

The nature of qualitative comments in evaluating professionalism

ALICE FROHNA & DAVID STERN

INTRODUCTION We analysed written comments on clerkship evaluation forms to determine if they furthered our understanding of students' professional behaviour.

METHODS We transcribed all written comments related to professional behaviour for 1 medical school class in 4 major clerkships, then categorised the comments. Comments were coded into categories by two raters. They were also coded as positive, negative or equivocal. The 3 types of comments were each correlated with a Likert-type score for professionalism for each student in these same clerkships.

RESULTS Most comments covered initiative, working well with others, patient skills and working hard. Of the 1845 comments, 1721 were positive, 106 were negative and 18 were equivocal, neither wholly positive nor wholly negative. The total number of positive comments correlated with the students' numerical professionalism score ($r = 0.29$, $P < 0.001$); negative comments correlated negatively with the numerical professionalism score ($r = -0.44$, $P < 0.001$). Equivocal comments were significantly negative ($r = -0.25$, $P = 0.002$). Qualitative analysis revealed that equivocal comments were focused on maturity, initiative and teamwork.

CONCLUSIONS Written comments contain unique indicators of students' professional behaviour and are largely positive in nature; negative comments stand out by their rarity and intensity. There is a set of negative comments masked by evaluators in language that appears neutral or even partly positive that statistically resemble negative comments and should be

regarded as such by students who receive this type of comments. Evaluators should be encouraged and trained to express their qualitative views of students.

KEYWORDS education, medical, undergraduate/ *methods; *professional autonomy; clinical clerkships; curriculum/ methods; students, medical/ psychology.

Medical Education 2005; **39**: 763–768

doi:10.1111/j.1365-2929.2005.02234.x

INTRODUCTION

While the medical community has long recognised the importance of professional behaviour among its members, the formal study of professionalism has been a more recent focus of attention. The American Board of Internal Medicine Foundation and the European Federation for Internal Medicine's 'Charter on Professionalism' outlines the professional values that physicians are expected to possess.¹ The General Medical Council has clearly defined the professional duties of a doctor.² Dr Jordan Cohen, of the Association of American Medical Colleges, argues that medical schools must identify students at the time of admission who will be likely to enact professional behaviours, and then ensure that students graduate having maintained these professional values.³

As students are learning how to behave with peers, superiors and patients, they are being evaluated on their professionalism. Clerkship evaluation forms usually include questions about students' professional characteristics, and often leave space for written comments. Clerkship evaluation forms have been shown to represent a reliable and valid method of evaluating both cognitive and non-cognitive aspects of medical students' clerkship performance.^{4–6}

University of Michigan Medical School, Ann Arbor, MI, USA

Correspondence: Alice Frohna PhD, G1116 Towsley Center, University of Michigan Medical School, 1500 E. Medical Center Drive, Ann Arbor, MI 48109-0201, USA. Tel: 734 647 3306; Fax: 734 936 1641; E-mail: azf@umich.edu

Overview

What is already known on this subject

Professionalism has been an important recent focus of attention, although the domains of professionalism observed by faculty may not match standard definitions.

What this study adds

Qualitative comments on evaluation forms may provide a richer, more complete view of students' professional behaviour.

Suggestions for future research

Clerkship evaluators should be encouraged to provide qualitative comments, and those who are responsible for issuing grades should consider these comments carefully. Further study would investigate whether using these comments helps the student improve their professionalism, and educators to identify students who may have difficulty in this area.

In addition, data from our institution establish convergent validity of faculty ratings of students' professionalism across four clerkships.⁷

Clerkship evaluations are overwhelmingly quantitative in nature, but there is inherent difficulty in reducing the complex construct of professionalism to a number or set of numbers. For that reason, many schools allow space on these forms for narrative comments. The comments are difficult to summarise and compare with the more numerical data, and are therefore challenging for clerkship directors to incorporate into the formal grading process.

However, from our previous research⁸ we noted that most evaluators do include narrative comments. Many of these comprise elaborated, rich examples of students' behaviour on the wards. We analysed these comments qualitatively to determine whether they furthered our understanding of students' professional behaviour. Our study investigates whether these written comments contain themes not otherwise captured in general 'professionalism' Likert-type items, and identifies which themes appear most

commonly. In addition, this work attempts to classify these comments as positive, negative or equivocal to determine if they could be used as a more formal measure of student performance.

METHODS

We examined the evaluation forms for 1 medical school class (approximately 153 students) from 4 of their third-year clerkships: internal medicine, paediatrics, surgery and obstetrics/gynaecology. These were chosen from the 7 required clerkships because they were the longest in duration (lasting from 6 to 12 weeks). This length provided greater opportunities for contact between evaluators and students, leading in turn to greater familiarity and potentially more evaluative comments.

