

# The validity of oral assessment (viva) that assesses specific and unique competencies in a post-graduate psychiatry examination

Varuni de Silva, Raveen Hanwella, Gominda Ponnampereuma

## Background

Studies have criticized oral assessments for having poor validity and reliability. There is limited research on structured oral assessments that assess specific competencies.

## Aims

To evaluate the validity of the oral assessment component of a postgraduate psychiatry examination.

## Methods

A retrospective analysis of the examination scores of 154 candidates from 12 postgraduate psychiatry examinations conducted during an 8 year period was carried out. Concurrent and construct validity was examined by correlating marks at the viva with the marks of theory, clinical long case and clinical short case components of the candidates. Separate multiple regression analyses were conducted to predict the scores of the different components of the examination.

## Results

Repeated measure ANOVA showed there was no significant difference in the means between the

different components of the examination ( $F=0.49$ ,  $p=0.486$ ). The viva was a sensitive method of assessment ( $79/88=89.7\%$ ) but the specificity was low ( $36/66=54.5\%$ ). Positive predictive value was 72.5% and the negative predictive value was 80.0%. Logistic regression analysis showed that the odds of passing the viva and passing the examination compared with passing the viva and failing the exam was 10.53 (95% CI 4.54- 24.47). There was a statistically significant, moderately high correlation between viva and theory components ( $r=.50$   $p<0.001$ ). Multiple regression models showed that viva marks were a predictor of performance at the clinical short case and theory components but not the clinical long case.

## Conclusion

Viva had good sensitivity and positive and negative predictive values. Instead of discontinuing the use of oral assessments, ways should be identified to improve the reliability and validity of the oral assessment.

SL J Psychiatry 2012; 3 (2):16-19

## Introduction

Oral assessment (viva voce) is a component of many undergraduate and postgraduate examinations in medicine. They are also used in assessments for recruitment to medical schools, jobs and career promotions for medical personnel in many countries.

The format of oral examinations is not uniform. Oral assessments can be used as a method of testing knowledge of a specific subject area. These types of oral assessments mainly test recall and are often unstructured. Other oral assessments are used in testing clinical competence. They use hypothetical case scenarios or case vignettes, video-taped patient encounters or real patients to test clinical competence. Variability of the examination is high when questions are not structured. The duration of the oral assessments too vary ranging from a few minutes to one to two hours.

Studies have criticized oral assessments for having poor validity and reliability (1-6). This criticism has led to many professional bodies abandoning or modifying the traditional oral examinations. The validity and reliability of oral examinations can be increased by the use of structured, standardized orals and by training examiners (7). Increasing the examination time and number of examiners is also known to improve reliability of the oral assessment. For oral assessments

based on clinical cases, reliability increases from 0.5 for a one hour assessment to 0.69 for a two hour assessment (8,9). Wass et al. have shown that increasing the duration of the oral assessment, thereby increasing the number of topics examined, and increasing the number of examiners could improve reliability (10). The reliability of using global judgments, appeared to be better than the reliability of averaged item scores (11). Kearny et al. reported that use of a structured oral examination format and global rating scales resulted in fair to good intra-rater and inter-rater reliability (12).

However, except for a few qualitative studies, we could not find quantitative studies on oral assessment that assessed specific competencies such as emergency and acute care management, which cannot be assessed using most other types of assessment such as written and clinical examinations (13).

The aim of this study was to evaluate the validity of a structured oral assessment that mainly assessed emergency and acute care management in the Postgraduate MD Psychiatry examination in Sri Lanka.

## Method

Oral assessment is a component of the Postgraduate MD Psychiatry examination in Sri Lanka. Each candidate is examined by two examiners. The candidate

is presented two clinical scenarios and questioned on differential diagnosis, investigation and treatment. Case scenarios are based on problems not commonly assessed in the clinical long and clinical short cases, such as emergency management and management of restless or confused patients. They are also questioned on other aspects, not dealt with, in detail, in the other components of the examination, such as psychological therapies. Consensus marking is carried out by the examiners. Each candidate is examined for 20 minutes. The candidate is given a global mark ranging from 0-100. The oral assessment contributes 10% to the overall final total.

