12-6-2013

Authoring Simulations for High Stakes Student Evaluation

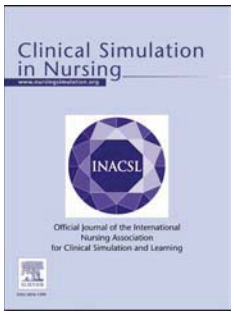
Janet Willhaus
Boise State University

Gail Burleson
Mott Community College

Janice Palaganas
Institute for Medical Simulation

Pamela Jeffries
Johns Hopkins University - School of Nursing

Accepted Manuscript



Authoring Simulations for High Stakes Student Evaluation

Janet Willhaus, PhD, RN Gail Burleson, MS, RN, CNE Janice Palaganas, PhD, RN, NP, CEN Pamela Jeffries, PhD, RN, FAAN, ANEF

PII: S1876-1399(13)00231-4

DOI: 10.1016/j.ecns.2013.11.006

Reference: ECSN 470

To appear in: *Clinical Simulation in Nursing*

Received Date: 23 October 2013

Accepted Date: 26 November 2013

Please cite this article as: WillhausJ., BurlesonG., PalaganasJ. & JeffriesP., Authoring Simulations for High Stakes Student Evaluation, *Clinical Simulation in Nursing* (2014), doi: 10.1016/j.ecns.2013.11.006.

Authoring Simulations for

High Stakes Student Evaluation

Janet Willhaus, PhD, RN
Assistant Professor
Boise State University School of Nursing, Boise, ID
janetwillhaus@boisestate.edu
208-426-3712 work
208-426-1370 fax
NLN Simulation Scholar in Residence during the time the scenarios were written

Gail Burleson, MS, RN, CNE
Nursing Instructor/Simulation Committee chairperson
Mott Community College Division of Health Sciences, Flint, MI
NLN Consultant

Janice Palaganas, PhD, RN, NP, CEN
Principal Faculty, Center for Medical Simulation
Assistant Director, Institute for Medical Simulation, Boston, MA
NLN Consultant

Pamela Jeffries, PhD, RN, FAAN, ANEF
Associate Dean for Academic Affairs
Johns Hopkins University School of Nursing, Baltimore, MD
NLN Consultant

The NLN received funding support for this project from Laerdal Medical

Authoring Simulations for High Stakes Student Evaluation

Abstract

Although simulation methods have primarily been used for teaching in nursing education, there is a growing interest in the use of simulation for student and program evaluation. Developing simulation scenarios for high stakes evaluation differs from traditional teaching/learning scenario authorship in a number of ways. This manuscript describes the process used to write, pilot test, and revise scenarios used in the National League for Nursing High Stakes Testing feasibility study. Observations and reported differences in scenario development and facilitation may provide insight to others regarding the best use of summative simulation scenarios.

Authoring Simulations for High Stakes Student Evaulation

Introduction

Today healthcare simulations are playing a major role in health professions education. The simulation movement is ever growing as evidenced by more sophisticated simulators, increasing numbers of simulation centers worldwide, and the Institute of Medicine (IOM, 2011) suggestion that simulations will play a more important role in health professions education today and in the future. The use of simulation has moved well beyond the point of asking about effectiveness (McGaghie, Issenberg, Petrusa, & Scalese, 2010) and is focused now on best practices supported by research (Norman, Dore, & Grierson, 2012).

With increasing construction of state-of-the-art simulation centers around the world, attention is turning to program evaluation and research to measure activities in newly created simulation centers. Researchers now strive to prove linkages between simulations in healthcare education and improved patient outcomes. One type of program evaluation that is increasing in frequency is the use of simulation for high stakes testing. High stakes simulations are those evaluated for significant consequences, impact, or students' grade (International Nursing Association for Clinical Simulation in Learning [INACSL], 2011). In 2010, the National League for Nursing (NLN) began a multi-site, feasibility study to investigate high stakes testing in simulation (Shultz, 2010). This article describes the development of high stakes simulation scenarios for the project. We also include insights on selection of topics, writing scenarios, peer review of scenarios, pilot testing, and final revisions that may provide guidance for scenario authors looking to standardize simulations for high stakes or summative assessment.

