



Predictive validity of measurements of clinical competence using the Team Objective Structured Bedside Assessment (TOSBA): Assessing the clinical competence of final year medical students

Frances M. Meagher, Marcus W. Butler, Stanley D.W. Miller, Richard W. Costello, Ronan M. Conroy & Noel G. McElvaney

To cite this article: Frances M. Meagher, Marcus W. Butler, Stanley D.W. Miller, Richard W. Costello, Ronan M. Conroy & Noel G. McElvaney (2009) Predictive validity of measurements of clinical competence using the Team Objective Structured Bedside Assessment (TOSBA): Assessing the clinical competence of final year medical students, *Medical Teacher*, 31:11, e545-e550, DOI: [10.3109/01421590903095494](https://doi.org/10.3109/01421590903095494)

To link to this article: <https://doi.org/10.3109/01421590903095494>



Published online: 12 Nov 2009.



Submit your article to this journal [↗](#)



Article views: 1435



View related articles [↗](#)



Citing articles: 1 View citing articles [↗](#)

Predictive validity of measurements of clinical competence using the Team Objective Structured Bedside Assessment (TOSBA): Assessing the clinical competence of final year medical students

FRANCES M. MEAGHER, MARCUS W. BUTLER¹, STANLEY D.W. MILLER¹, RICHARD W. COSTELLO¹, RONAN M. CONROY² & NOEL G. MCELVANEY¹

¹Department of Medicine, Royal College of Surgeons in Ireland, Beaumont Hospital, Dublin, Ireland, ²Department of Epidemiology and Public Health, Royal College of Surgeons in Ireland, Dublin, Ireland

Abstract

Background: The importance of valid and reliable assessment of student competence and performance is gaining increased recognition. Provision of valid patient-based formative assessment is an increasing challenge for clinical teachers in a busy hospital setting. A formative assessment tool that reliably predicts performance in the summative setting would be of value to both students and teachers.

Aim: This study explores the utility of the team objective structured bedside assessment (TOSBA), a novel ward-based formative assessment tool, in predicting student performance in the final clinical examination.

Methods: The performance of a cohort of final year students ($n=191$) in the TOSBA was compared with their subsequent performance in the final examination. A comparison was also made between student performance in the existing formative assessment tool, the objective structured long examination record (OSLER) and the final examination. We also examined the relationship between the TOSBA and the components of the final examination using clustering around latent variables analysis.

Results: There was a clear relationship between student performance in the TOSBA and performance in the final examination ($r^2=0.35$). Student performance in the OSLER showed a poor relationship with performance in the final examination ($r^2=0.15$) compared with the TOSBA. The TOSBA results showed particular correlation with specific components of the final examination which were clinically based.

Conclusion: TOSBA performance is a strong predictor of subsequent performance in the final examination. The clustering of the TOSBA with other assessments of clinical skills underlines its utility. Further research is required to determine whether performance in the TOSBA is predictive of subsequent performance during internship.

Introduction

The importance of valid and reliable assessment of student competence and performance is gaining increased recognition. Society and other stakeholders rightly demand that the final medical examination delivers doctors who are competent and fit to practice as an intern. The predictive value of measurements obtained in medical school and future performance in medical practice has been the subject of a recent Best Evidence Medical Education (BEME) review (Hamdy 2006). Prediction of future student performance is of interest both to students and their clinical teachers and formative assessment is an important aspect of student learning and professional development. A formative and authentic assessment tool that reliably predicts future student performance, in

Practice points

- The team objective structured bedside assessment (TOSBA) is a novel and feasible approach to formative ward-based assessment.
- Performance in the TOSBA is predictive of subsequent student performance in the final examination.
- Further research is required to evaluate the TOSBA as a summative assessment tool and to establish its utility in predicting clinical performance during internship.

a summative setting, would be of great value (Wilkinson & Frampton 2004).

