

The assessment of undergraduate psychiatry training: a paradigm shift

MYH Moosa, FY Jeenah

Department of Neurosciences, Division of Psychiatry, University of Witwatersrand, Johannesburg, South Africa

Abstract

Objective: The aim of this paper was to describe and compare the results of students in psychiatry at various training sites during the course of an academic year at the University of Witwatersrand following recent reforms in training and assessment methods. **Method:** All fifth year medical students (in a six year course) in 2005 were included in the study. Students were assessed at the end of each six-week block. At the end of the year the performances of all groups of students, in each of the various assessment tools, were captured. Each set of assessments over the seven blocks was checked to determine if all the objectives set out in the curriculum were assessed. Comparisons were made between the mean marks obtained in each of the assessment tools by students at the four clinical sites and over the seven blocks. **Results:** 191 students were assessed in 2005. The overall results of the various assessment tools were as follows: mean ward performance mark 14.6 (SD 1.07); mean case presentation mark 14.2 (SD 1.15); mean MCQ's mark 12.2 (SD 2.16); mean MEQ's mark 14.1 (SD 2.53); and mean OSCE mark 15.9 (SD 1.99). Three students failed (<60 marks) and had to repeat the block. Analysis of all the assessment tools in each of the seven blocks confirmed that all the objectives set out at the beginning of the course were assessed. One-way analysis of variance showed significant differences in the mean ward marks ($p < 0.05$); MCQ marks ($p < 0.05$); MEQ marks ($p < 0.05$) and OSCE marks ($p < 0.05$) between the groups in the seven blocks but no differences in the mean case presentation marks ($p = 0.52$). There were no significant differences in the mean ward marks ($p > 0.05$); case presentation mark ($p > 0.05$); MCQ marks ($p > 0.05$); MEQ marks ($p > 0.05$) and OSCE marks ($p > 0.05$) between the groups at the four clinical sites. **Conclusion:** The revised process by Division of Psychiatry at the University of Witwatersrand seems to have brought about improvements in the training and assessment of students by virtue of consistency at the various sites and in the different blocks in the given academic year.

Keywords: Medical, Education, Psychiatry, Assessment

Received: 24-10-2005

Accepted: 05-05-2006

Introduction

Assessment of Psychiatry in medical education at the University of Witwatersrand has, for many years, been largely educational and formative i.e. students learn from such assessment by receiving feedback on which to build their knowledge and skills. In recent years there has been increasing focus on the performances of the emerging doctors and public demand for assurance that these emerging doctors are competent. There is therefore a need to

conduct assessments accountably and according to the latest educational advances. These advances include that careful attention is paid to ensure that the validity and reliability of assessments meet with current educational expectations.¹ The assessment tools must be appropriate in terms of validity and reliability. Validity means ensuring that the assessment tools measure what it is supposed to measure and must contain a representative sample of what the student is expected to have achieved. Reliability expresses the consistency and precision of the tools measurements. There are a variety of factors which contribute to reliability eg. in a clinical examination there are three variables - the students, the examiners and the patients. For a reliable assessment, variability due to the patient and the examiner should be removed and wherever possible, a subjective approach to marking should be replaced by a more objective one.

Correspondence:

Dr MYH Moosa
PO Box 4581, Johannesburg 2000
Email: moosamy@medicine.wits.ac.za

Previous approach to assessment

The previous approach to assessment of Psychiatry at the University of Witwatersrand utilized non-standardised real patients (a single long case assessment) with an oral, by one or two examiners, to assess mainly clinical competence; multiple-choice questions (MCQ) and long essays to assess mainly knowledge. The validity and reliability of this form of assessment has recently been challenged. The long cases are often unobserved by the examiner and the examiner relies on the candidate's presentation of the findings, representing an assessment of "knows how" rather than "shows how", thus raising concerns about the validity of such an assessment format for the assessment of competence. In addition, the use of a single long case challenges reliability, because students do not always perform consistently.² Further, some clinical skills are content specific (i.e. cannot always be generalized) and cannot be assessed using the long-case format. Attempts have been made to improve the reliability and validity by moving to an 'Objective Structured Long Examination Record (OSLER)', which includes direct observation of the candidate interacting with the patient, with little success.³

Long essay formats were for many years the standard written assessment in medical schools throughout the world. They have lost favour in recent times and have been replaced by newer formats such as multiple choice questions (MCQs), and modified essay questions (MEQs). From a teacher's point of view essays are easy to set but are notoriously time consuming to mark. It is also difficult to achieve reliable marking between different sets of examiners marking the same papers. The standards of assessment expected of medical schools are now high, so unreliable and time-consuming assessments are being phased out. However, it may still be a useful method for assessing students' capacity for analytic thinking and their ability to present their ideas coherently (and legibly). Essays can also be used to assess students' attitudes and their aptitude for looking at a problem from different angles.

