

Validity and Reliability Evidence for a New Measure: The Evidence-Based Practice Knowledge Assessment in Nursing

Darrell Spurlock, Jr., PhD, RN, NEA-BC, ANEF; and Amy Hagedorn Wonder, PhD, RN

ABSTRACT

Background: Studies of evidence-based practice (EBP) among nurses often focus on attitudes and beliefs about EBP and self-reported EBP knowledge. Because knowledge self-assessments can be highly inaccurate, the authors developed and tested a new objective measure of EBP knowledge—the Evidence-Based Practice Knowledge Assessment in Nursing (EKAN). **Method:** Seven subject matter experts reviewed candidate items, resulting in a scale content validity index of 0.94. Rasch modeling was used to evaluate item–person performance on the proposed unidimensional trait of EBP knowledge. The candidate item pool was then tested among 200 undergraduate nursing students. **Results:** Strong evidence of unidimensionality was confirmed by narrow item infit statistics centering on 1.0. The item separation index was 7.05, and the person separation index was 1.66. Item reliability was 0.98, and person reliability was 0.66. **Conclusion:** The 20-item EKAN showed strong psychometric properties for an instrument developed under the Rasch model and is available for use in research and educational contexts. [*J Nurs Educ.* 2015;54(11):605-613.]

Received: February 11, 2015

Accepted: July 8, 2015

Dr. Spurlock is Director of Scholarship and Institutional Effectiveness, Mount Carmel College of Nursing, Columbus, Ohio; and Dr. Wonder is Assistant Professor, Indiana University School of Nursing, Bloomington, Indiana.

This research was funded in part by a grant from the Indiana University School of Nursing Research Incentive Fund.

The authors have disclosed no potential conflicts of interest, financial or otherwise.

Address correspondence to Darrell Spurlock, Jr., PhD, RN, NEA-BC, ANEF, Director of Scholarship and Institutional Effectiveness, Mount Carmel College of Nursing, 127 S. Davis Ave., Columbus, OH 43222; e-mail: dspurlock@mccn.edu.

doi:10.3928/01484834-20151016-01

Evidence-based practice (EBP) is widely recognized for its capacity to improve patient care quality and reduce medical errors and undesirable variability in health care (McGinty & Anderson, 2008; Melnyk, 2007; Pravikoff, Pierce, & Tanner, 2005). Yet, in nearly 15 years since the Institute of Medicine (IOM, 2001) called for the wide-spread implementation of EBP across health professions, full implementation at the point of care is still lacking (Melnyk, Fineout-Overholt, Gallagher-Ford, & Kaplan, 2012; Pravikoff, Tanner, & Pierce, 2005).

Research on EBP implementation presents a paradoxical picture: nurses often report feeling adequately prepared for and excited about EBP but continue to report barriers such as a lack of time, organizational resistance, and a lack of EBP-related knowledge and education (Melnyk et al., 2012). In a survey of 1,015 nurses from across the United States, the second most highly endorsed statement (from among 18 statements) was, “It is important for me to gain more knowledge and skills in EBP” (Melnyk et al., 2012, p. 412). Although knowledge has been identified as a consistent barrier in implementing EBP, it has received relatively little empirical study. One major factor limiting the study of knowledge is a lack of available EBP knowledge measures. Tanner (2011) highlighted the critical lack of well-tested, empirically supported measures for nursing education research, which is an important consideration, given that so much of one’s formal education for EBP takes place in academic settings. The purpose of the current research was to develop and generate initial validity evidence for a new objective measure of EBP knowledge—the Evidence-Based Practice Knowledge Assessment in Nursing (EKAN).

KNOWLEDGE AND COMPETENCE

The authors’ interest in the concept of objectively measured EBP knowledge is driven by the relationship of knowledge to the more general concept of competence—in this case, competence in EBP. Much has been written on the concepts of competence and competency in nursing, yet there is no universal consensus on a definition (Axley, 2008). The American Nurses Association (ANA, 2013) described competency as “an expected level of performance that integrates knowledge, skills, ability, and judgment”

(p. 3). The ANA suggested that knowledge includes thinking and understanding the sciences and humanities, standards of practice, and insights gained from experiences (p. 4). Those sentiments are magnified in the definition of professional competence in medicine described by Epstein and Hundert (2002), who wrote that competence is “the habitual and judicious use of communication, knowledge, technical skills, clinical reasoning, emotions, values, and reflection in daily practice for the benefit of the individual and community being served” (p. 227).

Melnyk, Gallagher-Ford, Long, and Fineout-Overholt (2014) used a Delphi approach to provide additional specificity to the construct of EBP competence by developing EBP competencies for two levels of nursing practice: entry-level RNs and advanced practice nurses (APNs). The 13 RN competency statements suggest that practicing RNs should be able to formulate searchable PICOT (Population, Intervention, Comparison, Outcome, Time) questions, search scholarly databases, conduct research appraisal, plan for EBP-based changes to practice, and disseminate findings from practice changes. The APN competency statements extend beyond those for RNs, reflecting expectations for increased knowledge and leadership to enact EBP (Melnyk et al., 2014). The competencies suggest that APNs should be able to conduct extensive literature searches, independently critically appraise clinical practice guidelines and original research, and mentor others in EBP, among other requirements. Those competencies may seem aspirational, but they clearly support the ambitious goals outlined in the IOM’s (2010) report on *The Future of Nursing*, with a core message that nurses must achieve higher levels of education and skill in order to be effective leaders in the rapidly changing, evidence-based health care environment.

