



ORIGINAL RESEARCH PAPER

The One Minute Mentor: A Pilot Study Assessing Medical Students' and Residents' Professional Behaviours through Recordings of Clinical Preceptors' Immediate Feedback

D Topps¹, RJ Evans², JE Thistlethwaite³, R Nan Tie², RH Ellaway¹

¹Northern Ontario School of Medicine, Ontario, Canada

²James Cook University School of Medicine and Dentistry, Townsville, Queensland,
Australia

³Warwick Medical School, University of Warwick, Coventry, United Kingdom

Published: 18 May 2009

Topps D, Evans RJ, Thistlethwaite JE, Nan tie R, Ellaway RH

The One Minute Mentor: A Pilot Study Assessing Medical Students' and Residents' Professional Behaviours through Recordings of Clinical Preceptors' Immediate Feedback
Education for Health, Volume 22, issue 1, 2009

Available from: <http://www.educationforhealth.net/>

A B S T R A C T

Introduction: The assessment of professional development and behaviour is an important issue in the training of medical students and physicians. Several methods have been developed for doing so. What is still needed is a method that combines assessment of actual behaviour in the workplace with timely feedback to learners.

Goal: We describe the development, piloting and evaluation of a method for assessing professional behaviour using digital audio recordings of clinical supervisors' brief feedback. We evaluate the inter-rater reliability, acceptability and feasibility of this approach.

Methods: Six medical students in Year 5 and three GP registrars (residents) took part in this pilot project. Each had a personal digital assistant (PDA) and approached their clinical supervisors to give approximately one minute of verbal feedback on professionalism-related behaviours they had observed in the registrar's clinical encounters. The comments, both in transcribed text format and audio, were scored by five evaluators for competence (the learner's performance) and confidence (how confident the



evaluator was that the comment clearly described an observed behaviour or attribute that was relevant). Students and evaluators were surveyed for feedback on the process.

Results: Study evaluators rated 29 comments from supervisors in text and audio format. There was good inter-rater reliability (Cronbach α around 0.8) on competence scores. There was good agreement (paired t-test) between scores across supervisors for assessments of comments in both written and audio formats. Students found the method helpful in providing feedback on professionalism. Evaluators liked having a relatively objective approach for judging behaviours and attributes but found scoring audio comments to be time-consuming.

Discussion: This method of assessing learners' professional behaviour shows potential for providing both formative and summative assessment in a way that is feasible and acceptable to students and evaluators. Initial data shows good reliability but to be valid, training of clinical supervisors is necessary to help them provide useful comments based on defined behaviours and attributes of students. In addition, the validity of the scoring method remains to be confirmed.

Keywords: Professionalism, educational assessment, digital audio, handheld computers

Introduction

The teaching and assessment of professionalism are important issues in medical education. Several authorities have suggested that professionalism should be taught as a specific subject area (Cruss & Cruss, 1997). Personal and professional development courses within undergraduate medical courses have become increasingly popular (Ginsburg *et al.*, 2003). In 1972 Johnson defined the six elements of a profession as: the presence of a skill based on knowledge; provision of training and education; a means of testing for competence; organisation of members, adherence to a code of conduct; and the members' focus on altruistic service beyond any financial reward (Johnson, 1972). Professional development within medicine highlights the code of conduct, including ethical behaviour, and altruism. Also included are accountability, respect and integrity, plus skills such as communication and lifelong learning (Van De Camp *et al.*, 2004).

There is good correlation between students' lack of professional behaviour and their likelihood of being reported to a medical board for unprofessional conduct later in their careers (Papadakis *et al.*, 2004). We, as medical educators, therefore need to develop robust assessment methods that faculty may act on to ensure that students challenged in their professional behaviours are identified and corrective steps are taken. More generally, educators cannot know that they are successful in teaching professionalism if they cannot measure its attributes and demonstrate that students have mastered its principles (Misch, 2002). Possible assessment methods include written examinations, OSCEs and portfolio-based assessment. Current evaluations of professionalism are often focused on students' self-assessments of how they would behave or act in a given situation (Klein *et al.*, 2003). None of these approaches are behaviour-based. Nor do they assess a student's actual performance in clinical settings, although there is a move towards work-based assessment and a call for research into tools that may reliably do this (Wass, 2005). A systematic review of instruments to measure professionalism concluded that there are few well-documented studies of such instruments for either formative or summative assessment (Veloski *et al.*, 2005).