Background

The use of healthcare simulations for high stakes testing has been controversial given its multiple confounding variables (Palaganas, 2012), and the argument that learning in simulation

requires a psychologically safe environment (Rudolph, Simon, Dufresne, & Raemer, 2006). High stakes testing in this context is defined as a scenario with the potential to fail students at the end of a course *or* program on the basis of a simulation experience (Kardong-Edgren, Hanberg, Keenan, Ackerman, & Chambers, 2011). The NLN promotes fair testing practices in nursing education and recommends that faculty should use multiple sources of evidence when evaluating basic nursing competence (NLN, 2010, 2012). One source of evaluation increasingly respected as a valuable method of assessing competence is healthcare simulation. Although simulation for teaching and learning has been the focus of many nursing research studies, the development of simulation as a valid and reliable evaluation instrument has received less attention.

In 2010, the NLN conducted a Think Tank on Simulation for High Stakes Evaluation in which participants identified nursing program outcomes and competencies able to be assessed using end-of-program, high-stakes simulation scenarios. The four outcome areas selected were deliberately chosen, extensively discussed, and brainstormed for feasibility in simulation-based assessment. They were *Assessment and Intervention*, *Nursing Judgment*, *Quality and Safety*, and *Teamwork and Collaboration.*

Four nurse educators, experienced in simulation design and operation, were recruited for the scenario development phase of this High Stakes Simulation Project. Author preparation materials included research results, best practice articles, the Think Tank report, and the report from a National Delphi Study to Developmentally Level Knowledge, Skills and Attitudes (part of the Quality and Safety Education for Nurses [QSEN] initiative).

The team of four authors met initially with the research team and an NLN simulation consultant to review scenario design features, further define the four topics, and explore patient situations that would represent each topic area. To ensure consistency and best practice in

scenario design, authors were asked to review a presentation on simulation design, a peer reviewed article on simulation design, and the NLN Simulation Design Template.

Each author selected one of the four topics and each wrote three end of program pilot scenarios designed to evaluate a graduating nursing student in the selected outcome area. Authors received feedback each other and from simulation consultants on objective writing, scenario progression, and proposed objective outcomes throughout scenario development. Details of each scenario were refined after several months of author collaboration.

Writing for evaluation

One advantage of using simulation as a teaching tool is the ability to allow students to practice and to make mistakes in a safe environment. Using simulation for high stakes testing removes the safety platform. This change in using simulation for summative rather than formative practice increases the need to employ evidence-based approaches and collaboration to guide scenario development for student evaluation.

McDonald (2013) defines formative evaluation as an appraisal of student achievement while the student is still learning. It asks "How are you doing?" Alternatively, summative evaluation provides a description of student achievement at the end of a course and asks "How did you do?" (Table 1.) The NLN High Stakes Simulation project entailed creating scenarios that tested a nursing student's ability to demonstrate effective patient care in one of four essential clinical performance areas at the end of the academic program. The results of this type of evaluation could potentially be used to determine whether a student would pass a course, graduate, or even gain licensure or certification. This study sought to determine the process, challenges, and potential solutions to the challenges of using healthcare simulation for summative evaluation of students in a fair and reliable manner.

Two models were used for scenario development guidance in this project: The NLN/Jeffries Framework (Jeffries, 2005, 2007), and the Simulation Design Scale evaluation instrument (Jeffries, 2007).

Rethinking the Model for Testing

The NLN/Jeffries Framework, developed in 2005, offers guidance for implementing simulation-based education. The framework was designed to best achieve student-learning outcomes; however designing scenarios for summative evaluation required a different approach to several items in the NLN/Jeffries Framework, and resulted in thoughtful rewriting of each scenario (Table 2). While the Simulation Design Characteristics section of the Framework provided an important and useful guide for teaching using simulation, designing a scenario for evaluating summative end of program proficiency required a different balance of characteristics.

