

## Assessing student reflection in medical practice. The development of an observer-rated instrument

Boenink, A.D.; Oderwald, A.K.; de Jonge, P.; van Tilburg, W.; Smal, J.A.

*Published in:*  
Medical Education

*Publication date:*  
2004

[Link to publication](#)

### *Citation for published version (APA):*

Boenink, A. D., Oderwald, A. K., de Jonge, P., van Tilburg, W., & Smal, J. A. (2004). Assessing student reflection in medical practice. The development of an observer-rated instrument: Reliability, validity and initial experiences. *Medical Education*, 38(4), 368-377.

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### **Take down policy**

If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.

## Assessing student reflection in medical practice. The development of an observer-rated instrument: reliability, validity and initial experiences

A D BOENINK,<sup>1</sup> A K ODERWALD,<sup>2</sup> P DE JONGE,<sup>1</sup> W VAN TILBURG<sup>1</sup> & J A SMAL<sup>3</sup>

**INTRODUCTION** This study describes the development of an instrument to measure the ability of medical students to reflect on their performance in medical practice.

**METHODS** A total of 195 Year 4 medical students attending a 9-hour clinical ethics course filled in a semi-structured questionnaire consisting of reflection-evoking case vignettes. Two independent raters scored their answers. Respondents were scored on a 10-point scale for overall reflection score and on a scale of 0–2 for the extent to which they mentioned a series of perspectives in their reflections. We analysed the distribution of scores, the internal validity and the effect of being pre-tested with an alternate form of the test on the scores. The relationships between overall reflection score and perspective score, and between overall reflection score and gender, career preference and work experience were also calculated.

**RESULTS** The interrater reliability was sufficient. The range of scores on overall reflection was large (1–10), with a mean reflection score of 4.5–4.7 for each case vignette. This means that only 1 or 2 perspectives were mentioned, and hardly any weighing of perspectives took place. The values over the 2 measurements were comparable and were strongly related. Women had slightly higher scores than men, as had students with work experience in health care, and students considering general practice as a career.

**CONCLUSIONS** Reflection in medical practice can be measured using this semistructured questionnaire built on case vignettes. The mean score allows for the measurement of improvement by future educational efforts. The wide range of individual differences allows for comparisons between groups. The differences found between groups of students were as expected and support the validity of the instrument.

**KEYWORDS** education, medical undergraduate/\*standards/\*methods; \*mental processes; clinical competence/\*standards; curriculum; psychometrics/\*method; reproducibility of results; Netherlands.

*Medical Education* 2004; **38**: 368–377

doi:10.1046/j.1365-2923.2004.01787.x

### INTRODUCTION

Reflection – the conscious weighing and integrating of views from different perspectives – is a necessary prerequisite for the development of a balanced professional identity. In medical education increasing attention is directed at issues concerning professional identity, communication, interpersonal skills, medical ethics, medico-legal aspects and the functioning of the health care system. However, once a student enters his or her clerkships, it is difficult to transfer this knowledge into clinical practice, as contradictory considerations will be thrust upon most clerks. For example, trying to be patient-centred will conflict with being on time for a presentation; trying to learn as much as possible, on the other hand, will conflict with patient interests. Leaving on time might result in conflicts with other clerks or superiors, and patient care problems may prove not as straightforward as they were in the preclinical years. In short, acquiring knowledge and practical skills alone are

<sup>1</sup>Department of Psychiatry, Vrije Universiteit Medical Centre, Amsterdam, the Netherlands

<sup>2</sup>Department of Metamedica, Section of Medical Ethics, Vrije Universiteit Medical Centre, Amsterdam, the Netherlands

<sup>3</sup>University Medical Centre, Utrecht, the Netherlands

*Correspondence.* Annette D Boenink, Psychiatrist, Consultation Liaison Psychiatry, Vrije Universiteit Medical Centre, PO Box 7057, 1007 MB Amsterdam, the Netherlands. Tel: 00 31 20 444 0196; Fax: 00 31 20 444 0197; E-mail: ad.boenink@vumc.nl

## Key learning points

Although the importance of reflection for medical students and doctors is often mentioned, no instrument to assess the ability to reflect in medical practice is readily available.

