2014-11-26

High-Stakes Evaluation: Five Years Later

Tonya Rutherford-Hemming

Suzan Kardong-Edgren
Boise State University

Teresa Gore
Auburn University

Patricia K. Ravert
Brigham Young University - Provo, patricia-ravert@byu.edu

Mary Anne Rizzolo

Featured Article

High-Stakes Evaluation: Five Years Later

Tonya Rutherford-Hemming, EdD, RN, ANP-BC, CHSE[a,*],
Suzan Kardong-Edgren, PhD, RN, ANEF, CHSE[b],
Teresa Gore, DNP, FNP-BC, NP-C, CHSE-A[c],
Patricia Ravert, PhD, RN, CNE, ANEF, FAAN[d],
Mary Anne Rizzolo, EdD, RN, FAAN, ANEF[e]

[a]*Senior Nurse Researcher, Office of Nursing Research and Innovation, Cleveland Clinic, Cleveland, OH 44195, USA*
[b]*Research Associate Professor, Jody DeMeyer Endowed Chair in Nursing, Boise State University, Boise, ID 83725, USA*
[c]*Associate Professor, Simulation Learning Coordinator, Auburn University, Auburn, AL 36849, USA*
[d]*Dean and Professor, Brigham Young University, Provo, UT 84602, USA*
[e]*Consultant, National League for Nursing, Washington, DC 20037, USA*

Abstract: This article compares the International Nursing Association of Clinical Simulation and Learning (INACSL) memberships' thoughts and discussions about the use of simulation for high-stakes evaluation in nursing education to a similar town hall discussion, five years ago. Data on the topic of high-stakes testing were collected through a survey before the INACSL conference in June 2014. During a town hall meeting at the conference, the survey data was presented, attendees shared their thoughts and reacted to prepared comments by selected simulation leaders. Half of the town hall attendees favored high stakes testing before the survey findings and discussion. This number dropped to a third of the attendees by the end of the town hall meeting, by straw poll. The survey findings, a summary of the town hall discussion and final thoughts from simulation leaders are shared in this article.

Cite this article:
Rutherford-Hemming, T., Kardong-Edgren, S., Gore, T., Ravert, P., & Rizzolo, M. A. (2014, December). High-stakes evaluation: Five years later. *Clinical Simulation in Nursing, 10*(12), 605-610. http://dx.doi.org/10.1016/j.ecns.2014.09.009.

At the 2009 International Nursing Association for Clinical Simulation and Learning (INACSL) conference, a round table discussion and a town hall meeting were held regarding the use of high-stakes testing in simulation. Prompted by discussion of the topic on the 2009 INACSL Listserv, five simulation educators led the round table discussion at that year's conference. Three key ideas came from that discussion: (a) the need to think of testing in a different way other than a 75-question multiple-choice examination for licensure, (b) the concern that high-stakes testing might negatively impact learning secondary to participant anxiety, and (c) questions of how to best prepare educators to implement high-stakes testing

* Corresponding author: ruthert@ccf.org (T. Rutherford-Hemming).

(Kardong-Edgren, Hanberg, Keenan, Ackerman, & Chambers, 2011).

Since that 2009 round table discussion, simulation has proliferated in use; 87% of nursing programs in the United States utilize simulation to some extent (Hayden, 2010). The use of simulation for high-stakes evaluation continues to evolve in both academia and practice settings. The term high-stakes evaluation is defined as

> "An evaluation process associated with a simulation activity that has a major academic, education, or employment consequence (such as a grading decision, including pass or fail implications; a decision regarding competency, merit pay, promotion, or certification)." High-stakes refers to the outcome or consequences of the process (Meakim et al., 2013, p. S7).

In the last 5 years, two landmark studies have been completed or are nearing completion related to simulation and high-stakes testing, the National Council of State Boards of Nursing (NCSBN) National Simulation Study (2014) and the National League for Nursing (NLN) High-stakes Simulation Project (2010). High-stakes testing continues to be used in medicine and dentistry in the United States (Hofer, Schuebel, Sader, & Landes, 2013; McGaghie, Issenberg, Petrusa, & Scalese, 2010) and is gaining respect and momentum in other countries (Harvey & Radomski, 2011; McIndoe, 2013). Several authors have suggested that simulation be used as an adjunct measurement to assess clinical competency of nurses (Decker, Utterback, Thomas, Mitchell, & Sportsman, 2011) and reported that high-stake testing allows students who do not demonstrate proficient

competency to be identified, remediated, and given individualized learning plans (Bensfield, Olech, & Horsley, 2012).

