



Validation of the SYDney Shared and Evidence-based decision-Making test (SYSStEM) among healthcare professionals

Kat Leung ^{a,*}, Lyndal Trevena ^a, Christopher Roberts ^a, Donna Waters ^b

^a Sydney Medical School, Faculty of Medicine and Health, University of Sydney, Australia

^b Sydney Nursing School, University of Sydney, Australia

Received 19 January 2019; revised 29 April 2019; accepted 8 May 2019

Available online 15 May 2019

Abstract

Purpose: Whilst competence in evidence-based practice (EBP) is a critical skill for healthcare professionals, there is a gap in the literature in the assessment of EBP. Current frameworks in Evidence-Based Practice education cite a five-step process (Ask, Acquire, Appraise, Integrate and Evaluate) yet many existing measures of competence focus on the first three, relying on self-report and cognitive testing, falling short of assessing skills. The Fresno Test has been validated for assessing the first three steps and the “Assessing Communication about Patient Preferences” tool is reliable in capturing the quality of information given by physicians in integrating evidence. The Evidence-Based Practice Implementation Scale is useful for the ‘evaluate’ step in the EBP process. We adapted components from all three of these tools to create the SYDney Shared and Evidence-based decision-Making test (SYSStEM) in order to measure EBP and Shared Decision Making (SDM) competencies in the context of patient care. The aims of this study were to examine the internal structure, criterion and consequential validity of the SYSStEM.

Method: Data were available from 37 healthcare practitioners who completed the SYSStEM online. We conducted a confirmatory factor analysis to determine the theoretical structure and a modified Angoff method to determine a minimum competence level for newly graduated health professionals.

Results: The final 10-item SYSStEM consisted of two subscales: ‘EBP knowledge’ and ‘SDM skills’ demonstrating good construct validity and reliability ($\alpha > 0.74$). The factorial validity of SYSStEM was shown by all items loading significantly on their expected factors and a good fit to the data. Although the benchmark for passing the SYSStEM was low, only 13/37 practitioners (35%) reached the minimum standard for EBP competence.

Discussion: This study provides preliminary evidence for the validity the SYSStEM tool for measuring EBP competence amongst health graduates. Further large-scale validation of SYSStEM is required.

© 2019 King Saud bin Abdulaziz University for Health Sciences. Production and Hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Competence in evidence-based practice; Construct validity model; Evidence-based decision making; Modified Angoff method; Shared decision making

* Corresponding author. Sydney Medical School, Faculty of Medicine and Health, University of Sydney, C/O Room 321B, Edward Ford Building (A27), NSW 2006, Australia.

E-mail address: kleu9486@uni.sydney.edu.au (K. Leung).

Peer review under responsibility of AMEEMR: the Association for Medical Education in the Eastern Mediterranean Region

Introduction

Medical, nursing and allied health clinicians need to commit to lifelong continuing professional development and providing evidence of optimal patient care.¹

<https://doi.org/10.1016/j.hpe.2019.05.002>

2452-3011/© 2019 King Saud bin Abdulaziz University for Health Sciences. Production and Hosting by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Abbreviations

ACEPP	assessing communication about patient preferences
CFA	confirmatory factor analysis
CFI	comparative fit index
CI	confidence interval
CREATE	classification rubric for EBP assessment tools in education
CTT	classical test theory
EBP	evidence based practice
EBPIS	evidence based practice implementation scale
EFA	exploratory factor analysis
EQS	structural equation modelling software
ICC	intraclass correlation coefficient
LM	lagrange multiplier
PCA	principle factor analysis
RMSEA	root mean square error approximation
SBX ²	Satorra–Bentler scaled chi-square
SDM	shared decision making
SEM	standard error of measurement
SIS ₂	stability index for standard 2
SPSS	statistical package for social science
SRMR	standardised root mean residual
SYStEM	SYdney Shared and Evidence-based decision-Making test

Yet, there is a critical gap in the literature in validated assessments of the fundamental skill of practising in an evidence-based and person-centred manner. We aimed to address this gap by validating an assessment tool, which integrates all of the steps of Evidence Based Practice (EBP), and includes the principles of Shared Decision Making (SDM). This study provides preliminary validity evidence for the new tool, which we named the SYdney Shared and Evidence-based decision-Making test (SYStEM).

Evidence based practice (EBP) is a process that integrates the best available evidence, with clinical expertise, whilst taking account of the context, circumstances, preferences and values of a person with health concerns.² The implementation of EBP has traditionally involved a 5-step process: asking focussed questions (Step 1 – Ask), acquiring research evidence (Step 2 – Acquire), appraising the quality of evidence (Step 3 – Appraise), applying evidence in practice (Step 4 – Integrate) and assessing the outcome of care

(Step 5 – Evaluate). Traditional EBP curricula that focus on the traditional five-step approach do not necessarily encourage health practitioners to effectively involve patients in decision-making.^{1,3} For many health decisions, the integration of shared-decision making (SDM) into the EBP model facilitates communication between health professionals and patients enabling informed choices that integrate research evidence, patient preferences and the clinical context into decisions.^{4–8} SDM is therefore an important aspect of evidence application,^{6,7} providing a unique and distinct attribute to the EBP competence of health professionals. The integration of EBP with SDM is therefore important when developing a new tool for the assessment of practise in an evidence-based and person-centred manner.

Appraising existing evidence based practice tools

There are few tools that fully align with the current Classification Rubric for EBP Assessment Tools in Education (CREATE) framework.⁹ The authors of this framework have already highlighted the lack of tools assessing EBP steps 4 and 5 (Integrate and Evaluate), and that tools need to go beyond self-report and cognition testing to include the assessment of skills, behaviour and even patient outcomes, if possible.⁹

The 12-item Fresno test¹⁰ has been validated for measuring the first three steps of EBP (Ask, Acquire, Appraise). It consists of seven scenario-based open questions measuring the use of PICO algorithms (P – Patient, Population or Problem, I – Intervention, exposure or test, C – Comparison interventions or exposure; and O – Outcomes of clinical importance) to build answerable clinical questions, literature searching and elements for evidence appraisal. Five short questions also include a series of epidemiological calculations for the effect of different studies. Answers are assessed using a scoring rubric which has been tested with medical residents and reported to have good discriminant validity and scale reliability. However, the content validity was assessed by an unknown number of evidence-based medicine teachers without measures of its reliability.¹⁰ The internal structure and variances between items in the test were not examined and the ‘pass’ score was assumed to be the midpoint of the total test score based on opinions within the research team. Despite these limitations, the Fresno test has subsequently been modified into discipline-specific EBP versions^{11–13} although these still do not measure all five steps of EBP competence.

