken voor het vak wiskunde en laten zien wat het vak allemaal te bieden heeft. Voor mij is een ding heel duidelijk. Wiskunde is niet alleen tafels en hoofdreken, maar een heleboel erom heen. Want je soms niet verwacht heeft toch heel veel met rekenen te maken. En als leraar moet je heel veel geduld hebben en of opbrengen om je leerlingen alles goed te laten begrijpen. Voor mezelf weet ik ook nog dat ik nog veel te doen heb aan de theorie om tot een goede leraar te komen.

Wat ook heel leuk was; ik had nooit gedacht dat je zoveel onbewust rekent. Je let daar eigenlijk helemaal niet op. Wanneer je opstaat ben je al met reken bezig. Je zet je wekker en even later gooi je alweer twee klontjes suiker in de koffie, waarom? Omdat je het altijd zo doet. Die dingen zijn zo logisch, maar als je nadenkt wordt het zo onlogisch. En dat vindt ik eigenlijk wel leuk.

En de theorie natuurlijk. Deze was veel moeilijker dan ik had verwacht. Het is toch even wat meer als alleen maar tafels, hoofdreken of cijferen. Rekenen is overal. Er zijn zoveel methoden, die ik ooit heb geleerd en die weet je nu niet meer. En daar moet ik nog veel aan leren.

**Reflectief verslag van een eerstejaars pabo-student.
Bron: Bert Imandt, pabo Hogeschool Zuyd.**

Het portfolio heeft enkele duidelijke voordelen boven het assessment center: Het is minder een momentopname dan het AC; het is organisatorisch minder belastend en het plaatst de student in een actieve, verantwoordelijke rol (je moet zelf aantonen dat er sprake is van competentie). Nadelen zijn er ook. Een groot risico is het gedachteloos en ongestructureerd verzamelen van alles waarvan men maar enigszins vermoedt dat het misschien wel iets te maken kan hebben met de betreffende competenties. Het is uiteindelijk wel aan de student om een selectie te maken uit zijn portfolio die een weerspiegeling is van zijn competenties, maar als het portfolio niet van meet af aan doelgericht is aangelegd, zal dit een

uiterst lastig karwei blijken te zijn.

Een tweede probleem heeft te maken met de betrouwbaarheid van de beoordeling op grond van het portfolio. De bewijsstukken in een portfolio kunnen zeer divers zijn, zowel naar inhoud als naar authenticiteit. Het trekken van verkeerde conclusies is groot als 'appels en peren' bij elkaar worden opgeteld. Mogelijk kan dit risico verkleind worden door van studenten te eisen elk bewijsstuk te beschrijven in termen van de competenties die het dekt en tevens aan te geven hoe het bewijs tot stand is gekomen (in welke situatie is wat gedaan en wie was daarbij om een beoordeling te geven?). Met die additionele informatie is de beoordelaar beter in staat de bewijsstukken naar waarde te schatten en te komen tot een competentieprofiel als uitgangspunt voor het beoordelingsgesprek met de andere beoordelaars en de kandidaat zelf.

3 · Aspecten en kwaliteitscriteria van competentie-instrumenten

Dit hoofdstuk bestaat uit twee onderdelen. In het eerste onderdeel worden aan de hand van vijf vragen vijf aspecten van competentie-instrumenten voor het beoordelen van docenten besproken. Zowel voor degenen die instrumenten willen ontwikkelen als degenen die reeds ontwikkelde instrumenten willen evalueren zijn de antwoorden op die vijf vragen van belang. Kwaliteitscriteria voor de evaluatie van competentie-instrumenten worden in het tweede onderdeel van het hoofdstuk besproken.

3.1 Aspecten van competentie-instrumenten voor docenten

3.1.1 Wie wordt beoordeeld?

De bespreking van competentie-instrumenten voor het beoordelen van docenten in dit hoofdstuk heeft betrekking op drie groepen (aspirant-)docenten:

- Zij-instromers. Hiermee worden sinds kort personen aangeduid die vanuit een ander beroep willen overstappen naar een baan in het primair of secundair onderwijs. Voor deze maar ook voor andere groepen geldt dat als gevolg van het tekort aan docenten andere opleidingsroutes naast de bestaande vol- en deeltijdopleidingen voor docenten mogelijk zijn. Voor de zij-instromers geldt dat zij bij gebleken geschiktheid direct voor de klas kunnen gaan staan.
- Studenten met minimaal HAVO- of MBO-diploma die een vierjarige HBO-voltijdopleiding volgen.
- Praktiserende, beginnende en ervaren docenten.

3.1.2 Waartoe wordt beoordeeld?

In het voorgaande hoofdstuk werden vier functies van competentie-instrumenten onderscheiden die omwille van de volledigheid van dit hoofdstuk hier nog eens kort herhaald worden:

- Nemen van instroom- of selectiebeslissingen. Hiervan is bijvoorbeeld sprake bij competentie-instrumenten voor zij-instromers waarmee op basis van een geschiktheidsonderzoek bepaald wordt of een kandidaat met of zonder extra scholing direct voor de klas kan gaan staan.
- Nemen van doorstroom- of voortgangsbeslissingen. Hiervan is bijvoorbeeld sprake als een student in de propedeuse van de lerarenopleiding beoordeeld wordt op de vraag of hij/zij hoofdfasebekwaam is.
- Nemen van uitstroom- of certificeringsbeslissingen. Hiervan is bijvoorbeeld sprake als een student van de lerarenopleiding aan het eind van de vierjarige opleiding beoordeeld wordt op het al of niet startbekwaam zijn.
- Nemen van beslissingen in het kader van professionele ontwikkeling. Hiervan is bijvoorbeeld sprake wanneer praktiserende docenten op basis van een zelfbeoordeling of een beoordeling door het schoolmanagement zich gaan scholen ter bevordering van de kwaliteit van het onderwijs op de school. Onderwijsinstellingen doen steeds meer aan kwaliteitszorg en het beoordelen van competenties van docenten gekoppeld aan een scholingsaanbod is een belangrijk onderdeel van die kwaliteitszorg.

3.1.3 Wat wordt beoordeeld?

Om competentie-instrumenten te kunnen ontwikkelen waarmee de geschiktheid of bekwaamheid van een zij-instromer, student of praktiserende docent beoordeeld kan worden, moet men weten welke competenties als essentieel beschouwd worden voor het uitoefenen van het beroep docent. Het document 'Startbekwaamheden leraar primair onderwijs (1997)' is een goed uitgangspunt voor het bepalen van wat de relevante competenties voor een docent in het primair onderwijs zijn. Aan dit document zullen hierna vele passages nagenoeg letterlijk ontleend worden.

De startbekwaamheden leraar primair onderwijs worden onderscheiden in vakspecifieke bekwaamheden en algemene startbekwaamheden. De startbekwaamheden van een specifiek vak, bijvoorbeeld rekenen/wiskunde, beginnen met een typering van de taak van een beginnende leraar en geven daarnaast een viertal bekwaamheidseisen die betrekking hebben op het eigen niveau van de beginnende leraar en zijn bekwaamheden met betrekking tot vakinhoud en vakdidactiek. Om onderwijs in de verschillende vakken te kunnen

geven, moet een leraar ook een aantal algemene startbekwaamheden beheersen. Het is duidelijk dat niet bij alle vakken in dezelfde mate een beroep gedaan wordt op dezelfde algemene startbekwaamheden.

In voornoemd document wordt een algemene startbekwaamheid beschreven door middel van een kernuitspraak en een specificatie van die kernuitspraak. De kernuitspraken hebben betrekking op: kenmerken en ontwikkeling (van de leerling), voorbereiding van onderwijs (door de leraar), creatie en organisatie van leeromgevingen, didactische werkvormen en mediagebruik, adaptieve begeleiding, pedagogisch handelen en evaluatie. Voor evaluatie zijn bijvoorbeeld drie kernuitspraken geformuleerd. De eerste luidt 'De beginnende leraar is in staat de resultaten van het onderwijs en de ontwikkelings- en leerprocessen te evalueren en te waarderen.' Deze kernuitspraak wordt nader gespecificeerd als 'Dat betekent dat de beginnende leraar: 1. inzicht heeft in principes en werkwijzen om resultaten van ontwikkelings- en leerprocessen vast te stellen; 2. over vaardigheden beschikt om recapitulatie van leerinhouden, toetsing van resultaten, observatie van verworven handelingen en ontwikkeld gedrag en

waardering van de uitslag uit te voeren.' De twee andere kernuitspraken luiden 'De beginnende leraar is in staat de kwaliteit van leerprocessen van de leerlingen vast te stellen' en 'De beginnende leraar is in staat gegevens over processen en resultaten van leerlingen in te voeren in een leerlingvolgsysteem en daarover met anderen te communiceren en daaruit conclusies te trekken.'

Behalve algemene en vakspecifieke startbekwaamheden leraar primair onderwijs, zijn in het document ook kenmerkende en kritische situaties voor het werk van een beginnend docent toegevoegd. Kenmerkende situaties zijn situaties waarvan het al dan niet beheersen wel van belang is voor het functioneren als docent, maar niet direct het succes of respectievelijk de mislukking als docent inhoudt. Kritische situaties zijn situaties waarvan het al dan niet beheersen ervan het succes respectievelijk de mislukking van het functioneren als docent inhoudt. In totaal worden elf kenmerkende en kritische situaties (gemarkt met een asterisk) onderscheiden:

- thematisch werken
- verschillende onderwijsaanpakken hanteren*
- werken met onderwijsconcepten en leerlijnen
- omgaan met leerproblemen*

- omgaan met groepsprocessen
- handelen in onverwachte situaties
- intercultureel onderwijs realiseren
- werken met nieuwe media
- contacten met ouders onderhouden*
- collegiale hulp gebruiken*
- collegiaal samenwerken

Zoals hiervoor is opgemerkt, vormen de startbekwaamheden een goed uitgangspunt om te komen tot een inventarisatie van de essentiële competenties die beoordeeld moeten worden. Om die inventarisatie te verkrijgen dient een competentie-analyse uitgevoerd te worden. Een competentie-analyse bestaat gewoonlijk uit drie stappen (zie ook Van Beirendonck, 1999, p. 46):

1. Het determineren van de essentiële competenties die voor het succesvol kunnen uitoefenen van een beroep, in dit geval docent, nodig zijn. Deze stap van de analyse bestaat uit het bestuderen van literatuur zoals de startbekwaamheden, het raadplegen van andere competentieanalyses, interviews met docenten en opleiders, het observeren van docenten etc.

2. Het omschrijven van de competenties in gedrags-indicatoren oftewel concreet observeerbaar gedrag.
3. Het van alle competenties vastleggen van de beheersingsstandaard, bijvoorbeeld wat als onvoldoende, voldoende en uitstekende beheersing aangemerkt moet worden.

3.1.4 Hoe wordt beoordeeld?

In het voorgaande hoofdstuk werden verschillende instrumenten of methoden besproken waarmee competenties van docenten beoordeeld kunnen worden. Voor het antwoord op de vraag welke concrete instrumenten in aanmerking komen of de voorkeur verdienen, zijn de competenties die beoordeeld moeten worden uiteraard de bepalende factor. Afhankelijk van de aard van de competenties en het aantal competenties dat beoordeeld moet worden, zal soms volstaan kunnen worden met de inzet van één instrument maar zal in de meeste gevallen de inzet van meerdere instrumenten gewenst zijn. Voor de bepaling van welke instrumenten voor welke competenties ingezet gaan worden, kan een competentiematrix goede diensten bewijzen. In een competentiematrix geeft men aan welke competenties met welke instrumenten beoordeeld worden.

Onderstaande tabel, ontleend aan Van Beirendonck (1999, p. 55), bevat het voorbeeld van een competentiematrix voor een commerciële functie in een telecommunicatiebedrijf. In deze competentiematrix staan de competenties die beoordeeld worden in de rijen van de matrix, terwijl in de kolommen de 'instrumenten' staan waarmee de competenties beoordeeld worden.

	In-basket	Analyse & presentatie	Groepsdiscussie	Gesprek klant	Interview	Test
PERSOONSGEBONDEN GEDRAG						
Stressbestendigheid		X		X		X
Optreden		X			X	
Prestatiemotivatie	X				X	X
INTERPERSOONLIJK GEDRAG						
Mondelinge communicatie		X	X			
Overtuigingskracht			X	X		
Empathie				X	X	
Sociale vlotheid			X	X		X
PROBLEEMVERWERKEND GEDRAG						
Probleemanalyse	X		X			
Oordeelsvorming	X	X	X			
Flexibiliteit	X		X	X		X

PROBLEEMOPLOSSEND GEDRAG						
Klantgerichtheid	X			X	X	
Initiatief	X		X			X
Creativiteit		X	X			
BEHEERSMATIGE VAARDIGHEDEN						
Plannen en organiseren	X		X			

De competentiematrix van een assessment-programma geeft een overzicht van welke competenties met welke instrumenten beoordeeld worden en kan dan ook goede diensten bewijzen bij het maken van een verantwoorde keuze van een assessment-programma. Met behulp van de competentiematrix kan men bijvoorbeeld nagaan of ook rekening gehouden is met aanbevelingen dat men eenzelfde competentie minstens tweemaal door verschillende instrumenten moet laten beoordelen.

3.1.5 Door wie wordt beoordeeld?

Bij nagenoeg alle instrumenten die ingezet worden voor de beoordeling van docentcompetenties, zijn beoordelaars of assessoren betrokken. In de litera-

tuur over het beoordelen van competenties wordt aan het onderwerp assessoren dan ook uitgebreid aandacht besteed. De bespreking in deze paragraaf beperkt zich tot wat van de assessor verwacht wordt, welke fouten beoordelaars kunnen maken bij het beoordelen en over welke eigenschappen beoordelaars zouden moeten beschikken.

