

# Kwaliteitscriteria voor portfolio-beoordeling van geneeskundestudenten

K. Overeem, E.W. Driessen, A. Drenthe, J. van Tartwijk, C.P.M. van der Vleuten

## Samenvatting

**Inleiding:** Portfolio's worden in toenemende mate geaccepteerd als leer- en beoordelingsinstrument. Echter, er is tot op heden weinig onderzoek gedaan naar de validiteit van beoordelingen aan de hand van portfolio's. Het is de vraag in welke mate beoordelaars zich laten leiden door de vormgeving en schrijfstijl van het portfolio wanneer zij het portfolio beoordelen. De validiteit van deze beoordeling is het onderwerp van de onderhavige studie: "Welke criteria gebruiken docenten bij het beoordelen van portfolio's en aan welke criteria wordt het meeste gewicht toegekend?"

**Methoden:** Voor deze studie zijn portfolio's van eerstejaars geneeskundestudenten in Maastricht gebruikt. Het stimuleren van reflectie op de eigen ontwikkeling door de student is het belangrijkste doel van het werken met portfolio's in de eerste studiejaar. In de latere studiejaar zullen de portfolio's ook worden gebruikt bij het plannen, begeleiden en beoordelen van de ontwikkeling van de medische competenties van de student. Op basis van interviews met mentoren/beoordelaars en de beschikbare literatuur, is een scoringslijst met vijftien criteria voor de kwaliteit van portfolio's samengesteld. De criteria in deze lijst kunnen worden verdeeld in twee groepen: inhoud (niveau van reflectie, realisatie van de leerdoelen, et cetera) en vorm (schrijfstijl en organisatie, et cetera). Twee onderzoekers hebben een gestratificeerde steekproef van veertig portfolio's gescoord op deze scoringslijst. De overeenstemming tussen deze beoordelaars bleek voldoende. Vervolgens is met behulp van regressieanalyse het verband bepaald tussen de scores op de kwaliteitscriteria en de beoordeling van reflectievaardigheden aan de hand van het portfolio.

**Resultaten:** Alle criteria samen blijken 78% van de variantie van het mentoradvies te verklaren. De sterkste predictor is de kwaliteit van reflectie. Dit criterium verklaart 66% van de variantie in het mentoroordeel.

**Conclusie en discussie:** Deze studie laat zien dat reflectievaardigheden op een valide manier kunnen worden beoordeeld middels een portfolio. Of ook medisch inhoudelijke competenties van een aanstaand arts valide kunnen worden beoordeeld met een portfolio moet blijken uit verder onderzoek. (Overeem K, Driessen EW, Drenthe A, Tartwijk J van, Vleuten CPM van der. Kwaliteitscriteria voor portfolio-beoordeling van geneeskundestudenten. Tijdschrift voor Medisch Onderwijs 2005;24(2):64-71.)

## Inleiding

Portfolio's worden gebruikt voor het volgen en begeleiden van de ontwikkeling van competenties van studenten en zijn ook bij uitstek geschikt voor het beoordelen van het resultaat van die ontwikkeling. Er zijn meerdere argumenten te noemen

die pleiten voor het beoordelen van competenties met portfolio's.

Het eerste argument is dat met portfolio's een authentiek beeld wordt gekregen van het functioneren in de beroepspraktijk. Van de lerende wordt verwacht dat hij in het portfolio bewijsmateriaal opneemt

dat een beeld geeft van de wijze waarop taken in de beroepspraktijk zijn uitgevoerd. Op basis daarvan kan worden beoordeeld in hoeverre de samensteller van het portfolio voldoende competent is. Portfolio's voorzien daarmee in de behoefte aan beoordelingsinstrumenten van functioneren in authentieke contexten.<sup>1</sup>

Een tweede argument is dat met behulp van het portfolio niet alleen het bereikte niveau van competentie kan worden vastgesteld, maar ook de ontwikkeling van die competentie zichtbaar wordt. Iets dat met andere beoordelingsinstrumenten moeilijk is te realiseren.<sup>2</sup>

