The State of the Science on Clinical Evaluation in Nursing Education

Lynne P. Lewallen and Elizabeth R. Van Horn

Abstract

AIM The purpose was to synthesize the published evidence to present the state of the science in clinical evaluation research in nursing education.

BACKGROUND Clinical evaluation is key to ensuring nursing students' clinical competence, application of knowledge, and critical thinking, all of which are important to patient safety and quality nursing care.

METHOD Cooper's research synthesis method was used.

RESULTS A comprehensive literature search resulted in 250 documents, of which 88 met study criteria. Topics were exhaustive but not mutually exclusive and included competence, instrumentation, congruence, teaching methods, objective structured clinical evaluation, faculty/preceptor issues with clinical evaluation, essential clinical behaviors, topic-based evaluation, decision-making about clinical grade, and clinical reasoning.

CONCLUSION Nursing education science is in its infancy in many areas. Two areas most in need of future research are the need to accurately define and efficiently measure competence in the clinical area and the need for reliable and valid instrumentation.

KEY WORDS Clinical Evaluation – Nursing Education – Systematic Review – Nursing Students

Clinical evaluation is ubiquitous in nursing programs, especially prelicensure programs and those graduate programs leading to advanced practice certification. Although licensure and certification exams are standardized in many countries, with published test plans (e.g., NCLEX) and based on a standard body of content knowledge, evaluation methods of clinical performance in nursing programs typically vary by program. Studies have been done on clinical evaluation for many years, but no literature is available that synthesizes this knowledge. Therefore, the purposes of this study were to gather the published evidence on clinical evaluation and synthesize this evidence to present the state of the science in clinical evaluation research in nursing education.

Clinical evaluation is key to ensuring nursing students' clinical competence, application of knowledge, and critical thinking, all of which are important to patient safety and the provision of quality nursing care (Walsh, Jairath, Paterson, & Grandjean, 2010). Indeed, evaluation of student competence in the clinical setting is pivotal in the professional development of nursing students in all types of nursing programs. The research literature is replete with nursing studies that have examined the effectiveness of a wide variety of clinical evaluation methods and tools. Examples include the use of e-portfolios (Garrett, MacPhee, & Jackson, 2013), faculty anecdotal notes (Hall, 2013), competency inventories (Hsu & Hsieh, 2013), clinical learning outcome tools (Skúladóttir & Svavarsdóttir, 2016), and objective structured clinical evaluation (OSCE) tools (Fan, Wang, Chao, Jane, & Hsu, 2015). Descriptive studies commonly include surveys about student perceptions of and satisfaction with clinical evaluation methods and instructor performance (Peyman et al., 2011; Severinsson & Sand, 2010). Qualitative studies of clinical evaluation methods include an exploration of clinical teaching behaviors through observation of nursing students (Cassidy et al., 2012) and a phenomenological study to explore the student evaluation process from the clinical instructor's perspective (Duke, 1996).

Several reviews of the literature on clinical evaluation have been published. Perhaps the first review of clinical evaluation and its development was published by Wood (1982), who examined several methods, including rating scales and student self-evaluation, and summarized the many inherent challenges in the clinical evaluation process. In 1994, Krichbaum and colleagues reviewed the history of clinical evaluation from the 1940s to the 1980s and described the development and testing of the clinical evaluation tool, a rating scale of 10 criteria that can be used across various levels of nursing programs.

More recently, two reviews examined the use of objective measures in clinical evaluation. Walsh, Bailey, and Koren (2009) conducted an integrative review on the use of the OSCE in nursing. This method, first used to evaluate medical education in the 1970s, has a standardized checklist that trained observers use to rate student performance (Rushforth, 2007). Walsh et al. (2009) reviewed 41 studies published from 1960 to 2008 that used the OSCE method

About the Authors *Lynne P. Lewallen, PhD, RN, CNE, ANEF, is professor and associate dean for academic affairs, School of Nursing, University of North Carolina at Greensboro. Elizabeth R. Van Horn, PhD, RN, CNE, is associate professor, School of Nursing, University of North Carolina at Greensboro.* *This Research was funded by a 2014 NLN Research in Nursing Education Grant, Ruth Donnelly Corcoran Research Award. The authors acknowledge the work of their research assistants, Dr. Catherine Moore, Dr. Wendasha Jenkins, and Dr. Sarah Abrams.* *For more information, contact Dr. Lewallen at lplewall@uncg.edu.* *The authors have declared no conflict of interest.*

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (www.neponline.net).

doi: 10.1097/01.NEP.0000000000000376

for nursing student clinical evaluation. Cant, McKenna, and Cooper (2013) examined 16 quantitative studies published from 2000 to 2011, all of which used an objective method of clinical evaluation, including an OSCE instrument. This review, which provided a more thorough assessment of the research, included reliability and validity data for each study, quality ratings, and assessments of risk of bias when using the study instrument. However, although useful in our understanding of clinical evaluation, the latter two reviews were limited to quantitative studies and a narrow form of clinical evaluation.

