

## ORIGINAL RESEARCH

# Reliability and validity of scenario-specific versus generic simulation assessment rubrics

Marian Luctkar-Flude\*, Deborah Tregunno, Kim Sears, Cheryl Pulling, Kayla Lee, Rylan Egan

School of Nursing, Queen's University, Kingston, Ontario, Canada

**Received:** January 5, 2020

**Accepted:** February 26, 2020

**Online Published:** May 17, 2020

**DOI:** 10.5430/jnep.v10n8p74

**URL:** <https://doi.org/10.5430/jnep.v10n8p74>

## ABSTRACT

**Background:** This study assessed reliability and validity of scenario-specific and generic simulation assessment rubrics used in two different deteriorating patient simulations, and explored learner and instructor preferences.

**Methods:** Learner performance was rated independently by three instructors using two rubrics.

**Results:** A convenience sample of 29 nursing students was recruited. Inter-rater reliability was similar but slightly higher for the generic rubric than the scenario-specific learning outcomes assessment rubric (ICC = .759 vs .748 and IRR = .693 vs .641) for two different scenarios. Most students found the scenario-specific rubric more helpful to their learning (59%), and easier to use (52%). Instructors (3/3) found the scenario-specific rubric more helpful to guide debriefing.

**Conclusions:** Scenario-specific rubrics may be more valuable for learners to help them identify their own knowledge and performance gaps and assist them in their preparation for simulation. Additionally, scenario-specific rubrics provide direction for both learners and instructors during debriefing sessions.

**Key Words:** Nursing education, Clinical simulation, Learning outcomes, Assessment rubrics, Reliability

## 1. BACKGROUND

Due to variations in clinical placements and experiences, not all nursing students have had the same practical experiences before graduation. Students may have limited exposure to clinical situations where they need to identify and manage a deteriorating patient.<sup>[1]</sup> Despite adequate knowledge, nursing students and postgraduate nurses may lack confidence and perform poorly in stressful or demanding situations, such as managing patient deterioration, due to a lack of real-life experience.<sup>[2]</sup> Well-designed clinical simulations allow all students to be exposed to a broader range of situations that may not occur during their clinical placements, can enhance nursing student critical thinking, and better prepare them to recognize and effectively respond to unfamiliar or emergency situations while witnessing the effects of their patient care

decisions in a setting where they do not need to be anxious about inflicting patient harm.<sup>[3-6]</sup>

Assessment is an important but often overlooked component of simulation education and research.<sup>[7]</sup> Simulation outcomes may be measured at the participant, patient or system level; key participant learning outcomes include changes in knowledge, skills, attitudes and behaviours.<sup>[8]</sup> The literature is overrepresented by studies that employ lower-level assessment such as participant-reported self-confidence and learning, and due to the lack of validated tools and metrics available for simulation evaluation, many nursing education programs develop and use evaluation instruments that are lacking in validity and reliability evidence.<sup>[9,10]</sup> Currently, there is a need to develop assessment methods that will en-

\*Correspondence: Marian Luctkar-Flude; Email: mfl1@queensu.ca; Address: School of Nursing, Queen's University, 92 Barrie Street, Kingston ON K7L 3N6, Ontario, Canada.

hance students' ability to adequately prepare for simulation as well as assess their performance.

The objectives of this study were: (i) to evaluate reliability and validity of a scenario-specific learning outcomes assessment rubric and a generic simulation performance assessment rubric; (ii) to describe instructor and learner satisfaction with the two types of rubrics; and (iii) to aid in selection of the most valid, reliable and acceptable assessment rubric for use in clinical simulations.

## 2. METHODS

This study used a mixed methods evaluation, consisting of faculty assessment, self-report scores, and narrative feedback. To evaluate the Learning Outcomes Assessment Rubrics, students participated in two different scenarios at two different time points. Face and content validity of scenario content and associated rubric criteria was conducted by clinical instructors in the nursing program. In addition, criterion descriptors (content validity) and external face validity was provided through peer review by three clinical content experts from the Ontario Simulation Alliance (OSA). Validity evaluations were conducted using the OSA Scenario Peer Review Form which is an adaption of the Loyalist College Simulation Peer Review Tool and the California Sim Alliance Scenario Validation Checklist. Feedback was reviewed, and adaptations to the rubrics were made through discussion. Finally, construct validity was reviewed by two educational developers from the Queen's University Centre for Teaching and Learning who reviewed the rubrics and provided feedback on the wording and levelling of the learning outcomes. Suggested measurement changes were also made prior to implementation of the study.

