Composite reliability of a workplace-based assessment toolbox for postgraduate medical education

Reference:
Moonen-van Loon, JM¹, Overeem K², Donkers HH¹, van der Vleuten CP¹, Driessen EW¹. Composite reliability of a workplace-based assessment toolbox for postgraduate medical education. Advances in health Sciences Education. 2013; 18(5):1087-102.

Author Institutions:
¹Department of Educational Research and Development, Faculty of Health, Medicine, and Life Sciences, Maastricht University, Maastricht, The Netherlands
²IQ Healthcare, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands

Tags
Clinical domain: N/A
Educational domain: Assessment (Post)graduate (Residency training)

Background
With the recent move towards a competency-based approach in residency education there has been a concurrent move to competency-based assessment, and the development of work-place or work-based assessment (WBA) tools. Commonly used ones include the Mini-CEX and other forms of encounter cards, DOPS and other forms of assessment of procedural skills, and multi-source feedback (MSF). These may be compiled in a portfolio, which might also include opportunities for reflection. As individual tools, these have been shown to be reliable for formative assessment, but there are mixed results when these are used in summative assessment, particularly in high stakes contexts

Purpose
To calculate the separate reliabilities of three commonly used WBA tools: mini-CEX, DOPS, and MSF, and the composite reliability of these WBA tools when combined in a portfolio.

Type of paper
Research: reliability study
Key Points on the Methods

Context – all residents in the Netherlands use WBAs that are collected in an e-portfolio. Performance reviews and decisions on academic advancement are based on the results. Data can be sorted by CanMEDS competencies, across tools.

Data: Residents in 12 medical / surgical and lab specialties ‘made their anonymous WBA data available for analysis’. These residents all had Mini-CEX, DOPS and MSF data available, with comparable scales used across disciplines. Data collected over 3.5 years. Analysis used generalizability theory to account for different sources of variance, and measured reliability within each of the 3 tools and across tools in the portfolio.

Key Outcomes

Over 500 residents and nearly 13000 data points (number of WBA’s). To assess the reliability of individual WBAs, the minimum number of assessments needed for a G coefficient of at least 0.80 was nine DOPS, eight mini-CEX, and nine MSF rounds. For composite reliability (looking at the items in a portfolio) the reliability coefficient threshold of 0.80 was reached with a minimum of seven mini-CEX, eight DOPS, and one MSF round, while for the first year resident portfolio six DOPS, five mini-CEX, and one MSF round were required. Two MSF rounds per year for each combination of WBA numbers increased the reliability coefficient by about 5 % for the complete data set and by 3 % for the first year portfolio.

Key Conclusions

The authors conclude that the ‘results indicate that it is possible to make a summative decision using a combination of mini-CEX, DOPS and MSF. [In fact all they showed was that there was good reliability over a certain number, not that summative decisions can be made.] Reliable judgement across all years and for year 1 alone proved achievable with numbers of assessments per individual tool that not only appeared to be feasible for residency programmes but would also ease residents’ and assessors’ workload.’

The authors also state ‘this study adds to the literature on the reliability of WBA tools by showing that reliability gains can be achieved by moving beyond single WBA tools to a composite in a ‘WBA tool box’ in a portfolio.’

Spare Keys – other take home points for clinician educators

If individual tools are used, the number needed to reach a decent reliability is higher than if greater than one tool looked at together. This is important as if fewer WBAs are needed it will likely increase feasibility and acceptability and may decrease cost and workload.

Given the need to have reliable tools to make high stakes judgments, the results of this paper suggest that even WBA tools that are quite variable can provide good reliability.
Using a theoretical framework in med ed research is important, if not essential ... but particularly in a general journal it may be necessary to describe the theory and why it was used.