Examples of behaviour and actions, along with their context, are important to obtain a fair overall picture of what occurs in the work place (Misch, 2002). Given their often limited time with each learner, many teachers are not comfortable judging learners in a clinical situation (Misch, 2002). Teachers may generally be insufficiently experienced in their dealings with learners to be able to judge how behaviour compares with an expected standard or with their peers. Rather than relying on single assessment statements



at the end of a rotation, our premise was that multiple recorded examples of observed behaviours, independently and blindly assessed by a panel of experienced evaluators, would provide a more objective and standardised approach to evaluating what are essentially subjective observations. Few clinical teachers have the time to sit down and create extensive notes on what they observe in their learners. A process that is less onerous for the teachers is needed.

Based on our experiences with personal digital assistants (PDAs) in residency training (Topps *et al.*, 2004), we developed this pilot project to evaluate the feasibility of using clinical supervisors' brief recorded comments on learners' professional attributes as a way to evaluate their professionalism. Based on discussions with faculty members experienced in student evaluation in clinical contexts, we chose to focus on four key features or attributes of professionalism: professional behaviour, attitudes, communication and ethics (PACE) (Van De Camp *et al.*, 2004).

Methods

Setting

James Cook University (JCU) medical school in Australia has offered a six year undergraduate programme (direct entry from secondary school) since 2000 and also contributes to postgraduate teaching in a general practice (GP) training programme. Ethics approval for this study was received from the university's Human Research Ethics Committee.

Subjects

Fifth year medical students were asked to volunteer for the project. In the fifth year, students spend eight week blocks in hospitals, rotating through specialities. For this study, learners were provided with free use of a PDA pre-loaded with relevant medical applications. Due to infrastructure and hardware limitations, we could recruit only the first six students who volunteered. For a pilot project, this sample number was considered adequate to provide insight into the benefits and barriers of this innovative approach, although insufficient for a definitive evaluation. A number of GP residents were also approached to participate and three were recruited. A team of five faculty with experience in assessing student professionalism were recruited as evaluators for this study. They were instructed on how to score the comments made by students' clinical supervisors. Supervisors were sampled on a convenience basis by students. No direct orientation to the study was provided to supervisors because of the large number of potential supervisors in the pool, and the study's limited resources made this unfeasible.

Data collection

Students and residents were each given a short 30-minute orientation session by one of the investigators, where they were shown how to make recordings and oriented to the general features of the PDA. We asked learners to obtain up to four comments in each of the eight weeks of the study period from clinical supervisors who oversaw their work with patients. We advised learners that the supervisors they approached should be in a position to comment on PACE attribute-related behaviours observed during their clinical placement. Learners were given prompt cards to present to supervisors which provided examples of the PACE attributes on which their comments were sought (Table 1). Using the voice recording feature of the PDAs, supervisors were asked by learners to dictate comments of not more than 60 seconds in length (thus 'One Minute Mentor') about PACE attributes demonstrated by learners in their care, and then to return the PDA to the learner. Although any digital audio recording device would have sufficed,



the use of the PDA enabled us to provide additional on-screen prompts and information. Recordings identified neither the supervisor nor the student to preserve anonymity. Students returned the recordings on a regular basis to the data coordinator via the PDA’s memory card or by email.

Table 1: Examples of PACE attributes presented to clinical supervisors on prompt cards

Professional behaviour	Attitudes	Communication	Ethics
Punctuality	Motivation	Introduction	Morality
Attendance	Commitment	Information sharing	Ethical awareness
Dress	Gender/sexuality/racial Tolerance	Explanation	Honesty
Courtesy	Empathy	Use of jargon	Integrity
Respects patients Dignity/autonomy	Trustworthiness/	Patient-centred	Appropriate sexual conduct
Confidentiality	Reliability	Summarising	
Teamwork	Altruism	With children	
Cultural sensitivity	Insight	With relatives/carers	
	Cynicism	With staff	
		With peers	

Data analysis

Submitted audio comments were distributed to evaluators through the medical school’s internal computer network both midway and at the conclusion of the data collection phase. Recorded comments transcribed in text format were also distributed, on average two weeks following the audio comments. We compared the text and audio formats in order to assess which was more acceptable to evaluators and determine whether scoring differed between the two formats. The insertion of time between evaluating the audio and the written comments was intended to reduce cross contamination of evaluations.

