

Web-assisted assessment of professional behaviour in problem-based learning: more feedback, yet no qualitative improvement?

Walther N. K. A. van Mook · Arno M. M. Muijtjens · Simone L. Gorter · Jan Harm Zwaveling · Lambert W. Schuwirth · Cees P. M. van der Vleuten

Received: 23 December 2010 / Accepted: 14 April 2011 / Published online: 1 May 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Although other web-based approaches to assessment of professional behaviour have been studied, no publications studying the potential advantages of a web-based instrument versus a classic, paper-based method have been published to date. This study has two research goals: it focuses on the quantity and quality of comments provided by students and their peers (two researchers independently scoring comments as correct and incorrect in relation to five commonly used feedback rules (and resulting in an aggregated score of the five scores) on the one, and on the feasibility, acceptability and perceived usefulness of the two approaches on the other hand (using a survey). The amount of feedback was significantly higher in the web-based group than in the paper based group for all three categories (dealing with work, others and oneself). Regarding the quality of feedback, the aggregated score for each of the three categories was not significantly different between the two groups, neither for the interim, nor for the final assessment. Some, not statistically significant, but nevertheless noteworthy trends were nevertheless noted. Feedback in the web-based group was more often unrelated to observed behaviour for several categories for both the interim and final assessment. Furthermore, most

W. N. K. A. van Mook (✉)

Departments of Intensive Care and Internal Medicine, Maastricht University Medical Centre (MUMC),
P. Debyelaan 25, 6202 AZ Maastricht, The Netherlands
e-mail: w.van.mook@mumc.nl

W. N. K. A. van Mook · A. M. M. Muijtjens · L. W. Schuwirth · C. P. M. van der Vleuten
Department of Educational Development and Research, Faculty of Health, Medicine, and Life
Sciences, University of Maastricht, Maastricht, The Netherlands

S. L. Gorter
Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Centre,
Maastricht, The Netherlands

J. H. Zwaveling
Department of Intensive Care Medicine, Maastricht University Medical Centre, Maastricht,
The Netherlands

J. H. Zwaveling
Maxima Medical Centre, Veldhoven, The Netherlands

comments relating to the category ‘Dealing with oneself’ consisted of descriptions of a student’s attendance, thereby neglecting other aspects of personal functioning. The survey identified significant differences between the groups for all questionnaire items regarding feasibility, acceptability and perceived usefulness in favour of the paper-based form. The use of a web-based instrument for professional behaviour assessment yielded a significantly higher number of comments compared to the traditional paper-based assessment. Unfortunately, the quality of the feedback obtained by the web-based instrument as measured by several generally accepted feedback criteria did not parallel this increase.

Keywords Problem based learning · Professional behaviour · Professionalism · Tutorial group · Assessment · Peer · Web-assisted · E-mail · Electronic

Abbreviations

APB Assessment of Professional Behaviours
NBME National Board of Medical Examiners
PB Professional behaviour

Introduction

Professionalism is becoming increasingly central in undergraduate and postgraduate training, and the herewith associated research results in a vast increase in the number of papers on the topic (van Mook, de Grave et al. 2009). Tools for assessing professionalism and professional behaviour have been developed to identify, counsel, and remediate the performance of students and trainees demonstrating unacceptable professional behaviour (Papadakis et al. 2005, 2008). Since validated tools are scarce (Cruess et al. 2006), combining currently available instruments has become the current norm (Schuwirth and van der Vleuten 2004; van Mook, Gorter et al. 2009). Self- and peer assessment and direct observation by faculty during regular educational sessions (Singer et al. 1996; Asch et al. 1998; Fowell and Bligh 1998; van der Vleuten and Schuwirth 2005; Cohen 2006) are some of these tools. Self-assessment is defined as personal evaluation of one’s professional attributes and abilities against perceived norms (Eva et al. 2004; Eva and Regehr 2005; McKinstry 2007). So far, there is a scarcity of published studies on self-assessment of professionalism (Rees and Shepherd 2005). Given the poor validity of self-assessment in general (Eva and Regehr 2005), it seems ill advised to use self-assessment in isolation without triangulation from other sources. Peer assessment involves assessors with the same level of expertise and training and similar hierarchical institutional status. Medical students usually know which of their classmates they would trust to treat their family members, which illustrates the intrinsic potential of peer assessment (Dannefer et al. 2005). However, a recent analysis of instruments for peer assessment of physicians revealed that none met the required standards for instrument development (Evans et al. 2004). Studies addressing peer assessment of professional behaviour of medical students are beginning to appear (Freedman et al. 2000; Arnold et al. 2005; Dannefer et al. 2005; Shue et al. 2005; Lurie, Nofziger et al. 2006a, b). Observation and assessment by faculty using rating scales is another commonly used method of professional behaviour assessment (van Luijk et al. 2000; van Mook, Gorter et al. 2009; van Mook and van Luijk 2010). Prior studies have revealed that such teacher-led sessions are highly dependent on the teacher’s attitudes, motivation and instructional skills

(van Mook et al. 2007). When teachers' commitment declines, assessment of professional behaviour may become more trivialised. This may misplace emphasis on attendance rather than participation and on completion of tick boxes rather than informative feedback and students' contribution and motivation (van Mook et al. 2007). In an attempt to further improve professional behaviour assessment, the triangulated teacher-led discussion of self- and peer-assessment of professional behaviour using a paper form is the contemporary practice at Maastricht medical school (van Mook and van Luijk 2010).