Tijdens het beoordelen van (aspirant-)docenten moeten de assessoren een aantal taken verrichten. Deze taken die door Van Beirendonck (1999, p.62-68) de vijf essentiële O's genoemd worden, zijn:

1. Observeren: het zo nauwkeurig mogelijk observeren van het gedrag van de docent. Met gedrag wordt bedoeld dat wat door de docent gezegd en gedaan wordt.
2. Opschrijven: het zo nauwkeurig mogelijk opschrijven van het geobserveerde gedrag van de docent. De taken observeren en opschrijven moeten nauwgezet uitgevoerd worden omdat het de taken in het beoordelingsproces zijn die veelal niet opnieuw gedaan kunnen worden.
3. Omzetten: het aangeven van welke concrete gedragsobservaties een uitdrukking zijn van welke competentie(s).
4. Oordelen: het beoordelen van de mate waarin de docent over de competentie beschikt. De beoordeling wordt meestal gekwantificeerd met behulp van schalen. Schalen met weinig schaalpunten verdienen in het algemeen de voorkeur boven schalen met veel schaalpunten. Deze laatste suggereren veelal ten onrechte dat fijne onderscheidingen gemaakt kunnen worden en leiden dan ook vaak tot (schijnbaar) verschillende oordelen van assessoren.
5. Overleggen: op basis van de oordelen die de assessoren toegekend hebben aan de competenties die met de verschillende instrumenten

beoordeeld zijn, moet een eindoordeel over de docenten gegeven worden. Om tot dat eindoordeel te komen, stelt Van Beirendonck (1999, p. 67) het volgende voor:

- Bespreek per instrument de oordelen van alle assessoren op de afzonderlijke competenties en stel in overleg een definitief oordeel per competentie vast.
- Stel in overleg een oordeel vast over elke competentie over alle instrumenten heen.
- Stel op basis van het voorgaande in overleg het definitieve eindoordeel vast.

Het beoordelen van het gedrag van anderen is geen eenvoudige taak waarbij veel fouten gemaakt (kunnen) worden. Wanneer assessoren kennis hebben van de fouten die ze kunnen maken, zijn die wellicht te vermijden. Beoordelingsfouten (zie onder andere Van den Broek e.a. , 2000, p. 44-47) die vaak door assessoren gemaakt worden zijn:

- Centrale tendentie: geen extreem hoge of extreem lage scores geven maar voornamelijk het midden van de beoordelingsschaal gebruiken.
- Halo-effect: door de positieve beoordeling van (aspecten van) een enkele competentie ook de (aspecten van) andere competenties positief beoordelen.

- Horn-effect: door de negatieve beoordeling van (aspecten van) een enkele competentie ook (aspecten van) de andere competenties negatief beoordelen.
- Generaliseren: op grond van slechts een enkele gedraging een algemeen oordeel geven.
- Strengheid en mildheid: consequent streng of mild beoordelen.
- Projectie: eigen gedrag tot norm verheffen, met name in het geval dat de leidinggevende de assessor is.
- Vooroordelen of stereotiepen: ervan uitgaan dat bepaalde eigenschappen of kenmerken van personen bij elkaar horen.
- Selectieve waarneming: zien of horen wat men wil zien of horen.
- Toegeeflijkheid: eerder hoger dan lager beoordelen, met name in het geval dat de leidinggevende de assessor is.
- Nabijheid: medewerkers waar men meer contact mee heeft hoger beoordelen, met name in het geval dat de leidinggevende de assessor is.
- Doelredenering: naar een eindresultaat toe redeneren om te komen tot een (te) positieve of negatieve beoordeling.

- Vroegere oordelen: vroegere oordelen een rol laten spelen.
- Relatie: ten onrechte positief beoordelen om de relatie niet te verstoren.
- Recentheid: te sterk laten meewegen van een recente gebeurtenis.
- Eerste indruk: laten meewegen van een eerste indruk.

Bij het formuleren van de eigenschappen waar over assessoren zouden moeten beschikken, nemen we de meest complexe beoordelingssituatie, namelijk de situatie waarbij de competenties van praktiserende docenten beoordeeld worden, als uitgangspunt. Als assessoren voor de beoordeling van praktiserende docenten komen in aanmerking: leidinggevend, hoofd vaksectie en collega's. Daarnaast kunnen bepaalde competenties ook door leerlingen beoordeeld worden. Niet alle competenties kunnen door alle in aanmerking komende assessoren beoordeeld worden. Zo zullen het hoofd van de vaksectie en collega's het beste kunnen beoordelen of een docent als vraagbaak en voorbeeld voor anderen fungeert, terwijl bij de beoordeling van de mate waarin een docent bij zijn lesgeven inspeelt op de leefwereld van de leerlingen het oordeel van de

leerlingen niet zou mogen ontbreken. Collega's en leerlingen zijn niet onafhankelijk in de zin van onafhankelijk van degenen die beoordeeld worden. Wanneer afhankelijke assessoren zoals collega's ingezet (moeten) worden, zal assessorentraining er voor dienen te zorgen dat de hiervoor genoemde beoordelingsfouten niet gemaakt worden. Bovendien zullen assessoren van praktiserende docenten over de volgende eigenschappen dienen te beschikken:

- Geloofwaardig: de assessor kent de docent en is bekend met de context waarin de docent werkt.
- Evident competent: de vakinhoudelijke competentie van de assessor is van een excellent niveau.
- 'Voorbeeldig': de assessor kan laten zien hoe het (wel) moet.
- Bekwaam: de assessor moet een bekwaam beoordelaar zijn.
- Overtuigend: de beoordeling van de assessor wordt door de docent geaccepteerd.
- Accuraat: de assessor trekt voldoende tijd uit voor de beoordeling.
- Betrouwbaar: de assessor bespreekt de beoordeling alleen met de betrokken docent en andere personen die bij de beoordeling betrokken zijn.

3.2 **Kwaliteitscriteria voor competentie-instrumenten voor docenten**

Competentie-instrumenten voor docenten moeten voldoen aan bepaalde kwaliteitscriteria. In de literatuur over zogenaamde authentieke instrumenten waartoe ook competentie-instrumenten voor docenten behoren, worden vele kwaliteitscriteria opgevoerd (zie ook Dierick e.a., 2001.) Onderstaande bespreking beperkt zich tot twee groepen kwaliteitscriteria namelijk psychometrische en procedurele kwaliteitscriteria. Onder psychometrische kwaliteitscriteria worden hier de validiteit en betrouwbaarheid van competentie-instrumenten bedoeld, terwijl de procedurele kwaliteitscriteria betrekking hebben op de uitvoerbaarheid en transparantie van competentie-instrumenten. In hoofdstuk 4 zullen bij het evalueren van instrumenten die momenteel voor het beoordelen van docentcompetenties gebruikt worden in voorkomende gevallen ook nog andere kwaliteitscriteria betrokken worden.

3.2.1 Psychometrische kwaliteitscriteria

VALIDITEIT

Bij de validiteit van een competentie-instrument gaat het om de vraag of dit aan zijn doel beantwoordt. Het doel van een instrument voor docentcompetenties is na te gaan of zij-instromers, studenten of docenten over de door de lerarenopleiding of de overheid gewenste competenties beschikken. Het valideren van een competentie-instrument bestaat er dan uit dat men gegevens verzamelt om te bewijzen dat de betreffende competenties gemeten worden. In dit verband verdient het onderzoek van Jansen (1993), uitgevoerd bij assessment centers voor managers, navolging. Uit dit onderzoek bleek dat verschillende competenties binnen een bepaald instrument, bijvoorbeeld de opdracht om een vergadering te leiden, wat als een kritische situatie aangemerkt kan worden, meer met elkaar samenhangen dan met dezelfde competenties gemeten met een ander instrument. Voor de competentiematrix van een commerciële functie die hiervoor gepresenteerd is, betekent dit dat 'overtuigingskracht' in een groepsdiscussie niet hetzelfde is als 'overtuigingskracht' in een gesprek met de klant en dat er dus geen

sprake is van dezelfde competentie. Zou men voor competentie-instrumenten voor docenten hetzelfde vinden, dan betekent dit, zowel voor competentie-instrumenten voor docenten als voor het onderwijs aan docenten, dat beter uitgegaan kan worden van kritische situaties dan van competenties.

De authenticiteit van de opdrachten die de beoordeelden moeten uitvoeren, wordt terecht als een sterk punt van competentie-instrumenten aangemerkt. Uit de bespreking in hoofdstuk 2 bleek echter dat aan authentieke opdrachten niet alleen voordelen maar ook nadelen verbonden zijn. Zonder nader onderzoek moeten daarom beweringen als '...hoe meer een toets aansluit bij de werkelijkheid die hij beoogt te benaderen, hoe groter de validiteit' (Elsen e.a., 2001, p. 36) dan ook betiteld worden als voorbarig.

BETROUWBAARHEID

Bij een betrouwbaarheidsstudie, of beter gezegd een generaliseerbaarheidsstudie (Sanders, 1998), van een competentie-instrument gaat het om de vraag hoe nauwkeurig de docentcompetenties gemeten kunnen worden met een bepaald competentie-instrument. Bij een competentie-instrument heeft men te maken met studenten of (aspirant-)docenten die opdrach-

ten moeten uitvoeren welke door assessoren beoordeeld worden. Hoe betrouwbaar of onbetrouwbaar een competentie-instrument is, kan pas bepaald worden nadat het competentie-instrument afgenomen is. Vooruitlopend op de bespreking in hoofdstuk 4 van competentie-instrumenten die momenteel beschikbaar zijn, kan opgemerkt worden dat er bij geen van die competentie-instrumenten onderzoek naar de betrouwbaarheid of generaliseerbaarheid heeft plaatsgevonden.

De betrouwbaarheid kan op twee verschillende manieren berekend worden. Bij de eerste manier gaat men uit van de scores die door verschillende assessoren aan dezelfde competenties op verschillende instrumenten toegekend zijn. Bij de tweede manier wordt eerst de betrouwbaarheid van elk instrument afzonderlijk berekend en daarna de betrouwbaarheid van alle instrumenten samen. Gezien wat hiervoor bij validiteit naar aanleiding van onderzoek bij assessment centers voor managers opgemerkt is, lijkt de tweede manier de voorkeur te verdienen. Dit betekent dat men zich beperkt tot het berekenen van de betrouwbaarheid van individuele instrumenten en die zal afhankelijk van het aantal 'items' of beoordelingspunten meer of minder hoog zijn. Berekening van de betrouwbaarheid van

alle instrumenten samen lijkt weinig zinvol indien (na onderzoek zou blijken dat) de verschillende instrumenten verschillende competenties meten. Wat betreft de toekenning van kwalificaties zoals een lage of hoge betrouwbaarheid zou men rekening moeten houden met de consequenties die de beslissingen voor de beoordeelden hebben. Van competentie-instrumenten die een belangrijke rol spelen bij de beslissing een student al of niet te certificeren, zou men een hogere betrouwbaarheid kunnen of moeten eisen dan van competentie-instrumenten voor het vaststellen van het competentieniveau van een docent in het kader van diens professionele ontwikkeling.

3.2.2 Procedurele kwaliteitscriteria

UITVOERBAARHEID

Bij dit criterium gaat het om de vraag of het competentie-instrument niet te ingewikkeld is om af te nemen of niet onevenredig veel tijd in beslag neemt van degenen die beoordeeld worden of van de assessoren.

TRANSPARANTIE

Bij dit criterium gaat het om de vraag of de beoordelers voldoende ingelicht zijn over de verschillende instrumenten die ingezet worden, de beoordelaars en de beoordelingscriteria.



4 · Beschrijving en evaluatie van competentie-instrumenten

In dit hoofdstuk worden competentie-instrumenten beschreven en geëvalueerd die momenteel ingezet worden bij de beoordeling van zij-instromers, studenten en docenten. Het eerste instrument, 'Bewezen Geschikt', is bedoeld voor de beoordeling van de competenties van zij-instromers. Het tweede instrument, 'Bekwaam Beoordeeld', wordt ingezet bij de beoordeling van pabostudenten in verschillende fasen van de opleiding maar met name bij derdejaars die zich voorbereiden op de LIO-stage. De twee andere instrumenten, het 'Integratief BeoordelingsMoment' en 'Als Docent in Beeld', zijn bedoeld om te beoordelen of eerstejaars studenten van de lerarenopleiding kunnen worden toegelaten tot de hoofdfase. Tenslotte wordt ingegaan op instrumenten die naar verwachting ingezet gaan worden bij de competentiebeoordeling van praktiserende docenten. Opgemerkt dient te worden dat de beschrijving van genoemde instrumenten uitsluitend gebaseerd is op schriftelijke documenten over die instrumenten en

niet op observaties van afnamen van de instrumenten. Uit de documenten zijn vele passages (noodzakelijkerwijs) letterlijk overgenomen of geparafraseerd. Het voorgaande is een van de redenen waarom bij de evaluatie van de instrumenten geen eindoordeel gegeven wordt. Een andere reden is dat de instrumenten nog steeds in ontwikkeling zijn en definitieve uitgaven van de instrumenten dan ook niet beschikbaar zijn.

4.1 Competentie-instrument 'Bewezen Geschikt'

4.1.1 Beschrijving van 'Bewezen Geschikt'

'Bewezen Geschikt' (Klarus e.a., 2000) is de naam van een assessmentprocedure voor zij-instromers in het primair en voortgezet onderwijs. Zowel de beschrijving als de evaluatie zijn gebaseerd op de instrumentenmap en het bijbehorende rapport voor het primair onderwijs.

WIE WORDT BEOORDEELD?