The University of Michigan clerkship evaluations include 3 Likert-style questions pertaining to professional characteristics. These questions deal with interpersonal relationships with physicians and other health professionals (evaluating co-operation and respect), interpersonal skills with patients (evaluating listening and speaking communication skills; empathy; and ability to deal with difficult patients) and professional responsibilities (evaluating willingness to complete prescribed and additional duties; and ability to work promptly, independently and efficiently). Evaluation methods differed slightly across clerkships. In internal medicine and surgery, each house officer and faculty member who had contact with the student, no matter how limited, completed an evaluation. The ob/gyn clerkship director completed a composite evaluation form himself, after conferring with physicians with whom the student had worked. In paediatrics, only the attending physicians completed evaluations. However, these evaluations sometimes reflected informal input from the house officers, as indicated in the written comments. Across all 4 clerkships, each student was evaluated by at least 15 and up to 45 residents and faculty, ensuring that each student's professional behaviour was observed and evaluated by many different people.

We examined each evaluation, read the written comments and transcribed verbatim all those that we felt were related to something other than academic performance. For example, a comment regarding the tendency to arrive late at conferences would be transcribed, while a comment about surgical technique would be ignored. A single comment was considered to be a phrase or clause as recorded by

the evaluator. For instance, the comment 'prompt, efficient and pleasant' would be considered to be one comment linguistically, but would be coded as 3 separate professionalism topics. This procedure yielded 1845 comments pertaining to professionalism for all students in all clerkships, from a total of 2997 comments on all topics.

We then categorised the comments further according to the specific traits of professionalism that were depicted therein. Our initial categories were culled from the qualities and behaviours described in the anchor text on the numerical scales of the evaluation form. These qualities included efficiency, promptness, respect for other professionals, empathy and listening skills. After performing this categorisation, there remained a substantial group of comments pertaining to behaviours other than those listed on the evaluation forms. We created a list of qualities and behaviours embodied in these comments, then condensed the list by collapsing similar behaviours into fewer categories. These categories fitted our gestalt impression of what professionalism entails, as well as those more formal definitions proposed by researchers and licensing boards. These categories are described in Results.

We used NUD*IST to search the clerkship comments and code each one into a category. Before proceeding with coding the entire data set, 2 raters each coded comments for slightly more than 10% of the class (181 comments for 17 students). Inter-rater reliability for coding of these comments into 6 categories was excellent ($\kappa = 0.84$). We discussed any discrepancies, checked our understanding of the categories and updated category definitions as necessary until we reached agreement about categorisation. One rater then proceeded with coding the remaining data. Additionally, comments were coded as positive or negative. Each comment unit was coded

as positive if it was wholly positive, or negative if it was wholly negative. The remaining comments were coded as equivocal.

Two further types of analyses were performed on the equivocal comments. First, qualitative analysis was performed to characterise the nature of the comments, as well as to assess the category of professionalism under which they fell. Secondly, the 3 types of comments were each correlated with an overall professionalism score for each student. This score was derived from the mean score across all raters for each clerkship, for each of the 3 Likert-type items on the clerkship evaluation forms pertaining to professionalism. This analysis was undertaken in order to determine whether equivocal comments resembled more closely positive or negative comments.

RESULTS

Thematic analysis of written comments

The most frequent categories of comments are listed and described in Table 1. The percentage pertains to the number of students who had a comment in that particular category. Some students had comments in most or all these categories; others had far fewer comments in the professionalism categories. Table 1 demonstrates the broad range of professionalism issues that were commented upon by evaluators.

Positive, negative and equivocal comments

Students as a group received far more positive than negative comments overall. Of the 1845 comments pertaining to professionalism, 1721 were positive. All students had some positive comments; there was a range of 2–43 positive comments per student. Thirty-seven students received 10 or fewer positive

Table 1 Comment categories

Category	Description	Percentage of students with a comment in this category
Initiative	Motivation, willingness to conduct patient care without prompting	50%
Relationships	Ability to work with superiors, subordinates, peers	46%
Patient skills	Compassion, listening skills, other humanistic qualities	38%
Work habits	Work ethic, attendance, participation	36%
Self-improvement	Self-assessment, receptiveness to critique	27%
Composure	Polish, confidence, ability to handle difficult situations with ease	21%

comments, 73 students received 11–20 positive comments, 33 students received 21–30 positive comments and 10 students received more than 30 positive comments.

It was far less common for students to receive negative comments; 106 comments were negative. Across students, there was a range of 0–16 negative comments. Ninety-nine students received no negative comments, 22 received only 1 negative comment, 25 students received 2–5 negative comments, 3 students received 6–9 negative comments, 3 students received 10–13 negative comments and 1 student received 16 negative comments.