The Evidence Based Practice Implementation Scale (EBPIS)¹⁴ is designed to monitor clinician behaviour and to evaluate the outcome of evidence-based care (EBP Step 5 – Evaluate). The EBPIS is based on the Transtheoretical model of health behaviour change¹⁵ in which the implementation of EBP is defined as clinicians' engagement in behaviours that are associated with four categories of clinical activities: (1) seeks and appraises evidence, (2) shares evidence with colleagues/patients, (3) collects and evaluates evidence-based outcome data, and (4) uses evidence to change practice. The 18-item EBPIS has been tested with a large cohort of nurses and reported to have good construct and criterion validity.¹⁴

The Assessing Communication about Patient Preferences (ACEPP) tool¹⁶ was designed to capture the quality and amount of information that clinicians provide patients when making an informed choice between treatment options. The tool consists of three domains measuring the occurrence and quality of communication about treatment effects of interventions, patient preferences, and clinical and social circumstances. The ACEPP tool has been tested with a group of family physicians and demonstrated a good reliability in capturing the quality of information given by these physicians.¹⁶

A new model of assessment in EBP

Health clinicians need to deliver evidence-based person-centred care within the context of an interdisciplinary team.^{9,17} Interdisciplinary care relies on health professionals from different disciplines collaborating and sharing information, streamlining and standardising processes to achieving optimal care outcomes.^{18,19} The recent development of consensus-based competencies in EBP for health professionals by Albarqouni et al. (2018) has further supported the education and assessment of generic EBP skills across disciplines.²⁰ In the development of SYStEM we therefore assessed generic skills for the first two steps of EBP and used a common health scenario that would be familiar to a range of health professionals and a pre-appraised evidence summary for testing the other three steps – appraise, integrate and evaluate. Since we found no existing tools that measured all five steps of EBP competence, our new competency-based assessment was validated using a contemporary construct validity framework.²¹

Validity theory is described as a unitary concept where multiple sources of evidence are required to support the interpretation of test scores for an intended measure. The construct validity model encompasses

five major sources of validity evidence (content, response process, internal structure, relationship to other variables, and consequences).^{21–23} This paper investigates the last three sources of validity evidence to support the construct validity of the SYStEM. We have previously reported on the content validity and response process of this tool²⁴ and briefly summarise this below under tool development.

This study aims to provide further evidence of the validity of SYStEM in the context of health professional practice by 1) examining its internal structure through assessing factorial validity and internal consistency, (2) demonstrating criterion validity to the degree of prior EBP education, and 3) applying a consensus about the level of competence for novice health graduates through the use of a modified Angoff standard setting procedure.

Method

Participants

The setting for this study was a large research intensive University. Recruitment of volunteer participants into the pilot study was via a web-based questionnaire survey emailed through the administration office of professional organisations and post-graduate medical, public health and nursing schools at a single research intensive University between March and July, 2015. A convenience sample of 93 registered and practising nurses, doctors and physiotherapists, with or without previous EBP training were eligible to complete the online survey. The questionnaire consisted of two main sections: (1) questions about characteristics of participants including exposure to prior EBP education and (2) the SYStEM. Approval was obtained from the Human Research Ethics Committee at the University of Sydney in Australia [Approval numbers: 2013/320 (Stage1); 2014/907 (Stage 2–3)].

Materials

The SYStEM ([Appendix A](#)) was specifically designed to assess two domains of the underlying construct of interest that underpins EBP competence of health professionals: EBP knowledge (Ask, Acquire, Appraise) and; SDM skills (Integrate) and Implementation (Evaluation). We adapted and combined three assessment tools – the Fresno test, ACEPP tool and EBPIS to support the measurement of five EBP steps through the two domains of EBP competence. The SYStEM contains five open questions with

subsequent closed-ended questions covering the five steps of the EBP process. We adapted Questions 1 and 2 from the Fresno test¹⁰ for measuring EBP steps 1 and 2 (Ask, Acquire) in the new tool. Question 1 measures the skills of participants' in writing a focused clinical question (Step 1) and asks "Please write a focused clinical question for this clinical scenario to help you organise a search of literature for an answer." This is followed by two close-ended questions about the study type and design, one of which asks "What type of study design would best answer your clinical question?" Question 2 tests the participants' skills for searching evidence sources (Step 2) while Question 3 measures their knowledge in understanding appraised evidence in the form of a Summary of Findings table (Step 3). Question 4 measures the application of integrated research evidence through meaningful clinical communication (Step 4) whereas Question 5 evaluates the quality of care for the clinical problem (Step 5). Participants answered questions on the SYStEM related to a clinical scenario for which an evidence summary was provided. A scenario about a 2-year old boy, named George with acute otitis media and the relevant Cochrane review²⁵ were used.

A marking rubric (Appendix B) using a similar scoring system to the Fresno test was developed by the research team for scoring both narrative and closed-ended questions. The rubric specifies explicit criteria for each item, allowing markers to quantify narrative answers into scores. For example, Question 1a asks the participant to write a focused clinical question with PICO (Population, Intervention, Comparison, and Outcome) components. Participants are scored based on their inclusion of each PICO component against specific descriptions required for the clinical question. The rows of the rubric represent four different categories and each of which is associated with a score. The SYStEM has 14 items covering five steps of EBP and the range of scores for each EBP step is plausibly weighted. The total score for EBP step 1 (Questions 1a–c) is 16, step 2 (Questions 2a–b) is 48, step 3 (Questions 3a–f) is 16, step 4 (Questions 4a–b) is 18; and step 5 (Question 5) is 15. Thus the possible maximum score for SYStEM is 113. For further details of each item score and marking criteria please refer to the rubric in Appendix B. The score for demonstrating a minimum level in mastery of EBP competence was set by the standard setting procedure, which we describe later.

Selected items from the ACEPP tool¹⁶ were used to systematically measure the coverage of meaningful clinical communication that supports a shared decision making process between the health professional and

client in Questions 4a (What is the evidence about benefits and harms of antibiotics that George's mother needs to weigh up?) and 4b (What preferences and values do you already know from George's mother? Can you suggest others that might be explored with her?). These two questions focus on the measurement of participant communication skills in explaining treatment options and helping clients explore preferences and make decisions as part of the process in evidence application (Step 4 – Integrate). We also adapted items from the EBPIs¹⁴ to assess answers for evaluation of quality of care in Question 5.