Het derde argument is dat door het gebruik van portfolio's formatieve en summatieve doelen kunnen worden gecombineerd. Het verzamelen van materiaal en het beschrijven van het eigen functioneren voor beoordelingsdoelen zal vaak impliceren dat het eigen handelen wordt geanalyseerd en geëvalueerd; een vorm van reflecteren. Deze reflectiebekwaamheid wordt gezien als een primaire voorwaarde voor professionele ontwikkeling en professioneel gedrag.<sup>3,4</sup>

Een vierde argument voor het beoordelen met portfolio's is de sturende werking die het portfolio heeft op de leeractiviteiten van de student. De vorm en inhoud van toetsen hebben een grote invloed op het leerproces van de student.<sup>5</sup> Bij het gebruik van portfolio's als beoordelingsinstrument kan een onderwijsorganisatie bewerkstelligen dat studenten zich richten op het adequaat uitvoeren van taken in de praktijk, bijvoorbeeld tijdens een co-schap of stage.<sup>6</sup>

Een vaak gehoord argument tegen het gebruik van portfolio's voor beoordeling betreft de matige betrouwbaarheid.<sup>7</sup> Hier is relatief veel onderzoek naar verricht. De validiteit van de portfoliobeoordeling is tot op heden weinig onderwerp van studie ge-

weest. Aangezien het portfolio een selectie is van materiaal uit de praktijk, wordt in de literatuur een grote meerwaarde van de validiteit van het portfolio verondersteld.<sup>8</sup>

Voor de beoordeling van portfolio's wordt een meer globale (holistische) wijze van beoordelen aangeraden.<sup>9-10</sup> Bij een dergelijke wijze van beoordelen bestaat het risico dat een beoordelaar zo onder de indruk is van de wijze waarop een portfolio is geschreven, gestructureerd of vormgegeven, dat hij of zij dit onbewust attributeert naar de te beoordelen competentie van de samensteller.<sup>8</sup> Een voorbeeld hier van is te vinden in een studie van Quinlan et al. bij portfolio's voor leraren in opleiding.<sup>11</sup> Uit de analyses van hardopdenkprotocollen van de in de studie betrokken beoordelaars bleek dat andere factoren dan beoogd (bij de beoordelaar reeds aanwezige informatie over de beoordeelde) het eindoordeel mede bepalen.

Voor de onderhavige studie vormden daarom de volgende twee vragen het uitgangspunt: "Welke criteria gebruiken de docenten bij het beoordelen van reflectievaardigheden aan de hand van portfolio's en aan welke criteria wordt het meeste gewicht toegekend?"

## Methoden

Om de onderzoeksvragen te beantwoorden zijn veertig reeds beoordeelde portfolio's gescoord op een scoringslijst van vijftien criteria voor de kwaliteit van vorm en inhoud van het portfolio. Met een regressieanalyse is nagegaan welke criteria het belangrijkste waren voor de beoordeling.

## Context

De portfolio's zijn afkomstig van eerstejaars geneeskundestudenten van de Universiteit Maastricht. Binnen de geneeskundeopleiding van de Universiteit Maastricht worden portfolio's op dit moment

ingezet voor het stimuleren van academische en professionele competentieontwikkeling van eerste-, tweede-, derde- en vierdejaars geneeskundestudenten.<sup>12</sup> De nadruk wordt in de eerste leerjaren gelegd op het leren reflecteren, zodat in latere jaren het portfolio op een verantwoorde manier kan worden ingezet voor het begeleiden en beoordelen van ervaringsleren. Studenten worden geacht sterkte/zwakte analyses op te stellen over hun functioneren en op basis daarvan leerdoelen te formuleren. Een handleiding met richtvragen en begeleiding door een mentor dienen als hulpmiddel om hierover systematisch na te denken.<sup>4</sup> Het portfolio wordt aan het einde van het studiejaar beoordeeld op basis van de volgende inhoudelijke criteria:

1. Kwaliteit sterkte/zwakte-analyse.
2. Duidelijkheid en haalbaarheid van de leerdoelen.
3. Mate van realisatie leerdoelen.
4. Mate waarin bewijsmateriaal de sterkte/zwakte-analyse ondersteunt.