Met het assessment worden personen met een hbo-opleiding, aangeduid als zij-instromers, beoordeeld die in het primair onderwijs willen werken maar daartoe niet bevoegd zijn.

WAARTOE WORDT BEOORDEELD?

Het assessment is in de eerste plaats bedoeld om na te gaan of een zij-instromer geschikt is om (met of zonder extra scholing) direct voor de klas te gaan staan. In de tweede plaats dient het assessment informatie op te leveren over de vraag welke competenties aanwezig zijn en welke competenties nog verder ontwikkeld moeten worden. In de derde plaats dient het assessment informatie op te leveren over het gewenste vervolg in de verdere ontwikkeling van de kandidaat. Het gaat bij dit instrument dus om het nemen van instroombeslissingen.

WAT WORDT BEOORDEELD?

Het assessment voor zij-instromers beoogt de geschiktheid van kandidaten te beoordelen om (met of zonder extra scholing) direct voor de klas te gaan staan. Uitgaande van de startbekwaamheden voor leraar primair onderwijs is een lijst bestaande uit

vier clusters met in totaal tien startcompetenties vastgesteld die als de gewenste startcompetenties beschouwd worden voor kandidaten die docent willen worden. Het accent ligt op de rol als leraar of begeleider en als lesgever.

Hieronder staan de tien startcompetenties met tussen haakjes achter elke startcompetentie het aantal observeerbare gedragsindicatoren dat gebruikt wordt voor de beoordeling van de startcompetentie:

Handelen in het onderwijsleerproces

1. Vermogen tot het begeleiden van leerlingen in hun leerproces (11).
2. Vermogen tot het overdragen van informatie (4).
3. Vermogen tot communicatie en interactie (4).
4. Vermogen klassenorganisatie en groepsmanagement (8).

Algemeen professioneel handelen

5. Vakdeskundigheid (7).
6. Beroepshouding (7).
7. Vermogen tot reflectie en initiatief (5).
8. Vermogen tot ontwikkelen van eigen professionele deskundigheid (3).

Werken binnen een organisatie

9. Met anderen in een team samenwerken (8).

Kennis van het onderwijs in het algemeen

10. Kennis van het onderwijs (3).

Ter illustratie staan hieronder voor de startcompetentie 'vakdeskundigheid' de zeven gedragsindicatoren die het resultaat zijn van de tweede stap van de competentieanalyse (zie paragraaf 3.1.3):

1. Is in staat een goede opbouw in de leerstof aan te geven.
2. Kan leermaterialen (boeken, CD-ROMs, etc.) op de juiste manier toepassen.
3. Kan leerproblemen binnen het vakgebied analyseren en adequaat daarop inspelen met gerichte opdrachten en/of vragen.
4. Kan leerstof/leeractiviteiten aanpassen aan de mate waarin de doelen zijn bereikt.
5. Weet vakinhoudelijke kennis en vaardigheden te koppelen aan leermethoden.
6. Kan leerlingen met vakspecifieke leerproblemen effectief ondersteunen.

De derde stap van de competentieanalyse, het op basis van de gedragsindicatoren vastleggen van de beheersingsstandaard van alle competenties, ontbreekt.

HOE WORDT BEOORDEELD?

De assessmentprocedure bestaat uit vier fasen:

- **Fase 1:** aanmelding, informatiebijeenkomst en invullen en opsturen van een portfolio. Als volgens de assessoren de bewijzen van geschiktheid voldoende zijn, krijgt de kandidaat een eindbeoordeling.
- **Fase 2:** simulatie en criteriumgericht interview. De kandidaat krijgt een eindbeoordeling als de bewijzen van geschiktheid voldoende zijn.
- **Fase 3:** voorbereiden, uitvoeren en reflecteren op een praktijkopdracht.
- **Fase 4:** eindbeoordeling, eventueel een eindgesprek en het verder uitwerken van het persoonlijk ontwikkelingsplan.

De tabel op pagina 70 en 71, ontleend aan Klarus (2000, p. 23), bevat de competentiematrix van 'Bewezen Geschikt'.

Voor een goed begrip van de assessmentprocedure worden de inhoud en de beoordeling van de zes opdrachten verder toegelicht.

1. Bij het portfolio kan het gaan om ervaringsverslagen, getuigschriften, ervaringsproducten of certificaten en cursusbeschrijvingen. De assessoren moeten beoordelen of het portfolio voldoende bewijs bevat dat de kandidaat over bepaalde startcompetenties beschikt.
2. Bij wat bij deze assesmentprocedure als simulatie aangeduid wordt, dient de kandidaat een aantal opdrachten uit te voeren op basis waarvan diens vakdidactische competenties beoordeeld worden. De simulatie die als voorbeeld gegeven wordt, bestaat uit vier opdrachten. De eerste opdracht is een probleemverkenkend gesprek waarbij de kandidaat zijn reactie op drie situaties met ouders moet aangeven. De voorbereiding en uitvoering van de opdracht bedragen 30 minuten. Bij de tweede opdracht moet de kandidaat op basis van door leerlingen geschreven dictees en uitgevoerde rekenopdrachten aangeven wat hij hiermee in een les zou gaan doen (30 minuten). Bij de derde opdracht moet de kandidaat een globale opzet van vier lessen van dertig minuten over eenzelfde thema maken (60 minuten).

De vierde opdracht betreft het analyseren en beschrijven van een leerprobleem van een kind (50 minuten).

3. Het criteriumgericht interview bestaat uit het bij de kandidaat navragen van bepaalde zaken op basis van de analyse van het portfolio en een bespreking van de vier opdrachten van de simulatie.
4. Bij de praktijkopdracht moet de kandidaat op een school één volledige les uit een bestaande reeks geven. De praktijkopdracht is zowel voor de kandidaten als de assessoren het meest intensieve onderdeel van de procedure.
5. De zelfreflectie op de praktijkopdracht is het startpunt van het reflectiegesprek. Hierbij wordt de kandidaat gevraagd aan te geven welke punten uit de zelfreflectie nog meer informatie kunnen geven over de eigen competenties.
6. De eindbeoordeling van de kandidaat en het persoonlijk ontwikkelplan waarmee de kandidaat een opleidings- en/of coachingstraject in kan gaan.

	Portfolio	Simulatie- opdrachten	Criteriumge- richt interview	Praktijk- opdracht	Reflectie- gesprek	Eindbeoor- deling + POP
A. Handelen in onderwijsleerproces						
1. Vermogen tot het begeleiden van leerlingen in hun leerproces		X		X	X	
2. Vermogen tot het overdragen van communicatie		X	X	X		
3. Vermogen tot communicatie en interactie		X	X	X	X	
4. Vermogen klassenorganisatie en groepsmanagement				X	X	
B Algemeen professioneel handelen						
5. Vakdeskundigheid	X	X	X	X	X	
6. Beroepshouding		X	X		X	X
7. Vermogen tot reflectie en initiatief	X	X	X	X	X	X
8. Vermogen tot ontwikkelen eigen professionaliteit	X				X	X

C Werken binnen een organisatie						
9. Vermogen om met anderen (in een team) samen te werken	X				X	X
D Kennis van het onderwijs in het algemeen						
10. Kennis van het onderwijs			X		X	

Twee assessoren beoordelen de competenties aan de hand van gedragsindicatoren. De assessor geeft bij ieder instrument per competentie op een checklist (zie bijlage 1 van de instrumentenmap) aan of de kandidaat al of niet over deze competentie beschikt. Afhankelijk van de competenties die wel of niet behaald zijn, geeft een assessor een eindwaardering voor het betreffende instrument: goed, voldoende of onvoldoende (Klarus, e.a. 2000, p. 51). Of een standaard voor deze eindwaardering gehanteerd wordt, wordt niet vermeld. Ook lijken er geen formele regels te zijn om uitgaande van de afzonderlijke assessorbeoordelingen van dezelfde competentie bij verschillende instrumenten tot de eindwaardering van een competentie te komen. Vermeld wordt dat de twee assessoren in overleg tot een definitieve eind-

waardering voor de verschillende instrumenten en een eindbeoordeling komen.

De eindbeoordeling kent drie kwalificaties:

- Volledig geschikt: de kandidaat kan direct in het onderwijs ingezet worden en kan bij een lerarenopleiding een aanvraag voor een bevoegdheid doen.
- Bijna volledig geschikt: de kandidaat kan direct in het onderwijs ingezet worden maar moet voor de bevoegdheidsverklaring afspraken maken met de lerarenopleiding om bepaalde competenties via scholing of anderszins aan te vullen binnen twee jaar.
- Niet (volledig) geschikt: de kandidaat kan niet of niet direct in het onderwijs ingezet worden en dient een opleiding te volgen tot docent.

DOOR WIE WORDT BEOORDEELD?

Het zij-instromers assessment duurt in de huidige vorm gemiddeld 15-20 uur per kandidaat. Per kandidaat zijn er twee assessoren, een docentopleider en een praktiserende docent. De assessoren zijn getraind in het afnemen van de diverse instrumenten en het geven van beoordelingsadviezen. Met uitzondering van het portfolio worden de instrumenten door de assessoren onafhankelijk van elkaar beoordeeld.

4.1.2 Evaluatie van 'Bewezen Geschikt'

VALIDITEIT

Uitgangspunt bij de ontwikkeling van de assessmentprocedure 'Bewezen Geschikt' was het document 'Startbekwaamheden leraar primair onderwijs (1997)' dat gelegitimeerd is door het onderwijsveld. Volgens Klarus e.a. (2000, p. 52) zijn de startbekwaamheden door de ontwikkelaars vertaald naar een competentiegerichte beoordelings-situatie waarbij enerzijds een selectie heeft plaatsgevonden naar relevantie voor de beroepspraktijk, anderzijds een groepering van startbekwaamheden naar competenties die werkelijk in de beroeps-

praktijk voorkomen. Het voorgaande heeft geresulteerd in de selectie van tien competenties met bijbehorende gedragsindicatoren. Geen informatie wordt gegeven over wie deze competenties gelegitimeerd heeft, zodat hier aangenomen wordt dat het de ontwikkelaars zelf geweest zijn. De ontwikkelaars gebruiken een combinatie van verschillende instrumenten per competentie om zodoende met 'verschillende brillen' dezelfde competentie te beoordelen (Klarus, 2000, p. 22). Hiervoor werd reeds naar aanleiding van onderzoek bij assessment centers (Jansen, 1993) opgemerkt dat zonder nader onderzoek niet kan worden aangenomen dat 'vermogen tot communicatie en interactie' bij het criteriumgericht interview dezelfde competentie is als 'vermogen tot communicatie en interactie' bij de praktijkopdracht. Aan de representativiteit van de assessmentprocedure is, zowel wat betreft het aspect volledigheid als het aspect authenticiteit, door de ontwikkelaars veel aandacht besteed. Een tekortkoming van de assessmentprocedure die waarschijnlijk niet de ontwikkelaars te verwijten valt, is dat het bij de assessmentprocedure uitsluitend gaat om het vaststellen van pedagogische en (vak)didactische competenties. Of de kandidaten een goede basiskennis hebben van vakken zoals Nederlands en rekenen

wordt in de assessmentprocedure niet nagegaan. Niet duidelijk is of deze basiskennis door de lerarenopleiding in een later stadium beoordeeld wordt voordat een bevoegdheid verstrekt wordt. Indien dat niet het geval is, wordt er voor gepleit dit wel te doen. Bij de rol van de assessoren in de assessmentprocedure kunnen vraagtekens gezet worden. Hoewel 'de criteria waaruit kan blijken of van succesvol functioneren sprake is' (Klarus e.a. 2000, p. 15) als een van de drie wezenlijke elementen voor een beoordeling beschouwd worden, vindt men in 'Bewezen Geschikt' nergens criteria of standaarden. Het ontbreken van standaarden kan afbreuk doen aan de validiteit van de beoordeling omdat zonder standaarden de oordelen van (subjectieve) assessoren een grotere rol kunnen spelen dan de (objectieve) instrumenten bij het beantwoorden van de vraag of kandidaten wel of niet geschikt zijn. Een gemiste kans is dat in het kader van de proefassessments geen onderzoek gedaan is naar de validiteit van de eindbeoordelingen door bijvoorbeeld een beperkt aantal, aan de assessoren onbekende, praktiserende docenten met deze assessmentprocedure te beoordelen. Verwacht zou mogen worden dat praktiserende docenten als 'volledig geschikt' beoordeeld zouden worden.

Indien dat ook inderdaad het geval geweest zou zijn, zou dat het vertrouwen in de validiteit van de eindbeoordelingen van de assessmentprocedure vergroot hebben.

BETROUWBAARHEID

Empirische gegevens over de betrouwbaarheid van de assessmentprocedure ontbreken zodat niet bepaald kan worden of beslissingen die op basis van de procedure genomen worden voldoende betrouwbaar zijn. Aan één aspect dat van invloed is op de betrouwbaarheid, de assessoren, is veel aandacht besteed. Door training van de assessoren en het gebruik van gedetailleerde checklists heeft men getracht de overeenstemming tussen de verschillende assessoren te verhogen. Of men daar ook in geslaagd is, is moeilijk te beoordelen omdat de rapportage (Klarus, 2000, p. 54) hierover te wensen overlaat. Om de overeenstemming tussen assessoren te kwantificeren worden correlaties gebruikt, maar correlaties zijn indicaties van relatieve overeenstemming en niet van absolute overeenstemming. Dit laatste is hier aan de orde omdat zo zou moeten zijn dat verschillende assessoren dezelfde, dat wil zeggen identieke, oordelen geven en niet dat de ene assessor systematisch lagere of

hogere oordelen geeft dan de andere assessor. De absolute overeenstemming kan nooit hoger zijn dan de relatieve overeenstemming en mogelijk vallen die twee hier samen maar dat kan op basis van de verstrekte gegevens niet beoordeeld worden.

