

Integratie van instructie en assessment in competentiegericht onderwijs

Gerard Straetmans

Dominique Sluijsmans

Bart Bolhuis

Jeroen van Merriënboer

Dr. G.J.J.M. Straetmans

(gerard.straetmans@citogroep.nl)

is toetsdeskundige bij de Citogroep, unit BVE en HO te Arnhem.

Dr. D. Sluijsmans is onderwijstechnoloog bij het Onderwijstechnologisch Expertisecentrum van de Open Universiteit te Heerlen.

B.G. Bolhuis is hoofdinspecteur van politie en werkt bij LSOP Politie Onderwijs en Kenniscentrum te Apeldoorn.

Prof. dr. J.J.G. van Merriënboer is hoogleraar en programmaleider bij het Onderwijstechnologisch Expertisecentrum van de Open Universiteit te Heerlen.

Het beroepsonderwijs is toe aan een ander onderwijsconcept. Decennia lang al worstelen zowel middelbaar als hoger beroepsonderwijs met een tegenvallend intern en extern rendement. De tijd van symptoombestrijding lijkt voorbij nu allerlei commissies en het ministerie van OC&W steeds vaker op één lijn zitten met het uitgangspunt dat de deelnemer centraal moet staan in een beroepsopleiding die opleidt tot responsief vakmanschap. De vraag is hoe dergelijk onderwijs gestalte moet krijgen. Er is grote behoefte aan de ontwikkeling van onderwijsmodellen die de vrij abstracte uitgangspunten kunnen helpen concretiseren. In dit artikel wordt beschreven hoe competentiegericht onderwijs kan worden opgezet door elementen uit een constructivistisch georiënteerd instructieontwerpmodel te integreren met moderne opvattingen over assessment. Eerst wordt aan de hand van een voorbeeld duidelijk gemaakt hoe competentiegericht onderwijs kan worden opgezet volgens het 4C/ID-model. Dit is een instructieontwerpmodel voor onderwijs waarin studenten complexe vaardigheden (competenties) verwerven door te werken aan authentieke taken. Onderwijs dat is opgezet volgens het 4C/ID-model lijkt bij uitstek geschikt voor een andere assessmentfilosofie, die ervan uitgaat dat het bewijs voor competentie tijdens het onderwijs wordt verzameld door studenten op gezette tijden authentieke taken onder assessmentcondities te laten uitvoeren. De resultaten worden opgeslagen in een zogeheten beoordelingsportfolio dat volgens een bepaalde methodiek, Protocol Portfolio Scoring, wordt aangelegd en beoordeeld.

INLEIDING

Alle vernieuwingen van de laatste decennia ten spijt is het heersende onderwijsconcept dat binnen het beroepsonderwijs gehanteerd wordt in die periode nauwelijks veranderd. Het is nog steeds de docent die de lijnen van het onderwijs uitzet en de deelnemer dwingt om de 'voorgebakken' leerstof te verwerven op een manier die eerstgenoemde het beste acht. Het is nog steeds de theorie die voor de praktijk uitgaat terwijl er genoeg aanwijzingen zijn dat deelnemers grote moeite hebben met deze kunstmatige scheiding. Beslissingen over zakken of slagen worden nog altijd voor een belangrijk deel gebaseerd op het verworven zijn van kennis en vaardigheden waarvan de relatie met het werk van de beginnend beroepsbeoefenaar lang niet altijd duidelijk is.

Dat dit traditionele onderwijsconcept, met name voor het beroepsonderwijs, zo langzamerhand aan vervanging toe is, blijkt onder andere uit het feit dat de niet aflatende stroom van maatregelen die in de afgelopen decennia zijn afgekondigd om twee kernproblemen aan te pakken, te weten een tekort schietend intern rendement en een gebrekkig extern rendement, te weinig effect hebben gesorteerd. Het interne rendement wordt gemeten als het percentage studenten van een bepaald instroomjaar dat de studie met een diploma afsluit. Het externe rendement heeft te maken met de waarde die het diploma heeft op de arbeidsmarkt. In de jaren '70 van de vorige eeuw waren werkgevers niet enthousiast over het niveau van beginnende beroepsbeoefenaars. De uitspraak dat deze nog geen hamer konden vasthouden was een aanklacht tegen de, in de ogen van de werkgevers, veel te algemene (theoretische) beroepsopleiding. De instelling van de commissie Wagner (1982) markeert het begin van een periode van onderwijsinnovaties die ertoe moesten leiden dat het beroepsonderwijs beter werd afgestemd op de eisen van de arbeidsmarkt. Maar ook deze veranderingen, die onder meer geleid hebben tot een kwalificatiestructuur met eindtermen voor elke beroepsopleiding in het mbo, hebben niet het verwachte resultaat opgeleverd. Zo constateert de Inspectie in 2002 dat bijna de helft van de mbo-opleidingen de studenten niet goed voorbereid op de beroepspraktijk.

Inmiddels is het voor sommigen duidelijk dat er meer substantiële veranderingen noodzakelijk zijn, wil Nederland een vooraanstaande rol kunnen spelen in de plannen van de Europese Unie om in 2010 de meest concurrerende regio van de wereld te zijn. Mertens (1998) en Meijers (1998) pleiten ervoor om niet de studenten aan de arbeidsmarkt aan te passen, maar deze de instrumenten te geven waarmee ze het lot in eigen hand kunnen nemen. De kwalificatiestructuur is in die opzet geen dwingend voorgeschreven einddoel maar hooguit een richtsnoer voor de persoonlijke opleiding en ontwikkeling. Het Ministerie van Onderwijs laat in 1999 soortgelijke geluiden horen met de introductie van het begrip *responsief vakmanschap*. In dit begrip komen diverse innovatieve thema's voor het beroepsonderwijs samen, zoals: de aansluiting onderwijs-arbeidsmarkt, een leven lang leren, werkend leren en de erkenning van elders of eerder verworven competenties (EVC). Responsief vakmanschap is vakmanschap dat tegemoetkomt aan de behoeften van de arbeidsmarkt en tegelijkertijd het individu de bagage biedt waarmee deze zichzelf en het vakmanschap verder kan ontwikkelen om te kunnen voldoen

aan de eisen van een snel veranderende maatschappij en arbeidsmarkt. Ook de commissie Boekhoud (Adviescommissie Boekhoud, 2002) zit op dat spoor met het benadrukken van de loopbaangedachte als leidend principe voor de inrichting van beroepskwalificerende trajecten van individuele studenten. De commissie hinkt echter op twee gedachten door tevens de versterking van de beroepskolom te propageren, wat toch weer meer past in een traditionele aanbodvisie. Vanuit het perspectief van de opleiders is dit wel verklaarbaar aangezien het voor hen op dit moment nog onduidelijk is hoe vraaggestuurd beroepsonderwijs in organisatorische zin zijn beslag moet krijgen. Het zal daarom nog wel even duren voordat vernieuwers en opleiders elkaar gevonden hebben in een nieuw evenwicht tussen het institutionele denken (versterking van de beroepskolom) en het individuele denken (bevorderen van de loopbaan).

Hoe het beroepsonderwijs van de toekomst er uit gaat zien, zal nog wel enige tijd onduidelijk blijven. De Bruijn (2003) pleit voor een snelle ontwikkeling van operationele onderwijsmodellen die de loopbaanbenadering kunnen concretiseren en zo de discussie daarover kunnen verhelderen. In die modellen zal aandacht besteed moeten worden aan een andere benadering van de student (de student wordt beschouwd als een beginnend beroepsbeoefenaar, niet als een naïeve student), aan een andere didactiek (geen hapklare brokken consumeren maar zelf sturing geven aan het eigen leerproces) en aan een andere planning van het beroepsonderwijs (niet de eindtermen of vakken vormen de bouwstenen van het opleidingstraject maar de zich ontwikkelende competenties die passen bij de ambitie van de individuele student). In deze bijdrage wordt beschreven hoe elementen uit een constructivistisch georiënteerd instructieontwerpmodel gecombineerd kunnen worden met nieuwe inzichten op het gebied van assessment om te komen tot competentiegericht onderwijs. Achtereenvolgens komt aan de orde wat verstaan wordt onder competenties en competentiegericht opleiden, hoe het leren op basis van authentieke taken georganiseerd kan worden en wat daarvan de implicaties zijn op het gebied van toetsing en afsluiting.

COMPETENTIEGERICHT ONDERWIJS ONTWERPEN VOLGENS HET 4C/ID MODEL

Sinds een jaar of 5 is onderwijskundig Nederland in de ban van de *competentiebeweging*. Zoals wel vaker gebeurt in de sociale wetenschappen is er in die eerste jaren vooral een strijd gevoerd over de betekenis van het begrip. Inmiddels heeft de Onderwijsraad met een publicatie geprobeerd hierover duidelijkheid te verschaffen en weten we dat er geen generieke definitie gegeven kan worden (Onderwijsraad, 2002). De betekenis en reikwijdte van het begrip competentie zijn sterk contextgebonden. Waar het in de HRM-wereld vooral verwijst naar meer stabiele persoonlijke eigenschappen zoals creativiteit, communicatief vermogen en leiderschap, wordt het begrip in het beroepsonderwijs meer geïnterpreteerd in de zin van beroepsspecifieke bekwaamheden. Nu deze discussie is afgerond wordt het tijd ons meer te concentreren op waar het natuurlijk om gaat: de waarde van het begrip voor het vormgeven van het beroepsonderwijs.