De beoordeling kent drie kwalificaties: onvoldoende, voldoende en goed. In een eindgesprek bespreekt de mentor zijn gemotiveerde advies met de betreffende student en de student wordt gevraagd of hij hiermee instemt of niet.

### *Onderzoeksinstrument*

Voor deze studie is een scoringslijst ontwikkeld met vijftien items, gebaseerd op de criteria voor de kwaliteit van vorm en inhoud van portfolio's. Bij het ontwikkelen van de lijst vormden de doelstellingen

van het gebruikte portfoliosysteem (leren reflecteren), interviews met mentoren en de literatuur over beoordelen aan de hand van portfolio's het uitgangspunt. Op basis daarvan is een eerste versie van de lijst ontwikkeld. De items hebben de vorm van vijftien uitspraken, waarvan telkens op een Likert-schaal gescoord wordt in hoeverre die van toepassing zijn op het betreffende portfolio. De score 1 staat voor 'duidelijk niet van toepassing' en 5 voor 'duidelijk van toepassing' (zie tabel 1).

De inhoudsvaliditeit van de lijst is getoetst door vijf deskundigen op het terrein van portfolio's en beoordelingen afkomstig van de Universiteit Maastricht en de Universiteit Utrecht.

### *Procedure*

Middels een gestratificeerde steekproef zijn veertig portfolio's verzameld uit het totaal van de 231 beschikbare portfolio's. De geselecteerde portfolio's vormen een redelijke afspiegeling van de totale populatie, zowel wat betreft de beoordeling ervan als wat betreft de verdeling over de mentoren (zie tabel 1).

Ter verdere validering van de scoringslijst heeft een voormeting plaatsgevonden door twee beoordelaars - de eerste en derde auteur van dit artikel - tien portfolio's te laten scoren met behulp van de scoringslijst. Alle items van de scoringslijst bleken tijdens de voormeting van toepassing op de portfolio's, maar op acht van de vijftien items was de overeenstemming tussen de beoordelaars onvoldoende. Van deze items zijn de

**Tabel 1.** *Overzicht van kwalificaties portfolio's in steekproef en totale populatie.*

Kwalificatie	Totaal jaar 1 (%)	Steekproef onderzoek (%)
Onvoldoende	25 (10,82)	5 (12,5)
Voldoende	124 (53,6)	20 (50)
Goed	82 (35,5)	15 (37,5)
Totaal	231 (100)	40 (100)

formuleringen aangescherpt en daarnaast is voor de hele scoringslijst een instructie met richtlijnen geschreven om vast te leggen wanneer bepaalde scores op de Likert-schaal moeten worden toegekend.

Vervolgens zijn deze veertig portfolio's met een scoringslijst gescoord door twee onafhankelijke beoordelaars; de eerste auteur van dit artikel en een niet bij het onderzoek en portfolio betrokken docent. Het scoren gebeurde 'blind', met andere woorden: de beoordelaars waren vooraf niet op de hoogte van het advies dat de mentor aan het portfolio had toegekend.

### Statistische analyse

Voor het berekenen van de interbeoordelaarsovereenstemming is een Pearson's produkt-momentcorrelatie gebruikt. Daarmee wordt niet de absolute maar de relatieve overeenstemming tussen beoordelaars bepaald. Omdat we voornemens waren het verband tussen criteria en beoordelingen vast te stellen met correlatie-

(regressie-)analyse, voldeed deze maat. De samenhang tussen de criteria en het beoordelingsadvies is vastgesteld met behulp van een multiple regressieanalyse. Op basis van een stepwise procedure is vastgesteld welke van de criteria de beste voorspeller vormde voor het mentoradvies.