Evaluators were instructed to score each comment, both audio and written, on two scales: competence (a reflection of the student’s performance in one or more PACE attributes) and confidence (indication of how confident the evaluator was in his/her judgement of the supervisor’s assessment of the student’s competence). Both audio and written comments were scored to see which of these formats was more acceptable to evaluators and if the two methods demonstrated similar reliability across evaluators. All scores were made on a 1-100 scale, where 1 corresponded with the poorest score and 100 with the best.

Quantitative Data Analysis

Analysis of data included production of box plots to provide a visual basis for discerning congruence amongst evaluators and between text and audio comment formats on both scales. Paired t-tests were then performed to look for significant differences between the text and audio comment formats. Cronbach’s alpha was used to compare how all evaluators scored each comment or recording on both the confidence and competence scales. All quantitative data analysis was performed on SPSS 13.0.1 for Windows (www.spss.com).

Feedback

At the conclusion of the study’s data collection phase students were asked to record their own 60-second responses to the following questions:



- What did you like about One Minute Mentor?
- What did you dislike about One Minute Mentor?
- What needs to change?
- Do you have any other questions or comments about One Minute Mentor?

Responses to these questions were transcribed and qualitatively analysed using Atlas.ti 5.0 (www.atlasti.com), a qualitative software package. As the clinical supervisors who provided comments were not identified in recordings or by the students, we could not obtain systematic feedback from this group; however, we took note of incidental recorded comments about the process made by clinical supervisors. Evaluators provided feedback on the acceptability and feasibility of the process after they had completed all scoring.

Results

The six students and three residents together provided recordings of 36 comments from clinical supervisors. By consensus, the team of evaluators eliminated seven of these comments that did not relate to any of the PACE attributes and were deemed to be irrelevant to the study’s purpose. Examples of relevant and irrelevant comments are given in Table 2.

Table 2: Examples of relevant and irrelevant comments for scoring

<p>Relevant comments</p> <p><i>Reinforcement of positive behaviour</i> “Right, well, you’re continuing to do very well in terms of communication skills and in particular sort of relating to children and sort of engaging children and all those types of things you’re doing very well...But I think your interaction with the children and putting mums at ease is very good and in terms of pediatrics, that’s half the battle – make friends with child and the mum. So that’s great!”</p> <p><i>Patient empathy</i> “[Student] has been an invaluable member of our team for over 12 weeks. She is always appropriately dressed and is keen to learn. It is great to work with such a motivated person who always shows empathy and tolerance to her patients. All of our elderly patients have appreciated the time, advice and kindness that she has given to them.”</p>
<p>Irrelevant comments</p> <p><i>Comments of a joking nature</i> Surgical registrar and I've just seen one of the med students examine a patient with an abdominal mass and trust me the patient is in tears after the examination, poor patient! But anyway it was a good demonstration of palpation.</p> <p><i>Irrelevant comments (i.e. not relating to PACE attributes)</i> I love your top today. I love the colour green and it looks very nice and cool.</p> <p><i>Comments relating to knowledge base rather than PACE attributes</i> Student is able to see patients by himself. Treats patients independently. Sutures independently. Has seen renal colic, tension headache, BCCs, children with lacerations, complex COAD with ischemic heart disease...</p>

Figure 1 is a box plot of the competence scores for each of the comments, showing the range, quartiles and means of scores given by the evaluators on the y-axis. The x-axis shows the internal identifier for each comment in the database. The figure visually depicts good agreement between evaluators for most comments, for example, comments numbered 2,5,11,16,25, but also the less agreement on comments numbered 12,13 and 24.