However, digital technologies have come to influence our ways of working and communicating, and created technology driven ways of teaching, learning and assessing (De Leng 2009). Adaptation to some of these changes can be useful (De Leng 2009), for example to reduce the time and expense involved in collecting self and peer ratings and facilitate anonymous information gathering and information analysis. In the National Board of Medical Examiners (NBME)' Assessment of Professional Behaviours (APB) program such web-based technology is contemporarily used (Mazor et al. 2007, 2008; National Board of Medical Examiners 2010). The study presented in this paper investigated the potential advantages of a web-based instrument versus a 'classic', paper-based method to assess professional behaviour in tutorial groups in a problem-based curriculum. In a comparison of these two approaches we focused on:

1. The quantity and quality of comments provided by students and the feedback provided by their tutor and peers, and on
2. The feasibility, acceptability and perceived usefulness of the two approaches.

Methods and research tools

The study involved *all* medical students enrolled in the second, ten-week course in year 2 at the Faculty of Health, Medicine and Life Sciences, Maastricht University, the Netherlands. During the bachelor programme of the six-year problem-based medical curriculum, professional behaviour is assessed on various occasions in tutorial groups during all regular courses (van Mook and van Luijk 2010). Each tutorial group consists of ten students on average and a tutor/facilitator, and each meeting lasts 2 h. For the purpose of this study, the students were divided into two groups: those in tutorial groups with even numbers and those in groups with odd numbers. The first group used a web-based instrument to assess professional behaviour and the other group used the usual method with a paper assessment form. We will first describe the two assessment methods in some detail.

The 'classic' paper-based professional behaviour assessment form

The working group Consilium Abeundi of the Association of Universities in the Netherlands, proposed a *practical* definition of *professional behaviour* (Project Team Consilium Abeundi van Luijk 2005; van Luijk et al. 2010; van Mook and van Luijk 2010). They framed professionalism as *observable* behaviours, reflecting the norms and values of the medical professional. Three categories of professional behaviour were distinguished: 'Dealing with work and tasks', 'Dealing with others', and 'Dealing with self-functioning' (van Luijk et al. 2000; Project Team Consilium Abeundi van Luijk 2005). These categories, together with the related clarifying descriptions, are the basis of the professional behaviour assessment form that is in use at Maastricht medical school since 2002 (Fig. 1) (van Mook and van Luijk 2009, 2010). Early in the curriculum students are familiarised

with its use. Professional behaviour is assessed at the start, halfway through and at the end of each regular course. For the halfway assessment each student prepares a self-reflective assessment and enters it in the form using the clarifying descriptions as a reminder and starting point. In the subsequent plenary session of the tutorial group, chaired by the tutor, the professional behaviour of each student is assessed by the group. All group members (students and tutor) are required to contribute to the discussion. All the comments and feedback are documented on each student's form by the tutor. At the end of the course this process is repeated, followed by a final summative assessment, resulting in a pass or fail (van Mook and van Luijk 2010).

The web-based instrument

The web-based instrument is based on an application that consists of a 360° feedback system specifically designed for higher education. Its development involved more than thirty pilot studies and evaluations by over 6,000 students. Prior to the current study, the tool was piloted at Maastricht in a group of first year students, which did not participate in the current study. Providing adequate practical information to students and tutors prior to using the application and rephrasing of items to achieve a more detailed focus on aspects of professional behaviour were considered prerequisite for the successful implementation of web-based assessment (unpublished data). The web-based instrument used for assessment of professional behaviour pertained to the same three categories (and clarifying descriptions) also used on the paper form (Project Team Consilium Abeundi van Luijk 2005). Ample information about background, confidentiality, timing and some practical matters was made available to students and staff electronically and in writing prior to, and during

Assessment form for professional behaviour in tutorial groups in years 1-3, academic year

