

Peerassessment en de ontwikkeling van reflectievaardigheden in de lerarenopleiding

Dominique M.A. Sluijsmans
(dominique.sluijsmans@ou.nl) is
werkzaam bij het
Onderwijstechnologisch Experti-
secentrum (OTEC) van de Open
Universiteit Nederland

In deze bijdrage wordt een studie gerapporteerd waarin de effecten van een geïntegreerde training in peerassessment-vaardigheden op de reflectievaardigheid en de beoordelingsvaardigheid zijn onderzocht. 110 eerstejaars pabo-studenten zijn binnen drie wiskundemodules getraind in vaardigheden als het definiëren van criteria, het geven van feedback en het schrijven van een kwalitatief peerassessment-rapport. Na elke module schreven studenten een reflectieverslag over hun leerervaringen. Dit verslag is na elke module telkens beoordeeld door een medestudent. De studenten werden in de gelegenheid gesteld hun reflectieverslag bij te stellen op basis van de ontvangen peerfeedback. De docent wiskunde beoordeelde het definitieve reflectieverslag twee weken na de laatste module. Uit de analyse van de geschreven peerassessment-rapporten en de behaalde resultaten met betrekking tot de kwaliteit van de reflectieverslagen na de derde module, kan worden geconcludeerd dat een training in peerassessment-vaardigheden een positieve uitwerking heeft op zowel de kwaliteit van de beoordelingsvaardigheid van de studenten als op de kwaliteit van hun reflectievaardigheid. Resultaten van het vragenlijst-onderzoek lieten daarnaast zien dat studenten de nieuwe aanpak positief waardeerden.

INLEIDING

Effectieve assessmentbenaderingen in het hoger onderwijs, gebaseerd op constructivistische uitgangspunten, staan de laatste jaren flink in de belangstelling. In deze assessmentbenaderingen staat de integratie van onderwijs- en toetsingsactiviteiten centraal, waarbij een actieve rol wordt weggelegd voor de student. De student wordt beschouwd als een mede-onderhandelaar over de inhoud van het onderwijs, waarbij in dialoog met de docent wordt gereflecteerd, gediscussieerd en samengewerkt. Assessment heeft al lang niet meer louter de functie van certificering, maar krijgt steeds meer een formatief karakter. Studenten worden in staat gesteld om deze formatieve momen-

ten te gebruiken ten behoeve van bijsturing van het eigen leerproces. De nadruk verschuift van een representatie van assessment als 'afrekenmechanisme' naar een representatie van assessment als een *tool for learning* (Arter, 1996; Boud, 1990; Dochy & McDowell, 1997).

Assessment is een belangrijk thema in huidige innovaties rondom meer studentgecentreerd onderwijs. Studenten worden geacht zich verantwoordelijk op te stellen ten opzichte van hun eigen leerpad en actief te participeren in onderwijsactiviteiten. Om dit te bereiken, zal een visie op assessment moeten worden gekozen waarbij toetsing, leren en instructie volledig in elkaars verlengde liggen (Sluijsmans, 2002). We zien immers nog te veel dat vernieuwingen aan waarde inboeten op het moment dat er aan de toetskant niets verandert. Om de implementatie van assessment als een *tool for learning* te realiseren, zullen op verschillende niveaus binnen een onderwijsorganisatie een aantal veranderingen moeten plaatsvinden. Zowel op het niveau van de student, als op het niveau van de docenten en het management zal een *commitment* aan een nieuwe aanpak in onderwijsontwerp en toetsontwerp nodig zijn. Dit artikel richt zich met name op de rol van de student en de docent. Meer specifiek is de rol van de individuele pabo-student als beoordelaar van zijn eigen werk en dat van zijn medestudenten nader onderzocht.

Het blijkt dat vooral in lerarenopleidingen veel behoefte bestaat aan het invoeren van peerassessment, aangezien deze werkvorm goed past binnen de meest recente opvattingen over het opleiden van leraren. Lerarenopleiders onderschrijven steeds meer het belang van discussie met medestudenten over de vereiste didactische vaardigheden en de toepassing hiervan in de onderwijspraktijk. Deze discussies worden vaak gevoerd door persoonlijke onderwijs- en leerervaringen. Het inzetten van studenten als medebestuurders in de vorm van peerassessment kan lerarenopleiders een waardevol instrument in handen geven om een meer actieve betrokkenheid van studenten te bewerkstelligen (Bangert, 1995). Maar hoe kan deze betrokkenheid door peerassessment nu concreet in de praktijk worden ingevoerd en wat zijn vervolgens effecten op leerresultaten van studenten en hun beeld ten aanzien van onderwijs en toetsing?

Ondanks een aantal pogingen richting studentgecentreerd onderwijs, is het in het hoger onderwijs immers nog steeds geen natuurlijk gegeven dat studenten meebeslissen over de kwaliteit van hun eigen werk en dat van medestudenten of dat zij het onderwijs meeontwerpen. Ook al wordt stellig beweerd dat elk assessment moet worden vormgegeven in partnerschap tussen studenten en docenten (Stefani, 1998), blijft de vraag bestaan hoe dit in de concrete onderwijspraktijk te realiseren.

Om meer inzicht te verkrijgen in toepassingen van peerassessments in het hoger onderwijs, is een aantal studies geanalyseerd (Sluijsmans, Dochy & Moerkerke, 1999). Peerassessment toepassingen blijken zich vooral te richten op *peer ranking* (iedere groepslid plaatst zijn medestudenten in een lijst van 'beste' tot 'slechtste' op één of meer criteria), *peer nomination* (ieder groepslid geeft aan welke student hij het beste beoordeelt op een bepaald criterium), of *peer rating* (ieder groepslid scoort elk ander groepslid op een set van karakteristieken op basis van één of meerdere beoordelingsschalen) (Kane & Lawler, 1978). De meerderheid van de studies beschrijven een overwegend kwantitatieve aanpak in peerassessment, waarbij vooral wordt gekeken naar de overeenstemming in scores die zijn gegeven door docenten en studenten (zie bijvoorbeeld Topping, 1998). Hoewel student-docentcorrelaties veelbelovende resultaten laten zien, blijkt ook

dat studenten het veelal niet prettig vinden om elkaar op een scoringsgebaseerde manier te beoordelen zonder de mogelijkheid te hebben tot het geven van feedback. Hanrahan and Isaacs (2001) bijvoorbeeld ondervonden een vijandige houding van studenten tegenover peerassessment in hun cursussen op de universiteit. Enkele onderzoekers vonden een toegenomen aversie tegenover peerassessment na ervaring ermee te hebben opgedaan (Rushton, Ramsey & Rada, 1993). Vooral het aspect van het geven van een cijfer wordt door studenten als risicovol en oneerlijk beschouwd (Kwan & Leung, 1996). De resultaten van het literatuuronderzoek en een pilot studie (Sluijsmans, Moerkerke, Dochy & Van Merriënboer, 2001) leidde derhalve tot drie assumpties die ondersteunend zijn voor een meer structurele implementatie van peerassessment.

De eerste assumptie is dat het uitvoeren van een peerassessment een complexe vaardigheid is die meer om het lijf heeft dan het geven van een cijfer aan een medestudent. Voordat studenten in de rol van medebeoordelaar worden geplaatst, is het voorwaardelijk dat zij begrijpen welke vaardigheden ermee gemoeid zijn om tot een goede beoordeling te komen. Studenten zullen derhalve expliciete training behoeven in assessmenttechnieken, om zo tot betrouwbare en acceptabele beoordelingen te komen (Boud, 1990; Hanrahan & Isaacs, 2001). Om een dergelijke training te ontwerpen, is het noodzakelijk de beoordelingsvaardigheid onder de loep te nemen en te bezien welke onderliggende vaardigheden van belang zijn. Op basis van het literatuuronderzoek en interviews met assessmentexperts, is een model ontworpen waarin deze onderliggende vaardigheden zijn uitgewerkt (Sluijsmans, Dochy & Moerkerke, 1999; Sluijsmans & Van Merriënboer, 2000). Drie hoofdvaardigheden zijn onderscheiden: (1) het kunnen definiëren van assessmentcriteria; (2) het kunnen geven van constructieve feedback; en (3) het kunnen schrijven van een kwalitatief assessment report. In het model wordt peerassessment opgevat als een *leeractiviteit* en niet als een scoringsinstrument, zoals nog vaak gebruikelijk is in huidige toepassingen (zie Falchikov & Goldfinch, 2001; Sluijsmans e.a., 1999).

