



WORKSHOP A

Examining the Reliability and Validity of the SPM Oral Proficiency Test

Pn Usha Diane Dennis
SMK Aminuddin Baki

Ms Goh Guay Khin
SMK Westlands, Penang

ABSTRACT

The recent reformulation of the SPM Oral English examination was done in order to address concerns about the authenticity of the previous model. The old model lacked face validity since reading a prepared dialogue aloud cannot be construed as a real-life situation. Furthermore the allocation of marks did not reflect the demands of the task, giving instead an inflated profile of the candidate. It also created a high-anxiety situation for the candidate.

The new SPM Oral English model has introduced four models of oral tasks. In this workshop, participants will evaluate the authenticity of the tasks as well as the set of criteria used to measure student-performance. In the first part of the workshop, participants will compare a sample of a task from both the old and new formats, to determine in what ways, if any, the tasks in the new format address the gaps found in the tasks contained in the old format. The second part of the workshop will involve participants in a simulation of any one of the models proposed in the new format. After the simulation, participants will collaboratively evaluate the performance against the criteria suggested. The final part of the workshop will draw conclusions on the validity and reliability of the criteria.

WORKSHOP SUMMARY

First, participants were asked to compare the tasks contained in the previous format with that in the new format. It was noted that the old format of reading a text and answering comprehension questions does not test speaking. In the second part of the test, candidates would be shown a pictorial stimulus and asked to answer three questions related to the stimulus. A single word answer could merit 6 marks. The tasks and allocation of marks did not ensure validity and reliability; the task was not meaningful in the real-world context, it was passive in nature not testing features of spoken discourse and the marks may not be an accurate measure of the candidates' true ability.

Participants were then asked to read the aim and objectives of the course and comment on whether the tasks contained in the models will allow for the candidate to demonstrate the learning outcomes. There was more flexibility in the new format. The management of the examination situation was less intimidating with single presentation, pair and group work. In addition, the assessor was the candidate's own class teacher whom they are familiar with. There was enough time for preparation which could be a cooperative effort with the teacher as well as classmates. The four-models also attempted to replicate the communicative features involved in speaking.

However, feedback during discussion suggests that implementation at the school level did not necessarily capture the communicative features built into the tasks. More often than not, texts were note 'prepared' taken off published materials, memorized and delivered as a one-way communication. Even pair-work dialogues or group discussion were scripted, memorized, rehearsed and performed. As such, it would be difficult to argue that the discourse was a sample of spoken English. The texts were unnaturally error-free and often delivered without understanding.



The assessment criteria did not appear to correspond with the demands of the task. In all the eight criteria listed, speaking was consistently mentioned as the criterion but in actual, there was very little speaking done by the candidates. Aspects of the performance recognize interaction which calls for unpredictability, reciprocity, turn-taking and adapting. Rehearsed scripts do not allow for these features. Based on the teachers' experience to date, opportunities for fielding and/ or handling questions were not created. From the language aspect in memorized scripts, the language used would not be candidates' own but that of the writer.

Another issue of concern was rater reliability. There had been no examiner training sessions and despite the appointment of the regional moderator, teachers' were not convinced about the reliability of marks awarded.