UITVOERBAARHEID

De planning en de uitvoering van de assessmentprocedure zijn arbeidsintensief. Zo varieerde de tijdsinvestering per assessor van 12 tot 20 uur per kandidaat waarbij voor de eindbeoordeling een dagdeel nodig is. Van alle opdrachten is de praktijkopdracht het meest intensieve onderdeel van de procedure. De praktijkopdracht vergt niet alleen veel werk van de kandidaten, maar ook van de assessoren. Die moeten een groot organisatievermogen aan de dag leggen. Ook de simulatieopdrachten vereisen veel aandacht. Voor de implementatie in het veld betekent het voorgaande dat docenten van opleidingen geschoold moeten worden in het opzetten en uitvoeren van de assessmentprocedure en opgeleid dienen te worden tot assessor. De instrumentontwikkelaars hebben een programma voor scholing en training van assessoren ontwikkeld.

De vraag of met minder instrumenten volstaan zou

kunnen worden, kan zonder kwantitatieve gegevens over de betrouwbaarheid van de assessmentprocedure niet beantwoord worden. Indien de betrouwbaarheid hoog zou blijken te zijn, zou op basis van de evaluatie van Tillema (2001, p. 35) het portfolio wellicht weggelaten kunnen worden, ook al omdat volgens de competentiematrix de competenties die met het portfolio beoordeeld worden ook reeds door andere instrumenten gedekt worden.

TRANSPARANTIE

Gezien de hoeveelheid informatie die aan kandidaten ter beschikking wordt gesteld, mocht worden aangenomen dat de kandidaten wisten wat van hen verwacht werd en hoe ze beoordeeld zouden worden. Dat dit niet het geval was, blijkt uit de kwalitatieve evaluatie van de assessmentprocedure door Tillema (2001, p. 35-36). Zo was voor kandidaten van buiten het onderwijs het invullen van de portfolio ondanks tussentijdse aanpassingen een probleem. De praktijkopdracht riep veel onzekerheid op omdat het de kandidaten niet duidelijk was wat precies van hen verlangd werd in de te geven les. Veel kritiek was er op wat er na de praktijkles gebeurde of juist niet gebeurde. De verkregen eindbeoordeling werd nogal eens als vaag en vrijblijvend ervaren doordat er ofwel

geen duidelijk advies gegeven werd ofwel het verdere traject niet duidelijk werd aangegeven. Wat betreft de assessoren waren er volgens Klarus e.a. (2000, p. 56) in eerste instantie problemen bij het scoren van de verschillende gedragsindicatoren. In de definitieve versie van de assessmentprocedure is dit aangepast.

4.2 Competentie-instrument ‘Bekwaam Beoordeeld’

4.2.1 Beschrijving van ‘Bekwaam Beoordeeld’

‘Bekwaam beoordeeld’ is de naam van een assessmentprogramma voor pabostudenten dat ontwikkeld is door de EOS projectgroep ‘Bekwaam Beoordeeld’. EOS staat voor Educatief Oostelijk Samenwerkingsverband waarin acht pabo’s en hogescholen participeren. De beschrijving en evaluatie zijn gebaseerd op de in 2000 verschenen publicatie getiteld ‘De Balans’.

WIE WORDT BEOORDEELD?

Het assessmentprogramma is gericht op de beoordeling van kerncompetenties van pabostudenten, met name studenten die het derde jaar van de opleiding nagenoeg afgerond hebben.

WAARTOE WORDT BEOORDEELD?

Het antwoord op deze vraag wordt niet expliciet in de documentatie gegeven. Volgens de ondertitel van ‘De Balans’ gaat het om ‘ontwikkelingsgerichte assessments voor pabostudenten’. Volgens ‘De Balans (p. 7)’ is het assessmentprogramma in eerste instantie gericht op derdejaars studenten die zich voorbereiden op de LIO-stage. In ‘De Balans (p. 23)’ wordt echter opgemerkt dat het resultaat van alle ontwikkelingsinspanningen een assessmentprogramma is dat operationeel is voor tweede, derde en vierdejaarsstudenten. Op basis van het voorgaande wordt hier geconcludeerd dat het assessmentprogramma bedoeld is voor het nemen van doorstroombeslissingen over studenten in verschillende fasen van de lerarenopleiding.

WAT WORDT BEOORDEELD?

Uitgangspunt voor het assessmentprogramma zijn de kenmerkende en kritische situaties die in ‘Start-bekwaamheden leraar primair onderwijs (1997)’ beschreven worden. Na literatuurstudie, analyse van representatieve praktijksituaties en de reeds eerder beschreven kenmerkende en kritische situaties en onderzoek via de proefafnamen, zijn er negen

zogenaamde kerncompetenties geselecteerd die met het assessmentprogramma beoordeeld worden.

Bij de beschrijving van een kerncompetentie kunnen de volgende stappen onderscheiden worden:

1. Het formuleren van een definitie waarbinnen de verschillende aanwezige componenten als handelingsdimensies binnen de competentie beschreven worden. Handelingsdimensies zijn clusters van gedrag die gezamenlijk en in bepaalde volgorde de gehele competentie vormen.
2. Op basis van deze handelingsdimensies worden in termen van gedrag in een vijfpuntsschaal de onderscheiden beheersingsniveaus beschreven. Niveau 4 wordt beschouwd als het beheersingsniveau aan het einde van het opleidingstraject.
3. Vanuit de vijfpuntsschalen worden voor elk onderscheiden niveau observatiepunten in termen van concreet waarneembaar gedrag benoemd met behulp waarvan een assessor het gedrag kan beschrijven en beoordelen.

Om een zo eenduidig mogelijke omschrijving van de verschillende kerncompetenties te krijgen, is een format ontwikkeld. Voor de kerncompetentie 'lesvoorbereiding (didactiseren)' worden hieronder delen van het format gepresenteerd (De Balans, p. 61-65).

Definitie

Onder 'lesvoorbereiding/didactiseren' verstaan we: de vaardigheid om leerstof geschikt te maken voor overdracht aan verschillende doelgroepen met specifieke kenmerken en dit middels een schriftelijke weergave te kunnen verantwoorden.

Handelingsdimensies binnen de competentie lesvoorbereiding

1. Leerstofselectie: het kiezen en formuleren van doelstellingen en leerinhouden aansluitend op een beschreven beginsituatie.
2. Didactische opbouw: het kiezen en beschrijven van structuur, activiteiten en werkvormen, inclusief middelen.
3. Adaptiviteit: het planmatig aanpassen van 1 en 2 aan individuele verschillen.
4. Evaluatie/toetsing: het kiezen en beschrijven van de afsluitingsprocedures.

De vijf beheersingsniveaus

Van de vijf beheersingsniveaus wordt hier beheersingsniveau 4 vermeld die luidt:

- Selecteert leerstof uit het aangeboden materiaal op basis van veronderstelde verschillen in beginsituatie en gegeven doelen.
- Stemt de les, instructie, werkvormen, begeleiding, af op de interesse, het verschil in tempo, niveau, en voorkennis in de groep.
- Plant een controle van de beginsituatie.
- Plant toetsing/evaluatie.

Observatiepunten

Voor beheersingsniveau 4 zijn de observatiepunten (gedragsindicatoren) bij leerstofselectie:

- Beschrijft beginsituatie in detail.
- Selecteert de leerstof uit het aangeboden materiaal.
- Kiest en beschrijft doelstellingen, gerelateerd aan specifieke situatie (leerinteresse, ontwikkelingskenmerken, niveau, omstandigheden).

Voor de observatiepunten voor de dimensies 'didactische opbouw', 'adaptiviteit' en 'evaluatie/toetsing' wordt verwezen naar de voornoemde documentatie.

Verantwoording keuzes

Voor het assessmentprogramma is de kerncompetentie 'lesvoorbereiding' van belang omdat:

- De competentie voorwaardelijk is voor het geven van goede lessen - het voorbereiden van lessen is een primaire beroepsfactor.
- Het vooraf (schriftelijk) verantwoording afleggen over de gemaakte keuzes, de planning en het vervolg van een les geeft onder andere informatie over de vraag of de student kan denken in kleine leerstappen en leerstofgehelen overziet, over de schriftelijke communicatie vaardigheden van de student

Begripsafbakening

Bij de competentie 'lesvoorbereiding' gaat het om het vormgeven van onderwijs, waarbij wordt overwogen hoe en in welke mate het gericht wordt op de specifieke kenmerken van de groep en de individuele kenmerken.

Relaties met andere kerncompetenties

Voor de beschrijving met welke andere kerncompetenties behalve voorwaardelijke kennis er een relatie is, wordt verwezen naar de voornoemde documentatie.

Relatie met de startbekwaamheden

Voor de beschrijving van de relatie met de algemene startbekwaamheden, de vakspecifieke startbekwaamheden en met de kenmerkende en kritische situaties, wordt verwezen naar de voornoemde documentatie.

Uitvoering

De student krijgt een 'at random' geselecteerd themapakket toegewezen, waaruit een les van 10 minuten over een specifiek onderwerp of vraagstuk binnen dit thema moet worden voorbereid voor overdracht aan een specifieke doelgroep (de groep medestudenten plus de rollenspeler). De opdracht is zodanig samengesteld, dat het voor de student mogelijk is de opdracht op elk van de beschreven beheersingsniveaus uit te voeren. Van de voorbereiding wordt een schriftelijke verantwoording gegeven die door een assessor beoordeeld wordt.

Voor een kerncompetentie zijn vijf beheersingsniveaus gedefinieerd. Hiervoor werden voor de kerncompetentie 'lesvoorbereiding' de gedragsindicatoren of observatiepunten van beheersingsniveau 4 aangegeven.

Of bij het bepalen van het beheersingsniveau een beheersingsstandaard gehanteerd wordt, wordt niet volledig duidelijk uit de documentatie. Het lijkt dat aan alle observatiepunten die bij een bepaald beheersingsniveau genoemd worden, voldaan moet zijn om met dat beheersingsniveau gewaardeerd te worden.

HOE WORDT BEOORDEELD?

In het assessmentprogramma worden de negen kerncompetenties beoordeeld met drie opdrachten, een reflectieverslag en een afrondend gesprek. Een vragenlijst vooraf en zelfbeoordelingen na elke opdracht en een schriftelijke verantwoording van de lesvoorbereiding complementeren het programma. De drie opdrachten zijn een samenwerkingsopdracht (duur 2 uur), een lesvoorbereiding (duur 70 minuten) en de presentatie van de voorbereide les (duur 15 minuten). Naast de kerncompetenties worden ook nog twaalf persoonskenmerken (motivatie, stressbestendigheid etc.) en de schriftelijke en mondelinge taalvaardigheid beoordeeld.

De competentiematrix van het assessmentprogramma staat in tabel hiernaast.

	Samenwerkings-opdracht	Lesvoorbereiding	Presentatie van de les	Reflectie-verslag	Afrondend gesprek
Samenwerken	X				
Leiding geven	X		X		
Didactiseren		X			
Presenteren			X		
Doelgericht handelen	X	X	X		
Plannen en organiseren	X	X	X		
Responsiviteit	X		X		
Flexibiliteit	X	X	X	X	X
Reflectief vermogen		X	X	X	X

De kerncompetenties worden met behulp van observatielijsten op vijf puntsschalen beoordeeld. Uit de documentatie blijkt niet of er formele regels zijn om de afzonderlijke assessorbeoordelingen van dezelfde kerncompetentie bij verschillende instrumenten samen te vatten in één eindwaardering van de betreffende kerncompetentie. Ook lijken formele regels die leiden tot een eind-

beoordeling van de student te ontbreken. Vermeld wordt dat twee assessoren over vier studenten rapporteren. De rapportage gaat naar de student en de opleidingsbegeleider. Op basis van deze rapportage bepalen begeleider en student samen of er sprake is van eventuele achterstand en stellen zij vervolgens samen een persoonlijk ontwikkelplan op.

DOOR WIE WORDT BEOORDEELD?

Het assessmentprogramma duurt anderhalve dag waarbij de studenten in groepen van acht, per onderdeel telkens door twee assessoren worden begeleid en geobserveerd. In voorkomende gevallen wordt de student ook door de rollenspeler beoordeeld. Om assessor te kunnen worden, moet men een training van acht dagdelen volgen. Het voornemen is een procedure voor de certificering van getrainde assessoren te gaan ontwikkelen.

4.2.2 Evaluatie van 'Bekwaam Beoordeeld'

VALIDITEIT

Het doel van 'Bekwaam Beoordeeld' is een beeld te krijgen van het beheersingsniveau van alle relevant geachte kerncompetenties en persoonskenmerken van pabostudenten. Hoewel volgens 'De Balans (p.5)' het assessmentprogramma (nog) niet alle kerncompetenties bevat, wordt geen informatie gegeven over de vraag hoe cruciaal die ontbrekende kerncompetenties zijn. Van de kerncompetenties wordt ook gezegd dat ze nog niet volledig en uitputtend omschreven zijn. Over hoe de kerncompetenties 'reflectief vermogen' en 'responsiviteit' benoemd

moeten worden, bestaan verschillende opvattingen. Overwogen wordt om de beoordeling van de kerncompetentie 'responsiviteit' bij de volgende versie van het assessmentprogramma te laten vervallen. Uit het voorgaande blijkt dat de ontwikkeling van het assessmentprogramma zich nog in een experimentele fase bevindt.