The use of high-stakes evaluation in medicine is completed using a process known as the objective structured clinical examination (OSCE) or "… short skills based simulations, usually evaluating different psychomotor and lower level cognitive skills …" (Meakim et al., 2013, p. S6). McWilliam and Botwinski (2012) identified several strengths and weaknesses associated with using the OSCE in a nursing program. These include (a) compromised data and the need to assign individuals to manage data collection, (b) inconsistencies in "patients" performances and the need to standardize patient training and include an evaluation of standardized patient performance, and (c) measurement instruments being amended by faculty poorly educated in high-stakes testing resulting in the loss of validation of the original instrument for new purposes. One of the biggest issues reported in the literature is the need to establish consistent inter-rater reliability in simulations for high-stakes testing (Boulet et al., 2011; Boulet & Murray, 2010; Feldman, Lazzara, Vanderbilt, & DiazGranados, 2012; Nunnink et al., 2014; Wind & Engelhard, 2012).

In light of these known problems and the increasing use of simulation in nursing, the INACSL conference planning committee decided to revisit the topic of high-stakes testing at the INACSL conference in 2014. Survey research was completed in the months before the conference to gain insight into where simulation educators and members stood on using simulation for high-stakes evaluation and to guide the town hall meeting forum at the conference.

Survey and Results

Institutional Review Board approval from Boise State University was received. The survey was sent by e-mail to all INACSL members, and a notice about the survey was placed on the NLN Simulation Innovation Resource Center (2014) Web site homepage so that non-INACSL simulation educators could also participate.

The survey consisted of nine questions. The survey was made available for almost 3 weeks; 605 individuals responded to the online survey. Table 1 outlines the results of the survey. Two hundred sixty-three individuals (43%) stated that they are currently using manikins or standardized patients for high-stakes evaluation. Respondents indicated a number of areas in which testing is occurring. Thirty-eight percent indicated that high-stakes evaluation is used at the end of clinical courses; 31% reported using high-stakes evaluation at the end of

Table 1 Survey Questionnaire

Question	Options	Frequency of Responses (%)
1. What region of the country are you in?	North Atlantic	120 (20)
	Midwest	144 (24)
	South	215 (36)
	West	91 (15)
	US Territories	2 (0)
	Outside United States	33 (5)
2. What type of program are you working in?	AD	148 (35)
	BSN	179 (42)
	MS	49 (12)
	Diploma	0 (0)
	Hospital	45 (11)
	Other	72 (17)
3. Are you currently using simulation with manikins or standardized patients for high stakes assessment?	Yes	263 (43)
	No	342 (57)
4. If yes, where in the curriculum/hospital educational time?	End of clinical courses	107 (38)
	End of fundamentals courses	89 (31)
	Other	155 (54)
5. How long have you (person filling out the questionnaire) been using high fidelity simulation?	0-2 years	136 (23)
	3-5 years	221 (37)
	6-8 years	157 (26)
	More than 8 years	82 (14)
6. Are you currently using OSCEs for high-stakes assessment?	Yes	166 (27)
	No	440 (73)
7. If yes, where in the curriculum/hospital training or maintenance of skills?	End of clinical courses	67 (35)
	End of fundamentals courses	60 (31)
	End of program/residency	29 (15)
	Other	84 (44)
8. If no, are there discussions going on with your faculty/educators about using	Simulations with manikins of standardized patients for high-stakes testing	109 (23)
	OSCEs for high-stakes testing	19 (4)
	Both	131 (27)
	Neither	225 (46)
9. Are you a member in INACSL?	Yes	356 (59)
	No	249 (41)

Note. INACSL = International Nursing Association for Clinical Simulation and Learning; OSCE = objective structured clinical examination.

fundamental courses. Respondents also reported using high-stakes evaluation at multiple and/or variable times in the program, throughout the clinical courses, or throughout the entire program. Only 27% of respondents said that they are currently using OSCEs for high-stakes evaluation; 46% of respondents indicated that there are no current discussions about high-stakes testing occurring in their programs.