*Marcus W. Butler, Weill Cornell Medical Centre-New York Presbyterian Hospital, NY, NY 10021, USA *Stanley D.W. Miller, St. James' University Hospital, Leeds LS9 7TF, UK *Current affiliation

Correspondence: Frances M. Meagher, Department of Medicine, RCSI Education & Research Centre, Royal College of Surgeons in Ireland, Beaumont Hospital, Dublin 9, Ireland. Tel: 353-1-8093815; fax: 353-1-8093765; email: fmeagher@rcsi.ie

Prior to 2005, the objective structured long examination record (OSLER) was the main assessment instrument used at the Royal College of Surgeons in Ireland medical school in the formative assessment of final year students (Gleeson 1997). We have recently described the team objective structured bedside assessment (TOSBA) which adapted the group objective structured clinical examination (GOSCE) (Biran 1991) and the team objective structured clinical examination (TOSCE) format (Singleton et al. 1999) as a formative ward-based teaching and assessment tool, using ward-based patients (Miller et al. 2007). The TOSBA has been introduced and is now used, in addition to the OSLE, as a formative assessment tool of final-year student clinical competence.

Aim

The aim of this study is to assess the utility of the TOSBA *in predicting performance in the final clinical examination*. Predictive validity is assessed by comparing the results of the TOSBA with the results of the final medical examination held later in the same academic year. In addition, we compare student performance in the TOSBA with that in the OSLE. Convergent and divergent validity is assessed by examining the relationship between student performance in the TOSBA and performance in the individual clinical components of the final examination.

Methods

The curriculum at our institution is a system-based, 5-year programme. The final examination is an assessment in medicine and surgery and comprises a number of components. These can be divided into those that assess knowledge (essay, short notes, multiple choice questions and a 10-station data OSCE) and those that assess clinical, patient-centred skills (an observed long case and a 10 station clinical OSCE). The clinical OSCE consists of eight patient-based stations (each 7.5 min duration) and two communication stations (each 10 min duration). Students who perform well in the final examination are awarded a distinction or 'Honours'. This is designated as a 'P+' (Appendix).

We have previously reported a detailed description of the TOSBA (Miller et al. 2007). Briefly, the TOSBA is a ward-based teaching and formative assessment during which three groups of five students rotate through three bedside stations in the same medical ward. Each station is comprised of an in-patient and an examiner. Consecutive students in each group are each given 5 min to perform one of the five different standardised clinical tasks: (i) take a brief but focused history, (ii) perform a targeted physical examination, (iii) generate a patient-specific differential diagnosis, (iv) outline an investigation and management plan and (v) answer questions pertaining to the patient's drug prescription chart. The students are directly observed performing the tasks, are graded on their performance (Table 1) and provided with educational feedback by the examiner. On completion of the TOSBA, all three examiners confer and an agreed final grade is awarded. Students are scheduled to attend two TOSBAs during their

Table 1. TOSBA and OSLE grading scheme.

Grade	Descriptor
P+	Honours standard
P/P+	Pass but potential to achieve honours standard
P	Pass standard
P/P- (TOSBA only)	Borderline standard
P-	Fail standard

4-week medicine rotation and therefore, ideally, see a patient with a problem in six different (organ) systems, depending on the available in-patient case-mix.

The OSLE is a 10-item analytical record of the traditional long case. The 10-item scale includes four items on *presentation of history*, (pace/clarity, communication process, systematic presentation and correct facts established), three items on *physical examination* (systematic, technique and correct findings established) and one item each on the *formulation of appropriate investigations in a logical sequence, management and clinical acumen* (Gleeson 1997). Unlike the TOSBA, students are not directly observed during the history-taking component of the OSLE. Educational feedback is provided and students are awarded one of the four grades on their performance – the P/P is not included (Table 1). The OSLE assessments were performed by members of the clinical team, to which a student was attached. An average of four (3.927) OSLEs were completed per student.

From September 2005 to March 2006, a total of 204 final-year medical students were exposed to TOSBAs over the course of the academic year. One hundred and ninety-one students sat for the final examination in medicine and surgery and were the subject of this study (11 students were not examined due to ineligibility and one student was absent). A core group of eight clinical faculty members who were familiar with curricular outcomes and expected level of student clinical competency carried out the TOSBA assessments. While no formal examiner training took place, the examiners had considerable experience in undergraduate teaching and assessment. An average of two (1.65) TOSBAs were completed per student. Data were analysed with Stata/SE release 10. Clustering around latent variables analysis was used to examine the patterns of association between assessment modalities.

