

Assessment en leren als een twee-eenheid: onderzoek naar de impact van assessment op leren

Prof. dr. Mien Segers
(segers@fsw.leidenuniv.nl) is
hoogleraar opleidingskunde aan
de Universiteit Leiden

In dit artikel worden de ontwikkelingen beschreven in het denken over assessment zowel in de schoolse context als in de context van professionele opleidings- en ontwikkelingsprogramma's. Voorbeelden van onderzoeksstudies naar de kwaliteiten van nieuwe vormen van assessment worden gepresenteerd. Deze zijn beperkt tot twee vormen van assessment die in verschillende onderwijs- en professionele opleidingsprogramma's worden geïmplementeerd: self- en peerassessment en case-based assessment. Het accent ligt hierbij op een kwaliteitskenmerk dat vanuit edumetrisch perspectief in toenemende mate aandacht krijgt: de consequentiële validiteit of het effect van assessment op verschillende aspecten van leren. De resultaten van deze studies indiceren dat, bij een zorgvuldige implementatie, self- en peerassessment positieve effecten hebben op cognitief, metacognitief en motivationeel vlak. Onderzoeksstudies naar de OverAll Toets als een vorm van case-based assessment, wijzen op het belang van de condities waaronder deze nieuwe vorm van assessment wordt geïmplementeerd. De percepties van de studenten van de kernelementen van de leeromgeving vormen een filter tussen de leer- en assessmentomgeving zoals ontworpen en het studiegedrag van studenten.

Tot slot wordt een assessmentonderzoeksagenda voor de toekomst voorgesteld. Gezien het toenemende belang van continue professionele ontwikkeling, wordt voorgesteld het onderzoek naar assessment uit te breiden naar professionele opleidings- en ontwikkelingsprogramma's. Drie relevante onderzoeklijnen worden besproken: onderzoek naar de interactie tussen eerdere ervaringen van de lerenden met assessment, hun perceptie van de werk-leeromgeving en de individuele leeropbrengsten van opleidingen; onderzoek naar condities waaronder vormen van peerassessment bijdragen tot de professionele ontwikkeling van lerenden; onderzoek naar assessment van probleemoplossende vaardigheden.

INLEIDING

De wijze waarop leerlingen, studenten en professionals in onderwijs-, opleidings- en ontwikkelingsprogramma's worden beoordeeld, speelt een cruciale rol in hun leerproces. Het is dan ook niet verwonderlijk dat allerlei vormen van beoordelingen zoals toetsen en examens een centrale rol spelen in hun denken over leren, onderwijs en opleiden.

Beoordelingen beïnvloeden het leerproces op tweeërlei wijze. Allereerst wordt een belangrijk deel van de school- of opleidingstijd besteed aan het uitvoeren van beoordelingen. De afgelopen tien jaar nam het gewicht dat wordt toegekend aan allerlei vormen van summatieve, certificerende toetsen zelfs toe. Dit is niet alleen het geval in Nederland waar bijvoorbeeld in het BVE, met de invoering van kwalificatiestructuren, voor wat betreft de toetsing de aandacht nadrukkelijk wordt gericht op examinering en externe legitimering. Dezelfde trends zijn terug te vinden in andere Europese landen en de Verenigde Staten. In de Verenigde Staten is het aantal staten dat gestandaardiseerde, certificerende externe toetsen hanteert, tussen 1995 en 1998 met 40% gestegen (Clarke e.a., 2000). Harlen & Crick (2001) verwijzen naar een onderzoek dat is uitgevoerd in 2000 door de Britse Qualifications and Curriculum Authority. De resultaten tonen aan dat de introductie van de nationale examens niet resulteerde in een afname maar in een toename van het gebruik van andere examens. Uit een schatting van de Professional Association of Teachers blijkt dat de gemiddelde Britse leerling tussen zijn vierde en achttiende levensjaar zestig examens aflegt. Beide auteurs concluderen: *'The USA and England now vie for the title of "most tested nation". When Resnick and Nolan (1995) claimed this title for the USA, noting that there were few countries today that gave these formal examinations to students before the age of 16, they were not taking into account of the rapid, and what may have seemed untypical, changes in the UK.'* (p. 171). Ook uit een onderzoek van Smith (2000) in de context van Vocational Education and Training (VET) in Queensland blijkt *'an excessive emphasis on summative assessment'* (p. 26), voor een belangrijk deel veroorzaakt door het gebruik van Training Packages die eerder focussen op wat, waar en wanneer te testen in plaats van op hoe de training te structureren en aan te bieden. Tegelijkertijd heeft dit geleid tot een significante daling in het gebruik van formatief assessment.

Naast de tijd die de lerende besteedt aan allerlei vormen van summatieve beoordelingen, heeft assessment ook invloed op het leerproces en op de leeruitkomsten. Uit onderzoek naar de effecten van summatieve toetsen en examens blijkt dat de scores van de lerenden op deze toetsen in de tijd toenemen. Echter, dit is eerder te danken aan de grotere vertrouwdheid van docenten en lerenden met de examens dan aan effectieve leerwinst (bijvoorbeeld Kohn, 2000; Koretz, 1988, 1991; Linn, 2000). Daarnaast blijkt uit een aantal onderzoeksstudies het negatieve effect van deze toetsen en examens op verschillende aspecten van het leren, zoals het leergedrag en de motivatie. Scouller (1996) bijvoorbeeld toonde aan dat het gebruik van multiplechoicetoetsen, een veelvuldig gebruikte vorm bij summatieve toetsen, leidt tot een oppervlakkig studiedrag. Ook Harlen & Crick (2001) wezen in hun overzichtsstudie op de negatieve effecten van nationale examens op verschillende aspecten van motivatie, namelijk (1) zelfsteem van de leerlingen (vooral zij die zwakker scoren); (2) angst en onzekerheid

over wat wordt verwacht en over eigen kunnen; (3) door de leerling verrichte inspanningen ('effort'); (4) *locus of control*; en (5) doelgerichtheid van de leerling. Deze negatieve effecten blijken toe te nemen met de leeftijd van de lerenden: *'Instead of increasing motivation and "test wiseness" with increasing age, older students feel more resentment, anxiety, cynicism and mistrust of standardised achievement tests.'* (p. 196). De effecten zijn bovendien meer uitgesproken negatief voor zwakker presterende lerenden. Uit dezelfde studie blijkt dat de mate waarin toetsen en examens een negatieve impact hebben, gerelateerd is aan de mate van *self-efficacy* van de lerenden, de mate van intrinsieke en extrinsieke motivatie, de mate waarin de leeromgeving zelfregulatie stimuleert en de druk uitgeoefend door voor de lerende belangrijke externe personen.

In reactie op de aandacht voor en de centrale plaats van summatieve toetsen in het onderwijs, zowel in de klas als op het niveau van de nationale examens, wordt in toenemende mate gepleit voor een evenwichtige plaats voor formatieve toetsing in het onderwijs. Het is niet verwonderlijk dat met name in het Verenigd Koninkrijk (bijvoorbeeld Black & William, 1998; Harlen & Crick, 2001) en de Verenigde Staten (bijvoorbeeld Stiggins, 2001) steeds meer nadruk wordt gelegd op de verschuiving van 'assessment of learning' naar 'assessment for learning'.

Tegelijk is er een groeiende aandacht voor onderzoek naar nieuwe vormen van assessment. *Peer- en selfassessment, portfolioassessment, performance-based assessment, case-based en projectbased assessment* zijn voorbeelden van assessmentmethoden die in veel onderwijs-, opleidings- en ontwikkelingsprogramma's zijn geïmplementeerd. Ze kunnen zowel een formatieve als een summatieve functie vervullen. Verwacht wordt dat deze nieuwe vormen van assessment een positieve invloed zullen hebben op het leren. Het onderzoek op het terrein van assessment richt zich in toenemende mate op deze vormen van assessment. Zowel vanuit psychometrisch als vanuit edumetrisch perspectief worden steeds meer verschillende kwaliteitsaspecten van deze vormen van assessment onderzocht.

EEN VERANDERDE VISIE OP LEREN

De huidige maatschappij stelt in toenemende mate hoge eisen aan professionals en aan de organisaties waar ze te werk zijn gesteld. De toenemende mate van globalisering, het snelle tempo waarin de technologie het bedrijfsleven en de maatschappij stuurt en de snelle kennisveroudering zijn slechts enkele voorbeelden van ontwikkelingen die een belangrijke invloed hebben op het professionele leven.

Dit leidt tot een groot aantal verwachtingen van, en kritieken op het onderwijs dat voorbereidt op het professionele leven. Een voorbeeld hiervan is de kritiek op de MBA-afgestudeerden: *'Beginning in the late 1980's MBA programs received intense criticism for being too focused on abstract learning. MBA graduates were viewed as: (1) too analytical, not practical and action oriented; (2) lacking interpersonal and in particular communication skills; (3) parochial, not global in their thinking and values; (4) having exceedingly high expectations about their first job after graduation... (5) not oriented toward information*

resources and systems; and (6) not working well in groups.' (Boyatzis e.a., 1995, p. 4). In reactie op deze en andere kritieken worden op velerlei niveaus kerncompetenties geformuleerd. De Britse National Skills Task Force (1999) bijvoorbeeld stelde zes kerncompetenties voor waarover professionals moeten beschikken: communicatieve vaardigheden; reken-/wiskundevaardigheden; kennis over en vaardigheden in informatie- en kennistechnologie; probleemoplossend vermogen; kunnen samenwerken met anderen en voldoende leervermogen.

Het spreekt voor zich dat de ontwikkeling van deze competenties niet alleen een zaak is van het initiële onderwijs. Gezien de genoemde ontwikkelingen en het tempo waarin deze plaatsvinden, neemt het belang van leren na de schoolperiode toe. De Grip (2000) wijst erop dat waar het postinitieel onderwijs in de jaren vijftig en zestig eerder het karakter had van tweedekansonderwijs, het volgen ervan nu eerder een noodzaak is om zich op de arbeidsmarkt te kunnen handhaven en mee te groeien met de eisen die er worden gesteld. Hij spreekt in dit verband over het voorkomen van kwalificatieveroudering.

In samenhang met de geschetste maatschappelijke ontwikkelingen, is het concept 'leren' de voorbije jaren gherdefinieerd. De huidige visie op leren is beïnvloed door inzichten en onderzoek vanuit verschillende aanverwante disciplines. Het constructivisme als leertheorie en de cognitieve psychologie en cognitieve wetenschappen, meer bepaald onderzoek naar de kenmerken van expertise, expertiseontwikkeling en cognitieve ontwikkeling, hebben hierbij een dominante rol gespeeld. Dit heeft geleid tot het beschrijven van het concept leren aan de hand van een set van principes (Birenbaum, 2003; Glaser, 1990; Pellegrino e.a., 2001; Tynjälä, 1998; Wolf e.a., 1991).

Allereerst wordt leren beschouwd als een actief constructief proces waarbij de lerende nieuwe informatie interpreteert op basis van zijn bestaande kennisstructuur (Glaser, 1990; Tynjälä, 1998; Wolf e.a., 1991). Nieuwe informatie wordt ingepast in de bestaande kennisstructuren, waardoor deze voortdurend worden aangepast en/of verbijzonderd. Dit proces van constructie en reconstructie wordt gedefinieerd als 'leren'. Een goed georganiseerd kennisbestand is een voorwaarde voor succesvol probleem oplossen. Uit expertiseonderzoek blijkt immers dat experts en novieten zich onderscheiden in de mate van georganiseerdheid van hun kennisbestand waardoor ze in staat zijn problemen te definiëren op basis van onderliggende patronen en principes. Dit leidt tot kwalitatief betere probleemanalyses en probleemoplossingen (Glaser, 1990).

Leren als een actief constructief proces waarbij de lerende een centrale rol speelt, impliceert het belang van cognitieve en metacognitieve strategieën (Birenbaum, 2003). Daarbij wordt van de lerende verwacht dat hij in staat is te reflecteren op zijn eigen denken en handelen, en dit zelf te sturen. Dit betekent dat deze benadering van leren zelfregulatie vooronderstelt: lerenden die in staat zijn in te schatten waar en wanneer bepaalde cognitieve strategieën relevant zijn, wanneer bepaalde procedures en regels moeten worden toegepast; leerlingen die vooruit kunnen plannen, hun leerproces kritisch kunnen monitoren waarbij ze op een efficiënte wijze hun inspanningen en tijdsinvestering inzetten en de inspanningen kunnen evalueren in functie van de leeropbrengsten (Birenbaum, 2003; Boekaerts, 2002; Glaser, 1990). Ook uit expertiseonderzoek blijkt dat experts beter dan novieten in staat zijn om de moeilijkheidsgraad

van problemen in te schatten en hun probleemoplossingproces kritisch te evalueren (Glaser, 1990).

