

Peer- en co-assessment als instrument voor diepgaand leren: bevindingen en richtlijnen

Prof. dr. F. Dochy (Filip.Dochy@Edit.unimaas.nl) is werkzaam bij de Universiteit Maastricht en de Katholieke Universiteit Leuven. Dr. W. Admiraal is verbonden aan het Expertisecentrum ICT in het onderwijs van het Interfacultair Instituut voor de Lerarenopleiding, Onderwijsontwikkeling en Studievoordigheden (IVLOS) van de Universiteit Utrecht. Prof. dr. A. Pilot is werkzaam bij het IVLOS van de Universiteit Utrecht.

Deze inleidende beschouwing over peer- en co-assessment geeft een kort overzicht van wat bekend is over de bruikbaarheid en de mogelijke opbrengsten van deze vormen van assessment. Voorts wordt aandacht besteed aan de kwaliteitskenmerken en de belangrijkste ingrediënten van peer- en co-assessment. Ten slotte wordt teruggeblikt op de voornaamste resultaten van onderzoek naar het gebruik van deze assessmentvormen en wordt kort ingegaan op de relatie met de onderzoeksresultaten van de in dit nummer verder gepresenteerde onderzoeken.

INLEIDING

De laatste tien jaar is het onderzoek naar assessmentvormen sterk toegenomen. Over de hele wereld wordt op brede schaal erkend dat assessment een sterke invloed heeft op leerprocessen van studenten (Birenbaum & Dochy, 1996) en dat recent ontwikkelde vormen van assessment daar een positieve invloed op hebben (Segers, Dochy, & Cascallar, 2003). *Assessment for learning* benadrukt in de Engelstalige landen de essentie van de hele inzet: nieuwe toetsvormen moeten in de eerste plaats worden gebruikt om het leren te bevorderen. In de term *constructive alignment* komt naar voren dat door de studenten gestelde leerdoelen, competenties van studenten, het onderwijsarrangement, de didactische aanpak en de assessmentvormen op elkaar dienen te worden afgestemd. Vooral bij de introductie van nieuwe vormen van assessment in opleidingen dient deze afstemming een belangrijk uitgangspunt te zijn. Naast de assessmentcultuur heeft zich ook de stroming ontwikkeld waarin assessment wordt gezien als een integraal onderdeel van het leerproces (en niet langer als een sluitstuk alleen). De waarde die assessment kan hebben in het leerproces voor wat betreft zowel de feedback op processen en producten als de beoordeling van prestaties in de vorm van cijfers wordt alom erkend. Vanuit die idee is gepoogd om meer accent te leggen

op de plaats en de functie van assessment in het onderwijsleerproces.

Uit het onderzoek naar assessment komt een duidelijke en onweerlegbare boodschap naar voren, met name dat maatregelen ter bevordering van de effectiviteit van assessment niet alleen het leren stimuleren, maar dientengevolge ook de prestaties van studenten verhogen (Black & Wiliam, 1998). Het onderzoek toont aan dat het bevorderen van kennisontwikkeling en leerprestaties door assessment bepaald wordt door 5 eenvoudige sleutelfactoren:

- het verschaffen van effectieve feedback aan studenten;
- actieve betrokkenheid van studenten bij het eigen leren;
- afstemming van het onderwijs aan de resultaten van assessment;
- erkenning van de sterke invloed die assessment heeft op de motivatie, de eigendunk en het welbevinden van studenten, die op hun beurt van invloed zijn op het leerproces;
- behoefte van studenten om in staat te zijn zichzelf te beoordelen, te begrijpen hoe ze hun prestaties kunnen verbeteren, en invloed te hebben op hun ontwikkeling.

In het onderzoek worden ook remmende factoren geïdentificeerd, waaronder bijvoorbeeld de neiging van docenten om de hoeveelheid werk en de inzet van studenten te beoordelen, eerder dan de kwaliteit van de verworven kennis en competenties.