Results

The performance of 191 students was analysed, and complete assessment data was available for 172 (90%) students. The relationship between the OSLE, TOSBA and final examination mark is illustrated in Figure 1. Student performance in the OSLE showed a poor relationship with performance in the total final examination ($r^2=0.15$). The OSLE showed a restricted grade distribution, with 56% of students achieving the same grade (P+) and a further 26% achieving a P/P+ (Table 2). Furthermore, there was only a two-mark mean difference in the final examination performance between students who received a P- grade and those who received a P/P+. While students who scored P- had a higher average mark than those who scored P, this difference was not

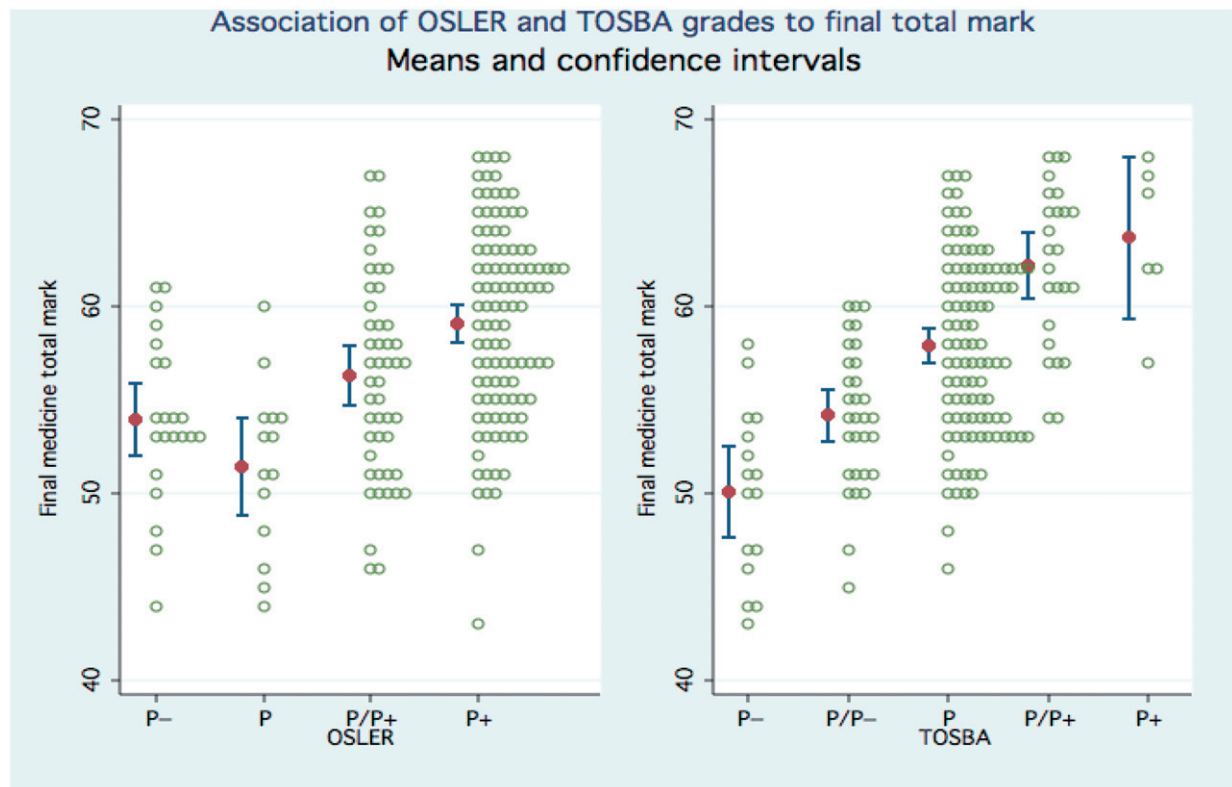


Figure 1. Association of OSLER and TOSBA grades to final total mark. OSLER $r^2=0.15$; TOSBA $r^2=0.35$.

statistically significant (Scheffé *post hoc* test, $p=0.444$). This pattern was repeated for the final clinical mark.