Simpel gezegd is een competentie de bekwaamheid van een persoon om dat wat hij geleerd heeft door formele scholing en ervaring op een bewuste en verantwoorde wijze en met goed resultaat toe te passen in bekende en onbekende taaksituaties. Competenties zijn geen eenvoudige optelsom van kennis- en vaardigheidselementen. Rijvaardigheid, bijvoorbeeld, is meer dan de optelsom van beheersing van de theorie, kunnen sturen, kunnen schakelen, beheersing van de bijzondere verrichtingen, anticiperend rijden, enzovoort. De ervaring maakt dat je *feeling* krijgt voor wat er in bepaalde verkeerssituaties moet gebeuren, ook als je nog nooit eerder in dergelijke situaties verkeerd hebt. Met andere woorden: competentie verwerf je door veel te oefenen. Competentiegericht opleiden moet daarom gebaseerd zijn op het uitvoeren van realistische taken. Dat kan door de praktijk meer naar de school te halen of door studenten de praktijk in te sturen. Maar in beide gevallen moet er een plan zijn om studenten op systematische wijze, via het werken aan taken, competenties te laten verwerven. Een plan dat heldere richtlijnen daarvoor geeft is het 4-componenten-instructieontwerpmodel (4C/ID model), oorspronkelijk ontwikkeld door Van Merriënboer e.a. (1992). De essentie van het 4C/ID-model wordt gevormd door voorschriften om te komen tot een curriculum waarin theorie en praktijk verregaand geïntegreerd zijn in zogeheten leertaken, die volgens het principe van de *hele-taakbenadering* worden aangeboden. De achtergrond voor deze specifieke aanpak is in beknopt bestek beschreven in Janssen-Noordman, e.a. (2000), waaruit de volgende passages zijn overgenomen:

'Beroepstaken worden niet "zomaar" als leertaken gebruikt. De complexiteit van de hele taak wordt gereduceerd op een manier waarbij de essentie van de hele taak zo veel mogelijk blijft gehandhaafd. Complexe real-life-situaties worden vereenvoudigd tot minder complexe maar nog steeds realistische situaties (...) Alle essentiële aspecten van de taak blijven in combinatie en coördinatie met elkaar in een leertaak vertegenwoordigd. (p. 39)'

'De eerste leertaken in een serie bieden meer ingebouwde ondersteuning dan de laatste. Deze benadering is te vergelijken met de zoomlensmetafoor (...) Als iemand een plaatje door de zoomlens van een camera bestudeert dan zal hij beginnen vanuit een groothoek. Deze laat het hele plaatje zien met de belangrijkste delen en de relatie tussen die delen, maar zonder detail. Inzoomen op de verschillende delen van het plaatje laat meer detail zien van de verschillende onderdelen en de relatie tussen die delen (...) Met deze hele-taakbenadering onderscheidt het 4C/ID model zich fundamenteel van de meer klassieke modellen, waarbij complexe taken worden opgedeeld in deeltaken die min of meer separaat worden aangeboden en geleerd.'

In deze bijdrage zullen we vooral ingaan op de eerste component van het model: de leertaken. Voor een uitgebreide uitwerking van de andere componenten verwijzen wij naar De Croock & Van Merriënboer (2003) en Jansen & Van Merriënboer (2001).

In het 4C/ID model worden competenties opgevat als complexe vaardigheden, waar een set van deelvaardigheden aan ten grondslag ligt. Voorbeelden van dergelijke com-

plexe vaardigheden zijn het diagnosticeren van röntgenfoto's, het bewaken van een chemisch proces, het ontwerpen van een huis en het geven van een bedrijfstraining. Elk van deze vaardigheden kan worden ontrafeld in een aantal deelvaardigheden. Voor de complexe vaardigheid 'het ontwerpen van een huis' bijvoorbeeld zijn belangrijke deelvaardigheden: het maken van een schetsontwerp op basis van de wensen van de klant, het uitwerken van een ontwerp in bouwtekeningen en een bestektekening maken voor de aannemer. Een competent ontwerper kan deze onderliggende vaardigheden *in samenhang* demonstreren. Competentiegericht leren gaat dus niet over het geïsoleerd leren van deelvaardigheden, maar over het leren coördineren en integreren van deelvaardigheden en over het aanpassen van geleerde procedures aan de specifieke vereisten van nieuwe, niet eerder uitgevoerde taken. Onderwijs dat volgens de 4 componenten is vormgegeven, heeft tot doel bij studenten het vermogen te ontwikkelen om het geleerde toe te passen in nieuwe situaties (transfer). Door het onderwijs aan te bieden in de vorm van integratieve en betekenisvolle leertaken zullen studenten de deelvaardigheden steeds opnieuw moeten coördineren en integreren wat hun transfervermogen ten goede komt.

Een beroepsopleiding waarin al in hoge mate gewerkt wordt volgens de principes van competentiegericht onderwijs is die van het LSOP Politie Onderwijs- en Kenniscentrum. Deze organisatie is in 2002 gestart met competentiegericht politieonderwijs. Daarbij is uitgegaan van de competentiedefinitie van de Adviescommissie Onderwijs-Arbeidsmarkt (ACOA, 1999), die een competentie beschrijft als 'het vermogen van een persoon om in beroeps-, (leer- en maatschappelijke) situaties op adequate wijze proces- en productgericht te handelen'. Voor het initiële politieonderwijs heeft de overheid 5 beroepsprofielen vastgesteld. Deze komen overeen met de niveaus in de reguliere kwalificatiestructuur: assistent politiemedewerker (niveau 2 mbo), politiemedewerker (niveau 3 mbo), allround politiemedewerker (niveau 4 mbo), politiekundige bachelor (niveau 5 hbo) en politiekundige master (niveau 6 hbo-master). Binnen elk beroepsprofiel zijn zogeheten kernopgaven geïdentificeerd. Dit zijn de beroepsmatige problemen waarmee een beroepsbeoefenaar regelmatig in aanraking komt en die kenmerkend zijn voor het beroep. Ze moeten aangepakt worden binnen een specifieke organisatorische context, waarbij van de beroepsbeoefenaar een oplossing en een passende aanpak worden verwacht. Voorbeelden van kernopgaven zijn: optreden bij milieuovertreding, arrestantenzorg, toezicht houden in publiek domein, aanpak van verkeersveiligheid, voorbereiding strategische aanpak bij rampen en calamiteiten en organiseren en uitvoeren van opsporingsconfrontatie (LSOP Politie Onderwijs- en Kenniscentrum, 2001).

In de volgende paragraaf wordt uitgelegd hoe competentiegericht opleiden gestalte kan krijgen volgens de principes van het 4C/ID-model. Dit gebeurt aan de hand van een concrete competentie, door LSOP aangeduid als kernopgave en binnen het 4C/ID-model complexe vaardigheid genoemd. De keuze is daarbij gevallen op toezicht houden in een publiek domein.¹ Deze complexe vaardigheid kan ontleed worden in een aantal deelvaardigheden. We hebben ons hier beperkt tot de vakmatige en methodi-

sche deelvaardigheden. Voor een volledig overzicht verwijzen wij naar Bolhuis & Van Royen (2003) en Grotendorst e.a. (2002).

Relevante deelvaardigheden voor het beroep van allround politiemedewerker zijn dat hij in staat is:

- zich zodanig op te stellen dat hij actief aanwezig en aanspreekbaar is voor publiek: 'kennen en gekend worden';
- waar te nemen en overzicht te houden met in achtneming van eigen veiligheid en die van anderen;
- afwijkend c.q. probleemgedrag te signaleren;
- ter plaatse de juiste informatie te kunnen verzamelen (vragen stellen en zelf waarnemen);
- een voorbeeldfunctie te vervullen;
- te handelen conform relevante wet- en regelgeving (proportioneel/discretionair);
- gesprekstechnieken te gebruiken;
- door houding en voorkomen gezag uit te stralen;
- de uitgangspunten van gemeenschappelijke veiligheidszorg in zijn handelen te tonen;
- bij het ontstaan van incidenten conflicten te hanteren, ook bij grotere mensenmassa's en, indien nodig, creatieve oplossingen te vinden;
- Integrale Beroepsvaardigheid Training (IBT) toe te passen (aanhoudingsprocedures en geweldtoepassing).

Uit de beschrijvingen van de deelvaardigheden blijkt dat het vooral gaat om concrete handelingen die de student geïntegreerd en gecoördineerd uit moet kunnen voeren om de complexe vaardigheid 'toezicht houden in een publiek domein' te demonstrenen.

COMPETENTIEGERICHT OPLEIDEN IS OPLEIDEN OP BASIS VAN LEERTAKEN

De kern van het 4C/ID-model is dat het onderwijs wordt vormgegeven op basis van een reeks leertaken (component 1) die studenten moeten uitvoeren. Hierbij staat de heeltaakbenadering centraal, dat wil zeggen dat studenten leertaken uitvoeren die 'hele taak' leerervaringen bieden in een gesimuleerde of echte taakomgeving. In het voorbeeld van 'toezicht houden in een publiek domein' wil dat zeggen dat studenten van meet af aan worden geconfronteerd met de noodzaak van integratie van deelvaardigheden die voor toezicht houden in publiek domein van belang zijn en niet primair de nadruk leggen op het leren beheersen van de afzonderlijke deelvaardigheden. Leertaken zorgen ervoor dat studenten zich bezig houden met activiteiten die van hen vereisen dat ze alle deelvaardigheden die voor een bepaalde competentie van belang zijn, in het vizier hebben. Bij het LSOP worden leertaken voor de politieschool leer- of instituutsofdrachten genoemd en voor het politiekorps korpsopdrachten.

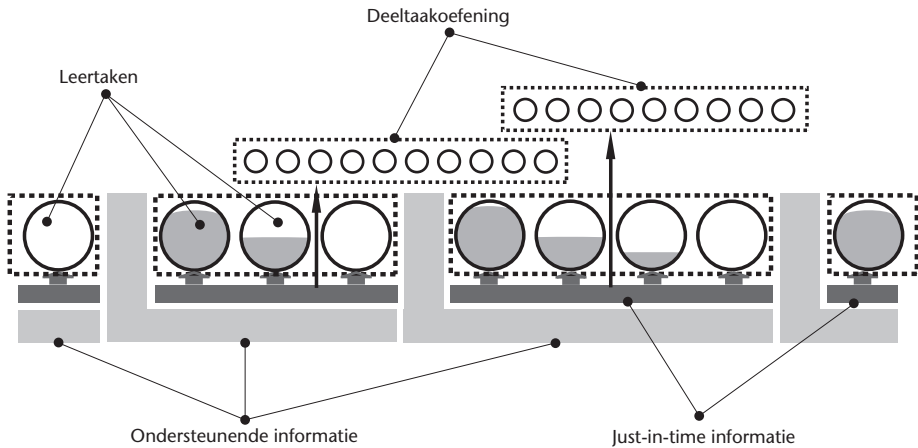
Uiteraard kunnen studenten niet al in het begin van hun opleiding leertaken uitvoeren die zeer complex zijn. Zo zou het bijvoorbeeld niet reëel zijn een beginnende student toezicht te laten houden in een publiek domein waar een politiek gevoelige manifestatie gaande is en waar kans is op dreigende escalatie met verstoring van de openbare orde tot gevolg. Verstandiger is het studenten te laten beginnen met het uitvoeren van relatief eenvoudige leertaken die, naarmate de competentie van een student zich ontwikkelt, steeds complexer worden. Om de student leertaken te laten uitvoeren die qua complexiteit van hetzelfde niveau zijn, maar wel van elkaar verschillen op al die dimensies waarop ook taken in de echte wereld van elkaar verschillen, worden binnen de opleiding *taakklassen* gedefinieerd. Leertaken binnen een bepaalde taakklasse vertonen dus variatie maar zijn wel van dezelfde complexiteit. Voor het bepalen van taakklassen is een aantal zogenaamde complexiteitsfactoren nodig. In tabel 1 worden voorbeelden van complexiteitsfactoren genoemd die van belang zijn voor de complexe vaardigheid 'toezicht houden in publiek domein'.