Een tweede assumptie is dat een training in peerassessment-vaardigheden positieve effecten kan hebben op de ontwikkeling van inhoudsgebonden vaardigheden, mits de training is ingebed in bestaande onderwijsactiviteiten (Mehrens, Popham & Ryan, 1998). Vanuit dit gezichtspunt wordt een beoordelingsvaardigheid niet getraind als een geïsoleerde vaardigheid, maar wordt deze direct gerelateerd aan de onderwijsinhoud. Als een docent bijvoorbeeld een training van de vaardigheid 'definiëren van criteria' integreert in zijn module over presentatievaardigheden, zullen studenten leren hoe te onderhandelen over criteria van goede presentaties. Het begrijpen van deze criteria helpt studenten vervolgens om zelf een betere presentatie te houden. De training van de beoordelingsvaardigheid en de ontwikkeling van een inhoudsgebonden vaardigheid werken zo versterkend op elkaar (Boud, 1995; Orsmond, Merry & Reiling, 1996; 1997; 2000).

De derde en laatste assumptie gaat ervan uit dat goed ontwikkelde reflectievaardigheden een voorwaarde zijn voor het uitvoeren van betrouwbare peerassessments. Om hun eigen werk of het werk van medestudenten te beoordelen, zullen studenten reflectievaardigheden aanwenden om sterktes en zwaktes te herkennen. Reflectie en peerassessment zijn in dit opzicht verweven: het analyseren van het werk van medestudenten kan leiden tot een verbeterde kijk op de kwaliteit van het eigen werk (Falchikov, 1995; Freeman, 1995).

Reflectie is een bekend begrip binnen de context van het opleiden van leraren (Korthagen & Wubbels, 2000; Kremer-Hayon & Tillema, 1999; Newman, 1996; Richert, 1999). De literatuur over reflectie is onuitputtelijk, maar het wordt algemeen onderschreven dat reflectie belangrijk is bij het onderkennen van het eigen leerproces om zo tot betere prestaties te komen (Reilly-Freese, 1999). Reflectie wordt binnen de lerarenopleiding vooral gekoppeld aan de stage-ervaringen (zie bijvoorbeeld Anderson & Freiberg, 1995). Schön (1987) onderscheidt hierbij *reflection on action*, waarmee het denkproces voor en na het geven van een les wordt bedoeld en *reflection in action*, dat verwijst naar het denkproces dat optreedt tijdens het geven van een les. Loughran (1996) ontwikkelde een driedelig conceptueel kader met betrekking tot reflectie op onderwijzen: reflectie tijdens het plannen van een les, tijdens een les en na een les. Een andere invulling van reflectie is het spiraalmodel van Korthagen (1985), waarbij vijf fasen worden onderscheiden: (1) actie, (2) kijken naar of terugkijken, (3) bewustwording van essentiële aspecten, (4) creatie van alternatieve oplossingen en (5) uitproberen. De vijfde fase vormt de eerste fase van een nieuwe cyclus.

Een specifieke vorm van reflecteren die meestal een terugkerende activiteit is van studenten in de lerarenopleiding, is het schrijven van een reflectieverslag na het volgen van een module, waarbij de stage slechts één aspect is. Lerarenopleiders hebben een verantwoordelijkheid om studenten te begeleiden in de principes van het schrijven van een reflectieverslag op een metacognitief niveau, zodat een proces van reflectie door te schrijven optreedt. Het ontbreekt echter vaak aan deze begeleiding, waardoor iedere student zijn eigen strategie ontwikkelt in het schrijven van een reflectieverslag.

Bovenstaande assumpties zijn toegepast in een studie die is uitgevoerd binnen de context van een pabo. Een training in beoordelingsvaardigheden is geïntegreerd in eerstejaars wiskundemodules. De betrokken wiskundedocent wilde graag het schrijven van goede reflectieverslagen – een belangrijke activiteit in deze modules – koppelen aan peerassessment (zie derde assumptie). Het gevoel was ontstaan dat studenten door het veelvuldig moeten schrijven van reflectieverslagen een soort routinegedrag vertoonden: studenten zouden vooral schrijven wat de docent wil horen, waardoor een troebel beeld ontstond over de werkelijke reflectievaardigheid van studenten. Om de training in assessmentvaardigheden te combineren met de ontwikkeling van reflectievaardigheden, is voor deze studie besloten het schrijven van een reflectieverslag te beschouwen als de inhoudsgebonden vaardigheid (zie de tweede assumptie). Volgens het model zijn studenten derhalve getraind in het definiëren en toepassen van geschikte criteria voor reflectieverslagen, hoe feedback te geven ten aanzien van reflectieverslagen van medestudenten en hoe een peerassessment-rapport te schrijven over reflectieverslagen van medestudenten.

Drie reeds bestaande eerstejaars wiskundemodules zijn in enige mate herzien ten behoeve van de assessmenttraining. Twee vragen zijn onderzocht. De eerste onderzoeksvraag richtte zich op de effecten van de assessmenttraining op twee typen leerresultaten: 1) de kwaliteit van de assessmentvaardigheid van studenten, geoperationaliseerd in drie delen: het gebruik van de juiste criteria, het geven van adequate feedback en de kwaliteit van het geschreven rapport, en 2) de kwaliteit van de reflectieverslagen. De hypothese was dat de training leidt tot een verbetering van beide resultaten.

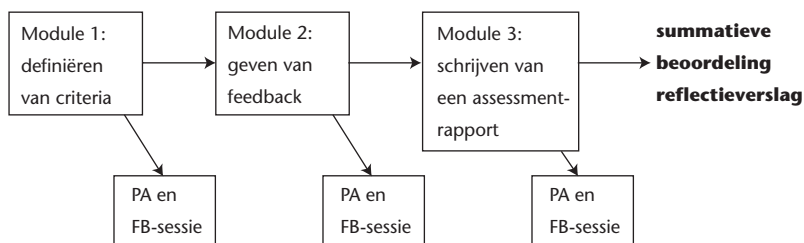
De tweede vraag richtte zich op de verandering in percepties van studenten ten aanzien van assessment en instructie als gevolg van de training. Graag expliciet formuleren

waarop de tweede onderzoeksvraag zich richtte Om deze vraag te beantwoorden is vóór de eerste en ná de derde module onderzocht wat de percepties van studenten waren ten aanzien van verschillende aspecten rondom wiskunde en assessment, zoals zelfbeeld en toetsangst. Trujillo & Hadfield (1999) concludeerden dat er bijna geen studies voorhanden zijn die interventies onderzoeken waarin de relatie tussen zelfbeeld en toetsangst in het wiskundedomein nader wordt bekeken. In deze studie werd verwacht dat het ontvangen van feedback en het schrijven van reflectieverslagen de beelden van studenten tegenover wiskunde positief zou veranderen en dat de toetsangst lager zou worden.

METHODE

Onderzoekspopulatie

De onderzoekspopulatie in deze studie waren 110 eerstejaars pabo-studenten van de Faculteit Onderwijs van de Hogeschool Zuyd (15 mannen, 95 vrouwen) met een gemiddelde leeftijd van 19.4 jaar. De groep studenten was heterogeen wat betreft onderwijsachtergrond: 79 studenten stroomden in vanuit het middelbaar onderwijs (havo/vwo), 21 studenten hadden reeds een beroepsopleiding op mbo-niveau gevolgd, 6 studenten stroomden in vanuit een andere hbo-opleiding en 4 studenten hadden al gewerkt in een andere beroepscontext.