We developed a semistructured, observer-rated instrument based on case vignettes to measure the degree of reflection of medical students in situations in medical practice.

This study shows that the instrument has acceptable psychometric properties.

The differences in total reflection scores that relate to sex, career preference for general practice, and previous health care work experience were supportive of the construct validity.

The results of our study raise the question of whether reflection as we have measured it is a skill that can be taught and learnt, or a personality trait that may be hard to change.

not enough to become a medical professional.<sup>1,2</sup> Reflecting on educational and clinical experiences in medical practice, including one's own behaviour, becomes crucial.

The medical school of the Vrije Universiteit (Free University), in Amsterdam, has initiated a reorganisation of the clerkships its students undertake. 'Reflection in medical practice' now spearheads the education given during the clerkships. Because this explicit attention on reflection was new in this curriculum, the importance of measuring the effects of the programme was acknowledged, and an evaluation procedure was developed in parallel with the development of the educational programme.

### Measuring reflection in medical practice: existing instruments

Although the importance of reflection for medical students and doctors is often mentioned, no instrument to assess the ability to reflect in medical practice is readily available. In our view, in everyday medical practice, reflection activity starts with undifferentiated

feelings or thoughts of unease, surprise or puzzlement. Then, reflection on all relevant perspectives may give an indication whether the issues to be considered include, for example, moral issues, legal or organisational issues, issues concerning personal feelings, or a combination of these. This type of reflection activity precedes more specific types of reflection such as more intensive self-reflection, or ethical reflection in a more restricted sense.

Niemi describes the quality of professional self-reflection and identity formation of medical students on the basis of qualitative material of learning logs and identity status interviews.<sup>3</sup> Sobral uses a scale of reflection-in-learning based on a self-report questionnaire.<sup>4</sup> In other fields (e.g. ethics, teaching and nursing) more research on reflection has been carried out. In ethics, instruments have been developed to measure 'moral reasoning', including the moral judgement interview devised by Kohlberg<sup>5</sup> and the defining issues test devised by Rest and Narvaez.<sup>6</sup> Both instruments build upon Kohlberg's theory of moral development, in which different levels of moral reasoning are defined. In teaching and in nursing, research has been carried out using journals written by students. This research builds on the work of Schön, who developed a theory about reflective practice and its meaning for the education of professionals.<sup>7,8</sup> In this theory, reflection is considered to be the central part of teaching and learning in the professions. Wong *et al.* developed a coding system for written reflective journals from nursing students.<sup>9</sup> They distinguished between non-reflectors, reflectors and critical reflectors, but the coding system failed to allocate students to finer levels of reflection.<sup>9</sup>

None of the existing instruments and measurements corresponds completely with our goal, either because they measure only 1 aspect of what we consider relevant to reflection in medical practice (e.g. self-reflection or moral reasoning), they have a different subject (reflection-in-learning) or they are too time-consuming and difficult to use in a repeated measurements study (coding reflective journals). Therefore, we decided to develop a new instrument to measure reflection in medical practice.

---

## DEVELOPMENT OF A NEW INSTRUMENT TO MEASURE REFLECTION OF MEDICAL STUDENTS

We specified that the instrument should have acceptable psychometric properties, high interrater

reliability, sufficient validity and sufficient interindividual distribution of scores. Furthermore, application of the instrument to larger groups of students had to be feasible and repeated measurements had to be possible.

We decided to use vignettes in the research instrument. Despite certain limitations, the use of vignettes is considered an appropriate tool for awareness and attitudinal research.<sup>10</sup> We composed a list with descriptions of situations (case vignettes) containing the type of problem that clerks/interns might possibly encounter in medical practice. The case vignettes were derived from several sources, including material that clerks brought into ethical seminars, as described by Huijter *et al.*,<sup>11</sup> and experiences we heard from clerks working in our department. After reading the description of each situation, students were asked to write down their reflections (described as 'considerations, feelings, thoughts and/or questions, even if they seem trivial or illogical'). We asked them to name at least 2 and at most 10 such reflections. The number of 10 might seem high, but it was chosen to encourage the students to continue to reflect after their first associations had emerged. The vignettes are shown in Table 1. A second question concerned the students' ideas about what is considered professional behaviour in the described situation. This second question will be addressed elsewhere.

