

Distribution Statement

Distribution A: Public Release.

The views presented here are those of the author and are not to be construed as official or reflecting the views of the Uniformed Services University of the Health Sciences, the Department of Defense or the U.S. Government.

RELIABILITY IN HEALTHCARE SIMULATION SETTING:
A DEFINITIONAL REVIEW

by

Sunny Yauger, BS, CHSE

Thesis submitted to the Faculty of the
Health Professions Education Graduate Program
Uniformed Services University of the Health Sciences
In partial fulfillment of the requirements for the degree of
Master in Health Professions Education 2020



UNIFORMED SERVICES UNIVERSITY, SCHOOL OF MEDICINE GRADUATE PROGRAMS
Graduate Education Office (A 1045), 4301 Jones Bridge Road, Bethesda, MD 20814



March 5, 2020

APPROVAL SHEET

Title of Dissertation: **Reliability in Healthcare Simulation Settings: A Definitional Review**

Name of Candidate: **Sunny Yauger, Master of Science in Health Professions Education**

Date **03/05/2020**

THESIS AND ABSTRACT APPROVED:

[Redacted Signature]

Dario Torre, PhD, MD
DEPARTMENT OF MEDICINE
Committee Chairperson

DATE:

3/5/20

[Redacted Signature]

Abigail Konopasky, PhD
DEPARTMENT OF MEDICINE
Thesis Adviser

3/5/20

[Redacted Signature]

Alexis Battista, PhD
DEPARTMENT OF MEDICINE
Committee Member

March 5, 2020

ACKNOWLEDGMENTS

I wish to express my gratitude to Abby and Alexis for their support and guidance.

DEDICATION

To my husband Joshua, and my children Matthew and Emily for your infinite love and support.

COPYRIGHT STATEMENT

The author hereby certifies that the use of any copyrighted material in the thesis manuscript entitled: Reliability in Healthcare Simulation Setting: A Definitional Review is appropriately acknowledged and, beyond brief excerpts, is with the permission of the copyright owner.



Sunny J. Yauger

March 19, 2020

DISCLAIMER

The views presented here are those of the author and are not to be construed as official or reflecting the views of the Uniformed Services University of the Health Sciences, the Department of Defense or the U.S. Government.

ABSTRACT

Reliability in Healthcare Simulation Setting: A Definitional Review

Sunny Jade Yauger, Master in Health Professions Education, 2020

Thesis directed by: Dario Torres, MD, PhD, Associate Professor, Graduate Programs in Health Professions Education (HPE)

Problem: Reliability in health professions education does not account for social interactions when using healthcare simulation as an assessment tool, thus challenging the traditional concept of reliability in health professions education.

Methods: This mixed method definitional review explored literature through a constructivist lens and grounded theory. Exploration was framed around the concept of healthcare simulation as social setting where participants interact with simulation setting attributes.

Results: Reliability is not directly defined but described by the types of interactions among simulation design attributes and simulated participant performance nuances that introduce errors into a simulation setting Training approaches are design to mitigate errors generated by simulation setting attributes.

Conclusion: Reliability terms are used to describe how training approaches are operationalized to achieve simulation setting reliability performance consistency and accuracy.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER 1: INTRODUCTION	1
BACKGROUND	1
THEORETICAL FRAMEWORK	3
PURPOSE STATEMENT AND RESEARCH QUESTIONS	3
REFERENCES	5
CHAPTER 2: RELIABILITY IN HEALTHCARE SIMULATION: A	
DEFINITIONAL REVIEW	6
SUBMITTED AS:	6
ABSTRACT	6
What is already known on this subject?	7
What this paper adds?	7
INTRODUCTION	9
The Construct of Reliability in HPE and Simulation.....	10
Designing for Reliability in Simulation.....	11
The Current Study.....	12
METHODS	13
Study Design.....	13
Search Strategy	13
Article Selection Method	14
Techniques to Enhance Trustworthiness	15
Reflexivity.....	16
RESULTS	16
Article Selection.....	16
Article Demographics	16
Word Frequency of Reliability Terms	17
Qualitative Findings.....	17
Terms tied to the concept of simulation reliability	18
Attributes interfering with reliability	19
Training approaches.....	19
Training for performance standardization via a standardized training strategy	20
Action-Based research development procces to convert children simulated	
patients into standarized patinets.....	21
Training for performance standardization consistency and accuracy via a	
standardize training strategy.....	21

Training for performance accuracy to mitigate inaccuracies of patient portrayal	22
Developing standardized patients as raters for performance consistency and accuracy	23
Developing unannounced standardized patient for consistency and accuracy in performance assessment	24
DISCUSSION	25
Implications for Reliability in Healthcare.....	25
Implications for differences in use of the terminology in HPE and healthcare simulation.....	26
Limitations	26
Conclusion	27
Research Recommendations	27
Acknowledgements.....	28
Prior Presentations	28
Funding	28
Ethics Approval and Consent to Participate Consent for Publication	28
Competing Interests	28
Figure 6. Training for performance standardization via a standardized training strategy	33
REFERENCES	39
CHAPTER 3: DISCUSSION.....	42
Military Relevance.....	46
REFERENCES	47
APPENDIX I: PUBMED ARTICLE SEARCH STRATEGY	49
APPENDIX II: CONSTRUCTS OF SIMULATION SETTING RELIABILITY.....	50
REFERENCES	59

LIST OF TABLES

Table 1. Demographic summary of healthcare simulation setting attributes.....	37
Table 2. Healthcare simulation setting reliability conceptualized by author.....	38
Table 3. Reliability terms used by authors.....	38

LIST OF FIGURES

Figure 1. Healthcare Simulation Setting Framework	29
Figure 2. Article Selection Flow Chart	29
Figure 3. Published healthcare article that describe a healthcare simulation setting framework	30
Figure 4. Reliability terms retrieved from a word frequency search within all PubMed articles searched	31
Figure 5. Interactions influencing simulation setting reliability	32
Figure 7. Action-Based research development proces to convert children simulated patients into standarized patinets	34
Figure 8. Training process to standardize patient performance accuracy (Baig, 2014) ...	34
Figure 9. Training for performance accuracy to mitigate inaccuracies of patient portrayal	35
Figure 10. Developing standardized patients as raters for performance consistency and accuracy	35
Figure 11. Training process to development unannounced standardized patients to rate practitioner performance	36

CHAPTER 1: INTRODUCTION

BACKGROUND

Reliability plays a central role in educational assessment and social science research where reliability in health professions education (HPE) provides the foundation for trustworthy evidence needed to inform and enhance effective practices.(1) The concept of reliability is regularly discussed in HPE and healthcare simulation-based education (SBE).(1-4) However, interpretations in these two fields may indicate an implicit assumption that the same idea of reliability in HPE transcends to healthcare simulation settings. Thus, further research is needed to clarify the terminology and concept of reliability when applied to a healthcare simulation setting. Untangling the lexical differences in “reliability” between HPE, and simulation-based education and research may avoid misinterpretations and assist those wishing to design reliable simulation settings. Furthermore, there is a lack of evidence on how to carry out simulation design and achieve operational reliability for simulation-based research.(5)

In medical education, reliability refers to the degree to which test or assessment data can be replicated or reproduced.(2) The concept indicates that the reproduced scores of an individual across numerous repeated measures will cancel out random error.(1) Through this process, you reach an individual’s “true score.” When an individual’s true score or true scores across groups of individuals are in agreement, the connections can be translated as a quantifiable measure to estimate reliability. For assessment scores to be interpreted or taken seriously, they must be reproducible; if not, then a meaningful interpretation is questionable.(1) In SBE, quantifying the reliability of high-stakes performance tests (e.g., objective structured clinical examinations) helps

identify the primary construct being measured as opposed to random errors generated by contextual factors.

In healthcare simulation, in comparison to the assessment literature, reliability has been defined as the “consistency of a simulation activity or the degree to which a simulation activity is measured, in the same way, each time it is employed under the same conditions with the same participants.”(6) In other words, in HPE more broadly, reliability refers to the consistency and accuracy of the measurement tool (7), while in the specific case of a simulation design, it refers to the consistency and accuracy of the simulation setting developed. Munroe et al. (5) contend, “to achieve reliability in simulation, settings must be designed and planned so they may be reproduced consistently to ensure all participants are exposed to the same conditions.”

In designing for reliability, healthcare simulation designers often consider the *characteristics* of a simulation. Characteristics are often discussed through such constructs as fidelity, scenario complexity, participant support, and debriefing.(8) Another construct designer consider are *mediating artifacts* influencing participant actions, like participant roles, clinical practices, and tools such as equipment and cognitive aids.(9) Lastly, designers consider the *environment*, for example, in-situ simulations or a mobile simulation which actively moves from one environment to another.(10)

According to Dieckmann and colleagues, healthcare simulation can be viewed as a *social practice* where participants interact in a goal-oriented fashion with the simulation setting’s constructs (e.g., characteristics, mediating artifacts, and the environment).(11) Implementation brings these constructs together and operationalizes a

dynamic setting. Given this dynamic setting, questions arise about whether reproducing a simulation setting can consistently yield the same outcomes or phenomena in simulation-based education or simulation-based research.