Apart from the viva the other assessment methods used in the MD Psychiatry examination are the theory examinations, clinical long case and clinical short case. The theory component consists of 60 multiple choice questions (MCQ) each with five true or false responses and an essay paper consisting of six essay questions. The MCQ and essay components each contribute 20% of the total marks. In the clinical long case component a candidate takes the history and examines the patient for one hour and presents the findings and discusses diagnosis and management. The candidate is examined by two examiners for 30 minutes. This component contributes 30% of the total marks. The clinical short case requires the candidate to examine two patients for 15 minutes each. They are assigned a task which may be a general assessment or a more specific task such as to carry out a cognitive assessment. Each candidate is examined by two examiners for 30 minutes. This component contributes 20% of the total marks.

### Data collection and analysis

A retrospective analysis of the examination scores of 154 candidates from 12 MD Psychiatry examinations conducted from 2000-2008 was carried out. Ethical clearance for the study was obtained from the Ethics Review Committee of the Faculty of Medicine, University of Colombo.

We first examined the validity, in general, by calculating the sensitivity, specificity, positive predictive value and negative predictive value of the viva examination in identifying candidates who pass the MD Psychiatry examination.

Next, in order to ascertain concurrent and construct validity studied the correlation of viva marks with the marks of theory, clinical long case and clinical short case marks of the candidates. We were able to obtain the details of the long case and short case marks separately only for 74 candidates. Separate multiple

regression analyses were conducted to predict the scores of the different components of Psychiatry MD examination. The components of the examination considered were theory, clinical long case, clinical short case and viva. Stepwise regression analysis was carried out. Assumptions of normality, homogeneity of variance, linearity of the residuals and independent errors for each model were met. For all models a linear model fitted the data best. SPSS version 13.0 was used in the analysis of data.

## Results

Eighty eight out of 154 candidates (57.1%) passed the MD Psychiatry examination during the period under study.

Marks of all components were calculated on a scale of 1-100. Table 1 shows the mean scores of the different components. A repeated measure ANOVA showed there was no significant difference in the means of the different components ( $F=0.49$ ,  $p=0.486$ ).

We first calculated the sensitivity, specificity, positive predictive value and negative predictive value of the viva by comparing the pass in the viva with an overall pass in the examination. Table 2 tabulates the ability of the viva to predict an overall pass in the examination.

Sensitivity is an indicator of the ability of the viva to identify candidates passing the examination. Specificity is an indicator of the ability of the viva to identify candidates failing the examination. While the viva is a sensitive method of assessment ( $79/88 = 89.7\%$ ) the specificity was low ( $36/66 = 54.5\%$ ). A candidate passing the viva has a 72.5% probability of passing the examination (positive predictive value). A candidate failing the viva has an 80.0% probability of failing the examination (negative predictive value).

Logistic regression analysis showed that the odds of passing the viva and passing the examination compared with passing the viva and failing the exam was 10.53 (95% CI 4.54- 24.47).

Table 3 illustrates Pearson's correlation between the different components of the examination. There is a moderately high correlation between the viva and the theory component which is statistically significant ( $r=.50$   $p<0.001$ ). More than 25% (square of 0.502 as a percentage) of the variation in viva marks can be explained by the theory marks. There is a moderate correlation between marks of the the viva and the short cases ( $r=.460$   $p<0.001$ ). Correlation between viva and the long case is low ( $r=.257$   $p<0.001$ ).

Table 1 – Mean and standard deviation of marks for the individual components

Component	Mean (%)	Standard deviation
Theory	55.30	6.7
Long case	55.33	10.2
Short case	52.53	12.8
Viva	55.21	11.0

Table 2 – Ability of the viva to predict an overall pass in the examination

	Total Pass	Total fail	Total
Viva pass	79	30	109
Viva fail	9	36	45
Total	88	66	154

**Table 3– Correlation of marks of different components of the examination**

	Theory	Long Case	Short Case
Theory	-		
Long Case	0.410 p<0.001	-	
Short Case	0.410 p<0.001	0.349 p=0.004	-
Viva	0.502 p<0.001	0.257 p=0.034	0.460 p<0.001

We fitted different multiple regression models to identify how the marks of each assessment component (excluding the viva marks) were predicted by the marks of the other components (including the viva marks) (Table 4). The models were able to predict 56% of the variance for the theory, 17% for clinical long cases, 27% for clinical short cases and 29% for the viva. The models indicate that the viva marks were a part of the models that significantly predicted the short case and theory components but not the long case.

The B values indicate the relationship between the outcome and the predictors. It also tells us to what degree each of the predictors affect the outcome if all other predictors are held constant.

