

# Onderzoek naar de inhoudsvaliditeit van het performance-assessment in de hoofdfase van de bacheloropleiding fysiotherapie

M. Ophey, M. Maas, J. de Beer

## Samenvatting

**Inleiding:** Binnen de opleiding fysiotherapie van de Hogeschool van Arnhem en Nijmegen bestaat al jaren een vorm van performance-assessment in het eerste en tweede studiejaar, waarin kennis, vaardigheden en attitudes integraal getoetst worden in een gereduceerde beroepscontext. Deze gereduceerde beroepscontext wordt vormgegeven door de student een getrainde simulatiepatiënt aan te bieden. Doel van dit onderzoek was na te gaan in hoeverre dit vernieuwde toetsinstrument valide resultaten oplevert.

**Methode:** Het onderzoek is verricht vanuit een tweetal perspectieven. Ten eerste is alle vier docenten die het leerproces van de studenten gedurende zeven weken voorafgaand aan het performance-assessment intensief gevolgd hebben, van te voren gevraagd een voorspelling te doen (docent expected) ten aanzien van het te verwachten resultaat per student. Ten tweede is aan alle 117 tweedejaarsstudenten gevraagd door middel van een vragenlijst ook een voorspelling te doen (student expected) ten aanzien van het te verwachten resultaat. Deze voorspellingen zijn vervolgens vergeleken met de daadwerkelijke, tijdens het assessment behaalde resultaten (observed).

**Resultaten:** Van de vier docenten zijn voor 107 van de 117 studenten de voorspellingen ingewonnen (92%). Onder de studenten lag de respons op 49 (42%). De door docenten met een voldoende beoordeelde studenten hebben voor 87% kans ook daadwerkelijk een voldoende te behalen (sensitiviteit). De door docenten met een onvoldoende beoordeelde studenten hebben voor 58% kans ook daadwerkelijk een onvoldoende te behalen (specificiteit). De studenten die zichzelf met een voldoende hebben beoordeeld, hebben voor 89% kans ook daadwerkelijk een voldoende te behalen (sensitiviteit). De studenten die zichzelf met een onvoldoende hebben beoordeeld, hebben voor 50% kans ook daadwerkelijk een onvoldoende te behalen (specificiteit).

**Conclusie:** Het huidige performance-assessment honoreert studenten, die naar eigen inschatting of naar de inschatting van de betrokken docenten een voldoende moeten behalen, ook daadwerkelijk met een voldoende. Het huidige performance-assessment keurt studenten, die naar eigen inschatting of naar de inschatting van de betrokken docenten een onvoldoende moeten behalen, in onvoldoende mate ook met een onvoldoende af. Vervolgonderzoek is noodzakelijk om mogelijke activiteiten te definiëren die de specificiteit van het assessment verhogen. (Ophey M, Maas M, Beer J de. Onderzoek naar de inhoudsvaliditeit van het performance-assessment in de hoofdfase van de bacheloropleiding fysiotherapie. Tijdschrift voor Medisch Onderwijs 2006;25(2):88-95.)

## Inleiding

Binnen de opleiding fysiotherapie van de Hogeschool van Arnhem en Nijmegen bestaat al jaren een vorm van performance-assessment in het eerste en tweede

studiejaar, waarin kennis, vaardigheden en attitudes integraal getoetst worden in een gesimuleerde en daarmee ook gereduceerde beroepscontext. Deze gereduceerde beroepscontext wordt vormgegeven door