Figuur 1 Opzet van de modules en de training

Onderzoeksmaterialen

- *Wiskundemodules en geïntegreerd trainingsprogramma.* Voor deze studie zijn drie elkaar openvolgende eerstejaars wiskundemodules uitgekozen, elk met een duur van zes weken. In deze modules zijn de studenten via studietaken geconfronteerd met een aantal didactische basisvaardigheden met betrekking tot het geven van wiskunde in de basisschool. Na elke module is de studenten gevraagd een reflectieverslag te schrijven ten aanzien van leerervaringen opgedaan tijdens de betreffende module. Om studenten meer te betrekken bij de invulling van het reflectieverslag en ze te ondersteunen in het beoordelen van elkaar, is in elk van de drie modules een training in assessmentvaardigheden geïntegreerd (zie figuur 1).

In de eerste module zijn vier zogenaamde peerassessment-taken (PA-taken) aan de reguliere studietaken gekoppeld. Alle PA-taken duurden een uur en werden tevens door de wiskundedocent geïnstrueerd. In de PA-taken stonden werkvormen als onderhandeling, discussie en dialoog tussen studenten en docent centraal. In de eerste module is vooral ondersteuning geboden voor het leren definiëren van criteria van een goed reflectieverslag. In de tweede module zijn twee PA-taken ingebed ten aanzien van het geven van feedback en in de derde module waren de PA-taken gericht op het schrijven van een kwalitatief peerassessment-rapport. Na elke module is een feedbacksessie georganiseerd waarin studenten de schriftelijke feedback mondeling rapporteerden aan de medestudenten.

- *Peerassessment-formulier.* Na elke module beoordeelden de studenten het reflectieverslag van één medestudent met behulp van een peerassessment-formulier. De studenten hebben dit in het totale onderzoek vijf keer gedaan: als voormeting, na elke module en als nameting. Het formulier was blanco bij de eerste en tweede module en bij de voor- en nameting. Het ontwikkelen van een gestructureerd peerassessment-formulier was onderdeel van de training in de derde module. Studenten waren dus in de gelegenheid deze kennis te gebruiken bij de derde module en de nameting. Na elke module werd aan de beoordeelde student het peerassessment-formulier overhandigd. Een kopie werd gebruikt voor analyse van de onderzoeksvragen. De studenten ontvingen geen studiepunten voor de kwaliteit van hun peerassessments. Studiepunten werden pas toegekend als studenten actief deelnamen aan de PA-taken en de feedbacksessie na elke module.

- *Beoordelingsformulier.* Om de kwaliteit van de peerassessments van de studenten te onderzoeken en te ontdekken of de training tot effecten zou leiden op de beoordelingsvaardigheid is een beoordelingsformulier ontwikkeld met een aantal dichotome variabelen. Een eerdere studie is gebruikt voor de definitie van dit beoordelingsformulier (Sluijtmans, Brand-Gruwel, Martens & Van Merriënboer, 2002). In het beoordelingsformulier zijn drie variabelen opgenomen die corresponderen met de drie beoordelingsvaardigheden in het peerassessment-model. Deze variabelen zijn: 1) het gebruik van criteria, 2) het geven van feedback en 3) het schrijven van een assessmentrapport. De variabele 'geven van feedback' is vervolgens geoperationaliseerd in vier variabelen: positieve commentaren, negatieve commentaren, constructieve commentaren en het stellen van vragen. De variabele 'schrijven van een assessmentrapport' is verder geoperationaliseerd in de variabelen structuur en naïef woordgebruik.

Om de variabele 'gebruik van criteria' te meten, is onderzocht of de student de juiste criteria toepaste. Elke student kon een maximum score van 19 behalen, aangezien in de eerste module 19 criteria voor een goed reflectieverslag door de studenten en de docent samen waren geformuleerd. De variabele 'geven van feedback' is gemeten door in het assessmentrapport van elke student de positieve, negatieve en constructieve commentaren te turven als ook het aantal gestelde vragen aan de beoordeelde student. In het beoordelingsformulier zijn zes items opgenomen om de variabele 'schrijven van een assessmentrapport' te meten. Eén item was bijvoorbeeld: 'Start de student zijn beoordeling met een positieve opmerking?' De variabele naïef woordgebruik is gemeten door het tellen van woorden als leuk, goed en prima.

Ten slotte zijn nog drie dichotome items opgenomen om additionele informatie te krijgen over de effecten van de training. Eén item is alleen gemeten bij de voor- en nameting: 'Beoordeelt de student het betreffende reflectieverslag als voldoende of onvoldoende?' De twee andere items waren: 'Werkt de student de criteria uit?' en 'Vermeldt de student eigen leerervaringen?' (zie verder Design en procedure).

Twee student-assistenten scoorden vervolgens onafhankelijk van elkaar alle kwalitatieve peerassessment-rapporten met behulp van het beoordelingsformulier. Vooraf zijn deze student-assistenten gedurende een dag in het gebruik van het formulier getraind. Vervolgens is de interraterbetrouwbaarheid bepaald. Deze bleek acceptabel voor elke variabele (Cohen's Kappa > .95).

- *Summatieve toets.* Om een effect op de kwaliteit van de reflectieve verslagen te meten, zijn de cijfers – gegeven door de docent – van de eerste en laatste reflectieverslagen geanalyseerd. De score kon variëren van 0 tot 100. Een score van 75 is gegeven als de student de criteria had verwerkt in het reflectieverslag. Extra punten zijn verwerkt als de criteria goed waren uitgewerkt. Als bepaalde aspecten rondom structuur ontbraken, zijn vijf punten per gemist aspect afgetrokken. Punten zijn tevens afgetrokken als het verslag te oppervlakkig was.

- *Studentvragenlijst.* Ter voor- en nameting hebben de studenten een vragenlijst ingevuld met betrekking tot hun percepties op wiskunde, onderwijs en assessment. Gebaseerd op het literatuuronderzoek en eerdere studies (Sluijsmans e.a., 1999; Sluijsmans e.a., 2001) zijn twaalf variabelen gedefinieerd met in totaal 116 items. De studenten hebben de vragenlijst beantwoord op een vijfpunts Likert-schaal, van 'helemaal mee oneens' tot 'helemaal mee eens'. De voormeting is georganiseerd om de ervaringen van studenten ten aanzien van wiskunde en toetsing tot dan toe te onderzoeken. De nameting richtte zich op de ervaringen van studenten na de drie modules en het peerassessment. Betrouwbaarheidsanalyses lieten zien dat elf items tot een aanzienlijke verlaging van de betrouwbaarheidscoëfficiënt leidden. Deze items zijn verwijderd uit de vragenlijst. De variabelen, aantal items, coëfficiënten en voorbeelden van items zijn opgenomen in Tabel 1.

- *Docentinterview, observaties en de reflectieve verslagen.* Additionele kwalitatieve data zijn verzameld ter ondersteuning van de kwantitatieve resultaten. Een gestandaardiseerd open interview is gebruikt om de reflecties van de wiskundedocent met betrekking tot de modules, het peerassessment en de feedbacksessies in kaart te brengen (Patton, 1990). Tijdens de feedbacksessies, waarin de studenten elkaar mondeling voorzagen van feedback, is door de onderzoeker geobserveerd. Enkele citaten uit de reflectieverslagen zijn gebruikt om de leerervaringen van studenten te illustreren.