Procedures

Seven experts were purposively selected for a marking panel²⁶ and included six academics from three disciplines (Medicine, Nursing, Health Science) at two Australian universities plus a clinical nurse specialist. Markers were sent an email with three completed SYStEMs to practise on marking then two weeks later attended a 2-hour training session. Markers were asked to score each question against the rubric to represent participants' level of competence on individual question. Results of the marking exercises were discussed during the training session. Markers were also given an opportunity to raise questions about the marking criteria and scoring mechanism, to understand more about the similarities and different expectations between health disciplines, as well as to gain insight about their own potential subjectivity in marking (leniency or stringency). In this way, markers were 'calibrated' towards similar and realistic expectations of actual performance on the SYStEM.²⁷ Based on the markers' feedback, the marking rubric and weighing of scores were then refined before the standard setting session (which we explain in the next section) and marking commenced. Completed SYStEMs were randomly distributed amongst the marking panel and each was marked 'blind' and independently by two panel members. Markers were given four weeks to complete marking in their own time.

Analysis

Content validity evidence and response process

Content validity for the SYStEM has been reported in detail elsewhere²⁴ but is summarised here. The SYStEM was evaluated against the EBP competency framework by a panel of 42 content experts.²⁴ Content experts' agreement on the representativeness (ICC and Cronbach's α : 0.77, $CI_{95} = 0.64–0.86$) and clarity

(ICC and Cronbach's α : 0.67, $CI_{95} = 0.50\text{--}0.80$) of items was good to substantial. Almost all experts ($n = 40, 95.4\%$) agreed with the comprehensiveness of construct items in the SYStEM. Response process for the SYStEM was determined as good.²⁴ We further used these performance criteria within the framework to devise the prototype SYStEM questions.

Validation of factorial validity and internal structure

For this study, we had no empirical data to support the combination of selected and adapted items from three different assessment tools into SYStEM would measure intended theoretical constructs (factors). Therefore, an exploratory factor analysis (EFA) was required to investigate to what extent the observed variables were linked to their underlying factors. We then conducted a confirmatory factor analysis (CFA)^{28,29} to determine the theoretical structure of SYStEM and a standard setting procedure to determine a minimum standard for EBP competence.

An EFA was performed in which the normality of 14 observed variables (items) was assessed for potential uni- and multivariate outliers on the predictors and their suitability for factor analysis. A Principle Component Analysis (PCA) based method of extraction for EFA, direct oblimin rotation, was used to explore the underlying constructs of the SYStEM, assuming a correlated relationship among items.²⁸ The pattern and structure of the two components/factors (EBP knowledge, SDM skills) indicated the best fit for the data. Therefore, the two-factor model was considered for testing the internal structure of the SYStEM.

To further examine the factor structure and dimensionality of the SYStEM, we conducted a CFA. The maximum likelihood estimation with Satorra-Bentler scaled chi-square (SBX^2) was used to compute standard errors adjusted to the extent of non-normality.^{30,31} We selected criteria for model fit indices that best estimated for small samples and non-normality,³² they were: a) the robust comparative fit index (CFI) > 0.95 ; b) the SBX^2 divided by degrees of freedom (df) < 2 and the p value should be insignificant; c) the root mean square error approximation (RMSEA) < 0.05 ; and the standardised root mean residual (SRMR) < 0.08 . Poor fitting models were re-specified if theoretically justifiable and also guided by examination of factor loadings and standardised residual covariance matrices. We employed the Lagrange Multiplier (LM) test to examine the usefulness of adding covariance to the model and if necessary the Wald test to examine item(s) that could be left out the model.³³

Relationship with other variables

Analysis of the relationship of test scores to relevant variables external to the test provides an important source of evidence about the degree to which these test-criterion relationships are consistent with the construct underlying the proposed test score interpretation.²³ Thus, the correlation of total and subscale variables to exposure of EBP education was examined using a second order analysis to the final CFA model, which also served as a cross-validation of the determined factor structure of the SYStEM.³⁴

Consequential validity through the standard setting procedures

Consequential validity describes the after-effects and possible positive or negative consequences of a particular assessment, test or measure. It can help identify tests that are not measuring constructs they are supposed to be measuring or that it is falsely measuring those actually taking it. Using the Kane's framework³⁵ with Messick's construct and consequence-focused conception of validity,³⁶ we systematically collected sources of consequential validity evidence through the standard setting procedures.

Standard setting is a process used to create boundaries between categories such as pass/fail for an assessment, which has consequences for those deemed to be lacking in acceptable competence. We employed the modified Angoff method³⁷ to determine the cut score on the SYStEM that separates the non-competent (Fail) from the competent (Pass) participants. The judges' decision on the EBP competence of health professionals was based on the participants' ability in using evidence to resolve a common health issue (otitis) in the community that would be understood by health professionals with and without medical training. The standard setting session was conducted four weeks after the marker training session, before marking commenced to avoid pre-knowledge and preconception of participant performance on the test. Panellists were briefed about the use of all documents (e.g. rating forms) needed for the standard setting procedure and to keep in mind the performance levels of a hypothesised borderline (minimally acceptable) person throughout the scoring process. All judges then independently determined the minimum score (cut score) required for a beginning practitioner to pass each SYStEM item in the first round. A 'beginner' health professional is expected to provide evidence-based health service/advice safely and independently where compensatory standards such as additional training or supervision are not assumed. We allowed two rounds of rating so all

judges were offered an opportunity to adjust their ratings after reviewing discrepancies from the first round; only if judges chose to do so. The final determination of the cut score was made by averaging estimates for each test item.