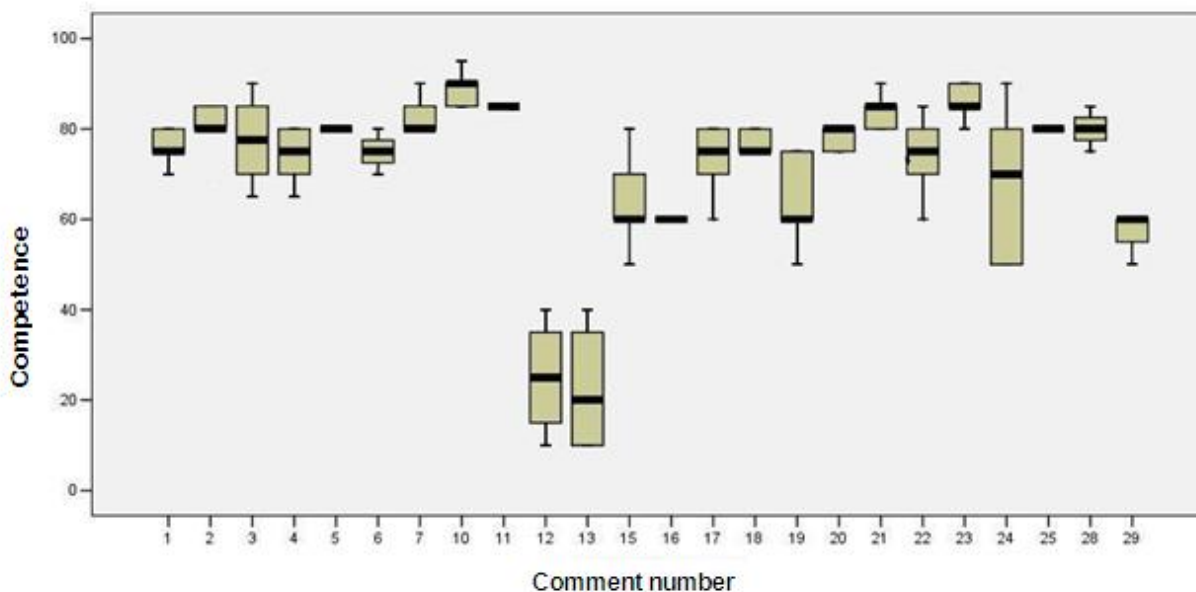


Figure 1: Box plots of competence scores to show congruence among ratings provided by the five evaluators for each of the 29 comments

We compared evaluators' competence scores made for comments presented to them in text versus audio formats. A scatter plot (Figure 2) visually suggests reasonably good correlation. Every evaluator scored every comment from both formats—once on the audio version and once on the transcribed version. A paired t-test for these scores showed no significant difference between these two methods for both the competence ($t = -1.25$, $p = 0.21$) and confidence scores ($t = -1.51$, $p = 0.135$). Looking at the evaluators' scores for each comment, the Cronbach's alpha coefficient showed good mean inter-rater reliability of 0.81 (text competence), 0.79 (audio competence) and 0.80 (audio confidence) but only 0.55 for text confidence.

Figure 3 is a box plot of the confidence scores for each comment, sorted according to rising competence scores. This graph suggests that there is less variance in the competence scores (AudComp) when the Evaluators expressed more confidence (AudConf) in the applicability of the comment. There were only 11 comments where all five evaluators felt able to apply a confidence score.

There were not enough comments to analyse and compare scores when they were broken down into the various categories of attributes. We looked in greater detail at what attribute the evaluators felt they were scoring. There was little or no agreement on this. For example, if we simply used a derived dichotomous variable to record whether or not the evaluator included just one of the four attributes, eg communication skills, the kappa was very poor when comparing pairs of evaluators, ranging from 0.216 to .572. It was also noted that, although evaluators were given a range of 1-100 for their scores, there was a decrease in granularity in that multiples of five were almost exclusively used.

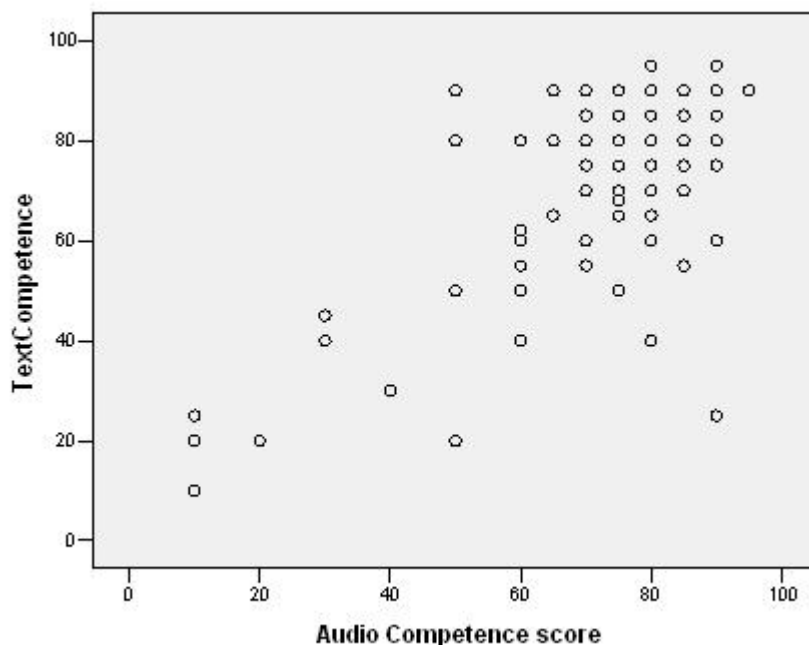


Figure 2: Scatter plot of competence scores made by assessments of audio versus text formats