NIEUWE ASSESSMENTVORMEN: BRUIKBAARHEID IN OPLEIDINGEN

Dat nieuwe vormen van onderwijs en nieuwe vormen van assessment alleen goed kunnen gedijen in een nieuwe assessmentcultuur ligt voor de hand (Dochy, Heylen, & Van de Mosselaer, 2000; 2002). De overgang van een testcultuur naar een assessmentcultuur is nu al doorgedrongen op verschillende niveaus (Dochy, 1999). Onderzoek naar het gedrag en naar de percepties van studenten (Sambell, McDowell, & Brown, 1997) toont aan dat in een assessmentcultuur:

- 1 studenten vormen van assessment als meer fair beschouwen dan traditionele toetsvormen;
- 2 studenten een kritische zin ontwikkelen en evaluatievaardigheden verwerven;
- 3 studenten diverse andere academische vaardigheden ontwikkelen;
- 4 kennis meer geïntegreerd wordt;
- 5 dat studenten meer samenwerken en gemotiveerd zijn om zich in te zetten in het onderwijsproces.

Ook zijn er studies die zich meer richten op andere mogelijke opbrengsten van assessment (vergelijk Brown & Dove, 1991) en die aantonen dat assessment:

- 1 bijdraagt aan het ownership van studenten voor het eigen leerproces;
- 2 studenten motiveert en aanmoedigt tot actieve betrokkenheid;
- 3 leidt tot een meer objectieve houding ten opzichte van samen leren;
- 4 bijdraagt aan een interactieve uitwisseling van kennis;
- 5 leidt tot meer direct en effectief leren;
- 6 signalen geeft aan studenten dat hun ervaringen en beoordelingen worden gewaardeerd;

- 7 transferabele vaardigheden ontwikkelen;
- 8 kan resulteren in een reductie van de tijd die docenten aan begeleiding besteden;
- 9 leidt tot een meer diepgaande manier van leren van studenten;
- 10 constructieve kritiek uitlokt.

In diverse studies is de kwaliteit in termen van de validiteit en de betrouwbaarheid van peerassessment onderzocht. In 72% van deze studies werd een voldoende hoge betrouwbaarheid vastgesteld, zoals uitgedrukt in verschillende statistische maten (Topping, 2003). Recentelijk is de discussie opgelaaid over het bepalen van de kwaliteit van assessment. Eerder hebben we beargumenteerd dat dit niet langer kan op basis van traditionele criteria (Dierick, Dochy, & Van de Watering, 2001, Dochy e.a., 2002). In nu lopende studies wordt nagegaan, zowel in de vorm van fundamenteel als in toegepast onderzoek in de onderwijspraktijk, welke de meest relevante kwaliteitsindicatoren zijn (Segers, Dochy, & Cascallar, 2003).

KWALITEITSKENMERKEN VAN ASSESSMENT

Relevantie en evenwichtigheid zijn twee aan elkaar complementaire aspecten van validiteit. In de klassieke literatuur over evaluatie worden ze meestal respectievelijk aangeduid met de termen begrips- en inhoudsvaliditeit. Traditioneel geven deze aan in hoeverre bijvoorbeeld alle aspecten van intelligentie in een intelligentietest aan bod komen (inhoudsvaliditeit) en of de items in een intelligentietest het begrip intelligentie meten (begripsvaliditeit). Door in een assessmentcultuur een nieuw woordgebruik te hantieren, wil men duidelijk maken dat validiteit op een enigszins andere manier wordt ingevuld. In de testpsychologie hield men zich bezig met het berekenen van de validiteit van een test, nadat de test ontwikkeld was. Validiteit werd benaderd als een criterium om achteraf reeds ontwikkelde evaluatie-instrumenten te beoordelen. Docenten die een toets of een assessment inzetten voor hun eigen groep studenten, hebben aan dergelijke berekeningen achteraf meestal geen boodschap. Zij hebben veeleer behoefte aan richtlijnen om die toets of dat assessment onmiddellijk zo valide mogelijk te maken. En daarbij moeten zowel evenwichtigheid als relevantie worden nagestreefd. In plaats van de validiteit achteraf te berekenen, beargumenteren docenten tijdens het construeren hoe en waarom bepaalde handelingsdimensies worden opgenomen en waarom ze juist op die manier in bepaalde beheersingsniveaus worden vertaald. Omdat deze validiteit niet steunt op een berekening, maar op een argumentatie vooraf, wordt de validiteit ook aangeduid met de term *argumentatieve* validiteit (zie Swanborn, 1996).