Leren is contextgebonden (Birenbaum, 2003; Glaser, 1990; Tynjälä, 1998). Niet alleen verschillen tussen culturen, maar ook de verschillen tussen individuen binnen een bepaalde cultuur als gevolg van verschillen in eerdere individuele ervaringen, leiden tot het toekennen van verschillende betekenissen aan een bepaald fenomeen.

Daarbij is leren een fundamenteel sociaal proces dat plaatsvindt binnen een sociale context (Birenbaum, 2003; Glaser, 1990; Tynjälä, 1998). Interactie met anderen bepaalt de wijze waarop individuen kijken naar fenomenen. Hierbij kan de groep waartoe de lerende behoort zowel de functie vervullen van aanbieder van voor de lerende nieuwe informatie, als van aanbieder van conflicterende informatie (vanuit het standpunt van de lerende bekeken). In beide functies draagt het sociale proces bij tot construeren of reconstrueren van de individuele kennisstructuren.

Leren is meer dan alleen kennisverwerving. Leren is een proces van enculturatie in een bepaalde gemeenschap, waarbij de wijze van denken en handelen die eigen zijn aan deze gemeenschap, worden overgenomen (Birenbaum, 2003; Billett, 1996).

Tot slot is leren niet enkel een cognitief proces. Motivationale, emotionele en sociale aspecten spelen een cruciale rol in het leerproces (Birenbaum, 2003; Boekaerts, 2003). Voorbeelden zijn: de doeloriëntatie van de lerende, zijn verwachtingen, de waarde die hij toekent aan de leertaak, de interacties met anderen en hoe de lerende zich voelt in het leerproces. Leren betreft dan ook niet alleen de cognitieve ontwikkeling, maar impliceert ook de ontwikkeling van deze aspecten.

Deze visie op leren heeft niet alleen de implementatie van verschillende zogenoemde innovatieve leeromgevingen in schoolse contexten bevorderd, maar tevens de vormgeving van leren en opleiden op de werkplek. In schoolse settings is ruime ervaring opgedaan met, en in meer beperkte zin onderzoek verricht naar, leeromgevingen zoals het probleemgestuurd en het projectgestuurd onderwijs (bijvoorbeeld Dochy e.a., 2003; Segers e.a., 2003). In werkpleksettings wordt niet alleen gepleit voor de implementatie van probleemgestuurde leeromgevingen (Stonyer & Marshall, 2002), maar zijn deze reeds in verschillende sectoren (bijvoorbeeld technologie, management, landbouw, geneeskunde) geïmplementeerd. Lohman (2002) identificeerde 106 publicaties over probleemgestuurde leeromgevingen in de context van opleidingen en professionele ontwikkeling. Naast deze specifieke leeromgevingen worden in toenemende mate in organisaties allerlei vormen van leren ondersteund die aansluiten bij de genoemde kenmerken van effectief leren. Norman Ashton (1994) beschrijft hoe IBM (UK) Ltd. deze veranderende visie op leren heeft geoperationaliseerd: *'In 1987/1988, we recognised that IBM had to transform. A key part of that transformation has been to recognise visibly that people are our key asset... Our people are our major resource, and we have to help them to develop themselves. The focus has moved to the individual being responsible for their own careers. A lot of people need help on this particularly if you go back to the company paternalism that existed in many companies well into the 80s. It is certainly having a very positive effect if we look the way our customer satisfaction index has moved up over the last two or three years.'* IBM implementeerde een aantal strategieën met als doel de werknemers maximaal te ondersteunen in hun professionele ontwikkeling. De functie van managers

was niet langer uitvoerend maar ondersteunend. Dit impliceerde voor de managers het delegeren van verantwoordelijkheden naar de werknemers. Dit bleek een leerproces voor zowel de managers als de werknemers. Zelfverantwoordelijkheid staat niet alleen centraal in het uitvoeren van taken maar ook en vooral in het ontwikkelen van de competenties om deze taken uit te voeren en dus de verwachte output te realiseren. De werknemer bepaalt zelf op welke manier hij aan deze competenties wil werken: trainingen, externe contacten, bestuderen van informatiebronnen, enz. Op basis van het competentieprofiel dat de werknemer heeft opgesteld, wordt een leercontract gemaakt waarin wordt aangegeven aan welke competenties wordt gewerkt en met welk verwacht resultaat. In dit proces van het opstellen en evalueren van competentieprofielen van individuele medewerkers speelt self- en peerassessment een cruciale rol. Om het leerproces van de werknemers te ondersteunen krijgt elk lid van het managementteam de functie van mentor waarbij hij de verantwoordelijkheid heeft voor het begeleiden van zes tot zeven werknemers van verschillende afdelingen. Daarnaast is teamleren een belangrijk aspect van het werken en leren bij IBM. In toenemende mate wordt gewerkt in projectteams waarin werknemers uit verschillende afdelingen samenwerken. Tot slot heeft IBM een Expertenraad (expert board) bestaande uit collega-werknemers ingesteld, die zowel de carrièreontwikkelingen van individuele medewerkers beoordeelt op basis van hun portfolio's als het opleidingsaanbod van IBM en de competentieprofielen die aan de basis van elke functie in het bedrijf liggen.

Deze herformulering van de kenmerken van leren heeft niet alleen geleid tot de implementatie van innovatieve leeromgevingen maar is tevens een belangrijke impuls geweest voor de revisie van de kenmerken van assessment.

VAN TESTEN NAAR ASSESSMENT

Leren, instructie en assessment zijn onlosmakelijk met elkaar verbonden en de afstemming of *'alignment'* (Biggs, 1999) is essentieel voor het bereiken van de doelen van onderwijs en professionele opleidings- en ontwikkelingsprogramma's (Birenbaum, 2003). De herformulering van het concept 'leren' induceerde een heroverweging van de wijze waarop lerenden worden beoordeeld. Birenbaum (2003) vat dit samen als de evolutie van een *'test culture'* naar een *'assessment culture'*. Wanneer ik in dit artikel verwijs naar toetsen en beoordelen in de assessmentcultuur zal ik het begrip 'assessment' gebruiken.

De grootste veranderingen in toetsen en beoordelen zijn de evolutie van *testing* naar meervoudig assessment, en van geïsoleerde toetsing naar geïntegreerd assessment. Deze evolutie is niet alleen zichtbaar in de schoolse context maar vindt, weliswaar nog in beperkte mate, steeds meer ingang in de context van opleidingen en professionele ontwikkeling (Kasworm & Marienau, 1997).

In figuur 1 wordt de evolutie van de testcultuur naar de assessmentcultuur geschetst door middel van zeven continua (Segers, 2003).

Toetsing	Assessment	
Authenticiteit	—————→	
	Gedecontextualiseerd, atomistisch	Gecontextualiseerd, in samenhang
Aantal meetinstrumenten en -momenten	—————→	
	Enkelvoudig	Meervoudig
Beheersingsniveaus	—————→	
	Laag	Hoog
Dimensies van competenties	—————→	
	Weinig/cognitief	Meerdere
Relatie tot het leerproces	—————→	
	Geïsoleerd assessment	Geïntegreerd assessment
Verantwoordelijkheid	—————→	
	Docent/examinator	Student

Figuur 1 De karakteristieken van assessment (Segers, 2003).

Het eerste continuüm verwijst naar de verschuiving van gedecontextualiseerde, atomistische tests naar authentieke gecontextualiseerde vormen van assessment. De testcultuur werd gedomineerd door testen waarbij de beheersing van separate kenniselementen, zoals de kennis van feiten en begrippen, werd gemeten. In het geval van vaardigheidsonderwijs werden de elementen van vaardigheden los van elkaar en van elke reële context gemeten. Een duidelijk voorbeeld hiervan is de toetsing in het vreemdetalenonderwijs. Shohamy (1995) verwijst in dit verband naar de evolutie van *'discrete point testing'* (bijvoorbeeld invuloefeningen waarbij een ontbrekend woord in een geïsoleerde zin dient te worden ingevuld) naar *'integrative language testing'* (toetsopdrachten zoals het schrijven van brieven en het samenvattend weergeven van de inhoud van een verhaal) naar *'communicative language testing'* (toetsopdrachten waarbij leerlingen in reële situaties de taal gebruiken als communicatiemiddel). In de context van volwassenenonderwijs en met name *'literacy education'*, wijst Kerka (1995) op het belang van dit kenmerk van assessment: *'Workplace literacy programs are particularly rooted in the context of the job site; Bousquet, Fleming and Lloyd (1994) describe a workplace assessment in which participants are given a scenario depicting a work-related situation - choosing among two job offers - and must make a choice, explain their strategy for choosing it, give supporting facts, and state why the alternative was not chosen.'* (p. 3) Ook Kasworm en Marienau (1997, p. 2) verwijzen naar het belang van authenticiteit in de context van opleiding en professionele ontwikkeling: *'Assessment of adult learners moves*

beyond the knowledge of abstract content to the world of situated cognition. Here the importance of context in establishing meaningful connections among knowledge, skills, and experiences becomes one of the cornerstones for the creation and implementation of assessment approaches. In de praktijk sluit het gebruik van assessmentmethoden zoals portfolioassessment, projectbased assessment en performance-based assessment nauw aan bij het authenticiteitskenmerk van assessment.

Het tweede continuüm toont de verschuiving van de beschrijving van de competenties van de lerende door middel van één maat, naar het beschrijven door middel van een competentieprofiel gebaseerd op meerdere maten en meerdere meetmomenten. Met name in de context van opleidingen en ontwikkeling wordt gewezen op het belang van dit kenmerk, omdat volwassenen in verschillende formele en informele situaties kennis opdoen en de verwerking van informatie wordt gekleurd door hun voorkennis (Kasworm & Marienau, 1997). Lantz en Friedrich (2000) bijvoorbeeld ontwikkelden en onderzochten een Instrument for Competency Development. Het doel van dit instrument is op basis van interviews in kaart te brengen wat de geïnterviewde heeft geleerd op een bepaald moment, intentioneel of niet. Op basis van een kwalitatieve en kwantitatieve analyse worden de resultaten van deze interviews (welke competentie is geleerd en op welk beheersingsniveau) uitgedrukt in een competentieprofiel. Hierbij wordt rekening gehouden met zowel reactief als pro-actief leren op de werkplek.

Het derde continuüm indiceert de evolutie van het meten en beoordelen van lage beheersingsniveaus naar hogere beheersingsniveaus. Traditionele testen werden in vele gevallen gekenmerkt door het meten van de reproductie van kenniselementen. Aansluitend bij de kenmerken van leren en de kenmerken van expertise spreekt het voor zich dat deze lagere beheersingsniveaus een voorwaarde zijn voor het kunnen hanteren van de betreffende kennis voor het oplossen van problemen. Echter, het beperken van de toetsing tot deze niveaus impliceert dat de toetsing een zeer beperkt beeld geeft van de mate waarin de lerende expertise heeft ontwikkeld. Het gebruik van performance-based assessment, projectbased assessment en case-based assessment zijn voorbeelden van deze verschuiving naar het meten van hogere beheersingsniveaus. Kasworm en Marienau (1997) verwijzen in dit verband naar het assessmentprogramma van het National External Diploma Program (American Council of Education), een highschooldiploma voor volwassenen die academische vaardigheden (bijvoorbeeld communicatievaardigheden, probleemoplossende vaardigheden, leervaardigheden, ...) buiten de schoolse context hebben verworven en deze competenties kunnen aantonen in een performance-assessmentproces. Dit proces vraagt de volwassenen deze vaardigheden aan te tonen in een aantal authentieke situaties. Voorbeelden zijn: kaarten interpreteren, authentieke wiskundige problemen oplossen en een opiniërend stuk schrijven voor een krant.

Het vierde continuüm betreft de multidimensionaliteit van de competenties die we nastreven in het onderwijs en professionele opleidingsprogramma's. Niet alleen cognitieve aspecten van competenties zijn belangrijk, maar in toenemende mate metacognitieve, affectieve en sociale dimensies, en in bepaalde disciplines en opleidingen psycho-

motorische aspecten. Deze moeten dan ook meegewogen worden in het assessment. Het gebruik van assessmentmethodes zoals self- en peerassessment en portfolioassessment sluit hier nauw bij aan.