Tabel 1 Voorbeelden van complexiteitsfactoren voor de vaardigheid 'toezicht houden in publiek domein'

Complexiteitsfactor	Taakklasse 1 (eenvoudige taken)	Taakklasse 2 (complexe taken)	Taakklasse 3
Aantal mensen Omgeving	20-80 openbare ruimte met normaal gebruik, bijv. kerkplein overdag geen risicoplekken	80-200 openbare ruimte met speciale manifestatie, bijv. kerkplein tijdens kermis overdag risicoplekken	>300 openbare ruimte met hoog veiligheidsrisico, bijv. rumoerig uitgaansgebied op zaterdagavond risicoplekken
Voorval	overtreding, bijv. fietsers in voetgangersgebied	diefstal, bijv. zakkenrollen of tasjesroof	geweldsmisdrijf, bijv. schietpartij
Optreden	signaleren	signaleren preventieadviezen geven regulerend	signaleren preventieadviezen geven regulerend
Briefing Onderzoek	geen briefing volgen geen opsporingsonderzoek opstarten	geen briefing volgen geen opsporingsonderzoek opstarten	briefing volgen onderzoek opstarten

Elke leertaak binnen een taakklasse varieert in die zin dat er een ander type probleem is of een andere context wordt gekozen. Op deze manier wordt in de opleiding transfer van leren nagestreefd, dat wil zeggen het vermogen om de complexe cognitieve vaardigheid toe te passen in een grote variëteit van nieuwe, levensechte probleemsitu-

aties. De complexiteit van de leertaken binnen een taakklasse blijft echter min of meer gelijk! Samengevat kan het 4C/ID-model worden geïllustreerd met figuur 1.



Figuur 1 De 4 componenten

De leertaken zijn als bolletjes weergegeven. De taakklassen worden aangeduid door de gestippelde lijnen rond de grote bolletjes. De grijze balken voor en onder een taakklasse geven de ondersteunende informatie (component 2) aan. Met ondersteunende informatie wordt de informatie bedoeld die door docenten veelal 'de theorie' wordt genoemd. De ondersteunende informatie is vooral bedoeld voor het leren en uitvoeren van niet-routinematige aspecten van leertaken. Ondersteunende informatie bestaat uit twee typen kennis: mentale modellen en cognitieve strategieën. Mentale modellen zijn ondersteunend bij vragen als: wat betekent iets? Hoe zit iets in elkaar? Hoe werkt het? Cognitieve strategieën helpen bij het systematisch aanpakken van een probleem, de zogenaamde aanpak kennis. Studenten zullen zowel mentale modellen als cognitieve strategieën moeten aanwenden om problemen op te lossen waarbij niet één vaststaande oplossing voorhanden is. Het is dan ook van belang veel variatie bij leertaken in dezelfde taakklasse aan te brengen, zodat studenten leren de modellen en strategieën in nieuwe situaties adequaat toe te passen.

De korte pijltjes staan voor de just-in-time informatie (component 3) die wordt aangeboden tijdens het uitvoeren van de leertaak; deze informatie is vooral gericht op de verwerving van routinematige vaardigheden. Dit zijn vaardigheden die vooral door herhaling kunnen worden geleerd en die bij juiste uitvoering altijd tot het gewenste gedrag leiden. Binnen het politiedomein is een aantal procedures aan te geven – zoals het uitschrijven van een bekeuring voor te hard rijden – die typische voorbeelden zijn van routinematige vaardigheidsaspecten. Het leren daarvan wordt het beste ondersteund door de benodigde informatie just-in-time, precies wanneer het nodig is, aan te bieden. Het gaat vooral om procedures waar just-in-time informatie kan worden gegeven in de vorm van bijvoorbeeld invulinstructies.

De routinematige vaardigheden waarbij just-in-time informatie wordt geboden (component 3) leert de student altijd tegelijkertijd met niet-routinematige aspecten om de hele-taak benadering te waarborgen. Voor bepaalde routinematige vaardigheden is echter een hoge mate van automatisering vereist. Het gaat hierbij om vaardigheden waarbij zeer snel en adequaat handelen cruciaal is. Om de benodigde hoge mate van automatisering te bereiken wordt dit type vaardigheden niet alleen binnen de leertaken maar ook apart getraind. Deze vierde component uit het model wordt deeltaakoefening genoemd. In figuur 1 zijn dit de kleine bolletjes. Voor opleidingen zoals door het LSOP verzorgd, is een vaardigheid die prima in deeltaakoefening kan worden getraind 'het trekken van een pistool'. Het is niet uitgesloten dat een politiemedewerker ooit zijn pistool moet gebruiken in een bepaalde situatie. De just-in-time informatie zou hierbij de procedure voor het grendelen van een pistool kunnen zijn. Omdat de politiemedewerker via deeltaakoefening uittentreuren heeft leren schieten vanaf verschillende afstanden op verschillende doelen (band van een auto, been, arm) zal hem dit, technisch gesproken, gemakkelijk afaan. Overigens wordt de deeltaakoefening nooit aangeboden aan het begin van een opleiding. Een belangrijke voorwaarde is dat de student de hele taak al een keer heeft bekeken met de zoomlens, zodat duidelijk is wat de relatie is tussen de deeltaakoefening en de leertaken.

Uit figuur 1 valt tevens af te lezen dat de bolletjes aanvankelijk grijs gekleurd zijn en dat deze grijze invulling vermindert na elke leertaak. De mate van inkleuring zegt iets over de hoeveelheid ondersteuning die aan de student geboden wordt. In de eerste leertaak van een taakklasse zal aan studenten vaak een uitgewerkt voorbeeld worden gepresenteerd. In ons voorbeeld over toezicht houden in een publiek domein kan dit bijvoorbeeld door studenten mee te laten lopen met een ervaren agent. In de tweede leertaak mogen studenten zelf in de praktijk bepaalde handelingen uitvoeren. Na verloop van tijd hebben de studenten geen ondersteuning meer nodig en kan er bij latere leertaken dus volledig zelfstandig gehandeld worden. Op deze manier blijft gewaarborgd dat de hele complexe vaardigheid 'toezicht houden in publiek domein' centraal staat in elke leertaak.

Binnen de opleiding tot all-round politiemedewerker zou de eerste taakklasse bijvoorbeeld uit drie leertaken kunnen bestaan.

De eerste leertaak in deze taakklasse is nog vrij eenvoudig en biedt de student veel ondersteuning. In het 4C/ID model wordt een dergelijke leertaak een *uitgewerkt voorbeeld* genoemd (Van Merriënboer, 1997). Als uitgewerkt voorbeeld zou een videofilm gebruikt kunnen worden die een ervaren politiemedewerker laat zien die gedurende een uur toezicht houdt op een rustig plein in een dorp. In de video komen de volgende situaties aan bod:

- de politiemedewerker spreekt een persoon aan en geeft hem een advies ter voorkoming van kleine criminaliteit;

- de politiemedewerker houdt een repressief (waarschuwend) gesprek naar aanleiding van een gepleegde overtreding of gedraging;
- de politiemedewerker wijst de weg aan een persoon die hierom vraagt;
- de politiemedewerker voert een gesprek met een slachtoffer (eerste opvang) van een eenvoudig misdrijf.

De student bekijkt de video en schrijft op basis van zijn observaties een verslag. Hij bespreekt dit verslag nadien met medestudenten en/of de begeleider waardoor een eerste beeld van de hele vaardigheid wordt gevormd.

Bij de tweede leertaak neemt de ondersteuning af. De student zal een situatie zoals die in de video is getoond deels zelfstandig moeten kunnen aanpakken. Een dergelijke taak wordt een *aanvulprobleem* genoemd: de probleemsituatie, het gewenste doel, evenals een deel van de oplossing is gegeven. Het is aan de student om verder te gaan op de aangegeven weg. Zo zou de student in de rol van assistent politiemedewerker kunnen meelopen met een ervaren politiemedewerker in een rustig publiek domein (zie tabel 1 voor de complexiteitsfactoren van taakklasse 1). Op bepaalde momenten zal de ervaren medewerker de student vragen om in actie te komen. Bijvoorbeeld om een fietser die een fietsverbod negeert aan te houden en te corrigeren.

In de derde leertaak van de taakklasse zal de student geen ondersteuning meer worden geboden en zal hij of zij zelfstandig de taak uitvoeren. De laatste leertaken in een taakklasse worden *conventionele problemen* genoemd: deze leertaken confronteren studenten met uitsluitend een probleemsituatie en een gewenste doelsituatie. In het geval van de opleiding tot all-round politiemedewerker zou dit bijvoorbeeld kunnen inhouden dat een student een uur toezicht houdt op een rustig plein. Gewenst gedrag, zoals hierboven aangegeven bij de bespreking van leertaak 1, wordt uitgelokt door doelgericht bepaalde probleemsituaties te creëren.

De 3 voorbeelden van leertaken laten zien dat de hele taak continu in het vizier blijft. Steeds is de student er zich van bewust dat hij leert om toezicht te houden in een publiek domein. Omdat de ondersteuning geleidelijk afneemt, zal de student in toenemende mate taken in een bepaalde taakklasse zelfstandig kunnen uitvoeren. De laatste leertaak van elke taakklasse kan als *assessmenttaak* worden beschouwd, aangezien de student daarbij volledig zelfstandig handelt.

VASTSTELLEN VAN COMPETENTIE

Competenties kunnen niet vastgesteld worden door een beoordeling van de componenten waaruit ze zijn opgebouwd (zoals kennis en (deel)vaardigheid). Ze moeten beoordeeld worden op basis van de geïntegreerde toepassing van die componenten bij een authentieke taak in een min of meer realistische context. In de Angelsaksische literatuur over onderwijskundig meten wordt in dat verband gesproken van *alternative assessment* of *authentic assessment*. Met die termen wordt duidelijk afstand genomen van die andere assessmentvorm die van kandidaten vraagt om dat wat ze weten en

kunnen te demonstreren in een antwoord op een gesloten vraag of kort-antwoordvraag. In Nederland wordt het onderscheid tussen assessment en alternative assessment eigenlijk niet gemaakt. Met assessment worden al die toetsvormen aangeduid waarin kandidaten nieuwe responsen moeten produceren met het oog op het uitvoeren van min of meer authentieke taken. Onder assessment verstaan we dus alle toetsvormen behalve die kunnen worden aangeduid als objectieve studietoetsen. Dochy en Struyven (2002) illustreren deze brede opvatting over assessment met de bespreking van een 'waaier' van assessmentvormen. Een aspect waarop de assessmentvormen zich nadrukkelijk van elkaar onderscheiden is de mate waarin de uitgevoerde taak overeenkomt met een taak in een realistische vervolgsituatie waarvoor de opleiding een voorbereiding is. Bij *performance assessments* is die overeenkomst groot. In Nederland worden dergelijke natuurgetrouwe assessmentvormen steeds vaker aangeduid met de term *proeve van bekwaamheid*. De meest tot verbeelding sprekende proeve van bekwaamheid is wel het rijexamen. Sinds jaar en dag moeten Nederlanders om het rijbewijs te verkrijgen een proeve van bekwaamheid afleggen. Dit praktisch deel van het rijexamen wordt door veel onderwijsvernieuwers gepropageerd als voorbeeld van een ideale assessmentvorm voor het beoordelen en vaststellen van competenties. Ze dragen daarvoor het volgende motief aan. De proeve van bekwaamheid is een assessmentvorm met een zeer hoog authenticiteitsgehalte; kandidaten demonstreren hun competentie door taken uit te voeren onder omstandigheden die zich in weinig tot niets onderscheiden van die in de reële werksituatie. Het rijexamen onderscheidt zich slechts van de reële werksituatie in het feit dat er een persoon in de auto aanwezig is die de route bepaalt en die, via de dubbele bediening, kan ingrijpen als er gevaar dreigt. Daarnaast wordt grote waarde gehecht aan het feit dat het rijexamen door praktisch iedereen als een nuttige afsluiting van de rijopleiding wordt beschouwd. In het Angelsaksisch jargon heet het dat de proeve van bekwaamheid een hoge *face validity* heeft (op het eerste gezicht valide lijkt te zijn).