Revised approach to assessment

The revised approach to assessment by the Division of Psychiatry at the University of Witwatersrand requires blueprinting i.e. the planning of test content to include all learning objectives (knowledge, skills, and attitudes etc.).^{4,5} It promotes the use of a larger variety of assessment formats (structured ward performance and logbooks assessment; MCQ; MEQ; objective structured clinical examination (OSCE), standardized long case presentation and orals) and the use of many examiners. The validity of the assessment process is improved by the use of assessment tools that are appropriate for the objectives being tested. For example, an MCQ assessment (used in the previous and revised approach) is a more appropriate and valid test of knowledge than communication skills, which is better assessed with an interactive tool like the OSCE (used in the new approach). MCQ's have previously shown high reliability because of the large number of items that can be easily tested and marked. These have been adapted in the revised approach to test knowledge as applied to problem solving or clinical reasoning, making them more versatile. The OSCE is a timed examination in which students moved from one station to the

next to demonstrate some combination of history taking, physical examination, counselling, or other aspect of patient management. At each station candidates' performances are rated on checklists and global rating scales. It provides reliability through adequate sampling and standardization. It also incorporates a structured assessment format, which allows the nature of problems and the level of difficulty to be standardized for all students.^{6,7}

The long essay format has been replaced by MEQ's to address some of the criticisms of essay style questions. They are designed to sample a student's knowledge covering a greater proportion of the content of a course or curriculum and to test problem solving and decision-making ability. It is also easier to achieve reliable (consistent) marking between different examiners (using standard answers and marking schedules) than is possible with essay style questions. The duration of the examination is often similar to that of an essay style question. The MEQ format usually involves a case history that is sequentially revealed to the candidate. An MEQ is usually introduced with a brief case scenario followed by a short answer style question. Having completed the first question the candidate turns to section two of the problem where a supplementary question is posed with or without further clinical information and so on. The marks for answers to each question are usually clearly indicated.

The use of multiple sets of examiners (inter-rater reliability) and standardised long cases and orals (inter-case reliability) also improves reliability⁸ and promotes consistency.

The current training of students in Psychiatry at the University of Witwatersrand is in small groups (about 30 students) throughout the year. Each student undergoes two weeks training at one of following clinical sites viz. Chris Hani Baragwanath Hospital, Tara Hospital, Helen Joseph Hospital and Johannesburg Hospital. At the sites students work closely with the nursing and medical staff and learning is clinically based with only a few didactic tutorials to help with some of the basic principles. The objectives and educational contents of the course are web-based and provided to the student at the beginning of the course. Pre-reading by the student is essential. In this current training schedule students are assessed at the end of each six-week block. Because the training and assessment of students occurred in various blocks and at various clinical sites, the purpose of this paper was to describe and compare the results of students in psychiatry at various training sites during the course of an academic year at the University of Witwatersrand following recent reforms in training and assessment methods.

Method

All fifth year medical students (in a six year course) in 2005 were included in the study. Students were assessed at the end of each six-week block (a total of seven blocks in the year). The breakdown of the assessment was as follows: structured long case and oral (20 marks), logbook and ward performance (20 marks); OSCE (20 marks), MCQ (20 marks) and MEQ (20 marks). The final 'block mark' (100 marks) was made up of a combination of all the marks.

At the end of the year the performances of all groups of students, in each of the various assessment tools, were captured. Each set of assessments over the seven blocks was checked to determine if all the objectives set out in the

Table I: Mean marks (SD) of the various assessment tools in the groups of students in the seven blocks

Assessment formats	Block 1 (n=30)	Block 2 (n=26)	Block 3 (n=26)	Block 4 (n=29)	Block 5 (n=28)	Block 6 (n=26)	Block 7 (n=26)
Ward	14.3 (0.9)	13.8 (1.1)	15.2 (1.1)	15.0 (1.1)	14. (1.0)	14.8 (1.1)	14.5 (0.9)
Case	14.0 (1.1)	13.9 (1.4)	14.2 (1.0)	14.5 (0.9)	14.2 (1.1)	14.4 (1.4)	14.3(1.0)
MCQ	13.1 (2.0)	10.68 (2.2)	10.6 (2.3)	11.8 (1.9)	12.4 (2.1)	13.0 (1.4)	13.4 (1.3)
MEQ	12.8 (2.5)	12.6 (2.0)	14.1 (1.8)	15.0 (2.4)	13.2 (3.2)	15.5 (1.8)	15.4 (2.0)
OSCE	15.3 (2.3)	15.5 (2.4)	15.5 (2.0)	15.9 (2.0)	17.3 (1.7)	16.2 (1.4)	15.6 (1.4)
Final mark	69.5 (4.7)	66.3 (5.6)	69.5 (4.6)	72.1 (4.5)	71.7 (6.2)	74.0 (3.2)	73.1 (4.5)