A more recent review by Wu, Enskar, Lee, and Wang (2015) examined the evaluation of clinical competency in undergraduate students. This review focused on the process of end-of-program assessment and described some reports of reliability and validity. It included only research with undergraduate students and studies published after 1999.

Clinical evaluation is multifaceted and includes evaluation of cognitive, affective, and psychomotor learning; furthermore, the components of clinical evaluation, especially critical thinking, can be complex and difficult to measure. Clinical evaluation is often subjective and is commonly carried out while the student is learning; the process has been compared to evaluation of the performing arts (Roberts, 2011). To date there have been no comprehensive reviews of research on clinical evaluation that include the spectrum of clinical evaluation forms, formats, or types. In addition, these limited examinations provide no tailoring to new and emerging patient care environments and student clinical experiences; they provide minimal contribution to nursing education for the future roles and responsibilities of nurses in the dynamic health care arena.

METHOD

Conceptual Framework

Cooper's (2010) method for research synthesis was used as the framework for this study. Cooper's seven steps in conducting a research synthesis are as follows: 1) formulate the problem, 2) search the literature, 3) gather information from studies, 4) evaluate the quality of studies, 5) analyze and integrate the outcomes of studies, 6) interpret the evidence, and 7) present the results. Although the theoretical frameworks of the studies were examined, the focus of the analysis was on the results of the studies.

Search

Inclusion criteria were reports of research studies that examined clinical evaluation for any level of nursing student, were accessible in full text, and were written in English. Database searches included the earliest dates of each database through the end of 2016. Articles were excluded if they did not report results of a study, if the study focused on practicing nurses rather than nursing students, if the study focused on human patient simulation or only on student or faculty perceptions of or satisfaction with clinical evaluation, and if the articles were not available in English.

With the assistance of a health sciences librarian, an electronic literature search was conducted using the following databases: EBSCO databases, Academic Search Complete, CINAHL with Full Text, Education Full Text (H.W. Wilson), ERIC, Dissertation Abstracts, Health Source–Consumer Edition, Health Source: Nursing/Academic Edition, PsycARTICLES, PsycINFO, Cochrane, and Scopus. The following search terms were used: *clinical evaluation AND nursing student NOT attitude NOT simulation NOT perception NOT satisfaction NOT review.*

The initial search resulted in 207 documents, including 21 unpublished doctoral dissertations. To expand the search, tables of contents of the major journals publishing nursing education research (*Nursing Education Perspectives, Journal of Nursing Education, Nurse Educator, International Journal of Nursing Education Scholarship, Nurse Education Today, Teaching and Learning in Nursing,* and *Nurse Education in Practice*) were searched manually, which produced an additional 27 unique articles. The reference lists of six literature review articles (Cant et al., 2013; Krichbaum, Rowan, Duckett, Ryden, & Savik, 1994; Walsh et al., 2009; Watson, Stimpson, Topping, & Porock, 2002; Wu et al., 2015; Yanhua & Watson, 2011) were also searched, resulting in another 16 unique articles.

Finally, the Registry of Nursing Research in the Virginia Henderson International Nursing Library's electronic database was searched to identify unpublished abstracts of relevant conference presentations. This step provided access to the so-called "gray" or unpublished research literature (Hawker, Payne, Kerr, Hardey, & Powell, 2002.) (A flow diagram describing the search is available as Supplemental Digital Content 2 at http://links.lww.com/NEP/A84.)

All these searches yielded 250 unique documents. Each article/dissertation was read and abstracted by the investigators and two trained doctoral research assistants. Periodic reliability checks were completed among the four readers, and any disagreements were resolved. As a result of this analysis, 162 documents did not meet the study criteria and were eliminated, leaving 88 articles, including three dissertations, for analysis.

Analysis

For analysis, a matrix summarizing key points of each article/dissertation zwas created along with individual narrative summaries. The analysis was conducted by the two investigators using an iterative process. Articles were read by both investigators, and topics were identified and agreed upon. The studies were then placed into categories. Several of the studies examined more than one topic; therefore, a decision was made to not make the categories mutually exclusive, so that all topics would be captured in this research synthesis.

