Use of item analysis to improve the quality of single best answer multiple choice question in summative assessment of undergraduate medical students in Malaysia

ABSTRACT

Background: Single best answer (SBA) as multiple-choice items are often advantageous to use for its reliability and validity. However, SBA requires good number of plausible distractors to achieve reliability. Apart from psychometric evaluation of assessment it is important to perform item analysis to improve quality of items by analysing difficulty index (DIF I), discrimination index (DI) and distractor efficiency (DE) based on number of non-functional distractors (NFD).

Objective: To evaluate quality of SBA items administered in professional examination to apply corrective measures determined by DIF I, DI and DE using students' assessment score.

Method: An evaluation of post summative assessment (professional examination) of SBA items as part of psychometric assessment is performed after 86 weeks of teaching in preclinical phase of MD program. Forty SBA items and 160 distractors inclusive of key were assessed using item analysis. Hundred and thirty six students' score of SBA was analysed for mean and standard deviation, DIF I, DI and DE using MS Excel 2007. Unpaired t-test was applied to determine DE in relation to DIF I and DI with level of significance. Item-total correlation (r) and internal consistency by Cronbach's alpha and parallel-form method was also computed.

Result: Fifteen items had DIF I = 0.31–0.61 and 25 items had DIF I (≤ 0.30 or ≥ 0.61). Twenty six items had DI = $0.15 - \geq 0.25$ compared to 14 items with DI (≤ 0.15). There were 26(65%) items with 1–3 NFD and 14 (35%) items without any NFD. Thirty nine (32.50%) distractors were with choice frequency = 0. Overall mean DE was 65.8% and NFD was 49 (40.5%). DE in relation to DIF I and DI were statistically significant with p = 0.010 and 0.020 respectively. Item-total correlation for most items was < 0.3. Internal consistency by Cronbach's alpha in SBA Test 1 and 2 was 0.51 and 0.41 respectively and constancy by parallel-form method was 0.57 between SBA Test 1 and 2.

Conclusion: The high frequency of difficult or easy items and moderate to poor discrimination suggest the need of items corrective measure. Increased number of NFD and low DE in this study indicates difficulty of teaching faculty in developing plausible distractors for SBA questions. This has been reflected in poor reliability established by alpha. Item analysis result emphasises the need of evaluation to provide feedback and to improve quality of SBA items in assessment.

Keyword: Single best items; Difficulty index; Discrimination index; Distractors efficiency; Functional distractors; Non-functional distractors