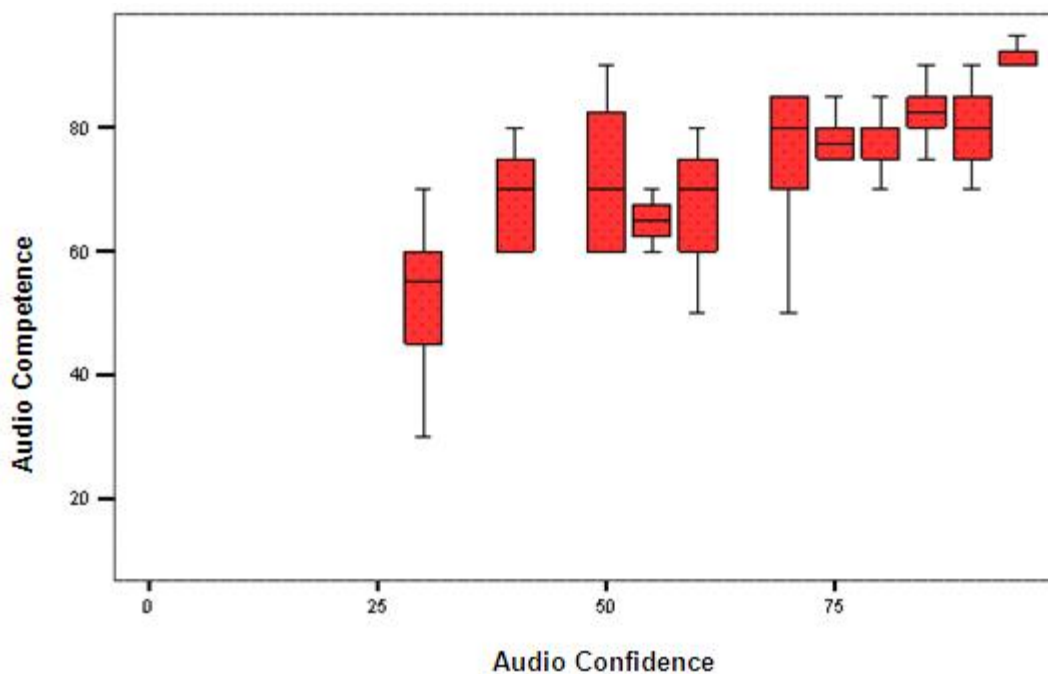


Figure 3: Confidence scores grouped by audio comment, sorted by rising competence scores



Comments from students indicated that they generally felt that this method of evaluation would be useful and an improvement over previous methods of evaluating attributes of professionalism (Table 3). Students found the approach acceptable and easy to use and encountered no difficulties in creating the recordings. However, several students commented that they found it difficult to convey to their preceptors and staff in a concise manner the type of comments that were wanted. They also felt awkward asking some of their supervisors for their time to complete this assessment.

Table 3: Examples of student feedback

<p>Supervisor difficulties</p> <p>“The thing that I didn’t enjoy about the One Minute Mentor was that I felt that I had to make time to approach my supervisors. Sometimes it was awkward and for those supervisors I hadn’t approached previously, which in this rotation was quite a few, it took extra time getting their attention and getting the aim of this project and then get the required information from them onto the One Minute Mentor.”</p> <p>“I think that with time, with more awareness around the hospital or with the doctors and supervisors and so forth, I think they’ll become accustomed to this form of feedback and they’ll have more understanding of what’s required of them and how to illustrate what they think of the students in the most appropriate and clear manner.”</p>
<p>Good contemporaneous feedback</p> <p>“The One Minute Mentor programme has been an enjoyable programme to be involved in, purely from the perspective of being able to have verbal comments about our progress in the hospital.”</p> <p>“Easy way to record and store feedback. Interesting to hear what doctors think about us.”</p>

