F A

A U R P S E I

K A A , P D, RN *, S K -E , P D, RN , J , MSN, RN

[a]University of Washington Tacoma, Tacoma, WA 98402-3100, USA
[b]Boise State University, WA, USA
[c]Washington State University, Spokane, WA 99210, USA

KE ORDS
evaluation;
instrument;
simulation

A : Interest in simulation as a teaching and evaluation strategy in nursing education continues to grow. Mirroring this growth, we have seen a proliferation of instruments designed to evaluate simulation participant performance. This article describes two frameworks for categorizing simulation evaluation strategies and provides a review of recent simulation evaluation instruments. The review focuses on four instruments that have been used extensively in the literature, objective structured clinical examinations (OSCE's) including four OSCE instruments, and an extensive list of new instruments for simulation evaluation.

C :
Adamson, K. A., Kardong-Edgren, S., & Willhaus, J. (2013, September). An updated review of published simulation evaluation instruments. *Clinical Simulation in Nursing, 9*(9), e393-e400. http://dx.doi.org/10.1016/j.ecns.2012.09.004

Simulation use continues to grow and develop in nursing and other programs educating health care providers around the world. DeVita (2009) argues that simulation should be a core educational strategy because it is "measurable, focused, reproducible, mass producible, and importantly, very memorable" (p. 46). Both the National Council of State Boards of Nursing and the National League for Nursing are conducting research about the use of simulation as a teaching and evaluation method (Hayden, 2011; Rizzolo, Oermann, Jeffries, & Kardong-Edgren, 2011). However, Tanner (2011) recently noted how "little investment there has been in developing suitable measures for the assessment of learning outcomes, particularly those relevant for a practice discipline" (p. 491). Recent reviews of the literature in nursing (Davis & Kimble, 2011; Yuan, Williams, Fang, & Ye, 2012), pharmacy (Bray, Schwartz, Odegard, Hammer, & Seybert, 2011), and medicine (Kogan, Holmboe, & Hauer, 2009) echo a continued quest for meaningful ways to evaluate participants in simulation activities.

In response to repeated requests for an updated and expanded list of evaluation instruments, this article provides a follow-up to the original instrument review article (Kardong-Edgren, Adamson, & Fitzgerald, 2010). The purposes for this article include (a) discussing existing frameworks for categorizing simulation evaluation strategies and (b) using an adaptation of these frameworks to provide the following:

1. An update on four instruments from our original review that have been cited extensively in the literature

* Corresponding author. kadamson@u.washington.edu (K. A. Adamson).

2. A review of objective structured clinical examinations (OSCEs), including the development of four OSCE instruments in undergraduate nursing education
3. A report on instruments that are either new or were not included in the original instrument review article (Kardong-Edgren, et al., 2010) and that are appropriate for simulation evaluation

FOCUS

Two useful frameworks that have emerged to categorize various evaluation strategies are translational science research (TSR; McGaghie, Draycott, Dunn, Lopez, & Stefanidis, 2011) and Kirkpatrick's (1994) levels of evaluation. The following is a brief overview of these frameworks, which will be used to categorize instruments in following sections.

TSR

The National Institutes of Health (2011, 2012) describe translational research as a continuum on which scientific discoveries move from preclinical (or bench) research to practical applications in patient care at the bedside and ultimately affect health care outcomes. In short, TSR can be thought of as research that takes new knowledge from "bench to bedside and beyond." Nomenclature in the field of TSR is somewhat contested (Woolf, 2008). However, the concept is highly applicable to simulation evaluation research. For the purposes of this article, we are adopting the overview provided by Dougherty and Conway (2008) and applied to simulation evaluation by McGaghie et al. (2011).

Translation Phase 1 designates preclinical activities (Woolf, 2008) that are meant to assess the efficacy of care. Relating this to simulation, we might say that this level of research demonstrates, in the simulation lab, whether students have learned something. Translation Phase 2 designates activities meant to assess who benefits from care. Relating this to simulation, we might say that these activities demonstrate whether what students learned in the simulation lab carries over to the actual patient care setting. Finally, Translation Phase 3 designates activities that are meant to assess whether improved care yields improved outcomes in the broader health care arena. Relating this to simulation, we might say that these activities demonstrate whether what was learned in the simulation lab and demonstrated in the patient care setting results in improved health outcomes. Phases 1 to 3 help describe the quality and applicability of simulation evaluation activities, with Phase 3 activities being the pinnacle of research because they describe how simulation affects health outcomes.