KWALITEITSKENMERKEN VAN EEN ASSESSMENT

Het praktisch deel van het rijexamen mag dan heel valide lijken, nadelen zijn er ook en die worden in brede lagen van de bevolking herkend. Zoals het eenmalige karakter en de beperkte omvang (40 minuten). Die maken dat het rijexamen een wankel basis is voor het trekken van solide conclusies over een veelomvattende competentie als rijvaardigheid. De gevolgen zijn ernstig: mensen die in feite competent zijn maar door pech tijdens het rijexamen worden afgewezen en mensen die de competentie niet beheersen maar door geluk tijdens het rijexamen niet 'door de mand vallen' en hun rijbewijs krijgen. Wat voor rijvaardigheid geldt, gaat ook op voor andere competenties. De proeve van bekwaamheid is aan de ene kant erg aantrekkelijk, maar aan de andere kant kleven er gevoelige problemen aan. Kane (1992) betoogt dat dit inherent is aan het beoordelen van competenties:

'Valid assessment of professional competence has proven to be an elusive goal. Objective test items are criticized as being unrealistic and therefore invalid. Direct observation tends to

be unreliable and therefore invalid. Simulations share both of these flaws to some extent. Basically, you can't win.'

Net als objectieve studietoetsen moeten de scores op assessmentopdrachten voldoen aan de gebruikelijke kwaliteitskenmerken voor betrouwbaarheid en validiteit. Echter, de (statistische) uitwerking die hieraan gegeven is voor objectieve studietoetsen, is voor assessmentopdrachten vaak niet bruikbaar. Dierick, c.s. (2002) signaleren in de literatuur een verschuiving van psychometrische naar edumetrische kenmerken voor het vaststellen van de kwaliteit van assessmentscores. Dit houdt onder andere in dat er aandacht komt voor kwaliteitskenmerken als: de nauwkeurigheid van de scores, de cognitieve complexiteit van de assessmentopdracht, de authenticiteit van de taakuitvoering, de transparantie van de assessmentprocedure, de eerlijke kans die studenten via de assessmentopdracht gekregen hebben om te laten zien dat ze de competentie verworven hebben, enzovoort. Crooks, c.s. (1996) zien deze kwaliteitskenmerken als schakels in een kwaliteitsketen die de verbinding vormt tussen de prestatie op een concrete assessmentopdracht en de conclusie over de bekwaamheid in een doeldomein. Als een schakel ontbreekt of onvoldoende solide is, komt het trekken van conclusies over de achterliggende bekwaamheid in gevaar. Deze opvatting over de validering van assessments wordt hieronder verder uitgewerkt voor een beperkt aantal kwaliteitskenmerken, die hun wortels voor een belangrijk deel in de psychometrische traditie hebben.

Om een solide conclusie te trekken over het al dan niet verworven zijn van een competentie dienen de assessmentscores in de eerste plaats *accuraat* te zijn. Een score is accuraat als die de ware score van de kandidaat dicht benadert. De ware score is een theoretisch begrip dat gedefinieerd wordt als de gemiddelde score van een oneindig aantal metingen door middel van equivalente toetsen of assessments, aangenomen dat daarbij geen oefeneffecten, of andere veranderingen in de persoon ontstaan (De Groot & Van Naerssen, 1977). Scores zijn nooit volledig accuraat. Dat wil zeggen dat de score nooit precies kan weergeven hoe het ervoor staat met de kennis, vaardigheid of competentie van de kandidaat. Elke assessmentscore is het product van het ware prestatieniveau plus een zekere meetfout. Globaal zijn er twee oorzaken voor het ontstaan van meetfouten: één is gerelateerd aan de kandidaat (interne fouten), de ander aan zaken die buiten de kandidaat liggen (externe fouten). Voorbeelden van bronnen van interne fouten zijn: lichamelijk en geestelijk welbevinden tijdens het assessment, de motivatie om te presteren en de ervaring met assessments. Deze aan de individuele persoon gebonden factoren kunnen een negatieve of positieve invloed hebben op de uiteindelijke score en daarmee het zicht op het ware prestatieniveau belemmeren. Voorbeelden van bronnen van externe fouten zijn: de kwaliteit van de aanwijzingen voor de kandidaat, de condities waaronder het gedrag vertoond moet worden, de taken die moeten worden uitgevoerd, de assessoren en het scoringsstelsel. Al met al zijn er zeer vele bronnen die alleen of gezamenlijk de prestatie van de kandidaat in positieve of negatieve zin kunnen beïnvloeden. De invloed van sommige bronnen was in de afgelopen decennia veelvuldig onderwerp van onderzoek. Dit geldt onder andere voor de invloed van taken en assessoren. Over de invloed van taken komen we hieronder bij het kwaliteitskenmerk 'generaliseerbaarheid' nog te spreken. Wat de invloed van assessoren

betreft: onderzoekers uit verschillende disciplines zijn tot de eensluidende conclusie gekomen dat de beoordelingsovereenstemming tussen assessoren met betrekking tot de kwaliteit van processen en/of de daaruit voortkomende producten te wensen overlaat (o.a. Sanders, 1980; Hofstee, 1983; Straetmans, 1985; Lunz, Wright & Linacre, 1990). De les die uit deze onderzoeksresultaten geleerd moet worden, is dat het onverstandig is om bij assessments met slechts één assessor te werken. In opleidingssituaties zal het echter om organisatorische en financiële redenen vaak niet haalbaar zijn om met meerdere, onafhankelijke assessoren te werken. Andere maatregelen die overwogen kunnen worden om te komen tot accurate assessmentsscores, zijn:

- Training van assessoren. Bij professionele assessmentcenters blijkt dit vaak een krachtig middel om de interpretatieverschillen tussen assessoren te reduceren (De Graaff, 1993).
- Gedetailleerde beoordelingsprotocollen in plaats van globale beoordelingsaspecten. Er is onderzoek bekend waaruit gebleken is dat assessoren meer accurate scores toekennen wanneer de beoordelingsaspecten de assessor minder vrijheid laten over wat er beoordeeld moet worden en over de wijze waarop dat moet gebeuren (Straetmans, 1985). De keerzijde hiervan is dat dergelijke beoordelingsprotocollen beperkt zijn in hun toepasbaarheid en bovendien meer tijd kosten. Een ander gevaar van gedetailleerde beoordelingsprotocollen is dat, in het streven naar een zo nauwkeurig mogelijke beoordeling, de nadruk te veel komt te liggen op de componenten van de competentie en daarmee voorbij schiet aan het doel.

Het tweede kwaliteitskenmerk betreft de *generaliseerbaarheid* van de scores. Bij het beoordelen van competenties gaat het nooit om die ene of enkele toevallig waargenomen prestatie(s) maar om het vermogen om soortgelijke prestaties onder gelijkwaardige omstandigheden te kunnen uitvoeren. 'Soortgelijk' is hier een misleidende term want de taken (en dus de prestaties) die onder een en dezelfde competentie vallen, kunnen aanzienlijk van elkaar verschillen. Zo doet het rijden op een snelweg een beroep op heel andere vaardigheden dan het rijden in de binnenstad of het uitvoeren van bijzondere verrichtingen met het voertuig, terwijl het alle drie taken zijn die gerangschikt worden onder de competentie 'rijvaardigheid'. Over het algemeen wordt uit generaliseerbaarheidstudies geconcludeerd dat de generaliseerbaarheid naar prestaties op soortgelijke taken beperkt is (Linn, Baker & Dunbar, 1991). De belangrijkste oorzaak daarvan is gelegen in het feit dat de assessmenttaken geen goede afspiegeling zijn van de taken die in principe aangeboden hadden kunnen worden. En dat vindt zijn oorzaak weer in tijd- en/of geldgebrek. Laten we, ter illustratie, het rijexamen nog eens opvoeren. Het is vermoedelijk een conservatieve schatting als we beweren dat de competentie rijvaardigheid bestaat uit het kunnen uitvoeren van zo'n 40 verschillende handelingen in zo'n 25 verschillende contexten (een context is in dit verband een combinatie van het verkeersbeeld, de kwaliteit van het wegdek en de weersgesteldheid). Dit brengt het aantal verschillende taken (combinatie van handeling en context) op 1000 stuks. Hoeveel verschillende taken zouden er in een examen van 40 minuten aan bod kunnen komen? Aangezien het rijexamen bijna altijd in een drukke stad wordt afgenomen en een deel van de tijd dus wachtend wordt doorgebracht, zal dit aantal niet veel meer

dan 30 bedragen. Het is niet erg waarschijnlijk dat het assessment dan een representatieve steekproef kan vormen uit het assessmentdomein van 1000 taken, te meer daar de selectie maar ten dele door de examinerator gepland kan worden. Immers, het leeuwendeel van de taken komt automatisch voort uit de verkeerssituatie van het moment.

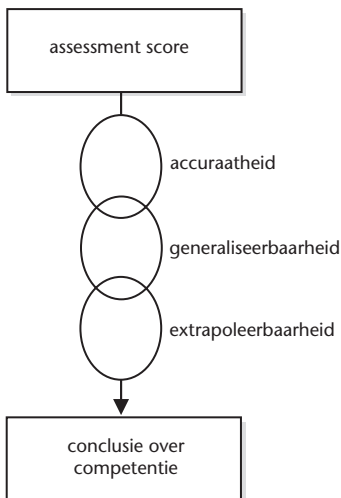
Het derde kwaliteitskenmerk houdt in dat de scores *extrapoleerbaar moeten* zijn. Dat wil zeggen dat de behaalde score het prestatieniveau moet weerspiegelen dat de betreffende kandidaat zou halen wanneer die dezelfde taak in een reële werksituatie zou uitvoeren. Soms is de extrapoleerbaarheid geen enkel probleem omdat de assessmenttaak niet afwijkt van de taak uitgevoerd in de reële werksituatie. Maar meestal wordt op de een of andere manier de werkelijkheid gereduceerd. Bijvoorbeeld omdat men de taakuitvoering te kostbaar (lancering Patriot) of te gevaarlijk (onschadelijk maken van een anti-tankmijn) vindt of omdat de geschikte problematiek niet op afroep beschikbaar is (aanhouding van een vuurgevaarlijke crimineel in een winkelcentrum). Naarmate de werkelijkheid meer gereduceerd is, wordt het moeilijker om staande te houden dat de score een correcte weergave is van het prestatieniveau in de reële werksituatie.

