

Title: Improving the reliability of standardized patient OSCE stations used during the Family Medicine MBChB6 end-of-block exams at the University of Limpopo (Medunsa Campus), Pretoria

Author: Gboyega A Ogunbanjo, M.B.B.S., M.Fam.Med.

Context and setting: At the University of Limpopo Medical School, the Family Medicine MBChB6 end-of-block exam includes two similarly paired OSCE stations testing consultation skills of students with ‘standardized’ patients.

Why the idea was necessary: Over the years, wide variations in the examiners’ scores were observed for these paired standardized patient OSCE stations. What is not known is whether the inter-examiner variability in scores is due to individual ‘*leniency*’ or ‘*stringency*’ or the ‘*assessment instrument*’ used for scoring the students.

What was done? A 5-step study was done as follows: Six standardized patient OSCE stations were video-recorded; Each video was shown to four examiners who examine students at these standardized patient OSCE stations to score independently using the “*current*” assessment instrument; Thereafter the same videos were shown to them to score independently but using a “*new*” assessment instrument (modified mini-CEX); Afterwards, they were trained on how to use the “*new*” assessment instrument; and after training they scored the same videos using the “*new*” assessment instrument. A focus group interview of the examiners on their experiences was done afterwards.

Evaluation of results and impact: Scores of the examiners with the *current* assessment instrument showed a very good level of agreement between examiners 1a and 1b ($\kappa = 1.00$) but poor between examiners 2a and 2b ($\kappa = 0.18$). With the *new* assessment tool (before training), these were poor between examiners 1a & 1b and 2a & 2b ($\kappa = 0.02$ and 0.18 respectively). After *training* on the new assessment tool, the levels of agreement remained poor between examiners 1a and 1b ($\kappa = 0.00$), while between examiners 2a and 2b, this became good ($\kappa = 0.67$). When intra-observer analysis was done between the *current* and *new* assessment tool, there was poor agreement in examiner 1a ($\kappa = 0.00$), but moderate to fair in examiners 1b ($\kappa = 0.57$), 2a ($\kappa = 0.40$) and 2b ($\kappa = 0.33$). The intra-observer analysis before and after training using the *new* assessment tool showed poor levels of agreement for examiners 1a ($\kappa = 0.00$) and 1b ($\kappa = -0.20$), but fair for examiners 2a ($\kappa = 0.40$) and 2b ($\kappa = 0.33$). Focus group interview of the 4 examiners had the following themes: *More training needed on the new assessment tool*, and *difficulty in assessing “professionalism” and “communication skills”* from the consultations. Study limitation due to small sample size of observations which resulted in poor levels of agreement in some cases. However, there was unanimous agreement among the examiners that the *new* assessment tool is an appropriate comprehensive assessment for students’ performance but more training needed for faculty before implementation.

Correspondence: Professor Gboyega A Ogunbanjo, Department of Family Medicine & Primary Health care, University of Limpopo (Medunsa Campus, Medunsa 0204, South Africa. Tel: 0027125214528; Fax: 0027 125214172; Email: gao@intekom.co.za