THEORETICAL FRAMEWORK

In this mixed method definitional literature review, this research was explored through a constructivist paradigm and grounded theory methods. A constructivist paradigm views knowledge as actively constructed and co-created through human interactions and relationships.(12) In this view, healthcare simulation is a social practice where participants interact in a goal-oriented fashion with setting attributes to elicit a behavioral response.(11)

PURPOSE STATEMENT AND RESEARCH QUESTIONS

A concern with simulation is its vulnerability to error stemming from its design flexibility and the impact of social factors when participants freely interact and engage.(13) Historically, efforts to account for measurement error have focused on reproducibility, therefore, there are assumptions that by implementing standardized and reproducible simulation settings, participants will be exposed to the same setting conditions, thus achieving reliability. The ecological complexity of a simulation setting creates a polarity between the traditional concept of reliability and its assumed implications that healthcare simulation settings are reproducible. Additionally, there are numerous ‘sister terms’ frequently used (e.g., fidelity) that make it even more complex to understand how reliability is realized in healthcare simulation. Therefore, the purpose of the thesis was to examine how reliability has been defined in healthcare simulation and what additional related terms are tied to it. The proposed research questions are:

1. How has reliability in healthcare simulation been defined and described in the literature?
2. What other relevant terms are tied to the concept of reliability in a simulation setting?

REFERENCES

1. Downing SM, Yudkowsky R. Assessment in Health Professions Education: Routledge; 2009 2009/5/7. 344 p.
2. Downing SM. Reliability: on the reproducibility of assessment data. Medical education. 2004;38(9):1006-12.
3. Brannick MT, Erol - Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. Medical education. 2011;45(12):1181-9.
4. Norman GR, Vleuten C, Newble D. International Handbook of Research in Medical Education: Springer Science & Business Media; 2002.
5. Munroe B, Buckley T, Curtis K, Morris R. Designing and implementing full immersion simulation as a research tool. Australasian Emergency Nursing Journal. 2016;19(2):90-105.
6. 2016. Healthcare Simulation Dictionary.
7. Polit DF, Beck CT. Essentials of nursing research: Appraising evidence for nursing practice: Lippincott Williams & Wilkins; 2009.
8. Groom JA, Henderson D, Sittner BJ. NLN/Jeffries simulation framework state of the science project: Simulation design characteristics. Clinical Simulation in Nursing. 2014;10(7):337-44.
9. Battista A. Activity theory and analyzing learning in simulations. Simulation & Gaming. 2015;46(2):187-96.
10. Riley RH. Manual of simulation in healthcare: OUP Oxford; 2008.
11. Dieckmann P, Gaba D, Rall M. Deepening the theoretical foundations of patient simulation as social practice. Simul Healthc. 2007;2(3):183-93.
12. Watling CJ, Lingard L. Grounded theory in medical education research: AMEE Guide No. 70. Medical teacher. 2012;34(10):850-61.
13. Downing SM, Yudkowsky R. Assessment in health professions education: Routledge; 2009 2009.

CHAPTER 2: RELIABILITY IN HEALTHCARE SIMULATION: A DEFINITIONAL REVIEW

SUBMITTED AS:

Yauger SJ, Konopasky AW, Battista A. Reliability in Healthcare Simulation: A
Definitional Review (manuscript under review)

ABSTRACT

Background: The construct of reliability in health professions education serves as a measure of the congruence of interpretation across assessment tools. When used as an assessment strategy, healthcare simulation serves to elicit specific participant behaviors sought by medical educators. In healthcare simulation, reliability often refers to the ability to consistently reproduce a simulation and that reproducing a simulation setting can consistently expose participants to the same conditions, thus achieving simulation reliability. However, some articles have expressed that simulations are vulnerable to error stemming from design conceptualization to implementation, and the impact of social factors when participants interact and engage with others during participation. The purpose of this definitional review is to examine how reliability has been conceptualized and defined in healthcare simulation, and how the attributes of simulations may present challenges for the traditional concept of reliability in health professions education.

Method: Data collection and analysis was approached through a constructivist perspective and grounded theory strategies. To identify articles for analysis, articles between 2009-2019 were filtered applying keywords related to simulation development

and design. Data winnowing was structured around a framework viewing simulation as a social practice where participants interact with simulation setting attributes.

Results: The central theme encompassing healthcare simulation reliability is how stimulation setting attributes influence simulation setting reliability. From this view, reliability is not directly defined but described by the interactions between simulation design attributes and simulated participant performance nuances. Training approaches aim to refine simulated participant performance to mitigate errors generated by setting attributes. Different lexical terms used to describe how to operationalize training approaches and relationship between reliability constructs are examined

Conclusion: Simulation setting reliability that is supposed to provide participants fair and trustworthy conditions to achieve performance competency set by medical educators. However, simulation setting reliability is dependent on the manner in which simulated participants are trained to achieve performance reliability.

Key terms: definitional review, simulation, reliability, reproducibility, consistency, replicability, performance, interference, attribute, design, setting, standardized patient, simulated patient, simulated participant

WHAT IS ALREADY KNOWN ON THIS SUBJECT?

- The use of healthcare simulation to assess performance competency is increasingly common
- The types of attributes incorporated into a simulation setting design
- The uses for incorporating simulated participants into a simulation design
- Simulated participants require training prior to simulation setting implementation

WHAT THIS PAPER ADDS?

- How the constructs of 'reliability' is conceptualized in healthcare simulation
- Identifies the attributes of a simulation setting that are most commonly thought to interfere with reliability in healthcare simulation
- Identifies additional relevant terms that are related to reliability in healthcare simulation
- Examination of training approaches to achieve s simulated participant performance reliability and training to mitigates error generated by simulation attributes strategies that have been employed to improve simulation reliability and accuracy

INTRODUCTION

A concern with healthcare simulation is its vulnerability to error stemming from its implementation flexibility and the impact of context related factors (e.g., availability of diagnostic tools, simulator delay or troubleshooting) when participants interact and engage in a simulation setting.(1) Adding to this are social interactions between participants and simulated participants. For instance, variability may arise from simulated participants performance. A simulated participant (SP) is an “inclusive term to include all role players in any simulation context” (2) with varying levels of experience or when they are selected from different population pools (e.g. faculty, medical students, children). These dynamic interactions can subject participants' goal-oriented activities to measurement errors when making critical decisions about their academic progression and licensure.

Viewed as a system, the ecological complexity of a simulation setting may not afford its capture by traditional concept of reliability, particularly the assumed implications that healthcare simulation settings can be replicated by others.(3) In health professions education (HPE), the construct of reliability traditionally focuses on education outcome measurement (e.g., education achievement), something that does not extend to include the social aspects of a simulation. Thus, a healthcare simulation setting challenges the traditional concept of reliability in HPE, raising an important question in simulation-based education and research: how do we define “simulation reliability”? In this study, I examine how reliability has been defined (explicitly and implicitly) in healthcare simulation and what additional related terms are tied to it.

The Construct of Reliability in HPE and Simulation

Reliability plays a central role in education assessment and social science research where “reliable data in HPE provides the foundation for trustworthy evidence needed to inform and enhance effective practices”.⁽¹⁾ This core concept in HPE indicates that the reproduced scores of an individual across numerous repeated measures will cancel out random error. ⁽⁴⁻⁸⁾ This process results in an individual’s “true score.” When an individual’s true score or collective true scores across groups are in agreement, the connections can be translated to a quantifiable measure to estimate reliability. For assessment scores to be interpreted or taken seriously, they must be reproducible; if not, then a meaningful interpretation is questionable.⁽¹⁾ In healthcare simulation, this same idea—quantifying the reliability of high-stakes performance tests (e.g., objective structured clinical examinations) helps identify the primary construct being measured (e.g., education achievement) as opposed to random errors generated by contextual factors (e.g., training location and the purpose of the simulation).⁽⁹⁾

In healthcare simulation, in comparison to the HPE assessment literature, reliability has been defined as the “consistency of a simulation activity or the degree to which a simulation activity is measured in the same way each time it is employed under the *same* conditions with the same participants”.⁽¹⁰⁾ This approach assumes that by implementing the same designed simulations, participants will be exposed to the same conditions, thus achieving a fair and trustworthy experience for each participant to optimally perform their activity driven goals towards competency achievement.

Furthermore, when discussing standardized test conditions, McGaghie indicates that examinees are the only “variable” and that “fixed” conditions pertain to setting attributes.(11) In other words, in HPE more broadly, reliability refers to the consistency and accuracy of the *measurement tool*, while in the specific case of simulation design, it refers to the consistency and accuracy of the simulation *setting* developed.(12, 13) Yet, uses of the concepts and terms of simulation reliability do not appear to be fixed across studies; further research is needed to clarify the concept and terminology of reliability when applied to a healthcare simulation setting. Moreover, there is a lack of evidence on how to carry out simulation design and achieve operational reliability for simulation-based research.(12) Untangling the lexical differences in “reliability” between HPE and simulation-based education and research may avoid misinterpretations and assist those wishing to design reliable simulation settings.

Designing for Reliability in Simulation

One of the most common ways the construct of reliability has been discussed in the healthcare simulation literature is with regards to simulation design practices. For example, Munroe et al. (12) contend that, “to achieve reliability in simulation, settings must be designed and planned so they may be reproduced consistently to ensure all participants are exposed to the same conditions”.(12) In designing for reliability, healthcare simulation designers often consider simulation attributes. Attributes are often discussed as artifacts that mediate participant activity goals and influence social practices, for example, the scope of simulated participants, medical equipment, simulator modality. (14, 15) Attributes, however, do not exist in isolation. Their properties permeate throughout a setting, interacting with each other and influencing participant interactions

as part of a dynamic system.(16) In other words, the design of a simulation can be controlled but the influence of certain attributes cannot. Consequently, a simulation can be designed from a standardized perspective, however, as a complex system, the whole setting is not equal to the sum of its parts. The dynamics of a simulation setting lead to activity variance challenging the concept of reliability in healthcare simulation as it is traditionally used in HPE.