Het vijfde continuüm legt de nadruk op de relatie tussen assessment en leerproces. Het betreft het belang van een evenwichtige balans tussen summatief en formatief assessment, waarbij assessment niet een leerproces afsluit, maar stimuleert en informeert. In de context van organisaties sluit de implementatie van ontwikkelingsprogramma's zoals het Persoonlijke OntwikkelingsPlan (POP) hierbij aan. Een voorbeeld is het Ameritech Corporation leadership development program: *'Each frontline manager is assessed on these competences prior to the start of a five-day program, using a 360-degree approach that solicits views from the person's supervisor, peers, and subordinates, and from the manager herself through self-assessment. At the conclusion of the instructional program, each manager is assessed again, using the same 360-degree approach, and the results are used to develop an individualized development plan targeting areas for further improvement and the strategies best suited to that individual. For at least a six-month period, each manager consults every thirty days with her or his instructor and supervisor to get feedback on progress and obtain additional support as needed.'* (Kasworm & Marienau, 1997, p. 4).

De zesde dimensie verwijst naar de verantwoordelijkheid voor het assessment. De toenemende verantwoordelijkheid van de lerende voor het eigen leerproces, impliceert een heroverweging van de rol van de lerende in het beoordelen van dit leerproces en de resultaten ervan. Het toenemend gebruik van self- en peerassessment in zowel schoolse context als de context van opleiding en professionele ontwikkeling zijn voorbeelden van deze verschuiving.

De beschreven verschuiving van de testcultuur naar de assessmentcultuur heeft de implementatie van nieuwe vormen van assessment bevorderd. Onderwijsgevend en opleiders hebben hoge verwachtingen van de invoering van nieuwe vormen van assessment. Met name self- en peerassessment en case-based assessment worden in toenemende mate geïmplementeerd in onderwijs- en opleidingssettings. Hoewel de literatuur veelvuldig de veronderstelde positieve effecten beschrijft van deze vormen van assessment, beperkt het onderzoek zich voornamelijk tot schoolse settings en is er bijzonder weinig onderzoek verricht in de context van opleiding en professionele ontwikkeling. Het onderzoek is meestal case-based. Bovendien wordt in een aantal van deze studies de kwaliteiten van deze vormen van assessment voornamelijk benaderd vanuit een psychometrisch perspectief met een accent op het meten van bijvoorbeeld de interbeoordelaarsbetrouwbaarheid. Vanuit de edumetrie wordt gewezen op het belang van de verbreding van de psychometrische kwaliteitseisen (Frederickson & Collins, 1989; Haertel, 1991; Linn e.a., 1991; Dierick & Dochy, 2001; Gielen e.a., 2003). De consequentiële validiteit of het effect van assessment op het leer- en instructieproces wordt in deze visie als een van de essentiële kwaliteitskenmerken geponeerd. In dit artikel beperk ik mij dan ook tot het presenteren van onderzoek naar self- en peerassessment en case-based assessment en dit vanuit edumetrisch perspectief. Centraal staat de vraag naar de mate waarin deze nieuwe vormen van assessment invloed hebben op het leren.

HET EFFECT VAN NIEUWE VORMEN VAN ASSESSMENT OP LEREN

Assessment heeft op verschillende momenten en via verschillende mechanismen invloed op het leerproces en het leergedrag. Gielen e.a. (2003) maken een onderscheid tussen pre-assessment-, post-assessment- en true-assessmenteffecten. Het pre-assessmenteffect, ook 'backwash effect' of 'the forward function' genoemd, betekent dat wanneer de lerende een assessment verwacht met bepaalde kenmerken, hij zich in zijn leerproces zal richten op deze kenmerken. Een voor de hand liggend voorbeeld is het gebruik van reproductieve toetsen die er toe leiden dat lerenden zich tijdens het leerproces richten op memorisatie. Het post-assessmenteffect is de impact op het leerproces van de feedback die de lerende ontvangt naar aanleiding van zijn prestatie op de assessmenttaken. Het true-assessmenteffect verwijst er naar dat de lerende leert door het uitvoeren van de assessmenttaken op zich.

Gielen e.a. (2003) verwijzen naar de rol van verschillende factoren op het effect dat assessment heeft op leren. Deze betreffen eerder genoemde kenmerken van assessment, namelijk: (1) de authenticiteit van de assessmenttaken; (2) de hogere 'beheersingsniveaus' die worden gemeten; (3) de integratie van assessment in het leerproces; en (4) de verantwoordelijkheid van de student bij het assessment.

Authenticiteit

Een eerste kenmerk van assessment, authenticiteit (figuur 1), heeft betrekking op het karakter van de gehanteerde assessmenttaken. Verondersteld wordt dat dit authentieke karakter bijdraagt tot de motivatie van de lerenden omdat ze de taak als meer interessant en betekenisvol ervaren. Ze zien de relevantie en het nut van de assessmenttaak in ('task-value of assessment') (Dochy & Moerkerke, 1997). Bovendien stimuleert het gebruik van gecontextualiseerde assessmenttaken de transfer van kennis en het leren van probleemoplossende vaardigheden waarbij het aanwenden van verschillende deelvaardigheden in hun samenhang wordt vereist.

Hogere beheersingsniveaus

Assessment impliceert dat meer dan kennisreproductie wordt vereist. Assessmenttaken hebben een hogere cognitieve complexiteit dan de traditionele testen (figuur 1). Er zijn indicaties dat lerenden meer diepgaand leergedrag ('deep learning strategies') zullen vertonen wanneer ze complexe assessmenttaken verwachten, dan wanneer ze reproductieve toetsen met gesloten vragen vooronderstellen. De lerenden zoeken additionele informatie op, reflecteren kritischer op de leerinhoud en richten hun aandacht op een diepgaande, persoonlijke verwerking van deze leerinhoud (McDowell, 1995; Ramsden, 1988; Sambell e.a., 1997; Thomas & Bain, 1984; Trigwell & Prosser, 1991; Scouller & Prosser, 1994; Scouller, 1998).

Integratie van assessment in het leerproces

Integratie van assessment in het leerproces moedigt de lerende aan gedurende de cursus of training te studeren in plaats van de dag of enkele dagen vóór de toets. Dit stimuleert leren op een meer diepgaand niveau in plaats van het reproductiegericht studeren net voor de toets (Askham, 1997; Dochy & Moerkerke, 1997; Sambell e.a., 1997; Thomson & Falchikov, 1998).

Eén van de kenmerken van een leeromgeving waarbij assessment een integraal deel uitmaakt van het leerproces, is dat een centrale plaats wordt toegekend aan feedback. Gielen e.a. (2003) verwijzen naar verschillende studies die de positieve invloed van feedback op het leerproces aantonen (post-assessmenteffect), wanneer deze tijdig is, op de juiste wijze wordt gegeven en gevraagd, en deze onderdeel is van een leeromgeving die aansluit bij de eerder genoemde kenmerken van leren. In een recente meta-analyse rapporteren Kluger en DeNisi (1996) zowel veld- als laboratoriumonderzoek naar de effectiviteit van feedback in de context van organisaties. Zij geven aan dat er substantiële verschillen zijn in de mate van effectiviteit van feedbackinterventies die gericht zijn op de performance en het leren in organisaties. In het algemeen zijn de resultaten positief, maar voor een derde van de studies verslechterde de performance en het leren in de organisatie na de feedbackinterventies. Kluger en DeNisi (1996) geven aan dat onvoldoende onderzoek is verricht op basis waarvan het mogelijk is deze resultaten te interpreteren.

De verantwoordelijkheid van de student

Studies naar het gebruik van self- en peerassessment tonen de effecten aan van het gebruik van deze vormen van assessment op verschillende aspecten van het leren. Hierop wordt verder ingegaan in de volgende paragraaf.

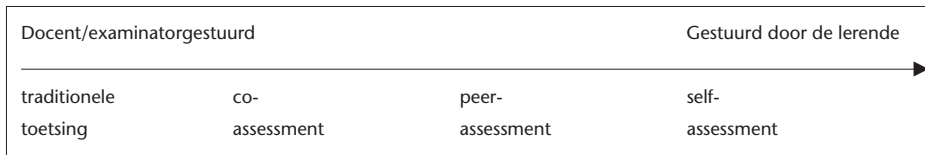
SELF- EN PEERASSESSMENT

Centraal bij self- en peerassessment is de betrokkenheid van de lerende bij de beoordeling. Deze vormen van assessment worden in toenemende mate geïmplementeerd in schoolse contexten. Ze zijn ook populair in de context van opleidingen en professionele ontwikkeling. Voorbeelden van hun toepassingsveld zijn: als onderdeel van portfolio assessment in Continuing Professional Development programma's (bijvoorbeeld Weindling, 2001), in professionele training programma's (Smith & Tillema, 1998, 2001; Tillema, 1998, 2001, 2003), in Nederland bij de EVC (Erkennen van Verworven Competenties) procedure, en bij de 360-graden competentiebeoordelingsprocedure.

Self-, peer- en co-assessment maken deel uit van het veranderingsproces naar een meer studentgecentreerde benadering (Somervell, 1993). Deze verschuiving eist een verandering van nadruk op de norm naar nadruk op het criterium van beoordeling, van de puur summatieve naar de formatieve en summatieve beoordeling, van externe naar interne beoordeling en van product- naar product-én-procesbeoordeling.

Kenmerken van self- en peerassessment

Traditionele vormen van assessment leggen de verantwoordelijkheid voor het assessment meestal in handen van de docent die de criteria opstelt, producten van de lerenden beoordeelt en feedback geeft: docentgestuurd assessment. Aan de andere kant van het continuüm staat de student die zijn eigen doelen stelt en zijn werk zelf beoordeelt: studentgestuurd assessment. Deze twee 'modellen' representeren de extremen van een continuüm dat de graad van autonomie van de student voor zijn leren weergeeft (figuur 2).



Figuur 2 Een continuüm van docentgestuurd assessment naar assessment gestuurd door de lerende (Dochy & Segers, 1999).

Bij selfassessment hebben de lerenden een organisatorische, sturende en inhoudelijke rol. Dit betekent dat ze (mede-) bepalen hoe, wanneer en op basis waarvan beoordeeld zal worden. Naast en tijdens het beoordelingsproces speelt reflectie een belangrijke rol. Aangenomen wordt dat dit reflectieproces leidt tot een verandering in denken, houding en gedrag (Sommervell, 1993). Selfassessment is in vele gevallen formatief van karakter (Boud, 1995).

Wanneer lerenden elkaar beoordelen, is er sprake van peerassessment (Falchikov, 1995). Bij deze vorm van beoordelen gaat het niet alleen om de uiteindelijke beoordeling, maar vooral om de wijze waarop peerassessment een bijdrage levert aan de kwaliteit van het leerproces. Peerassessment komt veelal voor in onderwijs- of opleidingsituaties waar het werken in een groep een centrale rol speelt. Leren in kleine groepen (6-15) wordt gebruikt als middel om 'dieper leren' aan te moedigen, en om meer autonomie aan de lerenden te geven door een deel van de verantwoordelijkheid voor doceren en leren aan henzelf te geven.

Bij co-assessment ten slotte, hebben de lerenden samen met de docent een rol in het beoordelingsproces. In overleg wordt een selectie van de beoordelingscriteria gemaakt. Bij co-assessment hebben de lerenden de gelegenheid elkaar te beoordelen, maar wordt het uiteindelijke oordeel uitgesproken door de docent (Hall, 1995). Co-assessment heeft dan ook de volgende twee bedoelingen. Op de eerste plaats biedt het de student de gelegenheid de rol van de docent in te nemen (rolwisselend onderwijs). Op de tweede plaats verwerft de student inzicht in het beoordelingsproces, waardoor hij beter in staat is anderen te beoordelen en meer vaardig wordt in het beoordelen van zichzelf (selfassessment). Co-assessment is een dialoog tussen docenten en lerenden. Deze samenwerking kan plaatsvinden op verschillende niveaus en in verschillende fasen. Het kan louter formatief zijn of deel uit maken van het summatieve beoordelingsproces. Daarbij kunnen twee modellen worden onderscheiden: het zwakke model en het sterke model. In het zwakke model liggen de criteria op voorhand vast. Dit wordt duidelijk gemaakt aan de staf. Er zijn dus geen aanpassingen mogelijk, noch door lerenden, noch door docenten. In het sterke model dragen lerenden zelf criteria aan. Indien de lerenden zelf moeilijk criteria kunnen vinden, presenteert de docent een lijst van criteria en laat de lerenden hiervan elk item bediscussiëren, verklaren, argumenteren voor en tegen, voorstellen doen, toevoegingen maken, enzovoort. Indien de lerenden zelf kritisch kunnen denken in termen van criteria, laat de docent de lerenden een lijst met criteria aanleggen. Daarna bediscussiëren de lerenden deze lijsten totdat er consensus bestaat.