Although many philosophical and epistemological positions on the definitions of knowledge exist, in the current research the generally accepted cognitive view of knowledge as the body of information possessed by a person is adopted (Reber, Reber, & Allen, 2009). Educational and cognitive psychologists have devised many systems to describe the types and qualities of knowledge, including categories such as declarative knowledge, procedural knowledge, situational knowledge, practical knowledge, and many others (De Jong & Ferguson-Hessler, 1996). A generally held view in modern psychometrics is of knowledge as a latent construct, the amount of which can be estimated using appropriate measurement procedures. In educational contexts, knowledge tests are the most frequently used and accessible way to measure knowledge (Thorndike & Thorndike-Christ, 2011). To guide the development of the EKAN, the current authors adopted the following definition of EBP knowledge: EBP knowledge is the body of information necessary for a nurse to integrate the best available evidence, clinical expertise, and patient–family preferences and values for the delivery of safe and effective health care. This definition specifies the relationship between knowledge and the well-accepted definition of EBP provided by Sackett, Rosenberg, Gray, Haynes, and Richardson (1996) and Melnyk and Fineout-Overholt (2005).

MEASURING EBP

Researchers have developed a variety of instruments to measure aspects of EBP. Most instruments measure EBP-related

attitudes, beliefs, perceived facilitators and barriers, or self-rated knowledge (Funk, Champagne, Wiese, & Tornquist, 1991; Nagy, Lumby, McKinley, & Macfarlane, 2001; Shaneyfelt et al., 2006; Upton & Upton, 2006). In a recent systematic review, Leung, Trevena, and Waters (2014) identified and examined 24 instruments measuring nurses’ EBP knowledge, skills, and attitudes. Leung et al. found the most evidence for the Evidence-based Practice Questionnaire (EBPQ; Upton & Upton, 2006); however, a key limitation of the EBPQ is that it is a self-report tool. Leung et al. noted, “While the measurement of clinicians’ EBP knowledge, skills, and attitudes remains important for educators and researchers, an instrument to objectively measure these constructs needs to be developed” (p. 2191).

The need to develop an objective measure of EBP knowledge centers mainly on concerns about the accuracy and validity of self-report measures. Concerns about self-reporting are not new, but considerable evidence has emerged over the past decade that suggests bias and inaccurate self-assessment should be considered to be the rule, rather than the exception (Krueger & Dunning, 1999). In a study of this phenomena in the health professions, Lai and Teng (2011) compared self-rated performance with objectively measured performance among 45 undergraduate medical students using the Fresno Test, a measure of evidence-based medicine knowledge and skill, and an investigator-developed competency tool. Lai and Teng found low, nonsignificant correlations between self-rating and objective ratings: $r = .13$ ($p = .4$) for searching ability and $r = .24$ ($p = .1$) for appraisal ability. Individuals consistently overestimated their capabilities in several domains of EBP-related knowledge and skill. Lai’s and Teng’s findings are consistent with those of Blanch-Hartigan (2011), who, in a meta-analysis of 35 studies comparing medical students’ knowledge self-assessments with objective measures, found a nonsignificant correlation of $r = -.004$.

In the most extensive analysis to date, Zell and Krizan (2014) combined data from 22 meta-analyses, including 357,547 participants, to compare objective measures with self-reported assessments. The skills compared included academic performance, clinical medicine skills, language competence, sports performance, and intelligence; the studies came from fields that included psychology, medicine, education, and sports science. Zell and Krizan found a correlation of $r = .29$ ($SD = .11$) between self-reported measures and objective measures (such as test scores and objective observations). These findings from across multiple fields and types of objectively measured tasks strongly suggested a more limited role for self-assessments of skills and abilities, especially when objective measurement is possible.

METHOD

Development of the Measure

To facilitate practical use by educators and researchers, the authors sought to design an efficient, easy-to-score instrument comprising 20 to 30 multiple choice items to measure the examinee’s EBP knowledge. Using the definition of EBP knowledge provided previously, two authoritative sources were identified to help specify the knowledge domains that should be

tested in an objective measure of nursing EBP knowledge: The American Association of Colleges of Nursing's (AACN) *The Essentials of Baccalaureate Education for Professional Nursing Practice* (2008) and the Quality and Safety Education for Nurses (QSEN) competencies described by Cronenwett et al. (2007). Competency frameworks and standards of practice from other professional nursing organizations were evaluated, but the authors found the AACN's *Essentials* and the QSEN competencies to be comprehensive, broadly applicable, and specific enough to define the knowledge domains and topics useful to test item development.

