Group simulation for "authentic" assessment in a maternal-child lecture course

Desiree Hensel[1] and Leah Stanley[2]

Abstract: The purpose of this pilot study was to explore student perceptions and outcomes surrounding the use of a labor and delivery simulation as a midterm exam in a maternal-newborn lecture course. An exploratory case study design was used to gain a holistic view of the simulation experience. Data from focus groups, written debriefings, simulation scoring rubrics, student course evaluations, and other course exams were analyzed using Stake's case study method. Qualitative analysis revealed four themes: confidence, fairness, reliability, and team effort. Students were able to accurately grade the performance of their group as a whole and complete a group self-debriefing, but quantitative analysis showed that the group scores were significantly higher than other individual course grades. The findings suggested that the group simulation was an authentic assessment of teamwork, but not individual performance. Future research is needed to determine what role simulation and collaborative testing should play in pre-licensure education.

Keywords: authentic assessment; case study; competency-based education; evaluation research; high-fidelity simulation; higher education; maternal-child nursing; nursing education; patient simulation; QSEN; quality and safety education; student performance appraisal; testing

Background

To practice in today's health care environment, experts believe that registered nurses (RNs) must possess specific knowledge, skills, and attitudes related to quality and safety, collectively known as the Quality and Safety for Nurses (QSEN) competencies (Cronenwett et al., 2007). These competencies include teamwork, patient-centered care, informatics, evidence-based practice, quality, and safety. The question of how best to assess students' mastery of these competencies is currently of great interest to nurse educators.

Educational Assessment Design

Assessment design can foster either deep or superficial learning (Tiwari et al., 2005). Frequent testing with traditional questions (e.g. multiple choice, true/false, etc.) has been shown to improve long-term retention (Roediger & Butler, 2011), but multiple choice questions in test banks and standardized exams are often flawed (Masters et al., 2001; Tarrant, Knierim, Hayes, & Ware, 2006). These flaws tend to penalize high-achieving students (Tarrant & Ware, 2008). Roediger and Butler also warn misinterpretation of multiple choice distractors may cause

---

[1] School of Nursing, Indiana University, 1033 E 3rd Street Sy444, Bloomington, IN 47405, dehensel@iu.edu
[2] Northwestern Memorial Hospital, Chicago, IL 60611

students to learn information incorrectly. Others argue that written examinations with conventional test questions are only indirect measures of student abilities used as proxies for real performance, and they do not necessarily predict workplace behaviors (Rodgers, Bhanji, & McKee, 2010; Wiggins, 1998).

"Authentic" assessments, on the other hand, are measures of student performance that require the same knowledge, skills, and attitudes that would be used when faced with the same situation in professional practice (Gulikers, Bastiaens, & Kirschner, 2004). While a traditional test requires only a response, an authentic assessment requires learners to perform or produce and to explain or justify their actions. When assessments are based on authentic tasks and use performer-friendly feedback, Wiggins (1998) asserts that assessments do more than test -- they serve an educative function.

Gulikers et al. (2004) provide a framework for understanding the validity of authentic assessments. Construct validity arises from the five elements of authenticity: (a) *task* or how the problem resembles a real practice situation, (b) *physical context* involving the resources and information available, (c) *social context* including performing collaboratively if that is the norm in practice, (d) *assessment form* or requiring that students observably demonstrate competencies, and (e) *criteria* or the professional standards used to judge the output. These elements are subject to student perceptions of their realism. The assessment's validity depends on the effects of the assessment on student motivation and learning. Gulikers et al. suggest that authentic assessment, along with authentic instruction, set the foundation for authentic learning that can be translated to practice.

Simulation

Simulation may serve as an optimal method to create authentic assessments for nursing education because they are grounded in an authentic task and can be used to both teach and assess learning (Jeffries, 2007). Competency testing using simulations is increasingly being used in clinical education with one study finding that 45% of undergraduate clinical courses in the United States used simulation to some degree to assess learning (Oermann, Yarbrough, Saewert, Ard, & Charasika, 2009). Students believe that assessments constructed using simulated scenarios can improve learning (Leung, Mok, & Wong, 2008). Simulation has also been shown to increase self-efficacy when used as a teaching strategy in lecture classes (Sinclair & Ferguson, 2009), but less is known about the use of simulation for competency testing in lieu of traditional examinations in such courses. The purpose of this pilot project was to create an authentic assessment midterm examination in a maternal-child lecture course using simulation and to evaluate the outcomes in terms of students' performance outcomes and perceptions. Our research question was how does a group simulation serve as an authentic assessment of competency for students enrolled in a maternal- child nursing lecture course?