A critical element encountered in developing scenarios for evaluation versus teaching was the requirement for standardization. Effective evaluation requires standardization of the simulation scenarios for all testers (McDonald, 2013). To discourage academic dishonesty and prevent students from sharing information about each scenario, three parallel scenarios were standardized and developed for each of the four selected topic areas. Feedback from collaborators was used to ensure three scenario variations were created with the same difficulty and behavioral anchors for each topic. The difficulties in standardization were apparent during video review of the simulations, including site-to-site comparisons. These challenges included facilitator actions, fidelity, and differences in regional practices.

A primary challenge to standardization was the potential variation in facilitator actions at each simulation site. Because the purposes of the scenarios were student evaluation, care had to be taken to prevent cueing or other measures of student support more appropriate for

teaching/learning. In addition, variations in the way facilitators conducted a simulation sometimes resulted in disparate understanding and performance by the student being tested. Variance in facilitator actions was common place in video recordings of scenario pilot versions from schools across the country. For example one facilitator misread the "Report to Students" and used her own words to sound more like a nurse giving report to the on-coming shift, stating that the patient needed a "neuro" exam rather than "neurovascular assessment." As a result, the students in that group did not perform a neurovascular assessment, even though that was one of the scenario objectives. To reduce this variability, scenario authors provided an audio recording of each scenario's "Report to Students" for final scenario testing which resulted in less facilitator variation.

Another challenge for designers involved fidelity. In local settings, faculty conducting simulations for specific learning objectives can utilize existing simulation equipment to meet student needs. However, because standardization was required for the multisite study, scenario developers had to specify types of manikins and other materials used. One example of this was the use of elastic compression stockings for a post-operative patient as opposed to the more current, but less available, automatic sequential compression devices (SCD) for thromboembolism prevention. The scenario report indicated that the automatic stockings had been ordered, but were not yet available in the "Report to students." This allowed schools without an available SCD device in their laboratory to participate in the pilot and final versions of the video recordings.

Differences in regional practices, especially concerning treatments and medications, were another source of variation during scenario composition. The authors, geographically located in different parts of the country, were familiar with medications and treatments used in patient care

in the local area. This required additional research and collaboration to minimize differences and ensure best practice in this multisite study. For example, one scenario author discovered different surgical practices for a particular injury, depending on region and surgeon preference. The literature did not reveal a preference for one method, so the author selected the most current practice for the scenario even though different methods were more commonly used in other areas.

The Challenges of Observing Student Performance

Each scenario design presented opportunities for students to demonstrate (or fail to demonstrate) competencies, and for evaluators to rate them as passing or failing an action using the Creighton Simulation Evaluation Instrument (CSEI) (Todd, Manz, Hawkins, Parsons, & Hercinger, 2008). Authors gave thoughtful consideration about how evaluators would know when a participant was addressing an objective. For example, if a participant looked at the chart it was impossible to determine what was being read and if the information was understood. To assist with understanding participant comprehension, each scenario ended with another nurse taking over care of the patient. The participant gave the nurse report about the patient condition, any changes, or care that had been given during the scenario. The scenario authors carefully scripted questions that the receiving nurse (often the facilitator) could ask the participant about the patient's care and condition to help identify student comprehension of a problem which might not have been clearly represented on the scenario recording.

The manikins used were programmed to exhibit signs and symptoms to give the student the best indication that a specific problem existed and allowed each student to demonstrate or fail to demonstrate the results of an intervention. Each author developed a programming outline for the scenarios and specialists from Laerdal Medical provided assistance if needed to all schools

who participated in the final phases of scenario recording. Programming aspects included both physiological signs and vocalized symptoms at both the initial state and in reaction to student care. Some physiological signs were easy to simulate and were effectively demonstrated by the mannequins. These included cues such as urine volume and color, vital signs, and pulse oximetry. Moulage or make up for the simulation was specified by the authors and also helped demonstrate patient conditions (pressure ulcer, bruising, etc.). Other signs, such as skin temperature, movement, and nonverbal expressions could not be provided by the mannequin type selected for the study and required facilitator involvement.