In comparison, as shown in Figure 1, a moderate correlation was found between the TOSBA and the final examination performance ($r^2=0.35$). The TOSBA had a similar concentration of grades to the OSLER, with 57% of students receiving the same grade (P) and only 12% receiving extreme grades (P– and P+) (Table 2). However, the relationship of TOSBA grades to the final examination result showed a better discrimination, with a 12-mark difference in average performance between those who received a P– grade and those who received a P/P+ and a graded association between TOSBA grade and an average final mark, though the two-mark difference between the final results of those who achieved P/P+ and P+ grades is not statistically significant (Scheffé *post hoc* test $p=0.990$). A similar pattern was again seen for final clinical mark.

The relationship between the TOSBA and the failure and honours rate in the final examination mark is illustrated in Figure 2. There is a clear relationship in both cases. A student who performed poorly in the formative TOSBA (P–) did not achieve an honours grade in the final examination and were likely to fail (38%). The converse is true for students who performed well in the TOSBA (i.e. scored either a P/P+ or P+ grade) – they had an 80% chance of achieving honours. In addition, no student who scored a P/P+ or P+ failed in the final examination. Of those students who scored a P in the TOSBA, 44% subsequently achieved an honours grade in the final examination.

Of the 11 students who failed in their final examination, nine also failed in the TOSBA (sensitivity 82%, 95%

Table 2. Comparative performance of the OSLER and TOSBA.

Assessment		Percentage	Average final mark (out of 100)	Average clinical mark (out of 100)
	Grade			
OSLER	P–	12	54	56
	P	7	51	49
	P/P+	26	56	57
	P+	56	59	60
TOSBA	P–	9	50	50
	P/P–	17	54	53
	P	57	58	59
	P/P+	14	62	62
	P+	3	64	64

CI 48–98%). Seven of these (77%) received a P– grade and two (23%) received a P/P– grade. There were 47 failures overall on the TOSBA giving a predictive value of failure of 19% (95% CI 9% to 33%). Of the 17 students who failed in the clinical component of their final examination, eight had failed in the TOSBA, giving a sensitivity of 47%, 95% CI 23–72%. The predictive value was 17% (95% CI 8–31%).

Correlation between the TOSBA and final examination components

We examined the relationship between the TOSBA and the components of the final examination using clustering around latent variables analysis. Clustering around latent variables

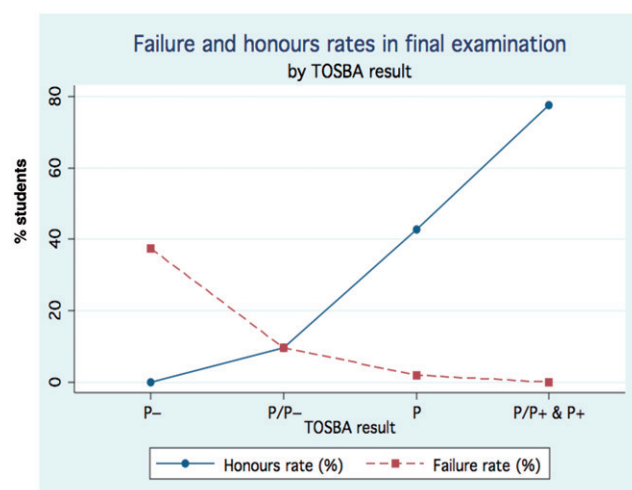


Figure 2. Association gradients for failure rates and honour rates.