ID number:	Name of the student:	Block	<input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	Only required for END OF BLOCK ASSESSMENT: Signature of the student (indicating that the student has read the completed form): _____ Signature of the tutor: _____
	Name of the tutor:	Group	<input type="checkbox"/> 0 <input type="checkbox"/> 1 <input type="checkbox"/> 2 <input type="checkbox"/> 3 <input type="checkbox"/> 4 <input type="checkbox"/> 5 <input type="checkbox"/> 6 <input type="checkbox"/> 7 <input type="checkbox"/> 8 <input type="checkbox"/> 9 <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	
Dealing with work*	Preliminary/interim assessment	Final assessment		
Areas requiring the student's attention: <ul style="list-style-type: none"> • preparation for tasks • thoroughness of completion of tasks • preliminary discussion of tasks • contribution to group effort • reporting findings from self study • other 	<input type="checkbox"/> requires attention <input type="checkbox"/> satisfactory comments: _____	<input type="checkbox"/> requires attention <input type="checkbox"/> satisfactory comments: _____		
Dealing with others*	Preliminary/interim assessment	Final assessment		
Areas requiring the student's attention: <ul style="list-style-type: none"> • teamwork • listening • leading the discussion • summarising • other 	<input type="checkbox"/> requires attention <input type="checkbox"/> satisfactory comments: _____	<input type="checkbox"/> requires attention <input type="checkbox"/> satisfactory comments: _____		
Dealing with oneself*	Preliminary/interim assessment	Final assessment		
Areas requiring the student's attention: <ul style="list-style-type: none"> • dealing with feedback • giving/asking for feedback • self reflection • honouring commitments • being on time for appointments/time management • other 	<input type="checkbox"/> requires attention <input type="checkbox"/> satisfactory comments: _____	<input type="checkbox"/> requires attention <input type="checkbox"/> satisfactory comments: _____		
PRELIMINARY/INTERIM ASSESSMENT: <input type="checkbox"/> unsatisfactory <input type="checkbox"/> satisfactory <input type="checkbox"/> reliable assessment not possible due to frequent non-attendance of the student		FINAL ASSESSMENT: <input type="checkbox"/> unsatisfactory <input type="checkbox"/> satisfactory <input type="checkbox"/> reliable assessment not possible due to frequent non-attendance of the student		
Requires attention: The areas referred to in the comments are in need of attention and improvement. Satisfactory: The student is functioning adequately within the present circumstances.		Unsatisfactory: below the expected average level of functioning in the tutorial group, clear areas of weakness where more work and improvement are required have been identified. Satisfactory: expected or higher level of functioning in the tutorial group, although areas where more work and improvement are required may be present. Cannot be assessed: reliable assessment is not possible because the student has frequently failed to attend group sessions.		
* Details and explanations of the categories are provided on the back of the form.				
Tutor: If final assessment is UNSATISFACTORY OR ASSESSMENT NOT POSSIBLE, please send 1st copy to Secretary of the Committee of Professional Behaviour, Department of Educational Development and Research, UNS 60, Room N4.13 Student: one copy should be inserted in the PORTFOLIO one copy should be kept by the student for his/her own records				

Fig. 1 The paper form used for professional behaviour evaluation and assessment at the Faculty of Health, Medicine and Life Sciences, Maastricht University, The Netherlands

the opening session of the course, as well as verbally to the tutors during the tutor instruction session. Halfway and at the end of the course each student in the web-based assessment group received an internet link in an e-mail. Clicking the link gave access to the web-based assessment instrument. The students were asked to complete the questions themselves and then invite five peers and the tutor of their group to evaluate their professional behaviour and provide feedback. Selection of the peer students was standardised to the five students listed immediately below the student's name on the centrally randomly generated list of the tutorial group members, resulting in a semi-anonymised feedback procedure. Ample space for narrative feedback relating to the three categories of professional behaviour was provided for each questionnaire item (Fig. 1). All items were also answered using a Likert scale (1 = almost never to 5 = almost always). The students received the results of the feedback process in the form of a printable report presenting the results of their self-assessment relative to the assessment by their peers as well as an overview of all the narrative comments. The web-based group used the printed reports and the paper-based group used the completed paper-based forms to discuss each student's professional behaviour during the end-of-course assessment in the final tutorial group of the course.

End-of-course questionnaire

At the end of the last tutorial group of the course, *all* students of the two groups were asked to complete a questionnaire addressing fourteen aspects of feasibility, acceptability and perceived usefulness of the two instruments. The tutors were invited to report their findings by e-mail. All data were recorded and analysed anonymously.