The Current Study

The purpose of this study was to examine how reliability has been defined in healthcare simulation and what additional related terms are tied to it. The proposed research questions were:

1. How has reliability in healthcare simulation has been defined and described in the literature?
2. What other relevant terms are tied to the concept of reliability in the context of a simulation setting?

This research aims to raise awareness and open a debate about how the attributes of a simulation setting challenge the traditional concept of reliability in HPE. Defining reliability in healthcare simulation could enhance communication and lead to more effective interactions among global medical educators, healthcare simulation educators, and researchers as they seek to create reliable simulations. Furthermore, these discussions may lead to research on how reliability for a simulation setting is achieved in simulation-based education and research and provide a foundation for gathering trustworthy evidence in the context of a simulation setting. This is an increasingly common focus for

HPE research and scholarship. To the best of our knowledge, this is the first definitional review of reliability in healthcare simulation.

METHODS

Study Design

In this mixed method definitional literature review, this research was explored through a constructivist paradigm and grounded theory methods. A constructivist paradigm views knowledge as actively constructed and co-created through human interactions and relationships.(17) In this view, healthcare simulation is a *social practice* where participants *interact* in a goal-oriented fashion with setting *attributes* to elicit a behavioral response (See Figure 1).(18) This simulation setting framework served to investigate the concept of healthcare simulation reliability. The application of grounded theory coding strategies allowed for iterative and constant comparison throughout the data collecting process. This strategic combination provided a flexible approach to draw from various perspectives in the literature. The Standards for Reporting Qualitative Research serves as the reporting guidelines. (19)

Search Strategy

With the assistance of a medical librarian, a search in PubMed targeted HPE simulation-users. Inclusion criteria were articles published between 2009-2019 regardless of article research design. Keywords related to simulation setting and design included simulation design, scenario design, scenario setting scenario development, simulation development, simulation environment, in-situ simulation, simulation milieu. Drawing from Norman (20) reliability terms agreement, reproducibility, association and correlation were included in the keyword search process. Exclusion criteria included grey

literature and articles related to gaming or virtual reality, systematic reviews, and non-English literature. Although virtual reality is a simulation modality, it was excluded on the basis that a virtual environment may engender a different sense of presence.(21) The medical librarian provided the search string used in PubMed for article results. To reveal initial article results, the search string can be copied and pasted into PubMed. (see Appendix I)

Article Selection Method

QSR International's NVivo data analysis software was used to support search processes, data collection efforts, and organization of data management. To initiate exploration of contextual data centered around reliability and its related terms, a stemmed-word query was used to reveal and quantify the most frequent synonyms associated with reliability from each article the abstracts resulting from the initial literature search. I set query specifications to generate words with a minimum of five characters to omit conjunction and determiner words (e.g., the, a, many, an, but, this). Reliability terms selected were added into a word search query that highlighted individual words throughout the articles for ease of detection. Highlighted words initiated in the first grounded theory cycle of data collection through open coding. Open coding "fractures or splits the data into individually coded segments" to use for future reference.(22) No new terms were identified during exploration.

Data collection included an iterative two-step process that sought to code discrete contextual data related to reliability and collect simulation setting metadata. To organize metadata and map each article's simulation setting, a spreadsheet was developed containing simulation setting classifications (e.g., modality, environment) and attributes

associated with simulation modality. Modality is a term used to refer to the type of simulation being used as part of the simulation activity such as simulated participants, human patient simulators, or live tissue or task trainers.(21). Open coding was operationalized followed by axial coding which supports the reassembly of fractured data with the aim to “link categories with sub-categories and ask how they are related”.(23) This approach uncovered themes and subthemes related to how healthcare simulation setting reliability is conceptualized and achieved.

The exploration of contextual data and interpretations of an article’s described simulation setting assisted in discerning the ontological difference among reliability when simulation is employed as an education assessment tool, the subject of clinical research, or as an environment of research.(11, 12, 24) The combination of discerning context and classifying setting attributes winnowed the data. Winnowing is a process of focusing on some of the data and disregarding other parts of it.(25) To access and analyze patterns between coded data, a matrix coding query was performed. Additionally, to check the consistency of coding across articles and the simulation setting demographic variables, I performed a crosstab query to check the spread of coding across coding and demographic data.

Techniques to Enhance Trustworthiness

Worked with a medical librarian enhanced search effort to maximize the identification of articles for possible inclusion. Furthermore, weekly meetings were held with two interdisciplinary researchers to discuss my findings and assumptions during the data collection and analysis processes. One researcher has expertise within healthcare simulation and has a specialization in simulation-based learning and the conduct of

systematic reviews. The second researcher has expertise in qualitative and mixed-methods research design. These meetings provided peer examination of the analysis as recommended by Saldana.(22) Accountability and an audit trail was captured using video recordings of each meeting.

Reflexivity

The first author conducted all literature reviews. Her own medical healthcare experience has taken place since 2008 until the present in simulation center operations and management.

RESULTS

Article Selection

PubMed produced 234 HPE simulation-user articles. Of these 234 articles, 107 did not include a description of a simulation setting framework and were excluded. The remaining 122 articles were screened and separated into two groups: studies that implemented a simulation setting for education achievement assessment and studies that implemented simulation to assess constructs other than education achievement (e.g., training, process development). Of these, 72 articles related to reliability in assessing education achievement were excluded, resulting in a corpus of 50 articles for the study related to reliability in simulation (See Figure 2).

Article Demographics

Article research locations included North America (34%), Europe (20%), United Kingdom (16%), Asia (10%), unspecified (8%), Australia (6%), and South America (6%). The majority of the articles were published between 2015 and 2018 (See Figure 3).

Sixty percent of the article did not specify the type of simulation environment.

Simulation modality included task trainers (34%) or simulated patients (32%), and 50% of articles assessed for simulator performance (See Table 1).

Word Frequency of Reliability Terms

Results from the word frequency search reveals eight terms and their synonyms related to reliability used across all articles from the PubMed search. Reliability terms include reliably (76%), standards (22%), correlations (19%), consistency (16%), agreement (8%), associations (8%), reproducible (2%), and replicate (2%) (See Figure 4). These terms were used to initiate text search query to highlights each word across all articles to operationalize the winnowing process.

Qualitative Findings

Open and axial coding reveals a theme and subthemes associated with healthcare simulation setting reliability. The central theme encompasses how setting attributes influence simulation setting reliability. From this view, reliability is not directly defined but described by the interactions between simulation design attributes (e.g., assessment tool items, timeframe, cases, SP performance training) and SP performance nuances (e.g., portrayal inaccuracies, content delivery, SP role assumed) (See Figure 5 and Table 2).

Drawing from these six articles, subthemes emerge relating to the different training approaches that aim to achieve simulation setting reliability. One of the training approaches entails training for SP performance to mitigate specific setting design attributes that introduce error to a simulation setting. The other training approach aims to mitigate errors introduced by SP performance nuances. This training approach aim to for SPs to perform in a standardized fashion. Essentially, simulation setting reliability is

dependent on the manner in which SPs are trained to achieve performance reliability. In the following section, lexical variations are discussed, attributes interfering with simulation setting and SP performance reliability are presented, training approaches are examined. A matrix was developed from open and axial coding to showcase the varying constructs related to simulation setting reliability (See Appendix I).

Terms tied to the concept of simulation reliability

A frequent term surrounding the concept of reliability is consistency, specifically in relation to SP performance (n = 3 articles).(26-28) MacLean (26) describes consistency as “SPs having a sound understanding of the purpose of the simulation and participant skills being targeted”. (26) In other words, SP performance reliability depends on how well SPs understand scenario objectives and the tools needed to elicit participant behaviors. To achieve setting reliability, SPs should consistently perform the designed activity requirements so that all participants face the same test situation.(29)

Authors use various terms geared towards training SPs to achieve standardization, accuracy performance and explore the implementation of training processes. (See Table 3). To strategically train for performance standardization, terms include calibration, agreement and concordance (n=2).(27, 29, 30) The term ‘concordance’ emerged from the winnowing process and not the word frequency query. Another term identified was accuracy (n = 2 articles). Accuracy, in contrast, is represented by evidence collected through observation of SP performance. Accuracy in performance is measured by the “proportion of essential clinical features presented correctly in each encounter”. (31) As a measurement criteria, accuracy validates that SPs consistently achieve performance reliability, thus “accuracy is another dimension of validity”. (32)

Reproducibility and replicability were two unique terms (n=1) that describe the accuracy of a particular type of performance. Although these terms may be considered synonyms, they have somewhat different meanings. Hill (33) defines reproducibility as the “consistent performance of each SP across student interviews” and replicability as the “consistent performance of a number of SPs across each scenario”. Ontologically, SP performance reliability pertains to how authors address achievement towards simulation setting reliability. Based on the ontology of attributes that influence simulation reliability, these lexical terms suggest the types of performance construct authors aim to refine to achieve simulation setting reliability.