Table 2 Benefits and Challenges to Using High-Stakes Testing in a Nursing Curriculum

Benefits	Challenges
• Provides another means of assessment • May highlight a gap in learning that is not seen with pencil-and-paper evaluation	• Valid and reliable scenarios • Inter-rater reliability • Large number of nursing students • Instructor bias

Town Hall Meeting during the International Nursing Association for Clinical Simulation and Learning Conference, 2014

A straw poll of the audience was taken after the survey results were presented; approximately 50% of attendees indicated that they were in favor of high-stakes testing. Patricia Ravert, Dean of the Brigham Young University School of Nursing, presented a dean's perspective of preparing to defend the outcomes of high-stakes testing. She provided a brief overview of the benefits and challenges to using high-stakes testing in a nursing curriculum (Table 2).

The discussion was then opened to the audience. Approximately 20 attendees spoke about their use or planned use of high-stakes testing using simulation. Several people spoke about the challenges of standardizing simulations that would be valid and reliable to use as part of the National Council Licensure Examination (NCLEX). One person questioned whether nursing faculty are ready to validate simulation scenarios when many faculty still struggle to write valid and reliable paper-and-pencil test questions. A few attendees raised concerns related to the feasibility of having schools include high-stakes evaluation in their criteria for graduation, noting the time it currently takes to have students complete simulations in groups and the increased time and faculty resources that would be required to test students individually. Furthermore, if high-stakes testing with simulation was made a part of NCLEX, an attendee pointed out that students would need to have repeated exposure in simulation to feel prepared for this type of evaluation, again raising questions related to adequate faculty resources and time for this endeavor.

Faculty bias in evaluating students was another concern raised by some, especially because nursing schools are often rated according to their NCLEX pass rate. Several attendees suggested that if high-stakes testing were to become a part of the graduation or NCLEX process, these scenarios would need to be evaluated by independent observers and not with faculty who were familiar with the student participants. Simulation testing centers or agencies might need to be developed to counter issues surrounding evaluation. One attendee suggested that faculty often equate simulations with skills assessments with students, suggesting that nursing faculty may not be discussing the same thing when they say they are "doing simulations." Another attendee encouraged the audience to think about the quality of simulations for low-stakes testing before negotiating ideas of high-stakes evaluation.

Throughout the open discussion, other attendees spoke about the value and benefit of using high-stakes testing with simulation. Several attendees voiced the need to test students' knowledge and abilities in other ways than merely a pencil-and-paper examination. Two attendees stated that simulation gives faculty more insight into students' knowledge and understanding related to patient concepts.

Audience participants also spoke about the use of OSCEs as part of high-stakes testing. One attendee who has used OSCEs at her facility stated that OSCEs highlight student comprehension better than grade point averages. Another attendee shared that OSCEs were previously used as part of nursing licensure in Canada but were discontinued because of the expense involved. The attendee stated that there was a high correlation between students who did not pass the OSCE and students who did not pass the board examination. A faculty member from the University of Michigan spoke about the success her school has found by using OSCEs to ensure competency before students entering the clinical area. While acknowledging that students are often anxious about this process, students are also more prepared when they enter the clinical setting. Other attendees acknowledged the anxiety that students often feel in simulation but reasoned that anxiety is likely to come with any type of high-stakes testing and pointed out that the current NCLEX is a high-stakes test as well. An attendee whose school participated in the NCSBN research study voiced her confidence that students can do well on OSCEs in high-stakes evaluation.

After a lively 20-minute audience discussion, the town hall meeting concluded with a second straw vote, asking attendees again, if they were in favor of high-stakes testing using simulation. Approximately 30% of the audience indicated that they were in favor of such evaluation, a loss of 20% support, after discussion. This suggests that additional insights into the actual difficulties associated with conducting high-stakes testing were provided by the town hall meeting. We noticed much less emotion in this discussion than five years ago and more thoughtful analysis about the actual benefits and pitfalls of high-stakes testing.

Where Do We Go from Here?

Dialog from the town hall meeting suggests that nurse simulationists (nurses who use simulation) are now more sophisticated in their awareness of the difficulties involved with the actual preparation of reliable and valid scenarios for high-stakes testing, the necessity of using trained evaluators who have been tested for inter-rater reliability, and the need to provide standardized testing experiences. The drop in the straw poll after discussion suggests a lack of awareness about the need to be prepared to legally defend scenarios, the need to provide evaluator preparation, and the need to promote rigorous research methodologies for high-stakes testing. The change in the polling is reassuring, demonstrating that faculty want to prepare for high-stakes testing in a responsible manner.

Final Thoughts from Three High-Stakes Experts

Patricia Ravert, PhD, RN, is the Dean and Professor at Brigham Young University.