The AACN's *Essentials* provide a framework to prepare baccalaureate graduates for the entry-level, generalist roles of caregiver and coordinator of care (AACN, 2008). Essential III, Scholarship for Evidence-Based Practice, outlines curricular expectations and expected outcomes for baccalaureate nursing program graduates. QSEN is a national collaborative of leaders in nursing and nursing education, originally funded in 2005 by the Robert Wood Johnson Foundation, whose goal in its first phase was to promote the transformation of nursing education curricula to one that is based on core concepts in quality, safety, and evidence-based practice (Cronenwett et al., 2007). Published in 2007, the QSEN knowledge, skills, and attitudes (KSAs) framework for prelicensure nursing education articulates the essential competencies required for evidence-based, safe, high-quality care (Cronenwett et al., 2007). In outlining the QSEN KSAs development process, Cronenwett et al. (2007) noted:

At each step, we sought feedback from nursing faculty. In contrast to the results of the survey, the nursing school faculty from 16 universities in the Institute for Healthcare Improvement Health Professions Education Collaborative reviewed the KSA draft, they uniformly reported that nursing students were not developing these KSAs. (p. 126)

Measurement Model and Content Validity

Because the EKAN was intended to measure EBP knowledge from respondents across a variety of educational backgrounds, settings, and populations, initial item evaluation and selection was conducted by fitting the data from the initial testing to the Rasch model. Tavakol and Dennick (2013) outlined the many advantages of the Rasch model over classical test theory (CTT), especially in the measurement of ability or achievement. Namely, psychometric evaluation using the Rasch model enables the separate evaluation of examinee ability and item difficulty and discrimination; in CTT, these factors are hopelessly confounded and cannot be separately examined (Tavakol & Dennick, 2013). The one-parameter Rasch model is represented mathematically as: $P_i(x_i = 1|b_i, \theta_j) = [1 + e^{-D(\theta_j - b_i)}]^{-1}$ where $P_i(x_i = 1)$ represents the probability of answering item i correctly, given the difficulty of the item (b_i), the ability of examinee (θ_j), e is the base of the natural logarithm (~2.7178) and D is an approximation of the normal ogive curve (~1.7; De Champlain, 2010). Item characteristics tend to demonstrate more stability across samples when evaluated using Rasch methods (Downing, 2003; Tavakol & Dennick, 2013), which is another value-added feature of measurement under the Rasch model.

One possible drawback to Rasch analysis is the required sample size of observations per item. Jones, Smith, and Talley

(2006) concluded that single-parameter Rasch models operate stably, and quality-of-item thresholds can be met with 100 to 200 observations per test item, whereas larger N s only enhance model robustness. Although the Rasch measurement model is popular in the psychometric literature, it is gaining in popularity in nursing, medicine, and the educational sciences. In fact, the Rasch model is the preferred measurement model in the movement to develop patient-reported outcome measures for use in population-based health research (Cappelleri, Jason Lundy, & Hays, 2014).

Using the AACN *Essentials* (2008) and the QSEN prelicensure competencies (2012) to organize the topics on which test items would be written, inspiration for the content of test items was drawn from a review of commonly used nursing research and EBP textbooks, the characteristics of data-based articles published in a range of scholarly publications, a review of syllabi for undergraduate EBP and nursing research courses, and the authors' own experiences as teachers of research and EBP in nursing education programs. Although there are no empirically derived gold standards for writing multiple choice test items, the authors wrote the initial item pool for the EKAN according to the well-regarded best practice guidelines by Haladyna, Downing, and Rodriguez (2002). Those guidelines focus the item writer on content validity, item construction, development of question options, and editing. Some of the recommendations include using three answer options instead of four, avoiding long question stems, avoiding the use of negatives (such as EXCEPT or NOT), keeping distractors exclusive when possible (or being clear that respondents should pick the BEST option), and making all distractors plausible.

The initial items were written to assess the full range of Bloom's revised cognitive levels (Anderson et al., 2000). The recommendation by DeVellis (2003) that an initial item pool contains 2 to 4 times the number of items desired in the final instrument was followed; the initial item pool consisted of 80 items. After the initial item pool was formulated, seven subject matter experts (SMEs) were enlisted to review each of the items. The SMEs were identified through professional affiliations and were required to possess demonstrated expertise, typically through a record of scholarly publication, recognized leadership, or direct involvement with either the AACN *Essentials* or the QSEN KSAs. In addition, the authors enlisted the review of two SMEs familiar with both EBP and demonstrated expertise in measurement and assessment. Following the procedures outlined by Polit, Beck, and Owen (2007), SMEs were asked to evaluate the relevance of the item to the construct of EBP knowledge (previously defined) on a scale from 1 to 4, where 1 = *not at all relevant* and 4 = *very/highly relevant*. SMEs were also asked to rate the clarity of the item and the congruence of the item to the suggested AACN or QSEN statement.

In the first round of SME reviews of the 80 candidate items, the scale content validity index was calculated as 0.90, which is a robust result (Polit et al., 2007). The authors examined individual items that reviewers rated poorly; five (of 80) items were deleted due to SME concerns that the items were peripheral to core EBP knowledge. Three other items were revised based on SME feedback. A second round of reviews was conducted on the three revised items, which were all found to be acceptable

TABLE 1
Description of Study Sample (N = 200)

Characteristic	n	%
Gender		
Female	181	90.5
Male	19	9.5
Race		
African American	12	6
American Indian/Alaskan Native	1	0.5
Asian/Pacific Islander	5	2.5
Hispanic/Latino	4	2
Multiracial	7	3.5
White/Caucasian	170	85
Prefer not to respond	1	0.5
Primary language		
English	195	97.5
Other	5	2.5
Highest non-nursing degree		
Associate	10	5
Bachelor of Arts	28	14
Bachelor of Science	52	26
Master of Arts	3	1.5
Master of Science	5	2.5
Other	4	2
None	98	49
Type of nursing education program enrolled		
Traditional Bachelor of Science in Nursing (BSN)	114	57
Accelerated BSN	86	43
Percent of current program completed		
< 25% complete	4	2
Approximately 25% complete	47	23.5
Approximately 50% complete	77	38.5
Approximately 75% complete	47	23.5
Almost 100% complete	25	12.5

by the SMEs. The final EKAN item pool scale content validity index was 0.94.