### Resultaten

De samenhang in de scores van de beide beoordelaars liep uiteen van 0,62 tot 0,94. De gemiddelde overeenstemming voor alle items was 0,82 (zie tabel 2).

De correlatiecoëfficiënt tussen alle items en het mentoradvies bedraagt 0,887. Dat betekent dat met behulp van de criteria 78% van de variantie in het uiteindelijke advies verklaard kan worden (significantie:  $p < 0.000$ ). De standaardschattingsfout, de maat voor de residuele variantie, bedraagt 0,394 (zie tabel 3).

Om het relatieve belang van elke onafhankelijke variabele te beoordelen, kunnen de bèta's, de gestandaardiseerde regres-

**Tabel 2.** *Overeenstemming tussen beoordelaars: significante samenhang (pmc,  $p < ,005$ ), gemiddelde scores (m) en standaarddeviatie (sd).*

	pmc	m	sd
1. Dit portfolio heeft een van de mooiste lay-outs.	,77	3,46	,85
2. Dit portfolio behoort tot de portfolio's met de beste spelling en zinsbouw.	,62	3,63	,74
3. De organisatie van het portfolio is correct (inhouden op de goede plaats, makkelijk om beschrijvingen, analyses en voornemens terug te vinden).	,78	4,58	,67
4. Het portfolio is compleet, d.w.z. er ontbreken geen rollen of delen van rollen.	,94	3,52	1,07
5. De student heeft kritisch naar zichzelf gekeken, dat wil zeggen voor iedere rol waaraan is gewerkt worden zowel sterke als zwakke punten aangegeven.	,83	3,81	1,00
6. In de sterkte/zwakte analyse worden verklaringen gezocht voor de aangegeven sterke en zwakke punten, zowel intern als extern. Het is meer dan een opsomming van feiten of situaties.	,87	3,10	1,14
7. De sterkte-zwakteanalyses bevatten voor iedere rol voldoende thema's.	,85	4,11	,88
8. Er wordt systematisch verwezen naar bewijsmateriaal (bewijsmateriaal ondersteunt de sterkte-zwakteanalyses).	,75	3,83	,85
9. De student heeft activiteiten buiten de studie in relatie gebracht met zijn competentie-ontwikkeling.	,86	2,66	1,22

## Vervolg tabel 2

	pmc	m	sd
10. Voor de inhoud van het portfolio heeft de student meer gedaan dan alleen de richtvragen gevolgd.	,80	2,28	,96
11. In het portfolio wordt teruggekeken naar eerdere versies (wat ging toen fout wat nu goed ging en waarom, welke uitspraken heb ik toen gedaan).	,85	2,95	1,19
12. De student heeft logische (voortvloeiend uit sterkte-zwakteanalyse) en heldere leerdoelen opgesteld.	,85	3,76	,99
13. De student heeft aangegeven hoe hij leerdoelen wil verwezenlijken.	,87	2,48	1,1
14. De student heeft geprobeerd zichtbaar te maken dat aan de leerdoelen is gewerkt.	,78	2,33	1,17
15. De student heeft meer aan het portfolio gedaan dan strikt noodzakelijk is.	,84	2,14	,94

**Tabel 3.** Regressiecoëfficiënten ( $R$ ) en verklaarde variantie ( $R^2$ ) van alle items en sterkste predictor.

Kwaliteitscriteria	R	R <sup>2</sup>	Standaard-schattingfout	Significantie
Alle	0,887	78%	0,394	0,000
Kwaliteit reflectie	0,803	64%	0,405	0,000

siecoëfficiënten, worden gebruikt. Enkel de regressiecoëfficiënt van item 6, waarin reflectie is geoperationaliseerd, blijkt significant te zijn (zie tabel 4). De correlatiecoëfficiënt tussen reflectie (item 6) en het mentoradvies is 0,803. Dit impliceert een verklaarde variantie van 64% (significantie:  $p < 0,000$ ). De standaardschattingfout bedraagt dan 0,405 (zie tabel 4).