Bovengenoemde kwaliteitskenmerken stellen niet alleen zware eisen aan een assessment, ze zijn ook problematisch in die zin dat optimalisatie van het ene kenmerk dikwijls een negatieve invloed uitoefent op de andere kenmerken. Het streven naar een rijexamen met een hoge extrapoleerbaarheid van de scores bijvoorbeeld, heeft ertoe geleid dat er slechts een beperkt aantal taken tijdens het examen aan bod kan komen (dus een lagere generaliseerbaarheid) en dat de beoordeling daarvan in hoge mate geschiedt op basis van een (subjectieve) interpretatie van de noodzakelijkerwijs globaal omschreven beoordelingsaspecten (een lagere accuraatheid). Gaan we naar een verregaande reductie van de werkelijkheid, bijvoorbeeld een computersimulatie van de Koninklijke Landmacht (Krol, 1992) waarin militaire artsen hun Triage-vaardigheid² moeten demonstreren, dan zien we precies het tegenovergestelde beeld. In de accuraatheid van de scores mogen we wat meer vertrouwen hebben omdat de beoordeling volledig gestandaardiseerd is en wordt uitgevoerd door een computerprogramma. Ook de generaliseerbaarheid is groot omdat in korte tijd veel verschillende calamiteiten kunnen worden aangeboden. Maar de vraag die opkomt is, of de kandidaat die een hoge Triage-score behaalt ook op het slagveld zijn werk goed doet, dus of de extrapoleerbaarheid ook voldoende is. De stress, hectiek en ellende van de reële werksituatie vind je niet of nauwelijks terug in de overigens realistische beelden van de computersimulatie. In de computersimulatie schuif je een ten dode opgeschreven gewonde wellicht gemakkelijk terzijde. In de werkelijkheid zal dit een stuk moeilijker zijn en al helemaal wanneer de zwaargewonde je in vertrouwen neemt. Wint het gevoel van medelijden het dan van de professionele beroepshouding die in de computersimulatie zo gemakkelijk te handhaven was?

Een tussenvorm zou hier een assessment kunnen zijn waarin simulatiepatiënten getraind zijn om zich te gedragen als oorlogsslachtoffers met specifieke verwondingen. In een dergelijke simulatie wordt ten opzichte van de computersimulatie gewonnen op het aspect extrapoleerbaarheid (de stress zal ongetwijfeld minder zijn dan op het slag-

veld, maar meer dan achter het beeldscherm). Ten aanzien van de accuraatheid wordt ingeleverd ten opzichte van de computersimulatie maar gewonnen in vergelijking met de reële werksituatie. Immers, de draaiboeken geven precies aan welk gedrag verwacht kan worden, wat van grote waarde is voor de beoordeling van dat gedrag. Hetzelfde geldt voor de generaliseerbaarheid. Ten opzichte van de reële werksituatie is die waarschijnlijk verbeterd omdat men de gewonden kan creëren en deze daarmee een representatieve steekproef kan laten vormen uit alle mogelijke verwondingen. Maar omdat een assessment met simulatiepatiënten erg arbeidsintensief en dus duur is, is het niet erg waarschijnlijk dat kandidaten in meer dan één calamiteit hun competentie moeten demonstreren. Een nadeel dus ten opzichte van de computersimulatie.

De conclusie moet zijn dat er geen beste assessmentmethode is, althans wanneer er vanuit gegaan mag worden dat er beperkingen zijn opgelegd aan het gebruik van tijd en middelen. Elke assessmentmethode heeft een zwakke schakel in de kwaliteitsketen die de assessmentscore verbindt met de conclusie over competentie (zie figuur 2 voor een voorbeeld van een kwaliteitsketen). Omdat de verschillende assessmentmethodes de zwakke schakel niet allemaal op dezelfde plek hebben, verdient het niet alleen aanbeveling om veel bewijs te verzamelen maar ook om dit te doen met verschillende methodes (Straetmans & Sanders, 2001).



Figuur 2 Voorbeeld van een kwaliteitsketen met een beperkt aantal schakels

Het Protocol Portfolio Scoring

Het 4C/ID-model komt tegemoet aan de hierboven gesignaleerde problemen met het vaststellen van competenties. Omdat in onderwijs dat op dit model gebaseerd is van meet af aan gewerkt wordt aan realistische taken (leertaken geheten), kan al tijdens de opleiding begonnen worden met het verzamelen van informatie die tegen het einde van de opleiding gebruikt wordt voor een afgewogen conclusie over het al dan niet verworven zijn van de competentie. Wanneer binnen een bepaalde taakklasse de onder-

steuning bij het uitvoeren van een leertaak wordt weggehaald en de student het dus 'op eigen kracht' moet doen, kan zo'n leertaak worden opgevat als een assessmentopdracht. Tijdens de opleiding wordt dus eigenlijk een portfolio aangelegd waarin steeds meer bewijsstukken worden opgeslagen voor de competentie in kwestie.

In veel beroepsopleidingen wordt inmiddels op de een of andere manier gebruikgemaakt van het portfolio. Volgens een recent onderzoek van de onderwijsinspectie worden portfolio's vooral gebruikt als instrument voor formatieve evaluatie (voor bijsturing van het onderwijs) en als middel om de student aan te sporen te reflecteren op zijn eigen leren. De inzet van het portfolio voor het beoordelen van competenties (*summatieve evaluatie*) werd in dit onderzoek niet aangetroffen (Inspectie van het Onderwijs, 2003)³. Tevens concludeert de onderwijsinspectie dat er nauwelijks of geen procedures zijn aangetroffen die de validiteit van beoordelingen op basis van een portfolio kunnen waarborgen. Om in die behoefte te voorzien is het Protocol Portfolio Scoring ontwikkeld.

Het Protocol Portfolio Scoring, kortweg PPS (Straetmans, in voorbereiding), is een reeks richtlijnen die gebruikt kan worden voor het aanleggen en beoordelen van een verzameling bewijsstukken en voor het nemen van beslissingen over het al dan niet verworven zijn van een competentie. Hoe PPS werkt, wordt hieronder besproken aan de hand van 5 kenmerken en verder toegelicht aan de hand van de fictieve output van een nog te ontwikkelen computerprogramma dat de PPS-methode ondersteunt (zie tabel 2). De lezer dient zich te realiseren dat bij de uitwerking van de PPS-methode, met name in de hieronder beschreven kenmerken 3, 4 en 5, bepaalde keuzes zijn gemaakt die niet noodzakelijk de meest haalbare of beste oplossing vormen. Nader onderzoek is nodig om na te gaan wat onder bepaalde omstandigheden de beste manier is om scores te aggregeren en prestatiestandaarden te bepalen.

Een beoordelingsportfolio dat is opgezet volgens PPS kan in principe deel uitmaken van een multifunctioneel, elektronisch portfolio zoals dat in sommige beroepsopleidingen al operationeel is. Met dat verschil dat het niet door studenten beheerd wordt maar door docenten; studenten hebben in het beoordelingsportfolio alleen leesrechten. Tabel 2 geeft een overzicht van de scores die G.R. Kromhoudt, fictief student aan de politie-school, behaald zou kunnen hebben op assessmentopdrachten in het kader van het onderwijsonderdeel 'toezicht houden in een publiek domein'. Het beoordelingsportfolio bevat 4 bewijsstukken. Bewijsstuk 1 en 2 zijn het resultaat van assessmentopdrachten in taakklasse 1. Bewijsstuk 3 kwam tot stand in taakklasse 2 en bewijsstuk 4 in taakklasse 3.

Tabel 2 Overzicht van assessmentscores

nr	assessment-methode	datum	asses-sor	beoordelingsaspecten (gemarkeerde kolom is kritisch) tussen haakjes: de maximaal haalbare score per aspect										correctiefactoren			score	horizontale standaard	beslis-sing	
				1 (5)	2 (5)	3 (6)	4 (5)	5 (4)	6 (5)	7 (3)	8 (4)	9 (5)	10 (5)	ruwe score	kwali-teit	moelijk-heid				
1	situational judgment test	050202	HK			4		3							7	.7	.7	$7 * .7 * .7 = 3.4$	8.4	onvol.
2	work sample test	080302	HK	3	2	5		4	5				5	24	.8	.4	$24 * .8 * .4 = 7.7$	23.6	onvol.	
3	work sample test	120402	GS	5	5	6		3	5		3		5	32	.9	.9	$32 * .9 * .9 = 25.9$	26.6	onvol.	
4	performance on the job	180502	HK	4		6	5	4	4	3		4	5	35	.85	1	$35 * .85 * 1 = 29.8$	29.6	vol.	
		gemiddelde score kritische aspecten				5.3		3.5	4.7			4.5								
		verticale standaard				4.8		3.5	4.2			4								
		beslissing						vol.	vol.	vol.		vol.								

Kenmerk 1 Methodemix

Tabel 2 laat zien dat de bewijsstukken in het beoordelingsportfolio van student Kromhoudt het resultaat zijn van de inzet van drie verschillende assessmentmethodes die met name van elkaar verschillen op het gebied van de natuurgetrouwheid. De *situational judgment test* (SJT) is een *hands off assessment* (Straetmans, 1998). Hierbij krijgt de kandidaat een realistische beschrijving voorgelegd van een situatie waarin hij/zij bij de latere beroepsuitoefening gemakkelijk terecht zou kunnen komen. De kandidaat moet daarna uit een reeks alternatieven de beste handelwijze kiezen. Zowel de situatiebeschrijving als de verschillende handelwijzen kunnen tamelijk realistisch gepresenteerd worden, bijvoorbeeld in de vorm van videofragmenten. Kijk op de volgende internetpagina voor een aardig voorbeeld van een SJT: <http://ergometrics.org/correct1.htm> Desondanks zullen maar weinig mensen een beslissing over competentie willen nemen die exclusief gebaseerd is op de uitslag van een situational judgment test. De test meet vooral beroepsgerelateerde kennis; een noodzakelijke maar onvoldoende voorwaarde voor competentie. Echter, gebruikt in combinatie met andere assessmentmethodes kan de SJT een nuttige bijdrage leveren aan de bewijskracht van het portfolio vanwege het feit dat in korte tijd veel verschillende probleemsituaties kunnen worden voorgelegd. Het zegt met andere woorden iets over het vermogen van de kandidaat om de beroepsgerelateerde kennis in steeds andere contexten te gebruiken (vermogen tot transfer).