Analysis

All narrative comments were independently coded and analysed by two blinded researchers (WvM and SG). Units of comments consisting of one grammatical clause but covering different topics were considered to be different units of comments. They used five generally accepted feedback rules to label the units of comments as incorrect or correct feedback (Table 1) (Pendleton and Schofield 1984; Branch and Paranjape 2002). Since feedback should preferably meet as many criteria as possible, an aggregated score across all five feedback categories was constructed (total scores of 0–3 were considered unsatisfactory and requiring improvement; total scores of 4 and 5 were considered satisfactory). The researchers (WvM and SG) discussed any discrepancies in their coding until agreement was reached. The number and nature of the comments from the interim and final evaluations were compared between the web-based (intervention) group, and the paper-based (control)

Table 1 Commonly used feedback rules (Pendleton and Schofield 1984; Branch and Paranjape 2002)

No.	Adequate feedback	Inadequate feedback
1	Is clear and concrete	Is vague and general
2	Is constructive and positive	Is destructive and negative
3	Is specific	Is non-specific
4	Comments on behaviour	Comments on personality
5	Is descriptive (formative)	Is evaluative (summative)

group using Pearson's Chi Square test. Independent samples t testing was used to perform quantitative survey analysis of the questionnaire scores. SPSS 16.0.1 was used for the statistical analyses (SPSS 2007). Effect sizes were computed where applicable as an indicator of the results' practical (clinical) relevance (independent of sample size). For between-group comparison of proportions p_{paper} , and p_{web} , effect sizes (ES) were calculated according to

$$ES = 2|\arcsin \sqrt{p_{\text{web}}} - \arcsin \sqrt{p_{\text{paper}}}|$$

For between-group comparisons of mean scores m_{paper} , and m_{web} , effect sizes ES were calculated according to

$$ES = |m_{\text{web}} - m_{\text{paper}}|/\sigma_{\text{paper}}$$

where σ_{paper} is the standard deviation in the paper-based group. $ES \cong 0.30$ was considered a small effect of negligible practical importance; $ES \cong 0.50$ was considered a medium effect of moderate practical importance, and $ES \cong 0.80$ large effect of considerable practical importance (Cohen 1987; Hojat and Xu 2004).

Results

Of the 307 (198 females, 109 males) medical students enrolled in the second course of the second year, 150 were assigned to tutorial groups with even numbers (web-based group) and 157 to the tutorial groups with odd numbers (paper-based group). Since assessment of professional behaviour is mandatory, the participation rate was 100%. We will first present the results of the quantitative and qualitative analyses of the narrative feedback provided at the interim and final assessment of professional behaviour in the two groups. After that we present the results for the feasibility, acceptability and perceived usefulness of the two assessment instruments.

Analysis of the amount of feedback

The total numbers of comments per category of professional behaviour for the interim and the final assessments are presented for the web-based and the paper-based group (Table 2). In the web-based group a mean of 4.5 invitations to peers to provide feedback were sent by each student (standard deviation 3). The number of comments was significantly higher in this group than in the paper-based group for *all* three categories (dealing with work, others and oneself). However, at the final assessment the total number of comments relating to the category 'Dealing with oneself' had halved compared to the interim assessment and this decrease was mainly attributable to a marked decrease in the comments provided by the web-based group.

Analysis of the quality of the feedback

The results for the quality of feedback at the interim and the final assessments are presented for the categories of professional behaviour (dealing with work, others and oneself) as correct and incorrect in relation to the aggregated score on the five feedback rules (Table 3). The inter-rater agreement during primary coding was 15834 out of 16165 codes (97.9%) for 3233 comments ($Kappa = 0.87$).

Table 2 The total number of comments per category of professional behaviour in the web-based and the paper-based group at the interim and final assessment

	Number of comments [n (%)]				<i>p</i> -value, χ^2	Effect size
	Total	Paper-based	Web-based			
Interim assessment						
Dealing with work	983 (100)	161 (16.4)	822 (83.6)	<0.001	0.76	
Dealing with others	692 (100)	105 (15.2)	587 (84.4)	<0.001	0.80	
Dealing with oneself	539 (100)	95 (17.6)	444 (82.4)	<0.001	0.73	
Final assessment						
Dealing with work	455 (100)	132 (29.0)	323 (71.0)	<0.001	0.46	
Dealing with others	337 (100)	92 (27.5)	245 (72.5)	<0.001	0.49	
Dealing with oneself	227 (100)	77 (33.9)	150 (66.1)	<0.001	0.35	

χ^2 = Chi-square

Table 3 Number of correct and incorrect scores in the web-based and paper-based groups regarding the aggregated scores on the five feedback criteria for the categories of professional behaviour at the interim and final assessment

	Aggregated score				<i>p</i> -value, χ^2	Effect size
	Sufficient (n)		Insufficient (n)			
	Paper-based	Web-based	Paper-based	Web-based		
Interim assessment						
Dealing with work	146	783	15	39	0.020	0.18
Dealing with others	86	458	19	129	0.372	na
Dealing with oneself	85	367	10	77	0.101	na
Final assessment						
Dealing with work	121	308	11	15	0.124	na
Dealing with others	84	210	9	35	0.261	na
Dealing with oneself	65	139	12	11	0.051	na

na = not applicable; χ^2 = Chi-square

When statistical significance was corrected for sample size no difference in the aggregated scores on all five feedback categories was found. Several differences of potential practical significance pertaining to the five generally accepted feedback rules were however, found, all favouring paper-based assessment. The feedback in the web-based group pertaining to the categories 'Dealing with others' and 'Dealing with oneself' was for example more often unrelated to observed behaviour. However, in the view of the statistical 'multiple comparisons' problem, the statistical significance regarding the between-groups differences revealed by these individual analysis is questionable. Finally, the majority of comments relating to the category 'Dealing with oneself' consisted of descriptions of a student's attendance to the neglect of other aspects of personal functioning.