Wat betreft de representativiteit van het assessmentprogramma, kunnen zowel bij het aantal en de omvang van de opdrachten als de authenticiteit van de opdrachten vraagtekens gezet worden. Wat betreft de omvang is de duur van de presentatie van de les nogal kort en wat betreft de authenticiteit lijkt de rollenspeler die bepaalde gedragingen probeert uit te lokken een te cruciale rol te spelen bij het realiseren van de beoogde authenticiteit. Deze afhankelijkheid van een rollenspeler maakt de authenticiteit van de opdrachten erg kwetsbaar. Wat ook afbreuk doet aan de authenticiteit van de opdrachten is dat ze niet uitgevoerd worden met kinderen maar met medestudenten. Het is zeer de vraag of het gedrag van medestudenten wel vergelijkbaar is met dat van kinderen.

Ook hier geldt wat bij de bespreking van de validiteit van 'Bewezen Geschikt' reeds opgemerkt werd, namelijk dat zonder nader onderzoek niet aangenomen

men mag worden dat een bepaalde kerncompetentie nog steeds dezelfde kerncompetentie is bij verschillende opdrachten.

Onderzoek naar de validiteit van het assessmentprogramma is gepland waarbij het functioneren in de stage vergeleken zal worden met de uitvoering van de opdrachten uit het assessmentprogramma.

BETROUWBAARHEID

De betrouwbaarheid van het assessmentprogramma is niet onderzocht. Van één aspect dat van invloed is op de betrouwbaarheid, de assessoren, wordt wel vermeld welke inspanningen naast een uitgebreide assessorentraining verricht zijn om de mogelijke verschillen tussen assessoren te minimaliseren, bijvoorbeeld het toevoegen van observatiepunten aan de vijfpuntsschaal, maar kwantitatieve gegevens over assessoren worden niet gepresenteerd.

UITVOERBAARHEID

De uitvoering van het assessmentprogramma vereist volgens de documentatie een uitstekende planning, organisatie en logistiek. Een draaiboek van het assessmentprogramma en handleidingen voor assessoren en studenten zijn beschikbaar. Over mogelijke problemen met betrekking tot de

uitvoerbaarheid wordt in de documentatie niet gerept.

TRANSPARANTIE

In voornoemd draaiboek staat dat een beschrijving van de inhoud van elke opdracht in de handleiding voor de kandidaat staat. Voorafgaand aan elke opdracht en zelfbeoordeling krijgt de kandidaat aanvullend een toelichting van één van de assessoren en is er gelegenheid tot het stellen van vragen. De zelfbeoordelingen, de verantwoording van de lesvoorbereiding en het reflectieverslag van de student zijn aan richtlijnen gebonden. Deze documenten zijn opgenomen in de kandidatenhandleiding zodat alle kandidaten over dezelfde en zo eenduidig mogelijke informatie beschikken. De assessoren beschikken over dezelfde informatie en hulpmiddelen zoals observatielijsten, observatiepunten, handelingsdimensies en vijfpuntsschaal e.d., alsmede richtlijnen voor het voeren van het gesprek en het schrijven van de rapportage. Er is een procedure ontworpen via welke de assessor tot een voorlopige beoordeling komt, voorafgaande aan het assessorenoverleg.

4.3 Competentie-instrument ‘Integratief BeoordelingsMoment’

4.3.1 Beschrijving van ‘Integratief BeoordelingsMoment’

Het ‘Integratief BeoordelingsMoment (IBM)’ is een door de Educatieve Faculteit Amsterdam (EFA) ontwikkeld assessment. De beschrijving en de beoordeling van het ‘IBM’ zijn gebaseerd op een publicatie van Oudkerk Pool en Snoek (2000), getiteld ‘Beoordeling als koersbepaler voor een competentiegericht curriculum’.

Wie wordt beoordeeld?

Met het ‘IBM’ worden eerstejaarsstudenten van drie opleidingen van de Educatieve Faculteit Amsterdam beoordeeld.

WAARTOE WORDT BEOORDEELD?

Met het ‘IBM’ wordt een doorstroombeslissing genomen, dat wil zeggen dat op basis van het ‘IBM’ beslist wordt of een student toegelaten wordt tot de hoofdfase van de opleiding.

WAT WORDT BEOORDEELD?

Volgens voornoemde publicatie (p. 12) worden met het ‘IBM’ de competenties van de student beoordeeld. Die competenties beschrijven de essentieel geachte bekwaamheden van een persoon die bepalend zijn voor het professioneel uitvoeren van taken in de concrete beroepssituatie. De competenties hebben daarom betrekking op een samenhangende en geïntegreerde verzameling van kennis, vaardigheden en attitudes.

Een HBO-er hoort over beroepsonafhankelijke en beroepsspecifieke competenties te beschikken.

De beroepsonafhankelijke competenties zijn:

- Het methodisch kunnen handelen in complexe situaties.
- Het in staat en bereid zijn tot het verantwoorden en legitimeren van gemaakte keuzes.
- Het willen bijdragen aan de ontwikkeling van methodiek en beroep.
- De behoefte om het handelen te verankeren in een wetenschappelijke kennisbasis.
- Het combineren van het geven van leiding en bieden van ondersteuning aan medewerkers.

De beroepsspecifieke competenties kunnen onderverdeeld worden in:

- Competenties die te maken hebben met het algemene gedrag van een docent als professioneel medewerker in een veranderende schoolorganisatie (mesoniveau van het beroep).
- Competenties die te maken hebben met het pedagogisch didactisch handelen van de docent (microniveau van het beroep).
- Competenties die te maken hebben met het vakgebied waar de docent les in geeft.

In het 'IBM' staan met name de competenties die te maken hebben met het algemene gedrag van een docent als professioneel medewerker centraal. Dit heeft te maken met het feit dat het 'IBM' beoordeelt of de student hoofdfasebekwaam is. Binnen het opleidingsconcept van de EFA betekent dit dat de student in staat moet zijn om zelfstandig en zelfverantwoordelijk vorm te geven aan zijn eigen leerproces.

Informatie over de tweede stap van de competentieanalyse, het omschrijven van de competenties in gedragsindicatoren, is ontleend aan de beschrijving van de kritische opdracht die een telefoongesprek betrof (zie hiernaast).

Tijdens het telefoongesprek, de evaluatie en de presentatie letten de assessoren op de volgende beoordelingsaspecten:

- Luisterende houding.
- Onderzoekende, vragende houding.
- Analyseren (verschillende gezichtspunten).
- Samenwerken.
- Koppeling praktijk en theorie.
- Initiatief nemen.
- Omgaan met (eventuele) weerstand.
- Planmatig werken.
- Telefoongesprek voeren.
- Samenvatten.

Bij deze opdracht is de volgende beheersingsstandaard, de derde stap van de competentieanalyse, vastgesteld:

- De student scoort onvoldoende wanneer hij op twee of meer van de beoordelingsaspecten niet handelingsbekwaam is en zich hiervan ook niet bewust is.
- De student scoort voldoende wanneer hij op één of twee beoordelingsaspecten niet handelingsbekwaam is maar zich hiervan wel bewust is en dit kan vertalen in leervragen.

- De student scoort goed wanneer hij handelingsbekwaam is op alle beoordelingsaspecten en kan aangeven hoe hij zich op deze gebieden verder kan ontwikkelen.

HOE WORDT BEOORDEELD?

Het 'IBM' bestaat uit de volgende onderdelen:

1. *Beoordeling van het portfolio (voorafgaand aan de IBM-dag).*

Met behulp van de in het portfolio verzamelde gegevens (eigen producten, feedback van docenten en medestudenten, reflecties op eigen leerproces, etc.) maakt de student een zelfevaluatie. Hierin laat hij het team van assessoren zien hoe ver hij is in zijn ontwikkeling op de door de opleiding vastgestelde competentiegebieden. In deze zelfevaluatie verwijst de student naar zijn portfolio als het gaat om bewijsvoering voor zijn beweringen.

2. *Feedback op het portfolio.*

Op basis van het portfolio geven de assessoren in een kort gesprek individuele feedback aan de student. Indien de assessoren bepaalde competenties onvoldoende bewezen achten, krijgt de student aanvullende opdrachten om alsnog te laten zien dat hij wel over deze competenties

beschikt. De student gebruikt de feedback om zijn inbreng in het rondetafelgesprek voor te bereiden.

3. *Kritische situatie.*

Aan een groep van drie studenten wordt een kritische situatie voorgelegd waarin de belangrijkste competenties op integratieve wijze aan de orde komen. Tijdens de uitvoering van deze opdracht observeren de assessoren. De assessoren bespreken hun beoordeling en bereiden op basis hiervan het rondetafel gesprek voor waarbij de resultaten uit voorgaande stappen en de beoordeling van de eventuele aanvullende opdracht worden meegenomen.

Een voorbeeld van een kritische situatie is een casus waarbij de mentor van een leerling de ouders opbelt omdat de betreffende leerling in korte tijd twee keer verzuimd heeft. Door de student wordt een telefoongesprek gevoerd met de leerling (jongerentelefoon), één van de ouders van de leerling (opvoedtelefoon) en de mentor van de leerling (onderwijstelefoon). In deze casus wordt een beroep gedaan op de handelingsbekwaamheid van de student als toekomstig HBO-pedagoog. De student krijgt voor deze opdracht tien minuten voorbereidingstijd en vijf minuten voor elk gesprek. Daarna krijgen de drie studenten twintig

minuten de tijd om met elkaar de gesprekken te analyseren. Deze analyse moet in een betoog van tien minuten aan de drie assessoren gepresenteerd worden.

4. *Ronde Tafel (RT).*

Het rondetafelgesprek is een evaluatief gesprek tussen assessoren en drie studenten aan de hand van voorgaande stappen. Assessoren en studenten beschikken hiervoor over een protocol. Assessoren en studenten zijn in dit gesprek gelijkwaardige gesprekspartners wat betekent dat van studenten ook feedback op elkaar, met name over functioneren tijdens de kritische situatie, gevraagd wordt

5. *Eindoordeel.*

Op basis van de oordelen op voorgaande stappen komen de drie assessoren tot een eensluidend eindoordeel. De student krijgt dit eindoordeel schriftelijk en kan dit opnemen in zijn portfolio ten behoeve van sturing van het eigen leerproces en van voorbereiding van de assessments in de hoofdfase.

Hoewel in de documentatie vermeld wordt dat in het 'IBM' met name de competenties centraal staan die te maken hebben met het algemene gedrag

van een docent als professioneel medewerker in een veranderende schoolorganisatie, wordt er geen overzicht gegeven van de competenties die met de verschillende instrumenten beoordeeld worden.

Vandaar dat van het 'IBM' geen competentiematrix opgesteld kon worden.

In de documentatie wordt ook geen informatie gegeven over de vraag hoe een eindwaardering van een competentie en hoe een eindbeoordeling van een student tot stand komt. Wel wordt vermeld dat de assessoren uiteindelijk tot drie verschillende beoordelingsuitslagen kunnen komen:

1. De student wordt door de assessoren voldoende competent beoordeeld om door te gaan naar de hoofdfase.
2. De student wordt als onvoldoende competent beoordeeld en krijgt een bindend negatief studieadvies.
3. De student krijgt een voorlopige beoordeling door de assessoren omdat hij nog niet voldoende competent geacht wordt maar naar verwachting binnen een half jaar wel de competenties heeft om toegelaten te worden tot de hoofdfase. De student sluit een individueel leercontract met zijn mentor af gericht op het alsnog voldoen aan de eisen van het 'IBM'.

Na een half jaar kan de student weer opgaan voor een 'IBM'.

DOOR WIE WORDT BEOORDEELD?

Het 'IBM' wordt afgenomen door professionele assessoren. Assessoren ondergaan een uitgebreide training waarin de assessor moet laten zien dat hij voldoende bekwaam is. De assessor is niet alleen beoordelaar maar ook ontwerper van beoordelings-situaties. Bovendien moet hij inhoudelijk deskundig zijn en op zo'n manier feedback kunnen geven dat studenten daar constructief mee omgaan. Verder moet de assessor ontwikkelingsgericht zijn ten opzichte van zijn eigen bekwaamheden en kunnen laten zien over de noodzakelijke reflectieve bekwaamheden te beschikken die essentieel zijn voor de eigen ontwikkeling op weg naar 'expert' assessor. De docent zou hiervoor gebruik moeten maken van zijn eigen docentportfolio waarin hij bewijst aan de gevraagde competenties te voldoen.

Bij het 'IBM' worden per zes studenten drie assessoren ingezet die gezamenlijk de studenten beoordelen. Er wordt gewerkt met gemengde teams van beoordelaars, bijvoorbeeld door minimaal één assessor afkomstig te laten zijn uit een andere opleiding.

4.3.2 Evaluatie van 'Integratief Beoordelings-Moment'

VALIDITEIT

Over de vraag welke competenties beoordeeld worden met het 'IBM', bevat de documentatie (Oudkerk-Pool & Snoek, 2000) weinig informatie. Van het 'IBM' kon dan ook geen competentiematrix gereconstrueerd worden. Ook ontbreken verwijzingen naar de start-bekwaamheden en andere bekende documenten over competenties van docenten.

De representativiteit van het assessment is beperkt. Er wordt slechts een enkel aspect van het functioneren van een docent beoordeeld maar wellicht wordt door de opleiding niet meer van studenten geëist en/of mag van eerstejaars studenten niet meer verwacht worden. Aan de authenticiteit van de opdracht is voldoende aandacht besteed.

BETROUWBAARHEID

Over onderzoek naar de betrouwbaarheid van het 'IBM' wordt niet gerapporteerd. Er is veel aandacht besteed aan maatregelen om de intersubjectiviteit van de assessoren te verhogen, maar kwantitatieve gegevens hierover ontbreken.

Indien men die gegevens wel verzameld zou hebben, had men evenals bij vergelijkbare beoordelingsinstrumenten wellicht kunnen constateren dat twee in plaats van drie assessoren voldoende zouden zijn geweest. Indien drie of meer assessoren nodig zijn voor een hoge intersubjectiviteit, zou nog eens kritisch naar de beoordelingsvoorschriften bij de verschillende onderdelen van het assessment gekeken moeten worden.