Vormen van assessment als self-, peer- en co-assessment worden verondersteld een aantal voordelen te hebben. Op de eerste plaats is er een hoge betrokkenheid van lerenden: wanneer de lerende de resultaten van het assessment en de ontwikkeling ervan mede heeft bepaald, zal hij deze eerder accepteren. Dit draagt er toe bij dat de lerenden de beoordeling als fair percipiëren (Sluijsmans e.a., 1998).

Tevens is het een leermoment voor de lerende, enerzijds door de feedback die de lerende krijgt en anderzijds door het zelf beoordelen. Ook versterken deze vormen van beoordelen de motivatie: de lerende krijgt verantwoordelijkheid, hij ondergaat niet slechts een programma (Somervell, 1993). De studie-ijver wordt bevorderd en uitval wordt voorkomen (Sluijsmans e.a., 1998).

Er is echter ook een nadeel. De lerenden van vandaag zijn vaak (nog) niet gewend kritiek te geven op hun medestudenten en voelen zich nog onervaren bij het beoordelen van anderen. Ze zijn bang om onfair beoordeeld te worden. Dit impliceert dat bij de introductie van deze vormen van toetsing de lerenden voldoende mogelijkheden moeten hebben om te leren kritisch te reflecteren, feedback te geven en met feedback op zichzelf om te kunnen gaan (Williams, 1992).

Samenvattend zijn de kenmerken van self-, peer- en co-assessment:

- het gaat hier niet meer om een eenzijdige beoordeling; de lerenden worden actief bij het beoordelingsproces betrokken;
- beoordelen is een leerproces: na een tijd vertonen de lerenden een authentieke nauwgezetheid en volledigheid in de manier waarop ze zichzelf beoordelen;
- *Meeting criteria improves learning*: lerenden zijn actieve participanten in hun eigen leerproces. Ze presteren beter als ze weten naar welke doelen ze toewerken, wanneer ze de gelegenheid krijgen om de relevantie van bepaalde criteria te onderzoeken en wanneer ze begrijpen hoe hun eigen kunnen vergeleken wordt met een set van gevestigde criteria;
- de ontwikkeling van interpersoonlijke vaardigheden: de lerenden leren via discussies, gesprekken, e.d. hun mening te uiten en voor zichzelf en eventuele medestudenten (bij de combinatie met self- and peerassessment) opkomen;
- de lerenden krijgen inzicht in de groepsdynamiek;
- de lerenden worden voorbereid op hun toekomstig professioneel functioneren waarbij self- en peerassessment in toenemende mate worden gehanteerd.

De effecten van self- en peerassessment op het leren

Het onderzoek naar self- en peerassessment is de afgelopen jaren in een stroomversnelling gekomen. Inmiddels zijn verschillende overzichtsstudies gepubliceerd: Boud en Falchikov (1989), Falchikov en Boud (1989), Brown en Dove (1991), Boud (1995), Dochy e.a. (1999), Falchikov en Goldfinch (2000), Topping (1998, 2003). Vele van deze overzichtsstudies richten zich op de meer traditionele, psychometrische kwaliteiten van self- en peerassessment. Ze presenteren voornamelijk studies naar kwaliteitscriteria zoals interbeoordelaarsbetrouwbaarheid. Daarnaast is een beperkt aantal studies gedaan naar de effecten van self- en peerassessment. Deze worden hier besproken.

Tot slot dient te worden opgemerkt dat, hoewel beide vormen van assessment steeds meer worden geïmplementeerd, tot nog toe nauwelijks tot geen onderzoek werd uitgevoerd in de context van opleidings- en professionele ontwikkelingsprogramma's.

Self-assessment

Een aantal studies onderzocht het effect van self-assessment op de leerprestaties van de lerenden. Een beperkt aantal studies zijn uitgevoerd in het hoger onderwijs. In een onderzoek van Hassmèn e.a. (1996, as cited in Dochy e.a., 1999) leerden 128 vrouwen het correcte antwoord op een specifieke taak door de taak zelf uit te voeren of de uitvoering te observeren. De deelnemers aan het onderzoek legden achteraf een vaardigheidstest of een schriftelijke toets af, met of zonder de self-assessment van de mate waarin ze zeker waren dat hun antwoord correct was. Uit de resultaten bleek dat de studenten die expliciet reflecteerden op hun prestaties (self-assessment) tijdens de verschillende oefensessies voor de toets, betere scores behaalden dan zij die niet reflecteerden. Davis en Rand (1980, as cited in Topping, 2003) vergeleken de prestaties van een klas waar de docent beoordeelde met een klas waar gebruik werd gemaakt van self-assessment. Hoewel in de klas met self-assessment duidelijk sprake was van overwaardering, waren hun uiteindelijke prestaties te vergelijken met deze van de klas met docentbeoordelingen. Echter, met veel minder inspanningen aan docentzijde. Sobral (1997, as cited in Topping, 2003) vergeleek een experimentele groep studenten die zelfgestuurde opdrachten beoordeelden via self-assessment met een controlegroep in een docentgestuurde leer- en assessmentomgeving. De experimentele studenten behaalden een hogere Grade Point Average dan de studenten uit de controlegroep. Gelijkwaardige resultaten werden gevonden in onderzoeksstudies uitgevoerd op middelbare scholen en basisscholen. Topping (2003) verwijst onder meer naar de studies van Sink e.a. (1991), Fontana en Fernandes (1994) en Fernandes en Fontana (1996). Sink e.a. (1991) vonden bij middelbare scholieren dat het betrokken zijn bij self-assessment leidde tot hogere prestaties op toetsen. Fontana en Fernandes (1994) en Fernandes en Fontana (1996) onderzochten het effect van self-assessment op de prestaties van lagereschoolkinderen op wiskundetoetsen. De experimentele groep presteerde beter dan de controlegroep.

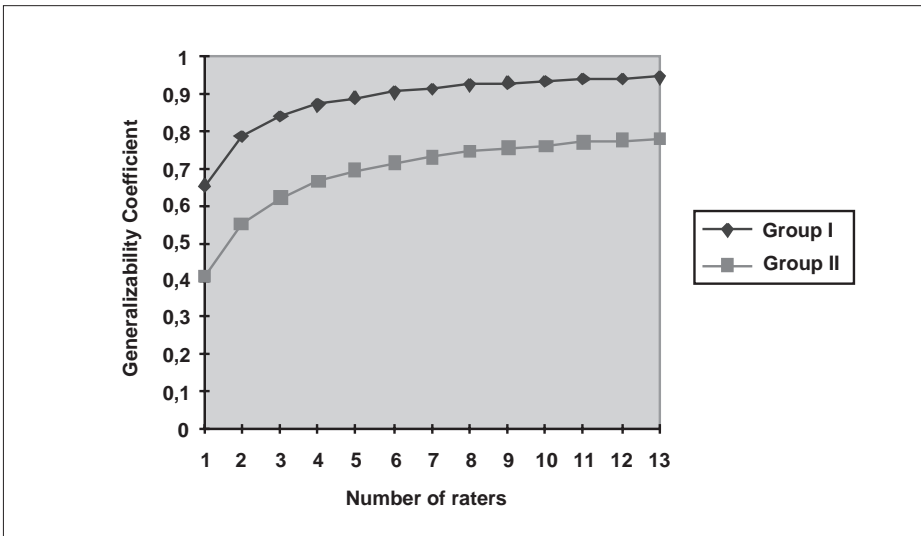
Topping (2003) verwijst ten slotte naar één studie uitgevoerd in het volwassenenonderwijs waarbij aan volwassenen werd gevraagd in welke mate zij dachten dat self-assessment hun denkvaardigheden positief had beïnvloed. Het bleek dat de volwassenen met name op langere termijn een positieve invloed percipieerden.

Naast het effect van self-assessment op prestaties op toetsen, is een aantal onderzoeken gericht op het effect op meta-cognitieve vaardigheden. Ook hier zijn positieve resultaten aangetoond in de overzichtsstudie van Topping (2003) naar self-assessment in vergelijking met docentgestuurd assessment, dit voor wat betreft management van het eigen leerproces, self-efficacy, diepe leerstrategieën in de plaats van oppervlakkige en self-assessmentvaardigheden zoals kritisch reflectievermogen.

Peerassessment

Segers en Dochy (2001) onderzochten de kwaliteit van peerassessment op basis van drie edumetrische kwaliteitscriteria: de generaliseerbaarheid van de assessment scores, de gepercipieerde effecten van peerassessment, en de mate waarin studenten deze vorm van assessment als fair beoordelen. In deze studie beoordeelden 27 derdejaarsstudenten (twee onderwijsgroepen) die participeerden in een probleemgestuurde onderwijskundemodule, zichzelf en elkaar over de mate waarin ze hadden bijgedragen aan

het coöperatief leerproces in de onderwijsgroep. De individuele peerassessmentscore werd gebruikt om de groepsscore voor het groepsproduct (een projectverslag) om te zetten in een individueel cijfer. Uit de *generalisability* studie bleek dat voor de eerste onderwijsgroep reeds op basis van de beoordeling door één student de generaliseerbaarheid van de scores acceptabel was (.653). Voor de tweede onderwijsgroep waren minimaal vier *raters* noodzakelijk om, op basis van de score van de student, generaliseerbare uitspraken te kunnen doen over zijn participatie in de onderwijsgroep. In de studie was de peerscore gebaseerd op de waardering door 12-14 studenten, wat in groep 1 leidde tot een generaliseerbaarheidscoëfficiënt van .90 en in groep 2 .75 (figuur 3).



Figuur 3 Generaliseerbaarheidscoëfficiënten voor beide onderwijsgroepen.

Uit de resultaten van de studentensurvey bleek dat de studenten percipieerden dat self-en peerassessment diep leergedrag (*deep-level thinking and learning*), kritisch denken, en gestructureerd werken in de onderwijsgroep stimuleren. De studenten gaven echter aan zich niet erg gemakkelijk te voelen bij deze methode van assessment. Dit was voor een deel te wijten aan het feit dat dit de eerste ervaring van de studenten was met deze assessmentmethode. Bovendien bleek uit de studentevaluatie van de leeromgeving dat de studenten oordeelden dat de docent te weinig feedback gaf op het functioneren in de onderwijsgroep. Deze bevindingen, en met name de rol van ervaring met deze vorm van assessment, sluiten aan bij de resultaten van verschillende studies die zijn besproken in de overzichtsstudie van Dochy e.a. (1999).

Topping (2003) verwijst naar verschillende studies die het gebruik van peerassessment in het schrijfvaardigheidsonderwijs onderzochten, zowel in het lager, middelbaar als het hoger onderwijs. Ze tonen alle een duidelijke verbetering in schrijfvaardigheden aan na

het peerassessment en een verhoogde interesse in schrijven. Dit blijkt zowel in het geval van individueel peerassessment als bij *peer response groups*. Ook bij het beoordelen van communicatieve vaardigheden en presentatievaardigheden en bij het beoordelen van beroepsvaardigheden (bijvoorbeeld in lerarenopleidingen) blijken studenten beter te presteren op latere beoordelingsmomenten. Het zelfvertrouwen is gegroeid en ze hebben een positief gevoel bij deze methode van assessment. Deze resultaten sluiten aan bij onderzoek van Ross (1995) in een coöperatieve leeromgeving, waarbij leerlingen uit de zevende graad het coöperatief gedrag van hun medeleerlingen moesten beoordelen. Uit de analyses van de audiotapes van het werken in de groepen ná peerassessment bleek een verhoging van zowel het aantal, als de kwaliteit van vragen naar hulp en van de geboden hulp. Bovendien bleken de leerlingen positiever te staan tegenover het vragen van hulp. Tot slot refereert Topping (2003) aan verschillende studies die een positief effect van peerassessment op de scores op summatieve toetsen indiceren.

Weinig onderzoek is uitgevoerd naar het gebruik van peerassessment in de context van opleidingen en professionele ontwikkeling. Een aantal aan elkaar gerelateerde onderzoekslijnen, voornamelijk gepubliceerd in de organisatiekundige literatuur, zijn in dit verband interessant. Het betreft onderzoek naar de rol van peerfeedback (bijvoorbeeld Kuchinke, 2000) en naar de rol van verschillende factoren bij het gebruik van peerassessment in de context van Multisource Assessment of 360-gradenbeoordeling (bijvoorbeeld Brutus & Derayah, 2002). In beide onderzoekslijnen wordt peerassessment, hoewel niet op deze wijze benoemd, als één van de beoordelingsinstrumenten gehanteerd. Het onderzoek van Kuchinke (2000) indiceert dat in leergerichte en studentgestuurde programma's meer feedback wordt gevraagd dan in prestatiegerichte en docentgestuurde programma's. In de context van Multisource Assessment (Brutus & Derayah, 2002) wordt gewezen op het belang van vertrouwen in het assessmentproces, de rol van transparantie van het proces, en ervaring met deze vorm van assessment (bijvoorbeeld via training).