The matrix data were analyzed by topic for study characteristics (sample size, sample characteristics, research method, country), significant or major findings, and patterns and themes. Based on the analyses, conclusions were drawn about the state of the research literature on clinical evaluation methods in nursing education.

RESULTS

Of the 88 studies that met criteria, 64 were quantitative, 10 were qualitative, and 14 used mixed methods. Mixed-methods studies were defined as those that reported substantive findings from both types of data. Studies that were quantitative but also included open-ended questions were not classified as mixed method. Of the quantitative studies, most were descriptive research (n = 33), followed in frequency by quasiexperimental design. Ten quantitative articles reported psychometric testing of instruments; the remainder used experimental (n = 4) and cohort or case-control designs (n = 4).

Levels of evidence were determined for the studies according to the criteria by Melnyk and Fineout-Overholt (2011). Studies focusing on psychometric testing were excluded from this analysis. Most of the studies represented the lower levels of evidence, especially

descriptive research or single qualitative studies. See Table 1 for details about levels of evidence.

Of the 88 articles, only 25 reported using a theoretical framework to guide the study. Twenty-nine articles reported funding for the study; the majority received intramural funding from universities or funding from governments (in the case of some studies conducted outside the United States). The types of student participants in the studies represented all levels of nursing education from diploma through master's education; no studies reported using practical nurse students or doctoral students as participants. Eighteen countries were represented.

Articles were grouped into 10 categories that were not mutually exclusive; rather, they were exhaustive to capture all the topics addressed in the studies. The researchers created definitions for each category; several studies examined more than one topic. (A table listing identified topics, study characteristics for each topic, and a list of included articles included for each topic is available as Supplemental Digital Content 1 at http://links.lww.com/NEP/A83.)

Topics are discussed individually, from the most common to the least common. A discussion of the current state of the science of clinical evaluation in nursing education follows.

Competence

Competence was the topic of 35 studies, by far the most common concept examined. Competence was defined as evaluation of competence in general or a specific area, such as a skill; most studies aimed to measure general competence, for example, at the end of a nursing program. The studies did not share a common definition of competence.

Most studies used researcher-developed instruments; 16 studies did psychometric testing of the instrument as part of the study.

Table 1: Levels of Evidence Reflected in the Studies

Level and Criteria	No. of Studies
Level 1. Systematic review/ meta-analysis of randomized controlled trials	0
Level 2. One or more randomized controlled trials	6
Level 3. Controlled trial (no randomization)	15
Level 4. Case-control or cohort study	5
Level 5. Systematic review of descriptive and qualitative studies	0
Level 6. Single descriptive or qualitative study	51
Level 7. Expert opinion	N/A (excluded from study)

Note. Levels of evidence as described by Melnyk and Fineout-Overholt (2011).

Only three instruments — the Nursing Competencies Scale, Clinical Nursing Competence Questionnaire, and the Competency Inventory of Nursing Students — were used in more than one study. The Nursing Competencies Scale is mentioned in two articles reporting different aspects of the same study (Norman, Watson, Murrells, Calman, & Redfern, 2002; Watson, Calman, Norman, Redfern, & Murrells, 2002). The Clinical Nursing Competence Questionnaire, developed as a measure of general nursing competence in one study (Lee-Hsieh, Kao, Juo, & Tseng, 2003), was subsequently translated into French; its psychometric properties were examined by Beogo, Rojas, Gagnon, and Liu (2016). The psychometric properties of the Competency Inventory of Nursing Students were reported in one study (Hsu & Hsieh, 2013) and were then used in a subsequent study conducted by one of the same authors (Hsieh & Shu, 2013).

For studies that measured specific competencies rather than general competence, topics included medication calculation, vaccination efficacy and knowledge, critical thinking, facility with psychiatric nursing skills, clinical perioperative competence, and interpersonal communication. All studies that measured general competence used some sort of evaluation that measured clinical behaviors; some also compared scores on specific knowledge exams or grades in didactic courses with clinical measures.

The person completing the competence measures varied. Several studies used student self-reports; others used faculty assessments, student assessments compared with faculty assessments, preceptor assessments, and comparison of preceptor or nurse manager and faculty assessments. Some studies looked at student progress over time (e.g., progress from beginning to end of a nursing program, beginning to end of a semester or rotation, or before and after an event such as a summer preceptorship). One study, which analyzed a school's archive of clinical assessments spanning more than a decade, examined the evolution of categories of behaviors measured over time.

Instrumentation

Twenty studies focused on the psychometric testing of specific instruments as the purpose of the study. To be classified in the topic of instrumentation, the researchers must have performed and reported psychometric analyses. Seventeen studies tested instruments with undergraduate students, two with graduate students, and one with both levels. Four studies reported the use of a conceptual framework.