Attributes interfering with reliability

Attributes influencing SP performance reliability were identified as the complexity of tasks SPs were asked to perform (i.e., performing a dual role as the patient and rater), rating tool checklist items used to assess participants, scenario timeframe, and fatigue (i.e., extended performance over time, performance of SPs to portray social-emotional skills). (26, 29, 30) Attributes that influence child SP performance include supporting SPs in the scenario (i.e. parent role), cognitive level of understanding case purpose, and fatigue.(27) Attributes influencing setting reliability include, inefficient SP training, content delivery, physical presentation, patient portrayal (i.e., tone, demeanor), and inconsistent adherence to performance training protocols.(26, 29, 33)

Training approaches

In addition to the emergence of terms and description, the operationalization of training approaches to achieve SP performance reliability were revealed. The follow axial coding model adapted from Morrow examines the relationship between constructs

associated with achieving SP performance reliability.(34) The model's pathway is ordered starting with the author's ontology of design attributes or SP performance nuances generating errors followed by phenomenological activities and training approaches thought to achieve SP performance reliability or processes to discover or test for achievement of SP performance reliability. In other words, this pathway to achieve SP performance reliability directly influence the manner in which simulation setting reliability is achieved. The following examination divides the training process between two subthemes, training for SP performance standardization and training to address specific attributes that generate setting error.

Training for performance standardization via a standardized training strategy

To mitigate errors generated by SP performance (i.e., body language, demeanor, case opening statement, Russell (29), implements a one-hour calibration session directly before setting implementation (See Figure 6). This calibration meeting is a strategic approach by which simulated patients performance is standardized. (29) A simulated patient "simply presents the symptoms and problems of an actual patient, where as a standardize patient is *trained* to present these symptom and problems in a consistent manner which does not vary from participant to participant or standardized patient to standardized patient".(29) Put differently, simulated patient and standardized patient are defined by training experience.

During the calibration meeting *agreement* must be reached by all SPs and quality assurance training. Agreement is made on expected performance skills in the delivery of the opening statement, the demeanor of the 'patient', and cueing (what information to give, how and when).(29) This meeting is "a short briefing for role players held at the

start of each day of summative examination to “remind SPs of expectations and also to follow up on any issues arising from the previous day(s).” (29) The rationale for the strategic training approach is that participants will lead the interactions based on the preliminary questions asked.

Action-Based research development process to convert children simulated patients into standardized patients

Using a three-phase action-based research process, the development to convert simulated child patients into standardized patients revealed that SP performance reliability was influenced by fatigue, SP parent role support, child’s cognitive level or understanding, performance embellishment over time and age appearance where “child actors need to look the correct age” (See Figure 7).(27) The plan was to implement child SP in future participant summative assessments. Trainers assessed the calibration meeting. However, the study does not discuss simulation setting implementation making the training unclear if performance accuracy was achieved.

Training for performance standardization consistency and accuracy via a standardized training strategy

Simulation setting reliability is influenced by setting errors stemming from simulation design attributes and SP performance (See Figure 8). Standardized patient performance errors may be elicited through physical presentation and facial expressions or performance portrayal.(35). To mitigate these errors, Baig (30) trains simulated patients to achieve performance accuracy by standardizing performance portrayal. Portrayal is defined as the “concordance between appearance, symptom representation, SP preparation and appropriateness of the case”.(30) Performance accuracy is achieved by “concordance between appearance, symptom representation, SP preparation and

appropriateness of the case”.(30) This description correlates to the definition of performance consistency where SPs should have a sound understanding of the purpose and the case and the skill address. Baig’s (30) maintains that concordance is how SP achieve a sound understanding of the purpose of the case and the skills being assessed. In other words, concordance is how SP performance consistency is achieved.(29, 30)

In contrast to Russell’s (29) strategy where a calibration meeting is performed one-hour prior to simulation setting implementation, this study’s SP performance consistency of emotions, facial expressions, and body language were validated via activity observations informed by guidelines developed for each summative assessment case.

Training for performance accuracy to mitigate inaccuracies of patient portrayal

Standardized patients may construct performance interference in more than one way by how they depict patient history, physical presentation and content portrayal (See Figure 9).(35) In this study, SPs assumed a parent or granparnt role of a speech delayed child.

Commonly, standardized patient performance training is sought to elicit participant performance behaviors. In contrast, the purpose of this study train and assess SPs for performance accuracy replicability , where SPs perform in a “consistent manner over a number of SP across each scenario”, or accuracy reproducibility, where SP present the “clinical scenario in the same way to a number of different participants.” (33). In this study, accuracy is defined as the combination of SPs presenting *all* designated scenario features to students in the same way to a number of different students.”(33) Additionally,

for trainer to assess performance accuracy, heavy socio-emotional cases were purposefully chosen..

Standardized patients were trained to depict a parent or grandparent of a speech delayed child via a standardized training approach. The SP's purpose was for participants to practice communication skills to acquire patient history. To achieve performance reproducibility for portrayal across cases, SPs were trained by the same trainer using station specific guidelines for the same case. Alternatively, other SPs recruited were trained to achieve performance replicability by using different SP across each case. However, it is unclear if the author describes *each* case as different SPs given training for the identical cases or different cases. To achieve standardized patient portrayal, SPs were trained to withhold information if when not specifically asked by participants. A concern may be that withholding information interferes with the amount of information participants need to make sense of situations.(31) Additionally SP were trained to memorize key features of a speech delayed child, "the gold standard for SP portrayal." (31) Post setting implementation revealed that fatigue from the heavy socio-emotional workload in this study interfered with both setting and SP performance reliability

Developing standardized patients as raters for performance consistency and accuracy

In this study, simulated patient were trained to assume a dual role as a standardized patient and rater. As a rater, SPs were expected to complete performance-based evaluation instrument to assess participant communication skills. Implementation of patient and rater roles adds to the layer of attribute that influence SP performance and simulation setting reliability (See Figure 10).

To mitigate attributes interfering with SP performance as a patient and setting reliability, simulated patients were trained for eight and a half hours using a four-phase training program incorporating experiential learning activities. Different simulated patients were trained on the same case to achieve performance replicability. To achieve performance to evaluate participant skills, simulated patients were trained by observing communication role-modelling behaviors using video recordings. The combined training process aimed to achieve dual-role performance reliability. Trainers evaluated SP performance and their ability to accurately and consistently complete the instrument rating over time. To evaluate SP performance using the performance-based evaluation instrument Bu using an inter-rater reliability process, SP's use of a performance-based evaluation over time validated that SPs understood the purpose for the case and the participant skills being assessed. In other words, SPs were able to consistently complete the performance-based evaluation instrument.

Developing unannounced standardized patient for consistency and accuracy in performance assessment

In this study, Shirazi aims to develop standardized patient to achieve performance consistency and accuracy in the role as unannounced SP assessor (See Figure 11). An unannounced SPs play their roles in situations where practitioners are unaware that the person is not a real patient.(21) SP's were to portray a patient with a depressive disorder combined with other illnesses (e.g., bipolar disorder, acute risk of suicide, pathologic grief) and assess the general practitioner's management of the depressive disorder performed.

Different SP were trained for a total of 13 hours on how to portray a patient with a depressive disorder and how to assess general practitioner performance. To assess for

performance consistency and accuracy, authors performed a test-retest one week later. However, simulation setting implementation in a live environment made it difficult to reassess SP for performance accuracy

DISCUSSION

The findings suggest that within healthcare simulation, reliability is often conceptualized by errors introduced by SP performance behaviors or specific simulation attributes. Articles place a heavy emphasis on directing SPs to perform to in a specific manner. The idea is for SP performance to mitigate errors introduced by setting attributes. From a social constructivist perspective, SPs are employed to facilitate social and behavioral interactions by providing verbal or nonverbal feedback. The idea of providing cues is to help navigate or clarify any gaps in the fidelity of the setting. (36) However, interactions mean that the effect of one factor depends on the level of the other factor creating performance inconsistencies that influences simulation setting reliability.

Implications for Reliability in Healthcare

The findings suggest that when terminology used to describe reliability in healthcare simulations are combined as they are in this review, simulations are subject to variability in implementation due to their social aspects. This finding is consistent with a study by Dieckmann, revealing that the experience of the simulation is influenced by the social characters of simulation, and the varying experiences of participants depend on how they interpret the same stimuli. (37-40) When designing for setting performance reliability, training approaches and strategies are not restricted to any type of design, therefore participant goal oriented activities are vulnerable to interference stemming from SP design.

Implications for differences in use of the terminology in HPE and healthcare simulation

In HPE and healthcare simulation, reliability is tailored to the construct researchers want to assess. Similarly, both domains seek to achieve reliability to provide fair and trustworthy conditions to observe education achievement. From a social constructivist perspective, reliability in healthcare simulation is heavily dependent on SP performance reliability. This dependency is due to the likelihood that SPs introduce performance errors that influence participant behavior. To achieve simulation setting reliability, SPs are trained to refine their performance using various training approaches. These approaches are tailored to address SP performance nuances (i.e., performance over time, performance standardization, performance accuracy). These approaches aim to reduce the amount of interference that influence participant behavior from achieving activity-goals.

Healthcare simulation setting is underpinned by SPs having a sound understanding of the purpose of the and the participant skills being assessed. The approach to setting reliability is refining SPs to achieve performance reliability, with the aim to reduce the amount of errors introduced into the simulation setting. Unlike the psychometric properties used to measure participant performance in HPE, SP performance reliability is measured for accuracy via observations where activities are the unit of measurement.

Limitations

My review covers 2009–2019 and includes 50 articles from the 234 initially identified. Despite my efforts to be as inclusive as possible, I may have missed some articles. Additionally, my search used specific terminology and focused on specific

databases, thus, this may have biased my included article sample. Searching a broader date range, database or terminology may have yielded additional articles.