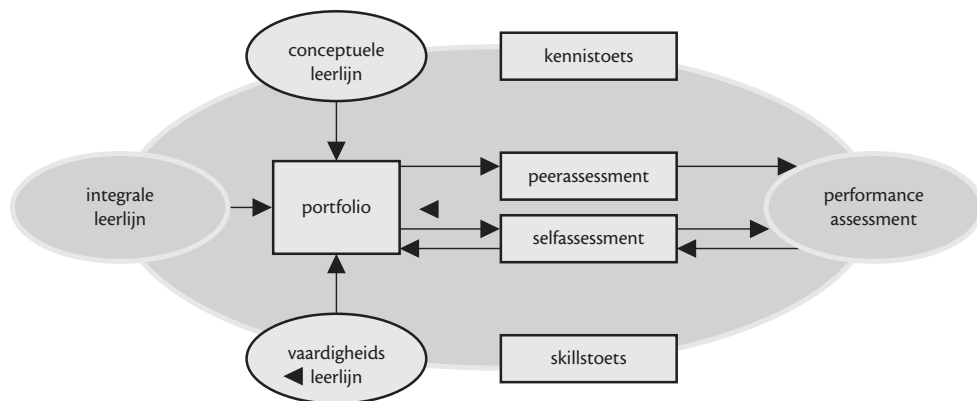
de student een getrainde simulatiepatiënt aan te bieden.<sup>1</sup>

Binnen de opleiding zijn drie leerlijnen te onderscheiden: de conceptuele leerlijn, de vaardigheidsleerlijn en de integrale leerlijn.<sup>2</sup> In de conceptuele leerlijn wordt achtergrondkennis aangeboden voor het oplossen van fysiotherapeutische problemen (anatomie, fysiologie, psychologie, et cetera) en onder andere getoetst middels een kennistoets. In de vaardigheidsleerlijn gaat het om het aanleren van praktische vaardigheden. Deze lijn is een aantal jaren geleden ontwikkeld volgens de skillslab-methode in samenwerking met het Transferpunt Vaardigheidsonderwijs Maastricht. Binnen de integrale leerlijn staat het fysiotherapeutisch methodisch handelen centraal. In deze lijn wordt volgens de probleemgestuurde onderwijsmethode gewerkt aan fysiotherapeutische problemen.<sup>3-4</sup> Hierbij wordt het meerdimensionale belasting-belastbaarheidmodel gehanteerd.<sup>5</sup> Het performance-assessment is het sluitstuk van de integrale leerlijn (figuur 1). De diverse leerlijnen komen samen in het portfolio waarin de student verantwoording aflegt over het gevolgde onderwijstraject.<sup>6</sup>

Vanuit de gedachte dat beoordelingscriteria tot op het kleinste detail geëxplici-

teerd moesten worden om tot een juiste beoordeling te kunnen komen, bestond het beoordelingsformulier aanvankelijk uit een grote (en daardoor ontmoedigende) hoeveelheid resultaats- en structuurcriteria op een dichotome schaal: voldaan of niet voldaan.<sup>7</sup>

Al ruim voordat het competentieren ook zijn intrede deed binnen de opleiding fysiotherapie werd duidelijk dat de som der delen het geheel niet kon dekken. Enerzijds moest de assessor veel aandacht besteden aan het beoordelen van de grote hoeveelheid criteria, waardoor deze nauwelijks op de redeneringen van studenten kon ingaan. Anderzijds waren de beschikbare criteria niet op redeneerstrategieën van toepassing. Dit was op zijn minst strijdig met de uitgangspunten van probleemgestuurd leren, waarin het hardop hypothetiseren een centrale plaats inneemt. Docenten voelden zich hierdoor miskend in hun intuïtieve oordeel over de competenties van de student en de essentie van het methodisch handelen, het klinisch redeneren, werd niet getoetst.<sup>8</sup> In dit verband is de uitspraak van O'Hearn: "We know what we do, but we do not know who we are" illustratief.<sup>9</sup> Een andere ontwikkeling waardoor de traditionele beoordelingsystematiek niet meer voldeed, is de evidence based practice.



Figuur 1 Integrale leerlijn, afdeling Fysiotherapie.