KIRKPATRICK'S LEVELS

In a similar fashion, Kirkpatrick's (1994) four levels of evaluation are helpful in describing what type of evidence different simulation evaluation strategies produce. The four levels, reaction, learning, behavior, and outcomes, are described in Figure 1, using language from Boulet Jeffries, Hatala, Korndorffer, Feinstein, & Roche, (2011, p. S50), along with the corresponding TSR nomenclature. In this combination of the TSR and Kirkpatrick frameworks for describing types of simulation evaluation evidence, learning at Level 2 (Translation Phase 1) may be subdivided into affective, cognitive, and psychomotor learning. Also, Kirkpatrick's Level 1, reaction, is not applicable to translational research.

RESEARCH

infarction, and a chest wound (Grant, Moss, Epps, & Watts, 2010). Grant et al. (2010) report interrater reliability findings from their modification of the CSET. The LCJR (Lasater, 2007) has been used for a variety of purposes, including debriefing (Mariani, Cantrell, Meakim, Prieto, & Dreifuerst, in press) and evaluation of technical skills such as IV insertion (Reinhardt, Mullins, De Blieck, & Schultz, 2012). Furthermore, Adamson, Gubrud, Sideras, and Lasater (2012) reported extensive reliability and validity findings from a range of studies used to assess the psychometric properties of the LCJR. Finally, the C-SEI originally developed and published by Todd et al. (2008)

T . 1 Updates on Instruments from Original Review Article

Articles: Original; Subsequent Publications	Related to the Instrument	Instrument	Reliability and Validity	Kirkpatrick and TSR	Special Notes
Original article, Clark, 2006, p. e76					

N P U I

Since the publication of our original review article, there has been a sharp increase in new simulation evaluation instruments in the literature (Kardong-Edgren et al., 2010). A sampling of these instruments and citations for the articles that cited them are included as Table 3 (view online extra at www.nursingsimulation.org). Several trends and other noteworthy information in the table deserve mention here.

Two articles cited in the table used the Spielberger State-Trait Anxiety Inventory to evaluate participant anxiety related to simulation activities (Gantt, in press; Gore, Hunt, Parker, & Raines, 2011). This represents an interesting exploration of the reactions of participants and the authenticity of their emotional responses related to simulated patient encounters. Additional research is under way about the biological markers related to stress and anxiety experienced by participants in simulation.

The National League for Nursing's Simulation Design and Student Satisfaction and Self-Con dence in Learning scales (Jeffries & Rizzolo, 2006) continue to be popular (Adamson, in press; Prentice, Taplay, Horsley, Payeur-Grenier, & Delford, 2011; Swenty & Eggleston, 2011). These, like most simulation evaluation instruments, focus on low-level learner reaction and learning (Kirkpatrick's Levels 1 and 2 and TSR Phase 1). Within the category of learning, most evaluation instruments focus on cognitive learning. This is disappointing because these low levels of evaluation may not re ect the effects simulation training has on the most important stakeholders in health care edu- of evtrainin tion inst6 0 heenm5(parin1(simul)er)s

T . 2 Objective Structured Clinical Examinations

Articles, Original and Subsequent	Instrument	Reliability and Validity

educators are selecting an instrument for use in performance evaluation, it is not enough to select a tool from a list with high marks reported in reliability and validity. It is important to consider whether the instrument is appropriate for the population and the activity to which it is being applied.

In research, when an instrument is used in a new population or for a measurement purpose different from what was originally intended, the researchers should report the process and statistics associated with validating the instrument for the new purpose. Using an instrument to evaluate populations and purposes beyond the original intent is like trying to measure a cup of milk with a yardstick. Although it is possible, without accurate knowledge about the vessel for the liquid, it would be dif cult to determine whether the amount of milk really equaled 1 cup. Before an instrument is used to evaluate student performance, consideration must be given to whether it is a valid and reliable measure for that population of participants and raters. Care should be taken to report any steps taken, such as a pilot project or content expert review.

Examples in the literature in which these kind of activities are demonstrated include a study by Reinhardt et al. (2012) in which the LCJR was adapted from its original purpose (evaluating clinical judgment) for evaluation of a clinical skill. Another example used instruments originally designed for undergraduate nursing students to evaluate an interprofessional simulation experience among already licensed health care professionals (Prentice et al., 2011). Although the report details the instruments' original reliability, it does not discuss how the Simulation Design Scale, the Educational Practices in Simulation Scale, and the Self-Con dence in Learning Scale were evaluated for use with this new population.