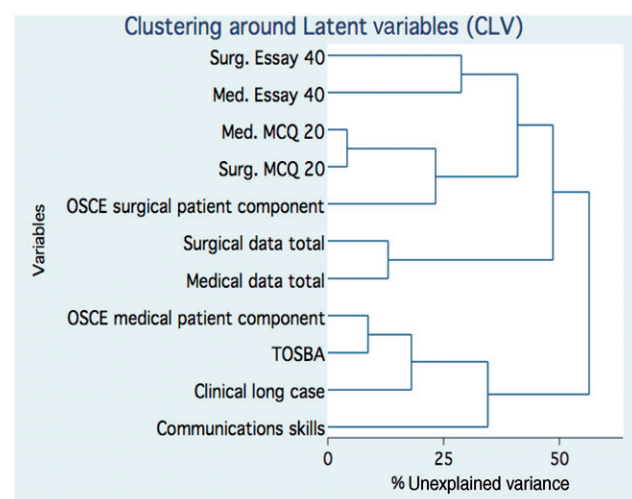


Figure 3. Association gradients for failure rates and honour rates.

sequentially groups the variables into clusters with the aim of minimising the variation within clusters and maximising the variation between clusters (Vigneau & Qannari 2003). Unlike factor analysis, it is a sequential process that makes it useful for detecting complex data structures such as clusters-within-clusters. This results in a hierarchical cluster analysis of the variables. The output of the analysis is shown in Figure 3. The assessments formed two broad clusters: the cluster at the top of the figure contains all of the written assessments. It can be seen that there is a strong link between assessments of the same type – the two multiple choice question (MCQ) exams cluster together, as do the data interpretation examinations (OSCE) and the essay examinations. Taken together, this cluster is dominated by assessments which test the knowledge domain and entail recognition, recall and organisation of material.

The lower cluster is the patient-centred assessments. The TOSBA clusters with the medical *in vivo* patient-based component of the OSCE, the clinical long case and the

communication skills assessment. It correlated less well with the medical *in vitro*, data-based component of the OSCE and poorly with the MCQ, essay paper and short notes assessments that were fact-orientated and measured knowledge and memory

Discussion

The analysis presented in this article represents an evaluation of the utility of the TOSBA in predicting subsequent student performance and identifies aspects of the TOSBA which could improve its value as a formative assessment tool of final-year student clinical competence.

Validity refers to the extent to which a measurement actually measures what it is intended to measure (Van der Vleuten 2000). Validity is not so much a property of a test, but rather refers to the usefulness of the test for a particular purpose. Multiple sources of evidence are required to evaluate the appropriateness of a test for a particular purpose (Sireci 2007). Although in the present evaluation we were limited to examine the relationship between performance in the TOSBA and other measures of student performance, we were also able to use these to examine convergent, divergent and predictive validity, all characteristics of the assessment process which are often neglected because of the difficulties inherent in their determination (Van der Vleuten 2000). Convergent validity was supported by the clustering of the TOSBA scores with other measures of clinical competence, and divergent validity supported by their distinctness from knowledge-based assessments.

Student performance in the TOSBA was predictive of performance in the final examination which, in itself, is unremarkable since it assesses skills which are also assessed in components of the final examination. In particular, however, 82% of those who would ultimately fail their in final examination failed in the TOSBA. However, given the failure rate in the TOSBA, this amounted to a predictive value of just under 20%, i.e. only 20% of students failing in the TOSBA went on to fail in the final examination. This may be, in part, explained by the subsequent remediation programme which was provided for those students who failed in their TOSBA, accounting for an improved final examination pass rate. Timely intervention based on early identification of poor clinical performance has been shown to help weaker students to improve their performance (Sayer et al. 2002). In addition, although poorly performing students have been shown not to seek out guidance and support (Malik 2000), we believe it is possible that students, who were identified as potential failures by the TOSBA, used this formative assessment as an incentive to improve their performance. Closer monitoring of the impact of remedial support on the subsequent performance of under-performing students will be an area of future research.

The TOSBA grades showed an ordered relationship with performance in the final medical examination which was superior to that of the OSLE, and a more impressive discriminant ability, with important differences in performance on the final examination being evident between the highest and lowest TOSBA grades. The ordered relationship with performance in the final medical examination shown by the

TOSBA was not seen with the OSLEP which demonstrated poor discriminant ability, with little differences in performance on the final examination being evident between the highest and lowest OSLEP grades. The poor performance of the OSLEP may reflect the junior status and inadequate examiner training of those who carried out these assessments – predominately interns or senior house officers. In addition, observation of the history-taking process in the TOSBA, unlike the OSLEP (unobserved), may improve the validity of the TOSBA grades. As a result of our data showing the clear superiority of the TOSBA over the OSLEP, we no longer use the latter as a formative assessment tool.