Method

Setting and Participants

The pilot study took place at a large public university in the Midwestern United States. All students enrolled in a junior-level, baccalaureate, maternal-child nursing course participated in a group simulation involving the care of a patient in labor in place of a written midterm

examination (*N*= 28 females & 2 males). The students participated in 6 groups of 5 students each in the pilot project.

Procedures

Students were given a study guide in advance to review the possible types of patients that they might encounter and were encouraged to study as a team. Groups were assigned in 30-minute blocks on the testing day. After orienting to the room and equipment, roles were randomly assigned for the brief simulation as nurses (*N=2*), evaluators (*N=2*), or video-recorder (*N=1)*. The scenario used a high-fidelity birthing simulator and involved caring for a patient in labor immediately after spontaneous rupture of membranes. The simulation design was consistent with the National League for Nursing (NLN)-Jeffries Simulation Framework (Jeffries, 2007). Students in the nurse role were expected to check the fluid, which was stained with meconium, and intervene for the abnormal patterns displayed on the electronic fetal heart rate monitor by at least repositioning the patient and notifying the physician. The simulation averaged approximately five minutes in length and ended when the students notified the physician.

The evaluators and instructor scored the simulation using a rubric with five categories: safety, communication, teamwork, assessment, and interventions. Each category spelled out performance criteria for 0, 1, or 2 points. After the simulation, students watched the video, and the evaluators shared their ratings. The groups were then asked to complete a written debriefing based on common debriefing questions and the QSEN competencies and to submit it to the instructor by the following day (see Figure 1).

---

1. Please summarize the simulation.
2. What went well in the simulation?
3. What would you have liked to have done better or differently?
4. The next set of questions addresses the QSEN competencies.
   a. Patient-centered care: Describe your communication with your patient. Was it therapeutic and respectful? How did it (or not) reflect caring?
   b. Teamwork & collaboration: How did (or not) the nurses work within their scope of practice?
   c. Evidence-based practice: Given the situation, describe the evidence-based-protocol that should be implemented to care for the baby at birth.
   d. Quality improvement: How would you describe the quality of this patient's care and why?
   e. Safety: Use the SBAR (Situation-Background-Assessment-Recommendation) format now to write out a report that should have been phoned to the physician. Did your group follow that format? If not, what were you missing?
   f. Informatics: How did technology play a role in your decision making and the provision of safe care?
5. What did you take away from this experience? Please include individual and group thoughts.

Figure 1. Group Debriefing Guide. Questions adapted from Cronenwett et al. (2007) and NLN Debriefing/Guided Reflection QSEN overview for Laerdal Simulations Volume II

Group scores were determined by the course instructor based on observed performance as scored by the student evaluators and instructor with the rubric (20%) and the group's collectively written self-debriefing (80%). All members in each group received the same final grade. Following the simulation, all students were invited to share their perceptions in one of two audio-taped focus groups led by a senior honors student (*N*=18).

Data Analysis

An exploratory case study design was used to evaluate student perceptions and performance outcomes. Approval was obtained from the university Institutional Review Board to retain and study all materials generated as part of normal course work and to conduct focus groups. Stake's (1978; 1995) case study analysis method was used to identify patterns. Student perceptions were identified using the focus group transcripts, debriefing guides, and course evaluations. Patterns found in the qualitative data were discussed between authors to arrive at predominant themes. The identified themes were then coded, placed in a matrix, and tabulated. Group performance was measured using the debriefing guide and simulation rubrics. The group simulation scores were compared to individual average course exam scores, scores on a nationally normed content proficiency exam (www.ATItesting.com), and course grades. Scores were analyzed using descriptive statistics and independent *t* tests. The objective and subjective data were cross-compared to draw conclusions.

Results

Student Perceptions

Four themes emerged: team effort, fairness, reliability, and confidence*.* All themes except confidence, about which all comments were positive, revealed both positive and negative perceptions (See Table 1).

Throughout the data, the participants expressed the recurring theme of *team effort* as a positive experience and as a way to feel less nervous than when taking an exam as an individual. "Nothing we do in the field is going to be an individual effort." "You would have at least one person there to help you out." Another student stated "You can stop each other if you're doing something wrong." However, some students also perceived some aspects of working in a group as detrimental, mostly in regard to dependence on others for one's grade. One student commented, "Your grade relies on what two people do, so…kind of scary." Furthermore, some students perceived that this type of group division may have hindered what they could have gained from the simulation. One student commented that "only two students got to benefit from participating."