curriculum were assessed. Comparisons were made between the mean marks obtained in each of the assessment tools by students at the four clinical sites and over the seven blocks.

Statistical analysis

The overall performances of all students and groups of students at the different clinical sites and in the different blocks for each assessment tools were computed as means with standard deviations. To determine if there were differences in the mean marks (continuous variable) of the groups of students in the seven blocks (categorical variable) and at the four clinical sites (categorical variable) for each of the five assessment tools, a one-way analysis of variance (ANOVA) was computed. All analysis was done using Statistical Package for Social Sciences 10.0 for windows (SPSS inc., Chicago, IL.). A value of $p < 0.05$ was considered significant.

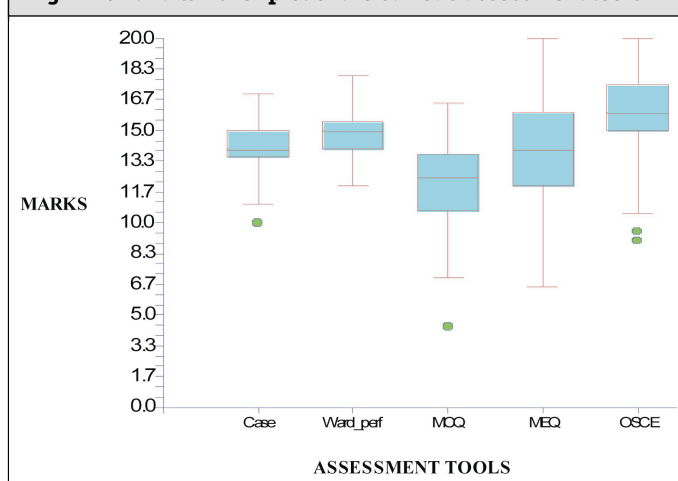
Results

191 students were assessed in 2005. The overall results of the various assessment tools were as follows: mean ward performance mark 14.6 (SD 1.07) (95% CI 14.37-14.69) (min-max: 12-18); mean case presentation mark 14.2 (SD 1.15) (95% CI 14.04-14.40) (min-max: 10-17); mean MCQ's mark 12.2 (SD 2.16) (95% CI 12.08-12.71) (min-max: 4.5-16.5); mean MEQ's mark 14.1 (SD 2.53) (95% CI 13.66-14.08) (min-max: 6.5-20); and mean OSCE mark 15.9 (SD 1.99) (95% CI 15.65-16.26) (min-max: 9-20) (Fig 1). Three students failed (<60 marks) and had to repeat the block.

Analysis of the contents all the assessment tools in each of the seven blocks confirmed that all the objectives set out at the beginning of the course were assessed.

One-way analysis of variance showed significant differences in the mean ward marks ($p < 0.05$); MCQ marks ($p < 0.05$); MEQ marks ($p < 0.05$) and OSCE marks ($p < 0.05$) between the groups in the seven blocks but no differences in the mean case presentation marks ($p > 0.05$) (Table I).

One-way analysis of variance showed no significant

Fig 1: Box and whisker plot of the various assessment tools.

differences in the mean ward marks ($p > 0.05$); case presentation mark ($p > 0.05$); MCQ marks ($p > 0.05$); MEQ marks ($p > 0.05$) and OSCE marks ($p > 0.05$) between the groups at the four clinical sites (Table II).

Discussion

The marked obtained by students in the revised assessment process adopted by the Division of Psychiatry in 2005, showed a normal distribution and spread and differentiated between students with differences in ability. Training and assessment tools had been reformed with the view towards improving validity and reliability. Blueprinting (assessment of all objectives set out at the beginning of the course) of each of assessments was achieved and was not affected by the fact that this had to be done in seven block assessments throughout the year. Overall, the criteria for assessment set out at the beginning of the year were achieved in this year.

Training of students did not appear to be affected by the fact that it was done at different clinical sites and by different

Table II: Mean marks (SD) of the various assessment tools in the groups of students at the four clinical sites.