Conclusion

This is the first definitional review of healthcare simulation reliability. The goal of this study is to open debate about how attributes of a simulation setting challenge the traditional concepts of reliability in health professions education. Findings were specifically based on the consistency of SP performance reliability, underpinned by the premise that SPs have a sound understanding of the purpose of the simulation and student skills being assessed. Thematically, authors believe that simulation setting attributes and SP performance are the leading cause interfering with achieving simulation setting reliability, however training approaches emphasize how refining SP performance can mitigate errors.

While this is not an emergent topic, clarifying these concepts is critical to effectively advancing healthcare simulation education and assessment. The focus on the psychometric tools used in HPE do not consider the social aspect of simulation. Social practices add an additional layer of complexity that interfere with a setting's reliability that is supposed to provide participants a fair and trustworthy conditions to achieve performance competency set by medical educators. Ultimately, simulation setting error subjects participants to unequal experiences, thus measurement error can occur in determining their progression in medical education or licensure attainment.

Research Recommendations

There should be further study to understand how and how often to measure accuracy of SP performance. More research should be done on the length and rigor of

training requirements for SPs. Additionally, more studies of the current nature should be done to verify the replicability of the information acquired.

Acknowledgements

I would like to acknowledge my research advisors for their guidance during this learning process.

Prior Presentations

Not applicable.

Funding

Not applicable

Ethics Approval and Consent to Participate Consent for Publication

Not applicable.

Competing Interests

None

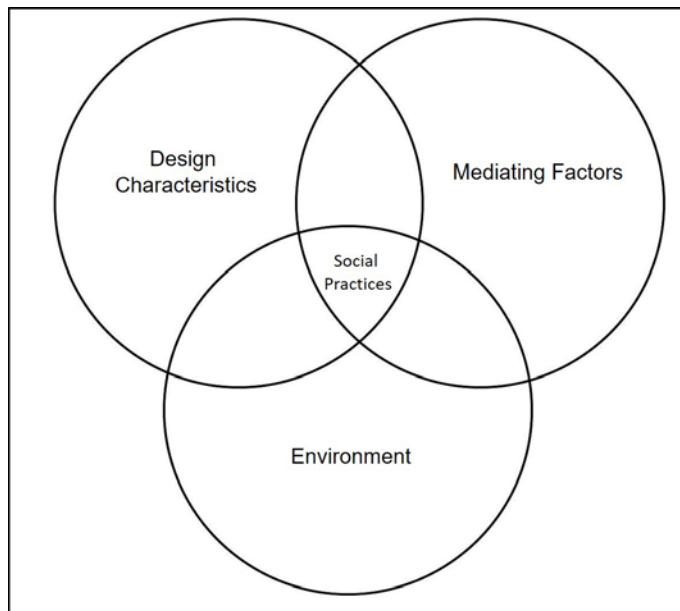


Figure 1. Healthcare Simulation Setting Framework

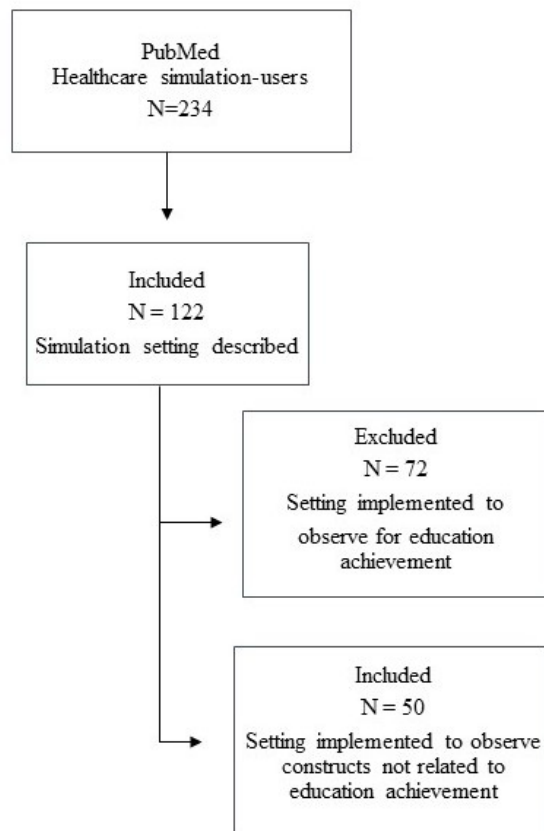


Figure 2. Article Selection Flow Chart

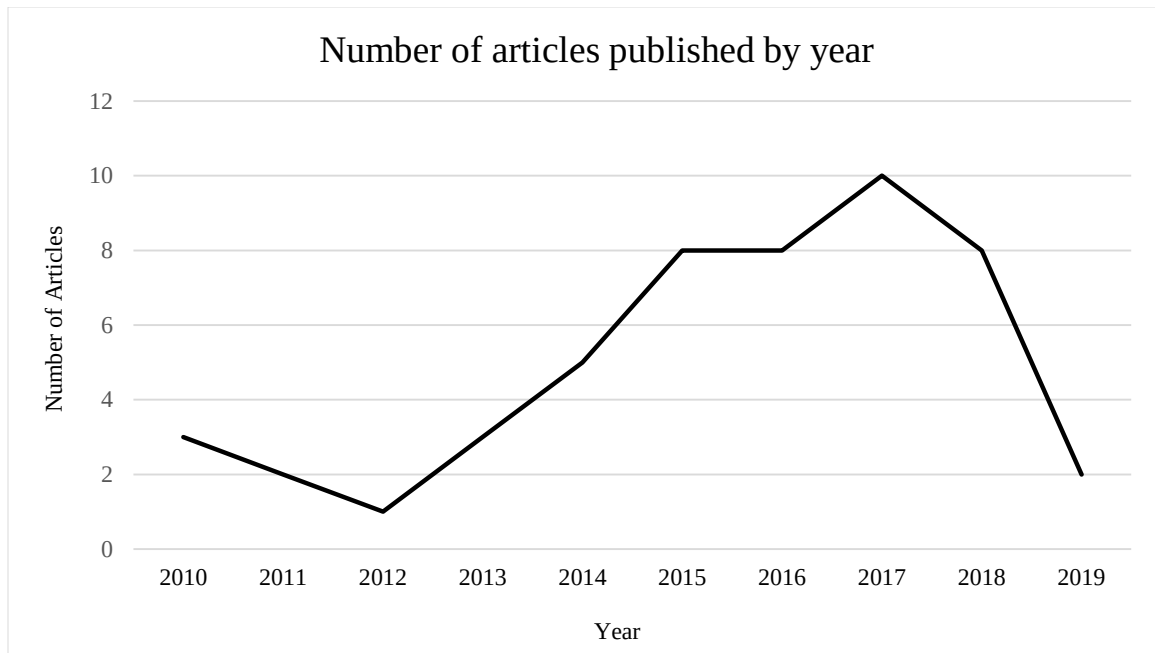


Figure 3. Published healthcare article that describe a healthcare simulation setting framework

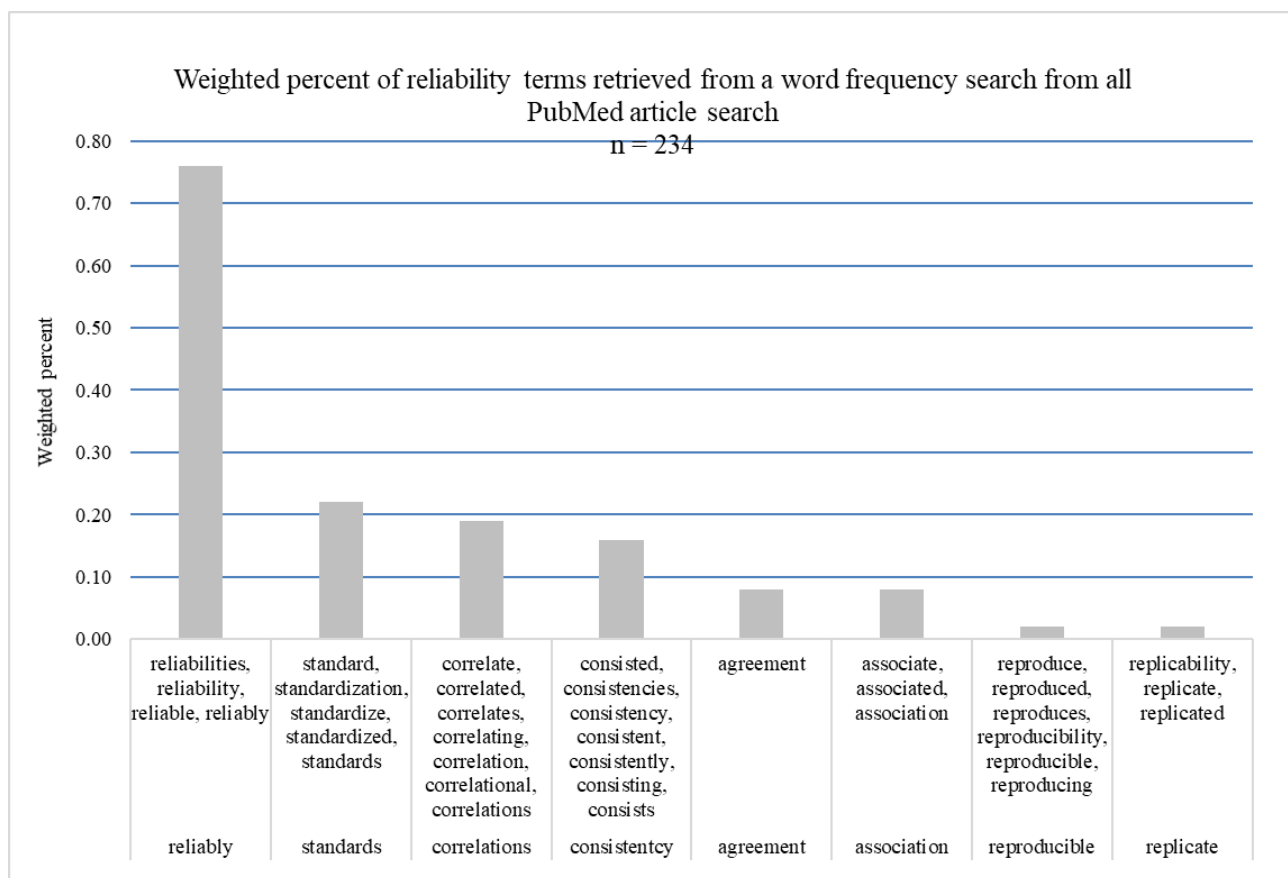


Figure 4. Reliability terms retrieved from a word frequency search within all PubMed articles searched