A useful assessment of the concept of reflection was developed by a group of senior staff from different specialisations, together with a medical ethicist and an expert in medical education research. Two types of assessment of the material were performed: a list was formulated of 12 potential perspectives that contribute to the quality of reflection: the medical perspective, medico-legal, ethical, professional norm, patient, family, educational, personal norm, personal feelings, position in hierarchy, and the perspective(s) of the doctor(s) and/or nurse(s) involved. Scores were allocated for each perspective according to whether a student mentioned it 'not at all' (= 0), 'once' (= 1) or 'extensively' (= 2). The second type of assessment concerned an overall score (of 1–10) given for the degree of reflection shown in the written response of the student (Table 2). The score was higher when the response included more evaluation, discussed contradictory arguments and dilemmas as such, showed a balance between personal and more general points of view, and when more perspectives were mentioned. Students did not have to mention all 12 potential perspectives to obtain a high score, but they had to

mention the perspectives relevant to the present case.

In the process of developing this assessment, a definition of what we thought of as excellent reflection in medical practice was constructed as follows: 'showing a balanced approach, considering all relevant perspectives, weighing up different interests, showing a keen eye for dilemmas and uncertainties, paying attention to the patient's viewpoint and demonstrating an evaluation of one's own position and latitude.' This corresponds with the description of the maximum score of 10.

A small pilot study was performed to test the feasibility of a prototype consisting of 11 vignettes. Filling in the instrument proved to be time-consuming for the students. In addition, we found that certain situations did not stimulate much reflection, especially when the students or clerks were asked to write their reflections from the perspective of a doctor. Some vignettes proved ambiguous: students picked up on a different dilemma from that which we had intended. With 7 remaining vignettes, a reliability study was undertaken (*n* varied from 10 to 15 per vignette). Two independent raters scored the students' reflections with a manual (using scores of 1–10). For 4 vignettes a sufficient interrater reliability was found (0.53–0.94, Pearson's *r*). The current study was performed using these 4 vignettes.

### Objectives and hypotheses of the study

The goal of the study was to evaluate the psychometric properties of the instrument. We assessed the distribution of scores, the consistency across multiple measurements, the relationship of the overall reflection score with scores on the different perspectives and its relation to gender, previous work experience in health care and career preferences. We hypothesised that women might be more reflective than men, because female medical students tend to have more patient-centred attitudes, which we consider an important reflection perspective.<sup>12</sup> We also thought that students considering general practice as a career would be more reflective because it is well known among Dutch medical students that reflection is an important part of general practice education and practice. Finally, we hypothesised that work experience in health care might lead to better reflection because of a greater familiarity with the different perspectives that are important in medical practice. Confirmation of these hypotheses would support the construct validity of the instrument.

**Table 1 Content of the case vignettes and the question applied to each**

## Vignette 1

As a clerk you have recently been involved in the admission of Mr R, who is almost certainly suffering from colorectal cancer. However, he has not yet been informed of this diagnosis because the resident who is in charge of this patient wants to wait until the information from additional examinations is available (information about pathology and staging).

In the hallway you are approached by Mr R. He asks you for information about his diagnosis: you were present when the examinations were performed and his case was discussed, weren't you? He is irritated and feels abandoned, and he says that the patient's right to information doesn't seem to be taken very seriously in this hospital.

## Vignette 2

You are a clerk in a surgical department. You've had a good day, because you were allowed to assist at a laparotomy, which was very interesting. Back at the department, there is a lot of work to do: seeing new patients and compiling their clinical records. Some X-ray photographs are missing and, at the request of the resident, you go looking for them. You have been working hard and you want to leave the hospital at 6.30 pm prompt, because tonight you are going out to a concert with your partner to celebrate being together for a year. Two weeks ago, you missed his/her mother's birthday party, because there was an acute admission at the end of the day.

Then a nurse shows up: could the resident talk to the family of Mrs B? They are quite upset because their mother had heard something was wrong. The nurse had not completely understood what the family meant, but as it happened, they were right, their mother was not doing very well after her operation. Mrs B was confused and, the nurse almost forgot to say, she had also developed a fever. The nurse says that she asked the other resident this afternoon to see to Mrs B, but he had not made any arrangements before leaving.