Participants and Procedures

After obtaining approval from the institutional review board for human subjects research, participants were recruited from among students enrolled in one of two baccalaureate nursing education programs offered in two large midwestern cities during the 2013-2014 academic year. Inclusion criteria required participants to be at least 18 years old and actively enrolled in a

Bachelor of Science in Nursing (BSN) program, with completion of or enrollment in at least one nursing course (no pre-nursing students). Participants were recruited to secured, face-to-face testing sessions using flyers, word-of-mouth invitations, and e-mail invitations. To incentivize participation, participants were offered a \$10 gift card to a popular retailer. One site administered the measures in paper-and-pencil format, whereas the other used an identical Web-based form using Qualtrics®, an online data capture and survey platform. Participants were not permitted to access other Web sites or use any reference material while completing the instruments. Considering a generous allotment of 2 minutes per multiple choice test item, and to provide time for responses to the demographic items, the testing sessions lasted until either the last student completed the instruments or 1.5 hours had passed. All participants finished the instruments within 1 hour.

ANALYSIS

Participants

Two hundred participants were included in the study, with an equal number recruited from each of the two study sites. Participants were primarily female (90.5%), self-identified as White/Caucasian (85%), and spoke English as their primary language (97.5%). Although most (57%) of the participants were enrolled in a traditional BSN program, a significant minority (43%) was enrolled in an accelerated BSN program. A substantial number of participants had degrees outside of nursing: 5% reported having earned an associate’s degree, 40% possessed a bachelor’s degree, and 4% a master’s degree. When asked to indicate the approximate point of program completion, the majority (38.5%) of participants indicated they had completed approximately 50% of their current programs, whereas equal proportions (23.5%) indicated they had completed approximately 25% or 75%. **Table 1** provides additional demographic details.

Participants were asked a range of questions to gain insight into experiences that could influence their performance on the knowledge items, including the recentness of completion and grades earned in research/EBP and statistics courses. Most (40.5%) of the participants were currently enrolled in a research/EBP course, whereas 30% of participants indicated they had not yet taken the course. Of those reporting a grade (n = 58) in their research/EBP course, 74% reported earning a grade of “A.” Most (56.5%) of the participants reported taking a statistics course more than 1 year ago, and 64% of those reported receiving a grade of “A.” The majority of participants (95.5%) reported receiving no special EBP education or training. Finally, participants were asked to rate the extent to which they agree with the statement, “I am sure I can deliver evidence-based care,” to which 80% either agreed or strongly agreed. **Table 2** contains additional details on the participants’ research and EBP-specific responses.

Construct Validation and Item Selection

Using Rasch analysis for selecting items for a knowledge scale is an iterative process, where item and scale analysis data inform theory-based judgments on the selection of items for a final scale. All participants provided responses for the 75 can-

didate EKAN items. The data were fitted to the Rasch model (1PL item-response theory) using jMetrik™ (Meyer, 2014) to evaluate item–person performance on the proposed unidimensional trait of EBP knowledge. jMetrik uses joint maximum likelihood (JML) in estimating parameters for Rasch. Iterative algorithm switching was set to allow for 200 iterations in the JML algorithm. The convergence criterion was left at the default value of 0.005. In the analysis, convergence was reached before 200 iterations. Item parameters were adjusted for bias using the $(n - 1)/n$ technique due to bias in parameter estimates from using the JML algorithm (Meyer, 2014).

Across all 75 candidate items, strong evidence of trait unidimensionality was confirmed by narrow item fit statistics, centering on 1.0. Infit and outfit statistics are parameters produced in Rasch analysis that provide information about how well a given item performs relative to an examinee's ability level (Meyer, 2014); values centering on 1.0 indicate good item–person fit. The weighted mean square infit was $M = 0.99$ (range = 0.91 to 1.16), standardized weighted mean square infit was $M = 0.08$ (range = -2.09 to 3.13), unweighted mean square outfit was $M = 0.99$ (range = 0.76 to 1.66), standardized unweighted mean square outfit was $M = 0.05$ (range = -2.25 to 3.15). Item difficulty for the candidate items ranged from -2.72 to 3.065 (mean difficulty level in Rasch is always 0). Under the Rasch model, separation indices reflect the extent to which a measure can consistently rank the person or the item on the trait continuum (Meyer, 2014). Meyer noted that separation values greater than 2.0 are desirable. The item separation index was 6.73; the person separation index was 1.42. Thus, item separation was robust, but person separation showed some restriction. Under the Rasch model, reliability is estimated for both items and persons. Person reliability parameters are interpreted similarly to reliability coefficients under CTT (Meyer, 2014). In the full item candidate pool, item reliability was 0.98; person reliability was 0.67. This again reflects strong item quality but some restriction in trait range among the sample.