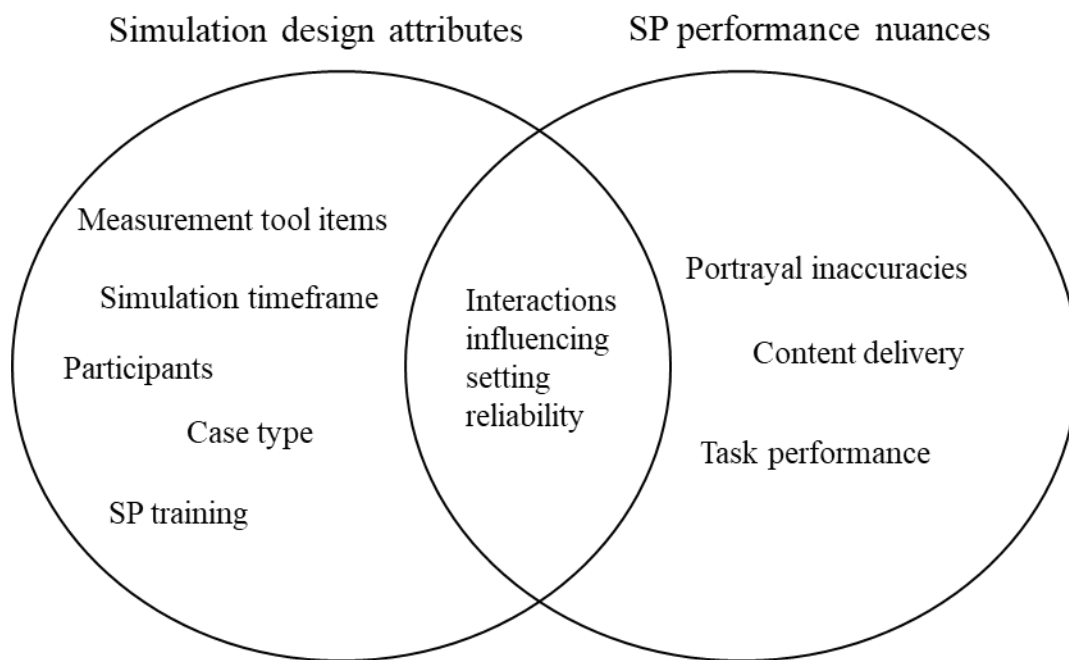


Figure 5. Interactions influencing simulation setting reliability

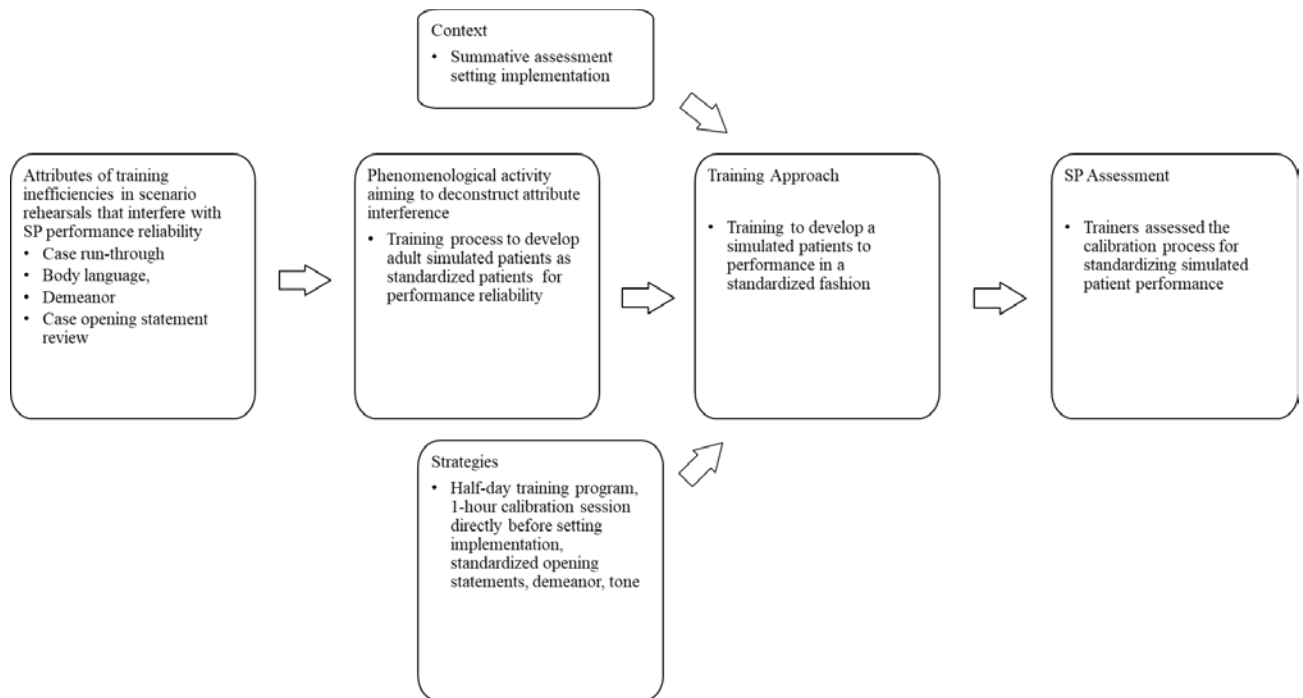


Figure 6. Training for performance standardization via a standardized training strategy

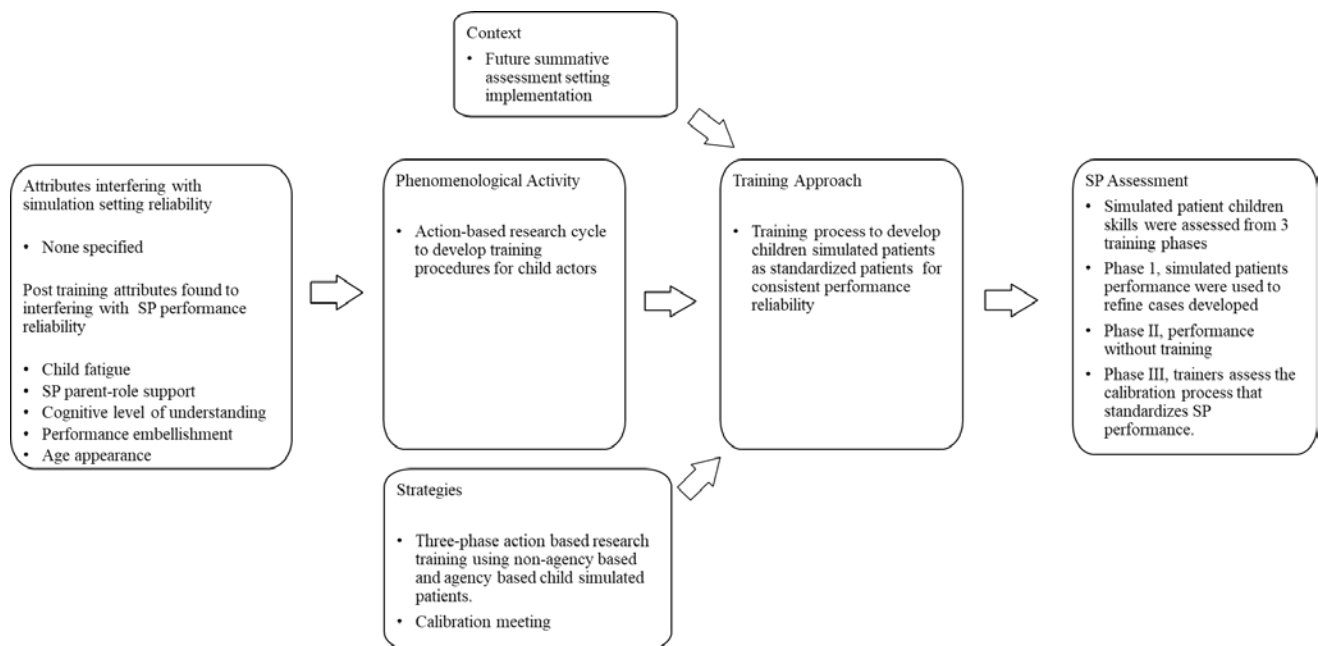


Figure 7. Action-Based research development process to convert children simulated patients into standardized patients