Bij een *work sample test* krijgt een kandidaat onder nagebootste omstandigheden een realistische taak uit te voeren. De taak wordt uitgevoerd in de reële werkomgeving met gebruikmaking van alle hulpmiddelen die de beroepsbeoefenaar normaal ook ter

beschikking staan. Het verschil met de normale uitoefening van het beroep is dat bepaalde gedragingen, die horen bij het repertoire van de competente beroepsbeoefenaar, explicieter worden uitgelokt door het aanbieden van bepaalde stimuli. Een direct gevolg hiervan is dat de kandidaat vaak alerter is dan tijdens een assessment waarin op minder nadrukkelijke wijze bepaald gedrag wordt uitgelokt. In het geval van de competentie 'toezicht houden in een publiek domein' zou een work sample test bijvoorbeeld het volgende kunnen inhouden:

Een aspirant-politiefunctionaris heeft te horen gekregen dat de stationschef van het plaatselijke station om frequenter politietoezicht heeft gevraagd in de onmiddellijke omgeving van het station omdat er geregeld berovingen zouden plaatsvinden. De aspirant-politiefunctionaris krijgt tegen de avondschemering opdracht om in de buurt van het station toezicht te houden. Aangekomen op het stationsplein ziet hij dat een zwerver kennelijk een rustplaats zoekt voor de nacht. De zwerver is een acteur die de rol moet spelen van agressieve, bijtgrage man die bovendien de beschermeling lijkt te zijn van een groepje opgeschoten jongelui (eveneens acteurs die een rol spelen) dat het stationsplein als skatebaan gebruikt.

De voorspellende waarde van de work sample test voor het functioneren in een beroep hangt onder andere af van de manier waarop het gedrag wordt uitgelokt. Een kenmerk van de competente persoon is dat die geheel zelfstandig de signalen oppikt dat er een probleemsituatie is ontstaan en dat er op de een of andere manier ingegrepen moet worden. Wanneer in een work sample test het verlangde gedrag te nadrukkelijk wordt uitgelokt, kan er niet worden vastgesteld of de kandidaat in staat is om de bewuste signalen te herkennen. Dit doet afbreuk aan de predictieve waarde van het instrument.

Bij *performance on the job* wordt er geen gedrag uitgelokt door gestuurde interventies. De aspirant politiefunctionaris doet zijn werk in een volledig authentieke setting. Het enige onderscheid met de reële werksituatie is de wetenschap van de kandidaat dat hij geobserveerd en beoordeeld wordt. Uiteraard levert dit instrument de beste voorspellingen over het toekomstig werkgedrag. De keerzijde vormt de inefficiëntie van de methode. Om bij ons voorbeeld te blijven: als zich geen ongeregelde heden voordoen in het gebied waar de aspirant-politiefunctionaris surveilleert, blijft het gedrag waarin men het meest geïnteresseerd is, achterwege.

Kenmerk 2 Standaardset van beoordelingsaspecten

Om op grond van een beoordelingsportfolio een conclusie te trekken over het al dan niet verworven zijn van de competentie moeten alle bewijsstukken in het beoordelingsportfolio dezelfde beoordelingsgrondslag hebben. Bedoeld wordt dat het gedrag van de kandidaat, ongeacht de gebruikte assessmentmethode, steeds beoordeeld is aan de hand van dezelfde set van beoordelingsaspecten. Alleen dan is het zinvol om de scores van verschillende bewijsstukken bij elkaar te tellen. Deze beoordelingsaspecten vormen tezamen een operationalisatie van de betreffende competentie. De beoordelingsaspecten betreffen die aspecten van het gedrag (of het resultaat daarvan) die het verschil maken tussen een competente en incompetent beroepsbeoefenaar. Het afleiden en formuleren van deze aspecten is arbeidsintensief. Het is de kunst om deze aspecten

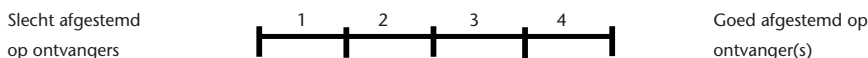
enerzijds globaal te formuleren, zodat ze in zo veel mogelijk situaties gebruikt kunnen worden, en anderzijds toch zo eenduidig te omschrijven dat assessoren de bedoeling begrijpen zonder verregaande interpretaties van hun kant. Dit laatste proberen we te bereiken door bij elk beoordelingsaspect indicatoren te geven. De indicator is een aanwijzing voor de aanwezigheid van het gewenste gedrag zoals gespecificeerd in het bovenliggende beoordelingsaspect. De assessor bepaalt welke indicatoren in een bepaalde taaksituatie 'aan de orde' waren en beoordeelt de kandidaat daarop. Via een scoringsvoorschrift worden de scores op de indicatoren omgezet in een score op het bovenliggende beoordelingsaspect. Figuur 3 is een voorbeeld van een globaal maar, dankzij de aanwezigheid van een reeks indicatoren, helder omschreven beoordelingsaspect dat gescoord wordt op een 6-puntsschaal.

Verstaat de kunst om door middel van een gesprek een conflict te beëindigen of een dreigend conflict af te wenden.

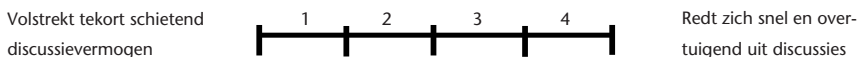


Het bovenstaande globale beoordelingsaspect wordt indirect, via onderstaande indicatoren, gescoord. (Vaak zullen er meer dan vier indicatoren geformuleerd zijn.) Assessoren bepalen welke aanwijzingen (indicatoren) zij in een bepaalde taaksituatie hebben gezien en vervolgens willen gebruiken om de prestatie te scoren op het betreffende beoordelingsaspect.

Niveau van het taalgebruik is afgestemd op de ontvanger



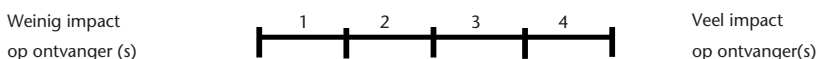
Reageert snel en passend op tegenwerpingen



De non-verbale communicatie is effectief en correct



Niveau van het taalgebruik is afgestemd op de ontvanger



Scoringsregel

De resultaten op de subschalen worden gesommeerd, waarna ze volgens onderstaande procedure worden omgezet in een score op de bovenliggende schaal van het beoordelingsaspect. (Deze scoringsregel is alleen geldig voor een situatie waarin vier indicatoren gebruikt zijn).

Aantal punten:	schaalscore
5	1
5 t/m 6	2
7 t/m 9	3
10 t/m 12	4
13 t/m 15	5
16	6

Figuur 3 Een beoordelingsaspect voor de competentie 'toezicht houden in een publiek domein'

Andere, hier niet verder uitgewerkte, beoordelingsaspecten voor de competentie 'toezicht houden in een publiek domein' zouden bijvoorbeeld kunnen zijn:

- neemt de juiste beslissingen in complexe probleemsituaties;
- heeft oog voor situaties die potentieel conflicterend zijn;
- straalt gezag uit door houding en voorkomen;
- baseert zijn handelen zo veel mogelijk op objectieve informatie.

Welke beoordelingsaspecten in een bepaalde beoordelingssituatie gebruikt kunnen worden, hangt af van de gebruikte assessmentmethode en van het verloop van de uitgevoerde opdracht. De PPS-methode gaat ervan uit dat een assessor tijdens de uitvoering van de opdracht zorgvuldig waarneemt en daarvan aantekeningen maakt en pas achteraf bepaalt welke indicatoren (en daarmee welke beoordelingsaspecten) gescoord kunnen worden op basis van wat is waargenomen. In het fictieve beoordelingsportfolio van student Kromhoudt (zie tabel 2) zien we bijvoorbeeld dat het eerste bewijsstuk op slechts twee beoordelingsaspecten gescoord is. Dat is niet verwonderlijk aangezien het bewijs gegenereerd is met een assessmentmethode die vaststelt of kandidaten *weten* hoe ze in bepaalde probleemsituaties zouden moeten handelen maar waarmee niet het handelen zelf kan worden beoordeeld. De gebruikte beoordelingsaspecten hebben betrekking op de kwaliteit van de gekozen oplossing, niet op het gedrag dat daartoe geleid heeft. De work sample test, de gebruikte assessmentmethode bij bewijsstuk 2 en 3, vraagt van de kandidaat om handelend op te treden. Hierbij gaat het niet meer alleen om het resultaat van het gedrag maar ook om het gedrag zelf. Daardoor konden meer beoordelingsaspecten gescoord worden. Het vierde bewijsstuk is gebaseerd op een observatie van gedrag in de reële werksituatie. Daarbij konden 8 van de 10 beoordelingsaspecten gescoord worden. Het is overigens niet zo dat het aantal gebruikte beoordelingsaspecten altijd het grootst is bij assessmentmethodes die de hoogste natuurgetroouwheid hebben. Aangezien er bij *work sample tests* mogelijkheden zijn om gedrag uit te lokken, is het in principe mogelijk om het gedrag zodanig te sturen dat zo veel mogelijk beoordelingsaspecten gebruikt kunnen worden.

Bepaalde beoordelingsaspecten zijn cruciaal voor de prestatie. Dat wil zeggen dat bij onvoldoende uitvoering van dat aspect van de prestatie de kans groot is dat het eindresultaat ook onder de maat is. Deze beoordelingsaspecten worden aangeduid als *kritische* beoordelingsaspecten.

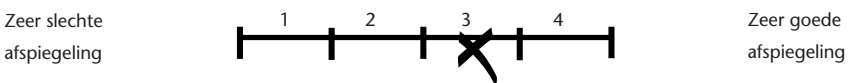
Kenmerk 3 Absolute standaarden

Wanneer is er voldoende bewijs verzameld om de conclusie te trekken dat de betreffende competentie verworven is? Het bepalen van een standaard is een van de moeilijkste zaken bij het ontwikkelen van tests en assessments. Experts zijn redelijk goed in het rangordnen van prestaties (wie presteerden het best, wie het slechtst?) maar bij het beoordelen van competenties gaat het in de eerste plaats om de vraag of de prestaties van een persoon voldoen aan een bepaalde kwaliteitsstandaard. Doorgaans kunnen experts wel een beschrijving geven van de gewenste kwaliteit maar daarmee is nog niet duidelijk door welke assessment score deze kwaliteit weerspiegeld wordt. In het computerprogramma van PPS zou een speciale module opgenomen moeten worden waarin assessoren de standaard voor een competentie kunnen bepalen. Dit gebeurt als volgt: een aantal assessoren krijgt opdracht om, onafhankelijk van elkaar, voor elk beoordelingsaspect de score in te voeren die minimaal behaald moet worden door een persoon die de competentie verworven heeft. Het is de bedoeling dat zo veel mogelijk assessoren hun mening geven. De scores van de assessoren op elk afzonderlijk beoordelingsaspect worden daarna gemiddeld. Deze gemiddelde score wordt de grensscore genoemd. Met de grensscores worden de standaarden bepaald die gebruikt worden om te beslissen of de betreffende competentie verworven is (zie kenmerk 5 Beslissingen nemen over competentie).