Equivocal comments were even less common, with most students receiving none of this type. There was a total of 18 equivocal comments, with a range of 0–4. Nine students received 1 equivocal comment each, while one student each received 2, 3 or 4 equivocal comments. Equivocal comments were spread across categories, but tended to focus on themes of maturity, initiative and teamwork. Because this group of comments was neither positive nor negative, we investigated their relationship to quantitative Likert-scale items to understand their meaning more clearly.

Quantitative analysis of comments

Our quantitative analysis of numerical clerkship ratings yielded a numerical professionalism score for each student.⁸ This professionalism score was derived by calculating the mean of the Likert scale scores for all raters for each of the 3 items pertaining to professionalism in a single clerkship, then calculating a mean for each student across the 3 items. This provided a professionalism score for each of the 4 individual clerkships. Finally, a mean of all 4 clerkship professionalism scores created a single numerical representation of the overall impression of a student's professional behaviour.

The number of positive, negative and equivocal comments for each student was calculated across all evaluation forms. The number of positive comments correlated moderately with the numerical professionalism score ($n = 1721$, $r = 0.29$, $P < 0.001$); negative comments correlated negatively with the numerical professionalism score ($n = 106$, $r = -0.44$, $P < 0.001$). Equivocal comments were significantly negative ($n = 18$, $r = -0.25$, $P = 0.002$).

Thematic analysis of equivocal comments

Because the equivocal comments were clearly not neutral in character, we subjectively analysed them to see if we could classify them with greater specificity. Based upon this analysis, we identified 3 general themes in the equivocal comments (Table 2).

Inexperience with clinical medicine

Many equivocal comments spoke to the student's newness to clinical medicine and the flow of work on the inpatient wards. Several such comments fell under the category of composure, describing issues regarding a student's polish or maturity as a clinician. These comments hinted both at student efforts toward fitting in with the team, as well as at problems in these attempts:

Impressed us with his strong work ethic. We would suggest, however, efforts at refining his clinical demeanour.

Seems excited about a career in surgery but at times was overzealous in conveying that.

Another cluster of comments pertained to motivation. Some such comments indicated that students were eager to fit in, but had difficulty because their efforts to fit in were incompatible with their role as students of clinical medicine:

Table 2 Themes present in equivocal comments

Theme	Characteristics	Professionalism themes
Inexperience with clinical medicine	Suggest students' efforts to 'fit in', largely unsuccessful	Composure, initiative, relationships
Change in behaviour during rotation	Usually indicates improvement throughout clerkship	Initiative, self-improvement, patient skills
Faint praise	Some element of praise, deeply buried	Initiative, work habits, relationships

Friendly, pleasant, gets along well with patients, house officers and faculty but did not seem genuinely interested in ongoing clinical activities.

Anxious to be involved in hands-on care of patients, but sometimes found himself to be not quite prepared for how technically difficult it was to perform suturing that appeared very easy in the hands of a resident or attending.

Other comments in this category spoke to students' lack of experience working with a team of other health-care providers. In both the following comments, the implied criticism of the student's efforts at teamwork is clear:

Although he can be a bit aloof at times, his interactions with patients were always pleasant and courteous.

House officers were impressed with his depth and wished that he had been more easily able to interact with them and with patients.

Changes during rotation

In a second major group of equivocal comments, evaluators indicated that the student's performance was satisfactory in one part of the clerkship, but lacking in another portion. In all but 1 case, the improvement occurred from beginning to end of the clerkship.

Several students were described as displaying greater initiative and responsibility as the clerkship progressed.

A little uninvolved in the flow of patient care initially in the rotation, but toward the end was more active in coming to the OR.

Needed some encouragement at the beginning of the rotation, but showed good improvement in skills and confidence.

Another set of comments referenced an improvement in patient interaction skills:

Had a little difficulty at first knowing how to get involved in labour and delivery, but soon developed a nice way of creating rapport with patients.

Faint praise

A third group of equivocal comments delivers 'faint praise' – a definite element of criticism that the

evaluator attempts to couch as praise. The following comments in this category pertain to initiative and hard work:

Recognised that he had days with more and others with less energy, effort and desire.

Performed at a very adequate level when specifically stimulated but otherwise did not make an effort to get involved.

Another group of 'faint praise' comments, largely in the areas of interpersonal interactions and participation, seems on the whole to have a more positive tone than the other types of equivocal comments described above. However, an element of negativity is apparent in these comments as well.

During structured teaching sessions rarely participated, not clearly due to a lack of interest.

Taciturnity may sometimes be misinterpreted as disinterest when this is not the intent.

Included in this general group of comments are several in which the evaluator predicts with confidence that positive changes will occur, but does not note any actual improvements. This is one example:

Will improve steadily as she gets more used to patient interaction and hospital culture.