The TOSBA was a useful predictor of an honours performance in the final examination. A student who achieved a P/P+ or P+ grade had an 80% chance of achieving honours. The converse is true for students who performed poorly in the formative TOSBA – they were very unlikely to achieve an honours grade in the final examination.

While the TOSBA is a good predictor of extremes in performance, it is less reliable for those students who received a P grade. Examination of the TOSBA grade breakdown suggests that the criteria for the P/P+ and P+ grades need to be reviewed. Forty-four percent of students who obtained a P grade in the TOSBA subsequently achieved honours (P+) in their final examination. Further analysis is required to evaluate this central cluster of grades.

The use of clustering around latent variables to examine convergent and divergent validity of the TOSBA is a new and potentially useful statistical tool (Vigneau & Qannari 2003). Clustering around latent variables shows the interrelationships between assessments as a tree diagram allowing the reader to see clustering at several levels. In the case of our final year assessments, it is clear that there are two broad categories, corresponding to knowledge-based and clinical skill assessments, and that the TOSBA fits into the latter category. However, within the assessments, a finer structure can be seen: assessments that test the same domain tend to cluster together – notably the MCQ examinations, whose close relationship indicates that there is an element of 'MCQ skill' underlying performance on these assessments.

We have not used the TOSBA in the final examination. This study is an exercise in the use of a formative assessment tool, with a high throughput, to determine its utility in predicting student performance and to evaluate its potential as a summative tool. In the context of increasing student numbers, the sustainability of the OSCE as an assessment tool is open to question (Harden & Gleeson 1979). The challenges of running an OSCE of psychometric integrity for large student groups are well documented (Van der Vleuten & Swanson 1990). This may be a circumstance where the TOSBA could be introduced with good effect. The use of real patients in the TOSBA confers an authenticity, which is not a feature of the OSCE, to the assessment of clinical competence. In addition, an earlier study of the TOSBA has shown that students appreciated the educational opportunity of learning from their peers in a team setting during the formative TOSBA assessment process (Miller et al. 2007).

Evaluation of test validity is not a static, one-time event (Sireci 2007). We are currently using the results of this

evaluation to improve the conduct of the TOSBA assessment, concentrating on examiner training and clearer grading guidelines for examiners with the aim of reducing the proportion of students clustered into the same grade.

Conclusion

The challenge of providing rigorous assessment of medical student cohorts of increasing size is one that faces all medical schools. The potential to reliably assess students in groups, in an authentic environment, represents a move away from traditional approaches to undergraduate assessment. If this approach can be implemented, without jeopardising the quality of the assessment, it is to be welcomed.

We believe the TOSBA has development potential as an assessment tool. The potential is shown in the good discriminatory ability of the grades as predictors of final examination performance. The clustering of the TOSBA with other assessments of clinical skills underlines its utility. Future analyses will address the current failure of the grading scheme to identify many of the students who subsequently achieved honours in the final examination. Our findings have education and research implications. We believe the TOSBA provides valid measurements of clinical competence that are useful in the formative assessment of medical students in an authentic setting, though it may benefit from further refinement. Further research is required to determine whether undergraduate performance as measured by the TOSBA is predictive of subsequent performance during internship.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Notes on contributors

F.M. MEAGHER is a Corrigan Lecturer in Medicine and Physician at the Royal College of Surgeons in Ireland (RCSI).

S.D.W. MILLER is a Lecturer in Medicine and Specialist Registrars in Respiratory Medicine, RCSI.

M.W. BUTLER is a Lecturer in Medicine and Specialist Registrars in Respiratory Medicine, RCSI.

R.W. COSTELLO is a Senior Lecturer in Medicine, RCSI, and Respiratory Physician in Beaumont Hospital, Dublin.