The resident asks you to have a look at Mrs B because she could be suffering from some postoperative complication, and as a clerk, you have to learn how to recognise that.

## Vignette 3

You are working as a clerk at the gynaecology department. You are only halfway through your clerkship, but for some reason you missed the opportunity to practise the pelvic examination. You have just had an interim assessment in which you were told to give special attention to this topic. You certainly intend to, because you are considering a career as a gynaecologist yourself.

However, your next patient, Mrs D, refuses a pelvic examination by a clerk. She has had this too often already, because she has an anomaly of the uterus. Several days later you are at the operating theatre. Mrs D is on the schedule and she has already been anaesthetised. You are working together with a sympathetic gynaecologist who knows about your interest in this specialty. He suggests that you do a pelvic examination now, so that you can feel the anomaly.

## Vignette 4

You are a clerk in surgical department in a medium-sized hospital. You like the work and there is a good atmosphere. You are hoping to get a job in this department after you've finished your study. You are on call on a Friday night. There is not much work and the resident who is also on call that night takes you with him to the Friday night get-together in the department library. He offers you a beer that you politely refuse because you are on call. He laughs and says, 'Come on, one beer never hurt anybody.'

**Table 1 Continued**

It is a lively get-together and you are talking with another clerk. You cannot see how much the resident has been drinking. From his behaviour you cannot tell that he has been drinking. After 1 hour both you and the resident are called to the ER. An 85-year-old woman has broken her hip. She has to be operated the same evening. Her husband is very upset and says to you, 'Doctor, you will take good care of my wife? We have been married for almost 60 years and she has never been to hospital, she was never ill.'

**Question** Write down as completely as possible which considerations, thoughts, feelings and/or questions arise in you in the described situation.  
(These may be contradictory, 'illogical' or 'trivial'! Try to name everything that comes up)  
Name at least 2 and at most 10

**Table 2 Scoring instructions for overall reflection scores**

Score	Scoring instruction overall reflection score
1–2	Oversimplified, intolerant opinion, only emotional reaction
3–4	Limited/restricted, narrow-viewed, one-sided reaction, mostly just 1 perspective, no weighing up or balancing, no attention paid to context
5	More than 1 perspective, but neither balancing nor attention paid to context
6–7	More perspectives, general as well as personal, some balancing between perspectives
8–9	Differentiated balancing, room for dilemmas and or doubt, explicit attention paid to the patient
10	A subtle/balanced approach, considering all relevant perspectives, weighing up of different interests, a keen eye for dilemmas and uncertainties, paying attention to the patient's viewpoint and an evaluation of one's own position and latitude

**Study design**

The study was performed in January and February 2000 with Year 4 medical students. In the Netherlands, Year 4 medical students have not yet entered the clerkships, and as such have limited clinical experience. The participants were attending a 3 × 3-hour, small group course in medical ethics. There were 20 groups, with 10–14 students in each. There were 3 weekly sessions, each lasting 3 hours, in which real-life ethical problems were introduced by doctors or the students themselves and discussed by the group, guided by staff from the Department of Medical Ethics. An analysis was made by the group using a model for cases, made by the

department, based on several existing models for ethical analysis.

For the purpose of our study, the instrument was divided into 2 halves (R1 and R2), each consisting of 2 vignettes. An alternate form design was used, with a control group that had no previous measurements, to evaluate the possible bias of having been pre-tested with an alternate form of the test. The groups of students were allocated to 1 of the 4 conditions on the basis of practical, organisational feasibility (Table 3). The vignettes were preceded by questions concerning the students' age, gender, year of entering medical education, previous work experience in health care and career preference after finishing

**Table 3 Design and number of students who completed the questionnaire in each condition**

Condition	Time		<i>n</i>
	T1	T2	
1	R1	R2	36
2	R2	R1	24
3		R1	52
4		R2	35

R1 = case vignettes 1 and 2; R2 = case vignettes 3 and 4.

medical study. The questionnaire was handed out by the medical ethics teaching staff, together with a letter explaining the purpose of the study. Students could fill in the questionnaire anonymously.