UITVOERBAARHEID

De planning en uitvoering van het 'IBM' zijn minder arbeidsintensief dan die van de twee assessments die hiervoor beschreven zijn. In de documentatie wordt geen melding gemaakt van problemen ten aanzien van de uitvoerbaarheid.

TRANSPARANTIE

Op basis van de informatie die in verband met de casus van de kritische situatie gegeven wordt, krijgt men de indruk dat de studenten goed geïnformeerd worden over wat van hen verwacht wordt bij de uitvoering van de opdracht en hoe zij beoordeeld zullen worden. Aan de assessoren worden hoge eisen gesteld zodat verwacht mag worden dat zij geen problemen zullen hebben bij het afnemen en beoordelen van het 'IBM'.

4.4 Competentie-instrument 'Als Docent in Beeld'

4.4.1 Beschrijving van 'Als Docent in Beeld'

'Als Docent in Beeld' is een assessment voor de afsluiting van de propedeuse, het eerste jaar van het reguliere voltijdsprogramma, dat ontwikkeld is door de Faculteit Educatieve Opleidingen (FEO) van de Hogeschool van Utrecht (HvU). De beschrijving en beoordeling zijn gebaseerd op de publicatie Proeve hoofdfase bekwaam 'Als Docent in Beeld (2000)' van voornoemde faculteit.

WIE WORDT BEOORDEELD?

Met het assessment worden eerstejaars studenten van de FEO beoordeeld.

WAARTOE WORDT BEOORDEELD?

Met het assessment wordt een doorstroombeslissing genomen, dat wil zeggen dat de uitslag van het assessment direct bijdraagt aan het geven van een (bindend) studieadvies aan het einde van het eerste jaar.

WAT WORDT BEOORDEELD?

Met de startbekwaamheden als uitgangspunt zijn door de FEO samen met andere educatieve partners competenties voor een startbekwame docent vastgesteld. Voor het onderhavige assessment is uitgegaan van de competenties zoals die in het EPS-project van de HvU/FEO voorgesteld zijn. Hiervan worden voor het assessment de volgende competenties gebruikt:

- Kan ik goed leiding geven aan een groep, stel ik grenzen? (Domein Pedagoog)
- Kan ik een goede presentatie geven? (Domein Didacticus)
- Kan ik samenwerken? (Domein Teamlid)
- Kan ik goed plannen? (Domein Professional)
- Kan ik evalueren en reflecteren? (Domein Professional)
- Heb ik een goed beeld van de taak en van mijzelf? (Domein Professional).

De tweede stap van de competentieanalyse bestaat hieruit dat bij een competentie een aantal rollen onderscheiden worden en dat bij die rollen gedragscriteria geformuleerd worden. Zo worden bij het 'domein pedagoog' drie rollen onderscheiden: opvoeder; begeleider van groepsdyna-

mische processen; verstrekker van zorg, gericht op individu en groep. In de documentatie wordt van beheersingsstandaards voor de onderscheiden competenties, de derde stap van de competentie-analyse, geen melding gemaakt. Uit de scorelijst voor de assessor blijkt dat er geen 'scores' worden gegeven (zie hieronder).

HOE WORDT BEOORDEELD?

Op het einde van het eerste jaar vindt een assessmentgesprek plaats tussen de opleiding en de student over zijn competenties op het gebied van de beroepsvoorbereiding (BVC). Het gesprek leidt tot een advies aan de vakexamencommissie (VEC). De VEC geeft een bindend studieadvies op basis van het BVC-advies en de behaalde resultaten bij cursussen uit het eerste jaar. In het gesprek toont de student dat hij klaar is voor de hoofdfase en hij doet dit aan de hand van bewijsmateriaal dat hij in de loop van het jaar opgebouwd heeft. Het materiaal is gebundeld in het portfolio met de titel 'Als docent in beeld'.

Nevenstaande competentiematrix van het portfolio laat zien dat het portfolio uit vijf onderdelen bestaat. Wat die vijf onderdelen inhouden, wordt na de presentatie van de competentiematrix toegelicht.

	Leiding geven	Presenteren	Samenwerken	Plannen	Evalueren	Beeld (zelf/taak)
1. Presentatie 1 en 2						
Vorbereidingsformulier				X		
Uitvoering verslag		X				
Feedback anderen	X	X				
Video (1) en verslag	X	X				
2. Educatieve activiteit						
Vorb.form./werkplan				X		
Verslag	X	X	X			
Feedback anderen		X	X			
3. Onderzoekstaken en samenwerking	X		X	X		X
4. Zelfevaluatie totaal						
Motivatieverslag	X				X	X
Feedback anderen	X	X	X	X	X	X
Sterkte/zwakte-analyse	X	X	X	X	X	
5. Assessmentgesprek					X	X

Het portfolio bestaat uit vijf onderdelen:

1. Twee presentaties bestaande uit een voorbereidingsformulier, een verslag van de uitvoering van de lessen en feedback van een medestudent en van de begeleider. Tevens een kort videofragment van één presentatie met een reflectieverslag waarbij de inzichten die hierbij zijn opgedaan, verwerkt worden in het verslag.
2. De educatieve activiteit bestaande uit een voorbereidingsformulier/werkplan, een verslag van de uitvoering van een les, feedback van medestudenten, van leerlingen en van de begeleider.
3. Opdrachten gemaakt bij de oriënterende stage waaronder een onderzoeksverslag over de taken van een docent in de klas en school en de mening van de student hierover in relatie tot het beroepsperspectief.
4. Een zelfevaluatie bestaande uit: een motivatieverslag waarbij de student zelf aangeeft zich wel/niet geschikt te achten voor het leraarsberoep en wel/niet toegang tot de hoofdfase te willen verkrijgen; feedback op papier over zijn competenties door een medestudent en de mentor; sterkte/zwakte analyse op basis van het voorgaande.

5. Een assessmentgesprek dat hiervoor reeds besproken is.

Het portfolio bestaat uit een verplicht deel en een vrij deel dat de student zelf kan vormgeven.

Twee assessoren beoordelen de student met betrekking tot de zes onderscheiden competenties. De beoordelingscategorieën zijn: goed; voldoende, nog aan werken; matig, nog veel aan werken; grote twijfel of deze competentie voldoende ontwikkeld kan worden; niet genoeg materiaal/ervaring om iets van te kunnen zeggen. Over (regels over) hoe die beoordeling tot stand komt, wordt geen informatie gegeven. De twee assessoren vergelijken hun beoordelingen en geven een eindbeoordeling. De eindbeoordeling kan zijn dat de student geschikt geacht wordt voor toelating tot de hoofdfase of dat er twijfel bestaat over zijn geschiktheid voor de hoofdfase.

DOOR WIE WORDT BEOORDEELD?

De beoordeling gebeurt door twee assessoren: een docent uit de hoofdfase (bij voorkeur uit de BVC) en een docent die tevens lid van de VEC is. Het streven is dat de student één docent al langer kent (les van gehad heeft) en de ander niet.

4.4.2 Evaluatie van 'Als Docent in Beeld'

VALIDITEIT

De competenties die geselecteerd zijn voor het assessment, zijn bedoeld om de competenties van studenten aan het einde van het eerste jaar van de opleiding te beoordelen. Hoewel niet duidelijk is door wie de keuze van de competenties gelegitimeerd is, worden een aantal van deze competenties ook in andere assessments gebruikt. Aan de representativiteit van het assessment, zowel de volledigheid als de authenticiteit, is de nodige aandacht besteed. Wat betreft de authenticiteit dient opgemerkt te worden dat niet alleen aan leerlingen van het voortgezet onderwijs les gegeven wordt maar ook dat aan de leerlingen expliciet om feedback gevraagd wordt. In de documentatie wordt niet vermeld of onderzoek naar de validiteit en de betrouwbaarheid van het assessment gepland is.

BETROUWBAARHEID

De documentatie bevat geen gegevens over onderzoek naar de betrouwbaarheid van het assessment. Ook wordt geen informatie gegeven over maatregelen die de intersubjectiviteit van de assessoren bij de

beoordeling van de competenties van de studenten kunnen verhogen. Als de assessoren echter uitsluitend met behulp van de in de documentatie vermelde evaluatieformulieren de studenten beoordelen, mag verwacht worden dat de assessoren een (te) groot aantal studenten verschillend zullen beoordelen.

UITVOERBAARHEID

Hoewel het samenstellen van het portfolio een arbeidsintensieve zaak moet zijn, bevat de documentatie geen informatie over problemen met betrekking tot de uitvoerbaarheid van het assessment.

TRANSPARANTIE

De instructies bij de instrumenten zijn van dien aard dat aangenomen mag worden dat de studenten goed op de hoogte zijn van wat van hen verwacht wordt en hoe zij beoordeeld zullen worden. De documentatie over het assessment bevat geen informatie over het trainen van assessoren of andere informatie over hoe de assessoren op hun beoordelingstaak voorbereid worden. De evaluatieformulieren waarmee de assessoren de competenties van de studenten moeten beoordelen zijn zeer summier.

4.5 Beschrijving van instrumenten voor de beoordeling van competenties van praktiserende docenten

Pelkmans (1998) heeft geïnventariseerd welke instrumenten in het onderwijs momenteel gebruikt worden bij het beoordelen van leraren en schoolleiders. Hij heeft zijn inventarisatie niet beperkt tot Nederland maar hij is ook nagegaan hoe beoordeling van leraren in verschillende andere landen plaats vindt. Interessant is in de tabel hiernaast, ontleend aan Pelkmans (1998, p. 26), die de resultaten bevat van een onderzoek uit 1996 bij 68 grote school-districten in de Verenigde Staten naar methoden en instrumenten die gebruikt worden bij het beoordelen van leraren.

Gebruikte methoden en instrumenten	Frequentie (%)
Systematische observatie van het lesgeven	94,1
Informele observatie	86,8
Waarderingscores van collega's	16,2
Waarderingscores van leerlingen	8,8
Leerlingprestaties	25,0
Schriftelijk examen	1,5
Portfolio van de leraar	23,5
Zelfevaluatie van de leraar	45,6

Uit het schooldistrictenonderzoek blijkt dat overal observatie toegepast wordt en dat daarnaast ook andere evaluatiemethoden gebruikt worden. De evaluatiemethoden worden bij beginnende docenten ingezet bij indiensttreding en bij docenten in vaste dienst voor het remediëren in geval van onvoldoende functioneren, ten behoeve van professionele ontwikkeling en bij het vernieuwen van aanstellingen. De beoordeling van een docent wordt veelal uitgevoerd door de schooldirecteur en een andere docent.

De situatie in het Amerikaanse onderwijs verschilt sterk van die in het Nederlandse onderwijs. Volgens Pelkmans (1998, p. 112) vindt beoordeling van zittende leraren in Nederland vooral plaats via een functioneringsgesprek, terwijl meer directe instrumenten zoals observatie, portfolio en assessment center slechts incidenteel toegepast worden. Wel worden de laatste jaren beginnende leraren meer begeleid en geëvalueerd door middel van observatie, terwijl bij de selectie van directieleden steeds vaker een assessment center ingezet wordt. Als gevolg van het lerarentekort zijn er recentelijk wijzigingen op het gebied van het personeelsbeleid geïntroduceerd. In de onderwijsbegroting 2001 wordt opgemerkt dat op het gebied van het personeelsbeleid scholen in het primair onderwijs nog te weinig eigen beleidsruimte hebben. Om die beleidsruimte te vergroten, worden extra middelen uitgetrokken om functie- en beloningsdifferentiatie mogelijk te maken. Scholen kunnen ook goed functionerende leraren of leraren die beschikken over bepaalde expertise, bijvoorbeeld taalcoördinatoren en interne begeleiders, extra belonen door hen in een hogere salarisschaal te plaatsen. Zo ontstaan meer carrièreperspectieven in het primair onderwijs. Als onderdeel van de CAO-afspraken zijn ook middelen beschikbaar gesteld

voor de bevordering van integraal personeelsbeleid en competentiebeoordeling en -beloning. Integraal personeelsbeleid wordt in de nota 'Maatwerk voor morgen' (1999) omschreven als 'systematische afstemming van de kennis en bekwaamheden van het personeel op geformuleerde inhoudelijke en organisatorische doelen van de onderwijsinstelling: de school als lerende organisatie.'

Voordat met competentiebeoordeling en competentiebeloning van praktiserende docenten begonnen kan worden, zal er eerst een competentieanalyse dienen plaats te vinden. Door Berenschot is een competentieanalyse uitgevoerd die heeft geresulteerd in het zogenaamde competentiemodel van Berenschot (Twisk e.a., 1999). Uit voornoemd document wordt hierna vrijelijk geciteerd en geparafraseerd. Het competentiemodel is gebaseerd op bestaande profielen en gelegitimeerd in panelbijeenkomsten met schooldirecteuren en leraren. Het competentiemodel bestaat uit tien competenties die elk uitgewerkt zijn in drie competentieniveaus (basis, ervaren en excellent niveau). Door per competentie het niveau van een bepaalde leraar vast te stellen, ontstaat een individueel competentieprofiel. Hieronder staat, ontleend aan Twisk e.a. (p. 17), een voorbeeld van een individueel competentieprofiel.