In beide onderzoekslijnen is er nauwelijks tot geen onderzoek gedaan naar de effecten van peerassessment op het leren van de deelnemers aan de programma's.

Samenvattend

Dochy e.a. (1999) onderscheiden op basis van de resultaten van hun literatuurstudie naar de effecten van self-, peer- en co-assessment, acht positieve effecten:

1. verhoogd vertrouwen in eigen kunnen bij de studenten;
2. verbeterd inzicht in de kwaliteiten van eigen prestaties;
3. een betere reflectie op eigen gedrag en prestaties;
4. een hogere leerwinst en, hiermee samenhangend, een hogere score op toetsen;
5. wat betreft gehanteerde leerstrategieën: verhoogd gebruik van dieptestrategieën;
6. het nemen van verantwoordelijkheid voor het leerproces, toenemende mate van zelfsturing;
7. meer tevredenheid van de studenten met betrekking tot het onderwijsleerproces;
8. een verbeterd leerklimaat.

CASE-BASED ASSESSMENT

Uit onderzoek van Field (2001) blijkt dat werkgevers probleemoplossende vaardigheden als één van de drie meest gewaardeerde kerncompetenties omschrijven. Het belang van deze competentie wordt internationaal erkend, zowel in het bedrijfsleven als door de overheid. Bijvoorbeeld in de eerder genoemde Britse National Skills Task Force (1999) wordt probleemoplossend vermogen als één van de zes kerncompetenties genoemd. Hoewel deze competentie algemeen als relevant wordt beschouwd in de context van het huidige maatschappelijk en economisch bestel, is er tot nog toe bijzonder weinig aandacht besteed aan assessment van probleemoplossende vaardigheden. Het aanpakken van nieuwe probleemsituaties is een complex geheel van cognitieve en metacognitieve processen gaande van (1) probleemherkenning, (2) -definiëring, (3) analyse en synthese van probleemkenmerken en oplossingsstrategieën, (4) probleemoplossing (uitvoeren van oplossingsstrategieën), naar (5) evaluatie. Een vorm van assessment om deze competentie te meten is case-based assessment. Beoogd wordt te meten, aan de hand van authentieke casus, in welke mate de lerende in staat is problemen te herkennen, definiëren, analyseren, synthetiseren, bij te dragen aan de oplossing ervan en de aangedragen oplossingen te evalueren.

In Nederland vinden we een voorbeeld van case-based assessment bij het LSOP Politie Onderwijs- en Kenniscentrum dat in 2002 is gestart met vernieuwd politieonderwijs. Hierbij werd ook de wijze waarop de bekwaamheid van politiestudenten werd beoordeeld, veranderd (Bolhuis, 2003). Er worden vier verschillende vormen van assessment gehanteerd: de arbeidsproef, een simulatie, een pen- en papertoets en ten slotte een authentieke opdracht. Deze laatste betreft complexe en van de praktijk afgeleide probleemsituaties, waarbij de studenten activiteiten ontplooiën, resulterend in producten. Een voorbeeld is een complexe casus rondom een calamiteit of ramp. De student wordt gevraagd een plan van aanpak te schrijven voor de betreffende specifieke situatie.

Curtis en Denton (2003) voerden ten behoeve van de Australian National Training Authority het Performance-based Assessment of Problem-Solving Project uit. Centraal in het project stond de ontwikkeling en validatie van een problem-solving assessmentinstrument. Het werd uitgevoerd in de Electronics and Information Technology Program aan het Torrens Valley Institute of TAFE. De studenten waren zowel jonge afstudeerders met slechts een beperkte werkervaring, als volwassenen met vele jaren werkervaring in het betreffende domein. Centraal in het assessment staat dat de studenten drie probleemtaken kiezen uit de modules die ze hebben gevolgd. Zowel de betreffende studenten zelf als de staf geven aan, via een *problem-solving inventory*, in welke mate ze *problem-solving skills* bezitten. Hun uitwerkingen van de probleemtaken dienen als bewijsmateriaal. Het onderzoek van Curtis en Denton (2003) richtte zich voornamelijk op de klassieke psychometrische kwaliteitsindicatoren. Daaruit bleek een voldoende interne consistentie en betrouwbaarheid. Wat betreft het effect van dit instrument op het leren van de studenten, indiceren de resultaten van de studentevaluaties dat de studenten, door het gebruik van het assessmentinstrument, zich beter in staat achtten de verschillende stappen in het probleemoplossingproces expliciet te benoemen en te onderbouwen. De studenten gaven ook aan dat de transparantie van het assessment wat betreft criteria kon worden verbeterd en dat deze vorm van assessment beter geïntegreerd diende te worden in het gehele assessmentsysteem.

De OverAll Toets

Aan de Faculteit der Economische Wetenschappen en Bedrijfskunde, Universiteit Maastricht, werd in 1991 de OverAll Toets ontwikkeld. Deze toets richt zich op het meten van de mate waarin de studenten in staat zijn authentieke probleemsituaties aan te pakken op basis van hun kennis van het betreffende domein (Segers, 1997).

De OverAll Toets meet het volgende: in welke mate kan de individuele student:

- een voor hem nieuw probleem of een aspect van een probleemsituatie definiëren
Kan hij het probleem of probleemaspect duiden op basis van de eerder bestudeerde relevante begrippen, modellen of theorieën;
- een voor hem nieuw probleem analyseren, zoals de relevante elementen onderscheiden of bepalen welke contextvariabelen welke rol spelen;
- de resultaten van de analyse synthetiseren;
- mogelijke oplossingen of te nemen beslissingen beargumenteren;
- oplossingen of beslissingen evalueren

De Overall Toets is ontwikkeld in een probleemgestuurd curriculum en intussen in verschillende zowel probleemgestuurde als projectgestuurde leeromgevingen geïmplementeerd. Enkele kenmerken van beide leeromgevingen zijn:

1. centraal in de opleiding staat het doel studenten op te leiden die in staat zijn reële problemen te definiëren, te analyseren, te synthetiseren en bij te dragen tot de oplossing ervan;
2. in het leerproces wordt uitgegaan van een probleemsituatie die studenten in een groep zelfstandig moeten aanpakken en oplossen. Daarbij wordt gebruikgemaakt van een systematische aanpak;
3. de probleemtaken zijn een middel om kennis te verwerven (probleemgestuurd onderwijs) en te gebruiken of toe te passen (probleemgestuurd onderwijs en projectonderwijs). Dit betekent dat de studenten door middel van het aanpakken van problemen de inhoud van de relevante disciplines, zowel wat betreft kennis als vaardigheden, verwerven en leren toepassen;
4. er wordt een variëteit aan problemen aangeboden aan de studenten. Steeds weer moeten ze hetzelfde theoretisch kader en dezelfde vaardigheden gebruiken om telkens nieuwe problemen aan te pakken. Op deze wijze wordt transfer van kennis en vaardigheden bewerkstelligd.

De OverAll Toets heeft een aantal specifieke kenmerken zowel wat betreft de toetsvorm als wat betreft de organisatievorm:

- elke toetsvraag is gebaseerd op en refereert aan authentieke probleemsituaties. De probleemsituaties zijn verschillend van aard, gaande van de rapportage van een wetenschappelijk onderzoek tot bijvoorbeeld het jaarverslag van een bedrijf;
- omwille van de generaliseerbaarheid is de toets gebaseerd op een set van probleemsituaties;
- de probleemsituaties worden in hun authentieke vorm aangeboden (dus niet bewerkt voor onderwijsdoeleinden);
- de probleemsituaties zijn niet bestudeerd tijdens het onderwijs;
- authenticiteit impliceert voor vele probleemsituaties dat ze multidisciplinair zijn;

- de probleemsituaties zijn vóór de toetsafname ter beschikking van de studenten zodat ze op voorhand kunnen worden bestudeerd;
- de toetsvragen richten zich op kernaspecten van de probleemsituaties;
- zowel de gesloten- als de openvraagvorm wordt gebruikt waarbij beide vraagvormen zich richten op een specifiek beheersingsniveau;
- de toets heeft een openboek karakter;
- de toets wordt door een multidisciplinair team geconstrueerd.

In figuur 4 is een voorbeeld gepresenteerd van twee OverAll Toetsvragen, gebruikt in het vierde jaar van de opleiding Onderwijskunde (Universiteit Leuven). De studenten hadden geparticipeerd in een module met als thema 'assessment'. De module was probleemgestuurd opgezet. Vóór de toetsafname was de casus 'Nederland, een multiculturele samenleving' (Pans & Gulikers, 1999) beschikbaar voor de studenten (naast vier andere casussen). De casus betreft een vaardigheidstraining voor 4-vwo, ontwikkeld rondom het thema van Nederland als multiculturele samenleving. De deelnemende vakken zijn Aardrijkskunde, Geschiedenis en Economie. Het casusmateriaal bestaat uit een handleiding voor docenten en een studentenhandleiding. In de docentenhandleiding worden de doelen en de werkvormen voor de betreffende training stapsgewijs toegelicht. De studentenhandleiding bevat eveneens de doelen en de uit te voeren leeractiviteiten. In de docenten- en studentenhandleiding is het assessment niet uitgewerkt. De OverAll Toetsvragen bij deze casus betreffen dan ook dit aspect.

CASUS 'NEDERLAND, EEN MULTICULTURELE SAMENLEVING'

Vragen 1 en 2 betreffen de casus 'Nederland, een multiculturele samenleving'.

Vraag 1

Onderstaande vragen betreffen de vraag 'waarom toetsen?'.

Als docent bent u verantwoordelijk voor de planning van het assessment in het project 'Nederland, een multiculturele samenleving'.

1a. Noem twee functies van assessment die relevant zijn voor dit project.

1b. Waarom zijn deze functies relevant in deze specifieke casus?

Vraag 2

Ten aanzien van de beoordeling wordt in de conclusie van dit project gesteld 'dat er onvoldoende werd gecommuniceerd ... op welke wijze getoetst of beoordeeld wordt.'

Onderstaande vragen betreffen dit punt, namelijk de vraag 'hoe toetsen?'.

2a. Beschrijf twee assessmentmethoden die geschikt zijn in deze casus.

2b. Geef bij elk van de door u gekozen assessmentmethoden twee (van elkaar onderscheiden!) argumenten waarom deze geschikt zijn in deze casus en licht toe.

Figuur 4 Een voorbeeld van een OverAll Toetsvraag.

De toetsvragen zijn authentiek omdat in de beroepspraktijk van een onderwijskundige, in toenemende mate, advies wordt gevraagd over de inrichting van assessment. Van de onderwijskundige wordt verwacht dat hij, gegeven een bepaalde onderwijskundige context, in staat is te beargumenteren hoe in de specifieke situatie aan een assessment vorm kan worden gegeven.

De kwaliteiten van de OverAll Toets onderzocht

Het onderzoek naar de kwaliteiten van de OverAll Toets richtte zich op verschillende validiteitaspecten: generaliseerbaarheid, constructvaliditeit, curriculum- en instructionele validiteit, criteriumvaliditeit en consequentiële validiteit. Ik zal me in dit artikel beperken tot het onderzoek naar de consequentiële validiteit van de OverAll Toets en enkel samenvattend de resultaten van de andere validiteitsonderzoeken weergeven.

Drie validiteitstudies

In 1991 startte een onderzoeksproject met als doel de verschillende kwaliteitsaspecten van de OverAll Toets te meten vanuit edumetrisch perspectief. Data werden verzameld over de generaliseerbaarheid van de OverAll Toetsscores, de constructvaliditeit, de inhoudsvaliditeit en de criteriumvaliditeit (Segers, 1997; Segers e.a., 1999). De resultaten van deze onderzoeken indiceerden:

1. Betreffende de generaliseerbaarheid van de scores bleek dat de meeste variantie in de toetsscores werd veroorzaakt door het interactie-effect van personen en casus (35% voor het essay-toetsgedeelte en 65% voor het mc-gedeelte). Dit impliceert dat studenten die beter presteren op één casus niet noodzakelijk degenen zijn die goed presteren op een andere casus. Met andere woorden: één casus heeft een lage predictieve waarde voor een andere casus. De resultaten van de *decision-study* indiceerden dat een OverAll Toets uit twaalf casussen moet bestaan om een acceptabele generaliseerbaarheidscoëfficiënt te bereiken (0.67) (Segers e.a., 1999).
2. De OverAll Toets heeft een acceptabele constructvaliditeit. Dit betekent dat de studenten die over een goed georganiseerd kennisbestand beschikken, zoals gemeten door twee verschillende instrumenten (een kennistoets en een conceptsorting-instrument), beter presteren op de OverAll Toets dan studenten met een zwak georganiseerd kennisbestand (Segers e.a., 1999).
3. Er is een belangrijke mate van overlap tussen het formele curriculum, het feitelijke curriculum en de inhoud van de OverAll Toets. Dit gegeven indiceert dat de OverAll Toets een hoge inhoudsvaliditeit heeft (Segers, 1997).
4. De OverAll Toets heeft een hoge criteriumvaliditeit: hij discrimineert tussen de studenten die goede en zwakke probleemoplossers blijken op basis van het criterium-instrument (Segers, 1997).