De implementatie daarvan in de integrale leerlijn veronderstelt metacognitieve vaardigheden om cruciale beslispunten in het fysiotherapeutisch handelen te benoemen, daarbij gebruikmakend van beschikbare wetenschappelijke evidentie.<sup>10</sup> Al met al werd het duidelijk dat er behoefte was aan meer ruimte voor het beoordelen van de totaliteit van de performance en aan de mogelijkheid om de criteria sterker te sturen op het expliciteren en legitimeren van het fysiotherapeutisch handelen binnen de gepresenteerde context.<sup>11</sup> Dit is de reden geweest om een nieuwe performance-assessment te ontwikkelen. In dit artikel worden het assessment en de resultaten van de validiteit van de methode beschreven. De onderzoeksvraag hierbij was of het nieuwe toetsinstrument valide resultaten oplevert. Met andere woorden: kan er met behulp van dit meetinstrument gedurende een performance-assessment onderscheid tussen bekwaame en niet-bekwaame studenten worden gemaakt?

## Methode

### *Beschrijving performance-assessment*

De eerste fase van het ontwikkelingsstraject had betrekking op de afstemming van het assessment op het competentieprofiel en het herformuleren en reduceren van beoordelingscriteria. De vijfpuntsschaal is ingevoerd om differentiatie in het beheersingsniveau te kunnen aanbrenge(n) (onvoldoende, twijfel, voldoende, ruim voldoende, goed). Het algemene beheersingsniveau is afgeleid van de fase-doelen van het tweede studiejaar. De complexiteit van het gesimuleerde probleem sluit daarbij aan. In deze praktijksituatie dient de student succesvol te handelen in een tweetal competentiegebieden met in totaal een vijftal competenties:

1. Competentiegebied 'diagnosticeren'
  - Hulpvraag verhelderen.

- Gegevens verzamelen door gericht lichamelijk onderzoek.
  - Fysiotherapeutische diagnose stellen.
2. Competentiegebied 'intervenieren'
    - Interventieplan opstellen.
    - Interventie uitvoeren.

Niet alle competenties worden in een assessment getoetst. Er wordt een keuze gemaakt voor een diagnostisch- of interventiegericht assessment.<sup>12</sup> In een diagnostisch scenario wordt de student bijvoorbeeld beoordeeld op de competenties 'gegevens verzamelen door gericht lichamelijk onderzoek' en 'fysiotherapeutische diagnose stellen'. Elke competentie is van een vijftal kritische kenmerken, inclusief een toelichting, voorzien. Bijvoorbeeld: de competentie 'gegevens verzamelen door gericht lichamelijk onderzoek' is geoperationaliseerd aan de hand van de volgende vijf kritische kenmerken: durven kiezen, zorgzaam handelen, systematisch handelen, vaardig handelen en beoordelen (tabel 1). Deze vijf kenmerken worden tijdens het assessment beoordeeld op een ordinale schaal van 1-5 (onvoldoende, twijfel, voldoende, ruim voldoende en goed). Daarmee kan de student per competentie maximaal 25 punten behalen. Het maximum aantal punten voor het assessment ligt daarmee bij 50 punten. Het absolute afkappunt voor de beoordeling ligt bij 30 punten (60% van de maximale score). Met andere woorden: de student dient minimaal 30 punten te behalen voor een voldoende beoordeling.

Deze kenmerken zijn in een tweetal audits gevalideerd. De eerste audit bestond uit een aantal bij het onderwijs betrokken docenten. In de tweede audit is het instrument voorgelegd aan een aantal fysiotherapeuten werkzaam in verschillende situaties in de eerste en tweede lijn. Het commentaar is verwerkt tot de huidige versie van het toetsinstrument.