To select items for the final 20-item EKAN form, individual candidate item parameters were examined. Items meeting Rasch parameter standards on difficulty, infit, and outfit were identified and ranked. Bond, Fox, and Bond (2007) suggested a rigorous range of 0.8 to 1.2 for infit and outfit mean square values. All items demonstrated acceptable infit and outfit mean square values. Because the current sample size was smaller than 300 participants, standardized mean square for infit and outfit were also examined. Meyer (2014) suggested that absolute values of greater than 3.0 for standardized mean square parameters indicate problems with fit that should prompt examination of the item. Because most candidate items fit the Rasch model well, a more conservative level of 2.0 was used to identify items for elimination. Using that method, nine candidate items were removed from consideration for inclusion on the final EKAN form, with most having absolute standardized infit or outfit parameters between 2.0 and 3.0. Only three items produced absolute values of 3.0 or greater.

As previously mentioned, item difficulty under Rasch analysis is indicated on a relative scale, where the mean difficulty level (representing items where 50% of participants answered the item correctly and 50% answered the item incorrectly) is

TABLE 2
Research and Evidence-Based Practice (EBP)
Characteristics of Sample ($N = 200$)

Characteristic	<i>n</i>	%
Recentness of research or EBP course		
Not yet enrolled	60	30
Currently enrolled	81	40.5
Completed < 6 months ago	35	17.5
Completed 6 months to 1 year ago	19	9.5
Completed > 1 year ago	5	2.5
Research course grade		
A	43	21.5
B	14	7
C	1	0.5
Not reported or not earned	142	71
Recentness of statistics course		
Not yet enrolled	10	5
Currently enrolled	30	15
Completed < 6 months ago	14	7
Completed 6 months to 1 year ago	33	16.5
Completed > 1 year ago	113	56.5
Statistics course grade		
A	101	50.5
B	47	23.5
C	10	5
Not reported/not earned	42	21
Special course or training in EBP		
No special course	191	95.5
Course of 1 day or less in length	5	2.5
Course of 2 to 3 days in length	3	1.5
More than 3 days in length	1	0.5
Response to "I am sure I can deliver evidence-based care."		
<i>Strongly disagree</i>	3	1.5
<i>Disagree</i>	2	1
<i>Neither agree nor disagree</i>	35	17.5
<i>Agree</i>	126	63
<i>Strongly agree</i>	34	17

0; positive values indicate higher levels of item difficulty and negative values indicate lower levels of item difficulty. The remaining 66 candidate items were ranked according to difficulty. An initial selection of 20 items from across the continuum of item difficulty were assembled into a draft final test form. The items were then examined for content. Although the authors did not write items based on hypothesized subdomains

of EBP knowledge, items for the final form were selected from a range of EBP topics. The 20-item form was analyzed and the results were compared with the performance of the initial candidate item pool. Because the initial goal was to produce a measure with 20 to 30 items, the authors analyzed a form with 10 additional items (30 items) to compare with the 20-item form and the original candidate item pool. No incremental improvement in person or item reliability over the 20-item form was noted and, as such, the 20-item form became the final form.

Final Scale Performance

For the final, 20-item EKAN measure, mean item difficulty was $M = 0.19$ (range = -2.0 to 2.8), weighted mean square infit was $M = 1.01$ (range = 0.95 to 1.06), standardized weighted mean square infit was $M = 0.33$ (range = -0.7 to 1.6), unweighted mean square outfit was $M = 1.02$ (range = 0.93 to 1.14), and standardized unweighted mean square outfit was $M = 0.34$ (range = -1.08 to 2.00). The item separation index was 7.05 ; the person separation index was 1.66 . Item reliability was 0.98 ; person reliability was 0.66 . Those values reflect strong item performance but indicate restriction in trait range, likely due to the homogeneity of the subject pool (Linacre, 2012; Meyer, 2014). Additional study in groups theoretically possessing a greater range of EBP knowledge is underway to confirm scale performance in groups with heterogeneous trait levels. **Table 3** provides details on individual item parameters and the topical content of the item.

On the final 20-item EKAN, scores ranged from 5 to 16 (of 20), and the mean score was $M = 10.4$ ($SD = 2.31$). No gender differences in mean score were noted ($t = 0.856$, $p = .393$). The correlation between responses to the attitude statement, "I am sure I can deliver evidence-based care," measured on a 5-point Likert-type scale, and total EKAN scores was not statistically significant ($r = .135$, $p = .057$). Group comparisons using other demographic and personal factors were not possible due to sample homogeneity on factors such as language, having special EBP training, or course grades in research/EBP or statistics courses; on these variables, subgroup sizes were often less than 10. To test for known-groups prior exposure or educational effects, participants who had not yet completed a nursing research/EBP course (a combination of participants not yet enrolled in a course or those currently enrolled in the first week of class) were compared with those who completed the course between 6 months and 1 year ago. An almost 2-point difference in mean EKAN scores between groups was noted (10.01 versus 11.47 ; $t = -2.53$, $p = .01$). A similar effect was seen in relation to the statistics course; those not having completed a statistics course scored statistically significantly worse on the EKAN than those who completed the course between 6 months and 1 year ago ($M = 8.8$ versus 10.9 , $t = -2.53$, $p = .015$). To further demonstrate this, the top and bottom decile of participants by EKAN score ($M = 6.5$ versus 14.1) were compared. Eighty percent of the top decile scorers had completed 75% or more of their educational programs, whereas only 20% of the bottom decile scorers had completed as much ($\chi^2(4,1) = 12.47$, $p = .01$). This provides further evidence of a prior exposure effect on EKAN scores.