### 2.1 Scenarios and assessment rubrics

For this study, we selected two deteriorating patient scenarios that were previously developed using the Ontario Simulation Alliance (OSA) standardized simulation design method which features an embedded scenario-specific learning outcomes assessment rubric within each scenario.<sup>[11]</sup> The first scenario focused on urosepsis, while the second scenario focused on respiratory distress. The respiratory distress scenario had previously been implemented within a critical care nursing course and learners reported that integration of the learning outcomes assessment rubric into the simulation experience enhanced their self-regulated learning and presimulation preparation.<sup>[12]</sup>

The learning outcomes assessment rubric for the urosepsis scenario consisted of five learning outcomes, and the rubric for the respiratory distress scenario consisted of six learning outcomes specific to each scenario. Each rubric included

three levelled sets of descriptors or indicators that aligned with the learner's performance as "demonstrating the competency," "needing some improvement," or "needing major improvement."

The OSA assessment strategy is comprised of two phases. During the pre-simulation preparation phase the rubrics provide learners with the criteria upon which they will be assessed during the scenario. Completion of the rubrics prior to simulation may promote self-reflection and self-regulated learning with regard to the specific learning outcomes and result in learners who are better prepared to participate in the simulation. Learners who self-identify learning gaps may be more likely to complete other pre-simulation preparation activities such as assigned readings when they better understand how they align with the upcoming simulation. Following the simulation, learners complete the self-assessment rubrics again and compare their results with the instructor's ratings during the debriefing phase. It was unknown whether the scenario-specific learning outcomes assessment rubrics were a valid and reliable method to assess whether learning outcomes were obtained. As well, as it is time-consuming to develop and validate a scenario-specific rubric, it is important to know whether they provide added value to the learner and instructors when compared to a validated generic simulation assessment rubric. The Sweeney-Clark Simulation Performance Rubric<sup>[13]</sup> was selected as the generic comparator, as it: 1) was familiar to instructors, and 2) has previously been used as a performance measure in our simulation program. This rubric consists of 8 categories rated over 5 levels that align with Benner's Novice to Expert theory. Each rubric was designed to fit on a single page to facilitate scoring by students and instructors.

### 2.2 Setting and sample

A convenience sample of 29 BNSc nursing students from a Canadian University (year 4, n = 14; year 3, n = 12; year 2, n = 3) was recruited based on guidelines for designing reliability studies.<sup>[14]</sup> To determine inter-rater reliability (IRR), an estimated k of 20 to 30 subjects are required when there are 3 raters. Study participant ages ranged from 19 to 33 years (mean = 22.9 years), and the majority were female (90%).

### 2.3 Measurement and analysis

Following training by the primary author, three instructors rated students' performance on both scenarios and at both time points using Sweeney-Clark's Simulation Performance Rubric<sup>[13]</sup> and the scenario-specific Learning Outcomes Assessment Rubrics. One instructor facilitated and evaluated students in the simulation lab while the other instructors evaluated the same students by watching a video recording.

Instructors were rotated through the simulation lab such that each evaluated some of the simulations live and some by video. Students also rated themselves on the same rubrics both before and after participating in each simulation scenario.

All learners participated in both scenarios. The order of completion of the two scenarios was randomly assigned to the students. Learners participated with an embedded actor who played the role of a second nurse in the scenario who was there to help if needed. Learner performance was rated independently by three instructors that were blinded to the level of student i.e. they were unaware of which year in the program each student belonged. To establish the reliability and validity of the Learning Outcomes Assessment Rubrics, a number of analyses were conducted. Intraclass correlation coefficients (ICC) were calculated to measure IRR and were interpreted as follows: excellent (> .75), fair to good

(0.40-0.75) and poor (< 0.40).<sup>[15]</sup> Learner and instructor satisfaction were assessed with open-ended questions and a 12-item researcher-developed satisfaction scale with good internal consistency (Cronbach’s alpha = .83). Ethics approval was obtained from the Queen’s University Health Sciences and Affiliated Teaching Hospitals Research Ethics Board. Informed consent was obtained from all participants.

### 3. RESULTS

The order of completion of the two scenarios was randomly assigned to the students resulting in 55% completing Scenario 1 (urosepsis) first, and 45% completing Scenario 2 (respiratory distress) first. Inter-rater reliability was similar but slightly higher for the generic rubric versus the scenario-specific one (see Table 1); however, reliability was excellent for both rubrics for scenario one (ICC = .759 vs .748), and good for both rubrics for scenario two (ICC = .693 vs .641).