Het tweede studiejaar heeft het generieke doel om te komen tot een beroepspe-

**Tabel 1** Kritische kenmerken van de competentie 'gegevens verzamelen door gericht onderzoek'.

	<b>Competentie: gegevens verzamelen door gericht lichamenlijk onderzoek</b> De student kan in een simulatiesituatie een lichamenlijk onderzoek uitvoeren om de gegevens en interpretaties van gegevens uit de anamnese te verifiëren, aan te vullen en/of te specificeren.
<b>Durven kiezen</b>	<i>Ontwerp een lichamenlijk onderzoek, afgestemd op de problematiek in de casus</i> • Kiest voor relevante lichamenlijke tests en meetinstrumenten op grond van casusgegevens.
<b>Zorgzaam handelen</b>	<i>Voer een onderdeel van je voorgenomen onderzoek uit</i> • Geeft ondubbelzinnige, bondige instructie. • Neemt hygiënische maatregelen. • Benadert de patiënt of proefpersoon zorgzaam en respectvol. • Neemt belevingsaspecten waar en neemt zonodig maatregelen.
<b>Systematisch handelen</b>	• Voert het onderzoek systematisch uit volgens de richtlijnen van het lichamenlijk onderzoek.
<b>Vaardig handelen</b>	• Kiest een passende onderzoekssituatie voor patiënt en therapeut, veilig en ergonomisch correct. • Maakt ondubbelzinnig lichamenlijk contact. • Heeft anatomisch correcte handvattingen en een vlot tempo.
<b>Beoordelen</b>	<i>Wat heb je waargenomen en hoe beoordeel je dat?</i> • Neemt aandachtig waar en verwoordt de bevindingen. • Beoordeelt bevindingen aan de hand van expliciete criteria.

cifieke verdieping van het primaire proces en voor te bereiden op de praktijkleerperiode in het derde studiejaar. In het kader van het mijlpalenbeleid van de opleiding ondergaat elke tweedejaarsstudent een performance-assessment halverwege en op het einde van het tweede studiejaar.

#### *Beoordeling van de validiteit van het toetsinstrument*

Het onderzoek naar de inhoudsvaliditeit is vanuit een tweetal perspectieven gedaan. Ten eerste zijn alle docenten uit de integrale leerlijn (n=4), die het leerproces van de studenten gedurende zeven weken voorafgaand aan het performance-assessment intensief gevolgd hebben, van tevoren gevraagd een voorspelling te geven (docent expected) ten aanzien van het te verwachten resultaat van de student. De voorspellingen zijn geformuleerd op een ordinale driepuntsschaal (onvoldoende,

twijfel, voldoende). Om voldoende contrast tussen de studenten te realiseren zijn bij de analyse van de data de 'twijfel'-beoordelingen buiten beschouwing gebleven.

Ten tweede is aan alle tweede jaarsstudenten (n=117) gevraagd om anoniem op basis van eigen inspanning in het voorgaande blok een voorspelling te geven ten aanzien van het te verwachten resultaat (student expected). Studenten konden zichzelf beoordelen met voldoende of onvoldoende.

De voorspellingen van docenten en studenten zijn vervolgens vergeleken met de daadwerkelijke tijdens het assessment behaalde resultaten (observed).<sup>13-14</sup>

De gegevens zijn in SPSS 11.5 verwerkt. Er is hierbij gebruik gemaakt van kruistabellen, correlatiecoëfficiënten en de Chi<sup>2</sup>-toets, waarbij een overschrijdingskans van 0,05 is aangehouden, tweezijdig getoetst is en bij kleinere celfrequenties gecorrigeerd is met de Fisher-Exact-toets.

## Resultaten

De 4 aan het onderzoek deelnemende docenten gaven een voorspelling voor 107 van de 117 studenten (92%). Bij 26 daarvan twijfelden zij over het te verwachten resultaat. Zodoende waren er voor 81 studenten voorspellingen in termen van 'voldoende' of 'onvoldoende'. Onder de studenten namen 49 van de 117 deel aan het onderzoek (42%).

In tabel 2 (kruistabel) wordt de relatie weergegeven tussen de voorspelling door de docent (docent expected) en het daadwerkelijke resultaat (result observed). Bij 52 van de met een voldoende beoordeelde studenten werd ook door de docenten verwacht dat deze studenten met een voldoende beoordeling het assessment zouden afronden. Bij 11 van de met een onvoldoende beoordeelde studenten werd ook door de docenten verwacht dat zij met een onvoldoende beoordeling het assessment zouden afronden.