Symptoms sometimes presented more of a challenge. These had to be expressed by the facilitator providing the "patient's" voice. It was essential that the facilitator remain true to the script so that the opportunities for student performance would be presented and standardization and fairness best ensured. Authors scripted the dialogues of some cases indicating specific responses in order to achieve the objectives. An example of a dialogue to demonstrate patient signs that the mannequin could not exhibit, a dialogue to enable evaluation of the student's ability to assess a symptom (pain), and portrayal of the demeanor of the patient are presented in Table 3.

After piloting the scenarios in diploma, associate degree, and baccalaureate programs across the U.S., the authors revised each scenario based on review of videos and facilitator feedback. Although the authors each wrote three scenarios they believed were the same level of difficulty for their topics, after the pilot it was determined at least one set of scenarios needed adjustment. Feedback from the piloting school facilitators indicated that one scenario required a call to the physician for further orders while the two parallel versions could be handled with nursing interventions. Additionally the pilot videos revealed the need for specific guidance to

each school on camera angles and fidelity for unobstructed viewing of critical tasks. Clear audio volume and fidelity were additional factors identified as important for evaluation. After editing and revising was completed, final versions of the scenarios were sent again to a variety of pre-licensure nursing programs across the nation where volunteer nursing students in their final semester of study were video recorded participating in the scenarios.

Recordings of the final versions of the scenarios were viewed and scored by evaluators who did not know any of the students. This limited the halo effect of having faculty evaluate their own students. The halo effect is a bias in which one's perceptions are influenced by the previous encounters or perceptions of others. Halo effect can influence judgments about performance (Walsh & Kapoor, 2009). Evaluators viewed the final recordings and scored the students using the CSEI.

Discussion

One obvious barrier to fair and reliable testing was the variation in simulation implementation and facilitation. Even in the final recordings, facilitators deviated from cuing guidance despite written instruction not to do so. This could have been the result of a desire to see students perform at higher levels or could have stemmed from a limited understanding of how to use simulations for evaluation purposes. It was clear from viewing both the pilot and final videos that simulation facilitators practice differently from location to location. Hayden (2010) surveyed faculty at 1,729 nursing programs nationwide and reported that the sources of education and training for simulation faculty differ from school to school or may not exist at all. Educational resources available vary widely with faculty reporting commercial vendors of manikin products as the main resource (Kardong-Edgren, Willhaus, Bennett, & Hayden, 2012)

Communication with evaluators about what actions were considered non-applicable for each scenario may have been helpful. While the authors did not participate in either evaluation or score norming, they communicated the actions that were considered essential or important in writing to the evaluators. The scenario authors did not, however, always give a clear indication of what items on the evaluation instrument might be considered not applicable. Evaluators who did not observe a participant action in the recordings were left to determine independently whether it was a failing omission or whether it was simply irrelevant. For example, if a participant failed to collaborate by phone with a provider, pharmacist, or family member, some evaluators indicated the tasks were not applicable because they had no way to judge the participant's communication skill. Other evaluators viewed the omission itself as critical and rated the participant as failing at that task. Better communication about the intent from the authors may have resulted in more reliable and consistent ratings.

Peer review and piloting of the scenarios was a critical element in producing parallel scenarios of the same user level. It also enhanced the accuracy and quality of materials available for facilitator use in the final simulation recordings. The National Council of State Boards (NCSBN) survey revealed that nationwide only 44% of simulation users pilot scenarios prior to use and only 12% used the NLN Simulation Design Scale as an evaluation method (Kardong-Edgren et al., 2012). The most common forms of review for quality are faculty peer (59%) and student (72%) evaluation.

Recommendations

Faculty writing simulation scenarios for high stakes evaluation should consider review of the Code of Fair Testing Practices in Education (2004). This document, prepared by the Joint Committee on Testing Practices, provides guidance for both test developers and test users in the

critical areas of developing and selecting appropriate tests, administering and scoring tests, reporting and interpreting test results, and informing test takers. While not all the recommendations for test developers fit the simulation format, they provide a checklist of best practices for any form of evaluation. The document emphasizes that both test writers and test users have both separate and overlapping roles.

Standardizing and streamlining education and practice standards for simulation facilitators are important steps for producing valid and reliable results with simulations for evaluation. Certification standards now exist for simulation health educators through the Society for Simulation in Healthcare (Society for Simulation in Healthcare Board, 2012). A reading list included in the newly released certification handbook may provide a logical starting point to promote standardization among simulation educators and facilitators.