A variety of statistical analyses were used to determine reliability, validity, and/or factor analysis of investigator-developed and established instruments. Most studies analyzed instruments that address student competence or clinical performance. Two studies evaluated the reliability and validity of OSCEs. Other topics included instruments that measure cultural competence and interpersonal communication.

Most studies determined that instruments were reliable; those testing validity found factors with substantial loadings that accounted for reasonable variance. However, some instruments lacked reliability or convergent validity with standardized outcomes. Although the studies were conducted in several countries, use of instruments cross-culturally should be performed with caution, as psychometric findings can be specific to sample and setting.

Congruence

Nursing students are rarely evaluated by only one person during their clinical education; often, multiple people will evaluate a student during

a single semester. Therefore, it is important that congruence be assessed among multiple evaluations of the same student. In our synthesis, 14 studies examined congruence, defined as the comparison of clinical evaluation outcomes between two or more types of evaluators. Comparisons included students' self-assessments compared to peer assessments; students' assessments compared with those of faculty; faculty assessments compared with preceptors'; preceptors' compared with students'; or patients' assessments of students compared with assessments made by preceptors, students, or faculty.

All studies examined undergraduate nursing students. Most reported incongruence among different types of evaluators. Although individual studies varied slightly, students tended to rate themselves lower than did preceptors or faculty; faculty tended to rate students lower than did preceptors. One study, which used a mixed-methods approach that included focus groups (Sedgwick, Kellett, & Kalischuck, 2014), found that a major impediment to achieving congruence was the differing views of each type of evaluator about the definition of the concept being measured (in this case, general nursing competence).

Teaching Methods

Twelve studies were grouped into the category of teaching methods, defined as the examination of selected student outcomes following a specific instructional method. These studies frequently measured outcomes pre- or postintervention or compared groups of students who received the intervention with a control group. All of these studies examined undergraduate students; one also included medical students in the sample.

Several different teaching methods were evaluated for effectiveness using two-group comparison or one-group pre- and posttest designs. Eight studies compared an innovative teaching method to traditional instruction; innovative methods included cooperative learning (two studies), analogy-guided learning, competency-based education, problem-based learning, mentoring, authentic assessment rubrics, and role play. Significant improvements in outcome measures were found in all of the innovative teaching methods except role play.

Two studies examined different configurations of clinical formats (designated hours or days) between two groups and the effects on student outcomes. Neither study found significant differences in groups that could be attributed to differing clinical hours or day configurations. One study compared health promotion counseling self-efficacy in nursing students and medical students in their respective clinical rotations; self-efficacy in nursing students was significantly higher at the end of the clinical experience. One study examined positive and negative faculty feedback and the effects on student grades and self-evaluation; positive feedback was significantly related to higher grades, and negative feedback was related to accuracy in student self-evaluation.

Objective Structured Clinical Evaluation

The OSCE was the topic of 11 studies; studies were classified in this category if the use of this clinical evaluation method was discussed in any way. OSCEs have been used for many years as an evaluation method, and most OSCE research has been conducted with graduate students. OSCEs are customized to the type of student being evaluated, for example, an OSCE for a primary care nurse practitioner student might consist of a structured simulated clinical visit using a trained standardized patient, and the student would be evaluated on aspects of that clinical encounter that the instructor deemed critical OSCEs are customized to the type of student being evaluated, for example, an OSCE for a primary care nurse practitioner student might consist of a structured simulated clinical visit using a trained standardized patient, and the student would be evaluated on aspects of that clinical encounter that the instructor deemed critical. All studies but one were conducted in countries other than the United States.

The use of OSCE was evaluated in four different ways. Two studies evaluated effectiveness with undergraduate students; student responses were mixed, with students noting the OSCE was a more stressful, but also more objective method of evaluation than others. One study tested the use of the OSCE as part of an educational intervention compared to traditional teaching; no significant difference was found for student outcomes. Another study used the OSCE as one measure of effectiveness of a competency-based teaching strategy versus traditional instruction; scores did not differ between groups.

Seven studies completed psychometric testing on the OSCE to determine reliability and validity or compared the OSCE with another established evaluation method. Studies testing reliability and validity had mixed results, with some finding it a reliable method and others finding a lack of concordance with scores from other methods. Student responses were also mixed. Some found the OSCE a useful and fair method of evaluation; many reported it to be a stressful experience.

Faculty/Preceptor Issues With Clinical Evaluation

Seven studies focused on faculty/preceptor issues with clinical evaluation. This topic emerged later than some of the others, with dates ranging from 1996 to 2014.