LIMITATIONS

Several limitations to this study warrant consideration. First, the construct of EBP knowledge is general and could include many concepts, topics, and skill areas not assessed by the EKAN. The EKAN was not designed to be a comprehensive nor exhaustive measure of the universe of concepts under the umbrella of EBP knowledge. The EKAN was designed to be a short, practical measure for use in research and educational contexts. Another potential criticism is that items on the EKAN tend to favor research appraisal knowledge over other areas of EBP. That is a valid claim, but given the central role of evidence in EBP and the limitations imposed by a 20-item test form, the knowledge necessary to accurately appraise published research evidence deserves priority consideration.

A second limitation in this study is the homogeneity of the study sample, which has possible implications for examining scale performance, conducting known-group comparisons, and using the scale in different populations of examinees. Concerns about participant homogeneity are partially addressed by using the Rasch model (instead of CTT) to evaluate and select items for the final EKAN form. In the Rasch model, the ability of the person and the function of the item are evaluated separately but on a common, logistic scale. This approach, although more complex than CTT approaches, overcomes several key sampling-related limitations of CTT (De Champlain, 2010; Tavakol & Dennick, 2013). On the second issue, the range of known-groups comparisons useful for demonstrating discriminant validity were limited in this sample of undergraduate nursing students. The authors' initial motivation for developing the EKAN was to have a tool that is useful for evaluating educational outcomes, given the uncertain validity of ability and knowledge self-assessments. As the need became clear for an objective EBP knowledge measure that is useful beyond educational contexts, the range of difficulty of the items written for the candidate item pool was expanded, while staying within the content domains specified by the AACN (2008) and QSEN (2012). Although additional frameworks are applicable to nursing education at more advanced levels (e.g., AACN, 2012), the knowledge content areas are consistent.

On the issue of using the measure in more diverse populations, because the authors endorse test-use recommendations reflected in both the *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, 1999) and the National League for Nursing's *Fair Testing Guidelines* (2012), additional research using the measure in more diverse populations—including practicing nurses, nurses with a wide range of educational preparation, and among geographically and demographically diverse nursing students in Associate Degree in Nursing through Doctor of Philosophy programs—is currently planned. The current article reports initial validity and scale performance results, fully recognizing that validity and reliability are not static concepts but rather should be supported by the ongoing collection of evidence.

TABLE 3
Evidence-Based Practice Knowledge Assessment in Nursing (EKAN) Final Form Item Description and Rasch Parameters

Item and Content Description	Difficulty (SE)	Infit: Weighted Mean Square Fit	Infit: Standardized Weighted Mean Square Fit	Outfit: Unweighted Mean Square Fit	Outfit: Standardized Unweighted Mean Square Fit
1. Purpose of regression versus other tests	-0.5 (0.15)	0.98 ^a	-0.41 ^b	0.96 ^a	-0.62 ^b
2. Sampling and study design	2.82 (0.27)	0.99 ^a	0.05 ^b	1.13 ^a	0.58 ^b
3. Purpose of institutional review board	-1.23 (0.18)	1.02 ^a	0.26 ^b	1.03 ^a	0.24 ^b
4. Distinguishing measures of central tendency	-0.8 (0.16)	0.99 ^a	-0.13 ^b	0.97 ^a	-0.3 ^b
5. Distinguishing validity, reliability, and generalizability	-0.58 (0.15)	0.99 ^a	-0.19 ^b	0.98 ^a	-0.31 ^b
6. Proper use of pre-appraised evidence	1.07 (0.16)	1.05 ^a	0.76 ^b	1.06 ^a	0.85 ^b
7. Role of judgment in EBP decision making	-1.59 (0.2)	1.02 ^a	0.18 ^b	1.08 ^a	0.52 ^b
8. Steps of the EBP process	0.24 (0.14)	1.03 ^a	0.85 ^b	1.02 ^a	0.69 ^b
9. Facilitating EBP in practice settings	1.15 (0.16)	1.07 ^a	0.98 ^b	1.09 ^a	1.14 ^b
10. Interpreting odds ratios	-0.12 (0.15)	0.99 ^a	-0.33 ^b	0.99 ^a	-0.28 ^b
11. Understanding credibility and bias	0.53 (0.15)	1.04 ^a	1.19 ^b	1.05 ^a	1.26 ^b
12. Identifying steps in plan-do-study-act cycle	1.95 (0.2)	1.07 ^a	0.55 ^b	1.15 ^a	1.00 ^b
13. Priority of evidence, patient values, and clinical judgment in EBP decision making	-0.28 (0.15)	1.01 ^a	0.18 ^b	1.00 ^a	0.02 ^b
14. Distinguishing causation from correlation in regression	0.53 (0.15)	1.06 ^a	1.68 ^b	1.09 ^a	2.04 ^b
15. Ranking of evidence quality (hierarchy)	0.38 (0.15)	1.05 ^a	1.68 ^b	1.07 ^a	1.80 ^b
16. Strength of measurement approaches	-0.49 (0.15)	0.96 ^a	-0.79 ^b	0.93 ^a	-1.09 ^b
17. Role of PICOT question in evidence searching	-0.12 (0.15)	1.00 ^a	0.05 ^b	1.00 ^a	0.00 ^b
18. Nurse-sensitive quality indicators	-2.01 (0.23)	1.00 ^a	0.04 ^b	0.95 ^a	-0.20 ^b
19. Understanding effect sizes	1.33 (0.17)	1.00 ^a	0.06 ^b	1.00 ^a	0.08 ^b
20. Statistical versus clinical significance	1.57 (0.19)	0.99 ^a	-0.11 ^b	0.95 ^a	-0.43 ^b

Note. EBP = evidence-based practice; PICOT = Population, Intervention, Comparison, Outcome, Time.