Feasibility, acceptability and perceived usefulness

The response to the questionnaire was 96% (143/157) in the paper-based group and 81% (121/150) in the 'web-based' group. Table 4 shows the mean scores per questionnaire item for the two groups. The differences between the groups were significant for all items and in favour of the paper-based form. The paper-based group yielded fourteen student comments on the use of the paper form. Nine comments emphasised the usefulness of the form, four offered suggestions for minor adaptations and one stated that the form was time consuming. The tutors did not comment on this form. The students in the web-based groups provided 179 comments (Table 5) and the tutors provided fifteen comments. The tutors emphasised the absence of interpersonal contact and the necessity of face-to-face contact during assessment of professional behaviour. They also thought that the web-based instrument was time consuming and they reported some technical problems (e.g. printing).

Discussion

Although other web-based approaches to assessment of professional behaviours have been studied (Mazor et al. 2007, 2008; Stark et al. 2008) as well as are contemporarily used(-National Board of Medical Examiners 2010), very few studies specifically address the amount and quality of feedback resulting from using such approach. The assessment method that is currently used at Maastricht medical school requires each student to reflect on their professional behaviour and requires all members of a tutorial group (tutor and students) to provide feedback on the professional behaviour of each student, which is then recorded by the tutor on the assessment form. This process was deliberately mimicked in the web-based instrument, which elicited feedback from students and tutor on the same three categories and items relating to professional behaviour that are included in the paper form.

The study reveals that the number of comments was significantly higher in the web-based group compared with the paper-based group. The quality of the feedback, however, did not parallel the quantitative increase. When considering the aggregated scores on the five feedback criteria no differences in quality of feedback was found between the groups (Table 3). Nevertheless, the feedback provided by the web-based group showed poorer quality in relation to several feedback criteria (e.g. was unrelated to the observed behaviour; data not shown). However, as previously mentioned, the statistical significance revealed by these individual analysis is questionable.

Moreover, the survey results on acceptability, feasibility and usefulness of the instruments were strongly in favour of the paper form. It should be noted that this result might be partly due to technical difficulties that were experienced with the web-based instrument despite adequate technical preparation and extensive tutor and student instruction. However, even when these limitations are taken into account, the web-based instrument did not show an *improvement* in educational impact compared with the existing method of assessing professional behaviour.

Another striking finding, which is unrelated to the nature of the assessment instrument, was the emphasis on attendance in the feedback relating to the category 'Dealing with oneself' and the relative absence of feedback on other aspects of self-functioning. An earlier analysis of 4 years of experience at Maastricht with paper-based assessment of professional behaviour had yielded similar findings (van Mook and van Luijk 2010). This suggests that the context (small group sessions) in the earlier years of medical school may be less suited to stimulate self-reflection (van Mook and van Luijk 2010). Perhaps

Table 4 Scores on the items of the questionnaire on the acceptability, feasibility and usefulness of web-based versus paper-based assessment of professional behaviour

No.	Questionnaire item	Web-based PB assessment (n = 121)		Paper-based PB assessment (n = 143)		Web-based versus paper-based	
		Mean	Standard deviation	Mean	Standard deviation	Significance* (p-value)	Effect size
1	The program was easy accessible using the link	4.19	1.16	na	na	na	na
2	The program/form was easy to use	3.84	1.30	4.18	0.93	0.02	0.37
3	The program/form was clear	3.53	1.32	4.15	0.92	<0.05	0.67
4	Completing the program/form contributed to self-reflection on my personal functioning	2.65	1.32	3.69	1.12	<0.05	0.93
5	The output (results, report) from the program was/were clear	2.61	1.2	na	na	na	na
6	The tutor has used the program's results (report)/form as basis for discussing professional behaviour in the tutorial group	2.27	1.56	4.20	0.96	<0.05	2.01
7	Discussing the program's results (report)/completed form in the tutorial group contributed to self reflection	2.18	1.28	3.89	0.98	<0.05	1.74
8	The program's results (report) increased the usefulness of the professional behaviour evaluation in the tutorial group	1.94	1.2	na	na	na	na
9	I recognize the strengths and weaknesses identified by my peers and/or tutor	2.93	1.08	3.96	0.93	<0.05	1.11
10	The time and effort needed to complete the program/form were worthwhile (agree = 1/disagree = 2)	1.85	0.70	1.26	0.70	<0.05	0.84
11	Time needed to complete the program/form (in minutes)	24.09	18.89	6.29	8.63	<0.05	2.06
12	Give a mark out of ten for ease of use (whole mark, 1–10)	6.57	1.69	7.60	1.21	<0.05	0.85
13	Give a mark out of ten for the usefulness of the evaluation of professional behaviour (whole mark, 1–10)	4.55	1.98	7.00	1.46	<0.05	1.68
		No.	Standard deviation	No.	Standard deviation	Significance* (p-value)	Effect size
14	Additional comments	179	na	14	na	na	na