### Conclusie en discussie

In deze studie is onderzocht welke criteria door mentoren worden gebruikt bij het beoordelen van portfolio's en aan welke criteria relatief het meeste gewicht wordt toegekend.

De onderzochte criteria verklaren 78% van de variantie in de beoordelingsadviezen. Hoewel alle criteria tezamen 78% van de variantie verklaren, is alleen de regressiecoëfficiënt van het criterium 'kwaliteit van reflectie' (item 6) significant. Hiermee kunnen 64% van de verschillen in de beoordelingen worden verklaard. Dat betekent dat we voor-

alsnog geen redenen hebben om aan te nemen dat andere criteria van (groot) belang zijn voor de beoordelingen dan de kwaliteit van de reflectie. Dit in tegenstelling tot de eerder genoemde studie van Quinlan et al.<sup>11</sup> Een belangrijk verschil met de studie van Quinlan et al. is dat de beoordelaars in onze studie niet bekend zijn met de beoordeelde. Dit komt overeen met de werkelijke portfolio-beoordelingsprocedure zoals die wordt toegepast in Maastricht. Ook daar zijn de beoordelaars onbekend met de beoordeelde.<sup>10</sup> Aangezien met het portfolio inderdaad beoogd werd de kwaliteit van reflectie te meten, kunnen we er dus ook vooralsnog van uitgaan dat de portfolio-beoordeling zoals die in Maastricht plaatsvindt in dit opzicht valide is.

Een beperking van deze studie vormt de relatief kleine steekproef. Verder onderzoek met een grotere steekproef leidt wellicht tot meer significante regressiecoëfficiënten.

Uit ons onderzoek komen ook een aantal aanwijzingen over de inhoudelijke kwali-

**Tabel 4.** *Bètagewichten en significantie per item in de vergelijking met beoordelingsadvies door de mentor als afhankelijke variabele.*

	Bèta	p
1. Dit portfolio heeft een van de mooiste lay-outs.	,164	,324
2. Dit portfolio behoort tot de portfolio's met de beste spelling en zinsbouw.	,009	,946
3. De organisatie van het portfolio is correct (inhouden op de goede plaats, makkelijk om beschrijvingen, analyses en voornemens terug te vinden).	-,330	,825
4. Het portfolio is compleet, dat wil zeggen er ontbreken geen rollen of delen van rollen.	,317	,116
5. De student heeft kritisch naar zichzelf gekeken, dat wil zeggen voor iedere rol waaraan is gewerkt worden zowel sterke als zwakke punten aangegeven.	,157	,492
6. In de sterkte-zwakteanalyse worden verklaringen gezocht voor de aangegeven sterke en zwakke punten, zowel intern als extern. Het is meer dan een opsomming van feiten of situaties.	,667	,036
7. De sterkte-zwakteanalyses bevatten voor iedere rol voldoende thema's.	-,360	,117
8. Er wordt systematisch verwezen naar bewijsmateriaal (bewijsmateriaal ondersteunt de sterkte-zwakteanalyses).	,272	,163
9. De student heeft activiteiten buiten de studie in relatie gebracht met zijn competentie-ontwikkeling.	-,147	,310
10. Voor de inhoud van het portfolio heeft de student meer gedaan dan alleen de richtvragen gevolgd.	-,295	,302
11. In het portfolio wordt teruggekeken naar eerdere versies (wat ging toen fout wat nu goed ging en waarom, welke uitspraken heb ik toen gedaan).	-,090	,575
12. De student heeft logische (voortvloeiend uit sterkte-zwakteanalyse) en heldere leerdoelen opgesteld.	,261	,161
13. De student heeft aangegeven hoe hij leerdoelen wil verwezenlijken.	-,139	,336
14. De student heeft geprobeerd zichtbaar te maken dat aan de leerdoelen is gewerkt.	-,180	,286
15. De student heeft meer aan het portfolio gedaan dan strikt noodzakelijk is.	,171	,567