Evaluators reported that they liked this approach and felt that it was more objective and standardised than previous, traditional approaches to evaluating professionalism since it was based on observations rather than supervisors’ opinions. By gathering multiple data points instead of one summative opinion it seemed less prone to bias. The time required for transcription of the comments was minimal since they were brief. The evaluators’ feedback revealed that all felt the audio evaluation of comments was more time-consuming than reading the transcribed text. However, several evaluators stated that, for some comments, the intonation, emphasis and timing included in the full audio version conveyed more information than the transcribed text version. Most felt that, in a full-scale production scenario, they could simply score on the basis of the transcribed text for most of the comments but that it would be useful to have the full audio version available for the occasional ambiguous comment. As indicated by the evaluators, it was clear that, despite the prompting cards and information sheets provided to our supervisor commentators, more effort was needed in coaching them to provide descriptions and comments on specific observed behaviours instead of continuing to provide the more traditional and familiar overall evaluation of a particular professional trait.

Discussion

Although this project was a pilot study focusing on the feasibility and acceptability of this new approach to evaluating students’ professional behaviours, we were also able to analyse supervisor’s scores and comments to derive meaningful statements about assessment reliability. Analysis of the competence scores showed reasonable levels of agreement between evaluators who were using either the written or audio comments. This suggests that in most instances, either format can be used, however, there are some comments for which evaluators’ scores differed between formats. For evaluators’ confidence scores, we found considerable variation both across evaluators and for scores based on text versus audio information. This variation may be, in part, due to the quality of the comments received, which might improve as staff become more familiar with the process and are provided with further training. Variability in scoring remains an issue. This is one of the reasons why we chose to score multiple comments or



observations for each learner rather than having the evaluator panel scoring a single recording in an end-of-rotation assessment. Our premise was that relying on assessments of multiple student-patient interactions would help reduce the random effects of a “bad day”, a single off-base comment or a misjudged score. A larger study would help to confirm this premise.

That decreasing variance in scoring seemed to be positively associated with increasing confidence scores suggests that evaluators might generally feel more confident when giving higher competence scores. This is not surprising. However, given the few data points assessed in this study, this finding needs confirmation. In order to clarify and standardise the attributes that they score, more guidance is needed for both evaluators and commentators. In our orientation sessions for both evaluators and clinical supervisors, we did not sufficiently clarify that we needed comments that applied to the PACE behaviours or attributes. Not surprisingly, therefore, there was poor agreement on their scoring.

Students and evaluators generally indicated that the One Minute Mentor approach was acceptable and easy to apply. The evaluators found the scoring of the comments to be relatively straightforward and no more or less time-consuming than conventional assessments. The time commitment required from busy clinical preceptors to record a comment was minimal. Through this method students received contemporaneous feedback on this important aspect of their education.

Although incentives were provided, the recruitment and ongoing encouragement of participants was at times difficult. This was felt to be due to the emphasis given to the staff and students that these comments would not be used as part of their overall summative evaluation. Accordingly, they took this process less seriously.

Students reported that they were generally happy with this method of assessment, particularly in that it gave them contemporaneous feedback. Nevertheless, some had concerns about the amount of time required to inform their supervisors about this new form of assessment and explain what the comment should address. Some felt uncomfortable about requesting time for this from their clinical supervisors. Discomfort in requesting time of busy clinical supervisors is not new (Kilminster & Jolly, 2000) and has also been seen for students who request comments in our current, paper-based method of professionalism assessment.

Informal observations suggested that the implementation of the project increased the awareness of professionalism among staff and students and promoted valuable discussion. This new approach generated immediate feedback on behaviour. Such timely feedback can have a powerful impact on student performance and it has some advantages over multisource or 360-degree assessment plus feedback (Department of Trade & Industry, 2005), which usually takes place at the end of a rotation. Moreover, at present there is limited evidence confirming the reliability and validity of the 360 degree method (Baker, 2005).

Supervisors seem to require additional training to better focus their comments on specific observed behaviours and focus their feedback on PACE attributes rather than on the more commonly assessed issues of students’ knowledge bases or clinical skills. This limited discrimination is often a feature of raters’ judgments (Albanese, 2000). Giving feedback is a skill that is sometimes taken for granted, but formal training on professionalism and on providing feedback to students should be provided to staff in any school-wide implementation of this programme for summative assessment. Before applying this programme to a large number of students, the collection and transcription of comments would also need to be streamlined and automated through an electronic database.