Items 1–7 were answered using a 1–5 Likert scale; 1 = completely disagree, 5 = completely agree

na = not applicable

* = Independent samples *t* test

attendance and time management as measures of responsible behaviour should be evaluated separately from feedback on other aspects of professional behaviour, a suggestion that was also put forward during the plenary discussion at a recent symposium on

Table 5 Remarks by students regarding web-based professional behaviour assessment

Category	Description	Number of comments
1	No added value	28
2	Difficult to use	25
3	Difficulty with internet access, including late receipt of link or report	29
4	Some questions need rephrasing	26
5	Electronic process of professional behaviour assessment too standardised	14
6	Time consuming	13
7	Hampers provision of feedback	11
8	Rest (including privacy issues, requests for assessment of professional behaviour of tutor, plea for paper-based professional behaviour assessment)	16
9	Problems with report, such as print problems and the presentation of results	10
10	Process of professional behaviour evaluation becomes less personal, more distant	7

professionalism (Centre for Excellence in Developing Professionalism 2010; van Mook and van Luijk 2010).

Comparison of the results of this study to results reported in the literature is difficult since few studies on web-based instruments to assess professional behaviour have been published. However, there are some published reports on the development and use of web-based assessment in general (Wheeler et al. 2003; Tabuenca et al. 2007). In one study, implementation of a web-based instrument resulted in a substantial reduction in administration and bureaucracy for course organisers and proved to be a valuable research tool, while students and teachers were overwhelmingly in favour of the new course structure (Wheeler et al. 2003). Another study described a successful multi-institutional validation of a web-based core competency assessment system in surgery (Tabuenca et al. 2007). However, the transferability of these more general studies to web-based self- and peer-assessment of professional behaviour seems limited. Studies addressing the NMBE's APB program however, report comparable promising results, with improved faculty comfort and self-assessed skill in giving feedback about professionalism as an example (Stark et al. 2008). It seems therefore advisable to conduct further studies to examine the effectiveness and optimal use of web-based assessment of professional behaviour.

Although the literature pertaining to web-based assessment is sparse, the peer assessment literature provides evidence of the importance of anonymity, or at least confidentiality, for the acceptance of peer assessment (Arnold et al. 2005; Shue et al. 2005). That is why we used a semi-anonymous feedback procedure in the web-based assessment in this study. Although reliability can be enhanced by increasing the number of raters (Ramsey et al. 1993; Dannefer et al. 2005), the desired number of raters may not be feasible or acceptable, for example due to time constraints. Consequently, in the current study we limited the number of peer raters to five randomly selected students (Ramsey et al. 1993; Dannefer et al. 2005). It seems reasonable for medical schools to base the selection of peer raters on practical and logistical considerations (Arnold et al. 1981; Arnold and Stern 2006; Lurie, Nofziger et al. 2006a, b), since bias due to rater selection has been shown not to affect peer assessment results (Lurie, Nofziger et al. 2006a, b). Although some anticipated problems could thus be adequately addressed, mention must be made of some remaining limitations of the current study.

Study limitations

In the preparation phase we were confronted with limited availability of tools for web-based multisource feedback. Because re-designing an existing tool proved costly, a for the purpose of this study superfluous feature, (the Likert scales), was left unchanged, and this may have unavoidably influenced the results, for instance those relating to time investment. The content of the web-assisted and paper versions of the instrument was otherwise identical. Furthermore, it cannot be excluded that the results were negatively affected by the participants' unfamiliarity with web-based assessment instruments, even though the implementation process was carefully prepared based on feedback from a pilot study. Ample time was spent on technical preparation and the participants received information and instruction on multiple occasions. Furthermore, the possibility of omitting redundant comments and concretizing comments before noting them by the tutor, or the more limited space may have contributed to the lower number and/or the higher quality of the comments in the paper based group. Finally, automated data extraction only enabled feedback analysis at the level of the whole year group, although analysis of data at individual (students, tutors) or tutorial group level would have been preferable.