Dr. Ravert: In educational settings, simulation has been used in a formative methodology and more recently in some programs as a summative assessment for high-stakes testing. Educational administrators are concerned in two main areas: (a) defensible measures of student success and (b) adequate resources of faculty and time. For simulation high-stakes testing to be a defensible measure, the scenarios must be validated, the evaluators must be trained, and inter-rater reliability established. A high-stakes test should be only part of a larger assessment plan. Assessment of student success is a process and should include data generated from multiple methods (triangulation) rather than one measure. If various measures lead to similar conclusions, then the level of uncertainty is reduced and would be defensible if a student questioned the validity of the test. Many resources are necessary to get to a point that simulation high-stakes testing could be considered valid and reliable. It takes time and money to develop appropriate scenarios and adequately train faculty to serve as raters. Most nursing programs conduct simulation experiences with groups of students rather than in a one-on-one situation, which would be required in a high-stakes testing environment—this would greatly increase the cost of the simulation program in each college. These concerns must be resolved before undertaking high-stakes testing.

Mary Anne Rizzolo, EdD, RN, is a Consultant at the NLN.

Dr. Rizzolo: The NLN Project to Explore the Use of Simulation for High-stakes Assessment is nearing completion. It began with a think tank of experts who shared their wisdom and recommended scenarios to assess students at the end of their program. Dr. Pamela Jeffries then assembled a team of expert simulation authors to design the simulations, whereas Drs. Marilyn Oermann and Suzan Kardong-Edgren examined evaluation tools and planned the training of raters. The scenarios were piloted and refined and then schools of nursing across the country ran the scenarios and sent video recordings of student performances. In the final phase, scenario authors scored the video recordings and then had repeated discussions on the rationales for their scores to come to consensus on scoring criteria. The final scoring criteria were given to raters, based on their expertise in both simulation and evaluation, who scored the videos at two points in time. Statistical analysis is underway to determine inter- and intra-rater reliability scores.

Many lessons have already been learned from our experiences to date on this project (Rizzolo, 2014; Willhaus, Burleson, Palaganas, & Jeffries, 2014), and the difficulty of implementing fair evaluation practices cannot be underestimated. Although general expectations of student performance can be generated and agreed upon by faculty, when simulation authors and raters were faced with applying them to an actual student performance, disagreement ensued. It was only through many hours of discussion and exposure to multiple different videos that consensus was reached.

As I listened to the discussion at the town hall, I was struck by the realization that no one questioned our current high-stakes assessment practice of having one faculty member evaluate students in the clinical area. In almost all programs, if a student fails the clinical portion of a course, the student fails the entire course, regardless of grades on written test. Clearly, we are already implementing high-stakes assessments in a poorly controlled environment with little consistency.

A well-designed and tested simulation, created from deliberative conversations among faculty that specify what behaviors students must demonstrate at the end of each course and at the end of the program is a worthy goal yet to be achieved by most. Having students demonstrate their competence with a simulation and administered in a controlled environment would certainly be an improvement over our current practice. Although it may be a difficult process to implement, it can lead to more fair and equitable testing practice. We hope that faculty will rise to the challenge. We look forward to revisiting this question in another 5 years to gauge our progress.

Dr. Teresa Gore is President Elect for INACSL.

Dr. Gore: INACSL's mission is to "promote research and disseminate evidence based practice standards for clinical simulation methodologies and learning environments." Our vision is to be "Nursing's portal to the world of clinical simulation pedagogy and learning environments" (www.inacsl.org). We are at such an important and interesting time in nursing education and simulation. The results of the NCSBN National Simulation Study (2014) and the NLN High-stakes Project (2010) will provide the nursing profession with empirical data to guide the education of prelicensure nursing students. Much of the discussion around the use of simulation has revolved around the need for validated scenarios, valid and reliable instruments for evaluating performance, and the appropriate "dose" of simulation to achieve the best outcome. The study methods developed in both of these studies provide excellent road maps for new simulation faculty and programs. Findings from the NCSBN study suggest that using The Standards of Best Practice: Simulation (2013) for program development can be instrumental in the development of a quality program.

Conclusions

This article discussed the use of high-stakes evaluation using simulation in nurse education today versus five years ago. Data on the topic of high-stakes testing were collected

through survey research before the INACSL conference in June 2014. Then during a town hall meeting at the INACSL conference, nurse educators spoke about the topic. This article described the survey findings, summarized the open forum at the conference, and highlighted perspectives from three nurse leader experts regarding future challenges with high-stakes evaluation.