The studies reported positive and negative experiences with evaluation tool use and confidence in the results of evaluation or studied an aspect of the process of evaluation. Several examined faculty and preceptor satisfaction with the process of grading students in clinical. Most identified some issues with evaluation tools, ranging from formatting to definitions of terms. Faculty and preceptors expressed difficulty in failing students and concern for objectivity in the process. Studies in this category also compared didactic and clinical grades and examined consistency among faculty in the same clinical course. Most studies reported the use of structured clinical evaluation tools, and one study explored the use of anecdotal notes.

It is interesting to note that all studies in this category examined undergraduate nursing students. Considering that graduate nursing programs, particularly those preparing advanced practice nurses, use preceptors very frequently, research with this population is needed.

Essential Clinical Behaviors

Seven studies aimed to examine or describe essential clinical behaviors, which were defined as behaviors related to safety, behaviors determining passing or failing grades, and descriptive characteristics of students' clinical behaviors. Studies in this category examined student behaviors deemed essential for success in clinical settings or behaviors indicative of student failure.

Two studies conducted in the 1980s were dissertations using quantitative methods; the others were published between 2011 and 2016. Of the published studies, four used qualitative methods and one used mixed methods. Two studies used only students in their sample, four used faculty only, and one used both. All studies focused on essential clinical behaviors of undergraduate students.

The research in this area examined a variety of concepts related to clinical behaviors and student characteristics. Three studies focused specifically on behaviors of failing students or unsafe clinical behaviors and the faculty decision-making regarding clinical evaluation. One study developed and tested a clinical assessment tool consisting of essential clinical behaviors scored with a rubric; the evaluation process in this study included a list of criteria for unsatisfactory performance. The remaining studies examined characteristics of successful and unsuccessful students and differentiating characteristics, the relationship between student personal skills and success, and examination of clinical behaviors evaluated by faculty as essential or desirable. Essential clinical behaviors often included attributes of the students that were not related to skill performance or critical thinking, such as timeliness, receptivity to feedback, and communication skills.

Topic-Based Evaluation

Six studies focused on evaluating students' clinical performance on a specific educational topic, such as empathy. This type of study has been an enduring topic with publications, from 1971 to the present day. All were conducted with undergraduate students.

All of the studies evaluated the effectiveness of a structured educational program on a specific topic in improving students' knowledge, skills, or behaviors. The studies focused on various topics: intrapartum care, empathy, caring, intimate touch, reminiscence therapy, and prevention of occupational exposure to blood-borne pathogens. The interventions were delivered via a variety of methods including web-based education, skills laboratory, formal classes, instructional cards, video, and others. Research designs were commonly two-group intervention and control or one-group quasi-experimental. All of the studies found significant improvement in student outcomes after the intervention. Studies in this category were carefully designed and used outcome measures sensitive to the effects of the intervention.

Decision-Making About Clinical Grade

Studies in this category were focused on the process the educator used in making decisions, that is, internal processes of decision-making versus evaluation of external criteria demonstrated by the student. Three studies used qualitative methods, and one use mixed methods. Three studies were conducted with undergraduate faculty; one study used OSCE examiners for master's level students. None of the studies identified if the final clinical grade for the course was pass/fail or a number grade.

Two studies interviewed nurse educators and asked them to describe a critical incident with a student or the experience of failing a student. Several factors were identified as important in the clinical evaluation process; they included faculty emotions, beliefs about the student, perceptions of cultural differences, and administrative support.

One study conducted focus groups to identify factors that influence examiners' decisions in using the OSCE as an evaluation tool. Common themes included examiner decisions about borderline performance, examiner experience, intuition, and faculty perceptions of student competence, communication ability, and examination approach. More experienced examiners allocated lower global rating scores.

Another study conducted a retrospective analysis of records of students who had failed a clinical course or withdrawn from a nursing program. Identified themes that influenced student failure included personal characteristics and student behaviors, student behaviors related to clinical practice, and faculty responses to student behaviors. The findings from these studies indicate that faculty and student factors play a role in influencing student evaluation methods and outcomes.

Clinical Reasoning

Studies in this category evaluated clinical reasoning skills using the Outcome Present State Test (OPT) Model of Clinical Reasoning. In these studies, critical reasoning was identified as a more specific type of critical thinking (Harmon & Thompson, 2015). Three studies used the OPT model with a quasiexperimental longitudinal approach; students completed the OPT model worksheets over several weeks of a clinical course. These studies found the OPT model to be effective in enhancing clinical reasoning skills, but the improvements approached a plateau after a few weeks.