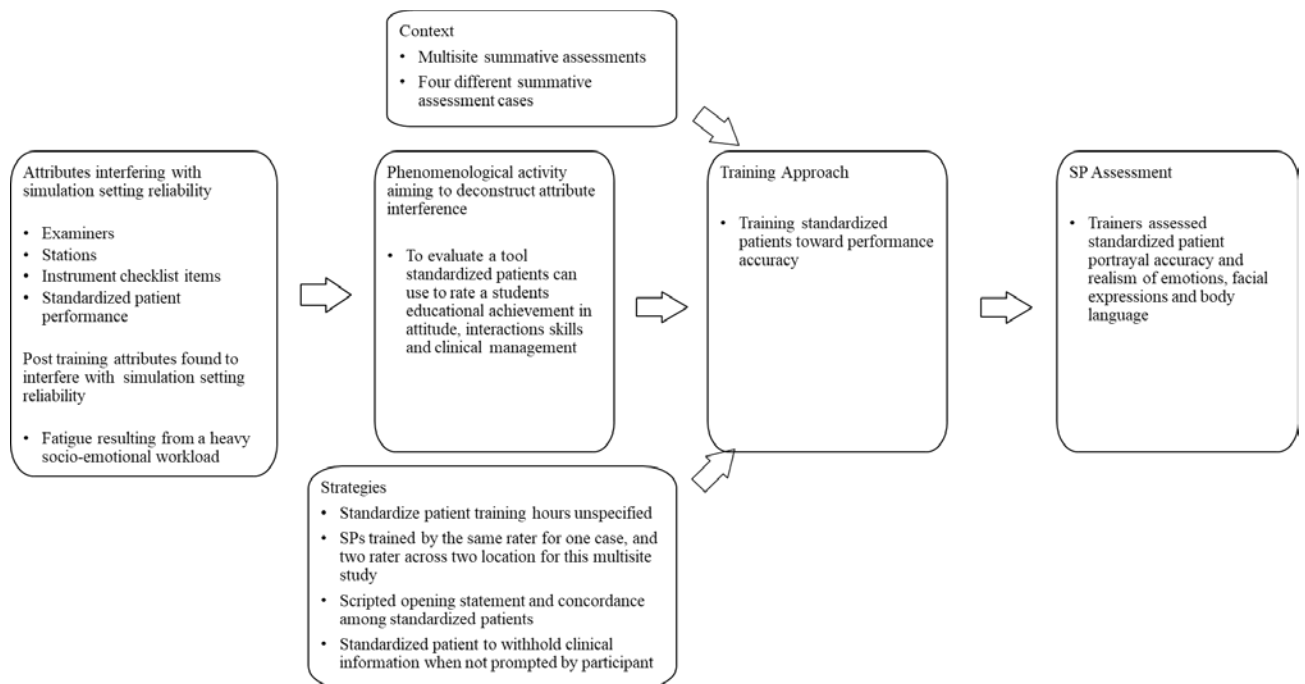


Figure 8. Training process to standardize patient performance accuracy (Baig, 2014)

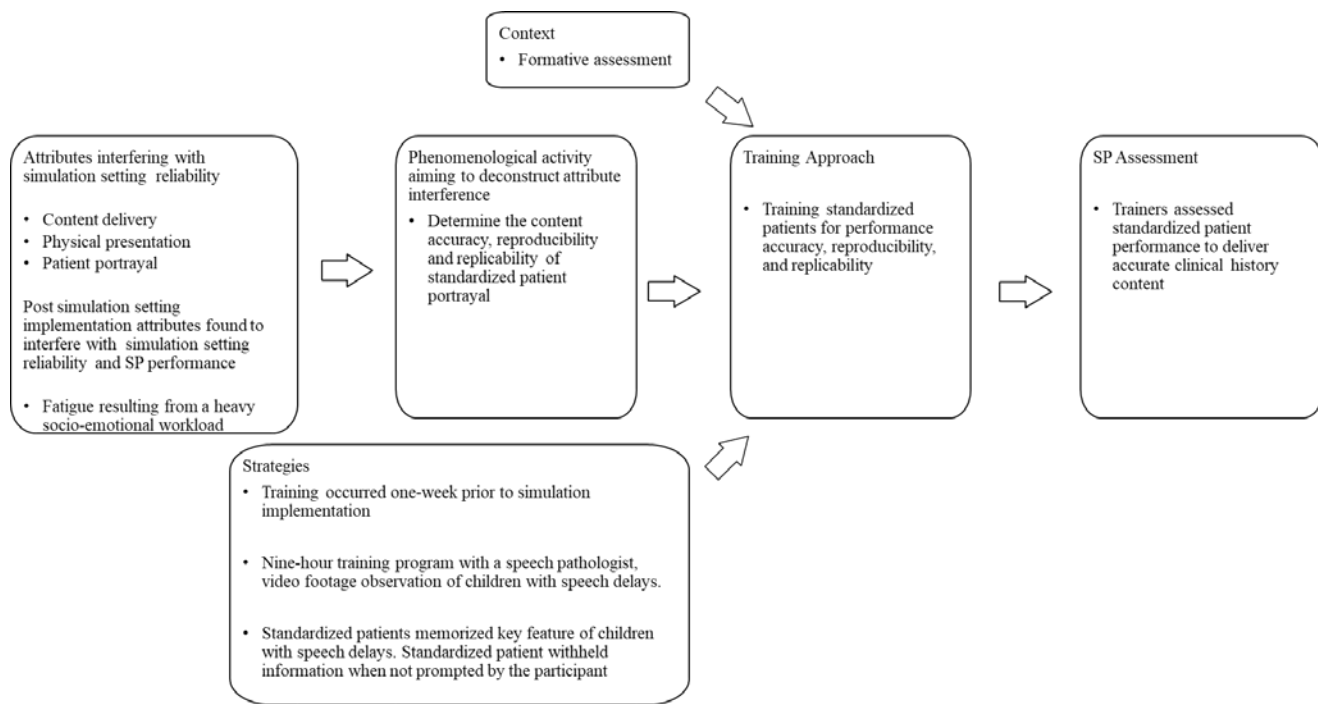


Figure 9. Training for performance accuracy to mitigate inaccuracies of patient portrayal

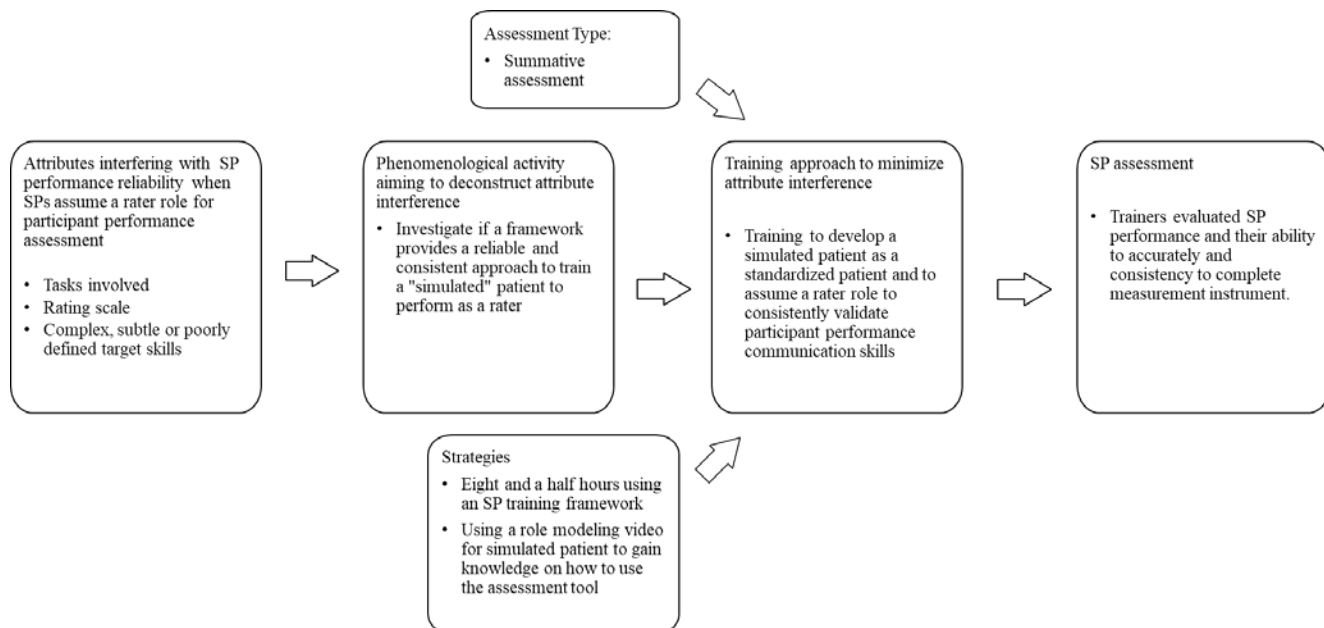


Figure 10. Developing standardized patients as raters for performance consistency and accuracy

