Approaches to professional behaviour assessment: Tools in the professionalism toolbox

Walther N.K.A. van Mook, Simone L. Gorter, Helen O'Sullivan, Valerie Wass, Lambert W. Schuwirth, Cees P.M. van der Vleuten

 Departments of Intensive Care and Internal Medicine, Maastricht University Medical Centre, P. Debyeelaan 25, 6202 AZ Maastricht, The Netherlands
 Department of Medical Education Development and Research, Maastricht University, Faculty of Health, Medicine, and Life Sciences, Maastricht, The Netherlands
 Department of Internal Medicine, Division of Rheumatology, Maastricht University Medical Centre, Maastricht, The Netherlands
 Centre for Excellence in Developing Professionalism, School of Medical Education, University of Liverpool, Liverpool, United Kingdom
 Manchester Medical School, University of Manchester, Manchester, United Kingdom

 A R T I C L E   I N F O

 Article history:
 Received 23 December 2008
 Received in revised form 22 July 2009
 Accepted 25 July 2009
 Available online 20 September 2009

 Keywords:
 Problem based learning
 Professional behaviour
 Professionalism
 Tutorial group
 Assessment
 Tools

 A B S T R A C T

 There is general agreement that professionalism and professional behaviour should be (formatively and summatively) assessed, but consensus on how this should be done is still lacking. After discussing some of the remaining issues and questions regarding professionalism assessment, this article discusses the importance of qualitative comments to the assessment of professional behaviour, focuses on the currently most frequently used tools, as well as stresses the need for triangulation (combining) of these tools.

© 2009 European Federation of Internal Medicine. Published by Elsevier B.V. All rights reserved.

1. Introduction

Assessing professionalism as an outcome measure of medical education mirrors the trend to measure outcomes of patient care as indicators of health care quality. Tools have been developed for measuring professionalism and professional behaviour to identify, counsel, and remediate students demonstrating poor professional behaviour [1]. Assessing professionalism during (internal medicine) specialist training programmes serves comparable purposes [2]. The challenge of assessing professionalism can be summarised by two statements: “If it can’t be measured, it can’t be improved,” and “They don’t respect what you expect, whereas they respect what you inspect” [3]. Assessment motivates students to learn what is important and informs teachers whether programme expectations are clear and whether students have learnt from these experiences.

Lack of formal assessment may send a conflicting message to students. Assessment of professional behaviour serves a clear formative purpose (‘no feedback, no learning’ or ‘assessment drives learning’), as well as a summative purpose (‘no learning, no pass,’ or ‘go/no-go decision’). To date the focus has largely been on the summative aspects, and identifying the few ‘bad apples.’ However the formative aspect is obviously essential and should benefit all students.

Many other issues and questions fuel the ongoing debate. Some debate whether professionalism or professional behaviour should be assessed. Should one global measure be used (as a comprehensive and integrated entity), or should professionalism be broken down into specific elements, which are measured independently? [4–7] Additional areas of discussion centre on: (i) how many observers are needed to make a reliable assessment? (ii) Who should perform it; faculty, peers, patients or even others? (iii) Where to perform assessments; only in authentic contexts or are simulated circumstances acceptable? And (iv) how often should assessments be done; frequently after short observations or less often after intensive observation? The perfect measure for professionalism lies most certainly in Utopia.

Any one single measure alone is not sufficient [8]. A combination (triangulation) of currently available instruments is necessary [9]. With any unstructured assessment tool, the quality of the information gathered is dependent on the user. Contemporary professionalism assessment methods, however, tend to be mildly structured. They include self and peer assessment, the objective structured clinical examination (OSCE), direct observation by faculty during regular educational sessions, critical incidents reports and learner maintained portfolios [3,10–12]. Often standardised checklists are used [13,14].
This article consecutively discusses the importance of qualitative comments and the tools commonly used for professionalism assessment.

2. The role of qualitative comments in assessing professionalism

Assessments are overwhelmingly quantitative in nature, with characteristically a “paucity of specific written comments” [15]. As professionalism is difficult to assess as a set of items or a single number, many schools reserve space on their forms for narrative comments [16,17]. Comments made during clerkships are mainly positive in nature. Negative comments stand out by rarity and intensity [16], and should thus be taken seriously. Research has shown that equivocal comments (not directly interpretable as clearly positive or negative), become negative ones when correlated with the numerical score, and should be regarded as such [16]. So, qualitative comments are extremely useful, but are not often provided. Assessors need instruction and training to do this [18].