Statistical tests

The IBM Statistical Package for Social Sciences (SPSS) Version 23.0³⁸ was used for analysis. Descriptive statistics such as means, standard deviations and percentages were used to explore characteristics of participants; as well as the pass–fail test score for each construct item, scale and subscales. EFA and CFA on the SYStEM were computed using the Structural Equation Modelling Software (EQS) Version 6.3.³⁹ The internal consistency for the total and each of the two subscales was examined with the use of Cronbach's α for which ≥ 0.70 was considered acceptable.⁴⁰ The internal validity of the standard setting exercise was evaluated by computing the agreement among judges in determination of the pass score and the Meskuskas Stability Index for Standard 2 (SIS₂).²⁷ The Meskuskas SIS₂ index is the ratio of the standard error of measurement of scores to the standard deviation of the judges' pass scores. The lowest acceptable value is 0.5, above 0.9 or more indicates a high stability (quality) of the standard setting method.^{41,42} Reliability coefficients of the total and two subscales were used to estimate the standard error of measurement (SEM) of cut scores, from which confidence intervals (CI) around pass scores were calculated.⁴³ The Inter-rater reliability among the SYStEM markers was calculated using the Intraclass Correlation Coefficient (ICC) where the consistency of judges' scores for each item was evaluated based on the range of each item score in the marking rubric. Values of 0.7 were interpreted as acceptable and moderate to excellent if ≥ 0.8 .⁴³

Results

A convenience sample of 93 health professionals expressed interest in participating in our study. Sixty (64.5%) of these (54 nurses, 5 doctors and 1 physiotherapist) commenced the SYStEM tool with thirty-seven (61.7%) (33 nurses, 3 doctors, 1 physiotherapist) full completions and 23 (38.3%) (21 nurses and 2 doctors) partial completions. The incomplete responders dropped out after EBP steps 1–2 without stating reasons for their withdrawal (although the acceptability to the SYStEM was substantial (ICC = 0.88, CI₉₅:0.81–0.93) in the pilot study.⁴⁴ Of the 33 nurses who fully completed

the SYStEM, 20 were practising nurses and 13 were educators/academics. Most participants completed the SYStEM within 60 minutes and the mean time for completion was 33 minutes.

Factorial validity and internal consistency

Exploratory factor analysis (EFA)

Thirty-seven completed SYStEMs were analysed. The 14-item dataset was first assessed for its normality and suitability for factor analysis. Items scores were proportionally adjusted to equal weight however, several variables such as Questions 3c (Probability of outcome) and 3f (Quantify harm) had a standardised skewness greater than 3.75. Correction for non-normality only slightly improved variables with severe univariate skewness and kurtosis. However, no multivariate outlier or multi-collinearity was detected through the examination of Mahalanobis distance. The Mardia-based kappa was 0.03, indicating no evidence of significant multivariate kurtosis. Thus, the outliers were included in order to maintain the stability of scale analysis and minimise the penalty for chi-square and other model fit indices.^{45,46}

The Bartlett's test of sphericity ($X^2 = 321.62$, $df = 91$) was highly significant at $P = 0.000$ and the Kaiser–Meyer–Olkin value was 0.68, which supported the factorability of the correlation matrix.²⁸ PCA revealed the presence of two components with eigenvalues exceeding 1, explaining 40.6% and 11.9% of the total variance that exceeded the criterion value of 95th percentile obtained from Parallel Analysis.⁴⁷ Inspection of the scree-plot showed a clear break after the second component, confirming the support for a two factor solution. Following Direct Oblimin rotation, the two factors showed an adequate inter-correlation between items ($r = 0.36$). The pattern matrix showed a relatively clear two-factor solution with all item loadings exceeding 0.4. Seven items loaded on the first component covered fundamental EBP concepts such as PICO (item 1a), research types and designs (items 1b–c), literature searching (items 2a–b), appraisal of treatment effect (item 3a) and evaluation of quality of care (item 5), we therefore named this a 'EBP knowledge' subscale. The second component loaded seven items relating to evaluations of treatment benefit and harms (items 3b–f), explanation of evidence (item 4a) and making decisions based on the clients' preferences and values (item 4b); which involves the integration of knowledge in evidence appraisals and effective shared decision making skills thus we named this subscale as 'SDM skills'.

Confirmatory factor analysis (CFA)

The CFA of a 14-item, two-factor (latent variable) model identified in the EFA was examined: (1) EBP knowledge (Q1a–c, 2a–b, 3a & 5) and (2) SDM skills (Q3b–f, 4a–b). Factor loadings in this hypothesised model were statistically significant. Three items (q3a, q3d, q3f) had loadings >0.8 and one item (q3e) only loaded 0.29 to the ‘SDM skills’ factor. The SBX² test was significant (CFA 1 in Table 1: SBX² = 177.37, Robust CFI = 0.59, RMSEA = 0.20), indicating a poor-fitting model. Post hoc model re-specification was performed in order to develop a better fitting model. Although the Wald test suggested that the exclusion of two items 3d (Quantify treatment benefit) and 3f (Quantify treatment harm) would significantly improve the model fitness, it was theoretically unjustified thus these two items were retained. Instead, two items 3e (Probability of treatment harm) and item 3c (statistical significance of intervention) were removed from the hypothesised model due to a low parameter estimate of 0.29, high residual covariance (0.96), and cross-loadings indicated in EFA. As a result, the model fit improved substantially but still lower than the expected criteria (SBX² = 90.40, Robust CFI = 0.81, RMSEA = 0.15). All items had loadings 0.4–0.8 except for two items (3d, 3f) loaded on the ‘SDM skills’ factor. These two items had loadings >0.9, indicating a systematic measurement error in item responses.²⁸

Inspection of the correlated error covariance items 3d and 3f (E9 and E11) through the LM test suggested that the SBX² would drop approximately 17.19 points. This is theoretically justified to account for the unique relationship between quantifying benefits and harms of treatment, which is conceptually similar but attribute different meanings to the ‘SDM skills’ factor. Therefore a covariance was added and the model re-estimated, confirming a strong correlated residuals ($r = 0.90$) between these two items. However, the model fit was slightly improved (SBX² = 77.6, Robust CFI = 0.87, RMSEA = 0.12) but parameter estimates

of these two items were good now (item 3e: 0.67, item 3f: 0.63). Factor loadings for other items were above 0.5 except for item q1b (0.38) and q1c (0.44) both indicated little relationship with the latent variable ‘EBP knowledge’. Therefore, these two items were dropped leading up to the final model with good fit indices (CFA 2 in Table 1: SBX² = 29.55, Robust CFI = 1.00, RMSEA = 0.00). The loading factor for each item exceeded 0.5 except for item PICO (0.48), nevertheless the factor correlation was 0.78 indicating a possible over-factoring.³³ Therefore, we compared fit statistics of the final model with the 10-item one factor model (CFA 3 in Table 1: SBX² = 40.90, Robust CFI = 0.95, RMSEA = 0.08), confirming the best model fit for this data was a two-correlated model representing ‘EBP knowledge’ and ‘SDM skills’ (Tables 1 and 2).