Students had mixed perceptions as to the *fairness* of the group simulation. Some students believed that the materials and resources provided beforehand were adequate and the amount of information the simulation focused on was fair. One student commented: "There were only three scenarios, and I feel like you could really focus on those three things and the things you needed to know how to do." There was a general consensus that the selection process was fair. "We were all prepared to be the nurse." On the other hand, many students viewed some aspects of the different roles as unfair. For example, one commented that "not everyone gets to be the nurse." Another said, "I feel like I wasn't contributing." Many comments were made about inadequate

preparation with the equipment beforehand, which made evaluation in this simulation unfair. Some students had not had the chance to be on a labor and delivery unit before participation in the simulation. One student said, "We had never seen any of the equipment, none of the machines, nothing." Many students believed that more exposure to the equipment would have greatly increased the fairness of the examination

Table 1

Frequency of Student Perception Themes

Theme		Debriefing Forms	Focus Groups	Course Evaluations	Total
Team Effort	*Beneficial*	14	3	2	19
	Detrimental	0	4	1	5
Fairness	*Fair*	2	13	1	16
	Not fair	0	14	3	17
Reliability	*Reliable*	6	3	2	11
	Not reliable	2	3	0	5
Confidence	*Gained*	6	5	0	11

Most students thought that simulation provided a way for poor test-takers to demonstrate their knowledge. One said, "[Simulation] is a better indicator to the instructor…to show them that you actually did prepare." Other students questioned the *reliability* of the assessment of their knowledge on the topic because the simulation only addressed one situation. One student noted, "We only did one simulation, so I feel like there may be other areas or other things we aren't as well versed in as we could have been."

The idea of gaining *confidence* was frequently mentioned as something gained from the experience. In the debriefing paperwork, one group wrote, "The biggest thing we took away from this experience is a gain in confidence that we are more prepared to work as a nurse than we previously believed." Students also addressed the confidence they gained when they were able to support each other as a group during the critiquing process. One student commented, "When your group watches it together I feel like they can kind of give you confidence." Being able to critique each other as a group after watching their video allowed students to hear supportive comments about their skills, giving them confidence.

Student Performance

Students were able to reliably grade their group's performance using the rubric, with student ratings matching the instructor's score 100% of the time. The most common performance deductions were for wearing dirty gloves when using the telephone and mistaking late for variable decelerations in fetal heart rate patterns. Groups were also able to complete a thoughtful self-briefing using the provided guide. All groups identified electronic fetal monitoring as a way that technology supported patient safety. All but one group were able to predict how the finding of meconium-stained amniotic fluid fit with national guidelines for neonatal resuscitation.

Still, the group scores on the simulation exam were significantly higher than those on any other individual course measure, averaging 94.7% (SD = 2.17). Specifically, the group simulation scores were higher than individual performance measures on the national content proficiency exam (M=71.1%, SD=8.41; p=.001), other course exams (M=85.7%, SD=4.53. p=.001), and the final OB course grade (M=91.7%, SD=2.74, p=.001).

Discussion

Evidence of Authenticity

The outcomes suggest that the group simulation fulfilled the criterion of a valid, authentic assessment as identified by Gulikers et al. (2004). The students perceived the task as being real nurse work. The social context involved collaboration as it would in the workplace. The performance was observable, and the students were evaluated against several professional quality and safety standards. Feedback suggested that physical context was perceived as the least authentic aspect of the simulation, mostly because some students believed they did not have enough prior experience using the labor and delivery equipment. Fetal heart patterns had been covered extensively in class, but viewing them on PowerPoint slides apparently felt very different to the students than reading them on a monitor.

Working in a team is an expectation of the RN (American Association of Colleges of Nursing, 2008; Cronenwett et al., 2007; Institute of Medicine [IOM], 2011), and one of the project's strongest themes revealed that students highly valued the teamwork aspect of the simulation. Since performing collaboratively is the norm in practice, the social context was perhaps the most compelling evidence that this simulation met Gulikers et al.'s (2004) standards for an authentic assessment. Students were encouraged to study together to begin to form a group identity before the actual simulation, and they volunteered that the group support provided a sense of comfort. This reinforced the findings by Elfrink, Nininger, Rohig, and Lee (2009) that the group setting and the group planning skills that are required may be some of the most beneficial aspects of the simulation experience.