Tabel 1 Variabelen, aantal items, betrouwbaarheidscoëfficiënten en voorbeelditems uit de studentvragenlijst

Variabele	#	?	Beschrijving	Voorbeeld item
Rol van de wiskundedocent	7	.6 .5	Perceptie op de activiteiten die een wiskundedocent in de lerarenopleiding vervult	De wiskundedocent moet creatief denken stimuleren
Zelfbeeld	16	.74	Perceptie op eigen functioneren binnen het wiskundedomein	Ik ben in staat om zelfstandig aan wiskundeopdrachten te werken
Visie op wiskunde en didactiek	16	.68	Perceptie op wiskunde en didactiek	Leerlingen in de basisschool leren wiskunde door samen te werken met andere leerlingen
Mathematische vaardigheden	7	.85	Perceptie op eigen mathematische vaardigheden	Ik ben in staat om andere studenten te helpen met wiskundige problemen
Toetsangst	6	.86	Perceptie op de mate van toetsangst	Ik ben doorgaans nerveus voor toetsen
Openbaarheid van assessment	5	.60	Perceptie op de mate waarin wiskundetentamens openbaar zijn	De toetsvragen moeten voor de toets plaatsvindt voor studenten bekend zijn
Voorspelbaarheid van assessment	4	.77	Perceptie op de mate waarin wiskundetentamens te voorspellen zijn	Ik kan voor de toets inschatten of ik deze zal halen of niet
Betrokkenheid in assessment	7	.82	Perceptie op de mate van betrokkenheid in assessment	Ik vind dat studenten meer moeten worden betrokken bij het vaststellen van assessmentcriteria
Groepsgedrag	4	.68	Perceptie op groepsgedrag van medestudenten	Ik vind het vervelend als medestudenten geen actieve bijdrage leveren aan een groepsproduct
Samenwerkend leren	8	.84	Perceptie op samenwerkend leren	Ik heb de voorkeur problemen met medestudenten uit te werken
Peerassessment	7	.74	Perceptie op peerassessment	Peerassessment is nuttig
Assessmentvaardigheid	18	.84	Perceptie op de assessmentvaardigheid	Ik ben in staat een product van een medestudent te analyseren

DESIGN EN PROCEDURE

Studenten participeerden voor een periode van een aantal maanden in het onderzoek. In deze periode volgden zij drie wiskundemodules met een geïntegreerde assessmenttraining met betrekking tot het schrijven van een reflectieverslag. Voorafgaand aan de eerste module is als voormeting een intake-bijeenkomst georganiseerd waarin de studenten drie activiteiten hebben uitgevoerd: het invullen van de vragenlijst, het schrijven van een persoonlijk reflectief verslag en het beoordelen van een anoniem reflectieverslag. Dit anonieme verslag van een eerstejaars student uit een voorafgaand schooljaar is vooraf aan het onderzoek door de wiskundedocent beoordeeld als onvoldoende. Het was een bewuste keuze om te kiezen voor een zwak reflectieverslag voor de voormeting, omdat studenten vaak geneigd zijn een product dat ze nog niet zo goed kennen hoger te beoordelen dan het daadwerkelijk is. Na de intake-bijeenkomst volgden de studenten de drie modules en de bijbehorende peerassessment-training. Deze modules waren de eerste wiskundemodules binnen de pabo-opleiding. Tijdens de modules werkten de studenten aan inhoudsgebonden taken en de PA-taken. Aan het einde van elke module schreven de studenten een reflectief verslag dat verbeterd kon worden na de tweede en derde module. Twee weken na de derde module leverde elke student zijn definitieve verslag in bij de docent voor een summatieve beoordeling. De assessmenttraining was gericht op drie thema's: wat zijn belangrijke criteria voor een reflectief verslag? (vier taken in module I), hoe geef ik feedback? (twee taken in module II) en hoe schrijf ik een goed assessmentrapport? (twee taken in module III). In de derde module bijvoorbeeld ontwikkelden de studenten een peerassessment-formulier op basis van een expertbeoordeling van de docent.

De output van het eerste deel van de training was een lijst van 19 criteria voor een verslag. Studenten kwamen in onderhandeling met de docent tot overeenstemming dat een goed verslag bijvoorbeeld zelfkritiek moet laten zien, stage-ervaringen, persoonlijke verwachtingen en sterktes/zwaktes. In het tweede deel van de training ontwikkelden studenten richtlijnen voor het geven van constructieve feedback. Eén richtlijn was bijvoorbeeld dat het positief zou zijn als studenten in hun beoordeling aangaven wat ze geleerd hadden van het verslag dat ze moesten beoordelen. In het laatste deel van de training werkten studenten aan een peerassessment-formulier en besloten ze wat belangrijk was in de structuur van een dergelijk formulier. Een expertbeoordeling vormde voor deze discussie de basis. De docent gaf duidelijk aan dat de geformuleerde criteria, feedbackregels en structuur-richtlijnen niet alleen behulpzaam zouden zijn bij het schrijven van de peerbeoordeling, maar ook bij het schrijven van het eigen reflectieverslag.

Na elke module hebben de studenten met behulp van de faciliteiten binnen Blackboard de reflectieverslagen uitgewisseld. Iedere student beoordeelde het verslag van een medestudent aan de hand van een 'wie-beoordeelt-wie'-schema, vastgesteld door de docent. Dit schema veranderde na elke module zodat elke student andere medestudenten beoordeelde en door andere medestudenten werd beoordeeld. De studenten haalden het verslag van de toegewezen medestudent af van Blackboard. Deze procedure maakte de studenten wederzijds afhankelijk, omdat de studenten geen beoordeling konden schrijven als het verslag niet tijdig in Blackboard stond.

De studenten schreven de beoordelingen thuis in eigen tempo. Na elke module werd in

groepen van tien tot twaalf studenten een feedbacksessie georganiseerd, onder leiding van de docent. In deze sessie presenteerde elke student zijn assessmentrapport. Dit rapport nam de beoordeelde student mee na de sessie, om het vervolgens te gebruiken bij het bijstellen voor een volgende versie. De peerfeedback kan worden beschouwd als een formatief assessment. Om toetsangst te verminderen en de periode te verlengen om een goed verslag te schrijven, ontvingen de studenten daarom nog geen cijfer voor hun verslagen tot het moment na de derde module. De rol van de docent in de feedbacksessie was die van voorzitter en begeleider.

Twee weken na de derde module leverden alle studenten hun definitieve verslag in bij de docent. Na de derde feedbacksessie is een outtake-bijeenkomst georganiseerd, in dezelfde opzet als de intake-bijeenkomst. In de outtake-bijeenkomst hebben de studenten wederom de vragenlijst ingevuld en hebben ze hetzelfde anonieme verslag als in de intake-bijeenkomst beoordeeld.

DATA-ANALYSES

Twee student-assistenten analyseerden 550 peerassessment-formulieren, verzameld op vijf testmomenten: 220 formulieren van de voor- en nameting en 330 verzameld na de modules. Op de eerste plaats zijn gemiddelden en standaarddeviaties berekend voor elke variabele van het beoordelingsformulier met betrekking tot de peerassessment-rapporten van de voor- en nameting. De data van de peerassessment-rapporten geschreven in de intake- en outtake-bijeenkomst zijn gescheiden geanalyseerd van de andere drie testmomenten. De reden hiervoor is dat de condities waaronder deze rapporten geschreven zijn anders waren bij de voor- en nameting dan bij de tussenliggende metingen. Studenten hadden bij de voor- en nameting minder tijd (half uur) terwijl de studenten op de tussenliggende tijdstippen hun eigen tijd konden bepalen.

Een non-parametrische test voor twee gerelateerde steekproeven (Wilcoxon) is gebruikt om significante effecten te onderzoeken. Frequenties zijn berekend voor de drie dichotome items. Chi-square testen (McNemar) zijn uitgevoerd om significante effecten op deze drie items te onderzoeken.

De scores op elke variabele van de tussenliggende peerassessment-rapporten ($n = 330$) zijn geanalyseerd met een binnen-subjecten variantieanalyse met herhaalde meting op de factor 'tijdstip van testen'. Univariate variantieanalyses zijn toegepast om significante effecten te identificeren na elk trainingsdeel. Gepaarde t -toetsen zijn uitgevoerd om verschillen tussen de leerresultaten van de reflectieverslagen na de voor- en nameting te onderzoeken.