Een tweede kwaliteitskenmerk is die van betrouwbaarheid en generaliseerbaarheid. Bij vele nieuwe vormen van assessment wordt de overeenstemming tussen twee of meerdere beoordelaars (interbeoordelaarsbetrouwbaarheid) van belang geacht. Dit wordt met name veroorzaakt door het feit dat bij veel vormen van assessment de te beoordelen processen en producten niet alleen of in het geheel niet wordt uitgedrukt in getallen, maar in formuleringen, beeldmateriaal of ander niet-cijfermatig materiaal. In het verlengde hiervan wordt bij het gebruik van assessmenttaken ook de generaliseerbaarheid onderzocht. Hierbij stelt men de vraag in welke mate een assessment gebaseerd op

authentieke taken, uitspraken kan doen over prestaties in andere realistische taken. Dit levert dan evidentie dat de instrumenten ook werkelijk de bedoelde competentie meten en de gekozen authentieke situatie voldoende representatief is. Het is duidelijk dat hier zeker beperkingen aan zijn en dat de specifieke context van de authentieke taken (casusspecificiteit) hierin een rol kan spelen. Het gaat hierbij om de generaliseerbaarheid van de beoordelingen, dat wil zeggen de vraag of het oordeel over deze student niet te zeer afhangt van de keuze van deze ene specifieke authentieke taak. De generaliseerbaarheid van de beoordeling kan worden verhoogd door de beoordeling te baseren op meerdere taken en meerdere beoordelingsmomenten. Daarnaast heeft ook het inschakelen van meerdere ervaren en/of getrainde beoordelaars en het gebruik van de juiste beoordelingsmethode een positieve invloed. Een gebruikelijke techniek voor het vaststellen van de generaliseerbaarheid, zoals een generaliseerbaarheidsstudie (Dochy e.a., 2002), is gericht op het ontdekken van de invloed van specifieke foutenbronnen. Ook hier kan men, om de eenduidigheid van de scoringsregels en de objectiviteit van de beoordelaars na te gaan, eenzelfde assessmenttaak door verschillende beoordelaars laten quoteren en nagaan in hoeverre de resultaten overeenstemmen.

Het belang van deze tweede eis, betrouwbaarheid en generaliseerbaarheid, dient enigszins gerelativeerd te worden. Een veelgehoorde opvatting in de onderwijswereld is dat een perfecte betrouwbaarheid niet haalbaar is en dat in de onderwijspraktijk validiteits-eisen belangrijker zijn (Segers e.a., 2003). Ook stelt men vaak vast dat men ten onrechte het streven naar betrouwbaarheid voorrang geeft op het streven naar validiteit, wellicht omdat in veel gevallen nagaan van de betrouwbaarheid eenvoudiger is. Volgens Gibbs (2003) heeft het ondoordacht streven naar betrouwbare testen in het verleden ons geleid naar een aantal hoogst onwenselijke onderwijssituaties, zoals studenten die reeksen items in testen proberen te herkennen, maar de onderwijsbijeenkomsten nooit hebben gevolgd. Er is volgens Gibbs te weinig rekening gehouden met het feit dat vormen van beoordeling dikwijls een sterke impact hebben op de activiteiten van studenten en op onderwijs in zijn totaliteit. Dit laatste wordt veelal benoemd door het concept 'consequentiele validiteit', waarvan in de testcultuur geen sprake was (Gielen, Dierick & Dochy, 2003). Binnen de testcultuur ging men ervan uit dat een evaluatie het onderwijs zo weinig mogelijk mocht verstoren. Evaluatie werd daarom vaak in de tijd afgescheiden van het onderwijs door toets- of examenperiodes in te richten. In recent onderzoek omtrent de invloed van evaluatie op het onderwijs en het leergedrag van studenten, heeft men echter ontdekt dat zelfs zulke afgescheiden toetsen een invloed uitoefenen op de dagelijkse klaspraktijken (pre-assessment-effecten) (Segers e.a., 2003; Gielen e.a., 2003). Bijgevolg is men gaan inzien dat het integreren van evaluatie binnen de leeromgeving een sterke ondersteuning kan betekenen van het leerproces (zowel pre- als post-assessment-effecten). Aangezien we dus onmogelijk nog de invloed van evaluatie op het onderwijs en het leren kunnen ontkennen, kunnen we er maar beter rekening mee houden en deze invloed op een positieve manier aanwenden, luidt de stelling in de assessmentcultuur. Met andere woorden, de didactische mogelijkheden van beoordeling (vaak meer dan alleen maar cijfers geven) worden ook onderkend in de assessmentcultuur.