De consequentiële validiteit van de OverAll Toets (Segers & Dochy, 2001)

Centraal staat de vraag: wat is het effect van de implementatie van de OverAll toets op het leergedrag van de studenten en het doceergedrag van de docenten? Met de implementatie van de OverAll Toets in het probleemgestuurd curriculum werd gepoogd een additionele stimulans in te bouwen voor studenten om zich niet te richten op memorisatie van informatie, maar op het aanpakken van authentieke problemen waarbij kennis

een instrument is. Tegelijk werd verondersteld dat de OverAll Toets ook het gedrag van de tutores positief zou beïnvloeden, namelijk dat zij meer aandacht in de onderwijsgroepen zouden hebben voor het probleemoplossingproces in plaats van voor memorisatie van informatie.

De consequentiële validiteit van de OverAll Toets werd de voorbije jaren in verschillende onderzoeksstudies gemeten. Sinds de invoering van de OverAll Toets werd jaarlijks een evaluatievragenlijst voorgelegd aan alle studenten. De vragen betreffen de mate waarin de studenten de doelen van de OverAll Toets herkennen; de helderheid, begrijpbaarheid van de casus, en de mate waarin ze aansluiten bij de thematieken van het leerproces tijdens het gevolgde onderwijs; de studiestrategieën bij het bestuderen van de casus en ten slotte de helderheid en moeilijkheidsgraad van de toetsvragen. Uit de resultaten van de vragenlijst bleek dat er vrij veel verschillen waren in de wijze waarop de studenten aan de casus werkten en in de hoeveelheid tijd die ze eraan spendeerden. Hoewel de studenten twee weken onderwijsvrij waren, besteedden ze gemiddeld slechts 27 uren aan de analyse van de casus. Daarbij bleek dat vele studenten slechts een oppervlakkige analyse uitvoerden: de vaak omvangrijke casus werd voornamelijk gelezen, de belangrijkste begrippen werden opgezocht en soms werden de kernbegrippen aangeduid. Opvallend was dat, hoewel uit een eerdere validiteitstudie een hoge inhoudsvaliditeit bleek, de studenten aanduidden dat de wijze van werken in de onderwijsgroepen niet goed aansloot op de wijze waarop ze werden bevraagd in de OverAll Toets. Vooral dit laatste aspect was verrassend. In de onderwijsgroepen staat het aanpakken van problemen centraal. Verondersteld werd dat de transfer van het werken aan problemen in de onderwijsgroep naar het bestuderen van nieuwe probleemsituaties bij de OverAll Toets, eenvoudig zou zijn.

Deze resultaten, die consistent bleken in de verschillende academiejaren waarin de vragenlijst werd afgenomen, waren de directe aanleiding voor de interviews. Er werden semi-gestructureerde interviews gehouden met groepen van studenten en, omwille van between-respondertriangulatie, met groepen van tutores.

De vragen betroffen het karakter van de OverAll Toets, de aansluiting tussen het onderwijsleerproces en de OverAll Toets, en de studiestrategieën van de studenten.

Uit de interviews kwamen vier thema's naar voren:

- Het concept van de OverAll Toets: de studenten percipieerden het doel van de OverAll Toets als het meten van de toepassing van kennis. Zoals Sebastian zei: 'In de OverAll Toets moet je kennis in de praktijk gebruiken'. Thomas expliciteerde het als volgt: 'De Overall Toets vraagt je kennis te gebruiken; je moet meer doen dan in een kennistoets. Voor de kennistoets lees je de tekstboeken en memoriseer je ze. Voor de OverAll Toets moet je zaken aan elkaar relateren, je moet omgaan met de context waarin de kennis wordt gebruikt. De OverAll toets gaat over de constructie van kennis, de kennistoets gaat over memoriseren'. Stephanie, een tutor, gebruikte het begrip 'het linken van kennis aan de praktijk'.
- De aansluiting van de toets bij het onderwijs: zowel de studenten als de tutores gaven aan dat de transfer van het probleemoplossingproces in de onderwijsgroepen naar het oplossen van de probleemsituaties gepresenteerd in de casus, een natuur-

lijke stap zou kunnen/moeten zijn. Echter, in de praktijk beperkte het werken in de onderwijsgroepen zich vaak tot het herhalen van de informatie die was gelezen in tekstboeken naar aanleiding van de initiële analyse van het probleem (die ook vaak zeer oppervlakkig gebeurde). Het te bestuderen probleem was dus veelal enkel een vertrekpunt voor de studie van de literatuur, waarbij deze laatste het doel werd in plaats van het middel. Als redenen hiervoor werden zowel door de staf als door de docenten gewezen op de vaardigheden (of het gebrek aan) van de tutor in het begeleiden van de probleemanalyse van de studenten; de motivatie (of het gebrek aan) van de studenten en de te grote hoeveelheid leerinhouden per module. Kurt vatte dit laatste punt als volgt samen: "Hoe kan je tijd vinden voor het analyseren van problemen wanneer je in zes weken 19 hoofdstukken moet verwerken? Dit is het punt waarop het probleemgestuurd onderwijs faalt!" Hij voegde hieraan toe: "In de onderwijsgroepen waren de grafieken minder complex en reeds getekend in de boeken die we gebruikten. Wat je moest doen was ze interpreteren. In de OverAll Toets moest je de grafieken zelf tekenen!" Hoewel een structureel onderdeel van de procedure in de onderwijsgroep, bleek daarbij de toepassing van de nieuwe verworven kennis op het uitgangsprobleem en op nieuwe gelijkaardige problemen, nauwelijks te gebeuren.

- De studenten wezen eveneens op het gebrek aan feedback in de onderwijsgroep. Omdat vele onderwijsgroepen niet verder gaan dan het opsommen van de informatie die ze in de literatuur hebben gevonden, vindt het eigenlijke leerproces waarbij discussie en feedback centraal staan, niet plaats. Een door de moduleleider expliciet ingebouwd feedbackmoment in het midden van de module werd als zinvol maar slecht gepland gepercipieerd. Het was te vroeg gepland in de module waardoor het niet 'primair relevant' was voor de studenten. Studenten gaven aan dat het voortdurend geven van feedback op het probleemoplossingsproces een structureel onderdeel van het onderwijsleerproces in alle modules zou moeten zijn.
- De studenten gaven aan dat, bij de voorbereiding op de OverAll Toets, velen onder hen zich beperkten tot het lezen van de casus. Sommige studenten vormden een groepje en gingen samen aan het werk. Dit werd als zeer effectief ervaren: "Het dwingt je kritisch om te gaan met wat je leest", zei David. Als reden voor het oppervlakkig bestuderen van de casus, gaven de studenten aan niet te weten hoe ze moesten omgaan met de vaak uitgebreide casuïstiek. Dit verwijst opnieuw naar het zwak functioneren van de onderwijsgroepen waarbij het probleemoplossingsproces wordt overschaduwed door het samenvatten van de informatie uit de literatuur.

Tot slot gaven zowel de studenten als de staf aan dat de OverAll Toets een essentieel onderdeel is van leeromgevingen zoals het probleemgestuurd onderwijs. Het knelpunt is niet de toetsing maar de praktijk van het probleemgestuurd onderwijs.

In dit verband wijst onderzoek van Nijhuis e.a. (2004) op de rol van de percepties als een filter tussen de leeromgeving en de studiestrategieën. In een module Internationale Bedrijfskunde, waarbij de studenten werkten aan en kennis verwierven door het werken aan opdrachten (*assignment-based learning*), bestond het assessment uit een kennis-toets, die de reproductie van, en het inzicht in begrippen en modellen op het terrein

van internationale bedrijfskunde mat (originele ABL-omgeving). In de betreffende module werd vervolgens de aanpak geoptimaliseerd door het gebruik van meer authentieke (dus minder gestructureerde) probleemtaken en het hanteren van meerdere informatiebronnen voor de analyse en oplossing van de probleemtaken (in de plaats van één handboek). Ook, en in overeenstemming met de wijzigingen in de leeromgeving, werd naast de kennistoets een OverAll Toets geïmplementeerd. Deze OverAll Toets meet, aan de hand van authentieke casus, de mate waarin de studenten in staat zijn kennis te hanteren als instrument om problemen aan te pakken (PBL-omgeving). In beide condities (leeromgevingen) werd de studenten (ABL: $n=362$; PBL: $n=248$) gevraagd naar hun verwachtingen van de toetsing. Hierbij werd gebruikgemaakt van de Scouller en Prosser (1994) Assessment Expectations vragenlijst. Daarnaast werden door middel van de Biggs Study Processes Questionnaire de studiestrategieën van de studenten gemeten. Aan het begin van de cursus werd gevraagd naar de studiestrategieën die ze verwachtten te hanteren; aan het einde van de module werd bevestigd welke studiestrategieën ze feitelijk hadden gehanteerd.

De resultaten toonden aan dat er geen significante verschillen waren tussen de studenten in beide leeromgevingen voor wat betreft de verwachte studiestrategieën.

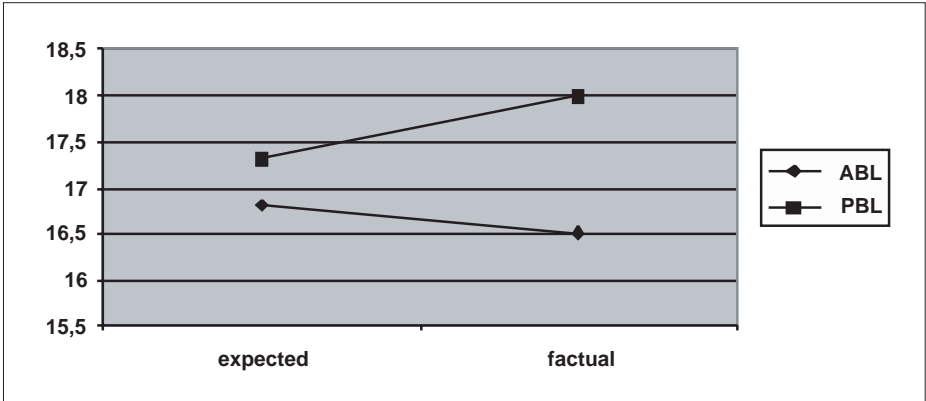
Daarnaast bleek uit de resultaten dat de feitelijke studiestrategieën tussen beide leeromgevingen significant verschilden ($p=0.000$). Dit gold zowel voor de diepte- als oppervlaktestudiestrategieën. De studenten vertoonden meer oppervlakkige studiestrategieën in de PBL-omgeving dan in de ABL-omgeving. Omgekeerd vertoonden ze meer dieptestrategieën in de ABL course dan in de PBL course.

Wat betreft de verwachte studiestrategieën bleek dat er geen significante verschillen waren tussen beide leeromgevingen (verwachte dieptestrategieën: $p=0.282$; verwachte oppervlaktestrategieën: $p=0.075$).

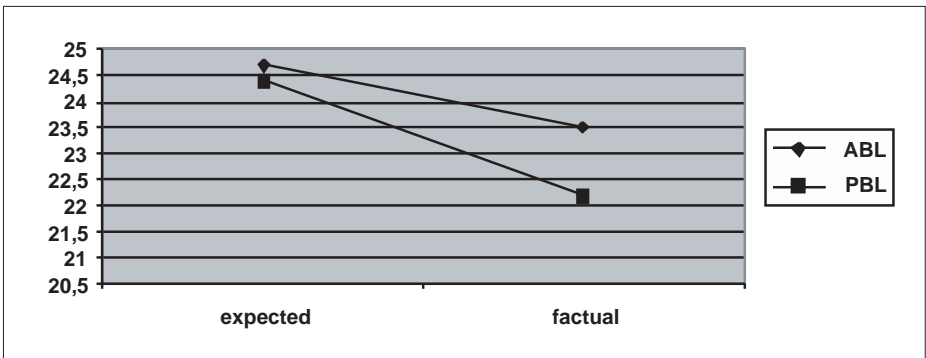
Wat betreft de verschillen tussen verwachtingen en feitelijk gehanteerde oppervlaktestrategieën, bleek uit de *paired sample test* dat in de ABL-omgeving de feitelijk gehanteerde studiestrategieën niet significant verschilden van de verwachte studiestrategieën. In de PBL-omgeving daarentegen werd significant meer oppervlaktestudiegedrag vertoond dan verwacht (*t-test*, $p = 0.001$) (figuur 5).