Answers were scored by 2 independent raters (JAS and PdJ), each scoring 1 vignette per list, and using the same vignette for each test form. The raters were unaware of all independent variables (sex, age, work experience, career choice and condition).

### Analysis

Relationships between overall reflection scores and the 12 perspectives were analysed for each of the 4 vignettes using Spearman rank correlations. By analysing the differences in overall reflection scores between conditions by *t*-tests, we were able to study the comparability of the measurements at different times, before and after the ethics course, and with or without pre-course measurement. Relationships between the independent variables (sex, work experience and career preference) and the dependent variable (total reflection score) were analysed by *t*-tests for each of the vignettes.

## RESULTS

A total of 195 students participated in the study. One group (*n* = 10) in condition 1 and 2 groups (*n* = 22) in condition 2 missed the second measurement because of organisational problems (e.g. cancelling of the last seminar because of tutor illness). They were excluded from analysis. Of the remaining 163 students, 16 students dropped out, mainly due to individual changing of groups, which implied a

**Table 4 Career preferences (in percentage of students, *n* = 190); further options were allowed**

Career preference	%	<i>n</i>
General practice	32	61
Social medicine	5	9
General hospital	82	155
Psychiatry	16	31
Research	12	23
Education	11	21
Tropical medicine	5	9

violation of the research protocol for these individuals, or due to non-attendance. These 16 students were then also excluded. Data from 147 students were used in the analysis. Table 3 shows the number of students in each condition.

Of the 147 students, 63% were female, reflecting the current trend in Dutch medical schools, and 53% (of *n* = 189) had previous work experience in health care. Their mean age was 23.5 years (SD 3; range 20–48 years). Career preferences are shown in Table 4.

Reflection scores for each vignette are shown in Table 5. The mean scores for each vignette on this measurement lie between 4.5 and 4.7 (SD 1.5–2, range 1–9 or 1–10). This means that all case vignettes have comparable 'reflection-triggering' properties, and that considerable interindividual differences exist.

Differences between the 4 conditions on scores of reflection on each of the vignettes as well as on the mean score of the 2 vignettes per questionnaire were not significant (1-way ANOVA). This means that we did not find an effect of pre-testing with an alternate form of the test, or any effect of the ethics course on the scores.

The correlations between scores on the 2 vignettes in 1 questionnaire were 0.35 for cases 1 and 2 (questionnaire R1) and 0.41 for cases 3 and 4 (questionnaire R2) (Pearson's *r*, *P* = 0.000, 2-tailed), which supports the internal validity of the instrument.

As no differences were found between conditions, we analysed the relation between the mean reflection scores on the first measurement and the second measurement to test the consistency across multiple

**Table 5 Mean scores and SD for total reflection on T1 and T2 per case vignette per condition**

Condition	Time T1				Time T2			
	Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
1 ( <i>n</i> = 36)	4.8 (± 1.9)	5 (± 1.6)					4.3 (± 1.6)	4.4 (± 1.6)
2 ( <i>n</i> = 24)			5 (± 1.3)	4.6 (± 2.2)	3.9 (± 1.7)	4.4 (± 1.4)		
3 ( <i>n</i> = 52)					4.7 (± 2.0)	4.2 (± 2.0)		
4 ( <i>n</i> = 35)							5.0 (± 1.6)	5.0 (± 2.0)

**Table 6 Scores on perspectives and relationship of perspective to overall reflection score. *a* represents the percentage of students (*n* = 147) who mentioned this perspective once (first number) or extensively (second number); *b* represents the correlation of mentioning this perspective with overall reflection score. All relations are positive**