teit van de portfolio's naar voren, die niet direct samenhangen met de onderzoeksvragen uit deze studie. Het scoren van de portfolio's door twee onafhankelijke beoordelaars geeft inzicht in de kwaliteit van de door eerstejaars studenten gemaakte portfolio's. Uit de gemiddelden op de scoringslijst (zie tabel 2) hebben we aanwijzingen dat portfolio's goed georganiseerd (item 3: m 4,6) en compleet (item 7: m 4,1) zijn, dat systematisch wordt verwezen naar bewijsmateriaal (item 8: m 3,8)

en dat logische en heldere leerdoelen zijn geformuleerd (item 12: m 3,8).

We hebben echter ook aanwijzingen dat de onderzochte portfolio's in mindere mate voldoen aan enkele andere kwaliteitscriteria. Ten eerste volgt uit de lage score op item 9 (m 2,6) dat studenten in hun reflecties vooral blijken te reflecteren op studiegerelateerde activiteiten. Activiteiten buiten de studie worden weinig in de reflecties opgenomen. Dit heeft waarschijnlijk te maken met de onwennigheid met het

portfolio in het eerste studiejaar. Competenties zoals presenteren, samenwerken, omgaan met kritiek en onzekerheid worden ook in grote mate in extracurriculaire activiteiten aangeleerd en geoefend. Ten tweede blijkt uit de lagere score op item 10 (m 2,3) – de mate waarin de student afwijkt van de voorgeschreven structuur – dat studenten over het algemeen weinig gebruik maken van de ruimte die het portfolio biedt voor een persoonlijke invulling. Dit is waarschijnlijk inherent aan het feit dat het eerste studiejaar nog relatief weinig ruimte biedt om een eigen invulling aan de studie te geven en reflectie op de eigen ontwikkeling voor veel studenten nog een nieuw fenomeen is.

Eveneens opvallend zijn de relatief lage scores op de items 11 (m 3,0), 13 (m 2,5) en 14 (m 2,1). Studenten geven in het portfolio vaak niet aan hoe ze leerdoelen wensen te verwezenlijken en besteden ook minder aandacht aan het verzamelen van bewijs om aan te tonen dat aan leerdoelen is gewerkt. Hier is een rol voor de mentor weggelegd. Het is belangrijk dat studenten in hun portfolio beschrijven hoe ze leerdoelen willen verwezenlijken, op welke termijn en welk bewijsmateriaal ze in een volgend portfolio gaan toevoegen om zichtbaar te maken dat aan de leerdoelen is gewerkt, zodat ze worden gestimuleerd tot een verandering in het functioneren.

## Literatuur

1. Roberts C, Newble DI, O'Rourke OJ. Portfolio-based assessments in medical education: are they valid and reliable for summative purposes? *Med Educ* 2002;36:899-900.
2. Gordon J. Assessing students' personal and professional development using portfolios and interviews. *Med Educ* 2003;37:335-40.
3. Wilkinson TJ, Challis M, Hobma SO, Newble DI, Parboosingh JT, Sibbald RG, et al. The use of portfolios for assessment of the competence and performance of doctors in practice. *Med Educ* 2002;36(6):918-24.
4. Driessen EW, Tartwijk J van, Vermunt JD, Vleuten CPM van der. Use of portfolios in early undergraduate medical training. *Med Teach* 2003;25(1):18-23.
5. Vleuten CPM van der, Driessen EW. Toetsing in probleemgestuurd onderwijs. Reeks Hoger Onderwijs Praktijk. Groningen: Wolters Noordhoff; 2000.
6. Beijaard D, Driessen EW, Tartwijk J van, Vleuten CPM van der. Portfolio verder uitgewerkt. In: Driessen EW, Beijaard D, Tartwijk J van, Vleuten CPM van der. *Portfolio's*. Groningen: Wolters-Noordhoff; 2002. p. 119-38.
7. Pitts J, Coles C, Thomas P. Educational portfolios in the assessment of general practice trainers: reliability of assessors. *Med Educ* 1999;33:515-20.
8. McMullan M, Endacott R, Gray MA, Jasper M, Miller CML, Scholes J, et al. Portfolios and assessment of competence: a review of the literature. *Journal of Advanced Nursing* 2003;41(3):283-94.
9. Ryan JM, Kuhs TM. Assessment of preservice teachers and the use of portfolios. *Theory into practice* 1993;32:75-81.
10. Driessen EW, Vleuten CPM van der, Schuwirth L, Tartwijk J van, Vermunt JD. The use of qualitative research criteria for portfolio assessment as an alternative to reliability evaluation: a case study. *Med Educ* 2005;39(2):214-20.
11. Quinlan KM. Inside the peer review process: how academics review a colleague's teaching portfolio. *Teaching and Teacher Education* 2002;18:1035-49.
12. Overeem K, Driessen EW, Tartwijk J van, Vleuten CPM van der. Portfolio om de basisarts-in-opleiding te begeleiden. *Tijdschrift voor Medisch Onderwijs* 2003;22(6):268-76.