**Tabel 2** Voorspelde resultaten door de docent en de daadwerkelijk behaalde resultaten.

	Docent expected		Totaal
	Voldoende	Onvoldoende	
<b>Result observed</b>			
Voldoende	52	8	60
Onvoldoende	10	11	21
<b>Totaal</b>	62	19	81

In tabel 3 (kruistabel) wordt de relatie weergegeven tussen de voorspelling door de student (student expected) en het daadwerkelijke resultaat (result observed). 31 van de met een voldoende beoordeelde studenten hebben ook verwacht een voldoende te behalen. Zeven van de met een onvoldoende beoordeelde studenten hadden ook verwacht een onvoldoende te behalen.

**Tabel 3** Voorspelde resultaten door de student en de daadwerkelijk behaalde resultaten.

	Student expected		Totaal
	Voldoende	Onvoldoende	
<b>Result observed</b>			
Voldoende	31	7	38
Onvoldoende	4	7	11
<b>Totaal</b>	35	14	49

Om uitspraken over de inhoudsvaliditeit van het assessment te kunnen doen, is gezocht naar de mate van overeenstemming tussen het daadwerkelijk behaalde resultaat en de voorspelling van de docent/studenten in de categorieën 'onvoldoende' en 'voldoende'. De mate van overeenstemming voor de docentvoorspellingen is 78%. Na correctie voor toevallige overeenstemming blijkt een Kappa van 0,40 met een  $p=0,003$ . De mate van overeenstemming voor de studentvoorspellingen is 78%. Na correctie voor toevallige overeenstemming blijkt een Kappa van 0,41 met een  $p=0,003$ .

Verder zijn nog sensitiviteit en specificiteit berekend. De door docenten met een voldoende beoordeelde studenten, hebben voor 87% kans ook daadwerkelijk een voldoende te behalen (sensitiviteit). De door docenten met een onvoldoende beoordeelde studenten hebben voor 58% kans ook daadwerkelijk een onvoldoende te behalen (specificiteit). De studenten die zichzelf met een voldoende hebben beoordeeld, hebben voor 89% kans ook daadwerkelijk een voldoende te behalen (sensitiviteit). De studenten die zichzelf met een onvoldoende hebben beoordeeld, hebben voor 50% kans ook daadwerkelijk een onvoldoende te behalen (specificiteit).

Tenslotte is statistisch getoetst op de samenhang tussen de verwachtingen van de docenten/studenten en de daadwerkelijke uitkomsten van het assessment: de

Chi<sup>2</sup>-toets geeft voor de samenhang tussen docentvoorspellingen en daadwerkelijk behaald resultaat een significant resultaat ( $p=0,001$ ;  $df=1$ ). De Chi<sup>2</sup>-toets geeft voor de samenhang tussen studentvoorspellingen en daadwerkelijk behaald resultaat een niet-significant resultaat ( $p=0,07$ ;  $df=1$ ).

### Discussie

In het huidige onderzoek stond de vraag naar de validiteit van het performance-assessment van de bacheloropleiding fysiotherapie centraal. Vanuit het perspectief van de docent waren de sensitiviteit en specificiteit van het instrument respectievelijk 87% en 58%, vanuit het perspectief van de student respectievelijk 89% en 50%. De conclusie is derhalve dat studenten die naar inschatting van hun docent en naar eigen inschatting een voldoende zullen halen, dat in het merendeel van de gevallen ook inderdaad halen. Voor hen die door zichzelf en door hun docent een onvoldoende voorspeld krijgen, is het instrument slechts matig betrouwbaar.

Bij deze conclusie dienen enkele kritische kanttekeningen te worden geplaatst. Aan de beoordeling van de inhoudsvaliditeit van het performance-assessment hebben slechts 42% van de studenten hun medewerking verleend. Mogelijke verklaring hiervoor is de onzekerheid in hoeverre hun medewerking aan het onderzoek invloed zou kunnen hebben op de uitkomst van het assessment. Er is gekeken naar de behaalde resultaten van studenten die niet mee hebben gewerkt aan dit onderzoek. Hierbij werd geen verschil in resultaten in termen van voldoende of onvoldoende beoordeling tussen de responders en non-responders gevonden. Desalniettemin kan een eventuele selectie-bias niet worden uitgesloten, omdat het onduidelijk is in hoeverre alleen studenten in het onderzoek geparticipeerd hebben die redelijk zeker waren van hun te verwachten resultaat.