Results of the second order CFA also supported the hypothesis that the measurement of EBP competence may include constructs for ‘EBP knowledge’ (0.96) and ‘SDM skills’ (0.78). Fit statistics indicated a good model fit (CFA 4 in Table 1: SBX² = 47.92, Robust CFI = 0.97, RMSEA = 0.05), including an association with EBP education/training ($r = 0.29$) (Fig. 1). Cronbach’s alpha showed that the reliability coefficient for the ‘EBP knowledge’ and ‘SDM skills’ subscales were 0.67 and 0.74 respectively, meaning that each subscale had a reasonably good internal consistency.

Standard setting procedures

Through the CFA, four items were dropped (items 1b, 1c, 3c and 3e) from the original 14-item SYSTEM due to unfit statistics. Therefore, the final version of the SYSTEM is a 10-item test, which consists of two 5-item subscales with a total score of 105 (Table 3). The ‘EBP knowledge’ subscale includes five items related to EBP steps 1–3, and step 5 (PICO question, search for evidence, appraisal of treatment effect, evaluate quality of care) whereas another five items in the ‘SDM skills’ subscale are indicators for EBP steps

Table 1
Model fit indices of four CFA SYSTEM models.

Model number and features	SBX ²	df	P-value	Robust CFI	RMSEA	SRMR
CFA 1 (hypothesised model)						
14-item, 2-factor	177.37	74	0.000	0.591	0.197	0.145
CFA 2 (final model)						
10-item, 2-factor with one pair of error covariance	29.55	31	0.541	1.000	0.000	0.079
CFA 3						
10-item, 1-factor with one pair of error covariance	40.90	33	0.162	0.949	0.082	0.092
CFA 4 (2nd order CFA)						
10-item, 2-factor with EBP education to EBP competence	47.92	42	0.245	0.971	0.053	0.083

Table 2
Pattern matrix for EFA of the final 10-item SYStEM.^a

Item no	SYSTEM item	EBP knowledge	SDM skills
1a	PICO Please write a focused clinical question for this clinical scenario to help you organise a search of literature for an answer	0.72	–
2a	Search information Where might you find answers to this clinical question? Name all possible sources of information you could think of; then briefly describe the advantages and disadvantages for each information source you list	0.56	–
2b	Literature search If you were to search for evidence using an electronic database, describe what your search strategy would be. Be specific about all search terms and search fields (categories) you would use. Describe how you might limit your search and explain rationale	0.47	–
3a	Treatment outcome What are the main outcomes in this table that described the potential benefit(s) of antibiotics compared with placebo?	0.51	0.46
5	Assess quality of care Can you suggest an activity (or more than one) which would help you assess the quality of care for this clinical problem?	0.52	–
3b	Probability of benefit Which (if any) of these are statistically significant?	–	0.80
3d	Quantify benefit How many children will benefit from pain reduction after 4–7 days of antibiotics compared to placebo? How did you work this out?	–	0.63
3f	Quantify harm How many children will have vomiting, diarrhoea, rash on antibiotics? How did you work this out?	–	0.59
4a	Explain evidence Can you use plain or lay language to explain the above evidence to George's mother? (i.e. What is the evidence about benefits and harms of antibiotics that she needs to weigh up?)	–	0.86
4b	Preference and value Based on the original clinical scenario, what preferences and values do you already know from George's mother? Can you suggest others that might be explored with her?	–	0.45

^a Bolded values indicate major loadings for each item (>0.40).

3 and 4 (appraise benefit and harm of treatment options, explain choices and identify values).

Table 3 summarises the pass scores, markers' agreement and participant performance on the SYStEM by items and subscales. The pass score for the 10-item SYStEM was 41. Based on the Classical Test Theory (CTT),⁴³ the true score could only be 31.9 after subtracting the value of SEM (9.0) and the pass score was somewhere between 23.16 and 58.84. All judges expected a minimum of 39% SYStEM items be answered correctly in this cohort participant but only 13 (35%) (12 nurses, 1 doctor) passed this benchmark, resulting in a 65% failing rate on the SYStEM. The pass score for the EBP knowledge subscale (item 1–5) was 31.93. Fourteen (38%) participants (13 nurses, 1 doctor) were deemed competent while 23 (62%) were not. The pass score for the SDM skills

subscale (item 6–10) was 9.08 so only nine (24%) participants (7 nurses, 2 doctors) claimed to have competence in SDM skills while 28 (76%) did not. Surprisingly in the nursing group, the passing rate of clinicians was three times higher than educators/academics. The inter-rater reliability among markers was substantial for all items and subscales scores (Table 3), which also reflects the reliability of the scoring rubric. The quality of judges' consensus (ICC = 0.92, CI₉₅:0.86–0.98) and the stability of the standard setting exercise were substantial (Meskauskas SIS₂ index: 1.28).⁴¹

Discussion

This study provides preliminary evidence to support the psychometric characteristics of the SYStEM by