PEER- EN CO-ASSESSMENT

Peerassessment kan omschreven worden als een proces waarbij studenten producten of andere opbrengsten van hun medestudenten evalueren, al of niet op basis van zelf-geselecteerde criteria. Bij deze procedure wordt gebruikgemaakt van beoordelings-schalen, meer kwalitatieve formulieren of andere instrumenten die ontworpen zijn door derden of zelf ontworpen werden door de gebruikers om aan hun specifieke behoeften tegemoet te kunnen komen (Dochy, Segers, & Sluismans, 1999). Topping (1998) omschrijft *peerassessment* als een onderwijsarrangement. Bovendien specificeert hij dat wat er beoordeeld wordt, namelijk niet de persoon, maar zijn/haar prestaties, processen of producten.

Het is belangrijk te benadrukken dat *peerassessment* niet alleen gaat om het toekennen van punten, maar ook een deel vormt van het leerproces waarbij bepaalde vaardigheden worden ontwikkeld. Uit onderzoek blijkt dat bij *peerassessment* een hoog niveau van verantwoordelijkheid bij studenten kan worden bekomen, dat studenten veelal eerlijk en accuraat zijn bij de beoordelingen die ze maken, dat de nadruk ligt op vaardigheden en dat *peerassessment* de mogelijkheid biedt tot het geven van feedback (Dochy, Heylen, & Van de Mosselaer, 2002). Sommige auteurs merken ook op dat *peerassessment* kan beschouwd worden als onderdeel van het zelfbeoordelingsproces doordat informatie wordt verschaft voor de zelfbeoordeling. Wanneer studenten wordt gevraagd producten van hun medestudenten te beoordelen, worden zij genoodzaakt de criteria waarop zij dit doen helder te formuleren. Hierdoor krijgen zij ook meer inzicht in de criteria waaraan hun eigen werk aan zou moeten voldoen (zie bijvoorbeeld Brown, Rust, & Gibbs, 1994). Self- en *peerassessment* kan dan ook niet helemaal los van elkaar gezien worden.