### De auteurs:

Mw. drs. K. Overeem is student-assistent bij de Capaciteitsgroep Onderwijsontwikkeling en Onderwijsresearch, Faculteit der Geneeskunde, Universiteit Maastricht.

Drs. E.W. Driessen is universitair docent bij de Capaciteitsgroep Onderwijsontwikkeling en Onderwijsresearch, Faculteit der Geneeskunde, Universiteit Maastricht en coördinator van een portfolio-project in Maastricht.

Mw dr. A. Drenthe is docent bij het Skillslab, Faculteit der Geneeskunde, Universiteit Maastricht, studie-adviseur en betrokken bij het portfolio-project in Maastricht.  
Dr. J. van Tartwijk is universitair hoofddocent bij het ICLON in Leiden.

Prof. dr. C.P.M. van der Vleuten is voorzitter van de Capaciteitsgroep Onderwijsontwikkeling en Onderwijsresearch, Faculteit der Geneeskunde, Universiteit Maastricht.

Correspondentieadres:  
Drs. K. Overeem, Universiteit Maastricht, Faculteit der  
Geneeskunde, Capaciteitsgroep Onderwijsontwikkeling

en Onderwijsresearch, Postbus 616, 6200 MD Maas-  
tricht, tel: 043-3885774, k.overeem@student.unimaas.  
nl of e.driessen@educ.unimaas.nl.

## Summary

**Introduction:** Portfolios are increasingly being accepted as a tool for learning and assessment. However, few studies have investigated the validity of portfolio assessment. The question is how much lay-out and writing style contribute to portfolio assessment. This study addresses the question "Which criteria do teachers use in portfolio assessment and which criteria should be weighted most heavily."

**Methods:** We examined portfolios of Year 1 students at Maastricht medical school. The main objective of the portfolio in the first few years is to stimulate students to reflect on their development. Later on portfolios will also be used for the planning, monitoring and assessment of the development of medical competencies. Fifteen quality criteria for portfolios have been derived from interviews with mentors/assessors and the available literature. The criteria can be divided into two categories: content (level of reflection, achievement of learning objectives, et cetera) and form (writing style, organisation, et cetera). Two researchers used the criteria in rating a stratified sample of forty portfolios. Inter-rater agreement was acceptable. Regression analysis was used to determine the association between the scores on the quality criteria and the ratings of reflective skills by the mentors.

**Results:** The fifteen criteria explained 78% of the variance in mentor ratings. The strongest predictor was the quality of reflection (66%).

**Conclusion and discussion:** This study shows that a portfolio can be used to make a valid assessment of reflective skills. Further studies will have to examine whether the portfolio is also a valid tool for the assessment of clinical competencies. (Overeem K, Driessen EW, Drenthe A, Van Tartwijk J, Van der Vleuten CPM. Quality criteria for portfolio assessment in undergraduate medical education. *Dutch Journal of Medical Education* 2005;24(2):64-71.)