^a Fit statistic falling within rigorous acceptability range of 0.8 to 1.2, recommended by Bond, Fox, and Bond (2007).

^b Standardized fit statistics falling within an acceptability range of -3.0 to 3.0, recommended by Meyer (2014).

DISCUSSION

The need for an objective measure of nurses' EBP knowledge manifests in a variety of ways. To this point, measurement of EBP knowledge in nursing has been limited to self-report questionnaires. Extensive research evidence suggests self-reports of knowledge and skill capabilities cannot be substituted

for objective measurement. Although nursing has made much progress in identifying the competencies necessary to enact EBP in practice settings (Melnik et al., 2014), a comprehensive, empirically derived model of the relationships among EBP knowledge, skills, attitudes, and beliefs is lacking. In developing the EKAN, leaders, researchers, and educators now have access to an objective measure of EBP knowledge that can be

used alongside existing EBP attitude and belief scales to examine these relationships. Despite the extensive efforts of educators and practice leaders over the past decade to integrate EBP into nursing curricula at all levels and to establish structures that support and enable EBP in practice settings, EBP has not yet become the rule; too often, it remains the exception. The EKAN will provide for new insights into understanding the extent to which examinees possess the knowledge necessary for enacting EBP. Making sense of participants' performance on the EKAN will require an openness on the part of researchers and nursing faculty alike to question the effectiveness of existing EBP educational strategies, the durability of EBP knowledge over time, and the role of deliberate practice in sustaining EBP knowledge and skills.

CONCLUSION

Additional research is needed to establish normative score information for nurses with a variety of educational backgrounds. In this research, the authors reported differences in scores for participants who had completed their research/EBP and statistics courses, compared with those who had not. The authors hypothesized that additional score improvements in nurses with more EBP educational exposure and practical experience will be seen. Significant insights will be gained from a large, multisite study of students and practicing nurses enrolled in Associate Degree in Nursing, BSN, Master of Science in Nursing, Doctor of Nursing Practice, and Doctor of Philosophy programs across the United States. In the current study, the authors found a small, positive but statistically non-significant relationship between an EBP belief measure and EBP knowledge scores, which is consistent with findings from other fields. An important line of inquiry, also in progress, is to further examine the correspondence between self-reported and objectively measured EBP knowledge. There is little research on this topic in the nursing literature, but the need for it is clear in light of findings from Blanch-Hartigan (2011) and Zell and Krizan (2014).