Bij *co-assessment* hebben studenten de gelegenheid elkaar te beoordelen maar het eindoordeel wordt uitgesproken door de docent (Dochy & Segers, 1999). Best, Carswell en Abott (1990) definiëren *co-assessment* als een leer- en onderwijsproces waarbij de student en docent elkaar ontmoeten om de objectieven en standaarden te verduidelijken. Volgens hen is het hierbij niet noodzakelijk dat de student verantwoordelijk is voor de beoordeling, maar is het vooral belangrijk dat de student heeft meegewerkt aan het bepalen van wat moet worden beoordeeld en door wie (Sommervell, 1993). Bij *co-assessment* hebben de studenten dus samen met de docent een rol in het beoordelingsproces. Dochy en Segers (1999) wijzen erop dat *co-assessment* de volgende drie bedoelingen heeft. *Co-assessment* biedt de student de gelegenheid om de rol van de docent in te nemen (rolwisselend onderwijs); de student verwerft inzichten in het beoordelingsproces, waardoor hij beter in staat is anderen te beoordelen, en ten slotte is de student meer vaardig in het beoordelen van zichzelf (self assessment). *Co-assessment* is een dialoog tussen docenten en studenten. Het kan louter formatief zijn, maar ook deel uit maken van het summatieve beoordelingsproces. Daarbij kunnen criteria op voorhand vastliggen, maar studenten kunnen ook zelf criteria aandragen. Indien de studenten zelf moeilijk criteria kunnen vinden, presenteert de docent een lijst van criteria en laat de studenten elk item bediscussiëren, argumenteren voor en tegen formuleren, voorstellen doen, toevoegingen maken, enzovoort.

Samenvattend zijn de belangrijkste aspecten van peer en co- assessment (Dochy e.a., 2002):

- actief betrekken van studenten bij het beoordelingsproces;
- beoordeling van de vaardigheid van probleemoplossen en het proces van leren leren;
- ontwikkeling van een authentieke nauwgezetheid en volledigheid in de manier waarop studenten zichzelf beoordelen;
- meeting criteria improves learning. Studenten 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 competenties vergeleken worden met een set van vastgestelde criteria;
- ontwikkeling van interpersoonlijke vaardigheden. Studenten leren via discussies en dialogen hun mening uiten en voor zichzelf en eventuele medestudenten (bij de combinatie met self- and peerassessment) opkomen;
- studenten krijgen inzicht in groepsdynamiek;
- studenten worden voorbereid op hun toekomstig professioneel functioneren;
- studenten leren de vakinhouden op een diepgaande manier.

ONDERZOEK NAAR PEERASSESSMENT

Niet alleen in de onderwijspraktijk, maar ook in onderzoek is de aandacht voor peer-assessment sterk toegenomen. Het grootste deel van dit onderzoek richt zich op toepassingen in het hoger onderwijs, hoewel er ook steeds meer op het niveau van basis- en middelbaar onderwijs en volwassenenonderwijs wordt gedaan (Dochy, Segers & Sluijsmans, 1999; Falchikov & Goldfinch; 2000; Segers & Dochy, 2001; Topping, 2003). Zo blijkt ook in programma's gericht op off-the-job training, bij leren op de werkplek, dat peerassessment ook steeds meer aan belang wint (Smith, 2003). Voorbeelden hiervan zijn het gebruik van peerassessment in de context van Personal Development Plans, leercontracten, MBA programma's gericht op Self-Managed Learning, en 360° Competency Feedback systems.

Recentelijk is veel onderzoek gepubliceerd over de effecten van peerassessment op het leer- en onderwijsproces. Vele onderzoeken zijn *case studies* in specifieke vakgebieden en dikwijls exploratief van aard. Sinds kort zijn er ook wat meer generatieve reviewstudies verschenen (Dochy, Segers & Sluijsmans, 1999; Falchikov & Boud, 1989; Falchikov & Goldfinch, 2000; Topping, 1998; 2003). Topping (2003) en Dochy e.a. (1999) concluderen dat er evidentie is voor de stelling dat peerassessment leidt tot leerwinst, die op zijn minst vergelijkbaar is met de leerwinst op basis van traditionele toetsen, die door Topping als *gains from teacher assessment* beschreven (Topping, 2003, p. 77). De auteurs refereren naar cognitieve en metacognitieve effecten, zowel voor de ontvanger van de beoordeling als voor de student die de beoordeling uitbrengt. Deze studies geven aan dat peerassessment het hogere-ordenen stimuleert, tot een beter begrip van de taken leidt en tot betere transfer naar nieuwe situaties leidt, vooral omdat peerassessment onmiddellijke, *just-in-time* en op het individu

afgestemde feedback inhoudt. Belangrijke pluspunten zijn ook dat studenten meer tijd aan de eigenlijke taak besteden (*higher rates of productive time on task*) en minder vaak dezelfde fouten blijven herhalen (*reduction of cumulative error*) (Dochy e.a., 1999; Topping, 2003). Daarnaast wijzen beide overzichtsstudies ook op de effecten van peerassessment op de ontwikkeling van sociale en communicatieve vaardigheden: vaardigheden om in teams te leren en te werken, actief leren, onderhandelen, evalueren, feedback geven en diplomatisch handelen.