Perspective	Case 1		Case 2		Case 3		Case 4	
	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>
1 Medical	35 + 2	0.31*	10	0.23*	12	ns	3	0.25*
2 Legal	10	ns	0		7	0.29*	2	ns
3 Professional norm	27 + 4	0.23*	18 + 4	0.39*	14	ns	33 + 4	0.24†
4 Ethical	9 + 4	0.18†	3	0.20†	24 + 3	0.46*	23 + 1	0.33*
5 Patient	41 + 8	0.37*	8	0.33*	33 + 3	0.31*	17 + 2	0.35*
6 Family	1	0.16†	4 + 1	0.22*	0		8	0.35*
7 Educational	0		27 + 2	0.35*	33 + 5	ns	1	ns
8 Personal norm	34 + 8	0.35*	30 + 15	0.42*	45 + 15	0.20†	45 + 17	ns
9 Personal feelings	26 + 14	0.16†	32 + 43	ns	35 + 21	0.25*	34 + 14	0.25*
10 Position/hierarchy	36 + 4	0.28*	10 + 1	0.25*	21 + 6	0.41*	24 + 2	0.45*
11 Doctor	5 + 1	0.19†	0		0		6	ns
12 Nurse	0		0		0		0	

All correlations Spearman's rho.

\*  $P < 0.01$ ; †  $P < 0.05$ , 1-tailed.

measurements. We were able to do so because although the questionnaires were filled in anonymously, a large majority of the students agreed to write down their student number. Pearson's  $r$  was 0.38 ( $P < 0.01$ , 2-tailed,  $n = 53$ ). However, there was a remarkable difference between the 2 conditions in which 2 measurements took place. In the second condition (R2 before and R1 after the course) there was a strong relationship (Pearson's  $r = 0.75$ ,

$P < 0.000$ ,  $n = 21$ ), but in the first condition (R1 before and R2 after the course) the relationship was not significant (Pearson's  $r = 0.20$ ,  $P = ns$ ,  $n = 32$ ). This difference will be addressed in the Discussion.

Table 6 shows the percentage of students who mentioned a perspective once or extensively. The table also shows the relationship between the mentioning



of this perspective with the total reflection score (Spearman's rho).

The perspectives mentioned by the students differed per case, as the cases differed in the order of importance of the possible perspectives. Some of the perspectives were hardly ever mentioned, or not at all. Students did not spontaneously empathise with the doctors or nurses involved in the situation. They did not often mention considerations about the patients' families, but if they did this had a positive impact on the overall reflection score. There was a consistent correlation between the mentioning of the patient perspective with a higher overall reflection score. Medical considerations were not often mentioned but were related with higher overall reflection. Ethical considerations were mentioned in cases 3 and 4, because of the more explicit moral dilemmas featured in these 2 case vignettes. These were highly correlated with overall reflection scores. The same was true for mentioning the position of the clerk in the hierarchy. Most prevalent were the perspectives concerning personal norms and personal feelings. In some, but not all, cases, this had a positive correlation with overall reflection score.

As there were no differences between the 4 conditions, we put them together for the analyses of the

other independent variables (Table 7). We found that female students gained higher scores on reflection than male students. This difference was significant in 3 of the 4 vignettes. In the vignette 4, there was no difference between female and male students. There was no relationship with age. Students with previous work experience in health care had higher reflection scores. This resulted in a significant difference in 2 vignettes but only in 1 vignette if controlling for sex (partial correlation), because more female than male students had previous work experience in health care.

In terms of the relationship between reflection scores and career preferences, we found that students who were considering general practice as a career obtained higher scores on all vignettes (significant at the  $P < 0.05$  level in 2 vignettes) than students who did not rate general practice as an option.

## CONCLUSIONS AND DISCUSSION

We have developed a semistructured, observer-rated instrument based on case vignettes to measure the degree of reflection of medical students in situations in medical practice. The interrater reliability proved sufficient, as did the internal consistency:

**Table 7 Differences in reflection scores between groups (concerning sex, previous health care work experience and general practice as career option)**