	Basis	Ervaren	Excellent
Gedifferentieerd werken			X
Groepsmanagement		X	
Onderwijskundige wendbaarheid		X	
Vakmatige beheersing		X	
Vernieuwend denken		X	
Stimuleren tot vernieuwing		X	
Samenwerking		X	
Longitudinaal plannen			X
Collegiale consultatie		X	
Externe contacten			X

Uitgangspunt bij de ontwikkeling van het competentiemodel was de gedachte dat het voor het onderwijs van essentieel belang is dat leraren een loopbaan kunnen doorlopen in het primaire proces van het onderwijs. Het competentiemodel richt zich op een loopbaanontwikkeling van juniorleraar tot ervaren of zelfs excellent leraar.

Het competentiemodel speelt een cruciale rol als instrument om onderwijsinstellingen in staat te stellen de gewenste bekwaamheden van leraren te formuleren en te beoordelen. Integraal personeelsbeleid impliceert dat het management van een onderwijsinstelling eerst haar visie moet ontwikkelen op de doelstellingen van de instelling. Deze doelen kunnen vervolgens vertaald worden in competenties die van de medewerkers worden verlangd. Als een gewenst competentieprofiel is vastgesteld, wordt dit met de docent besproken in een gesprekkencyclus die uit een doelstellingsgesprek, een aantal functioneringsgesprekken en een beoordelingsgesprek bestaat. Naar aanleiding van het doelstellingsgesprek waarbij het huidige profiel van de docent vergeleken wordt met het gewenste profiel, wordt een persoonlijk ontwikkelingsplan van de docent opgesteld. De ontwikkeling van de docent en de realisatie van het persoonlijk ontwikkelingsplan is onderwerp van het functioneringsgesprek. Realisatie van het persoonlijk ontwikkelingsplan is mogelijk door het volgen van opleidingen, het laten begeleiden door ervaren of excellente collega's (coaching) etc. Het doel van het beoordelingsgesprek is vast te stellen in hoeverre de in het functioneringsgesprek gemaakte afspraken zijn nagekomen en

of de gestelde doelen bereikt zijn. Naar aanleiding van het beoordelingsgesprek wordt voor elk van de tien competenties aangegeven welk beheersingsniveau de docent heeft bereikt. De competenties worden beoordeeld als: uitstekend, goed, voldoende en onvoldoende. Op basis van voorgaande beoordelingen worden de volgende competentieniveaus van docenten onderscheiden: basis, ervaren en excellent niveau. Bij verschillende competentieniveaus horen ook verschillende taken en verschillende salarissen. Uit het voorgaande blijkt dat het hier om het nemen van competentie-managementbeslissingen gaat. Door Twisk e.a (1999, p. 48) wordt aangegeven dat onderwijssectoren en instellingen het model op verschillende manieren kunnen aanpassen aan hun eigen behoeften. De aanpassing die wij hier zouden willen voorstellen is dat bij het beoordelen van competenties van docenten expliciet rekening gehouden wordt met evaluaties van docenten door leerlingen en met door leerlingen behaalde resultaten op toetsen, tentamens en examens. Voor evaluaties door leerlingen bestaan verschillende leerling-vragenlijsten die hun waarde bewezen hebben. Het is opmerkelijk dat geen van de in dit hoofdstuk beschreven competentie-instrumenten de door leerlingen behaalde resultaten bij de beoordeling van de

docent betreft terwijl het onderwijs van de docent toch met name beoordeeld zou moeten worden op diens bijdrage aan het leren van de leerlingen.

Tot slot

Afgaande op de programmering van congressen en studiedagen, op de inhoud van onderwijstijdschriften, op de gesprekken in veel docentenkamers en op de aard van in gang gezette onderwijsinnovaties, lijkt er geen andere conclusie mogelijk dan dat competentiegericht opleiden 'leeft' in kringen van onderwijsgeevenden en opleiders. De levensvatbaarheid moet echter nog blijken en zal mede afhangen van de vraag of men er in zal slagen om de instrumenten te ontwikkelen die nodig zijn om de intake, voortgang en afsluiting van competentiegerichte onderwijstrajecten in goede banen te leiden. In dit boekje hebben we proberen duidelijk te maken dat het vaststellen van competenties geen eenvoudige zaak is. Dat zal voor menigeen een bittere teleurstelling zijn. Dacht men op basis van competenties eindelijk op eenvoudige en eenduidige wijze de bokken van de schapen te kunnen scheiden, blijken vraagstukken van toetsing en afsluiting eigenlijk nog lastiger te liggen dan bij de meer traditionele vormen van opleiden.

We zijn ons ervan bewust dat dit boekje hiervoor

geen oplossing kan bieden; het kan de complexiteit van de problematiek niet verkleinen. Wèl hopen we dat het een bijdrage kan leveren aan de bewustwording van de problematiek en ook dat het aanleiding mag geven tot discussie onder docenten en, in het verlengde daarvan, samenwerking op het gebied van de ontwikkeling van instrumenten. Zeker als dat laatste het geval zou zijn, is het verstandig rekening te houden met de volgende aanbevelingen die alle voortkomen uit de evaluaties van bestaande instrumenten in hoofdstuk 4.

- Meer aandacht voor het formuleren van competenties. Het is ons opgevallen dat ontwikkelaars op basis van hetzelfde uitgangsmateriaal (startbekwaamheden) vaak tot verschillende formuleringen van competenties komen. En ook dat er zaken worden genoemd die volgens ons meer te maken hebben met persoonlijkheidseigenschappen dan met competenties, zoals bijvoorbeeld 'flexibiliteit' en 'reflectief vermogen'. (Dimensies en situaties van het AC worden dan niet goed uit elkaar gehouden; zie paragraaf 2.4.1). Waarschijnlijk wordt met 'flexibiliteit' bedoeld dat de (aanstaande) docent in bepaalde (relevante) situaties in staat is signalen op te pikken die nopen tot het

afwijken van geleerde en beproefde standaard-procedures. Dergelijke situaties zouden gespecificeerd moeten worden en in één of meerdere instrumenten aan bod moeten komen. Dat vergroot de kans om de bedoelde competentie te 'meten'.

- Aandacht voor leerlingprestaties. In geen van de besproken en beoordeelde instrumenten wordt aandacht besteed aan de resultaten van de inspanningen van docenten. Toch vinden wij, met name bij het beoordelen van praktiserende docenten, dat dit aspect aandacht zou moeten krijgen. Niet dat docenten beoordeeld moeten worden door de prestaties van hun leerlingen te vergelijken met een of andere (nationale) standaard, wel dat van tijd tot tijd gekeken zou moeten worden naar de vorderingen (ten opzichte van een eerder meetmoment) van de leerlingen als onderdeel van het bewijs voor competentie.
- Aandacht voor kennis van het vak of vakken die de docent moet verzorgen. Waar men er vroeger vaak vanuit ging dat gedegen kennis van het te onderwijzen vak voldoende was om een goed docent te zijn, lijkt nu de aandacht te eenzijdig gericht op didactische en algemeen-professionele competenties.

In de reguliere lerarenopleiding speelt dit probleem niet, maar bij zij-instromers wel degelijk.

Interessante websites

De volgende websites bevatten informatie over onder andere startbekwaamheden, zij-instromers, competentiebeoordelingen en andere assessments.

<http://www.stoas.nl>

<http://www.toetswijzer.nl>

<http://www.kenniscentrumevc.nl>

<http://www.citogroep.nl>

<http://www.feo.hvu.nl/stage>

<http://portfolioinfo.efa.nl>

<http://www.minocw.nl>

<http://www.educatiefpartnerschap.nl>

<http://www.wordleraar.nl/>

Literatuur

Ackley, B.C., & Arwood, E.L. (1999). The goals of effective preservice teacher education candidates. A longitudinal study. Unpublished manuscript.

Als Docent in Beeld (2000). Utrecht: Faculteit Educatieve Opleidingen (Hogeschool van Utrecht).

Beirendonck, L. van. (1999). Beoordelen en ontwikkelen van competenties. Leuven: Acco.

Bie, D. de, & Mostert, P. (2000). Competentie, over de overbodigheid van een modieus begrip. Onderzoek van Onderwijs, juni.

Bos, E.S. (1998). Competentie: verheldering van een begrip. OTEC-rapport 98/Ro1. Heerlen: Open Universiteit.

Broek, L. van den, Giessen, R. van der, & Oers-van Dorst, A. van. (2000). Performance management. Groningen: Samson.

De Balans. (2000). Arnhem: Pabo Arnhem (HAN).

Dierick, S., Dochy, F., & Van de Watering, G. (2001). Assessment in het hoger onderwijs: over de implicaties van nieuwe toetsvormen voor de edumetrie. Tijdschrift voor Hoger Onderwijs, 19, 1, 2-18.

Dietze, A., Jansma, F., & Riezebos, A. (2000). Een kijkkader voor competenties voor de tweedegraads lerarenopleidingen (concept). Programma-management Educatief Partnerschap.

Elsen, M., Wolters, I., Pilot, A., & Nedermeijer, J. (2001). Geschiktheid van toetsvormen voor het meten van academische competenties. Tijdschrift voor Hoger Onderwijs, 19, 1, 33-43.

Fitzpatrick, R., & Morrison, E.J. (1971). Performance and Product Evaluation. In R.L. Thorndike (Ed.). Educational Measurement (pp. 237-270). Washington DC: American Council on Education.

Frederiksen, N. (1984). The real test bias: influences of testing on teaching and learning. American Psychologist, 39, 193-202.

Gagné, R.M. (1985). The conditions of learning and theory of instruction. New York: Holt, Rinehart & Winston.

Glaser, R. (1991). Expertise and assessment. In M.C. Wittrock & E.L. Baker (Eds.). Testing and Cognition. Englewood Cliffs, NJ: Prentice Hall.

Gonczi, A. (1994). Competency based assessment in the professions in Australia. *Assessment in Education*, 1, 1, 27-44.

Groot, A.D. de, & Naerssen, R.F. van (1977). Studietoetsen: construeren, afnemen, analyseren. Deel II. Den Haag: Mouton Publishers.

Hager, P. & Gillis, S. (1995). Assessment at higher levels. In Hall, W.C. (Ed.). Key aspects of competency-based assessment. Adelaide: NCVER.

Hooft, E.A.J. van. (1999). 360-graden beoordeling bij GITP. Een onderzoek naar de betrouwbaarheid en validiteit. Tilburg: Katholieke Universiteit Brabant, Departement Personeelspsychologie.

Instituut voor Leraar en School (2000). Projectbrief 2. DDC Docenten Development Center. <http://www.ils.han.nl>.

Houston, W.R. (1974). Exploring Competency Based Education. Berkeley (CA): McCutchan.

International Task Force on Assessment Center Guidelines (2000). Guidelines and Ethical Considerations for Assessment Center Operations. Endorsed by the 28th International Congress on Assessment Center Methods. San Francisco, Ca.

Jansen, P.G.W. (1991). Het beoordelen van managers. Effectiviteit van assessment-center-methodes bij selectie en ontwikkeling van managers. Baarn: Nelissen.

Jansen, P.G.W. (1993). Opleiding en ontwikkeling met de assessment-center methode. In: H. Tillema (Red.). Assessment en opleiden in organisaties. Opleiders in organisaties. Capita Selecta afl. 14. (pp. 23-45) Deventer: Kluwer Bedrijfswetenschappen.

Jellema, F. (2000). Toepassing van 360-graden feedback in Nederlandse organisaties. *Opleiding & Ontwikkeling*, 13, 7/8, 21-25.

Kammen, D. van, Melisse, R., & Oudkerk-Pool, I. (2000). Beoordeling als koersbepaler voor een competentiegericht curriculum. In Idsinga, Tj. (Red.). *Assessments in de lerarenopleidingen*. Brochure naar aanleiding van de conferentie op 22 februari 2000. (pp. 34-43) Leeuwarden: Repro NHL.

Kane, M.T. (1992). The validity of assessments of professional competence. (ERIC Document Reproduction Service No. ED 343 958).

Kane, M.T., Crooks, T., & Cohen, A. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice*, 18 (summer), 5-17.

Klarus, R., Schuler, Y., Eising, M., & Groenen, N. (2000). *Bewezen geschikt: instrumentenmap*. Wageningen: Stoas.

Klarus, R., Schuler, Y., Wee, E. ter, Eising, M., Drijfhout, P., & Tillema, H. (2000). *Bewezen geschikt*. Wageningen: Stoas.

Krol, J.D. (1992). Computer en triage: gewonden uit een databank. *COKL-blad*. Maandblad van het commando opleidingen Koninklijke Landmacht, maart 1992.

Maatwerk voor morgen: het perspectief van een open arbeidsmarkt. (1999). Den Haag: Ministerie van Onderwijs, Cultuur en Wetenschappen.

Madaus, G.F. (1988). The influence of testing on the curriculum. In Tanner, L. (Ed.). *Critical Issues in Curriculum*. The 8th Yearbook of the National Society for the Study of Education. Chicago: University of Chicago Press.

Maessen de Sombreff, P. van der, Westen, G., & Veer, J. de (2000). De validiteit van de SQ-test. *Gids voor Personeelsmanagement*, 7/8, 38-43.

Oudkerk-Pool, I., & Snoek, M. (2000). *Beoordeling als koersbepaler voor een competentiegericht curriculum*. Amsterdam: Educatieve Faculteit Amsterdam.

Pelkmans, A. (1998). Evaluatie van leraren en schoolleiders. Wageningen: Vakgroep Agrarische Onderwijskunde.

Richert, A.E. (1990). Teaching teachers to reflect: A consideration of program structure. *Journal of Curriculum Studies*, 22, 6, 509-527.

Rijke, T.R. de, & Dochy, F.J.R.C. (1995). Assessment Centers in bedrijf en onderwijs: stand van zaken. In Dochy, F.J.R.C. & Rijke, T.R. de (Red.). *Assessment Centers. Nieuwe toepassingen in opleiding, onderwijs en HRM.* (pp. 17-38) Utrecht: Uitgeverij LEMMA BV.