**Table 1.** Inter-rater reliability of simulation assessment rubrics (n = 3 raters; n = 29 nursing students)

Deteriorating Patient Scenario	Scenario-Specific Rubrics		Generic Rubric
	Learning Outcomes Assessment Rubric for Scenario 1	Learning Outcomes Assessment Rubric for Scenario 2	Sweeney-Clark Simulation Performance Rubric
Scenario 1: Urosepsis	ICC = .748, <i>p</i> < .001, (95% CI .49, .87)		ICC = .759, <i>p</i> < .001 (95% CI .55, .88)
Scenario 2: Respiratory Distress		ICC = .641, <i>p</i> < .001 (95% CI .33, .82)	ICC = .693, <i>p</i> < .001 (95% CI .43, .85)

\*ICC = Intraclass correlation coefficient

Use of the assessment rubrics was highly acceptable to learners for both pre (94%) and post (100%) assessment with the majority finding the scenario-specific rubric more helpful to their learning (59%), and easier to use (52%) than the generic rubric. Instructors (3/3) rated the scenario-specific rubric more helpful to guide debriefing but rated the generic rubric (2/3) easier to use.

Qualitative feedback from both instructors and learners highlighted the usefulness of the scenario-specific rubrics to support pre-simulation preparation (see Tables 2 and 3). Feedback generally focused on the ease of use and contributions to learning of the two different types of rubrics. The generic rubric was seen as “easier to use” and made it “easier to compare scores between scenarios,” whereas the scenario-specific rubric better “identifies expectations of learners,” but was also “more time consuming to complete” and “more difficult to grade.” Instructors noted the scenario specific rubrics provided “areas to reflect on during debrief,” “promoted self-regulated learning,” and “may decrease anxiety going into simulation.”

### 4. DISCUSSION

This study compared two types of simulation assessment rubrics in terms of their reliability, validity and acceptability to learners and instructors. Reliability and validity of the generic Sweeney-Clark Simulation Performance Rubric was previously reported, with IRR established using standardized simulation videos resulting in an overall ICC of 0.92.<sup>[13]</sup> Inter-rater reliability for the Sweeney-Clark rubric in our study was lower but still good (ICC = .759 and .693) and was similar to the IRR for our scenario-specific rubric (IRR = .748 and .641). Additional rater training with standardized simulation videos may improve IRR with both types of rubrics.

Based on the quantitative survey results as well as the qualitative feedback, it was clear that the majority of students preferred the scenario-specific rubric. They appreciated the added details that “shows expectations for simulation” versus the generic rubric that was perceived as more “vague.” Both learners and instructors found the generic rubric “easier to use.” Thus, it is not possible to recommend one type of

rubric over another based on the results of one study. Instructor preference, level of the learner, and learning outcomes for the scenario may be taken into account when selecting a rubric. Novice learners may benefit more from having a scenario-specific learning outcomes assessment rubric for each simulation, whereas, senior learners may be better able to self-assess using a generic rubric that is used for all their simulations. Further research is needed to clarify whether

this is the case. Whichever rubric is selected, it is important that instructors are trained in its use and that IRR is established prior to using it for any summative assessments. Both types of rubrics may be used to guide self-reflection and provide a structure for formative assessment during debrief. Repeated use of a rubric throughout a nursing program may promote a better understanding of the learner’s competency level as it evolves.<sup>[13]</sup>