One study used OPT model worksheets to evaluate the effect of collaborative learning activities on enhancing clinical reasoning skills and found significant increases in student scores; however, final worksheets scores on the OPT model were low. Since its original publication, the OPT model has had little documented use as a method of clinical evaluation and has demonstrated modest effectiveness in improving clinical reasoning skills in undergraduate students.

DISCUSSION AND IMPLICATIONS

This extensive review and analysis of research on clinical evaluation in nursing education revealed several important findings. For nearly five decades, a large number of studies have addressed a large variety of topics associated with clinical evaluation in nursing education. The state of the science about clinical evaluation shows breadth, but not much depth, with few studies building on previous research.

The research was conducted globally, in 18 countries, with the largest percentage of studies (36 percent) conducted in the United States. The largest category was competence, totaling 35 studies (40 percent of the total). Although studies were conducted with students of various levels, most (88 percent) were conducted with undergraduates. Clearly, there is a need for research on the evaluation of graduate students in the clinical setting, especially with regard to the use of preceptors, who are likely to evaluate the students they teach.

Many areas need development in order to advance the science of nursing education research related to clinical evaluation. First, more rigorous study designs are needed. The study designs were evaluated using Melnyk and Fineout-Overholt's (2011) levels of evidence (Table 1). This method of categorization aids in determining the overall rigor of a research study and the usefulness of the findings as evidence to guide practice based on its design, with systematic reviews and meta-analysis of randomized control trials representing the highest level of evidence (level 1) and expert opinion representing the lowest (level 7). Studies focusing only on psychometric testing of instruments were not included in the levels of evidence categorizations.

None of the studies we reviewed for this study were at level 1, the highest level of evidence. Our search found no meta-analyses, and systematic reviews were part of the exclusion criteria. We had hoped to locate groups of studies that would be amenable to meta-analysis. Upon examination of the articles that were retrieved, it was determined that none of the groups of studies could be meta-analyzed with strong quality at this time; many differences in outcome measurement, rating scales, and numbers and types of students and populations prohibited fruitful pooling.

Another goal of the study was to find groups of qualitative studies that lent themselves to qualitative metasynthesis. Only three qualitative studies were on the same topic; of these, two were reports of different aspects of the same study, and the other collected data using a Q-sort methodology instead of in-depth interviews, making the findings impossible to metasynthesize. This speaks to the state of the science — much of our research in nursing education is piecemeal and does not build on earlier work.

Most of the studies in our review were categorized at levels 3, 4, and 6, indicating moderate to low levels of evidence. Only a few studies were randomized controlled trials. Most studies used single setting descriptive designs, with a few using quasiexperimental, case-control, or cohort group designs.

Another large weakness of the body of research is the common use of clinical evaluation instruments that were developed by the study investigator(s) and not tested for validity and reliability. With rare exceptions, evaluation instruments in this study were not used more than once, which lends little evidence to support the use of these instruments in evaluating diverse student populations and settings. Compounding this issue is the lack of replication studies to provide support for the reliability of the research findings.

Additional weaknesses include the relative lack of use of theoretical frameworks to guide the research, with only 28 percent of studies reporting use of a theoretical framework. Most studies were small and conducted in a single site. The lack of multisite studies limits the generalizability of most of the study findings. In addition, all of the studies were conducted in a single program level (associate, bachelor's, or graduate); thus, the efficacy of clinical evaluation instruments for differing levels of students is unknown.

RECOMMENDATIONS FOR FUTURE RESEARCH

Based on the findings of this study, the following recommendations are provided for nursing education and research related to clinical evaluation of nursing students. First, nursing education should explore adopting standardized measures of competence. This can only be accomplished if competence can be consistently defined so that measures can be developed to measure the defined concept. Nursing students are evaluated on a national standardized competency exam (NCLEX) in order to obtain a license to practice nursing. Therefore, nursing education should also adopt valid, reliable, and standardized methods of clinical evaluation. Instruments should be developed and tested that evaluate general and specific competencies that include inquiry, use of evidence-based practice, quality and safety standards, critical thinking, knowledge, professional behaviors, and caring.

In alignment with this recommendation, the National League for Nursing (NLN) Nursing Education Research Priorities include "Build the science of nursing education through the discovery and translation of innovative evidence-based strategies" (NLN, 2016, p. 1). Proposed strategies for achieving this directive include instrument development and testing, use of multisite and multimethod research designs, and use of meta-analysis and metasynthesis to inform the state of the science (NLN, 2016, p. 1). Nursing education needs to build and strengthen the science of nursing as it relates to clinical evaluation. Strategies include evaluating the efficacy of instruments in diverse populations using large multisite studies. The replication of small studies with promising findings should be conducted to provide support for instrument rigor and generalizability; nursing education leaders have called for replication studies to build the science (Morin, 2016). Measures of competence should also be related to patient care outcomes. These strategies will lend evidence toward the standardization of instruments for clinical evaluation.