CONCLUSIONS

Our results suggest that written comments contain unique indicators of students' professional behaviour felt to be important by faculty and residents. While some categories were identified in both the numerical and written portions of the evaluation forms (patient skills, work habits), many other categories were identified only in the narrative portion of the evaluations (initiative, maturity/composure, self-improvement). These categories could be used to cue the evaluators' comments on the students' professional behaviour, and capture in numerical form behaviours that faculty identify as important.

In light of the rich information that we found in the written comments, these data support the routine use of written comments as a standard part of evaluation forms and inclusion as a consideration for assigning clerkship grades. Evaluators could be asked to provide illustrative examples to flesh out their

numerical evaluations. Our work demonstrates that narrative comments can be used to form a more complete view of a student's professional character. Much of this information is captured only when the evaluators happen to comment on these qualities. Thus, only students who are particularly weak or particularly strong in this area may be noted.

Written comments designed to give feedback about professional behaviour to third-year medical students are overwhelmingly positive in tone. Negative comments stand out by their rarity and intensity. As most evaluators do not make negative comments, those who do should be taken seriously. We located an additional set of negative comments masked by evaluators in language that appears neutral or even partly positive. While this category of comments is linguistically unique, as indicated by our analyses, statistically they resemble negative comments and should be regarded as such by students who receive this type of comment. Better yet, evaluators should be encouraged to express their less positive views of students as a growth opportunity for the student.

Although written comments proved to be interesting and added information about professional behaviour beyond that provided by the numerical ratings, it may be difficult to incorporate them routinely into standard evaluation practices. It will be much more time-consuming for evaluators to collect and interpret written comments, and to provide feedback on these comments to the students being evaluated. Furthermore, although we were intrigued by the equivocal comments, as a group they appear infrequently and generalisations from this small sample should be made with caution. The setting of these comments was also unique, in that only 3 general numerical questions about professional behaviour are listed on evaluation forms. Other schools with more domains for professionalism items might find a different or possibly smaller set of written concerns.

In conclusion, this study found that faculty will comment on professional behaviours that are provided on the quantitative scales. Increasing the number of dimensions of professionalism on our scale to include work habits, rapport with patients, self-improvement, composure and initiative would appear warranted from the qualitative comments of faculty across multiple clerkships at this institution. In addition to increasing the quantitative dimensions to be sampled, attention to the subtleties of written comments in the narrative portions of evaluation forms can provide additional and important evidence of professional behaviour in medical students.

Contributors: AF collected the data upon which the paper is based. She carried out the quantitative analysis of the clerkship comments and wrote the manuscript. DS was the principal investigator on the grant that supported this work. He was instrumental in the conception of the larger study on predicting professional behaviour of third-year medical students. He assisted in the preliminary qualitative analysis about which this paper reports.

Acknowledgements: we would like to thank the Greenwall Foundation for their support of this research project.

Funding: this work was supposed by a grant from the Greenwall Foundation: 'Predictors of Behavioral Competence: Developing a Diagnostic Test'; D T Stern, Principal Investigator, A Z Frohna, Co-Investigator, 1999–2001, \$75 587.

Ethical approval: when this project was started in 1999, we sought advice from our institution's Academic Review Board as to whether the project required approval. The review board responded that because our analysis involved anonymous student data, and that no students would be identifiable in any final products, that this project was exempt. Thus, our status is IRB exempt from approval.

Conflicts of interest: none.

REFERENCES

- 1 ABIM Foundation, ACP-ASIM Foundation, European Federation of Internal Medicine. Medical professionalism in the new millennium: a physician charter. *Ann Int Med* 2002;**136**:243–6.
- 2 General Medical Council (GMC), Good Medical Practice, 3rd edn. 2001. Available at: http://www.gmc-uk.org/med_ed/. Accessed 21 December 2003.
- 3 Cohen JJ. Our compact with tomorrow's doctors. *Acad Med* 2002;**77**:475–9.
- 4 Dawson-Saunders B, Paiva REA. The validity of clerkship performance evaluations. *Med Educ* 1986;**20**:240–5.
- 5 Mavis B, Doig K. The value of noncognitive factors in predicting students' first-year academic probation. *Acad Med* 1998;**72**:201–3.
- 6 Pangaro L. A new vocabulary and other innovations for improving descriptive in-training evaluations. *Acad Med* 1999;**74**:1203–7.
- 7 Frohna A, Gruppen LD, Stern DT. Faculty provide valid assessments of professional behaviour. *Presented at the Association of American Medical Colleges Research in Medical Education Conference*, 2000, Chicago, October.
- 8 Stern DT, Frohna AZ, Gruppen LD. The prediction of professional behaviour. *Med Educ* 2005;**39**:75–82.

Received 19 January 2004; editorial comments to authors 4 June 2004, 7 September 2004, 15 October 2004; accepted for publication 19 November 2004