In this pilot study, the students reported gaining confidence from the simulation. Through the debriefing process the student groups were able to begin the work of quality improvement, defined as observing care outcomes and implementing new methods to improve care (Cronenwett et al., 2007). Student learning gives an authentic assessment its consequential validity (Gulikers et al., 2004), and the written group self-debriefings showed evidence of student learning. For instance, students had an opportunity to say what they would have liked to have done differently and what they took away from the experience. SBAR (Situation, Background, Assessment, and Recommendation) format is recommended to improve collaborative communication (Beckett & Kipnis, 2009). During this simulation, no student used the SBAR communication format, but all groups were able to provide a corrected version with the debriefing.

Authentic assessments arise from and inform authentic instruction (Gulikers et al., 2004). In this case, we expected that the most challenging part of the task might be to interpret the electronic fetal monitor data, but we did not expect to see so many violations of standard infection control precautions. Specifically, soiled gloves are to be removed after patient contact (World Health Organization, 2009), but students routinely touched personal items and made

telephone calls while wearing soiled gloves. These repetitive errors led us to conclude that the methods we were using to teach the use of personal protective equipment were not effective.

Issues with Group Testing

It is not unusual for nursing students to be focused on grades (Oermann & Gaberson, 2009), and by far the most negative comments about the simulation were related to the issue of grades. Elfrink et al. (2009) found negative attitudes arose from being "singled out" to be the nurse in group simulations, but we found students were disappointed when they did not get to be the nurse. This was partially due to feeling like they were not contributing and disliking the fact that their grade was linked to the performance of others, even though the grading process was heavily skewed to favor the group-think debriefing and yielded higher average grades than other course assessments.

Although the students agreed that the simulation provided a real-world assessment of group skills, students did not perceive the simulation as being a reliable and accurate measure of individual abilities. Analysis of all quantitative measures supported the students' perceptions. Group performance scores were significantly higher than other individual course performance measures. Others also have found that when group testing is utilized, scores tend to be higher than individual scores (Michaelsen & Sweet, 2011; Sandahl, 2009).

The increasing complexity of the healthcare environment calls for a greater emphasis on the nurse's ability to work collaboratively (Cronenwett et al., 2007; IOM, 2011). This includes working collaboratively during patient care and in the quality improvement process. If teamwork is a practice competency, the question for nursing educators becomes what role should collaborative testing should play in pre-licensure education? Nurse educators may fear group testing because they want to prepare students to take national licensure examinations, which are individual efforts. Still, leaders in nursing education assert that multiple methods of assessment give a clearer picture of student abilities (NLN, 2010). Sadahl (2009) argues that students learn from the group-think process in collaborative testing and retain the information longer than from traditional individual testing.

Feasibility

Availability of resources has to ultimately factor into testing format decisions. While Roedinger and Butler (2011) suggested that frequent testing promotes learning, we feel resource availability would challenge the ability to use frequent testing with simulation for large groups often seen in lecture courses. In this study, 30 minutes was allotted per group for an orientation, to complete the scenario, and to watch the video-tape within the lab. Thus testing six groups took three hours to administer. Previous traditional midterm examinations in this course took only one hour to administer.

Conclusions

We concluded that this group simulation was an authentic assessment of teamwork that increased student confidence and promoted learning. Thus it was an appropriate method to test attainment of course objectives related to collaboration. However, it was not a measure of individual performance. Limitations of this pilot project included the narrow demographics;

participants did not reflect the entire population of nursing students demographically and therefore may not have provided widely applicable results. Another limitation was that the testing scenario itself was focused on a very particular situation and only the performance of the two students in the nurse direct patient care role could actually be assessed. Finally, the nature of case study design provides room for researcher bias due to the nature of the analysis, although cross-comparisons using of multiple types of measures added credibility to the findings and helped to minimize bias in this study.

Despite the pilot study limitations, the findings provide direction for future studies. More research is needed to understand the feasibility and outcomes of simulation testing compared to traditional testing in larger groups and different nursing content applications. Since simulations and traditional assessment methods seem to have different but complementary strengths and weaknesses, future research is also needed to identify the best mix of testing methods to predict and improve practice performance.