Gemiddelden en standaarddeviaties zijn berekend voor de twaalf variabelen van de studentvragenlijst. Gepaarde t -toetsen zijn uitgevoerd om verschillen tussen de voor- en nameting te onderzoeken.

RESULTATEN

Onderzoeksvraag 1: De effecten van de assessmenttraining

Effecten op de beoordelingsvaardigheid. De peerassessment-rapporten uit de voor- en nameting zijn geanalyseerd om te kijken of over de drie modules heen effecten waren. In tabel twee zijn de gemiddelden en standaarddeviaties van elke variabele met betrekking tot de voor- en nameting weergegeven. De non-parametrische test voor twee gerelateerde steekproeven (Wilcoxon) liet zien dat de training een positieve invloed heeft gehad op zes van de zeven variabelen. In de voormeting bijvoorbeeld gebruikten studenten minder dan een criterium ($M=.57$), terwijl dit in de nameting significant hoger is ($M=5.0$). Omdat er geen significant effect gevonden is op de variabele 'naïef woordgebruik' is deze variabele uitgesloten van verdere analyses.

Tabel 2 Gemiddelden, standaarddeviaties en z-waarden van de assessmentformulieren na de voormeting en nameting

Variabele	Voormeting		Nameting		
	M	SD	M	SD	z
Definiëren van criteria					
- gebruik van criteria ^a	0.57	.70	5.00	4.32	-8.15*
Geven van feedback					
- positieve commentaren	0.86	.34	1.41	1.36	-3.77*
- negatieve commentaren	1.17	.73	2.73	2.22	-6.37*
- constructieve commentaren	0.29	.46	1.96	1.88	-7.25*
- gestelde vragen	0.12	.33	1.69	0.81	-4.26*
Schrijven van een assessmentrapport					
- structuur ^b	.075	.45	2.46	1.21	-8.77*
- naïef woordgebruik	1.62	.76	2.17	1.86	-.17
* $p < .001$.					
^a maximumscore = 19					
^b maximumscore = 6					

Resultaten van de berekende frequenties op het dichotome item 'Beoordeelt de student het reflectief verslag als voldoende of onvoldoende?' laten zien dat in de voormeting 28,2% van de studenten het verslag als onvoldoende beoordeelt ten opzichte van 66,4% in de nameting. Resultaten van de Chi-square testen (McNemar) tonen aan dat dit verschil significant is ($p=.000$). De studenten zijn negatiever over het anonieme verslag in de nameting dan in de voormeting.

In tabel 3 staan de gemiddelden en standaarddeviaties na elke module vermeld van de variabelen waarop significante resultaten zijn gevonden in tabel 2. Significante hoofdeffecten zijn gevonden op alle variabelen voor de factor 'tijdspit van testen'. Univariate ANOVA's laten zien welke toename of afname in de gemiddelde aantallen significant was.

Tabel 3 Gemiddelden en standaarddeviaties op de variabelen van het scoringsformulier na elke module en de mean square errors, vrijheidsgraden en F-waarden voor het hoofdeffect op factor 'tijdstip van testen'

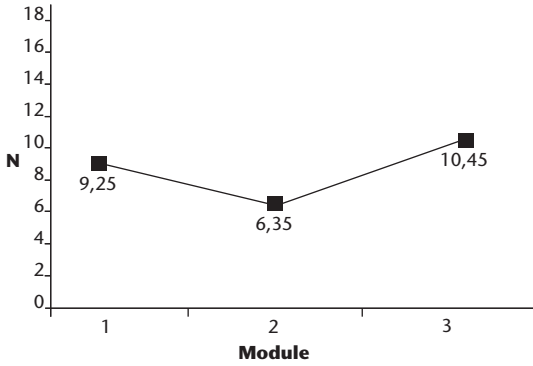
Variabele	Module I		Module II		Module III				
	M	SD	M	SD	M	SD	MSE	df	F
Definiëren van criteria - gebruik van criteria ^a	9.25	4.60	6.35	4.68	10.45	5.27	19.31	(2, 240)	24.97*
Geven van feedback									
- positieve commentaren	2.75	2.94	3.31	2.67	6.08	3.72	7.60	(2, 210)	45.64*
- negatieve commentaren	4.28	4.03	2.50	2.73	4.04	2.73	8.26	(2, 208)	13.23*
- constructieve commentaren	2.05	1.01	2.28	1.81	2.96	2.72	4.41	(2, 210)	24.79*
- gestelde vragen	1.71	.69	1.92	.88	3.56	1.96	5.07	(2, 210)	10.56*
Schrijven van een assessmentrapport ^b									
- structuur	2.07	.97	2.55	.90	4.19	1.44	1.18	(2, 210)	109.54*

* = $p < .001$
^amaximumscore = 19
^bmaximumscore = 6

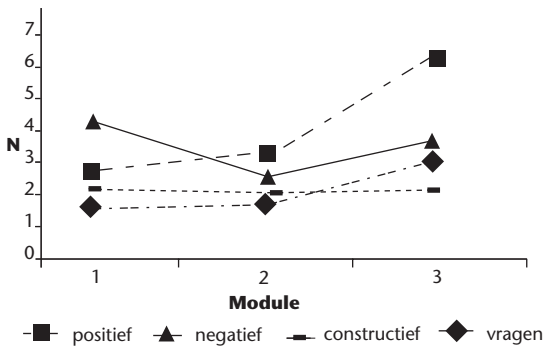
De patronen over de tijd heen van alle variabelen die uitspraken doen over de ontwikkeling van de drie assessmentvaardigheden zijn weergegeven in de figuren 2a, 2b en 2c. In figuur 2a is te zien hoe de ontwikkeling is met betrekking tot de vaardigheid 'gebruik van criteria'. In figuur 2b is het verloop van de vier variabelen met betrekking tot het geven van feedback te zien en in figuur 2c kan worden afgelezen hoe de ontwikkeling is met betrekking tot het schrijven van een assessmentrapport.

Voor de twee resterende dichotome items zijn frequenties berekend. Uit analyse van de beoordelingen na de eerste module blijkt dat 63,6% van de studenten de criteria uitwerkten oftewel beargumenteerden waarom een criterium al dan niet op het reflectieverslag van toepassing was, maar dat dit percentage zakte naar 28.2% na de tweede module. 36,4% respectievelijk 71,2% van de studenten noemden de criteria vaak wel, maar lichtte deze niet verder toe in de beoordeling. Chi-square testen (McNemar) toonden aan dat deze afname significant was ($p = .000$). Het percentage steeg weer na de derde module tot 69.1% ($p = .000$). Met betrekking tot het item 'Vermeldt de student eigen leerervaringen?' is een positief antwoord gevonden bij 3.6% van de studenten na de eerste module, 41.8% na de tweede module ($p = .000$) en 56.4% na de derde module ($p = .036$).

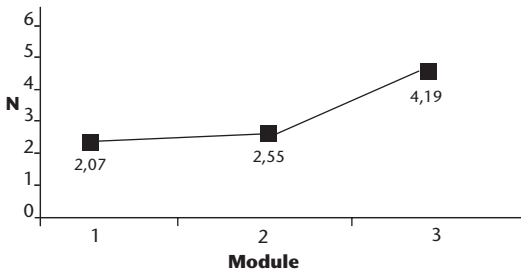
Effect op de kwaliteit van de reflectieverslagen. De wiskundedocent heeft de reflectieverslagen die de studenten hebben geschreven in de intake-bijeenkomst en de uiteindelijke verslagen na de derde module beoordeeld. De gemiddelde score van de laatste verslagen was 70.77 ($SD = 7.44$) vergeleken met een gemiddelde score van de verslagen geschreven in de intake-bijeenkomst van 46.80 ($SD = 13.19$), $t(1,96) = -16.59$, $p < .001$. De training en de peerfeedback zouden geleid kunnen hebben tot betere reflectieverslagen.