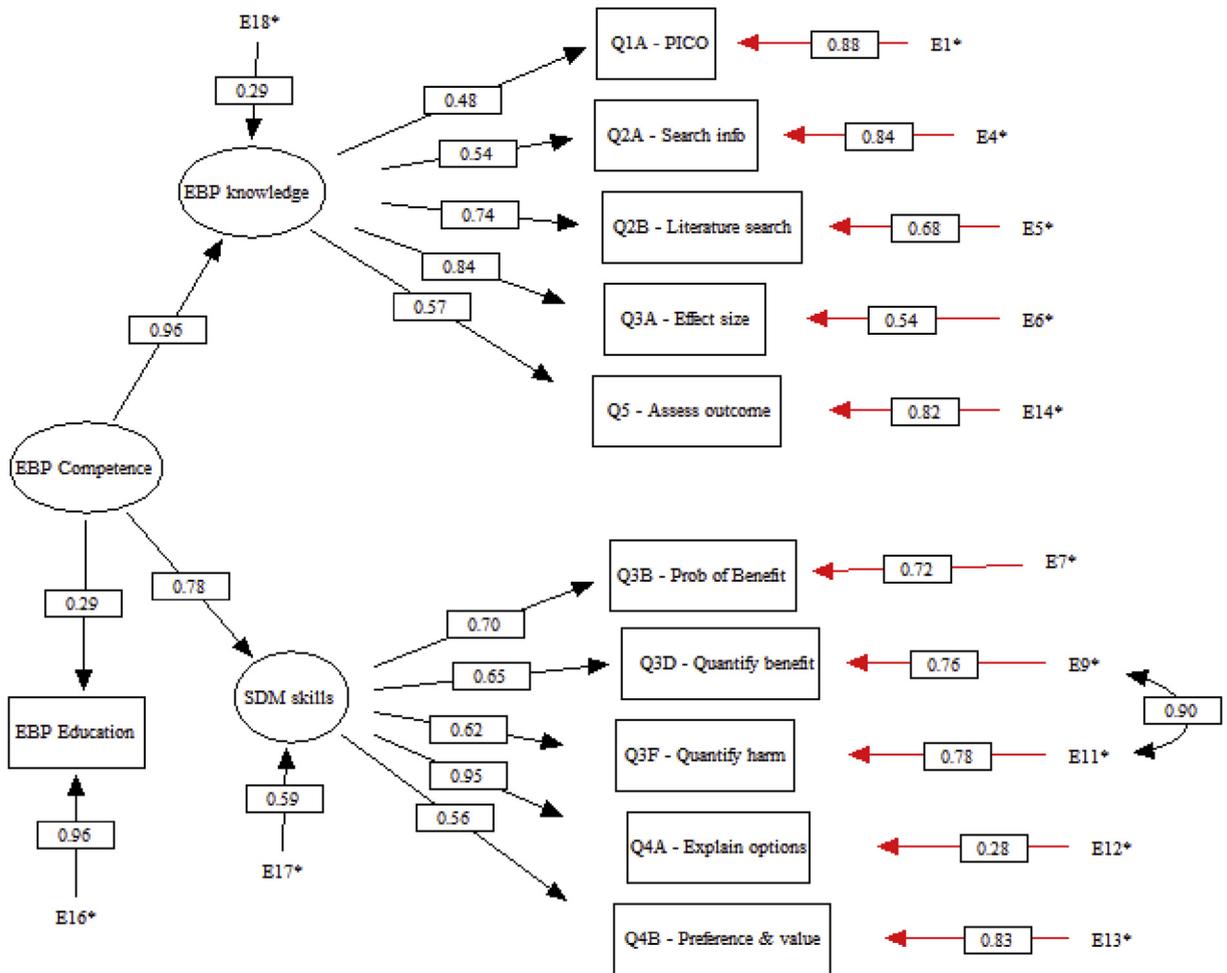


Fig. 1. Second-order confirmatory factor analysis model of measurement in EBP competence (SYStEM) showing standardised factor loadings and its relationship with EBP education.

demonstrating its internal structure through assessing its factorial validity and internal consistency, and providing consequential validity evidence through a standard setting procedure. The final 10-item SYStEM was found to have good internal consistency ($\alpha = 0.74$) with all items loaded significantly on their expected factors and a good fit to the data. It consisted of two subscales which both had good internal consistency and strongly correlated with each other: ‘EBP knowledge ($\alpha = 0.67$; $r = 0.96$)’ and ‘SDM skills ($\alpha = 0.74$; $r = 0.78$)’. The criterion validity of SYStEM was demonstrated by the positive correlation ($r = 0.29$) between test scores and participants’ exposure to post-graduate EBP education. Our internal consistency was slightly lower than the 12-item Fresno test ($\alpha = 0.88$) but the latter only measures the first three steps of EBP and cannot be used as a direct comparison.¹⁰ The

consequential validity was demonstrated through a modified Angoff standard setting exercise and although the benchmark for passing the SYStEM was low, only 13/37 practitioners (35%) reached the minimum standard for EBP competence. We note that the overall poor performance of participants in our study was similar to the passing rate for EBP knowledge (Steps 1–3) in two modified Fresno test studies with novice physiotherapists¹³ and occupational therapists.¹² Approximately 50% participants in our study were able to correctly answer questions for EBP steps 1–2 (Q1a–2b). Nevertheless, performance on the evidence appraisal questions (Q3a–f) was poor (11–27% pass rate), limited ability to explain evidence correctly in Q4a (16% pass rate), consideration of preferences and values in Q4b (32% pass rate); and to our surprise, a low ability to suggest clinical activity that enables the

Table 3
Summary of scale reliability, pass score, marker consensus and participant performance on the SYStEM.

Item no	SYStEM Subscale (scale reliability)	Question – theme	Full score	Pass Score (95% CI)	Marker agreement ICC (95% CI) ^c	Total participants (n = 37) (33 nurses, 3 doctors, 1 physiotherapist)			
						Mean	SD	Pass n (%)	Fail n (%)
1	EBP	1a – PICO	12	5.50	0.77 (0.40–0.86)	5.70	2.89	19 (51.4)	18 (48.6)
2	Knowledge (Cronbach's α = 0.66)	2a – info source	24	9.17	0.84 (0.59–0.93)	8.39	5.90	16 (43.2)	21 (56.8)
3		2b – search evidence	24	9.17	0.84 (0.58–0.94)	8.95	6.66	20 (54.1)	17 (45.9)
4	Subscale total	3a – effect size	6	2.17	0.73 (0.44–0.85)	1.39	1.23	8 (21.6)	29 (78.4)
5		5 – evaluate care	15	5.92	0.79 (0.49–0.92)	2.63	2.40	5 (13.5)	32 (86.5)
6	SDM skills (Cronbach's α = 0.74)	3b – prob of benefit	81	31.93 ^{a,b}	(15.83–48.03)	27.06	14.09	14 (37.8)	23 (62.2)
7		3d – quantify benefit	2	1.33	0.86 (0.66–0.94)	0.34	0.61	4 (10.8)	33 (89.2)
8	Subscale total	3f – quantify harm	2	0	0.80 (0.47–0.92)	0.34	0.69	10 (27.0)	27 (73.0)
9		4a – explains evidence	12	4.92	0.81 (0.54–0.92)	0.37	0.71	9 (24.3)	28 (75.7)
10	Subscale total	4b – preference, value	6	2.83	0.98 (0.96–0.99)	2.87	2.75	6 (16.2)	31 (83.8)
1–10		Scale total (Cronbach's α = 0.74)	24	9.08 ^{a,b}	(6.49–11.67)	1.93	1.71	12 (32.4)	25 (67.6)
		105	41 ^{a,b}	(23.16–58.84)	0.94 (0.86–0.98)	32.91	5.08	9 (24.3)	28 (75.7)
							17.83	13 (35.1)	24 (64.9)

^a Pass score percent for EBP knowledge subscale = 39.42%, SDM skills subscale = 37.83%, total score = 39%.

^b Standard error of measurement (pass score) for EBP knowledge = 8.2, SDM skills = 1.32, total score = 9.1.

^c Statistically significant at P < 0.005.

measurement for EBP implementation in Q5 (EBP step 5) – 13% pass rate. These results may indicate the priority for future planning of EBP curricula for health science programs.