To aid in the development of replication studies, the authors suggest that a repository of clinical evaluation instruments be created for nursing faculty. This repository should include a description of all instruments that have been tested in clinical evaluation of students, citation of the research that documents and supports their use, and information on how to obtain the instrument for use in the clinical setting. Such a repository would be invaluable to nurse faculty in identifying valid and reliable clinical evaluation instruments that could be used and tested in future research. This first and important step will aid in advancing the science of nursing education and ultimately the profession of nursing.

REFERENCES

Beogo, I., Rojas, B. M., Gagnon, M. P., & Liu, C-Y. (2016). Psychometric evaluation of the French version of the Clinical Nursing Competence Questionnaire (CNCQ-22): A cross-sectional study in nursing education in Burkina Faso. *Nurse Education Today*, *45*, 173-178. doi:10.1016/j.nedt.2016.08.007

Cant, R., McKenna, L., & Cooper, S. (2013). Assessing preregistration nursing students' clinical competence: A systematic review of objective measures. *International Journal of Nursing Practice*, *19*, 163-176. doi:10.1111/ijn.12053

Cassidy, I., Butler, M. P., Quillinan, B., Egan, G., McNamara, M. C., Tuohy, D., … Tierney, C. (2012). Preceptors' views of assessing nursing students using a competency based approach. *Nurse Education in Practice*, *12*, 346-351. doi:10.1016/j.nepr.2012.04.006

Cooper, H. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Los Angeles, CA: Sage.

Duke, M. (1996). Clinical evaluation — Difficulties experienced by sessional clinical teachers of nursing: A qualitative study. *Journal of Advanced Nursing*, *23*(2), 408-414. doi:10.1111/1365-2648.ep8542553

Fan, J. Y., Wang, Y. H., Chao, L. F., Jane, S. W., & Hsu, L. L. (2015). Performance evaluation of nursing students following competency-based education. *Nurse Education Today*, *35*(1), 97-103. doi:10.1016/j.nedt.2014.07.002

Garrett, B. M., MacPhee, M., & Jackson, C. (2013). Evaluation of an eportfolio for the assessment of clinical competence in a baccalaureate nursing program. *Nurse Education Today*, *33*(10), 1207-1213. doi:10.1016/j.nedt.2012.06.015

Hall, M.A., & Thompson, C. (2013). An expanded look at evaluating clinical performance: Faculty use of anecdotal notes in the U.S. and Canada.*Nurse Education in Practice, 13*,271-276. doi:10.1016/j.nepr.2013.02.001

Harmon, M. M., & Thompson, C. (2015). Clinical reasoning in pre-licensure nursing students. *Teaching and Learning in Nursing*, *10*(2), 63-70. doi:10.1016/j.teln.2014.12.001

Hawker, S., Payne, S., Kerr, C., & Hardey, M., Powell, J. (2002). Appraising the evidence: Reviewing disparate data systematically. *Qualitative Health Research*, *12*, 1284-1299. doi:10.1177/1049732302238251

Hsieh, S. I., & Shu, L. L. (2013). An outcome-based evaluation of nursing competency of baccalaureate senior nursing students in Taiwan. *Nurse Education Today*, *33*(12), 1536-1545. doi:10.1016/j.nedt.2013.01.012

Hsu, L. L., & Hsieh, S. I. (2013). Development and psychometric evaluation of the competency inventory for nursing students: A learning outcome perspective. *Nurse Education Today*, *33*(5), 492-497. doi:10.1016/j.nedt.2012.05.028

Krichbaum, K., Rowan, M., Duckett, L., Ryden, M. B., & Savik, K. (1994). The clinical evaluation tool: A measure of the quality of clinical performance of baccalaureate nursing students. *Journal of Nursing Education*, *33*(9), 395-404.

Lee-Hsieh, J., Kao, C., Kuo, C., & Tseng, H. F. (2003). Clinical nursing competence of RN-to-BSN students in a nursing concept-based curriculum in Taiwan. *Journal of Nursing Education*, *42*(12), 536-545.

Melnyk, B. M., & Fineout-Overholt, E. (2011). *Evidence-based practice in nursing and healthcare: A guide to best practice*. Philadelphia, PA: Lippincott, Williams & Wilkins.

Morin, K. H. (2016). Replication: Needed now more than ever. *Journal of Nursing Education*, *55*(8), 423-424.