Kenmerk 4 Correctiefactoren

Niet alle bewijsstukken in het beoordelingsportfolio hebben dezelfde bewijskracht. Zoals hiervoor duidelijk is gemaakt, moeten scores uit een assessment aan allerlei educatieve kwaliteitskenmerken voldoen. Over het algemeen schieten de scores op één of meer van deze kwaliteitskenmerken tekort. Naarmate de tekortkomingen ten aanzien van de kwaliteitskenmerken groter zijn, zou het betreffende bewijsstuk een kleinere bijdrage moeten leveren in de uiteindelijke beslissing over competentie. Hoeveel kleiner wordt bepaald door een correctiefactor die op zijn beurt het resultaat is van de inschatting van een assessor over de mate waarin de scores voldoen aan de kwaliteitskenmerken. Deze inschatting moet voor elk assessment opnieuw gemaakt worden door de beantwoording van een aantal vragen. Figuur 4 laat een voorbeeld zien van zo'n vraag.

In hoeverre beoordeelt u de uitgevoerde taak (taken) in het assessment als een representatieve afspiegeling van alle taken die in principe voorgelegd hadden kunnen worden aan de kandidaat?



Figuur 4 Inschatting van het kwaliteitsaspect 'generaliseerbaarheid'

Om te voorkomen dat de oordelen over de kwaliteitskenmerken, die *niet* geverifieerd worden, zullen leiden tot een te grote correctie van de ruwe scores, zou het bijbehorende computerprogramma van PPS de maximale correctiefactor standaard moeten instellen op 0,7. Afhankelijk van de oordelen van een assessor over de kwaliteitskenmerken varieert de correctiefactor dan tussen 0,7 (maximale correctie van de ruwe score) en 1 (geen correctie).

Naast de correctiefactor voor de kwaliteit van het bewijs is er ook nog een correctiefactor voor de moeilijkheidsgraad van de uitgevoerde taak. Onderwijs dat is ingericht volgens het 4C/ID-model ordent de taken waaraan studenten werken in zogeheten taakklassen. De taakklassen verschillen van elkaar in complexiteit en dus in moeilijkheidsgraad van de aangeboden taken. Ook de taken die voor assessmentdoelen worden gebruikt, zijn aanvankelijk eenvoudiger dan later in de opleiding. Binnen de gehanteerde PPS-methode zou dit kunnen leiden tot een voortijdige, positieve beslissing over het verworven zijn van de competentie. Correctie van de ruwe score op een assessment is nodig als de uitgevoerde taak eenvoudiger is dan de taken die een competent persoon in de reële werksituatie moet kunnen uitvoeren (de zogeheten criteriumtaak). Daartoe dient de assessor bij elk assessment een inschatting te geven van de moeilijkheidsgraad van de uitgevoerde taak. Als de uitgevoerde taak eenvoudiger is dan de criteriumtaak, wordt de correctiefactor kleiner dan 1. Op basis van een classificatie van drie of vier verschillende moeilijkheidsgraden, helder geoperationaliseerd aan de hand van eenduidige indicatoren voor de moeilijkheid, zou bepaald kunnen worden hoeveel kleiner dan 1 de correctiefactor in concrete gevallen moet worden.

In tabel 2 is de uitwerking van de correctiefactoren zichtbaar gemaakt. In het tweede assessment heeft de student een ruwe score van 24 gehaald. Dat is meer dan de standaard voor dit assessment (23,6). Maar na toepassing van beide correctiefactoren is de ruwe score gereduceerd tot 7,7 en moet er een negatieve beslissing genomen worden. Er zal derhalve nog meer bewijs verzameld moeten worden.

Kenmerk 5 Beslissingen nemen over competentie

Een beslissing over het al dan niet verworven zijn van een competentie wordt volgens de PPS-methode gebaseerd op een horizontale én een verticale evaluatie van de scores in het portfolio. Beide evaluaties moeten positief zijn om de conclusie te kunnen trekken dat er sprake is van competentie. We zullen dit toelichten aan de hand van tabel 2. Telkens wanneer een nieuw bewijsstuk wordt toegevoegd aan het portfolio, vindt er een evaluatie plaats door de tot dan toe verzamelde scores te vergelijken met de horizontale en verticale standaard. Horizontale evaluatie na toevoeging van een bewijsstuk houdt in dat de ruwe score, na correctie, vergeleken wordt met de som van de grensscores van de beoordelingsaspecten die gebruikt werden bij de beoordeling van de betreffende prestatie; in het geval van bewijsstuk 4 dus van alle beoordelingsaspecten behalve nummer 2 en nummer 8. Correctie houdt in dat de ruwe score (in dit geval 35), vermenigvuldigd wordt met de correctiefactoren voor kwaliteit en moeilijkheid. De gecorrigeerde score (29,8) wordt vervolgens vergeleken met de horizontale standaard (29,6). Aangezien de gecorrigeerde score groter is dan de horizontale standaard, wordt een

positieve beslissing genomen. De verticale evaluatie wordt uitgevoerd voor elk kritisch aspect en behelst een vergelijking van de grensscore met het gemiddelde van alle tot dan behaalde scores voor het betreffende kritische aspect. Als deze gemiddelde score groter of gelijk is aan de grensscore wordt een positieve beslissing genomen. Wanneer alle beslissingen 'positief' zijn, wordt de conclusie getrokken dat de kandidaat de betreffende competentie verworven heeft.

Met name de verticale standaard kan een lastig te nemen horde vormen. Een zeer lage score, behaald op een of meerdere kritische aspecten, in combinatie met een relatief hoge grensscore, vergt veel en relatief hoge scores ter compensatie. Dit zou kunnen betekenen dat een student erg veel assessments zou moeten doen. Een mogelijke oplossing hiervoor zou kunnen zijn dat het aan de student wordt overgelaten of een bewijsstuk in het beoordelingsportfolio moet worden opgenomen.

DISCUSSIE

Dit artikel beschreef een innovatieve, competentiegerichte aanpak voor het beroepsonderwijs. Kenmerkend voor zo'n aanpak is het centraal stellen van integratieve leertaken. Conform het 4C/ID-model worden deze leertaken geordend van relatief eenvoudig tot het gewenste eindniveau in zogenaamde taakklassen. Binnen elke taakklasse voeren de studenten leertaken uit die qua moeilijkheid equivalent aan elkaar zijn, maar die wel van elkaar verschillen op alle dimensies waarop taken in de beroepspraktijk ook van elkaar verschillen. Bovendien krijgen studenten veel ondersteuning en begeleiding bij het uitvoeren van de eerste leertaken in een taakklasse. Maar de laatste leertaken in een taakklasse worden door de studenten zelfstandig uitgevoerd. Deze taken kunnen dan ook goed gebruikt worden als assessmenttaken.

Deze aanpak verschilt nogal van de aanpak zoals die dikwijls in het beroepsonderwijs gehanteerd wordt. Een eerste verschil betreft de nadruk op het gebruik van 'hele' integratieve leertaken. Dit raakt het hart van het 4C/ID-model en wordt noodzakelijk geacht om de beoogde integratie van vaardigheden, kennis en attitudes te kunnen bereiken. Deze integratie is noodzakelijk om een goede transfer van het geleerde naar de beroepspraktijk mogelijk te maken. Wanneer men redeneert vanuit deze beroepspraktijk ligt het gebruik van 'hele' taken, alsmede de integratie van verschillende vakinhouden, natuurlijk voor de hand. Maar in een curriculum dat is opgezet vanuit vakken of vakinhouden vereist dit een compleet nieuwe manier van denken en werken.

Een tweede verschil betreft de continue afwisseling tussen theorie en praktijk. Het oude adagium 'eerst de theorie, dan pas de praktijk', met als uiterste implicatie alleen een stage in het laatste opleidingsjaar, wordt losgelaten. De informatie die het werken aan de leertaken in een bepaalde taakklasse ondersteunt, wordt voorafgaand aan deze taakklasse gepresenteerd of met de studenten besproken en/of kan door de studenten tijdens het werk aan deze taakklasse worden geraadpleegd. De nieuwe, aanvullende informatie die nodig is om de taken in een taakklasse van een hogere moeilijkheidsgraad te kunnen uitvoeren, komt pas bij de volgende taakklasse aan de orde zodat er

sprake is van een voortdurende afwisseling tussen theorie en praktijk. Zowel de theorie als de praktijk worden steeds op een hoger plan getild. Bovendien wordt informatie die betrekking heeft op het uitvoeren van routineaspecten van een vaardigheid (just-in-time informatie) bij voorkeur *tijdens* het werken aan de leertaken aangeboden (bijvoorbeeld door een docent in een coachende rol, een 'job aid' of een Elektronisch Performance Support Systeem).

Een derde belangrijk verschil betreft de betekenis van studievoortgang en studiesucces. De opvatting dat studievoortgang beschreven kan worden als de accumulatie van studiepunten, of als een groeiende verzameling tentamenbriefjes, wordt losgelaten. Studievoortgang betekent nu dat studenten steeds opnieuw in hun handelen moeten bewijzen competent te zijn, onder omstandigheden die in complexiteit toenemen. Dit waarborgt een geleidelijke competentieontwikkeling tot het niveau dat vereist is voor de beroepspraktijk. Naar verwachting is deze onderwijsmethode, waarbij geleerd wordt door te werken aan authentieke taken, uitdagender en motiverender voor studenten.

De in dit artikel voorgestelde procedure voor het aanleggen en beoordelen van een beoordelingsportfolio, aangeduid als *Protocol Portfolio Scoring (PPS)*, kan bijdragen aan deze nieuwe visie op studievoortgang. Bewijsmateriaal over competentieontwikkeling wordt in het beoordelingsportfolio verzameld, zo veel mogelijk rekening houdend met kwaliteitseisen op het gebied van accuraatheid, generaliseerbaarheid en extrapoleerbaarheid. Een variatie aan assessmentmethoden (de *methodemix*) wordt gebruikt om de competentie zo goed mogelijk te beoordelen. Elk nieuw bewijsstuk geeft informatie over de competentie, uitgedrukt in termen van scores op een standaardset van beoordelingsaspecten. Bij de aggregatie van de scores wordt rekening gehouden met de kwaliteit van het beoordelingsinstrument alsmede met het complexiteitsniveau waarop de student, in een bepaalde fase van de opleiding, functioneert. Beslissingen over het al dan niet verworven zijn van een competentie berusten op een verticale en horizontale evaluatie van de tot dan behaalde scores. Onderzoek zal moeten uitwijzen of en in hoeverre de verticale en horizontale standaarden leiden tot accurate beslissingen over competentie en welke meerwaarde de scorecorrecties hebben voor de accuraatheid van de beslissingen over competentie.