Voorts zijn bij het beoordelen van de inhoudsvaliditeit van het performance-assessment studenten, die door de docent vooraf met een 'twijfel' ten aanzien van hun te verwachten prestaties tijdens het performance-assessment zijn beoordeeld, buiten beschouwing gebleven. Van deze 'twijfel-studenten' blijkt de helft het performance-assessment wel te hebben behaald en de andere helft niet. Achteraf is het de vraag of zij niet betrokken hadden moeten worden bij de analyse.

Toch menen wij dat de methodologische keuze om aan de hand van student- en docentvoorspellingen performance-assessmentresultaten te valideren, ondanks de bovengenoemde beperkingen, een meerwaarde heeft geboden. Opvallend is dat in een groot aantal gevallen er een overeenkomst was in de voorspelling door de docent en de eigen voorspelling van de betreffende student. Dit blijkt ook uit het feit dat de docenten 62 van de 81 studenten (77%) in staat achtten het assessment te behalen (tabel 2). Van de studenten achtten zich 35 van de 49 studenten (71%) in staat het assessment met succes af te ronden (tabel 3).

Uit dit onderzoek blijkt dat het performance-assessment in het tweede studiejaar van de opleiding fysiotherapie een hoge sensitiviteit en een matige specificiteit kent. Tot op heden was er wel een vermoeden dat de specificiteit matig was, omdat een redelijk aantal studenten in het derde studiejaar gedurende hun praktijkleerperiode in de problemen komt voor wat klinisch redeneren en handelen in de complexe beroepssituatie betreft. In een vervolgonderzoek zal nagegaan worden hoe het hier onderzochte cohort studenten in de praktijkleerperiode zal functioneren.

Nieuw daarentegen is dat het assessment de goede studenten ook daadwerkelijk selecteert. Tevens is het nieuw dat het

oordeel van de docenten gedurende het onderwijsblok van hoge waarde blijkt te zijn. Dit is in dit onderzoek ook statistisch aangetoond.

Blijft de vraag welke oorzaken er zijn voor de lage specificiteit van het performance-assessment. Deze oorzaken dienen zowel in het instrument zelf als ook in de toepassing van het instrument te worden gezocht. Er zal een factoranalyse uitgevoerd moeten worden om te achterhalen welke kritische gedragskenmerken een bijdrage leveren aan het verhogen van de specificiteit. Uit een eenvoudige item-analyse konden daarover geen conclusies getrokken worden.

In de toepassing van het instrument bleken zowel de training van de assessoren in het hanteren van het meetinstrument als een training van de simulatiepatiënten in het spelen van hun rol succesfactoren bij implementatie te zijn. Uit de evaluaties per assessor bleek een zekere variatie in sensitiviteit en specificiteit te bestaan, veroorzaakt door enkele uitschieters. In het vervolg zal er een vast team van assessoren geformeerd worden bestaand uit zowel docenten uit andere fasen van de opleiding als externen, die in regelmatige trainingen bij elkaar komen en in gesimuleerde toetsituaties gezamenlijk geschoold zullen worden. De verwachting is dat deze activiteiten de validiteit zullen verhogen. Toekomstige evaluatieonderzoeken moeten dit uitwijzen.