Piloting and reviews by others are powerful tools in the development of any simulation scenario, but are imperative for writing scenarios for testing. This step should not be rushed and may take one or more semesters before a scenario is considered valid and reliable.

Student preparation is also a key to fair testing. Students with little simulation experience will be understandably anxious, and students who are not oriented to the room, equipment, and available resources would be unfairly tested. Test anxiety is not limited to written examination. Research demonstrates that students in summative and high risk simulations also demonstrate anxiety and physiological stress (Manderino, Yonkman, Ganong, & Royal, 1986; Gantt, 2013)

Conclusion

Writing high stakes simulation scenarios to evaluate student performance differs significantly from writing scenarios for formative or learning experiences. This is only one lesson learned from the NLN High Stakes Simulation Testing investigation. It is projected that

other experiences and recommendations will be forthcoming in the literature for other phases of the project.

References

INACSL (International Nursing Association of Clinical Simulation and Learning Board of Directors) (2011). 2011. *Clinical Simulation in Nursing, 7*(supplement 4), s3 - s7. doi:Doi:10:1016/j.ecns.2011.05.005.

Gantt, L. T. (2013). The effect of preparation on anxiety and performance in summative simulations. *Clinical Simulation in Nursing, 9*(1), e25-e33. doi:10.1016/j.ecns.2011.07.004

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

IOM (Institute of Medicine). (2011). *The Future of Nursing: Leading Change, Advancing Health.* Washington, DC: National Academies Press.

Jeffries, P. R. (2005) A frame work for designing, implementing, and evaluating simulations used as teaching strategies in nursing. *Nursing Education Perspectives,* 26(2), 96-103.

Jeffries, P. R. (2007) *Simulation in nursing education: from conceptualization to evaluation.* New York: National League for Nursing

Joint Committee on Fair Testing Practices in Education. (2004). *Code of fair testing practices in education.* Washington, DC. Retrieved from http://apa.org/science/programs/testing/fair-code.aspx

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

Kardong-Edgren, S., Willhaus, J., Bennett, D., & Hayden, J. (2012). Results of the national council of state boards of nursing national simulation survey: Part II. *Clinical Simulation in Nursing, 8*(4), e117-e123. doi:10.1016/j.ecns.2012.01.003

Manderino, M. A., Yonkman, C. A., Ganong, L. H., & Royal, A. (1986). Evaluation of a cardiac arrest simulation. *Journal of Nursing Education, 25*(3), 107-111.

McDonald, M. E. (2013). *The nurse educator's guide to assessing learning* (3 Ed.) Burlington, MA: Jones and Bartlett

McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical education 2003-2009. *Medical Education, 44*, 50-56. doi:10.1111/j.1365-2923.2009.03547.x

NLN (National League for Nursing) (2010, December). *Reflection and Dialogue: High Stakes Testing.* Retrieved from National League for Nursing: http://www.nln.org/aboutnln/reflection_dialogue/refl_dial_7.htm

NLN (National League for Nursing) (2012) *Fair Testing Guidelines for Nursing Education.* Retrieved from National League for Nursing: www.nln.org/fairtestingguidelines

Norman, G., Dore, K., & Grierson, L. (2012). The miminal relationship between simulator fidelity and transfer of knowledge. *Medical Education, 46*, 635-647. doi:10.1111/j.1365-2923.2012.04243.x

Palaganas, J. C. (2012). *Exploring healthcare simulation as a platform for interprofessional education.* (Order No. 3547895, Loma Linda University). *ProQuest Dissertations and Theses,* 217. Retrieved from http://search.proquest.com/docview/1278062785?accountid=9649. (1278062785).

Rudolph, J. W., Simon, R., Dufresne, R. L., & Raemer, D. B. (2006). There's no such thing as a "non-judgmental" debriefing: a theory and method for debriefing with good judgment. *Simulation in Healthcare, 1*, 49-55.

