Nurse Education Today 64 (2018) 108–114

Contents lists available at ScienceDirect

Nurse Education Today

journal homepage: www.elsevier.com/locate/nedt

Validation of time to task performance assessment method in simulation: A comparative design study☆

Mary Ann Shinnick*, Mary A. Woo

University of California at Los Angeles, United States

ARTICLE INFO

Keywords:
Simulation-based assessment
Evaluation
Nursing education
Clinical competency
High stakes evaluation
Time to task
Lasater
Eye tracking

ABSTRACT

Background: There is a lack of objective and valid measures for assessing nursing clinical competence which could adversely impact patient safety. Therefore, we evaluated an objective assessment of clinical competence, Time to Task (ability to perform specific, critical nursing care activities within 5 min), and compared it to two subjective measures, (Lasater Clinical Judgement Rubric [LCJR] and common "pass/fail" assessment).

Design/Methods: Using a prospective, "Known Groups" (Expert vs. Novice nurses) comparative design, Expert nurses (ICU nurses with > 5 years of ICU experience) and Novice nurses (senior prelicensure nursing students) participated individually in a simulation of a patient in decompensated heart failure. Fourteen nursing instructors or preceptors, blinded to group assignment, reviewed 28 simulation videos (15 Expert and 13 Novice) and scored them using the LCJR and pass/fail assessments. Time to Task assessment was scored based on time thresholds for specific nursing actions prospectively set by an expert clinical panel. Statistical analysis consisted of Medians Test and sensitivity and specificity analyses.

Results: The LCJR total score was significantly different between Experts and Novices ($p < 0.01$) and revealed adequate sensitivity (ability to correctly identify "Expert" nurses; 0.72) but had a low specificity (ability to correctly identify "Novice" nurses; 0.40). For the subjective measure 'pass/fail', sensitivity was high (0.90) but specificity was low (0.47). The Time to Task measure had statistical significance between Expert and Novice groups ($p < 0.01$) and sensitivity (0.80) and specificity (0.85) were good.

Conclusion: Commonly used subjective measures of clinical nursing competence have difficulties with achieving acceptable specificity. However, an objective measure, Time to Task, had good sensitivity and specificity in differentiating between groups. While more than one assessment instrument should be used to determine nurse competency, an objective measure, such as Time to Task, warrants further study.

1. Background

Establishing clinical competence of health care practitioners is essential to patient safety. This is imperative as premature deaths linked to preventable causes in United States (US) hospitals has increased to over 400,000 annually (James, 2013). Competence is defined in many ways. It has been described as "the ability to perform a task with desirable outcomes" (Benner, 1982 p. 303) and includes knowledge, skills, performance, attitudes and values (Cowan et al., 2007). Current methods of evaluation of clinical competence in nursing education is by knowledge and/or skill tests and subjective opinion (subjective evaluation by clinical preceptors/instructors) which repeatedly have been shown to be invalid and unreliable (Oermann et al., 2009; Watling and Lingard, 2012). As many clinical or academic faculty have never had training in assessment (Rizzolo et al., 2015), these subjective assessments have problems with interrater reliability and validity (Bensfield et al., 2012) and lack standardization between and within institutions.

Competency assessment is often dependent on the evaluator's experience, perception, knowledge and training. However, most evaluations of a nurse by a preceptor in the clinical setting relies on the preceptor's recall of performance and consists predominately of subjective grading. Not surprisingly, analyses of competency vary among instructors (Yanhua and Watson, 2011). Unfortunately, many times nurses are selected to be preceptors by their managers based upon their experience as a nurse, but not necessarily upon their knowledge or skills as an educator or clinical evaluator (Hemman et al., 2007). Many institutions provide some training for nurses assigned as preceptors but there is scant data or standardization between or even within institutions on this instruction. A survey of US Intensive Care Units found 80%

☆ This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
* Corresponding author at: 700 Tiverton Room 3-956, Los Angeles, CA 90095, United States
E-mail address: mshinnic@sonnet.ucla.edu (M.A. Shinnick).

https://doi.org/10.1016/j.nedt.2018.02.011
Received 18 October 2017; Received in revised form 13 January 2018; Accepted 6 February 2018

provided formalized preceptor training program for an average of 6 classroom hours (Thomason, 2006). Training topics included "professional behavior, knowledge base, completion of competencies, and organizational skill", but did not address "how to" assess or evaluate clinical competency. Patient safety usually is not mentioned and there is no mention of a valid and reliable instrument or clinical standard by which to assess clinical competence. This is a vital omission as the preceptor is the last "quality control measure to ensure those nurses who are about to enter the professional work environment are competent to do so" (Earle-Foley et al., 2012 pg. 28).

Objective, valid and reliable instruments which are easy to use with little to no formal training of raters would be an asset for assessment. Currently, there are barriers to interrater reliability with existing instruments in part due to multiple training sessions required to use currently available tools and the difficulty in scoring participants in situations that include complex, time-sensitive, sequential actions (Boulet et al., 2008). The US National League of Nursing (NLN) recommends valid and reliable instruments be used to assess nurses as well as multiple approaches for assessment of knowledge and clinical abilities, especially with high-stakes assessment (National League of Nursing, 2012).