Case	Case reflection score			
	1	2	3	4
A Male	4.0* ( $\pm$ 1.7) <i>n</i> = 43	4.5 ( $\pm$ 1.8) <i>n</i> = 43	4.2* ( $\pm$ 1.4) <i>n</i> = 32	3.9* ( $\pm$ 2.0) <i>n</i> = 31
Female	4.9* ( $\pm$ 2.0) <i>n</i> = 74	4.6 ( $\pm$ 1.7) <i>n</i> = 76	5.0* ( $\pm$ 1.6) <i>n</i> = 65	5.0* ( $\pm$ 1.8) <i>n</i> = 63
B Health care work experience	4.9† ( $\pm$ 2.0) <i>n</i> = 60	4.5 ( $\pm$ 1.6) <i>n</i> = 62	5.0 ( $\pm$ 1.6) <i>n</i> = 48	5.0† ( $\pm$ 1.9) <i>n</i> = 46
No health care work experience	4.2† ( $\pm$ 1.8) <i>n</i> = 57	4.5 ( $\pm$ 1.9) <i>n</i> = 57	4.4 ( $\pm$ 1.5) <i>n</i> = 46	4.2† ( $\pm$ 2.0) <i>n</i> = 45
C General practice as career option	5.0 ( $\pm$ 2.2) <i>n</i> = 41	4.9 ( $\pm$ 1.7) <i>n</i> = 42	5.4‡ ( $\pm$ 1.5) <i>n</i> = 30	5.3‡ ( $\pm$ 1.9) <i>n</i> = 30
General practice not career option	4.4 ( $\pm$ 1.8) <i>n</i> = 77	4.3 ( $\pm$ 1.7) <i>n</i> = 78	4.4‡ ( $\pm$ 1.5) <i>n</i> = 64	4.3‡ ( $\pm$ 1.9) <i>n</i> = 61

\* Case 1:  $t = -2.4$ , d.f. = 115,  $P = 0.017$  (2-tailed); Case 3:  $t = -2.5$ , d.f. = 95,  $P = 0.014$ ; Case 4:  $t = -2.7$ , d.f. = 92,  $P = 0.009$ .

† Case 1:  $t = 2.0$ , d.f. = 115,  $P = 0.046$ ; Case 4:  $t = 2.1$ , d.f. = 89,  $P = 0.039$ .

‡ Case 3:  $t = 3.0$ , d.f. = 92,  $P = 0.004$ ; Case 4:  $t = 2.2$ , d.f. = 89,  $P = 0.033$ .

respondents had comparable scores on overall reflection on different vignettes. The large range of the scores enabled analysis of differences, and the rather low mean score allowed for measurement of improvement by interventions such as reflection-directed educational programmes in a subsequent study. There was no effect of pre-testing with an alternate form of the test and there was significant correlation of scores between 2 measurements with an interval of 4 weeks. This is a first indication that we measured a reasonably stable construct. However, we cannot yet rule out the possibility that reflection performance, as is known for professional competence, is task-specific.<sup>13</sup> The somewhat deviant results found in vignette 2 could be an indication of possible task-specificity. Moreover, from our study, no inferences can be made as to how many items are needed to achieve a stable score for the individual student.

We found there to be no effect on this instrument by a 9-hour course in clinical ethics. This may be an indication of the divergent validity of the instrument, as improvement in overall reflection was not the primary goal of this ethics course. However, there may be other reasons for this lack of effect. Firstly, the amount of education in this course could have been too small to have an effect. Self *et al.* found that at least 20 hours of small group discussion was needed to improve moral reasoning.<sup>14</sup> Secondly, the circumstances of the first and the second measurements were not completely identical. The first measurement took place in the first half-hour of the course, when students were fresh and interested. The last measurement took place during or even after the last hour of the course, at the end of the day, when students were probably less motivated to spend time completing the instrument. Mean reflection scores were in fact slightly lower at the second measurement. Although this was not statistically significant, it might be an indication that circumstances can influence the measurement. We measured a decrease in time spent on the second measurement compared to the first. In the first condition there was an additional handicap at the second measurement because these students had to fill in another questionnaire before the reflection measurement. This might have impaired motivation and so account for the lack of relationship between first and second measurements in this group, compared to the second group, which did not have to fill in an additional questionnaire and in which a strong relationship between first and second measurements was found. In future research, the context of the measurements has to be comparable in order to avoid this type of confounding.

The differences in total reflection scores concerning gender, career preference for general practice, and previous health care work experience were supportive of the construct validity. Case vignette 2 showed somewhat different results: in this case women and future general practitioners did not show a higher degree of reflection. A possible explanation could be that this was the case where the dilemma was between work – including education and patient care – and private life. This dilemma might have posed a more emotional dilemma than the other 3 case vignettes and this possibly reduced the degree of reflection. These somewhat deviant results could be indicative of task-specificity, but could also stimulate hypotheses on factors that might interfere with reflection. Personal (negative) emotional involvement might be such a factor.