Tot slot bleek dat in beide leeromgevingen de studenten minder dieptestrategieën hanteerden dan ze hadden verwacht (figuur 6). Uit de *paired sample test* bleek, dat voor de PBL-omgeving het verschil tussen de verwachtingen van de studenten en de feitelijk gehanteerde dieptestrategieën significant was, dit in tegenstelling tot de ABL-omgeving. In de PBL-omgeving werden significant minder dieptestrategieën gehanteerd dan verwacht.

Nader onderzoek indiceerde een aantal verklaringen voor de verschillen tussen beide omgevingen in gehanteerde studiestrategieën. Daartoe werden, naast de studiestrategieën en de verwachtingen ten aanzien van het assessment, door middel van de Ramsden Course Experiences Questionnaire, de studentpercepties van de leeromgeving gemeten. Tabel 1 presenteert de resultaten van de logistische regressieanalyse, waarbij het gepresenteerde model is gebaseerd op de variabelen gemeten door de Ramsden Course Experiences Questionnaire. Hieruit blijkt dat op basis van dit model 88.7% van de studenten correct geplaatst kan worden in één van beide leeromgevingen (Nagel-



Figuur 5 Verwachte en feitelijk vertoonde oppervlakte studiestrategieën in de originele (ABL) en geoptimaliseerde (PBL) leeromgeving.



Figuur 6 Verwachte en feitelijk vertoonde diepte studiestrategieën in de originele (ABL) en geoptimaliseerde (PBL) leeromgeving.

kerke $R^2 = 0.69$). In dit model spelen 4 variabelen een significante rol ($p < 0.01$): duidelijke doelen (-0.20), geschikte assessment (-0.21), acceptabele studielast (-0.66), en de geschiktheid van het studiemateriaal (-0.31). De studenten ervoeren dat in de geoptimaliseerde PBL-omgeving de doelen minder helder zijn, de studielast te hoog is en het studiemateriaal onduidelijk. Bovendien percipieerden de studenten dat een groter accent werd gelegd op het beoordelen van de mate waarin men in staat is kennis te reproduceren dan op het toepassen van kennis. Op basis van eerder onderzoek (Trigwell & Prosser, 1991) kan worden verondersteld dat deze negatieve percepties hebben geleid tot het oppervlakkiger studiegedrag.

Tabel 1 Resultaten van de logistische regressieanalyse op basis van de variabelen gemeten door de Ramsen Course Experiences Questionnaire.

Variable	Coeff	Model 1 t-stat
CONSTANT	8.82	0.000***
Good teaching	0.08	0.131
Clear goals	-0.20	0.005***
Generic skills	0.11	0.062*
Appropriate assessment	-0.21	0.009***
Appropriate workload	-0.66	0.000***
Independent study	0.10	0.068*
Group work	0.03	0.623
Usefulness of the textbook	-0.31	0.000***
Chi-square (df)	326.92 (8)	0.000
% correct predicted	88.7	
Nagelkerke R2	0.69	

CONCLUSIE: EEN TOEKOMSTIGE ASSESSMENTONDERZOEKSAGENDA

Uitgaande van de verdere uitbouw van competentiegerichte opleidingen waarbij de eigen verantwoordelijkheid van de lerende centraal staat, is er in toenemende mate behoefte aan een aangepast assessment van de mate waarin en de wijze waarop de lerenden de beoogde competenties verwerven. Dit geldt niet alleen voor de schoolse context, waar in toenemende mate onderzoek wordt verricht naar verschillende aspecten van assessment. Echter, ook in de praktijk van formele en niet-formele *on-the-job* leermogelijkheden en *off-the-job* opleidingen wordt steeds meer gebruikgemaakt van beoordelingen door de lerende zelf of door collegae. Dit gebeurt bij opleidingen en professionele ontwikkelingsprogramma's bijvoorbeeld in de context van 360-gradenbeoordelingen en Continuous Professional Development Programs met portfolio's. In de context van de Proeven van Bekwaamheid bij professionele opleidingen klinkt steeds meer de wens om de traditionele kennistoetsen te verlaten en te zoeken naar geschikte assessmentvormen. Naast het gebruik van portfolio's wordt in dit verband onder meer gezocht naar instrumenten die meten in welke mate de professional in staat is voor zijn beroep authentieke probleemsituaties aan te pakken.

Vanuit onderwijskundig en opleidingskundig perspectief is alsnog weinig aandacht besteed aan de optimalisatie van assessment in deze context.

Voor de verdere ontwikkeling van assessment in de context van opleidingen en professionele ontwikkelingsprogramma's zijn drie onderzoekslijnen relevant.

In toenemende mate worden opleidingen en professionele ontwikkelingsprogramma's geïmplementeerd die nauw aansluiten bij de kenmerken van constructivistische leeromgevingen. Tegelijk, en in congruentie met de assessmentcultuur die voornamelijk in de

schoolse context is ontwikkeld, worden de principes van assessment in het volwassenenonderwijs geherdefinieerd (Kasworm & Marienau, 1997). Uitgegaan wordt van de veronderstelling dat deze leer- en assessmentomgevingen tot meer effectief leren zullen leiden. In praktijk worden, als integraal deel van deze leeromgevingen, nieuwe vormen van assessment ingevoerd waarbij self- en peerassessment, en allerlei vormen van case-based assessment een rol spelen. Verwacht wordt dat deze een extra impuls vormen voor effectieve leerprocessen. De vraag is echter of de relatie tussen deze leeromgevingen en de leerprocessen en -opbrengsten wel zo direct is als wordt verondersteld. Uit onderzoek in schoolse settings is gebleken dat een complexe interactie van individuele factoren de leerprestaties van studenten beïnvloeden (bijvoorbeeld Trigwell & Prosser, 1991). Op basis van inzichten in het leren van volwassenen (bijvoorbeeld Knowles, 1980; Kolb, 1984; Merriam & Caffarella, 1991) kan worden verwacht dat dit op meer significante wijze het geval is voor het volwassenenonderwijs. Wanneer volwassenen deelnemen aan opleidings- en ontwikkelingsprogramma's, in welke vorm ook, brengen ze een grote hoeveelheid ervaringen met zich mee. Dit betreft niet alleen inhoudelijke aspecten, maar ook ervaringen met het leer- en instructieproces. Dit is in belangrijke mate het geval voor assessment. Alle participanten in volwasseneneducatie hebben een rijk onderwijsverleden waarin assessment een cruciale rol speelt. In een aantal gevallen betreffen het negatieve ervaringen waarbij assessment wordt geassocieerd met reproductiegerichte, selectieve toetsen. Deze ervaringen kleuren de wijze waarop de volwassenen omgaan met leren en assessment in nieuwe settings (het langetermijn-post-assessmenteffect), meer bepaald de wijze waarop ze met leer- en assessmenttaken omgaan. Echter niet alleen eerdere ervaringen met assessment spelen een rol. Ook de wijze waarop de leeromgeving (inclusief assessment) wordt gepercipieerd of, in het geval van werkplekleren, de mate waarin de werkomgeving als een leeromgeving wordt gepercipieerd, beïnvloedt het leergedrag. Uit onderzoek (bijvoorbeeld Nijhuis e.a., 2004) is gebleken dat constructivistische leeromgevingen, waarbij de participanten de leerdoelen als vaag en de werkdruk als groot percipiëren, eerder leiden tot oppervlakkig leergedrag in plaats van diepgaand leergedrag. Onderzoek naar de mate waarin de interactie van deze variabelen een significante rol speelt in de individuele leeropbrengsten van opleidingen en professionele ontwikkelingsprogramma's, zou een belangrijke bijdrage kunnen leveren aan het optimaliseren van deze programma's.

Een tweede onderzoekslijn betreft onderzoek naar een veel gehanteerde vorm van assessment in opleidingen en professionele ontwikkelingsprogramma's, namelijk self- en peerassessment. Uit eerder onderzoek blijkt dat een aantal factoren een belangrijke rol spelen in de mate waarin peerassessment het leren beïnvloedt: de mate waarin studenten het assessment fair vinden, meer bepaald de mate waarin de peerassessor als objectief, accuraat en bekwaam wordt gepercipieerd; de validiteit van het assessmentproces, meer bepaald de transparantie; de rol van eerdere ervaringen met self- en peerassessment; de door de lerende gepercipieerde afstemming tussen de leeromgeving en de vorm van assessment. Dit onderzoek is voornamelijk uitgevoerd in schoolse settings. Onderzoek in de context van opleidingen en professionele ontwikkelingsprogramma's is schaars, maar wijst in dezelfde richting als onderzoek in schoolse settings. Relevante onderzoeksvragen zijn: in welke mate beïnvloeden (1) de gepercipieerde validiteit van

peerassessment, en (2) het vertrouwen van de beoordeelde in de peerassessors de effecten van peerassessment op het leren, hierbij controlerend voor eerdere ervaringen met peerassessment? In deze context is leren gedefinieerd als de ontwikkeling van professionele competenties.

Een derde onderzoekslijn betreft een uitbreiding van het onderzoek naar assessment van probleemoplossende vaardigheden naar opleidings- en ontwikkelingsprogramma's. Hoewel deze programma's in toenemende mate kerncompetenties als probleemoplossende vaardigheden centraal stellen, zijn er tot nog toe weinig instrumenten ontwikkeld voor het beoordelen er van. De OverAll Toets is ontwikkeld en geïmplementeerd in voornamelijk schoolse settings. Het Authentic Performance-Based Assessment Instrument is ontwikkeld en als pilot onderzocht in professioneel volwassenenonderwijs (Curtis & Denton, 2003). Het verder ontwikkelen en onderzoeken van deze vorm van assessment in andere contexten biedt perspectieven voor de optimalisatie van het assessment van het probleemoplossend vermogen van professionals in opleidings- en ontwikkelingstrajecten. Dit veronderstelt onderzoek op het kruisvlak tussen cognitieve psychologie en opleidingskunde. Cognitief psychologisch onderzoek kan een bijdrage leveren aan het definiëren van het construct van de te beoordelen probleemoplossende vaardigheid. Vanuit onderwijskundig en opleidingskundig onderzoek kan een bijdrage worden geleverd aan de ontwikkeling en validering van het assessmentinstrument ter beoordeling van de betreffende probleemoplossende vaardigheid.

Verschillende onderzoeksvragen zijn hierbij relevant: (1) de fairness van het instrument voor verschillende groepen van participanten (bijvoorbeeld laag- en hogeschoolden); (2) de relevantie van het instrument in verschillende disciplines; (3) de condities waaronder de implementatie van het instrument bijdraagt aan de verdere ontwikkeling van probleemoplossende vaardigheden.

Tot slot: ontwikkelingen op het terrein van leren en assessment hebben een internationaal en interdisciplinair karakter. Vooral wanneer we ons op de markt van professionele opleidings- en ontwikkelingsprogramma's begeven, is een lokale monodisciplinaire onderzoeksagenda een *contradictio in terminis*. Internationale, interdisciplinaire samenwerkingsverbanden kunnen de kwaliteit en de impact van het voorgestelde onderzoek positief beïnvloeden.

LITERATUUR

- Ashton, N. (1994) Employers for Capability: Autonomous Development in IBM. *Capability Volume, 1,2, 1-5*.
- Askham, P. (1997) An instrumental response to the instrumental student: assessment for learning. *Studies in Educational Evaluation, 23, 299-317*.
- Biggs, J. (1999) *Teaching for Quality Learning at University*. Buckingham: SRHE and Open University Press.
- Billett, S. (1996) Situated learning: Bridging sociocultural and cognitive theorising. *Learning and Instruction, 6, 263-280*.