Conclusions

The results revealed that a confidential web-based assessment instrument for professional behaviour yielded a significantly higher number of comments compared to the traditional paper-based assessment. The quality of the feedback obtained by the web-based instrument was comparable as measured by several generally accepted feedback criteria. However, judging by the questionnaire results students strongly favoured the use of the traditional paper-based method. The interpersonal nature of professional behaviour prompted comments that professional behaviour was eminently suitable for 'en-groupe', face-to-face discussion and assessment. Although teachers and students are nowadays preferably 'wired for learning' it seems that, so far, professional behaviour assessment does not necessarily require the use of advanced assessment technologies, although such new 'innovative' electronic and/or web-based assessment methods thus do result in more feedback of comparable quality. Their exact position among the currently used, labour-intensive traditional assessment armamentarium needs to be subject of further study.

Acknowledgments The authors wish to thank the following colleagues for their contributions. Kirsten Thijssen, medical student, University of Maastricht for her contributions to data entry from the paper forms; Renee Stalmeijer, Department of Educational Development and Research, Maastricht University for her help in developing the survey and all students and tutors in block 2.2 in the 2008–2009 academic year for participating in this study. Furthermore, the authors thank ms. Mereke Gorsira, Department of Educational Development and Research, Faculty of Health, Medicine and Life Sciences, Maastricht University, Maastricht, The Netherlands, for critically reviewing the manuscript regarding use of the English language. Finally, the authors are indebted to Mrs. Anita Legtenberg, data manager MEMIC, Maastricht University, The Netherlands, for assistance in managing the dataset from the web-based groups.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

- Arnold, L. and D. Stern (2006). Content and context of peer assessment. In: D. T. Stern (ed.), *Measuring medical professionalism*. New York: Oxford University Press. ISBN-13: 978-0-19-517226-3.
- Arnold, L., Shue, C. K., et al. (2005). Medical students' views on peer assessment of professionalism. *Journal of General Internal Medicine*, 20(9), 819–824.
- Arnold, L., Willoughby, L., et al. (1981). Use of peer evaluation in the assessment of medical students. *Journal of Medical Education*, 56(1), 35–42.
- Asch, E., Saltzberg, D., et al. (1998). Reinforcement of self-directed learning and the development of professional attitudes through peer- and self-assessment. *Academic Medicine*, 73(5), 575.
- Branch, W. T., & Paranjape, A. (2002). Feedback and reflection: Teaching methods for clinical settings. *Academic Medicine*, 77(12 Pt 1), 1185–1188.
- Centre for Excellence in Developing Professionalism (2010). Professionalism or post-professionalism? Four years of the centre for excellence. Liverpool, UK (chair H. O'Sullivan).
- Cohen, J. (1987). *Statistical power analysis for behavioral sciences*. Hillsdale, NJ: Erlbaum. ISBN 0-8058-0283-5.
- Cohen, J. J. (2006). Professionalism in medical education, an American perspective: From evidence to accountability. *Medical Education*, 40(7), 607–617.
- Cruess, R., McIlroy, J. H., et al. (2006). The professionalism mini-evaluation exercise: A preliminary investigation. *Academic Medicine*, 81(10 Suppl), S74–S78.
- Dannefer, E. F., Henson, L. C., et al. (2005). Peer assessment of professional competence. *Medical Education*, 39(7), 713–722.
- De Leng, B. (2009). Wired for learning: How computers can support interaction in small group learning in higher education. Thesis. Mediview, Maastricht. ISBN:978-90-77201-35-0.
- Eva, K. W., Cunnington, J. P., et al. (2004). How can I know what I don't know? Poor self assessment in a well-defined domain. *Advances in Health Sciences Education: Theory and Practice*, 9(3), 211–224.
- Eva, K. W., & Regehr, G. (2005). Self-assessment in the health professions: A reformulation and research agenda. *Academic Medicine*, 80(10 Suppl), S46–S54.
- Evans, R., Elwyn, G., et al. (2004). Review of instruments for peer assessment of physicians. *British Medical Association*, 328(7450), 1240.
- Fingertips <http://www.fngtps.com/work>. Accessed 16 Mar 2011
- Fowell, S. L., & Bligh, J. G. (1998). Recent developments in assessing medical students. *Journal of Postgraduate Medicine*, 74(867), 18–24.
- Freedman, J. A., Lehmann, H. P., et al. (2000). Web-based peer evaluation by medical students. *Academic Medicine*, 75(5), 539–540.
- Hojat, M., & Xu, G. (2004). A visitor's guide to effect sizes: Statistical significance versus practical (clinical) importance of research findings. *Advances in Health Sciences Education: Theory and Practice*, 9(3), 241–249.
- Lurie, S. J., Nofziger, A. C., et al. (2006a). Effects of rater selection on peer assessment among medical students. *Medical Education*, 40(11), 1088–1097.
- Lurie, S. J., Nofziger, A. C., et al. (2006b). Temporal and group-related trends in peer assessment amongst medical students. *Medical Education*, 40(9), 840–847.
- Mazor, K. M., Canavan, C., et al. (2008). Collecting validity evidence for an assessment of professionalism: Findings from think-aloud interviews. *Academic Medicine*, 83(10 Suppl), S9–S12.
- Mazor, K., Clauser, B. E., et al. (2007). Evaluation of missing data in an assessment of professional behaviors. *Academic Medicine*, 82(10 Suppl), S44–S47.
- McKinstry, B. (2007). BEME guide no 10: A systematic review of the literature on the effectiveness of self-assessment in clinical education. <http://www.bemecollaboration.org/beme/files/BEME%20Guide%20No%2010/BEMEFinalReportSA240108.pdf>. Accessed October 10th 2008.
- National Board of Medical Examiners (2010). Assessment of professional behaviors program. <http://www.nbme.org/schools/apb/index.html>. Accessed September 24th 2010.
- Papadakis, M. A., Arnold, G. K., et al. (2008). Performance during internal medicine residency training and subsequent disciplinary action by state licensing boards. *Annals of Internal Medicine*, 148(11), 869–876.
- Papadakis, M. A., Teherani, A., et al. (2005). Disciplinary action by medical boards and prior behavior in medical school. *he New England Journal of Medicine*, 353(25), 2673–2682.
- Pendleton, D., & Schofield, T., et al. (1984). A method for giving feedback. In: *The consultation: An approach to learning and teaching*. Oxford: Oxford University Press, pp. 68–71.
- Project Team Consilium Abeundi van Luijk, S. J. e. (2005). Professional behaviour: Teaching, assessing and coaching students. Final report and appendices. Mosae Libris.