**Table 2.** Simulation rubric instructor feedback

	Helpful aspects of rubrics	Least helpful aspects of rubrics
Generic rubrics	<ul style="list-style-type: none"> <li>• Headings in the generic rubric are much easier to mark and can spark discussion during debrief about areas for improvement</li> <li>• Includes a strengths and area for improvement section</li> <li>• Easier to use</li> <li>• Easier to compare scores between scenarios</li> </ul>	<ul style="list-style-type: none"> <li>• Not enough detail for the learners to prepare</li> <li>• Learners may have difficulty rating themselves</li> </ul>
Scenario specific rubrics	<ul style="list-style-type: none"> <li>• Rubric descriptors help to define learner level of performance</li> <li>• Learning outcomes help to guide debrief</li> <li>• Identifies expectations of learners</li> <li>• Helpful to both instructors and learners</li> <li>• Directs areas to reflect on during debrief so the student can recognize areas requiring improvement themselves, then also based on opinions of others</li> <li>• Students should have them beforehand to review so it is clear what is expected in the scenario</li> <li>• Scenario specific categories may be more tailored to each scenario</li> <li>• Promotes self-regulated learning</li> <li>• May decrease anxiety going into simulation if students know better what to expect</li> </ul>	<ul style="list-style-type: none"> <li>• Time consuming to complete when there are too many objectives being evaluated, this should become easier as instructor becomes familiar with the objectives &amp; descriptors</li> <li>• Difficult to follow at times while observing the simulation if there is a lot of information on the rubric</li> <li>• Some students fall under more than one category for each competency</li> <li>• They may start out well and end up with difficulties, then it is hard to rate</li> <li>• I am not sure how grading themselves before is helpful</li> <li>• I find it more difficult to grade when a student has some points in each learning outcome</li> </ul>

**Table 3.** Simulation rubric learner feedback

	Helpful aspects	Least helpful aspects
Generic rubrics	<ul style="list-style-type: none"> <li>• Much easier if they’re to the point</li> <li>• The second rubric (generic) was easier because it was more specific for each point, and less to read</li> <li>• Identifying where I should be at &amp; comparing that to my performance</li> </ul>	<ul style="list-style-type: none"> <li>• Prepare you for general skills but not specifics</li> <li>• Sometimes they are too general and so they are not specific to the scenario</li> <li>• It is hard if it is a general assessment because certain scenarios it is hard to apply to</li> <li>• The vague (generic) rubric was not as helpful</li> <li>• Priorities in one simulation may differ from another</li> <li>• The lack of specifics, how you lose marks in simulations</li> </ul>
Scenario specific rubrics	<ul style="list-style-type: none"> <li>• Knowing what is expected of you</li> <li>• I find the competencies listed before going into the simulation helpful for what I will be performing and to know what skills to perform</li> <li>• Simulation assessment rubrics outline specific tasks and expectations to be met making it easy to see where the nurses strengths and weaknesses lie</li> <li>• I found it was helpful to assess myself because then I could identify areas that I need to improve on myself</li> <li>• Getting to think critically and respond to the patient’s symptoms and complaints using the rubric as a guide</li> <li>• Specific rubric is superior</li> <li>• Shows expectations for simulation</li> <li>• Allows students to see competency level, what exactly they are being assessed on</li> <li>• Only 3 categories, straightforward</li> <li>• Helpful when it is more specific to the individual scenario</li> </ul>	<ul style="list-style-type: none"> <li>• Lengthy descriptions</li> <li>• Had a lot to read and was a little confusing</li> <li>• If you can only circle one box when it had multiple points may not truly represent your abilities</li> <li>• Too few boxes (e.g. least competent, competent, exceeds expectations) may be too black and white, most wouldn’t fall in one box, multiple boxes help find middle ground</li> <li>• Many words/options led me to skim a lot and not fully read them by the end</li> <li>• Simulation assessment rubrics may not take into account every action (either good or bad) that a nurse makes</li> </ul>

A limitation of the study was the requirement for the instructors to evaluate the learners using two different assessment rubrics at the same time. This was less of an issue for instructors when they reviewed video recordings of the simulations as they were able to pause and/or rewind the videos as needed. Despite this limitation, good IRR was obtained, and the feasibility of using either rubric for real-time assessment was demonstrated.

## 5. CONCLUSION

This study addresses a gap in the nursing education literature by illustrating that simulation assessment can be used to scaffold learner pre-simulation preparation, and self-assessment while remaining both valid and reliable. Study results suggest scenario-specific rubrics may be more valuable for learners as they help learners identify their own knowledge and performance gaps and assist them in their preparation for simulation. Additionally, scenario-specific rubrics provide direction for both learners and instructors during debriefing

sessions. Using valid, reliable and acceptable assessment rubrics for both instructors and nursing students can provide us with valuable assessment data on both the learning outcomes and the quality of the simulation scenarios in promoting achievement of learning outcomes. Integration of learning outcomes assessment rubrics into simulation design will enhance our ability to evaluate the contribution of clinical simulation. It will also allow us to better prepare nursing graduates to transition to practice with the skills and experiences they need to manage deteriorating patient scenarios. Further research is needed to determine to what extent generic and scenario-specific assessment rubrics contribute to learner motivation and self-regulation during the presimulation preparation phase, and how to better engage learners. The scenarios and validated rubrics are available to nursing faculty from across the province of Ontario through a centralized shared repository.