Sanders, P.F. (1998). Generaliseerbaarheidstheorie. In Van den Brink, W.P., & G.J. Mellenbergh (red.). *Testleer en testconstructie* (p. 93-131). Amsterdam: Boom.

SLO/VSLPC (1997). *Startbekwaamheden leraar primair onderwijs.* Utrecht: APS.

Smith, H.A., & Pratt, D. (1996). The use of biodata in admissions to teacher education. *Journal of Teacher Education*, 47, 1, 43-52.

Straetmans, G.J.J.M. (1993). Het vaststellen van competentie met vaardigheidstoetsen. *Tijdschrift voor Hoger Onderwijs*, 11, 3, 188-204.

Straetmans, G.J.J.M. (2000). *Assessment van vaardigheden door middel van zelfbeoordeling.* Arnhem: Citogroep.

Thijssen, J. (2000). Competentiemanagement en HR-beleidsfuncties. *Opleiding & Ontwikkeling*, 13, 7/8, 13-19.

Tillema, H. (2001). Gewaardeerd meesterschap. *Velon tijdschrift voor lerarenopleiders*, 22, 2, 36-43.

Twisk Th. F., Daniëls, J.J.M.C., & Bolweg J.F. (1999). *Competenties van docenten.* Utrecht: Berenschot.

Wiggins, G.P. (1993). *Assessing Student Performance. Exploring the Purpose and Limits of Testing.* San Francisco: Jossey-Bass Publishers.

is

e hecta-
komen
kensap-
et plan,
den ge-
hampig-

nden de
r woor-
l gek ge-
lf gigan-
le bio-in-
rvolgens
en in een
bio-indu-
gt Driek
erste Ka-
wordt ge-
tfabriek.
de dieren

cht bij te
ens bege-
choloog,
erzoeken

Delft: spoedcursus nieuwe leraren

DELFT — Met een spoedcur-
sus didactiek wil de Technische
Universiteit Delft een bijdrage
leveren aan de bestrijding van
het lerarentekort. De eerste ze-
ven nieuwe vwo-leraren, die na
de zomervakantie met een cur-
sus zijn begonnen, staan al voor
de klas. Een groep van 35 deel-
nemers begint binnenkort met
de basisopleiding, die twee we-
ken duurt.

Derdejaarsstudenten en do-
centen van de TU, die een deel-
tijdbaan als leraar ambiëren,
komen in aanmerking voor de
spoedopleiding. Zij zetten hun
opleiding voort via internet. De
TU wil de opleiding ook open-
stellen voor academici in de
exacte wetenschappen uit het
bedrijfsleven die voor een deel
van hun werktijd leraar willen
zijn. (ANP)

Boeing-747 toestellen kunnen
zou halen (ANP)

Pabo'ers willen geld voor stage

DEN HAAG — Pabo-studen-
ten eisen geld voor hun werk
als leraar-in-opleiding (lio). Zij
komen op 9 februari in actie en
boeden minister Hermans van
Onderwijs een petitie aan.

Sommige lio-studenten krij-
gen betaald voor hun werk,
maar anderen krijgen niets.
Dat verschil is mogelijk omdat
de basisscholen zelf beslissen of
ze de lio's een salaris betalen.

Tijdens hun periode als le-
raar in opleiding werkt een pa-
bo-student 21 weken lang full-
time in dienst van een school.
Daarvan staat hij drie dagen
per week voor de klas en woont
hij net als de leraren vergade-
ringen en oudersavonden bij.

Om die reden vinden de stu-
denten dat er een einde moet
komen aan de in hun ogen
"rechtsongelijkheid". Zij eisen
dat de lio-stage wordt betaald
met een arbeidscontract.

Zij hebben inmiddels bijval
gekregen van diverse Pabo-
directies. Ook de Landelijke Stu-
denten Vakbond en de Algeme-
ne Onderwijshoud scharen zich
achter de studenten. (HOP)

Twee Limburgse

maandag 12 februari 2001

Bijcholingsplicht voor docenten

DEN HAAG — De tijd dat leraren hun papiertje haalden en vervolgens een leven lang voor de klas mochten staan is voorbij. Docenten moeten na- en bijscholing volgen om bij te blijven. Werkgevers zijn verplicht de scholing aan te bieden.

De kwaliteitseisen voor onderwijzend en ondersteunend personeel, worden vastgelegd in de Wet op de beroepen in het onderwijs, die in de loop van dit jaar naar de Tweede Kamer wordt gestuurd. Het kabinet heeft afgelopen vrijdag met de uitgangspunten voor deze wet ingestemd.

Het kabinet wijkt met het voorstel af van het regeerakkoord, waarin stond dat er een register van leraren zou komen. Minister Hermans voelt niet voor zo'n lijst. (ANP)

VVD
espro-
erak-
erpre-
kama
m van
it ge-
ans
seve
an-
be-
De
nan
wer-
Ge-
van

Opleidingen hebben geen plaats voor zij-instromers

ONDERWIJS Minister Hermans van Onderwijs roept de hulp van commerciële instellingen zoals de LOI in om bij te springen bij het omscholen van zij-instromers. De lerarenopleidingen kunnen de stroom van werknemers uit het bedrijfsleven die leraar willen worden niet aan. Er staan nu 11000 gegadigden ingeschreven bij bemiddelingsorganisaties. In de eerste maanden van dit schooljaar waren er ongeveer 250 vacatures in het basisonderwijs, en 900 in het voortgezet onderwijs. (ANP)

Vier grote steden krijgen € 3,6 miljoen voor zij-instromers

AMSTERDAM — De vier grote steden krijgen € 3,6 miljoen (1,6 miljoen) extra om het tekort aan leraren in het primair onderwijs aan te pakken. In tien maanden zullen Amsterdam, Rotterdam, Den Haag en Utrecht tweehonderd mensen uit het bedrijfsleven omscholen tot leraar. Dat staat in het plan Vijfwant dat minister P...

Pabo's minder in trek

GRONINGEN — De pogingen om het beroep van leraar aantrekkelijker te maken, hebben nog niet tot enthousiasme geleid onder jongeren met een middelbareschooldiploma. Het aantal studenten dat zich heeft gemeld voor een opleiding tot leraar in het basisonderwijs (pabo) ligt dit jaar liefst 16 procent lager dan vorig jaar begin augustus.

In totaal hebben 6620 voltijdstudenten zich aangemeld. Het aantal deeltijd eerstejaars blijft met 1824 praktisch gelijk. Dat blijkt uit ge-

Plan Hermans voor lerarenopleidingen

DEN HAAG — Minister Hermans van Onderwijs komt voor 15 mei met een plan voor de lerarenopleidingen. Hierin geeft hij aan wat hij met kleine, noodlijdende faculteiten wil en hoe de opleidingen een effectievere rol kunnen spelen bij het omscholen van mensen uit het bedrijfsleven tot leraar.

PvdA-Kamerlid Hamer had gisteren tijdens een debat om een noodplan gevraagd. Zij hoopt dat Hermans met geld over de brug komt, zodat de minst populaire opleidingen hun hoofd boven water houden. Verder moet de minister in kaart brengen hoe hij de instroom wil verhogen.

Weinig studenten kiezen voor de tweedegraads opleiding tot leraar in het voortgezet onderwijs. (ANP)

Ne

VOORE
zijn we
In janua
particul
jard (/ 6,1
keningen.

maandbedr
blijkt uit cij
Bureau voor

Volgens een
het CBS heeft
ouderwetse
te maken m
fectenber
gers nog
zekere
naar

Opleiding tot
stand,un

van HTP-Kamerlid Stellingwerf. Nieuwambus kondigt maatregelen aan. Ze ontwikkelt een taalproef. Dat kunnen schouwers tonen als bij aanhouding in het buitenland een verklaring voor het blauwe kaartje wordt gevraagd.

Beginnende docenten drukken prestaties

ERIST — Middelbare scholen die veel beginnende docenten of invallers in dienst hebben, presseren slechter. De examenresultaten zijn lager en het aantal uitvallers en zitverlopers is hoger. Dat constateert Trouw-journalist Marjan Agerbeek in de 'Kortsgedrukte Middelbare Scholen 2001/2002'. Hiermee kunnen ouders, leerliniaars en scholen de onderwijsinstellingen op diverse criteria met elkaar vergelijken. Het is de tweede keer dat Agerbeek een gids publiceert. (ANP)

MÁXIMA — Het Tweede-Kamerlid Heilken (CDA) wil weten wat het kabinet vindt van...

Hermans wil minimale kwaliteit van leraren in de wet regelen

De minimale kwaliteit die een leraar moet hebben, komt in de wet te staan. Minister Hermans (Onderwijs) wil dit vastleggen in de Wet op de Ouderwetse Opleiding van Leraren (WOL).

Hermans wil de kwaliteit van de leraars vaststellen en de instellingen die hen opleiden en aanstellen op de hoogte brengen van de wet. Dit kan de kwaliteit van de leraars verbeteren en de kwaliteit van de onderwijsinstellingen.

Minister Hermans en staatssecretaris...

Lerarentekort piekt vanaf 2002

DEN HAAG — Tussen 2002 en 2004 is de behoefte aan leraren het grootst. In 2002 wordt de piek in het primair onderwijs verwacht; twee jaar later kent het voortgezet onderwijs het grootste aantal vacatures.

Volgens de Arbeidsmarkt-effectrapportage 2000, die minister Hermans (Onderwijs) gisteren naar de Tweede Kamer stuurde, wordt een kleinschalig beroepsgebaar op herintreders en mensen uit het bedrijfsleven die geen nog geen specifieke opleiding als leraar hebben. De behoefte aan de laatste categorie neemt toe omdat de stille reserve van rederwijskrachten langzaam aan afneemt.

De spanning op de arbeidsmarkt neemt toe doordat er meer leraren komen, verwacht Hermans. In het basisonderwijs stijgt de werkgelegenheid tot 2004 met 7,5 procent en zijn jaarlijks gemiddeld 8000 nieuwe leraren nodig. Als gevolg van klassenverkleining is dat in 2002 al 10 miljoen. De werkgelegenheid in het voortgezet onderwijs stijgt met 6 procent, waardoor jaarlijks 1000 nieuwe leraren moeten worden aangetrokken. (ANP)

Kok oversteelt met e-mails klimaatrapport

Managers worden schooldirecteur

AMSTERDAM — Managers die hun kennis inzetten voor de school, worden schooldirecteur. Dit is de bedoeling van de Tweede Kamer. De Tweede Kamer heeft de Tweede Kamerlid Heilken (CDA) gevraagd om te onderzoeken hoe de schoolleiding kan worden verbeterd. De Tweede Kamer heeft de Tweede Kamerlid Heilken (CDA) gevraagd om te onderzoeken hoe de schoolleiding kan worden verbeterd. De Tweede Kamer heeft de Tweede Kamerlid Heilken (CDA) gevraagd om te onderzoeken hoe de schoolleiding kan worden verbeterd. (ANP)

Pabo moet studenten weigeren

MEPPEN — De pabo in Meppel moet studenten weigeren die niet voldoen aan de eisen. De Tweede Kamer heeft de Tweede Kamerlid Heilken (CDA) gevraagd om te onderzoeken hoe de schoolleiding kan worden verbeterd. De Tweede Kamer heeft de Tweede Kamerlid Heilken (CDA) gevraagd om te onderzoeken hoe de schoolleiding kan worden verbeterd. (ANP)

hoo meer te stimuleren. (ANP)

Universiteit Delft test docenten

DELFT — Docenten aan de Technische Universiteit Delft moeten elke vijf jaar bewijzen dat ze goed lesgeven. De studentenraad is blij, want aan de gemiddelde docent valt "een boel te verbeteren".

Dat zegt Roel van Raak, lid van de studentenraad. Volgens deze raad schenkt de TU Delft te weinig aandacht aan onderwijs. De universiteit vindt het belangrijker of iemand een goede onderzoeker is. Daardoor maken Delftse docenten nogal eens een foutje. Voortaan moeten nieuwe docenten eerst op examen. Dat heeft het college van bestuur toegezegd. Ze moeten bovendien eens in de vijf jaar 'op herhaling'. (HOP)

'Pabo doet nog te weinig aan internet'

MEPPEL — Terwijl veel leerlingen al surfen op het web, moeten veel leraren nog leren e-mailen. De lerarenopleidingen doen nog te weinig om hen beter te scholen. Dat blijkt uit onderzoek dat minister Hermans (Onderwijs) bij de opening van de Educatieve Kinderstudio in Meppel in ontvangst nam. Scholen zijn te veel gericht op de techniek en te weinig op de inhoud, is de conclusie.

De educatieve studio is onderdeel van de pabo in Meppel. Kinderen, begeleid door pabo-studenten, kunnen daar werken met computers. Zo doen zowel kinderen als aankomend leraren ervaring op. (HOP)

Eerder in deze reeks verschenen:

EPS-reeks 01 - oktober 2000 - "Advies Werkgroep Verkenning OPLEIDING LERAAR FUNDEREND ONDERWIJS"

EPS-reeks 02 - maart 2001 - "Het Zweedse model - Een nieuw systeem voor het opleiden van leraren"

EPS-reeks 03 - april 2001 - "Landelijk Programmamanagement Educatief Partnerschap - Voortgangsrapportage 2000"

EPS-reeks 04 - april 2001 - "Naar een vernieuwde samenwerking tussen lerarenopleiding en school"

Colofon

publicatie: Programmamanagement EPS (i.s.m. HBO-raad)

(eind)redactie: Boezeman Teksten en trainingen · Lex Boezeman

vormgeving en productiebegeleiding: Jack of all Trades · Marion E. Vegter

druk: Drukkerij van Mechelen · Utrecht

uitgave: oktober 2001