3. Rating scales

Rating scales are used in a variety of assessments of professional behaviour. The most commonly used tool is the global performance rating, in which assessors retrospectively rate general categories of behaviour, with or without a numerical scale. A study has indicated significant similarity between the different instruments [17]. Elements appearing in almost all rating scales relate to: how well the student deals with other people (peers, staff, patients), the level of student self awareness (dealing with critique, self-criticism), communication skills (giving and asking information, listening to others), and the demonstration of personal qualities (tolerance, empathy, leadership, respect for others, active performance, and efficient task performance) [17]. Rating forms are readily available and easy to use. One drawback is the “paucity of comments” [15]. In addition they are plagued with “halo” or “horn” effects [19] and “leniency error” [20]. The “halo” effect occurs when an assessor allows one strong point or impression of the candidate to influence all consequent decisions. Being aware that someone has scored well on a knowledge test can positively influence the tutor’s view on this person’s professional behaviour. The “horn” effect is the exact opposite — one weak point can overshadow all other judgments. “Leniency error” occurs when ratings are used generously regardless of actual performance, in an attempt to avoid conflict or to make a good impact. Despite these limitations, rating scales are used in many medical schools.

4. Observation by faculty

Faculty observation of students’ professional behaviour offers many of the factors needed for effective assessment [21]. Different assessors can be used, in realistic settings (such as naturally occurring conflicts in authentic context), across a number of different activities and over a period of time. The complex relationship between faculty and students can however influence the results. Three issues may affect the quality of observational assessments: (i) role conflict: the interchanging roles of teacher and assessor can be difficult to combine, (ii) the stakes: the consequences or outcomes for students will differ across assessments, (iii) equivalence: the circumstances and activities during which students are observed may differ [21]. The mini-clinical examination (mini-CEX), is an example of a method which uses a rating form to evaluate professional behaviour [22]. The professionalism mini-evaluation exercise (P-MEX, a mini-CEX variant, see Fig. 1) for example evaluates 21 specific behaviours and is performed by a trained faculty member during a 15 to 20 minute observation period. It provides a feasible method for formative assessment of professionalism in a variety of settings, promoting feedback and encouraging self-reflection [23].

Fig. 1. Example of rating scale filled out by faculty: the professionalism mini-evaluation exercise (P-MEX) (courtesy of Prof. Dr. Y. Steinert, McGill University, Montreal, Canada).

5. Self-assessment

Self-assessment is defined as a personal evaluation of one’s professional attributes and abilities against perceived norms [24–26]. Self-assessment is integral to many appraisal systems, and has been proposed as an important aspect of personal professional behaviour by several regulatory bodies and by those developing learning outcomes for students [24]. Medical students are sometimes perceived to be poor at self-assessing performance. Correlations between self-assessment, and external measures are indeed poor [27,28]. Its accuracy can be enhanced by feedback, explicit assessment criteria and benchmarking guidance [24]. Whereas rigorous studies on self-assessment are on the whole rare [24], a few published studies on self-assessment of professionalism exist [29]. A study studying the accuracy of medical students’ self-assessment of performance on clerkships revealed that students with higher grades tended to underestimate their performance, whilst poorer performing students tended to overestimate their performance [29,30]. However, some students have too high expectations of their own performance, perhaps explained by previous academic success and lack of adequate feedback [29]. In general however, the least competent were least able to self-assess accurately [24]. Furthermore, a diminished capacity for self-improvement of a student (e.g. by not accepting constructive criticism) has been associated with a risk three times as high as for controls of subsequent disciplinary action as a practicing physician [1]. Self-assessment should always be triangulated with other methods of assessment.
6. Peer assessment

Medical students usually know which of their classmates they would trust to treat their family members, illustrating the intrinsic potential of peer assessment. Dannefer et al. reported that the item, “I would refer my own family or patients to this future physician or ask this person to be my own doctor” in their rating scale questionnaire was a good proxy for two facets of professionalism, namely ‘work habits’ and ‘interpersonal skills’ [31]. Peer assessment involves assessors with the same level of expertise and training, and who share the same hierarchical institutional status. The information gathered through peer assessment may differ from that from other sources [28,32]. It has been used to measure the performance of medical students [31,33–35], residents [36], and practicing doctors [37,38], provides useful insight into learners’ cognitive and interpersonal skills as well as professional behaviour [39–41], and stimulates the development of self-motivated, responsible and reflective students [42]. A recent review by Evans et al. was far less positive. None of three identified instruments for peer assessment met the required standards for instrument development. All focussed mainly on reliability and feasibility, omitting work on validity and rater training [43].