Standardized beta values are measured in standard deviations and it is a better indicator of the contribution of the predictors to the model. The β value indicates that both the long case and the viva make almost an equal contribution in the model predicting the theory marks. For the short cases the viva was the most important predictor; i.e. when viva marks are added to model C, one standard deviation change in the viva score results in a 0.41 standard deviation change in the clinical short case score. Clinical long case marks (0.26) too were a significant predictor of clinical short case marks. For the long case, the only significant predictor was the theory marks (0.41).

## Discussion

Our study found that a pass in the viva was a sensitive indicator of an overall pass in the examination (89.7%) but it was not very specific (54.5%). Therefore viva performance was better at predicting a pass in the examination than a failure.

Similar means for theory, long case, short case and viva examinations indicate that candidates seem similarly competent in the competencies that each category of examination has assessed. The standard deviation (SD) of theory was narrower than the SD of other components, indicating a narrower spread of marks in the theory component. This means that it is least possible to discriminate candidates using the theory component as opposed to the clinical short case, clinical long case or viva.

**Table 4– Prediction of MD psychiatry examination component scores using stepwise multiple regression analysis**

	B	95% CI	β	R2
<b>Theory<sup>a</sup></b>				0.31***
Long case	0.22	0.09-0.36	0.34	
Viva	0.23	0.11-.36	0.38	
<b>Long Case<sup>b</sup></b>				0.17***
Theory	0.622	0.28-.97	0.41	
<b>Short case<sup>c</sup></b>				0.27***
Long Case	0.33	0.06-0.60	0.26	
Viva	0.47	0.22-0.72.	0.41	
<b>Viva<sup>d</sup></b>				0.29***
Short case	0.28	0.08-.048	0.33	
Theory	0.51	0.13-0.89	0.31	

- a. Dependent variable theory-predictor variable short case was excluded
  - b. Dependent variable long case-predictor variables short cases and viva were excluded
  - c. Dependent variable short case-predictor variable theory was excluded
  - d. Dependent variable viva-predictor variable long case was excluded
- \*\*\*p<0.001

The moderately high correlation between viva and the theory marks indicates that viva component has mostly assessed theoretical knowledge. This confirms the findings in previous studies (3,14-15). However because the oral examination marks only explain 25% of the theory marks, it leaves room for speculation as to whether the oral examination assesses competencies that the theory examination does not. This query is, in particular, of value in the context of this study, as the viva questions were aimed at assessing the ability to handle emergencies which was not specifically assessed in the theory component. The evidence for assessing a separate ability that is not assessed by the theory component or other components would have been stronger if the correlation was lower. A moderate correlation implies that candidates who are good at dealing with emergencies have good theoretical knowledge as well.

The viva does not assess the candidate’s response to an actual emergency situation. It only assesses the candidate’s knowledge about dealing with an emergency. The viva can be modified to take into account the speed of a candidate answering a given question when assessing responses to emergency situations. If the viva specifically considers the speed of response to such questions, even with a correlation coefficient of 0.5 the inclusion of a viva examination could be arguably be justified. In this particular viva, the global mark of the examiners may have reflected the candidate’s speed of responding. Oral assessments have the advantage of human interaction and this interaction can be used positively to assess areas such as the process of clinical decision making of the candidates. However examiner variability needs to be taken into consideration. This variability could be minimized by a structured, standardized oral assessment with detailed examiner guidelines.

The regression analysis indicates the strong influence of the theoretical component in all the components except the clinical short cases. This is strong evidence that the short cases assess a set of abilities or competencies that is least dependent on theory. However, short cases seem to have a strong overlap with viva. In both the oral assessment (viva) and the short cases the candidates have to make quick clinical decisions and this may explain the overlap.

There are several limitations to this study. The small sample size taken in to account when interpreting the results of this study. We did not have access to the individual examiner marks and therefore could not assess reliability. However several studies have reported on the reliability of oral examinations and how oral assessments can be made reliable by including more cases and examiners and structuring the oral assessments (7). Although the sensitivity of the oral assessment and positive and negative predictive values were impressive it should be noted that the non-independence of data (i.e. viva scores being a part of the overall score) used for the calculation of specificity and sensitivity, may have inflated these results.