De PPS-methode beoogt de opbouw en beoordeling van een beoordelingsportfolio op een systematische en verantwoorde manier te laten plaatsvinden om zo te komen tot accurate beslissingen over de verwerving van competenties. Onderzoek zal moeten uitwijzen of die belofte kan worden waar gemaakt. Zo ja, dan maakt het beoordelingsportfolio het gebruik van meer traditionele vormen van toetsing, met name voor summatieve evaluatie, in feite overbodig. Toch kan een opleiding goede redenen hebben om ook conventionele kennis- en vaardigheidstoetsen te (blijven) gebruiken. Kennistoetsen kunnen bijvoorbeeld informatie verstrekken over de mate waarin ondersteunende informatie wordt beheerst; een zekere mate van beheersing kan nodig zijn om vruchtbaar te kunnen werken aan de leertaken in een volgende taakklasse. En vaardigheidstoetsen kunnen in het kader van deeltaakcoëfening informatie geven over de mate

waarin bepaalde routineaspecten van een complexe vaardigheid beheerst worden. Maar dergelijke vormen van toetsing zullen dan primair een *formatieve* functie vervullen en dus ondergeschikt zijn aan de beslissingen die op basis van het beoordelingsportfolio genomen worden.

Een ander interessant punt van onderzoek op het gebied van het beoordelen van competenties betreft de rol van de student. Vooral nog is er van uitgegaan dat het beoordelingsportfolio niet door studenten beheerd zal worden maar door docenten. Studenten hebben alleen leesrechten. Een volgende stap zou kunnen zijn dat studenten in staat worden gesteld om hun zelfbeoordelingen of beoordelingen door medestudenten (Sluijsmans, 2002), aan het beoordelingsportfolio toe te voegen. Dergelijke informatie zou, nadat de docent dit geaccordeerd heeft, meegenomen kunnen worden bij de summatieve beoordeling van competenties.

In het begin van dit artikel werd betoogd dat het traditionele onderwijsconcept in het beroepsonderwijs aan verandering toe is. Het Nederlands onderwijs is al enige jaren in de ban van de *competentiebeweging*, maar voornamelijk heeft dit niet in sterke mate bijgedragen aan het realiseren van de gewenste verandering. Een belangrijke reden daarvoor is dat er nog nauwelijks praktisch bruikbare onderwijsmodellen beschikbaar zijn voor het ontwerpen van competentiegericht onderwijs en competentiegerichte vormen van toetsing (Onderwijsraad, 2002). Wij hopen dat dit artikel enigszins in deze leemte voorziet. Het 4C/ID-model en de PPS-methode bieden praktische richtlijnen voor het ontwikkelen van competentiegericht onderwijs. De onderwijskundige waarde moet echter nog worden bewezen. Experimenten in proeftuinachtige opzetten zijn daarom nodig. In de eerste plaats om na te gaan wat het betekent voor de direct betrokkenen om onderwijs te verzorgen dan wel te krijgen dat zo totaal anders is georganiseerd. In de tweede plaats om vast te stellen in hoeverre de aanpak geschikt zou zijn voor toepassing op grote schaal. In de derde plaats om te onderzoeken welke onderdelen uit het instructie- en assessmentplan voldoende functioneren en welke voor verandering of vervanging in aanmerking komen. Wij hopen en verwachten dat de hier geschetste aanpak de discussie over de vormgeving van het competentiegerichte beroepsonderwijs kan bevorderen.

LITERATUUR

- Adviescommissie Boekhoud (2002) *Doorstroomagenda beroepsonderwijs*. Zoetermeer: Ministerie van Onderwijs, Cultuur en Wetenschappen.
- Adviescommissie Onderwijs-Arbeidsmarkt (ACOA) (1999) *Een wending naar kerncompetenties: De betekenis van kerncompetenties voor de versterking van de kwalificatiestructuur secundair beroepsonderwijs*. Den Bosch: ACOA.
- Bolhuis, B., & Van Royen, R. (2003) *Competent blauw op straat. Proeven van bekwaamheid in het vernieuwde politieonderwijs*. Apeldoorn: LSOP/Centraal Examenbureau Politie.
- Crooks, T.J., Kane, M.T., & Cohen, A.S. (1996) Threats to the valid use of assessments. *Assessment in Education*, 3, 3, 265-285.

- De Bruijn, E. (2003) *De pedagogisch-didactische benadering in de beroepsonderwijskolom. Bouwstenen voor een herontwerp*. 's-Hertogenbosch: Cinop.
- De Croock, M.B.M., & Van Merriënboer, J.J.G. (2003) Een trainingsblauwdruk voor het leren van complexe cognitieve vaardigheden: het 4C/ID*-model. In: P.W.J. Schramade (Red.), *Handboek Effectief Opleiden (5.1-4.01-4.42)* 's-Gravenhage: Delwel Uitgeverij B.V.
- De Graaff, M.J.C. (1993) Managers als beoordelaars. In: P.G.W. Jansen & F. de Jong (Red.), *Assessment Centers: Een open boek* (pp. 62-73). Utrecht: Spectrum.
- De Groot, A.D., & Van Naerssen, R.F. (1977) *Studietoetsen: construeren, afnemen, analyseren. Deel II*. Den Haag: Mouton Publishers.
- Dierick, S., Watering, G. van de, & Muijtjens, A. (2002) De actuele kwaliteit van assessment: ontwikkelingen in de edumetrie. In: F. Dochy, L. Heylen, & H. van de Mosselaer (Red.), *Assessment in onderwijs: nieuwe toetsvormen en examinering in het studentgericht onderwijs en competentiegericht onderwijs* (pp. 91-122). Utrecht: Lemma.
- Dochy, F., & Struyven, K. (2002) Assessment: betekenis en assessmentvormen. In: F. Dochy, L. Heylen, & H. van de Mosselaer (Red.), *Assessment in onderwijs: nieuwe toetsvormen en examinering in het studentgericht onderwijs en competentiegericht onderwijs* (pp. 33-60). Utrecht: Lemma.
- Grotendorst, A., Jellema, M., Stam, I., Vegt, M. van der, & Zandbergen, C. (2002) *Leren in veiligheid. Het nieuwe politieonderwijs in maatschappelijk perspectief*. Apeldoorn: LSOP Politie Onderwijs- en Kenniscentrum.
- Hofstee, W.K.B. (1983) Beoordeling van subsidie-aanvragen voor onderwijsresearch: een psychometrische evaluatie. *Tijdschrift voor Onderwijsresearch*, 8, 273-284.
- Inspectie van het Onderwijs (2003) *Het gebruik van het portfolio als nieuwe toetsvorm in het hoger onderwijs. Eindverslag portfolio-onderzoek*. Rapport nr. 2003-10. Utrecht: Inspectie van het Onderwijs.
- Janssen-Noordman, A.M.B., & Van Merriënboer, J.J.G. (2002) *Innovatief onderwijs ontwerpen*. Groningen: Wolters-Noordhoff.
- Janssen-Noordman, A.M.B., Van Merriënboer, J.J.G., van der Vleuten, C., & Bouhuijs, P. (2000) Het leren van complexe taken. *Onderzoek van Onderwijs*, november 2000, 37-40.
- Kane, M.T. (1992) *The validity of assessments of professional competence*. (ERIC Document Reproduction Service No. ED 343 958).
- Kessels, J.W.M. (1996) *Succesvol ontwerpen: Curriculumconsistentie in opleidingen*. Deventer: Kluwer Bedrijfswetenschappen.
- Linn, R.L., Baker, E.L., & Dunbar, S.B. (1991) Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20, 8, p.15-21.
- LSOP Politie Onderwijs- en Kenniscentrum (2001) *Functioneel ontwerp van Politieonderwijs 2002: De inrichting van het vernieuwde politieonderwijs* (oktober 2001). Apeldoorn: LSOP Politie Onderwijs- en Kenniscentrum.
- Lunz, M.E., Wright, B., & Linacre, M. (1990) Measuring the impact of judge severity on examination scores. *Applied Measurement in Education*, 3, 4.
- Mertens, F. (1998) Beroepsonderwijs zonder beroepen. *Gevraagd is een herontwerp van technische. Meso-magazine*, 100, pp.12-19.
- Meijers, F. (1998) The development of a career identity. *International Journal for the Advancement of Counseling*, 20, pp. 191-207.

- Onderwijsraad (2002) *Competenties: van complicaties tot compromis*. Den Haag: Onderwijsraad.
- Sanders, A.J. (1980) *Evaluatierapport blok 155 studiejaar 1977-1978*. Instituut Conserverende Tandheelkunde voor Volwassenen. Nijmegen: Katholieke Universiteit.
- Sluijsmans, D. (2002) *Student involvement in assessment*. Academisch Proefschrift. Heerlen: Open Universiteit Nederland.
- Straetmans, G.J.J.M. (1985) *Evaluatie in het tandheelkundig onderwijs: beoordelen van practicumwerkstukken en meten van probleemoplosvaardigheid*. Academisch proefschrift. Nijmegen: Katholieke Universiteit.
- Straetmans, G.J.J.M. (1998) Toetsing van competenties. In: P.W.J. Schramade (Red.). *Handboek Effectief Opleiden (9.1-3.01-3.36)*. 's-Gravenhage: Delwel Uitgeverij B.V.
- Straetmans, G.J.J.M., & Sanders, P.F. (2001) *Beoordelen van competenties van docenten*. Den Haag: Programmamanagement EPS/HBO-raad.
- Straetmans, G.J.J.M. (in voorbereiding). *Protocol Portfolio Scoring. Een methode voor het systematisch scoren en vaststellen van competenties*. Perspectief op Assessment. Arnhem: Citogroep.
- Van Merriënboer, J.J.G., Jelsma, O. & Paas, F.G.W.C. (1992) Training for reflective expertise: A four-component instructional design model for training complex cognitive skills. *Educational Technology, Research and Development*, 40, 23-43.
- Van Merriënboer, J.J.G. (1997) *Training complex cognitive skills: A four-component instructional design model for technical training*. Englewood Cliffs, NJ: Educational Technology Publications.

NOTEN

- ¹ Vermeld moet worden dat het LSOP in werkelijkheid niet volgens het 4C/ID-model heeft gewerkt, maar volgens de ontwerpmethodiek 'competentiegericht curriculum' (Kessels, 1996).
- ² Triage verwijst in dit verband naar het selecteren van oorlogsgewonden met het oog op de aard en snelheid van behandeling om op die manier zo veel mogelijk slachtoffers te kunnen redden.
- ³ Om terminologische verwarring te voorkomen, gebruiken we in het vervolg de term 'beoordelingsportfolio' als het gaat over een portfolio dat voor summatieve evaluatie wordt ingezet.