## CONFLICTS OF INTEREST DISCLOSURE

The authors declare that there is no conflict of interest.

## REFERENCES

- [1] Bogossian F, Cooper S, Cant R, et al. Undergraduate nursing students' performance in recognising and responding to sudden patient deterioration in high psychological fidelity simulated environments: An Australian multi-centre study. *Nurse Education Today*. 2014; 34(5): 691-6. PMID:24183634 <https://doi.org/10.1016/j.nedt.2013.09.015>
- [2] Cooper S, Cant R, Porter J, et al. Managing patient deterioration: assessing teamwork and individual performance. *Emergency Medicine Journal*. 2013; 30(5): 377-81. PMID:22609696 <https://doi.org/10.1136/emermed-2012-201312>
- [3] Buykx P, Kinsman L, Cooper S, et al. FIRST2ACT: Educating nurses to identify patient deterioration — A theory-based model for best practice simulation education. *Nurse Education Today*. 2011; 31(7): 687-93. PMID:21481992 <https://doi.org/10.1016/j.nedt.2011.03.006>
- [4] Chua WL. Simulation training appears to improve nurses' ability to recognise and manage clinical deterioration. *Evidence Based Nursing*. 2017; 20(4): 122-3. PMID:28794005 <https://doi.org/10.1136/eb-2017-102742>
- [5] Connell CJ, Endacott R, Jackman JA, et al. The effectiveness of education in the recognition and management of deteriorating patients: A systematic review. *Nurse Education Today*. 2016; 44: 133-45. PMID:27429343 <https://doi.org/10.1016/j.nedt.2016.06.001>
- [6] Lewis CB, Vealé BL. Patient Simulation as an Active Learning Tool in Medical Education. *J Med Imaging Radiat Sci*. 2010; 41(4): 196-200. PMID:31051879 <https://doi.org/10.1016/j.jmir.2010.09.002>
- [7] Diekelmann N, Ironside PM. Developing a science of nursing education: Innovation with research. *Journal of Nursing Education*. 2002; 41(9): 379-80.
- [8] Jeffries PR, Rodgers B, Adamson K. NLN Jeffries Simulation Theory: Brief Narrative Description. *Nursing Education Perspectives*. 2015; 36(5): 292-3. PMID:26521496 <https://doi.org/10.5480/1536-5026-36.5.292>
- [9] Adamson KA. Evaluation tools and metrics for simulations. In: Jeffries PR, editor. *Clinical simulations in nursing education: Advanced concepts, trends and opportunities*. Baltimore, MD: National League for Nursing; 2014. p. 145-64.
- [10] Haddeland K, Slettebø Å, Carstens P, et al. Nursing Students Managing Deteriorating Patients: A Systematic Review and Meta-Analysis. *Clinical Simulation in Nursing*. 2018; 21: 1-15. <https://doi.org/10.1016/j.ecns.2018.05.001>
- [11] Egan R, Luctkar-Flude M, Sears K, et al. Enhancing Registered Nurse job readiness through constructive alignment and standardization in simulation design. 34th Society for Teaching and Learning in Higher Education Conference: Transforming our Learning Experiences; Kingston, ON: Society for Teaching and Learning in Higher Education; 2014.
- [12] Luctkar-Flude M, Tregunno D, Egan R, et al. Integrating a learning outcomes assessment rubric into a deteriorating patient simulation for undergraduate nursing students. *Journal of Nursing Education and Practice*. 2019; 9(8): 65. <https://doi.org/10.5430/jnep.v9n8p65>
- [13] Sweeney NL, Clark Rollins M, Gantt LT, et al. Development and reliability testing of the Sweeney-Clark Simulation Evaluation Rubric. *Clinical Simulation in Nursing*. 2020; 41: 22-32. <https://doi.org/10.1016/j.ecns.2019.04.002>
- [14] Shoukri MM, Asyali MH, Donner A. Sample size requirements for the design of reliability study: review and new results. *Statistical Methods in Medical Research*. 2004; 13(4): 251-71. <https://doi.org/10.1191/0962280204sm365ra>
- [15] Oremus M, Oremus C, Hall GBC, et al. Inter-rater and test-retest reliability of quality assessments by novice student raters using the Jadad and Newcastle-Ottawa Scales. *BMJ Open*. 2012; 2(4): e001368. PMID:22855629 <https://doi.org/10.1136/bmjopen-2012-001368>