Studies specifically addressing peer assessment of professional behaviour are nevertheless beginning to appear. They focus on acceptability and practical feasibility, reliability, validity, effects of rater selection and temporal trends [31,35,44–46]. Shue et al. reported that 66% of students agreed that peer assessment of professionalism was acceptable as long as it reflected their preferences on how the assessment should take place. Students generally agreed on the requirements of peer assessment. They valued reporting to an impartial counsellor, a 100% anonymous process, and ensuring the classmates receive corrective instruction [44]. Students are reluctant to report poor professional behaviour of their peers since it might harm the individual, themselves, or the clinical team. Many students prefer that such assessment is used for formative purposes only.

Anonymity (or at least confidentiality) is important to encourage participation. The educational environment also plays a role in students’ willingness to assess peers [45]. Competitiveness should also be taken into account; a small number of students feel that they give their ‘competitors’ for post graduation jobs an advantage by providing (positive) feedback on their performance (HOS, University of Liverpool, personal communication). Several studies addressed the issue of reliability related to the number of raters. From these the acceptable number is estimated to be around 8 to 10 [31,37]. Typically, assessments of interpersonal skills and professionalism require more ratings than assessment of cognitive skills [47]. Biases in rater selection do not appear to affect the results of peer assessment [35]. No relationship was found between peer ratings, race, gender, geographical origin, or social class of the raters [32,33]. Medical schools can thus select the method of peer-rater assignment based on practicality and logistical considerations [35].

7. Multi source feedback or 360-degree evaluation

This popular approach expands the assessor pool markedly, for example with peers, staff members, nurses, physician assistants, standardised patients (SP), and real patients [28,48–50]. The professionalism evaluation form (called APB, developed as a part of the National Board of Medical Examiner’s Assessment of Professional Behaviours Field Trial), for example includes observable behaviours, relational items, global performance ratings as well as free text options [51]. The impact of 360-degree professionalism assessment on faculty comfort and feedback skills during an internal medicine residency programme was recently published [52]. Although sample size was small, the faculty reported increased skill in giving feedback in general and both increased skill and comfort on professionalism issues specifically [52].

8. Objective structured clinical examinations (OSCEs) and standardised patients (SPs)

The OSCE can validly and reliably assess clinical knowledge and technical skills [53], as well as students’ humanism [54], communication [55] and empathy [56] skills provided that a range of contexts is sampled using different assessors. Performance is generally reflected in an overall score, representing a combination of history taking, physical examination, technical skills, organisation, and interpersonal and communication skills [53]. Interpersonal skills scores are sometimes reported separately because of their importance in assessing overall performance, their capacity to differentiate between residents at the same as well as different skill levels [57], and their strong association with students’ and residents’ competency [58,59]. Several studies have investigated the question of whether interpersonal skills should be assessed by faculty or by standardised patients [53,60,61]. SPs based assessment of doctors’ professionalism, addressing attentiveness, respect and empathy, and interpersonal skills appears promising [11,58,62,63]. SPs tend to give slightly higher scores than faculty [53,64], but otherwise appear to be interchangeable with physicians as assessors of interpersonal skills [53]. Global ratings are at least as valid and reliable as checklist scores [65,66] for OSCEs in general, and specifically on the interpersonal skills domains [53].

9. Real patient perceptions of doctors’ professional behaviour

Patients’ perceptions of physicians’ conduct gained from examining unsolicited complaints can help to identify those demonstrating poor professional behaviour and at higher risk of malpractice [67]. A small number of physicians generates a disproportionate share of complaints [67]. Specific types of behaviour were, however, not predictive of malpractice activity [67]. Nevertheless, physicians without malpractice claims offered patients more constructive and facilitating comments, used more humour, and were perceived as being more willing to answer questions. As well as avoiding complaints, they were arguably demonstrating a higher level of professional behaviour [68]. When individual performance data are correctly disclosed, physicians’ practice patterns and behaviours can change [69,70]. Apart from analysing unsolicited complaints, patient questionnaires have been designed to assess practicing physicians’ professionalism [71,72].