REFERENCES

- American Association of Colleges of Nursing. (2008). *The essentials of baccalaureate education for professional nursing practice*. Retrieved from <http://www.aacn.nche.edu/education-resources/baccessential08.pdf>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- American Nurses Association. (2013). *Competency model*. Washington, DC: Author. Retrieved from <http://www.ana-leadershipinstitute.org/Doc-Vault/About-Us/ANA-Leadership-Institute-Competency-Model-pdf.pdf>
- Anderson, L.W., Krathwohl, D.R., Airasian, P.W., Cruikshank, K.A., Mayer, R.E., Pintrich, P.R., . . . Wittrock, M.C. (2000). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York, NY: Pearson.
- Axley, L. (2008). Competency: A concept analysis. *Nursing Forum*, 43, 214-222. doi:10.1111/j.1744-6198.2008.00115.x
- Blanch-Hartigan, D. (2011). Medical students' self-assessment of performance: Results from three meta-analyses. *Patient Education and Counseling*, 84, 3-9. doi:10.1016/j.pec.2010.06.037
- Bond, T., Fox, C.M., & Bond, T.G. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed.). Mahwah, NJ: Routledge.
- Cappelleri, J.C., Jason Lundy, J., & Hays, R.D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clinical Therapeutics*, 36, 648-662. doi:10.1016/j.clinthera.2014.04.006
- Cronenwett, L., Sherwood, G., Barsteiner, J., Disch, J., Johnson, J., Mitchell, P., . . . Warren, J. (2007). Quality and safety education for nurses. *Nursing Outlook*, 5, 122-131.
- De Champlain, A.F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical Education*, 44, 109-117. doi:10.1111/j.1365-2923.2009.03425.x
- De Jong, T., & Ferguson-Hessler, M.G.M. (1996). Types and qualities of knowledge. *Educational Psychologist*, 31, 105.
- DeVellis, R.F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.
- Downing, S.M. (2003). Item response theory: Applications of modern test theory in medical education. *Medical Education*, 37, 739-745.
- Epstein R.M., & Hundert, E.M. (2002). Defining and assessing professional competence. *Journal of the American Medical Association*, 287, 226-235. doi:10.1001/jama.287.2.226
- Funk, S.G., Champagne, M.T., Wiese, R.A., & Tornquist, E.M. (1991). BARRIERS: The Barriers to Research Utilization Scale. *Applied Nursing Research*, 4, 39-45.
- Haladyna, T.M., Downing, S.M., & Rodriguez, M.C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15, 309-334.
- Institute of Medicine. (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academies Press.
- Institute of Medicine. (2010). *The future of nursing: Leading change, advancing health*. Washington, DC: National Academies Press.
- Jones, P., Smith, R., & Talley, D.M. (2006). Developing test forms for small-scale achievement testing systems. In S.M. Downing & T.M. Haladyna (Eds.), *Handbook of test development* (pp. 487-525). Mahwah, NJ: Lawrence Erlbaum.
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77, 1121-1134.
- Lai, N.M., & Teng, C.L. (2011). Self-perceived competence correlates poorly with objectively measured competence in evidence based medicine among medical students. *BMC Medical Education*, 11, 25. doi:10.1186/1472-6920-11-25
- Leung, K., Trevena, L., & Waters, D. (2014). Systematic review of instruments for measuring nurses' knowledge, skills and attitudes for evidence-based practice. *Journal of Advanced Nursing*, 70, 2181-2195. doi:10.1111/jan.12454
- Linacre, J.M. (2012). *A user's guide to WINSTEPS Rasch-model computer programs*. Chicago, IL: MESA Press.
- McGinty, J., & Anderson, G. (2008). Predictors of physician compliance with American Heart Association guidelines for acute myocardial infarction. *Critical Care Nursing*, 31, 161-172.
- Melnyk, B.M. (2007). The evidence-based practice mentor: A promising strategy for implementing and sustaining EBP in healthcare systems. *Worldviews on Evidence-Based Nursing*, 4, 123-125.
- Melnyk, B.M., Fineout-Overholt, E., Gallagher-Ford, L., & Kaplan, L. (2012). The state of evidence-based practice in US nurses: Critical implications for nurse leaders and educators. *The Journal of Nursing Administration*, 42, 410-417.
- Melnyk, B.M., & Fineout-Overholt, E.F. (2005). *Evidence-based practice in nursing and healthcare: A guide to best practice*. Philadelphia, PA: Lippincott.
- Melnyk, B.M., Gallagher-Ford, L., Long, L.E., & Fineout-Overholt, E. (2014). The establishment of evidence-based practice competencies for practicing registered nurses and advanced practice nurses in real-world clinical settings: Proficiencies to improve healthcare quality, reliability, patient outcomes, and costs. *Worldviews on Evidence-Based Nursing*, 11, 5-15.
- Meyer, J.P. (2014). *Applied measurement with jMetrik*. New York, NY: Routledge.

- Nagy, S., Lumby, J., McKinley, S., & Macfarlane, C. (2001). Nurses' beliefs about the conditions that hinder or support evidence-based nursing. *International Journal of Nursing Practice, 7*, 314-321.
- National League for Nursing. (2012). *NLN fair testing guidelines for nursing education*. Retrieved from <http://www.nln.org/docs/default-source/default-document-library/fairtestingguidelines.pdf?sfvrsn=2>
- Polit, D.F., Beck, C.T., & Owen, S.V. (2007). Focus on research methods: Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*, 459-467.
- Pravikoff, D.S., Pierce, S.T., & Tanner, A. (2005). Evidence-based practice readiness study supported by academy nursing informatics expert panel. *Nursing Outlook, 53*, 49-50.
- Pravikoff, D.S., Tanner, A.B., & Pierce, S.T. (2005). Readiness of U.S. nurses for evidence-based practice: Many don't understand or value research and have had little or no training to help them find evidence on which to base their practice. *American Journal of Nursing, 105*(9), 40-51.
- Quality and Safety Education for Nurses. (2012). *Pre-licensure KSAs*. Retrieved from <http://qsen.org/competencies/pre-licensure-ksas/>
- Reber, A.S., Reber, E., & Allen, R. (2009). *The Penguin dictionary of psychology*. New York, NY: Penguin Books.
- Sackett, D.L., Rosenberg, W.M., Gray, J.A., Haynes, R.B., & Richardson, W.S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal, 312*(7023), 71-72.
- Shaneyfelt, T., Baum, K.D., Bell, D., Feldstein, D., Houston, T.K., Kaatz, S., . . . Green, M. (2006). Instruments for evaluation education in evidence-based practice: A systematic review. *Journal of the American Medical Association, 296*, 1116-1127.
- Tanner, C.A. (2011). The critical state of measurement in nursing education research. *Journal of Nursing Education, 50*, 491-492. doi:10.3928/01484834-20110819-01
- Tavakol, M., & Dennick, R. (2013). Psychometric evaluation of a knowledge-based examination using Rasch analysis: An illustrative guide: AAME Guide No. 72. *Medical Teacher, 35*, e838-e848.
- Thorndike, R.M., & Thorndike-Christ, T.M. (2011). *Measurement and evaluation in psychology and education* (8th ed.). Boston, MA: Pearson.
- Upton, D., & Upton, P. (2006). Development of an evidence-based practice questionnaire for nurses. *Journal of Advanced Nursing, 53*, 454-458.
- Zell, E., & Krizan, Z. (2014). Do people have insight into their abilities? A metasynthesis. *Perspectives on Psychological Science, 9*, 111-125. doi:10.1177/1745691613518075