Therefore, we tested an objective assessment of clinical competence, Time to Task (ability to perform specific, critical nursing care activities within a set time period/5 min), and compared it to two commonly employed assessments: Lasater Clinical Judgement Rubric (LCJR) (Lasater, 2011), used frequently in simulation formative assessment, and the common 'pass/fail' assessment (subjective measure of clinical competence) used in clinical assessment of nursing students/staff nurses. The hypothesis was that the Time to Task assessment would be able to differentiate between two known groups (Expert and Novice nurses) and have acceptable sensitivity (ability to identify expert/experienced nurses) and specificity (ability to identify novice/student nurses).

2. Methods

2.1. Sample

Using a prospective, 'known groups' (Expert vs. Novice nurses) comparative research design, Expert nurses (ICU or Emergency Department nurses with > 5 years of clinical experience) and Novice nurses (senior prelicensure nursing students nearing graduation) participated individually in a manikin-based simulation of a patient with decompensated heart failure (HF). Expert nurses were recruited from two different institutions, a large tertiary hospital and a smaller community hospital. Novice nurses were recruited from one School of Nursing. Students were recruited in a class not taught by the researchers and nurses were recruited by blanket e-mail to nursing staff. All research activity was done outside of work and school time and there was no grade impact or extra credit given for study participation. Each participant received a $25 gift card for their study participation. The Principal Investigator was not blinded to group membership.

Inclusion criteria for Expert nurses was employment as an ICU or Emergency Department nurse of ≥ 5 years. Inclusion criteria for Novice nurses was nursing student status in their senior prelicensure year who had successfully completed coursework in the care of decompensated heart failure (HF).

For video Reviewers, inclusion criteria were experience in evaluating prelicensure nursing students in the clinical setting as either a faculty member at a college of nursing or a clinical preceptor who works with and assesses nursing students. The study had University of California at Los Angeles Institutional Review Board approval.

2.2. Data Collection Instruments

2.2.1. Video Capture

Video of each participant's point of view was achieved with eye tracking glasses (ETG; SensoMotoric Instruments version 2.7 [Teltow, Germany]) ETG have excellent audio and video (24 Hz audio; 1280 × 960 video resolution) without traditional blind spots seen with static cameras. As they are worn on the face, the resultant video includes only the simulation area from the wearer's point of view and the video Reviewer cannot identify the participant by sight. In other words, a viewer cannot see the subject at all. While audio recording is clear, none of the Reviewers knew the subjects in this study. This method of video recording was chosen based on earlier research reporting problems with traditional video capture in high stakes testing (Forbes et al., 2016; Rizzolo et al., 2015).

2.2.2. Simulation Design

Simulation of decompensated HF is important in the development of proficient nurses, as HF is a common hospital discharge diagnosis (Schocken et al., 2008). The study simulation depicts an adult male admitted for dyspnea. During the simulation, the patient complains of increasing dyspnea and physical exam reveals pulmonary crackles and symptoms do not improve until both oxygen and furosemide have been given. This simulation was content-validated by nurses and a cardiologist who are HF experts (Shinnick et al., 2011). The facilitator of each simulation was a trained and certified healthcare simulation educator (CHSE).

2.2.3. Demographic Questionnaire

A questionnaire including the participant's age, gender, history of personal or family HF experience, prior simulation exposure, years and type of employment and previous employment of the nursing students as a nurse helper (i.e., nurse's aide, care partner, etc.) was collected after the simulation event.

2.2.4. Post Simulation Questionnaire

A 5-item, multiple choice format questionnaire was administered after the simulation to evaluate HF knowledge. It addressed nursing interventions applicable to care of a patient who was dyspneic that would apply to the HF patient but HF was not explicitly identified. This questionnaire was content validated by expert nurses in HF.

2.3. Assessments Completed by Video Reviewers

2.3.1. Lasater Clinical Judgement Rubric

LCJR was chosen as it was designed as an objective tool to evaluate clinical judgment of a student or licensed nurse in a single simulation or clinical episode. It has been used in several simulation studies which reported content and construct validity and faculty were able to consistently and accurately recognize the correct level of student competence (senior or junior nursing students) (Gubrud-Howe, 2008; Lasater, 2011; Sideras, 2007). Reliability from three different LCJR studies is reported as 57%–100% (Adamson et al., 2012; Hallin et al., 2016).

LCJR measures an array of proficiency in four phases based on Tanner's Clinical Judgment Model and range from Exemplary to Novice (Lasater, 2007). It has a total of 11 domains. Each domain is scored as either Exemplary (4 points), Accomplished (3 points), Developing (2 points) or Beginning (1 point) for a total possible score of 44 points (Table 1). The instrument is lauded as a valid and reliable instrument in assessing a student's clinical judgment skill in a simulation but some find it cumbersome to use (Prion et al., 2017). Reliability results are affected by the differing characteristics of individual simulations as well as any instrument instruction given to Reviewers, which vary among institutions as well (Adamson et al., 2012). Also, not all domains pertain to every simulation. In this study, 4 domains ("information seeking" "clear communication", "evaluation/self-analysis" and