10. Surveys to assess professionalism and its components

Questionnaires can be used to assess prevalence and subjective experience, relating to individuals, observations or perceptions of others, or to the climate of a group or institution. However, assessing only the formal curriculum is insufficient. Surveys of the organisational professionalism and peer/multi-source feedback provide windows into the more informal setting, into the hidden curriculum. Baldwin has written in depth on such surveys and their use in professionalism assessment [73].

11. Critical incident reports

“A critical incident report is a detailed description of a one-off occurrence of specific sentinel event, which identifies a student at the extremes of performance” [74]. Since every professional’s behaviour includes occasional lapses, identification of critical incidents does not necessarily mean that a student is unprofessional by character. Faculty should be able to report even minor incidents to provide the opportunity to initiate remediation strategies and without any
concern that these will necessarily lead to dismissal. Single occurrences should not be ‘blown out of proportion.’ Longitudinal tracking of critical incidents contributes to monitoring the effectiveness of remediation strategies. Multiple incidents over time appear to identify those students with significant deficits in professional development. Such students are twice as likely to be subsequently disciplined by a state medical board [13]. Their identification necessitates formal action, as ‘the paper trail associated with repeated reports reinforces the persistent nature of their deficiencies’ [75]. Some students do not accept responsibility for the problems identified. If unprofessional behaviour persists, despite remediation attempts, these students are most likely to be dismissed in the US [74,75]. Different universities have developed and implemented different strategies and faculty reporting systems [74–76].

12. Portfolio

Portfolios effectively assess day to day performance [77,78]. A comprehensive range of information, collated in this way, can produce a well founded summative judgment [77]. Doctors or students can also simultaneously formatively analyse their own performance. They can reflect on and improve their practice and set realistic objectives for further learning [79,80]. Mentoring is the single most decisive success factor [81]. Secondly, the portfolio must be smart and lean [77]. Careful implementation is crucial. A strong resistance to the portfolio can be unleashed when learners are forced to stick to a rigidly prescribed format [79]. The dynamic and flexible qualities of portfolios are well suited for the assessment of the multifaceted and complex nature of professionalism. The portfolio cultivates skills such as reflection and self-assessment, elements essential to professionalism [82]. Its developmental nature allows longitudinal follow-up, provision of feedback, and monitoring of progress over time [83]. There is little doubt that the use of portfolios in the assessment of professionalism will continue to develop [84].

13. The use of humanism “connoisseurs” and other methods to assess of professionalism

Michis has suggested the use of humanism “connoisseurs,” who have expertise in the interpersonal components of the medical art, to provide formative and summative feedback on competency in medical professionalism and humanism, in an attempt to counteract the marginalisation of professionalism in the informal curriculum [85]. Another innovation, the “virtual practicum,” uses recent technological advances to provide a simulation mirroring daily practice. This focuses on difficult ethical situations and communication dilemmas. It is an interactive CD-ROM programme, for example based on patients with HIV/AIDS [86].

14. Conclusions

Assessment of professional behaviour clearly serves a formative purpose for all students, as well as a summative purpose for a few. Qualitative, constructive comments are essential, but infrequently provided. In contemporary medical education the methods most commonly used to assess professional behaviour are self and peer assessment, the OSCE, direct observation by faculty during regular educational sessions, critical incidents reports and learner maintained portfolios. Standardised checklists are often used. The perfect measure for professionalism is almost certainly unachievable so, for now, a combination of the available tools must suffice, a process called triangulation. During clinical training, the combined use of portfolios, short clinical examination assessments (mini-CEX) specifically designed for or incorporating aspects of professionalism, and/or self- and multi-source feedback on professionalism issues are practically achievable and must suffice. The assessor makes an important contribution to the quality of the information gathered by these tools. Staff must be trained to ensure they provide students with meaningful oral and written feedback on their professional behaviour.

15. Learning points

• All students should be assessed, both formatively and summatively. Formative assessment however is essential for all students, whereas summative assessment serves a purpose for only identifying a few ‘bad apples.’
• Staff should agree on which definition of professionalism to uniformly use. Staff training on providing oral and written feedback is essential to optimise benefit from the assessment.
• Self and peer/multi-source assessment, the OSCE, direct observation by faculty during regular educational sessions, critical incidents reports and learner maintained portfolios are frequently used professionalism assessment tools.
• Combination of the available tools, so-called triangulation, is needed to reliably and validly assess professionalism.
• The assessor is the main determinant of the quality of the information gathered when using the assessment tools.

References