Shultz, C. M. (2010). High stakes testing? Help is on the way. *Nursing Education Perspectives, 31*(4), 205. doi:http://dx.doi.org/10.1043/1535-5026-31.4.205

Society for Simulation in Healthcare Board. (2012). *SSH certified simulation healthcare educator handbook.* Society for Simulation in Healthcare. Retrieved from http://ssih.org/uploads/static_pages/PDFs/Certification/CHSE_Handbook.pdf

Todd, M., Manz, J. A., Hawkins, K.S., Parsons, M. E., & Hercinger, M. (2008). The development of a quantitative evaluation tool for simulation in nursing education. *International Journal of Nursing Education Scholarship, 5*(1). Article 41.

Walsh, K. & Kapoor, S. S. (2009). Online feedback to medical education resources: Is there a halo effect? [Letter to editor] *Education for Health, 22* (1). Available from: http://www.education for health.net/

Table 1

Comparison of formative and summative evaluation (McDonald)	
Formative Evaluation "How are you doing?"	Summative Evaluation "How did you do?"
Occurs during the process of learning	Occurs at the completion of instruction
Assesses progress in a course	Summarizes achievement in a course
Directs learning to achieve objectives	Assesses objective achievement
Grades not assigned	Assigns grades
Provides feedback	Provides feedback

Comparison of formative and summative evaluation using simulation (Jeffries)		
	Formative Evaluation	Summative Evaluation
Purpose	Gain data to help future development of learner, scenario	Gain data to determine competency of the participant
Time Frame	Before the end of the course, examines parts of the course	Occurs at the end of the course; focus on the whole event
When to evaluate	When there is still time to improve performance	To document a specific competency level

Table 2

Teacher		
	Teaching/learning	High stakes testing
Role as facilitator in the simulation	May provide support and encouragement to the learner throughout the simulation	Manages simulation in a defined role: introduction of scenario; voice of patient and healthcare collaborators.
Scenario control	May adjust scenario "on the fly" to promote learning	Manages a standardized scenario
Familiarity with the simulation scenario	May be the author of the scenario, may have developed the objectives	Facilitates a standardized simulation scenario developed by others
Most important preparation	Expertise in the scenario subject is critical to guide students' learning	Ability to "go by the script" is critical for evaluation
Student		
Level of students	At various stages of a program (same program)	End of program (many different programs)
Experience with simulation	Usually the same for the cohort	Different amounts and types, maybe none
Participation	May work as a team, collaborate	Single participant
Preparation	May be given in advance to augment learning and performance	Preparation materials given immediately before the scenario

Table 3

Scenario Design Characteristics		
Feature	Importance, level of inclusion	
	Learning model	Assessment model
Objectives	High	High
Fidelity	Depends on objectives	High
Problem solving	Various levels of complexity Depends on objectives	High
Peer support	High	Low
Debriefing	High	Low

Table 4

Neurovascular Assessment Dialogue

Anticipated question from Nurse (Participant)	**Mrs. Jasper (Facilitator)**
Any numbness or tingling in your right foot?	"No, my foot's fine"
Can you feel my touch?	"Yes, Stop bothering me"
Can you move your right foot?	***Facilitator:*** Say "Yes, she moves it , wiggles toes"
Do you have pain in your foot when I move it?	"No"
Examines color, temperature of right foot	***Facilitator:*** Both normal, same as unaffected foot (May need to state this)
Palpates right pedal pulse	Pulse is present. ***Facilitator:*** (If manikin does not have pedal pulses, need to say, "Pulse is present")

Hip Pain Dialogue

Anticipated question from Nurse (Participant)	**Mrs. Jasper (Facilitator)**
Are you having any pain?	It's terrible, starts to cry
Could you rate the pain of on a scale of 0-10	I can't deal with numbers today, it just hurts
Describe the pain	Throbbing
Where is the pain	It's in my hip
Does anything make it better or worse?	It hurts more when I move. If I lie still, it's not so bad. So don't make me move!
Is it constant or does it come and go?	It's constant

Note to Facilitator:

Mrs. Jasper is very confused and agitated, especially concerning her cat. Please see dialogue pages, and project confusion and agitation frequently.