National League for Nursing. (2016). *NLN research priorities in nursing education 2016-2019*. Retrieved from http://www.nln.org/docs/default-source/professional-development-programs/nln-research-priorities-in-nursing-education-single-pages.pdf?sfvrsn=2

Norman, I. J., Watson, R., Murrells, T., Calman, L., & Redfern, S. (2002). The validity and reliability of methods to assess the competence to practise of pre-registration nursing and midwifery students. *International Journal of Nursing Studies*, *39*(2), 133-145.

Peyman, H., Darash, M., Sadeghifar, J., Yaghoubi, M., Yamani, N., & Alizadeh, M. (2011). Evaluating the viewpoints of nursing and midwifery students about their clinical educational status. *Iranian Journal of Medical Education*, *10*(5), 1-10.

Roberts, D. (2011). Grading the performance of clinical skills: Lessons to be learned from the performing arts. *Nurse Education Today*, *31*, 607-610. doi:10.1016/j.nedt.2010.10.017

Rushforth, H. E. (2007). Objective structured clinical examination (OSCE): Review of literature and implications for nursing education. *Nurse Education Today*, *27*(5), 481-490. doi:10.1016/j.nedt.2006.08.009

Sedgwick, M., Kellett, P., & Kalischuck, R. G. (2014). Exploring the acquisition of entry-to-practice competencies by second-degree nursing students during a preceptorship experience. *Nurse Education Today*, *34*(3), 421-427. doi:10.1016/j.nedt.2013.04.012

Severinsson, E., & Sand, A. (2010). Evaluation of the clinical supervision and professional development of student nurses. *Journal of Nursing Management*, *18*, 669-77.

Skúladóttir, H., & Svavarsdóttir, M. H. (2016). Development and validation of a Clinical Assessment Tool for Nursing Education (CAT-NE). *Nurse Education in Practice*, *20*, 31-318. doi:10.1016/j.nepr.2016.06.008

Walsh, M., Bailey, P. H., & Koren, I. (2009). Objective structured clinical evaluation of clinical competence: An integrative review. *Journal of Advanced Nursing*, *65*(8), 1584-1595. doi:10.1111/j.1365-2648.2009.05054.x

Walsh, T., Jairath, N., Paterson, M. A., & Grandjean, C. (2010). Quality and safety education for nurses clinical evaluation tool. *Journal of Nursing Education*, *49*(9), 517-522. doi:10.3928/01484834-20100630-06

Watson, R., Calman, L., Norman, I., Redfern, S., & Murrells, T. (2002). Assessing clinical competence in student nurses. *Journal of Clinical Nursing*, *11*, 554-555.

Watson, R., Stimpson, A., Topping, A., & Porock, D. (2002). Clinical competence assessment in nursing: A systematic review of the literature. *Journal of Advanced Nursing*, *39*(5), 421-431.

Wood, V. (1982). Evaluation of student nurse clinical performance: A continuing problem. *International Nursing Review*, *29*(1), 11-18.

Wu, X. V., Enskar, K., Lee, C. C., & Wang, W. (2015). A systematic review of clinical assessment for undergraduate nursing students. *Nurse Education Today*, *35*, 347-359. doi:10.1016/j.nedt.2014.11.016

Yanhua, C., & Watson, R. (2011). A review of clinical competence assessment in nursing. *Nurse Education Today*, *31*, 832-836.

CE

Instructions:

- Read the article. The test for this CE activity can only be taken online at www.NursingCenter.com/CE/NEP. You will need to create (its free!) and login to your personal CE Planner account before taking online tests. Your planner will keep track of all your Lippincott Professional Development online CE activities for you.
- There is only one correct answer for each question. A passing score for this test is 13 correct answers. If you pass, you can print your certificate of earned contact hours and access the answer key. If you fail, you have the option of taking the test again at no additional cost.
- For questions, contact Lippincott Professional Development; 1-800-787-8985.

Registration Deadline: December 3, 2021.

Disclosure Statement:
The authors and planners have disclosed that they have no financial relationships related to this article.

Provider Accreditation:
Lippincott Professional Development will award 1.5 contact hours for this continuing nursing education activity.

Lippincott Professional Development is accredited as a provider of continuing nursing education by the American Nurses Credentialing Center's Commission on Accreditation.

This activity is also provider approved by the California Board of Registered Nursing, Provider Number CEP 11749 for 1.5 contact hours. Lippincott Professional Development is also an approved provider of continuing nursing education by the District of Columbia, Georgia, and Florida, CE Broker #50-1223.

Payment:

- The registration fee for this test is $17.95.

For more than 169 additional continuing education articles related to Education topics, go to NursingCenter.com/CE.