References

Bensfield, L. A., Olech, M. L., & Horsley, T. L. (2012). Simulation for high-stakes evaluation in nursing. *Nurse Educator*, *37*(2), 71-74.

Boulet, J. R., Jeffries, P. R., Hatala, R. A., Korndorffer, J. R., Feinstein, D. M., & Roche, J. P. (2011). Research regarding methods of assessing learning outcomes. *Simulation Healthcare*, *6*, S48-S51.

Boulet, J. R., & Murray, D. J. (2010). Simulation-based assessment in anesthesiology: Requirements for practical implementation. *Anesthesiology*, *112*, 1041-1052.

Decker, S., Utterback, V. A., Thomas, M. B., Mitchell, M., & Sportsman, L. (2011). Assessing continued competency through simulation: A call for stringent action. *Nursing Education Perspectives*, *32*, 120-125.

Feldman, M., Lazzara, E. H., Vanderbilt, A. A., & DiazGranados, D. (2012). Rater training to support high-stakes simulation-based assessments. *Journal of Continuing Education in the Health Professions*, *32*(4), 279-286.

Harvey, P., & Radomski, N. (2011). Performance pressure: Simulated patients and high-stakes examinations in a regional clinical school. *Australian Journal of Rural Health*, *19*, 284-289.

Hayden, J. (2010). Use of simulation in nursing education: National survey results. *Journal of Nursing Regulation*, *1*(3), 52-57.

Hofer, S. H., Schuebel, F., Sader, R., & Landes, C. (2013). Development and implementation of an objective structured clinical examination (OSCE) in CMF-surgery for dental students. *Journal of Cranio-Maxillo-Facial Surgery*, *41*(2013), 412-416.

International Nursing Association of Clinical Simulation and Learning. (2013). Standards of best practice: Simulation. *Clinical Simulation in Nursing*, *9*(6S), S1-S32.

Kardong-Edgren, S., Hanberg, A. D., Keenan, C., Ackerman, A., & Chambers, K. (2011). A discussion of high-stakes testing: An extension of a 2009 INACSL conference roundtable. *Clinical Simulation in Nursing*, *7*(1), e19-e24. http://dx.doi.org/10.1016/j.ecns.2010.02.002.

Meakim, C., Boese, T., Decker, S., Franklin, A. E., Gloe, D., Lioce, L., … Borum, J. C. (2013). Standards of best practice: Simulation standard I: Terminology. *Clinical Simulation in Nursing*, *9*(6S), S3-S11. http://dx.doi.org/10.1016/j.ecns.2013.04.001.

McIndoe, A. (2013). High stakes simulation in anaesthesia. *Continuing Education in Anaesthesia, Critical Care & Pain*, *12*(5), 268-273. http://dx.doi.org/10.1093/bjaceaccp/mks034.

McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education research: 2003-2009. *Medical Education*, *44*(1), 50-63.

McWilliam, P. L., & Botwinski, C. A. (2012). Identifying strengths and weaknesses in the utilization of an Objective Structured Clinical Examination (OSCE) in a nursing program. *Nursing Education Perspectives*, *33*(1), 35-39.

National Council of State Boards of Nursing. (2014). *NCSBN National Simulation Study*. Retrieved from https://www.ncsbn.org/2094.htm.

National League for Nursing. (2014). *SIRC-Simulation innovation resource center*. Retrieved from http://sirc.nln.org/.

National League for Nursing. (2010). *Reflection and dialogue: High-stakes testing*. Retrieved from http://www.nln.org/aboutnln/reflection_dialogue/refl_dial_7.htm.

Nunnink, L., Foot, C., Venkatesh, B., Corke, C., Saxena, M., Lucey, M., & Jones, M. (2014). High-stakes assessment of the non-technical skills of critical care trainees using simulation: Feasibility, acceptability and reliability. *Critical Care Resuscitation*, *16*(1), 6-12.

Rizzolo, M. A. (2014). Chapter 9: Developing and using simulation for high-stakes assessment. In P. R. Jeffries (Ed.), *Clinical simulations in nursing education: Advanced concepts, trends, and opportunities*. Washington, DC: National League for Nursing.

Willhaus, P., Burleson, G., Palaganas, J., & Jeffries, P. (2014). Authoring simulations for high-stakes student evaluation. *Clinical Simulation in Nursing*, *10*, e177-e182.

Wind, S. A., & Engelhard, G. (2012). Examining rating quality in writing assessment: Rater agreement, error, and accuracy. *Journal of Applied Measurement*, *13*(4), 321-335.