Toch geven de genoemde studies ook aan dat peerassessment niet altijd tot voordelen leidt. Diverse variabelen kunnen een negatieve invloed hebben op het leerproces en de leeropbrengst. Met name drie variabelen komen hierbij naar boven: de accuraatheid van de beoordeling, de transparantie van de beoordeling en de ervaring in het gebruik van peerassessment. Studies naar de *accuraatheid van studentbeoordelingen* tonen geen consistente resultaten. Vriendjespolitiek komt soms voor, leidend tot hogere scores en *collusive marking* blijkt voor te komen, leidend tot een gebrek aan differentiatie in scores binnen de groep. Verder komt *decibel marking* soms voor, waarbij dominante groepsleden hogere scores krijgen; en ook *parasite marking* komt voor, waarbij meelifters profiteren van een groepsscore. Het omzetten van groepsscores naar individuele scores door middel van co-assessment blijkt evenwel een goede methode om deze mogelijke verschijnselen te voorkomen (zie Dochy e.a., 2002). Daarbij gaat veel onderzoek in op de percepties van studenten op de *fairness* van peerassessment. Studenten blijken zich niet altijd vertrouwd te voelen in hun rol van beoordelaars en het geven van kritiek op medestudenten blijkt soms moeilijk. Er blijkt ten aanzien van dit aspect echter een positieve ontwikkeling in de tijd, die inhoudt dat meer ervaring met peerassessment leidt tot meer vertrouwen in eigen kunnen. Studenten geven aan begeleiding en training nodig te hebben om de rol van beoordelaars op te kunnen nemen. Dit wordt ook aangetoond in het onderzoek van Sluijsmans (dit nummer).

Ten tweede benadrukt onderzoek het belang van de *transparantie* van de peerassessment in termen van duidelijkheid van de procedure, de doelen, de criteria en de norm. Hoewel er is aangetoond dat bekendheid van studenten met, en *ownership* over, de criteria leidt tot grotere overeenkomst tussen de beoordeling van peers en experts, is er toch nog weinig bekend over de invloed hiervan op de leeropbrengst. Tenslotte is er aangetoond dat meer *ervaring* in het gebruik van peerassessment een positieve invloed heeft op het effect. Dit omvat eveneens het geven van trainingen waarin studenten vertrouwd worden gemaakt met de doelen van peerassessment, de procedures, het ontwikkelen en het gebruik van criteria en het proces van het geven en krijgen van feedback van medestudenten. Ervaring van docenten met het modeleren van feedback blijkt ook van belang te zijn: De studie van Van den Berg, Admiraal & Pilot in dit nummer suggereert dat een verklaring voor het feit dat studenten zo zijn gericht op de stijlkenmerken in hun feedback, mede veroorzaakt wordt door het model dat docenten geven. Docenten zijn in hun feedback ook vooral gericht op stijlkenmerken en grammaticale zaken.

De studies van Sluijsmans en Van den Berg e.a. (dit nummer) bevestigen de bovenstaande onderzoeksresultaten en brengen belangrijke nuanceringen aan die voor de bruikbaarheid van peerassessment in de praktijk relevant zijn.