However, our use of the modified Angoff method to determine the pass scores for each item, adding to the validity of the SYStEM by giving meaning to the observed test scores.^{27,37,48} It is interesting to see that all item pass scores were lower than 50% of item full scores (except for Q3b). We do not know whether this relates to the high expectations of judges, or that the test was too difficult for a newly graduated health professional.

Implications

The SYStEM is an improvement over existing EBP tests with good validity and reliability established through EBP academics, clinicians and researchers from a range of health disciplines. A scenario about a common childhood illness and the relevant pre-appraised evidence were used in this pilot to substantially minimise the need for epidemiological calculations. The SYStEM is comprehensive in the five EBP steps and is inclusive of shared-decision making thus it can also be used to assess the change in capability of health graduates as part of evaluations for interventional EBP education. Further testing of the SYStEM across different health disciplines would provide valuable information on the validity of the SYStEM in graduates, and investigate whether there are discipline specific differences to be addressed through education. The potential utility of the SYStEM is not limited to the assessment of competence in any specific discipline of health practice.

Strengths and limitations

This is the first instrument to specifically aimed at addressing the contemporary need for the assessment of EBP practice inclusive of shared decision making in healthcare professionals. We believe this study adds meaningful evidence to support the proposed interpretations and uses of the assessment scores.³⁵ This includes a process for the development of an acceptable and defensible pass score for the SYStEM, providing further meaningful interpretations of the test results. We acknowledge that researchers may have some concerns as to whether our best estimates of this pilot study serve as a proof of principal before going on to larger studies. These relate to the sample size and heterogeneity, as well as unsupervised test completion.

We acknowledge that in this naturalistic pilot study, the strength of the validity evidence is compromised by the small and uneven convenience sample in terms of the mix of health disciplines, and the possibility of variations in EBP training. All of these variables may translate to levels of difficulty with SYStEM items and contribute to the overall low passing rate on the SYStEM. In terms of the data analysis methods, we acknowledge that there was unequal weighing of rubric scores, and only one small sample was used for EFA, CFA and model modifications. With the use of CTT, we cannot precisely interpret observed item scores due to the high SEM and resulting wide confidence intervals for the EBP knowledge subscale and full scale, which may be the consequence of high differential values in this small sample.

Since the SYStEM was developed for measuring the generic EBP competence of medical, nursing and allied health professionals, we combined the data of all participants from different health disciplines for CFA. Unfortunately, the majority of participants who completed the test were nurses and only a few doctors and physiotherapist. Due to this small and heterogeneous sample, we were unable to conduct further CFAs for each health discipline subgroup to affirm the same model fits. However, we strategically overcome these problems by using the polychoric correlation matrices to compute EFA, using the Satorra-Bentler chi-square test to perform CFA as it is considered as robust in dealing issues with univariate non-normality³⁴; also a second order CFA was used to cross-validate the determined factor structure of SYStEM questions, this may not reflect the context of typical work based type assessments. We also acknowledge that unsupervised online testing may allow some participants to access external resources for answers, which may have inflated estimates of competence.

Conclusion

This study offers further validation evidence of the SYStEM and provides health care education researchers with a new instrument with which to measure the EBP competence of health graduates across all five steps of EBP model, inclusive of shared decision making. The two domains of attributes underlying this scale, EBP knowledge and SDM skills have a good fit to the data. EBP knowledge is related to the understanding of PICO, literature searching and appraisal of treatment outcome. SDM skills are related to the evaluation of treatment benefit and harms that best suits the preferences and values of health consumers

through a person-centred care approach. The SYStEM is therefore a contemporary instrument, which meets the requirements for measuring the enhanced EBP competence of health graduates in modern healthcare. This study has provided valid evidence to support the psychometric characteristics of the SYStEM and forms the basis for further research in this field. Further large-scale validation evidence of the SYStEM is required.

Ethical approval

Ethical approval has been granted from the Human Research Ethics Committee at the University of Sydney in Australia [Approval numbers: 2013/320 (Stage1), 8 May 2013 and 2014/907 (Stage 2–3), 17 November 2014].

Funding

None.

Declaration of interests

The authors have no competing interests to declare.

Other disclosure

None.

Acknowledgements

The authors wish to sincerely acknowledge the work of all survey participants who spent their precious time and effort in completing the questionnaire for this study. We would also like to express our sincere gratitude to the panel markers especially for Dr. Henry Ko, Research Fellow at the NHMRC Clinical Trials Centre, Sydney Medical School; and Ms. Zoe Cheng, GP Liaison Nurse, Clinical Nurse Specialist in Child and Family Health, Western Sydney Local Health District; for their assistance with marking and participation with the standard setting procedures.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.hpe.2019.05.002>.

References

1. Dawes M, Summerskill W, Glasziou P, et al. Sicily statement on evidence-based practice. *BMC Med Educ.* 2005;5(1):1–7.