### Conclusions

The moderate correlations of viva marks and with marks of other components of the MD examination indicate that there is a considerable percentage variation of viva marks which other examination marks cannot account for. Both the correlation and regression analyses provide sufficient evidence to initiate further studies on the validity of oral examinations which assess specific and unique competencies such as handling emergencies. This observation is further supported by the impressive sensitivity, and positive and negative predictive values attained by the oral examination. Instead of discontinuing the use of oral assessments ways should be identified to improve reliability and validity of the oral assessment and to consider its use for evaluating decision making rather than testing knowledge.

### Acknowledgments

We wish to thank Prof. Rezvi Sheriff, former Director, Postgraduate Institute of Medicine, University of Colombo, Sri Lanka for the support and for granting us permission to access the marks of the examination.

### Declaration of interest

First and second authors have examined at the MD (Psychiatry) examination.

**Varuni de Silva**, Senior Lecturer, Department of Psychological Medicine  
**Raveen Hanwella** Senior Lecturer, Department of Psychological Medicine  
**Gominda Ponnampereuma**, Senior Lecturer, Medical Education Development and Research Centre (MEDARC), Faculty of Medicine, University of Colombo, Sri Lanka  
**Corresponding author:** Varuni de Silva,  
**E mail:** varunidesilva2@yahoo.co.uk

## References

1. Davis MH, Karunathilake I. The place of the oral examination in today's assessment systems. *Med Teach* 2005; 27:294-297. [CrossRef](#)
2. Weingarten MA, Polliack MR, Tabenkin H, Kahan E. Variations among examiners in family medicine residency board oral examinations. *Med Educ* 2000;34:13-17. [CrossRef](#)
3. Jayawickramarajah PT. Oral examinations in medical education. *Med Educ* 1985;19:290-293. [CrossRef](#)
4. Sierles FS, Daghestani A, Weiner CL, deVito R, Fichtner CG, Garfield DA. Psychometric Properties of ABPN-Style Oral Examinations Administered Jointly by Two Psychiatry Residency Programs. *Acad Psychiatry* 2001; 25:214-222. [CrossRef](#)
5. Levine HG, McGuire, C.H. The validity and reliability of oral examinations in assessing cognitive skills in medicine. *Journal of Educational Measurement* 1970;7:63-74. [CrossRef](#)
6. Kelley PR, Jr., Matthews JH, Schumacher CF. Analysis of the oral examination of the American Board of Anesthesiology. *J Med Educ* 1971;46:982-988.
7. Schubert A, Tetzlaff JE, Tan M, Ryckman JV, Mascha E. Consistency, inter-rater reliability, and validity of 441 consecutive mock oral examinations in anesthesiology: implications for use as a tool for assessment of residents. *Anesthesiology* 1999;91:288-298. [CrossRef](#)
8. Swanson DB. A measurement framework for performance-based Tests. In: *Further Developments in Assessing Clinical Competence*, Edited by I.R. Hart IR Harden RM. Montreal, Can Health Publications pp. 1987; 13-45.
9. van der Vleuten CP, Schuwirth LW. Assessing professional competence: from methods to programmes. *Med Educ* 2005;39:309-317. [CrossRef](#)
10. Wass V, Wakeford R, Neighbour R, Van der Vleuten C. Achieving acceptable reliability in oral examinations: an analysis of the Royal College of General Practitioners membership examination's oral component. *Med Educ* 2003; 37:126-131. [CrossRef](#)
11. Daelmans HE, Scherpbier AJ, Van Der Vleuten CP, Donker AJ. Reliability of clinical oral examinations re-examined. *Med Teach* 2001; 23:422-424. [CrossRef](#)
12. Kearney RA, Puchalski SA, Yang HY, Skakun EN. The inter-rater and intra-rater reliability of a new Canadian oral examination format in anesthesia is fair to good. *Can J Anaesth* 2002; 49:232-236. [CrossRef](#)
13. Simpson RG, Ballard KD. What is being assessed in the MRCGP oral examination? A qualitative study. *Br J Gen Pract* 2005; 55:430-436. PMID:15970066
14. Nunnink L, Venkatesh B, Krishnan A, Vidhani K, Udy A. A prospective comparison between written examination and either simulation-based or oral viva examination of intensive care trainees' procedural skills. *Anaesth Intensive Care* 2010; 38:876-82. PMID:20865872
15. McGuire CH. The oral examination as a measure of professional competence. *J Med Educ* 1966; 41:267-274. PMID:5907853