Het lijkt belangrijk te zijn om studenten eerst formatief peerassessment te laten ervaren, een goede transparantie van de assessment te organiseren en studenten vertrouwd te maken met procedures en criteria. In concrete onderwijssituaties blijken ook duidelijke richtlijnen te vinden voor een maximaal effect van peerassessment voor het beoordelen van schriftelijke producten (Van den Berg, e.a., dit nummer). Opvallend hierbij is dat wederzijdse feedback in groepjes van drie of vier studenten die elkaar ook mondelinge toelichtingen geven binnen de contacturen goed blijkt te werken. Al even opvallend is de vaststelling dat het niet volledig bekend zijn met elkaars onderwerp niet hoeft te leiden tot een meer oppervlakkige beoordeling. Verder komt uit deze studie naar voren dat docenten niet noodzakelijk ook feedback moeten geven; eerder is het aan te bevelen dat de docent zich op het moment van bilaterale feedback beperkt tot de rol van procesbewaker.

AFSLUITING

Peer- en co-assessment hebben de laatste tien jaar in het onderwijs een hoge vlucht genomen. Duidelijk is geworden dat de sleutel van een goed assessment door peers ligt in de actieve betrokkenheid, de effectieve feedback, de afstemming van onderwijs op assessmentresultaten, de erkenning van de invloed van assessment op motivatie en welbevinden, en de behoefte van studenten om mee hun leerproces te sturen. Er is inmiddels aangetoond dat peer- en co-assessment velerlei opbrengsten kunnen hebben: meer ownership, hogere motivatie, meer diepgaand leren van inhouden, betere academische en communicatieve vaardigheden, enzovoort. Uit internationaal onderzoek blijkt dat schrik van docenten voor het gebruik van deze assessmentvormen vaak niet gegrond is: studenten blijken een hoog niveau van verantwoordelijkheid aan de dag te leggen, en zijn veelal eerlijk en accuraat. Toch hangt veel af van de precieze organisatorische inrichting, de context, het niveau van de studenten, het product of het gedrag dat wordt beoordeeld, en de duidelijkheid van de criteria. Dergelijke variabelen verklaren zeker de soms gevonden contradictorische onderzoeksresultaten: aspecten van betrouwbaarheid en generaliseerbaarheid blijken hoger te zijn bij groepen van gevorderde studenten in een opleiding, en lager voor assessment van professionele praktijk dan voor assessment van academische producten. Discussie, onderhandelen en samen construeren van de assessmentcriteria met de studenten leidt tot meer ownership, een meer diepgaand begrijpen van de inhoud en tot hogere generaliseerbaarheid. De kwaliteit van peerassessment blijkt ook kwalitatief beter wanneer dit wordt ondersteund door training van studenten, checklists, voorbeelden, coaching door docenten en monitoring.

Concluderend kan worden gezegd dat het gebruik van peer- en co-assessment zijn plaats verdient in het onderwijs. Uiteraard naast diverse andere assessmentvormen; diversiteit van assessmentvormen verhoogt het leereffect bij studenten en verhoogt de kwaliteit van de assessment. Praktijk en onderzoek zullen uitwijzen in welke context en in welke organisatorische inbedding peer- en co-assessment het meest effectief zijn.

