Featured Article

Examining Validity, Fidelity, and Reliability of Human Patient Simulation

Deb Shelestak, PhD, RN[a],*, Barbara Voshall, DNP, RN, CNS[b]
[a]Assistant Professor, Nursing, Kent State University at Stark, North Canton, OH 44720, USA
[b]Professor, Nursing, Graceland University, Independence, MO 64050, USA

Abstract: Human patient simulation is increasingly being used in summative or high-stakes performance evaluation. Given the significance of the decisions made, based on performance, it is essential that there is demonstrated content validity of the human patient simulation scenario, implementation fidelity by the facilitators, and reliability of the scoring rubric and raters before its use in decision making. The focus of this article is to define these three measurement concepts and describe methods commonly used to examine them. Resources are also suggested for those who are interested in more information about these concepts.

The use of human patient simulation (HPS) to examine various aspects of participant behavior, that is, content knowledge, skills, and teamwork, has been rapidly increasing. According to the recently established International Nursing Association for Clinical Simulation and Learning Standard 7, when using HPS in summative or high-stakes assessment, it is essential to ensure that the decisions made about a participant's performance are accurate and consistent (Sando et al., 2013). As you would in intervention studies, attention to the HPS design and evaluation instruments are necessary to control for extraneous, confounding variables, and intervening variables.

When designing an experimental intervention study, a researcher needs to consider the internal validity of the design. Internal validity is the degree to which the results of the study can be attributed to the intervention of interest and no other extraneous variables. There are several potential threats to internal validity, one being poor instrumentation. In research studies, reliability and validity of an instrument must be examined and found acceptable before the beginning of data collection (Polit & Beck, 2012). Fidelity of the intervention is also essential to ensure the findings are valid interpretations (Santacrose, Maccarelli & Grey, 2004).

The same concerns apply in situations where high-stakes decisions are made based on performance. In the design and implementation of simulation exercises, attention to reliability and validity issues is essential. Those involved in HPS should carefully consider these issues in the

simulations they develop and facilitate. The purpose of this article is to describe techniques to confirm the content validity of the scenario, fidelity of the HPS, and reliability of the observations and/or ratings before the simulation.

<div style="border:1px solid">

Key Points

- Whenever judgments are linked to performance during an HPS, it is essential that the validity of the scenario, fidelity of the intervention, and reliability of the decisions are examined before the simulation.
- In this era of high-stakes testing and evaluation of student outcomes, it is paramount that faculty take the necessary steps to ensure the simulation experience provides an accurate and consistent measure of student ability.
- The purpose of this article is to describe techniques to confirm the content validity of the scenario, fidelity of the HPS, and reliability of the observations and/or ratings before the simulation.

</div>

Content Validity

There are three types of instrument validity, each of which is more complex in nature: content, criterion-related, and construct (Hand, 2004). An instrument demonstrates content validity when it measures the designated concept, that is, depression, self-esteem, health belief behaviors, and is usually the first type of instrument validity established (Polit & Beck, 2012). In HPS, content validity of the simulation experience is also important to consider. This includes all components of the simulation exercise. For example, does the scenario describe the situation accurately and sufficiently? Do the diagnostic cues lead to appropriate actions or interventions by the participants?

There are two common techniques used to determine content validity. The first is to send the scenario to experts for feedback. It is important to consider the participants when defining "expert." For example, if this scenario was designed for students in a nursing foundations course, faculty who routinely teach that level of student would be most appropriate. On the other hand, if this was a simulation for competency testing of clinical employees, experts would be practitioners familiar with the facility's policies and procedures. Based on the feedback, adjustments to the scenario should be made before the HPS exercise.

In situations where a more precise measure of content validity is needed, a content validity index (CVI) can be calculated. The CVI is normally used during instrument development and is an index based on multiple experts' rating of relevance of the instrument (Polit & Beck, 2012).

Table	Example of CVI Ratings			
	Rater 1	Rater 2	Rater 3	Average Rating
Q1	4	4	4	4
Q2	3	3	4	3.3
Q3	3	3	3	3
Q4	3	2	2	2.3[*]
Q5	3	3	3	3

Based on 1 = strongly disagree to 4 = strongly agree.
 * Average rating <3 would revise the scenario based on rater feedback.

Typically the expert is asked to judge the relevance of each item (or question) using a 0 to 3 scale: 0 = not relevant, 1 = slightly relevant, 2 = somewhat relevant, and 3 = very relevant. The level of agreement for each item is then calculated. An excellent discussion of the CVI can be found in Polit, Beck, and Owen (2007).

In HPS, there are no individual items or questions to be judged. There are, however, critical questions that can be asked and rated by the experts. At any point where participant behavior is expected and assessed, information provided during the scenario should guide the participant to the desired judgment or intervention. For example, in a scenario in which a vasovagal incident results in severe bradycardia, the following statements can be evaluated to determine the content validity, using a four-point Likert-type scale (strongly disagree to strongly agree): (a) There was sufficient information in the patient history; (b) The diagnostic information was appropriate for the situation; (c) The "patient" demonstrated significant changes to warrant specific action; (d) The following interventions are appropriate for the situation. (Provide the list of actions you want to see from participants.); (e) There was a sufficient change in the "patient's" condition to evaluate impact of interventions; and (f) The simulation scenario is an accurate representation of the described clinical condition.

Using the vasovagal example from above, you receive the following results from three experts (Table).

By averaging the scores for each statement, you can determine the level of agreement among the experts. For each question, the score should be ≥3 (if using a 1 to 4 scale). In other words, your experts should agree or strongly agree with each statement (Polit & Beck 2006). It is important for the experts to provide feedback or justification for a less than perfect score. This feedback allows you to make changes to the scenario before using the HPS.

Implementation Fidelity

Intervention fidelity is defined as the degree to which the intervention is delivered as designed. In randomized controlled trials, fidelity of the intervention is essential to ensure credibility of the research findings. Protocols,