## Literatuur

- Berkel A van, Hofman K, Kinkhorst G, Lintelo L te. Competentie-assessment. De ontwikkeling en toepassing van self-, peer- en expert assessments in het hbo: een praktijkvoorbeeld. Utrecht: Lemma; 2003.
- Bie D de, Gerritse J. Onderwijs als opdracht. Overwegingen en praktische suggesties voor de ontschooling van het hoger onderwijs. Houten/Diegem: Bohn Stafleu van Loghum; 1999.
- Moust J, Schmidt HG. Probleemgestuurd onderwijs. Praktijk en theorie. Groningen: Wolters-Noordhoff; 1998.
- Jefferson JR. Problem-based learning and the promotion of problem solving: choices for physical therapy curricula. *Journal of Physical Therapy Education* 2001;15(1):26-31.
- Hagenaars L, Bernards ATM, Oostendorp RAB. Over de kunst van hulpverleners. Het meerdimensionale belasting-belastbaarheidsmodel. Amersfoort: Nederlands Paramedisch Instituut; 2003.
- Bie D de, Kleijn J de. Wat gaan we doen? Het construeren en beoordelen van opdrachten. Houten: Bohn Stafleu van Loghum; 2001.
- Dochy F, Heylen L, Mosselaer H van de. Assessment in onderwijs. Nieuwe toetsvormen en examinering in studentgericht onderwijs en competentiegericht onderwijs. Utrecht: Lemma; 2002.
- Kuiper C, Balm M, editors. Paramedisch Handelen. Utrecht: Lemma; 2001.
- O'Hearn MA. The elemental identity of physical therapy. *Journal of Physical Therapy Education* 2002;16(3):4-7.
- Kuiper C, Verhoef J, Louw D de, Cox K, editors. Evidence-based practice voor paramedici. Methodiek en implementatie. Utrecht: Lemma; 2004.
- Doody C, McAteer M. Clinical reasoning of expert and novice physiotherapists in an outpatient orthopaedic setting. *Physiotherapy* 2002;88(5):258-68.
- Higgs J. Developing clinical reasoning competencies. *Physiotherapy* 1992;78(8):575-9.
- Redelmeier DA, Guyatt GH, Goldstein RS. Assessing the minimal important difference in symptoms: a comparison of two techniques. *J Clin Epidemiol* 1996;49(11):1215-9.
- Alderson JC, Wall D. Does washback exist? *Applied Linguistics* 1993;14:115-29.

### De auteurs:

Dhr. Martin Ophey, MSc is hoofddocent fysiotherapie.

Mw. Marjo Maas is fysiotherapeut.

Drs. Joost de Beer is hoofddocent fysiotherapie en logopedie.

Allen zijn werkzaam bij het Taakteam Toetsing en Evaluatie van de Hogeschool van Arnhem en Nijmegen.

### Correspondentieadres:

Dhr. Martin Ophey, MSc, PT, Hogeschool van Arnhem en Nijmegen, afdeling Fysiotherapie, Kapittelweg 33, Postbus 6960, 6503 GL Nijmegen, martin.ophey@han.nl.

## Summary

**Introduction:** A form of performance assessment has been used in the first and second years of the bachelor programme in physiotherapy of the Arnhem, Nijmegen College of Higher Education (Hogeschool Arnhem, Nijmegen) for many years. The assessment comprises integrated assessment of knowledge, skills and attitudes in a 'reduced' professional context with trained simulated patients. The purpose of this study was to examine the validity of this type of performance assessment.

**Method:** Validity was examined from two perspectives. First, the four teachers who taught the students during the fourteen weeks prior to the assessment were asked to predict assessment outcomes for each of the 117 students ("teachers expected"). Second, the students were asked to predict their own outcomes ("student expected"). Teachers' and students' expectations were compared with the actual assessment outcomes.

**Results:** Teachers' predictions were obtained for 107 out of 117 students (response 92%). Of the students, 49 out of 117 responded (42%). The students for whom the teachers predicted a pass had a 87% chance of actually passing (sensitivity) and the students for whom the teachers predicted a fail had a 58% chance of actually failing (specificity). The same results for the pass and fail scores predicted by the students were 89% and 50%, respectively.

**Conclusion:** Students whom teachers and students expected to pass did actually pass the current performance assessment. However, predictions of failure by students and teachers were less accurate. Follow-up research will have to investigate what measures should be taken to enhance the specificity of the assessment. (Ophey M, Maas M, Beer J de. Validity of performance assessment in a bachelor degree course in physiotherapy. *Dutch Journal of Medical Education* 2006;25(2):88-95.)