The results of our study raise the question of whether reflection as we have measured it is a skill that can be taught and learnt, or a personality trait that may be hard to change. We think at least 3 factors are involved. Firstly, some people show a greater tendency to be reflective, which could be a personality disposition. The better reflection scores of female students and future GPs are indications of the contribution of this factor. Secondly, people show varying levels of skill or capability to perform reflective analyses, which probably reflects intellectual capacities and personal sensitivity, and also knowledge and experience. The contribution of work experience in health care is a first indication of this factor. Last but not least, the context and circumstances of the required reflection, as real or experienced pressure of time, may have an influence on reflection.

This study shows that our instrument has acceptable psychometric properties. Whether it is sufficiently sensitive to measure changes as a result of an educational programme will be the subject of further research.

---

## CONTRIBUTORS

ADB developed the protocol for the study, served as main researcher and wrote the first draft of the manuscript. PdJ assisted in development of the protocol, advised on methodological and statistical issues, participated in rating the material and commented on the manuscript. AKO assisted in development of the protocol, enabled and organised the research to be carried out at the Section of Medical Ethics, Vrije Universiteit Medical Centre, participated

in rating the material and commented on the manuscript. JAS assisted in development of the protocol, gave advice and expert information on the methodology of medical education research, participated in rating the material and commented on the manuscript. WvT assisted in development of the protocol, commented on the manuscript and actively supervised the research process.

---

## ACKNOWLEDGEMENTS

This research was made possible by the kind participation of the staff of the Department of Metamedica of the Vrije Universiteit Medical Centre.

---

## FUNDING

This research was made possible by a grant from the Onderwijs Kwaliteitsfonds (OKF).

---

## REFERENCES

- 1 American Association of Medical Colleges. Physicians of the 21st century. Report of the Working Group on Personal Qualities, Values and Attitudes. *J Med Educ* 1984;**59**:177–89.
- 2 Metz JCM, Pels Rijcken-Van Erp Taalman Kip EH, van den Brand-Valkenburg BWM. Blueprint 1994: training of doctors. Objectives of undergraduate medical education in The Netherlands. Nijmegen: University of Nijmegen, University Publication Office 1994.
- 3 Niemi PM. Medical students' professional identity: self-reflection during the preclinical years. *Med Educ* 1997;**31** (6):408–15.
- 4 Sobral DT. An appraisal of medical students' reflection-in-learning. *Med Educ* 2000;**34** (3):182–7.
- 5 Kohlberg L. *The Philosophy of Moral Development*. San Francisco: Harper & Row 1981.
- 6 Rest JR, Narvaez D. *Moral Development in the Professions: Psychology and Applied Ethics*. Hillsdale, New Jersey: Lawrence Erlbaum Associates 1994.
- 7 Schön DA. *The Reflective Practitioner. How Professionals Think in Action*. London: Temple Smith 1983.
- 8 Schön DA. *Educating the Reflective Practitioner*. San Francisco, London: Jossey-Bass 1987.
- 9 Wong FKY, Kember D, Chung LY, Yan L. Assessing the level of student reflection from reflective journals. *J Adv Nurs* 1995;**22** (1):48–57.
- 10 Schoenberg NE, Ravdal H. Using vignettes in awareness and attitudinal research. *Int J Soc Res Methodology* 2000;**3** (1):63–74.
- 11 Huijter M, van Leeuwen E, Boenink A, Kimsma G. Medical students' cases as an empirical basis for teaching clinical ethics. *Acad Med* 2000;**75** (8): 834–9.
- 12 Batenburg V, Smal JA, Lodder A, de Melker RA. Are professional attitudes related to gender and medical specialty? *Med Educ* 1999;**33** (7):489–92.
- 13 Van der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ* 1996;**1**: 41–67.
- 14 Self DJ, Olivarez M, Baldwin DC Jr. The amount of small-group case-study discussion needed to improve moral reasoning skills of medical students. *Acad Med* 1998;**73** (5):521–3.

*Received 7 January 2002; editorial comments to authors 16 October 2002, 25 March 2003, 30 May 2003; accepted for publication 7 August 2003*