Assessment formats	CHB (n=63)	Tara (n=64)	HJH (n=31)	JH (n=33)
Ward performance	14.6 (1.1)	14.9 (1.0)	14.3 (1.2)	14.4 (0.9)
Case presentation	14.2 (1.3)	14.2 (0.9)	14.4 (1.0)	14.1 (1.4)
MCQ	12.7 (1.7)	12.4 (2.1)	11.9 (2.3)	12.2 (2.3)
MEQ	14.4 (2.5)	14.0 (2.6)	13.0 (2.5)	14.8 (3.0)
OSCE	15.9 (1.8)	15.9 (2.3)	15.9 (2.0)	16.3 (1.6)
Final Mark	71.7 (4.2)	71.2 (6.1)	69.5 (5.2)	71.7 (6.0)

trainers as evidenced by the fact that there were no significant differences between marks at the four clinical sites. Because this was of concern to the department, it pre-empted this by having regular workshops of all trainers in 2004. It is possible that this vigorous pre-training ensured that the performance of students was not dependent on the site of training or by the trainers.

All the paper based and the OSCE tools were set by the same examiners to control for variations in the difficulty. Despite this, students tended to perform better as the year progressed as evidenced by the difference in mean marks in the different blocks. It is possible that students, who did their psychiatry block later in the year, had spent more time doing clinical medicine and were more familiar and comfortable with dealing with patients in general and ward procedures and consequently performed better. However, the differences may be due to other factors such as students' abilities, etc. that did not form part of this study. Notwithstanding this, there is reasonable consistency from block to block throughout the year so as not to affect previously validated and reliable assessment tools and the assessment process as a whole.

There are unique aspects in the training and assessment at the University of Witwatersrand which make a comparison with other locations difficult. Further evaluation of the process needs to be done on a regular basis to confirm / refute these findings. One study in Africa looked at commonly used methods of assessment to evaluate each method in terms of its ability to test students' performance and the cost and suitability thereof. The author concluded that OSCEs are the most effective assessment method but lack of funding in Africa makes their introduction impossible.⁹

Conclusion

The revised process by Division of Psychiatry at the University of Witwatersrand seems to have brought about improvements in the training and assessment of students by virtue of consistency at the various sites and in the different blocks in the given academic year.

However, the repeated assessments are labour-intensive and therefore costly. It is demanding for clinicians with joint academic appointments who are entrusted with the responsibility of medical education, to continue to assess

undergraduate students in this manner within the country's limited resources. If the revised assessment process is to be continued this is an area that needs to be addressed.

The challenge, however, in the revised assessment process still lies in predicting the student's actual performance on the wards or in the consulting room and the development of reliable assessments with predictive value of subsequent clinical competencies and a simultaneous educational role are yet to be achieved.¹⁰

References

1. Neufeld VR, Norman GR. *Assessing clinical competence, vol 7*, Springer, New York. 1985
2. Swanson DB, Norman GR, Linn RL. *Performance-based assessment: lessons learnt from the health professions. Educational Researcher* 1995; 24: 5-11.
3. Gleeson F. *The effect of immediate feedback on clinical skills using the OSLEER. In: Al Rothman and R Cohen, Editors, Proceedings of the sixth Ottawa conference of medical education 1994, University of Toronto Bookstore Custom Publishing, Toronto, 1994, 412-415.*
4. Dauphinee, D. *Determining the content of certification examinations. In: D Newble, B Jolly and R Wakeford, Editors, The certification and recertification of doctors: issues in the assessment of clinical competence, Cambridge University Press, Cambridge, 1994, 92-104.*
5. Hays RB, Van der Vleuten C, Fabb WE, Spike NA. *Longitudinal reliability of the Royal Australian College of General Practitioners certification examination. Medical Education* 1995; 29: 317-321.
6. Newble DI, Swanson DB. *Psychometric characteristics of the objective structured clinical examination. Medical Education* 1996; 22: 325-334.
7. Harden RM, Gleeson FA. *ASME medical educational booklet no 8: assessment of medical competence using an objective structured clinical examination (OSCE). Journal of Medical Education* 1979; 13: 41-54.
8. Swanson DB. *A measurement framework for performance based tests. In: IR Hart and RM Harden, Editors, Further developments in assessing clinical competence. Can-Heal, Montreal, 1987, 13-45.*
9. Iasi HI. *African medical schools need to consider problem based learning. Student British Medical Journal* 2003;11:393-436
10. Van Der Vleuten C. *Validity of final examinations in undergraduate medical training, British Medical Journal* 2000; 321: 1217-1219.