Figuur 2a Ontwikkeling van de vaardigheid 'definiëren van criteria'



Figuur 2b Ontwikkeling van de vaardigheid 'geven van feedback'



Figuur 2c Ontwikkeling van de vaardigheid 'schrijven van een assessmentrapport'

Onderzoeksvraag II: Verandering in perceptie

In tabel vier staan de gemiddelden en standaarddeviaties op de twaalf variabelen van de studentvragenlijst weergegeven.

Tabel 4 Gemiddelden en standaarddeviaties op de vragenlijstresultaten van de voor- en nameting (vijfpunts Likert-schaal)

		Voormeting	Nameting	<i>t</i>
Rol van de wiskundedocent	<i>M</i>	3.32	3.59	-4.76*
	<i>SD</i>	.46	.47	
Zelfbeeld	<i>M</i>	2.30	2.41	-1.87
	<i>SD</i>	.48	.49	
Visie op wiskunde en didactiek	<i>M</i>	2.35	3.14	-13.38*
	<i>SD</i>	.40	.48	
Mathematische vaardigheden	<i>M</i>	3.33	3.59	-2.58*
	<i>SD</i>	.84	.74	
Toetsangst	<i>M</i>	2.98	2.58	3.80*
	<i>SD</i>	.88	.84	
Openbaarheid van het assessment	<i>M</i>	3.35	3.56	-1.70
	<i>SD</i>	.67	1.21	
Voorspelbaarheid van het assessment	<i>M</i>	2.73	3.04	-2.77*
	<i>SD</i>	.85	.90	
Betrokkenheid in assessment	<i>M</i>	3.14	3.26	-1.29
	<i>SD</i>	.74	.68	
Groepsgedrag	<i>M</i>	4.22	4.37	-2.42
	<i>SD</i>	.52	.48	
Samenwerkend leren	<i>M</i>	3.40	3.25	1.67
	<i>SD</i>	.72	.67	
Peerassessment	<i>M</i>	4.23	4.27	-.77
	<i>SD</i>	.42	.42	
Assessmentvaardigheid	<i>M</i>	3.62	3.87	-4.61*
	<i>SD</i>	.43	.41	
* <i>p</i> <.001				

Resultaten op de gepaarde *t*-toetsen laten zien dat de studenten veranderden in hun perceptie op tien van de twaalf variabelen. Na de training waren ze positiever over de rol van de docent, hadden ze een positiever beeld ten aanzien van wiskunde en didactiek, hadden ze een positiever zelfbeeld en voelden ze zich vaardiger in wiskunde. Hun toetsangst is significant afgenomen en zij waarden het assessment als voorspelbaarder en opener. De perceptie ten opzichte van het groepsgedrag is positief veranderd. Tevens voelden de studenten zich meer vaardig in het beoordelen van elkaar. Vreemd genoeg zijn de studenten niet positiever gaan staan tegenover samenwerkend leren. Ten aanzien van de variabelen peerassessment en de betrokkenheid in assessment is het opvallend dat geen verandering is opgetreden. Dit kan echter te maken hebben met de reeds hoge score op de voormeting (resp. *M*= 4.23 en *M*= 3.14).

CONCLUSIE EN DISCUSSIE

Het gepresenteerde longitudinale onderzoek was opgezet om meer inzicht te krijgen in de ontwikkeling van beoordelingsvaardigheden bij eerstejaars pabo-studenten en hun vaardigheid in het schrijven van reflectieverslagen als effecten van een geïntegreerde training in peerassessment binnen de context van het wiskundedomein. Tevens zijn de percepties van studenten voor en na de training vastgesteld. De verwachtingen waren dat de training zou leiden tot een verbeterde beoordelings- en reflectievaardigheid. In het volgende worden de resultaten op een rijtje gezet en besproken.

Samenvatting resultaten

Het meest interessante deel van het onderzoek was het onderzoek naar de effecten op de beoordelingsvaardigheid, aangezien dit het primaire doel was van de training. Vergelijking tussen de resultaten van de voormeting en nameting toonden significante effecten op bijna alle variabelen. De tijd die beschikbaar was voor het schrijven van het assessmentrapport in de intake- en outtake-bijeenkomst was beperkt, wat een verklaring kan zijn voor het feit dat de studenten maar vijf criteria gebruikten in de nameting, terwijl een maximum van 19 kon worden gehaald. Dit geldt ook voor de variabele 'structuur', waarbij slechts een score van 2.46 in de nameting werd gevonden, terwijl een maximum van zes kon worden behaald.

Uit de eerste analyses bleek dat er geen significant effect was op de variabele 'naïef woordgebruik'. Een eerste verklaring hiervoor kan zijn dat als studenten langere assessmentrapporten schrijven, het een logisch gevolg is dat er meer naïeve woorden worden gebruikt. Op de tweede plaats kan het worden toegewezen aan het beperkte woordgebruik van eerstejaars studenten in dit type activiteiten, waardoor al snel woorden als 'leuk' en 'goed' worden toegepast. Een derde verklaring kan zijn dat naïef woordgebruik niet noodzakelijkerwijs betekent dat studenten minder goed beoordelen. Als een student immers de juiste criteria toepast, constructieve feedback kan geven en een gestructureerd assessmentrapport aflevert, is het van secundair belang of daar veel of weinig naïeve woorden bij worden gebruikt. Om bovenstaande drie redenen is het te overwegen om de variabele 'naïef woordgebruik' als overbodig te beschouwen in het bepalen van de kwaliteit van de beoordelingsvaardigheid.

Uit de resultaten bleek tevens dat studenten in de nameting negatiever waren over het anonieme reflectieverslag dan bij de voormeting. Dit impliceert een meer kritische houding na de training. Tijdens het schrijven van het assessmentrapport van het reflectieverslag dat zij ook al hadden beoordeeld in de intake-bijeenkomst, maakten studenten opmerkingen als: *'Deze student heeft niet de criteria gebruikt die wij bepaald hebben voor een goed reflectieverslag'*.

De figuren 2a, 2b en 2c illustreerden de ontwikkeling van studenten in de drie beoordelingsvaardigheden. De effecten zijn duidelijk zichtbaar. Studenten gebruikten meer criteria na het eerste trainingsdeel over het definiëren van criteria. Dit daalde echter weer na het tweede trainingsdeel in de tweede module. Een verklaring kan zijn dat de criteria wat wegzakten na de tweede module, omdat in de tweede module vooral het accent lag op het geven van feedback. Samenvattend kan worden gesteld dat de studenten een hoger niveau op alle vaardigheden hebben bereikt in vergelijking met het begin. Daarnaast bleken ze ook betere reflectieverslagen te schrijven.

Toetsangst en wiskunde

Positieve effecten zijn gevonden met betrekking tot de veranderde perceptie van studenten. Opmerkingen van de studenten geven aan dat zij zich steeds gemakkelijker voelden om het peerassessment uit te voeren. Een belangrijk resultaat is dat de toetsangst verminderde. Studenten die reeds een negatieve houding tegenover wiskunde hadden ontwikkeld veranderden in hun mening door de reflectieverslagen en het peerassessment. De houding van studenten tegenover wiskunde verschilt vaak aanzienlijk met hun houding ten opzichte van andere domeinen, zoals bijvoorbeeld pedagogiek en geschiedenis. Eerstejaars studenten worden behoorlijk beïnvloed door hun ervaringen met wiskunde in de middelbare school. Sommige studenten vinden wiskunde oprecht interessant, maar vele andere hebben minder positieve ervaringen (Ashcraft, Kirk, & Hopko, 1998; Fennema & Sherman, 1978; Jacobs, Watson & Sutton, 1996). Aanstaaende leerkrachten met negatieve gevoelens met betrekking tot wiskunde, zijn moeilijk veranderbaar in hun perceptie (Sullivan, 1989). Op zich verrassend, aangezien ze best wel weten dat ze kiezen voor een beroep waar ook het geven van wiskunde en rekenen een belangrijk aspect is (Kelly & Tomhave, 1985; Trujillo & Hadfield, 1999; Watson, 1987). De feedbacksessies in het onderzoek kunnen een ondersteunende rol hebben gespeeld in het reduceren van de toetsangst en het verbeteren van mathematische vaardigheden.