Finally, researchers should consider their options for evaluation in light of the TSR and Kirkpatrick frameworks for categorizing simulation evaluation. The literature is saturated with reports of low-level participant evaluations, including reaction (Kirkpatrick's Level 1). It is time to step up and focus on what really matters: how simulation affects learning, behaviors, and ultimately patient outcomes.

Conclusion

Researchers can assist the continued maturation of the simulation pedagogy by aspiring to higher levels of evaluation and reporting psychometric measures and steps taken to assure validation with new populations. This report included instruments developed in several countries. Sharing the results of study replication from different cultural and international environments is an essential part of the further development of valid and reliable measures for simulation instruments. Replication studies using existing instruments with new populations and venues will be part of the process to turn tentative belief into accepted knowledge. Replications help further establish reliability, validity, and practice (Haller & Reynolds, 1986).

References

Ackermann, A. D. (2009). Investigation of learning outcomes for the acquisition and retention of CPR knowledge and skills learned with the use of high- delity simulation. *Clinical Simulation in Nursing*, *5*(6), e213-e222 http://dx.doi.org/10.1016/j.ecns.2009.05.002.

Adamson, K. A. (2012). Piloting a method for comparing two experiential teaching strategies. *Clinical Simulation in Nursing*, *8*

Hemman, E. A., Gillingham, D., Allison, N., & Adams, R. (2007). Evaluation of the combat medic skills test. *Military Medicine*, *172*(8), 843-851.

Hutton, M., Coben, D., Hall, C., Rowe, D., Sabin, M., Weeks, K., et al. (2010). Numeracy for nursing, report of a pilot study to compare outcomes of two practical simulation tools—An online medication dosage assessment and practical assessment in the style of objective structured clinical examination. *Nurse Education Today*, *30*, 608-614. http://dx.doi.org/10.1016/j.nedt.2009.12.009.

Jeffries, P., & Rizzolo, M. (2006). Designing and implementing models for the innovative use of simulation to teach nursing care of ill adults and children: A national multi-site study. Retrieved October 8, 2012, from. http://www.nln.org/researchgrants/LaerdalReport.pdf

Kardong-Edgren, S., & Adamson, K. A. (2009). BSN medical-surgical student ability to perform CPR in a simulation: Recommendations and implications. *Clinical Simulation in Nursing*, *5*(2), e79-e83. http://dx.doi.org/10.1016/j.ecns.2009.01.006.

Kardong-Edgren, S., Adamson, K., & Fitzgerald, C. (2010). A review of currently published evaluation instruments for human patient simulation. *Clinical Simulation in Nursing*, *6*(1), e25-e35. http://dx.doi.org/10.1016/jecns.2009.08.004.

Kirkpatrick, D. L. (1994). *Evaluating training programs: The four levels*. San Francisco, CA: Bernett-Koehler.

Klakovich, M. D., & Dela Cruz, F. A. (2006). Validating the interpersonal communication assessment scale. *Journal of Professional Nursing*, *22*(1), 60-67. http://dx.doi.org/10.1016/j.profnurs.2005.12.005.

Kogan, J. R., Holmboe, E. S., & Hauer, K. E. (2009). Tools for direct observation and assessment of clinical skills of medical trainees. *Journal of the American Medical Association*, *302*(12), 1316-1326. http://dx.doi.org/10.1001/jama.2009.1365.

Lasater, K. (2007). Clinical judgment using simulation to create an assessment rubric. *Journal of Nursing Education*, *46*(11), 496-503.

Levett-Jones, T., McCoy, M., Lapkin, S., Noble, D., Hoffman, K., Dempsey, J., et al. (2011). The development and psychometric testing of the Satisfaction with Simulation Experience Scale. *Nurse Education Today*, *31*, 705-710, doi 10.1016.j.nedt.2011.01.004

Liaw, Y. S., Sherpbier, A., Rethans, J.-J., & Klainin-Yobas, P. (2012). Assessment of simulation learning outcomes: A comparison of knowldege and self-reported con dence with observed clinical performance. *Nurse Education Today*, *32*(6), e35-e39. http://dx.doi.org/10.1016/j.nedt.2011.10.006.

Manz, J. A., Hercinger, M., Todd, M., Hawkins, K., & Parsons, M. E. (in press). Improving consistency of assessment of student performance during simulated experiences. *Clinical Simulation in Nur70,-cT103-7aslon-6-710,*

28iStructur94s2-32dente55ct-253opment6fofof c3inicale224asse496bilitie-22944E-31715ddolicSe