- Birenbaum, M. (2003) New Insights into Learning and Teaching and Their Implications for Assessment. In M. Segers, F. Dochy, & E. Cascallar (eds.). *Optimising New Modes of Assessment: In Search of Qualities and Standards* (pp. 13-37). Dordrecht, Boston, London: Kluwer Academic Publishers.
- Black, P., & William, D. (1998) Assessment and classroom learning. *Assessment in Education*, 5, 7-74.
- Boekaerts, M. (2002) Bringing about change in the classroom: Strengths and weaknesses of the self-regulated learning approach. *Learning and Instruction*, 12, 589-604.
- Boekaerts, M. (2003) Towards a model that integrates motivation, affect, and learning. In L. Smith, C. Rogers, & P. Tomlinson (Eds.). *Development and motivation: joint perspectives* (pp. 173-189). Leicester: British Psychological Society.
- Bolhuis, S. (2003) Towards process-oriented teaching for self-directed lifelong learning: a multidimensional perspective. *Learning and Instruction*, 13, 327-347.
- Boud, D. (1995) *Enhancing Learning through Self-assessment*. Kogan Page. London: Philadelphia.
- Boud, D. & Falchikov, N. (1989) Quantitative studies of self-assessment in higher education: A critical analysis of findings. *Higher Education*, 18, 529-549.
- Boyatzis, R. E., Cowen, S. S., & Kolb, D. A. (1995) *Innovations in professional education: Steps on a journey from teaching to learning*. San Francisco: Jossey-Bass.
- Brown, S., & Dove, P. (1991) *Self and Peerassessment*. Birmingham: Standing Conference on Educational Development (SCED).
- Brutus, S., & Derayeh, M. (2002) Multisource assessment programs in organizations: An insider's perspective. *Human Resource Development Quarterly*, 13, 187-202.
- Clarke, M., Madaus, G. F., Horn, C. J., & Ramos, M. A. (2000) Retrospective on educational testing and assessment in the 20th century. *Journal of Curriculum Studies*, 32, 159-181.
- Curtis, D., & Denton, R. (2003) *The authentic performance-based assessment of problem solving*. Station Arcade: National Centre for Vocational Education Research.
- De Grip, A. (2000) *Van tweedekansonderwijs naar een leven lang leren* (inaugurale rede). Maastricht: Universiteit Maastricht.
- Dierick, S., & Dochy, F. (2001) New lines in edometrics: new forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation*, 27, 307-329.
- Dochy, F., & Moerkerke, G. (1997) Assessment as a major influence on learning and instruction. *International Journal of Educational Research*, 27, 415-432.
- Dochy, F., & Segers, M. (1999) Innovatieve toetsvormen als gevolg van constructiegericht onderwijs: op weg naar een assessment-cultuur (pp. 181-206). In P. De Boeck & M. Lacante (eds.). *Meer kansen creëren voor het hoger onderwijs*. Dordrecht: Kluwer.
- Dochy, F. J. R. C., Segers, M. S. R., & Sluysmans, D. (1999) The use of self-, peer and co assessment in higher education: a literature review. *Studies in Higher Education*, 24, 331-350.
- Dochy, F., Segers, M., Van den Bossche, P., & Gijbels, D. (2003) Effects of problem-based learning: a meta-analysis. *Learning and Instruction*, 13, 533-568.
- Falchikov, N. (1995) Peer feedback marking: Developing peerassessment. *Innovations in Education and Training International*, 32, 175-187.
- Falchikov, N., & Boud, D. (1989) Student self-assessment in higher education: A meta analysis. *Review of Educational Research*, 59, 395-430.

- Falchikov, N., & Goldfinch, J. (2000) Student peerassessment in higher education: A meta analysis comparing peer and teacher marks. *Review of Educational Research*, 70, 287-322.
- Field, L. (2001) *Employability skills required by Australian workplaces*. Sydney: Field learning.
- Frederickson, J. R., & Collins, A. (1989) A system approach to educational testing. *Educational Researcher*, 18, 9, 27-32.
- Gielen, S., Dochy, F., & Dierick, S. (2003) Evaluating the Consequential Validity of New Modes of Assessment: The Influence of Assessment on learning, Including Pre-, Post-, and True Assessment Effects. In M. Segers, F. Dochy & E. Cascallar (eds.). *Optimising New Modes of Assessment: In Search of Qualities and Standards* (pp. 37-54). Dordrecht/Boston/London: Kluwer Academic Publishers.
- Glaser, R. (1990) Toward new models for assessment. *International Journal of Educational Research*, 14, 475-483.
- Haertel, E. H. (1991) New forms of teacher assessment. *Review of Research in Education*, 17, 3-29.
- Hall, K. (1995) *Co-assessment: participation of students with staff in the assessment process. A report of Work in Progress*. 2nd European Electronic Conference On Assessment And Evaluation.
- Harlen, W., & Crick, R.D. (2001) Testing and Motivation for Learning. *Assessment in Education*, 10, 169-207.
- Kasworm, C., & Marienau, C. A. (1997) Principles for assessment of adult learning. *New Directions for Adult and Continuing Education*, 75, 5-16.
- Kerka, S. (1995) *Techniques for Authentic Assessment*. <http://www.cete.org/acve>
- Knowles, M. (1980) *The Modern Practice of Adult Education: From Pedagogy to Andragogy*. New York: Cambridge Books.
- Kluger, A. N., & DeNisi, A. (1996) The effects of feedback interventions on performance: A historical review, a meta-analysis and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.
- Kohn, A. (2000) *The Case Against Standardized Testing*. Portsmouth, NH: Heinemann.
- Kolb, D. (1984) *Experiential Learning: Experience as the Source of learning and Development*. Englewood Cliffs, N.J.: Prentice Hall.
- Koretz, D. (1988) Arriving at Lake Wobegon: are standardized tests exaggerating achievement and distorting instruction? *American Educator*, 12, 2, 8-15.
- Koretz, D. M. (1991) State comparisons using NAEP: Large costs, disappointing benefits. *Educational Researcher*, 20, 3, 19-21.
- Kuchinke, K. P.(2000) The role of feedback in management training settings. *Human Resource Quarterly*, 11, 381-401.
- Lantz, A., & Friedrich, P. (2000) How to evaluate the effects of learning at the workplace; presentation of an instrument for Competence Assessment. http://www.hiceducation.org/Edu_Proceedings/Annika%20Lantz.pdf
- Linn, R. (2000) Assessment and accountability. *Educational Researcher*, 29, 2, 4-16.
- Linn, R. L., Baker, E., & Dunbar, S. B. (1991) Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 16, 1-21.

- Lohman, M. (2002) Cultivating problem-solving skills through problem-based approaches to professional development. *Human Resource Development Quarterly*, 13, 243-261.
- Merriam, S., & Caffarella, R. (1991) *Learning in Adulthood: a Comprehensive Guide*. San Francisco: Jossey Bass.
- McDowell, L. (1995) The impact of innovative assessment on student learning. *IETI*, 32, 302-313.
- National Skills Task Force (1999) *Delivering Skills for All. Second Report*. Sudbury: Department for Education and Employment/Prolog.
- Nijhuis, J., Segers, M., & Gijssels, W. (2004) Redesigning a learning environment: the influences on students' perceptions and their learning strategies. *Learning Environment Research*. (accepted with revisions)
- Pans, R. H. M. C., & Gulikers-Dinjens, M. T. H. (1999) *Nederland, een multiculturele samenleving. Een integratief, multidisciplinair project voor 4 VWO*. Maastricht: Expertisecentrum Actief Leren.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.) (2001). "Knowing what students know: The science and design of educational assessment." Washington, DC: National Academy Press.
- Ramsden, P. (1988) *Improving learning. New perspectives*. London: Kogan page.
- Resnick, L. B., & Nolan, K. L. (1995) Standards for education. In D. Ravitch (ed.). *Debating the Future of American Education: do we need national standards and assessment?* (pp. 168-193). Washington DC: Brookings Institution.
- Ross, J. A. (1995) Effects of feedback on student behavior in cooperative learning groups in a grade-7 math class. *Elementary School Journal*, 96, 125-143.
- Sambell, K., McDowell, L., & Brown, S. (1997) But is it fair?: An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation*, 23, 349-371.
- Scouller, K. (1996) Influence of assessment method on students' learning approaches, perceptions and preferences: The assignment essay versus the short answer examination. *Different Approaches: Theory and Practice in Higher Education*. Proceedings HERDSA Conference 1996. Perth, Western Australia, 8-12 July.
<http://www.herdsa.org.au/confs/1996/scouller.html>
- Scouller, K. (1998) The influence of assessment method on students' learning approaches: Multiple choice question examination versus assignment essay, *Higher Education*, 35, 453-472.
- Scouller, K., & Prosser, M. (1994) Students' Experiences in Studying for Multiple Choice Question Examinations, *Studies in Higher Education*, 19, 267-279.
- Segers, M. S. R. (1997) An alternative for assessing problem-solving skills: the OverAll Test. *Studies in Educational Evaluation*, 23, 373-398.
- Segers, M. (2003) Evaluating the OverAll Test: Looking for Multiple Validity Measures. In: M. Segers, F. Dochy & E. Cascallar (eds.). *Optimising New Modes of Assessment: In Search of Qualities and Standards* (pp. 119-140). Dordrecht/Boston/London: Kluwer Academic Publishers.

- Segers, M., & Dochy, F. (2001) New assessment forms in problem-based learning: The value-added of the students' perspective. *Studies in Higher Education*, 26, 327-343.
- Segers, M., Dochy, F., & De Corte, E. (1999) Assessment practices and students' knowledge profiles in a problem-based curriculum. *Learning Environments Research*, 2, 191-213.
- Segers, M., Van den Bossche, P., & Teunissen, E. (2003) Evaluating the effects of redesigning a problem-based learning environment. *Studies in Educational Evaluation*, 29, 315-334.
- Shohamy, E. (1995) Language Testing: Matching Assessment Procedures with Language Knowledge (pp. 143-161). In M. Birenbaum & F. Dochy (eds.). *Alternatives in assessment of achievements, learning processes and prior knowledge*. Boston/dordrecht/London: Kluwer Academic Publishers.
- Sluijsmans, D., Dochy, F., & Moerkerke, G. (1998) Creating a learning environment by using self-, peer- and co-assessment. *Learning Environments Research*, 1, 293-319.
- Smith, L. R. (2000) *Issues impacting on the quality of assessment in vocational education and training in Queensland*. Queensland: Department of Employment, Training and Industrial Relations.
- Smith, K., & Tillema, H. (1998) Evaluating Portfolio use as a learning tool for professionals. *Scandinavian Journal of Educational Research*, 42, 193-205.
- Smith, K., & Tillema, H. (2001) Long-term influences of portfolios on professional development. *Scandinavian Journal of Educational Research*, 45, 183-203.
- Somervell, H. (1993) Issues in assessment, enterprise and higher education: The case for self-, peer and collaborative assessment. *Assessment and Evaluation in Higher Education*, 18, 221-233.
- Stiggins, R. (2001) *Student-Involved Classroom Assessment*. Upper Saddle River, NJ: Merrill Prentice Hall.
- Stonyer, H., & Marshall, L. (2002) Moving to problem-based learning in the NZ engineering workplace. *Journal of Workplace Learning*, 14, 190-197.
- Thomas, P., & Bain, J. (1984) Contextual dependence of learning approaches: The effects of assessments. *Human Learning*, 3, 227-240.
- Thomson, K., & Falchikov, N. (1998) Full on until the sun comes out: the effects of assessment on student approaches to studying. *Assessment and Evaluation in Higher Education*, 23, 379- 390.
- Tillema, H. (1998) Design and validity of a portfolio instrument for professional training. *Studies in Educational Evaluation*, 24, 263-278.
- Tillema, H. (2001) Portfolios as developmental assessment tools. *International Journal of Training and Development*, 5, 126-135.
- Tillema, H. (2003) Integrating developmental assessment with student-directed instruction: A case in vocational education in the Netherlands. *Journal of Vocational Education and Training*, 55, 113-125.
- Topping, K. J. (1998) Peerassessment between students in college and university. *Review of Educational Research*, 68, 249-276.

- Topping, K. J. (2003) Self and Peerassessment in School and University: Reliability, Validity and Utility. In M. Segers, F. Dochy & E. Cascallar (eds.). *Optimising New Modes of Assessment: In Search of Qualities and Standards* (pp. 55-89). Dordrecht/Boston/London: Kluwer Academic Publishers.
- Trigwell, K., & Prosser, M. (1991) Relating learning approaches to teaching: A relational perspective. *Higher Education*, 22, 251-266.
- Tynjälä, P. (1998) Traditional studying for examination versus constructivist learning tasks: Do learning outcomes differ? *Studies in Higher Education*, 23, 173-189.
- Weindling, A. M. (2001) Education and training: continuing professional development. *Current Paediatrics*, 11, 369-374.
- Williams, E. (1992) Student attitudes towards approaches to learning and assessment. *Assessment and Evaluation in Higher Education*, 17, 45-58.
- Wolf, D., Bixby, J., Glenn, J., & Gardner, H. (1991) To use their minds well: Investigating new forms of student assessment. *Review of Research in Education*, 17, 31-74.