Een andere verklaring voor de afname van de toetsangst kan zijn dat studenten niet meteen na de eerste module werden afgerekend op hun reflectieverslag. De peerfeedback is gebruikt als een vorm van formatieve assessment, wat minder bedreigend voor ze was. Hoewel de docent de feedbacksessies wel bijwoonde, hadden studenten het gevoel dat ze niet meteen op een voldoende niveau moesten scoren bij de eerste versie. Het grootste knelpunt in structurele veranderingen in instructie en assessment is vaak dat studenten leren voor de toets en dat docenten doceren voor de toets. Madaus (1988) illustreert dit fenomeen op het niveau van het curriculum: *'Het is de toetsing, niet zozeer het officieel vastgestelde curriculum, dat in toenemende mate bepaald wat wordt onderwezen, hoe het is onderwezen, wat wordt geleerd en hoe het wordt geleerd'* (p. 83). Wanneer gekozen wordt voor een meer formatieve assessmentbenadering, wordt toetsgedrag gereduceerd. Dalziel (1998) voegt hieraan toe dat het van belang is dat studenten worden geïnformeerd over de wijze waarop de toetsing is ingericht in modules en dat zij worden betrokken in beslissingen hierover. Met andere woorden, assessment is veel meer dan alleen een pen-en-papiertoets.

Mondelinge feedback

Tijdens de feedbacksessies na elke module, rapporteerden de studenten in groepen van tien tot twaalf studenten hun assessmentrapporten mondeling aan de betreffende medestudent. In de eerste feedbacksessie voelden studenten zich nog erg onzeker. Enkele studenten waren erg nerveus en vertoonden fysiologische verschijnselen als zweten, stotteren en rode wangen. Duidelijk was dat studenten niet gewend waren elkaar op deze wijze van feedback te voorzien. In de tweede en derde sessie werden deze signalen beduidend minder. Zoals één student aangeeft: *'In het begin was het erg vreemd om feedback te geven aan een medestudent, maar langzamerhand werd het gemakkelijker om het te doen'*.

Studenten bleken in de tweede en derde feedbacksessie meer in staat te zijn om hun feedback direct te richten aan de beoordeelde student en stelden zich minder afhankelijk op van de docent. Het is opvallend dat de studenten hun verantwoordelijkheid namen in het beoordelingsproces en telkens een uitgebreid assessmentrapport schreven, vooral omdat dit niet meetelde voor studiepunten. De docent verklaart dit als volgt: *'De studenten werkten erg serieus aan hun taken, ik weet niet precies waarom. Collega's klagen juist vaak over de werkhouding van studenten. Misschien kwam het wel omdat ik vertrouwen uitstraalde naar de studenten door hun te laten weten dat hun input waardevol was. Een andere verklaring kan zijn dat ze zich wederzijds afhankelijk voelden. Als bijvoorbeeld een student zijn assessment rapport niet schreef, zijn reflectieverslag niet op tijd in Blackboard plaatste of geen enkele constructieve bijdrage in de groep leverde, werd deze student aangesproken door zijn medestudenten. Wat mijn rol betreft, was het fijn om te merken dat ze niet meer meteen naar mij toe renden. De rol die ik vaak had, verschoof nu naar de studenten zelf en dat is een goede zaak'.*

Beperkingen van het onderzoek

Hoewel er veelbelovende resultaten uit het onderzoek zijn gekomen, kan er ook een aantal beperkingen van de gepresenteerde studie worden genoemd. Op de eerste plaats hebben de studenten in de periode dat ze de wiskundemodules volgden en de assessment taken uitvoerden, ook een aantal andere activiteiten binnen andere domeinen en binnen de stage-omgeving verricht. Het is derhalve moeilijk om met zekerheid vast te stellen of vooral de effecten op de kwaliteit van de reflectieverslagen toe te schrijven zijn aan de trainingsinterventie. Een eerdere studie toonde echter juist aan dat een korte trainingsperiode maar tot een kleine vooruitgang in beoordelingsvaardigheid en inhoudsgebonden vaardigheden leidde tussen controle- en experimentele groepen (Sluijsmans e.a., 2002), vandaar de langere opzet in deze studie. Dezelfde beperking geldt voor de verandering in perceptie. Deze kan ook zijn opgetreden door de invloed van andere leerervaringen buiten de betreffende modules of simpelweg door rijping.

Een tweede beperking ligt in de assessmentbenadering. Peerassessment is in deze studie vooral beschouwd als een leermiddel, niet als toetsvorm. Dit bleek echter niet even waardevol voor elke student. Een student die bijvoorbeeld al behoorlijk vaardig was in het schrijven van reflectieverslagen zag niet de meerwaarde in van het beoordelen van het reflectieverslag van een medestudent die nog niet zo vaardig was. Het omgekeerde kwam ook voor: een student die te weinig inzicht had in de criteria van een verslag of regels van goede feedback, was minder goed in staat een goed assessmentrapport te schrijven. De mening van de docent met betrekking tot deze observatie was dat studenten dan misschien niet direct voordeel zouden hebben van de feedback die ze *ontvingen* van studenten, maar dat ze zeker hebben geleerd van het *geven* van feedback. Een interessante vraag is daarom ook in hoeverre het schrijven van betrouwbare assessments kennis over het te beoordelen product vereist.

Op de derde plaats is nog onvoldoende zicht verkregen in de relatie tussen reflectie – dat kan worden gezien als een vorm van selfassessment – en peerassessment. In de inleiding wordt aangenomen dat reflectievaardigheden voorwaardelijk zouden zijn voor het uitvoeren van goede peerassessments. Het zou te ver voeren als gesteld zou worden dat die assumptie met deze studie is aangetoond. Het is wellicht denkbaar dat iemand

zonder reflectievaardigheid prima peerassessments kan uitvoeren waar de beoordeelde student veel aan heeft, of dat iemand met veel reflectievaardigheden niet in staat is het werk van andere studenten te beoordelen. De mogelijke causale relatie tussen reflectie en peerassessment is interessant genoeg om verder onderzoek in te verrichten. Een laatste beperking is de generaliseerbaarheid van deze studie. Studenten hebben geleerd hoe reflectieverslagen te schrijven binnen het wiskundedomein, maar er is niet onderzocht of deze bevindingen ook gelden voor andere inhoudsdomeinen. Het is interessant om te onderzoeken of studenten na een dergelijke interventie ook betere verslagen schrijven in andere domeinen.

TOT SLOT

Studenten ondersteunen in de ontwikkeling van reflectie- en beoordelingsvaardigheden is een belangrijk aspect in verdere onderwijsvernieuwingen binnen lerarenopleidingen richting meer competentiegebaseerde curricula. De resultaten van deze studie helpen bij het ontwerpen van het onderwijs. Peerassessment-taken en feedbacksessies stimuleren studenten meer betrokken te zijn bij de invulling van hun eigen onderwijs, in plaats van het allemaal over zich heen te laten komen. Leren elkaar te voorzien van feedback maakt studenten verantwoordelijker ten opzichte van elkaar. Juist dit verantwoordelijkheidsbesef is van belang voor het functioneren in de beroepscontext.

Samenvattend is door deze studie weer een stap genomen in de invulling van studentbetrokkenheid in het hoger onderwijs. Eén van de mogelijke, zoals ook uit de bijdragen van Joosten-ten Brinke e.a. en Van den Berg e.a. (dit nummer) blijkt. Een meer integratieve benadering van peerassessment betekent echter een investering op vele fronten van vele partijen. Het onderkennen van het belang van het definiëren van goede beoordelingscriteria door studenten en docenten is al een belangrijke stap. Criteria ten aanzien van gewenste output zijn vaak onduidelijk en worden nog te veel door individuele docenten bepaald. Zeker in meer competentiegerichte omgevingen, waarbij performance-assessment van integratieve beroepstaken steeds meer opkomt, is dit niet langer een individuele taak. In lerarenopleidingen is men het er bijvoorbeeld over eens dat studenten in staat moeten zijn te reflecteren (Boud, 1990; Reilly-Freese, 1999; Hayon-Kremer, & Tillema, 1999; Korthagen, 2001). Er blijven echter te veel uiteenlopende meningen bestaan over hoe dit concreet vorm moet krijgen, waardoor de studenten maar doen wat de docent vraagt. Een interventie zoals toegepast in deze studie is misschien een goede start de verschillende perspectieven helder te maken.