2. Straus SE, Glasziou P, Richardson WS, Haynes RB. *Evidence-based Medicine: How to Practice and Teach It*. 4th ed. Edinburgh: Church Livingstone Elsevier; 2011.
3. Grad R. Teaching evidence-based medical care: description and evaluation. *Fam Med*. 2001;33:602–606.
4. Trevena L, Barratt A. Integrated decision making: definitions for a new discipline. *Patient Educ Couns*. 2003;50(3):265–268.
5. Hoffmann T, Legare F, Simmons MB, et al. Shared decision making: what do clinicians need to know and why should they bother? *MJA*. 2014;201(1):35–39.
6. Barratt A. Evidence based medicine and shared decision making: the challenge of getting both evidence and preferences into health care. *Patient Educ Couns*. 2008;73:407–412.
7. Elwyn G, Frosch D, Thomson R, et al. Shared decision making: a model for clinical practice. *J Gen Intern Med*. 2012;27(10):1361–1367.
8. Hoffmann T, Bennett S, Mar CD. *Evidence-based Practice Across the Health Professions*. 2nd ed. Chastwood Australia: Churchill Livingstone, Elsevier; 2013.
9. Tilson JK, Kaplan SL, Harris JL, et al. Sicily statement on classification and development of evidence-based practice learning assessment tools. *BMC Med Educ*. 2011;11(78):1–10.
10. Ramos KD, Schafer S, Tracz SM. Validation of the Fresno test of competence in evidence based medicine. *Br Med J*. 2003;326(7384):319 (313).
11. Lewis LK, Williams MT, Olds TS. Development and psychometric testing of an instrument to evaluate cognitive skills of evidence based practice in student health professionals. *BMC Med Educ*. 2011;11(77):1–11.
12. McCluskey A, Bishop B. The adapted Fresno Test of competence in evidence-based practice. *J Contin Educ Health Prof*. 2009;29(2):119–126.
13. Tilson JK. Validation of the modified Fresno test: assessing physical therapists' evidence based practice knowledge and skills. *BioMed Cent Med Educ*. 2010;10:38.
14. Melnyk BM, Fineout-Overholt E, Mays MZ. The evidence-based practice beliefs and implementation scales: psychometric properties of two new instruments. *Worldviews Evid Based Nurs*. 2008;208–216. fourth quarter.
15. Prochaska JO, Velicer WF. The transtheoretical model of health behavior change. *Am J Health Promot*. 1997;12(1):38–48.
16. Shepherd HL, Barratt A, Trevena LJ, et al. Three questions that patients can ask to improve the quality of information physicians give about treatment options: a cross-over trial. *Patient Educ Couns*. 2011;84:379–385.
17. Institute of Medicine (US) Committee on the Health Professions Education Summit. The core competencies needed for health care professionals. In: Greiner ACKE, ed. *Health Professions Education: a Bridge to Quality*. Washington, DC: National Academies Press; 2003.
18. Nancarrow S, Booth A, Ariss S, Smith T, Enderby P, Roots A. Ten principles of good interdisciplinary team work. *Hum Resour Health*. 2013;11(1):1–11.
19. Department of Human Services. *Health Independence Programs Guidelines*. Melbourne: State Government; 2008.
20. Albarqouni L, Hoffmann T, Straus S, et al. Core competencies in evidence-based practice for health professionals: consensus statement based on a systematic review and delphi survey. *JAMA Netw Open*. 2018;1(2):e180281.
21. Kane MT. Validation. In: Brennan RL, ed. *Educational Measurement/Sponsored Jointly by National Council on Measurement in Education and American Council on Education*. 4th ed. USA: Praeger Publishers; 2006:17–64.
22. Downing SM, Haladyna TM. Validity and its threats. In: Downing SM, Yudkowsky R, eds. *Assessment in Health Professions Education*. 1st ed. New York: Routledge, Taylor and Francis group; 2009:21–55.
23. American Educational Research Association, American Psychological Association, Education NCoMi. *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association; 2014.
24. Leung K, Trevena L, Waters D. Content validation of the evidence-based nursing practice assessment tool. *Nurse Res*. 2018;26(1):33–40.
25. Roderick VP, Sharon SL, Glasziou PP, Del Mar CB, Maroeska RM. Antibiotics for acute otitis media in children. *Cochrane Database Syst Rev*. 2015;6. <http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD000219.pub4/abstract>.
26. Fowell S, Fewtrell R, McLaughlin P. Estimating the minimum number of judges required for test-centered standard setting on written assessments. Do discussions and iteration have an influence? *Adv Health Sci Educ*. 2008;13(1):11–24.
27. Yudkowsky R, Downing SM, Tekian A. Standard setting. In: Downing SM, Yudkowsky R, eds. *Assessment in Health Professions Education*. 1st ed. New York: Routledge, Taylor and Francis Group; 2009:119–148.
28. Tabachnick B. Principal components and factor analysis. In: *Using Multivariate Statistics*. 6th ed. Boston, USA: Pearson Education Inc; 2013:612–678.
29. Byrne BM. *Structural Equation Modeling with EQS: Basic Concepts, Applications, and Programming*. 2nd ed. USA: Lawrence Erlbaum Associates, Inc.; 2006.
30. Satorra A, Bentler PM. *Scaling Corrections for Chi-square Statistics in Covariance Structure Analysis*. 1988.
31. Bentler PM, Dijkstra T. Efficient estimation via linearization in structural models. In: Krishnaiah PR, ed. *Multivariate Analysis VI*. Amsterdam: North-Holland; 1985:9–42.
32. Fan X, Thompson B, Wang L. Effects of sample size, estimation methods, and model specification on structural equation modeling fit indexes. *Struct Equ Model*. 1999;6(1):56–83.
33. Ullman JB. Structural equation modeling. In: Fidel L, ed. *Using Multivariate Statistics*. 6th ed. Boston, USA: Pearson Education Inc; 2013:681–785.
34. Byrne BM. *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. 3rd ed. New York, USA: Routledge, Taylor & Francis Group; 2016.
35. Kane MT. Validating the interpretations and uses of test scores. *J Educ Meas*. 2013;50(1):1–73.
36. Messick S. Standards of validity and the validity of standards in performance assessment. *Educ Meas Issue Pract*. 1995;14(4):5–8.
37. Ben-David MF. AMEE Guide No. 18: standard setting in student assessment. *Med Teach*. 2000;22(2):120–130.
38. *IBM SPSS Statistics for Windows, Version 23.0 [computer program]*. Armonk, NY: IBM Corp; 2016.
39. *EQS 6.3 Structural Equations Program [computer program]*. California, USA: Multivariate software Inc; 2015.
40. Cicchetti D. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*. 1994;6(4):284–290.
41. Meskuskas JA. Setting standards for credentialing examinations: an update. *Eval Health Prof*. 1986;9(2):187–203.
42. Lyson ML, Downing SM, Gruppen LD, Yudkowsky R. Applying the bookmark method to medical education: standard

- setting for an aseptic technique station. *Med Teach*. 2013;35(7): 581–585.
43. Streiner D, Norman G, Cairney J. *Health Measurement Scales: a Practical Guide to Their Development and Use*. 5th ed. Oxford, UK: Oxford University Press; 2015.
 44. Leung K. *Measurement of Competence in Evidence-based Practice* [Thesis], [Chapter 9]. Australia: Sydney Medical School, University of Sydney; 2017.
 45. Gao S, Mokhtarian PL, Johnston RA. *Non-normality of Data in Structural Equation Models*. Washington, USA: Transportation Research Board's 87th annual meeting; 2007.
 46. Henly SJ. Robustness of some estimators for the analysis of covariance structure. *Br J Math Stat Psychol*. 1993;46(2): 313–338.
 47. Pett MA, Lackey NR, Sullivan JJ. An overview of factor analysis. In: Pett MA, Lackey NR, Sullivan JJ, eds. *Making Sense of Factor Analysis*. Thousand Oaks: SAGE Publications, Inc; 2003:2–29.
 48. Dylan W. Meaning and consequences in standard setting. *Assess Educ Princ Policy Pract*. 1996;3(3):287–308.