This pilot study has its limitations. This small study was principally intended to examine whether this evaluation approach is technically and procedurally feasible. Involving greater numbers of participants and collecting more observations over a longer



period of time are needed to establish the validity of this assessment tool. Further, our study lacked feedback from supervisors, whose views are key in fully understanding acceptability and feasibility. We did not query supervisors because of our blinding process. There would need to be some method of commentator authentication or sign-off as part of the process.

While our project made use of PDAs to collect the digital audio recordings, this technique does not necessarily require the use of PDAs. Any device or recording mechanism capable of capturing digital audio, including centrally stored phone-accessible dictation systems, could also be used. Many programmes elsewhere are now using PDAs in various ways as a student resource or evaluation tool. The valuation method we tested would be one more way in which these devices could be incorporated into the daily workflow of the teachers and learners. Additionally, the PDA can store written and audio examples, instruction pointers and the attribute template, all of which may help facilitate user compliance.

It is possible that in this method students could selectively delete recordings that they felt were 'negative' in tone or might lower their marks. This could easily be controlled by modifying the software of the PDA or of any centrally-stored recording mechanism used.

Conclusions

The One Minute Mentor project demonstrates the feasibility of using brief, structured verbal comments by clinical supervisors to assess professionalism in students and residents. This method may be used for formative assessment but would require further evaluation before it can be used in summative assessment. Most of the difficulties encountered in this pilot study were related to the need to develop a 'culture of use' and familiarity with this new method of assessing professionalism. Supervisors require more training. With time and continued use, we anticipate that supervisors, students and even other health staff would become familiar with this method. More rigorous and larger studies and comparisons against known metrics to assess validity are required to more fully establish the One Minute Mentor model as a useful adjunct for assessing professionalism.

References

- Albanese, M. A. (2000). Challenges in using rater judgements in medical education. *Journal Evaluation Clinical Practice*, 6(3), 305-319.
- Baker, R. (2005). Can poorly performing doctors blame their assessment tools? *British Medical Journal*, 330(7502), 1254.
- Cruess, S. R., & Cruess, R. L. (1997). Professionalism must be taught. *British Medical Journal*, 315(7123), 1674-1677.
- Ginsburg, S., Regehr, G., & Lingard, L. (2003). To be and not to be: the paradox of the emerging professional stance. *Medical Education*, 37(4), 350-357.
- Department of Trade and Industry. (2005). 360 degree feedback. *Best practice guidelines.*, from www.dti.gov.uk/bestpractice.
- Johnson, T. (1972). *Professions and Power*. London: Macmillan Press Limited.



- Kilminster, S. M., & Jolly, B. C. (2000). Effective supervision in clinical practice settings: a literature review. *Medical Education*, 34(10), 827-840.
- Klein, E. J., Jackson, J. C., Kratz, L., Marcuse, E. K., McPhillips, H. A., Shugerman, R. P., *et al.* (2003). Teaching professionalism to residents. *Academic Medicine*, 78(1), 26-34.
- Misch, D. A. (2002). Evaluating physicians' professionalism and humanism: the case for humanism "connoisseurs". *Academic Medicine*, 77(6), 489-495.
- Papadakis, M. A., Hodgson, C. S., Teherani, A., & Kohatsu, N. D. (2004). Unprofessional behavior in medical school is associated with subsequent disciplinary action by a state medical board. *Academic Medicine*, 79(3), 244-249.
- Topps, D., Evans, R., Khan, L., & Hays, R. B. (2004). *Personal Digital Assistants in Vocational Training* (Report, Adobe PDF). Townsville, Queensland: James Cook University.
- Van De Camp, K., Vernooij-Dassen, M. J., Grol, R. P., & Bottema, B. J. (2004). How to conceptualize professionalism: a qualitative study. *Medical Teacher*, 26(8), 696-702.
- Veloski, J. J., Fields, S. K., Boex, J. R., & Blank, L. L. (2005). Measuring professionalism: a review of studies with instruments reported in the literature between 1982 and 2002. *Academic Medicine*, 80(4), 366-370.
- Wass, V. (2005). The changing face of assessment: swings and roundabouts. *British Journal General Practice*, 55(515), 420-422.
-