Acknowledgements

References

American Association of Colleges of Nursing. (2008). *The Essentials for Baccalaureate Education for the Professional Practice of Education*. Retrieved from http://www.aacn.nche.edu/education/pdf/baccessentials08.pdf

Beckett, C. D. Y. & Kipnis, G. (2009). Integrating SBAR to Improve Quality/Patient Safety Outcomes. *Journal for Healthcare Quality, 31* (5), 19-28. doi: 10.1111/j.1945-1474.2009.00043.x

Cronenwett, L., Sherwood, G., Barnsteiner, J., Disch, J., Johnson, J., Mitchell, P., …Warren, J., (2007). Quality and Safety Education for Nurses. *Nursing Outlook, 55* (3), 122-131. doi: 10.1016/j.outlook.2007.02.006

Elfrink, V.L., Nininger, J., Rohig, L., Lee, J. (2009). The Case for Group Planning in Human Patient Simulation. *Nursing Education Perspectives, 30*(2), 83-86.

Gulikers, J. T. M., Bastiaens, T. J., & Kirschner, P. A. (2004). A Five-Dimensional Framework for Authentic Assessment. *Etr&D-Educational Technology Research and Development, 52*(3), 67-86. doi: 10.1007/bf02504676

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

Jeffries, P. R. (2007). *Simulation in Nursing Education: From Conceptualization to Evaluation.* New York: National League for Nurses.

Leung, S. F., Mok, E., Wong, D. (2008). The impact of assessment methods on the learning of nursing students. *Nurse Education Today, 28* (6), 711-719. doi: 10.1016/j.nedt.2007.11.004

Masters, J.C., Hulsmeyer, B.S., Pike, M.E., Leichty, K., Miller, M.T., & Verst, A.L. (2001). Assessment of multiple-choice questions in selected test banks accompanying text books used in nursing education. *Journal of Nursing Education, 40*(1), 25-31.

Michaelsen, L.K. & Sweet, M. (2011). Team-based leaning. *New Directions for Teaching and Learning*, *2011*(128), 41. doi: 10.1002/tl.467

National League for Nursing. (2010). National League for Nursing task force tackles high stakes testing. Retrieved from http://www.nln.org/newsreleases/highstakes_testing_061110.htm

Oermann, M.H. & Gaberson, K.B. (2009). *Evaluation and testing in nursing education*, 3rd ed. New York: Springer.

Oermann, M.H., Yarbrough, S.S., Saewert, K. J., Ard. N., & Charasika, M. (2009). Clinical evaluation and grading practices in schools of nursing: National survey findings Part II. *Nursing Education Perspectives, 30* (6), 352–35. doi: 10.1043/1536-5026-30.6.352

Roediger, H.L. & Butler, A.C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences, 15*(1), 20-27. doi: org/10.1016/j.tics.2010.09.003

Rodgers, D. L., Bhanji, F., & McKee, B. R. (2010). Written evaluation is not a predictor for skills performance in an Advanced Cardiovascular Life Support course. *Resuscitation, 81*(4), 453-456. doi: 10.1016/j.resuscitation.2009.12.018

Sandahl, S.S. (2009). Collaborative testing as a learning strategy in nursing education: A review of the literature. *Nursing Education Perspectives, 10* (1), 171-175

Sinclair, B. & Ferguson, K. (2009). Integrating simulated teaching/learning strategies in undergraduate nursing education. *International Journal of Nursing Education Scholarship, 6* (1), Article7. doi: 10.2202/1548-923X.1676

Stake, R.E. (1978). The case study method in social inquiry. *Educational Researcher, 7* (2), 5-8. doi: 10.2307/1174340

Stake, R.E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage Publications, Inc.

Tarrant, M., Knierim, A., Hayes, S. K., & Ware, J. (2006). The Frequency of Item Writing Flaws in Multiple-Choice Questions Used in High Stakes Nursing Assessments. *Nurse Education Today, 26*(8), 662-671. doi: 10.1016/j.nedt.2006.07.006

Tarrant, M. & Ware, J. (2008). Impact of Item-Writing Flaws in Multiple-Choice Questions on Student Achievement in High-Stakes Nursing Assessments. *Medical Education, 42*(2), 198-206. doi: 10.1111/j.1365-2923.2007.02957.

Tiwari, A., Lam, D., Yuen, K. H., Chan, R., Fung, T., & Chan, S. (2005). Student Learning in Clinical Nursing Education: Perceptions of the Relationship Between Assessment and Learning. *Nurse Education Today, 25*(4), 299-308. doi: 10.1016/j.nedt.2005.01.013

Wiggins, G. (1998). *Educative Assessment.* San Francisco, CA: Jossey-Bass Inc.

World Health Organization (2009). Glove Use Information Leaflet. Retrieved from http://www.who.int/gpsc/5may/Glove_Use_Information_Leaflet.pdf