LITERATUUR

- Best, M., Carswell, R., & Abott, D. (1990) Self-evaluation for nursing students. *Nursing outlook*, 38, 4, 172-177.
- Berg, B.A.M. van den, Admiraal, W., & A. Pilot (2003) Peerassessment in universitair onderwijs. Een onderzoek naar bruikbare ontwerpen. *Tijdschrift voor Hoger Onderwijs*, 21, 4, 251-271.
- Birenbaum, M., & Dochy, F.J.R.C. (1996) (Eds.) *Alternatives in assessment of achievements, learning processes and prior knowledge*. Boston, Dordrecht, London: Kluwer Academic Publishers.
- Black, P., & Wiliam, D. (1998) Assessment and classroom learning. *Assessment in Education*, 5, 1, 7-74.
- Brown, S., Rust, C., & Gibbs, G. (1994) Involving students in the assessment process. In *Strategies for diversifying assessments in higher education* Oxford: Oxford Centre for Staff Development. (<http://www.lgu.ac.uk/deliberations/ocsd-pubs/div-ass5.html>)
- Dierick, S., Dochy, F., & Van de Watering, G. (2001) Assessment in het hoger onderwijs: Over de implicaties van nieuwe toetsvormen voor de edumetrie. *Tijdschrift voor Hoger Onderwijs*, 19, 1, 2-18.
- Dochy, F. (1999) Een structureel beleid voor toetsing en assessment in constructiegericht onderwijs (CGO): Kern voor het slagen van universitaire onderwijsvernieuwing. Inaugurale rede uitgesproken bij aanvaarding van de L. Verhaegen leerstoel. Limburgs Universitair Centrum: Universitaire Pers.
- Dochy, F., Heylen, L. & Van de Mosselaer, H. (2000) *Coöperatief leren in een krachtige leeromgeving: Handboek probleemgestuurd leren*. Leuven: Acco.
- Dochy, F., Heylen, L. & Van de Mosselaer, H. (2002) *Assessment in Onderwijs*. Utrecht: Lemma.
- Dochy, F.J.R.C., Segers, M.S.R., & Sluijsmans, D. (1999) The use of self-, peer and co-assessment in higher education: a literature review. *Studies in Higher Education*, 24, 3, 331-350.
- Dochy, F., & Segers, M. (1999) Innovatieve toetsvormen als gevolg van constructiegericht onderwijs: op weg naar een assessment-cultuur, 181-206 In: P. De Boeck en M. Lacante (Red) *Meer kansen creëren voor het hoger onderwijs*. Dordrecht: Kluwer.
- Falchikov, N. & Boud, D. (1989) Student self-assessment in higher education: A meta-analysis. *Review of Educational Research*, 5, 4, 395-430.
- Falchikov, N., & Goldfinch, J. (2000) Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70, 3, 287-322.
- Gibbs, G. (2003) *Improving student learning through changing assessment. A conceptual and practical framework*. Paper presented at the SIG Invited Symposium of the EARLI Conference, Padova, Italy, August 26-30, 2003.
- Gielen, S., Dierick, S., & Dochy, F. (2003) *Do new assessment forms meet the expectations with regard to consequential validity?* Paper presented at SIG Invited Symposium of the EARLI Conference, Padova, Italy, August 26-30, 2003.

- Sambell, K. & McDowell, L., & Brown, S. (1997) "But what is fair?" An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation*, 23(4), 349-371.
- Segers, M., & Dochy, F. (2001) New assessment forms in problem-based learning: the value-added of the students' perspective. *Studies in Higher Education*, 26, 3, 327-343.
- Segers, M., Dochy, F., & Cascallar, E (2003) The Era of assessment engineering: Changing perspectives on teaching and learning and the role of new modes of assessment 1-13. In: M. Segers, F. Dochy & E. Cascallar (Eds.) *Optimising New Modes of Assessment: In Search of Qualities and Standards*. Dordrecht, Boston, London: Kluwer Academic Publishers.
- Sluijsmans, D. (2003) Peerassessment en de ontwikkeling van reflectievaardigheden in de lerarenopleiding. *Tijdschrift voor Hoger Onderwijs*, 21, 4, 230-250.
- Smith, P.J. (2003) Workplace Learning and Flexible Delivery. *Review of Educational Research*, 73, 1, 53-89.
- Sommervell, H. (1993) Issues in assessment, enterprise and higher education: The case for self-, peer and collaborative assessment. *Assessment and Evaluation in Higher Education*, 18(3), 221-33.
- Swanborn, P. G. (1996) A common base for quality control criteria in quantitative and qualitative research. *Quality & Quantity*, 30, 19-35.
- Topping, K. J. (1998) Peer assessment between students in college and university. *Review of Educational Research*, 68, 3, 249-276.
- Topping, K.J. (2003) Self and Peer Assessment in School and University: Reliability, Validity and Utility (pp. 55-89) In: M. Segers, F. Dochy & E. Cascallar (Eds.) *Optimising new modes of assessment: In search of qualities and standards*. Dordrecht, Boston, London: Kluwer Academic Publishers.