LITERATUUR

- Anderson, J.B. & Freiberg, H.J. (1995) Using self-assessment as a reflective tool to enhance the student teaching experience. *Teacher Education Quarterly*, 22, 77-91.
- Arter, J. (1996) Using assessment as a tool for learning. In: R. Blum & J. Arter (Eds.) *Student performance assessment in an era of restructuring* (pp. 1-6.) Alexandria, VA: Association for Supervision and Curriculum Development.

- Ashcraft, M.H., Kirk, E.P., & Hopko, D. (1998) On the cognitive consequences of mathematics anxiety. In: C. Donlan (Ed.), *The development of mathematical skills*. Hove, England: Taylor & Francis.
- Bangert, A.W. (1995) *Peerassessment: an instructional strategy for effectively implementing performance-based assessments*. University of South Dakota.
- Berg, B.A.M. van den , Admiraal, W., & Pilot, A. (2003) Peerassessment in university teaching: an exploration of useful designs. *Tijdschrift voor Hoger Onderwijs*, 4, 21, 251-271.
- Boud, D. (1990) Assessment and the promotion of academic values. *Studies in Higher Education*, 15, 101-111.
- Boud, D. (1995) *Enhancing learning through self-assessment*. London: Kogan Page.
- Dalziel, J. (1998) Using marks to assess student performance: some problems and alternatives. *Assessment and Evaluation in Higher Education*, 23, 351-366.
- Dochy, F.J.R.C. & McDowell, L. (1997) Assessment as a tool for learning. *Studies in Educational Evaluation*, 23, 279-298.
- Falchikov, N. (1995) Peer feedback marking: developing peerassessment. *Innovations in Education and Training International*, 32, 175-187.
- Falchikov, N., & Goldfinch, J. (2001) Student peerassessment in higher education: a meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70, 287-322.
- Fennema, E., & Sherman, J. (1978) Sex related differences in mathematics achievement and related factors: A further study. *Journal for Research in Mathematics Education*, 9, 189-203.
- Freeman, M. (1995) Peerassessment by groups of group work. *Assessment and Evaluation in Higher Education*, 20, 289-300.
- Hanrahan, S., & Isaacs, G. (2001) Assessing self- and peerassessment: the students' views. *Higher Education Research and Development*, 20, 53-70.
- Joosten-ten Brinke, D., Verhoeven, N., & Van Buuren, H. (2003) Peer- and selfassessment in the methodology curriculum: An innovation project of the Open University of the Netherlands. *Tijdschrift voor Hoger Onderwijs*, 4, 21 272-285.
- Kane, J.S., & Lawler III, E.E. (1978) Methods of peerassessment. *Psychological Bulletin*, 85, 555-586.
- Kelly, W.P., & Tomhave, W.K. (1985) The study of math anxiety/math avoidance in pre-service elementary teachers. *Arithmetic Teacher*, 32, 51-53.
- Korthagen, F.A.J. (1985) Reflective teaching and pre-service education in the Netherlands. *Journal of Teacher Education*, 36, 11-15.
- Korthagen, F.A.J. (2001) *Linking practice and theory: the pedagogy of realistic teacher education*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Korthagen, F., & Wubbels, T. (2000) Are reflective teachers better teachers? In: G. M. Willems, J. H. J. Stakenborg & W. Veugelers (Eds.), *Trends in Teacher Education* 131-141. Leuven-Apeldoorn: Garant.
- Kremer-Hayon, L., & Tillema, H.H. (1999) Self-regulated learning in the context of teacher education. *Teaching and Teacher Education*, 15, 507-522.
- Kwan, K., & Leung, R. (1996) Tutor versus peer group assessment of student performance in a simulation training exercise. *Assessment and Evaluation in Higher Education*, 21, 205-214.
- Loughran, J.J. (1996) *Developing Reflective Practice*. UK: Falmer Press.

- Madaus, G.F. (1988) The influence of testing on the curriculum. In: L.N. Tanner and K.J. Rehege (Eds.), *Critical issues in curriculum* (pp. 83-121) Chicago: The University of Chicago Press.
- Mehrens, W.A., Popham, W.J., & Ryan, J.M. (1998) How to prepare students for performance assessments. *Educational Measurement: Issues and Practice*, 17, 1, 18-22.
- Newman, S.J. (1996) Reflection and teacher education. *Journal of Education of Teaching*, 22, 297-310.
- Orsmond, P., Merry, S. & Reiling, K. (1996) The importance of marking criteria in the use of peerassessment, *Assessment and Evaluation in Higher Education*, 21, 239-249.
- Orsmond, P., Merry, S. & Reiling, K. (1997) A study in self-assessment: tutor and students' perceptions of performance criteria, *Assessment and Evaluation in Higher Education*, 22, 357-369.
- Orsmond, P., Merry, S. & Reiling, K. (2000) The use of student derived marking criteria in peer and self-assessment, *Assessment and Evaluation in Higher Education*, 25, 23-38.
- Patton, M.Q. (1990) *Qualitative evaluation and research methods*. Newbury Park, CA: Sage Publications.
- Reilly Freese, A. (1999) The role of reflection on preservice teachers' development in the context of a professional development school. *Teaching and Teacher Education*, 15, 895-909.
- Richert, A.E. (1999) Teaching teachers to reflect: a consideration of programme structure. *Journal of Curriculum Studies*, 22, 509-527.
- Rushton, C., Ramsey, P., & Rada, R. (1993) Peerassessment in a collaborative hypermedia environment: a case study. *Journal of Computer-Based Instruction*, 20, 75-80.
- Schön, D.A. (1987) *Educating the reflective practitioner: towards a new design for teaching and learning in the professions*. San Francisco, CA: Jossey-Bass.
- Sluijsmans, D., Brand-Gruwel, Martens, R., Van Merriënboer, J. (2002) Peerassessment training in teacher education. *Assessment and Evaluation in Higher Education*, 27, 5, 443-454.
- Sluijsmans, D., Dochy, F., & Moerkerke, G. (1999) Creating a learning environment by using self- peer- and co-assessment. *Learning Environments Research*, 1, 293-319.
- Sluijsmans, D., Moerkerke, G., Dochy, F., & Van Merriënboer, J.J.G. (2001) Peerassessment in problem based learning. *Studies in Educational Evaluation*, 27, 153-173.
- Sluijsmans, D., & Van Merriënboer, J.J.G. (2000) *A peerassessment-model*. Heerlen: Open University of the Netherlands, Center for Educational Technology and Expertise.
- Sluijsmans, D.M.A. (2002) *Student involvement in assessment. The training of peerassessment skills*. Unpublished doctoral dissertation, Open University of the Netherlands, Heerlen.
- Stefani, L. (1998) Assessment in partnership with learners. *Assessment and Evaluation in Higher Education*, 23, 339-350.
- Sullivan, P. (1989) Factors inhibiting change: A case study of a beginning primary teacher. *Research in Mathematics Education in Australia*, August, 1-9.
- Topping, K. (1998) Peerassessment between students in colleges and universities. *Review of Educational Research*, 68, 3, 249-276.

Trujillo, K. M., & Hadfield, O. D. (1999) Tracing the roots of mathematics anxiety through in-depth interviews with preservice elementary teachers. *College Student Journal*, 33, 219-232.

Watson, J. (1987) The attitudes of preservice primary teachers toward mathematics: Some observations. *Research in Mathematics Education in Australia*, August, 48-5.