- Ramsey, P. G., Wenrich, M. D., et al. (1993). Use of peer ratings to evaluate physician performance. *The Journal of the American Medical Association*, 269(13), 1655–1660.
- Rees, C., & Shepherd, M. (2005). Students' and assessors' attitudes towards students' self-assessment of their personal and professional behaviours. *Medical Education*, 39(1), 30–39.
- Schuwirth, L. W., & van der Vleuten, C. P. (2004). Changing education, changing assessment, changing research? *Medical Education*, 38(8), 805–812.
- Shue, C. K., Arnold, L., et al. (2005). Maximizing participation in peer assessment of professionalism: The students speak. *Academic Medicine*, 80(10 Suppl), S1–S5.
- Singer, P. A., Robb, A., et al. (1996). Performance-based assessment of clinical ethics using an objective structured clinical examination. *Academic Medicine*, 71(5), 495–498.
- SPSS, Inc. (2007). SPSS 16.0.1.
- Stark, R., Korenstein, D., et al. (2008). Impact of a 360-degree professionalism assessment on faculty comfort and skills in feedback delivery. *Journal of General Internal Medicine*, 23(7), 969–972.
- Tabuenca, A., Welling, R., et al. (2007). Multi-institutional validation of a web-based core competency assessment system. *Journal of Surgery Education*, 64(6), 390–394.
- van der Vleuten, C. P., & Schuwirth, L. W. (2005). Assessing professional competence: From methods to programmes. *Medical Education*, 39(3), 309–317.
- van Luijk, S., Gorter, R., et al. (2010). Promoting professional behaviour in undergraduate medical, dental and veterinary curricula in the Netherlands: Evaluation of a joint effort. *Medical Teacher*, 32, 733–739.
- van Luijk, S. J., Smeets, S. G. E., et al. (2000). Assessing professional behaviour and the role of academic advice at the Maastricht Medical School. *Medical Teacher*, 22(2), 168–172.
- van Mook, W. N., de Grave, W. S., et al. (2007). Factors inhibiting assessment of students' professional behaviour in the tutorial group during problem-based learning. *Medical Education*, 41(9), 849–856.
- van Mook, W., de Grave, W., et al. (2009a). Professionalism: Evolution of the concept. *European Journal of Internal Medicine*, 20, e81–e84.
- van Mook, W., Gorter, S., et al. (2009b). Approaches to professional behaviour assessment: Tools in the professionalism toolbox. *European Journal of Internal Medicine*, 20, e153–e157.
- van Mook, W. N., van Luijk, S. J., et al. (2010). Combined formative and summative professional behaviour assessment approach in the bachelor phase of medical school: A Dutch perspective. *Medical Teacher*, 32, e517–531.
- van Mook, W., van Luijk, S., et al. (2009c). The concepts of professionalism and professional behaviour: Conflicts in both definition and learning outcomes. *European Journal of Internal Medicine*, 20, e85–e89.
- Wheeler, D. W., Whittlestone, K. D., et al. (2003). A web-based system for teaching, assessment and examination of the undergraduate peri-operative medicine curriculum. *Anaesthesia*, 58(11), 1079–1086.