R.M. CONROY is a Senior Lecturer in Epidemiology and Public Health, RCSI.

N.G. McELVANEY is a Professor of Medicine and Chairman of the Department of Medicine, RCSI, and Respiratory Physician, Beaumont Hospital, Dublin.

References

- Biran LA. 1991. Self-assessment and learning through GOSCE (group objective structured clinical examination). *Med Educ* 25(6):475–479.
- Gleeson F. 1997. The objective long examination record (OSLER). *Med Teach* 19:7–14.
- Hamdy H, Prasad K, Anderson MB, Scherpbier A, Williams R, Zwierstra R, Cuddihy H. 2006. BEME systematic review: Predictive values of measurements obtained in medical schools and future performance in clinical practice. *Med Teach* 28(2):103–116.

- Harden RM, Gleeson FA. 1979. Assessment of clinical competence using an objective structured. *Med Educ* 13(1):41–54.
- Malik S. 2000. Students, tutors and relationships: The ingredients of a successful support scheme. *Med Educ* 34:635–641.
- Miller SDW, Butler MW, Meagher F, Costello RW, McElvaney NG. 2007. Team objective structured bedside assessment (TOSBA): A novel and feasible way of providing formative teaching and assessment. *Med Teach* 29:156–159.
- Sayer M, Chaput De Saintonge M, Evans D, Wood D. 2002. Support for students with academic difficulties. *Med Educ* 36(7):643–650.
- Singleton A, Smith F, Harris T, Ross-Harper R, Hilton S. 1999. An evaluation of the team objective structured clinical examination (TOSCE). *Med Educ* 33(1):34–41.
- Sireci SG. 2007. On validity theory and test validation. *Educ Res* 36(8):477–481.
- Van der Vleuten CPM. 2000. Validity of final examinations in undergraduate medical training. *Brit Med J* 321:1217–1219.
- Van der Vleuten CPM, Swanson DB. 1990. Assessment of clinical skills with standardized patients: State of the art. *Teach and Learn Med* 2:58–76.
- Vigneau E, Qannari EM. 2003. Clustering of variables around latent components. *Commun Stat-Simul C* 32(4):1131–1150.
- Wilkinson TJ, Frampton CM. 2004. Comprehensive undergraduate assessments improve prediction of clinical performance. *Med Educ* 38:1111–1116.

Appendix: OSLER grading system.

Extended criterion		Extended marking scheme	
Referenced grading scheme			
P+	80	Outstandingly clear and factually correct presentation of the patient history, demonstration of physical signs and organisation of the case management. Clearly a candidate displaying outstanding communication skills and clinical acumen. First class honours.	
	75	Excellent overall case presentation, communication skills, examination technique and demonstration of the correct facts and physical signs of the case. The candidate may even display outstanding attributes in some but not all measurable criteria. First class honours.	
	70	Excellent in most aspects of overall case presentation, communication skills, examination technique and demonstration of the correct facts and physical signs of the case. Also excellent communicator and demonstrates the ability to investigate	
	65	Very good overall presentation covering all major aspects; few omissions, good priorities. Very clearly an above average candidate in terms of communication and clinical acumen. Second class honours, grade 1	
	60	Very good in most respects of presentation and communication but not in all aspects. However, a good solid performance in most areas assessed with a well developed clinical acumen. Second class honours, grade 2	
P	55	Good sound overall presentation and communication of the case without displaying any attributes out of the ordinary. The candidate displays an overall adequate standard of examination technique. The patient's problems are identified and a reasonable management outline suggested.	
	50	Adequate presentation of the case and communication ability. Nothing to suggest more than just reaching an acceptable standard in physical examination and identification of the patient's problems.	
P–	45	Poor performance in most areas of case presentation, communication with the patient and demonstration of physical signs. Inadequate attempt at a clear identification of the patient's problems. The candidate may display some adequate attributes but does not reach an acceptable pass standard overall.	
	40	Poor performance overall The candidate does not reach an acceptable pass standard in almost all categories. The candidate's performance in all areas is 'so poor' that the standard required is not even remotely approached. Quite clearly this candidate requires a further period of training	