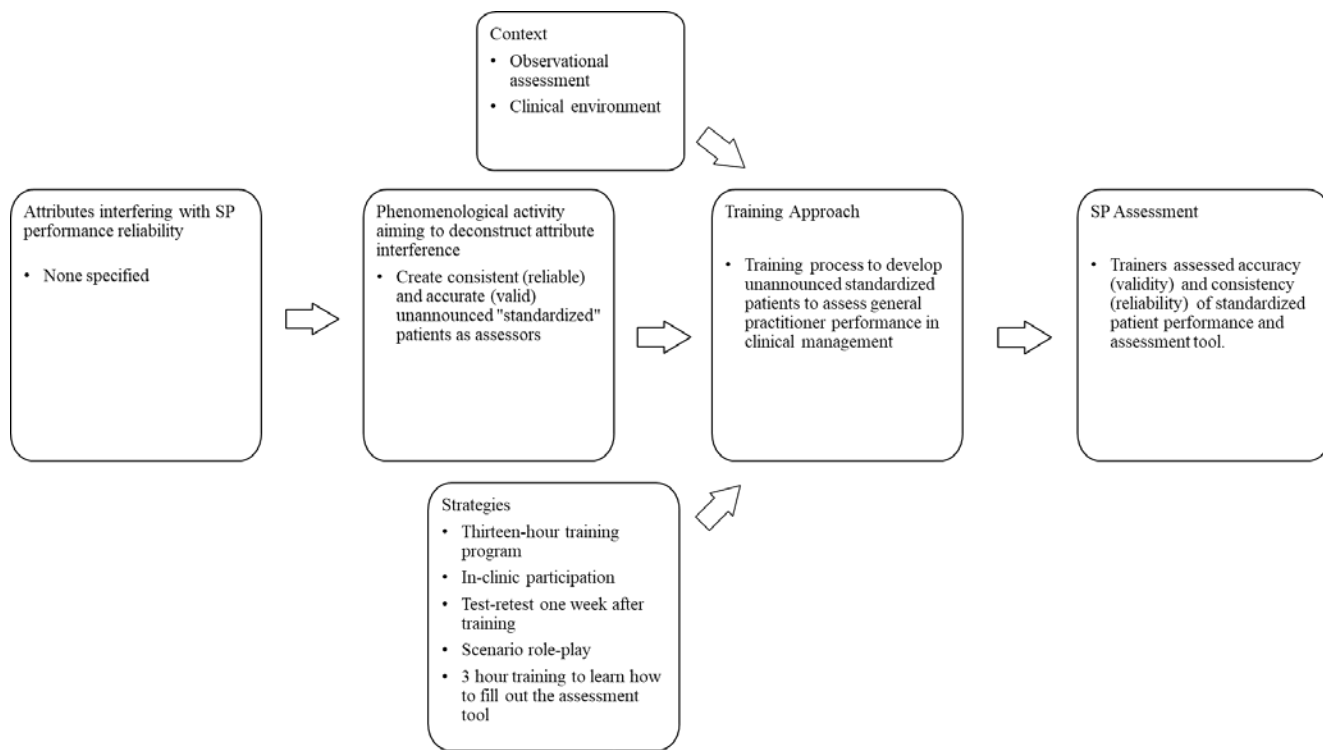


Figure 11. Training process to development unannounced standardized patients to rate practitioner performance

Table 1. Demographic summary of healthcare simulation setting attributes

Healthcare simulation setting attributes	Percentage of studies
Environment Type	
Unspecified	60%
In-house lab	16%
In-situ	12%
Multisite	6%
Non-hospital	4%
Portable	2%
Modality Type	
Task Trainer	34%
Simulated Patient	32%
Human Patient Simulator	12%
Animal Tissue	8%
Hybrid	6%
Augmented Reality	2%
Cadaver	2%
Inflatable Environmental Shell	2%
Unspecified	2%
Construct Assessed	
Simulator Performance (i.e., task trainers, SPs)	50%
Subjective Interpretations	20%
Setting Dynamics	8%
Tool Performance	6%
Scenario Design	4%
Facilitator Performance	2%
Human Factors	2%
Process effectiveness	2%
Task Development	2%
Technical Skills	2%

Table 2. Healthcare simulation setting reliability conceptualized by author

Russell (2011)	Reliability as rehearsal inefficiency to standardize performance (29)
Russell (2015)	Setting reliability as consistency of "simulated" patient performance between candidates (27)
Baig (2014)	Reliability as setting variance due the dynamic interactions of setting attributes (i.e., examiners, stations, items, standardized patients) (30)
Hill (2013)	Setting Reliability as inaccuracies in "standardized" patient portrayal (33)
MacLean (2018)	Reliability relates to the consistency of the SP's rating performance using a measurement tool overtime (26)
Shirazi (2011)	Reliability as consistency of "standardized" patient performance and tool assessment

Table 3. Reliability terms used by authors

Authors	Reproducibility	Replicability	Consistency	Accuracy/Validity	Agreement	Concordance
MacLean (2018)			X	X		
Russell (2015)			X		X	
Baig (2014)				X		X
Hill (2013)	X	X		X		
Russell (2011)			X		X	
Shirazi (2011)			X	X		

REFERENCES

1. Downing SM, Yudkowsky R. Assessment in Health Professions Education: Routledge; 2009 2009/5/7. 344 p.
2. Lewis KL, Bohnert CA, Gammon WL, Hölzer H, Lyman L, Smith C, et al. The association of standardized patient educators (ASPE) standards of best practice (SOBP). *Advances in Simulation*. 2017;2(1):10.
3. Mulugeta L, Drach A, Erdemir A, Hunt CA, Horner M, Ku JP, et al. Credibility, replicability, and reproducibility in simulation for biomedicine and clinical applications in neuroscience. *Frontiers in neuroinformatics*. 2018;12:18.
4. Downing SM. Reliability: on the reproducibility of assessment data. *Medical education*. 2004;38(9):1006-12.
5. Downing SM, Yudkowsky R. Assessment in health professions education: Routledge; 2009 2009.
6. Brannick MT, Erol - Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. *Medical education*. 2011;45(12):1181-9.
7. Norman GR, Vleuten C, Newble D. *International Handbook of Research in Medical Education*: Springer Science & Business Media; 2002.
8. Roberts J, Norman G. Reliability and learning from the objective structured clinical examination. *Medical education*. 1990;24(3):219-23.
9. Jeffries PR, Rodgers B, Adamson K. NLN Jeffries simulation theory: Brief narrative description. *Nursing Education Perspectives*. 2015;36(5):292-3.
10. Lopreiato JO. *Healthcare simulation dictionary*: Agency for Healthcare Research and Quality; 2016.
11. McGaghie WC, Issenberg, S. Simulation in Assessment. *Assessment in Health Professions Education*. New York, NY: Routledge; 2009. p. 245-68.
12. Munroe B, Buckley T, Curtis K, Morris R. Designing and implementing full immersion simulation as a research tool. *Australasian Emergency Nursing Journal*. 2016;19(2):90-105.
13. Polit DF, Beck CT. *Essentials of nursing research: Appraising evidence for nursing practice*: Lippincott Williams & Wilkins; 2009.

14. Battista A. Activity theory and analyzing learning in simulations. *Simulation & Gaming*. 2015;46(2):187-96.
15. Vygotsky LS. *Mind in society: The development of higher psychological processes*: Harvard university press; 1980.
16. Von Bertalanffy L. *General system theory*. New York. 1968;41973(1968):40.
17. Watling CJ, Lingard L. Grounded theory in medical education research: AMEE Guide No. 70. *Medical teacher*. 2012;34(10):850-61.
18. Dieckmann P, Gaba D, Rall M. Deepening the theoretical foundations of patient simulation as social practice. *Simul Healthc*. 2007;2(3):183-93.
19. O'Brien BC, Harris IB, Beckman TJ, Reed DA, Cook DA. Standards for reporting qualitative research: a synthesis of recommendations. *Academic Medicine*. 2014;89(9):1245-51.
20. Norman G. When I say... reliability. *Medical education*. 2014;48(10):946.
21. 2016. *Healthcare Simulation Dictionary*.
22. Saldaña J. *The coding manual for qualitative researchers*: Sage; 2015.
23. Charmaz K. *Constructing grounded theory*: sage; 2014.
24. Cheng A, Auerbach M, Hunt EA, Chang TP, Pusic M, Nadkarni V, et al. Designing and conducting simulation-based research. *Pediatrics*. 2014;133(6):1091-101.
25. Creswell JW, Creswell JD. *Research design: Qualitative, quantitative, and mixed methods approaches*: Sage publications; 2017.
26. MacLean S, Geddes F, Kelly M, Della P. Simulated patient training: Using inter-rater reliability to evaluate simulated patient consistency in nursing education. *Nurse Educ Today*. 2018;62:85-90.
27. Russell D, Simpson R, Eden M, Foreman P. The use of child actors to simulate paediatric patients in the MRCGP Clinical Skills Assessment (CSA): developing an effective model. *Educ Prim Care*. 2015;26(4):242-7.
28. Shirazi M, Sadeghi M, Emami A, Kashani AS, Parikh S, Alaeddini F, et al. Training and validation of standardized patients for unannounced assessment of physicians' management of depression. *Academic Psychiatry*. 2011;35(6):382-7.
29. Russell D, Simpson R, Rendel S. Standardisation of role players for the Clinical Skills Assessment of the MRCGP. *Educ Prim Care*. 2011;22(3):166-70.

30. Baig LA, Beran TN, Vallevand A, Baig ZA, Monroy-Cuadros M. Accuracy of portrayal by standardized patients: results from four OSCE stations conducted for high stakes examinations. *BMC medical education*. 2014;14(1):97.
31. Tamblyn RM, Klass D, Schanbl G, Kopelow M. Factors associated with the accuracy of standardized patient presentation. *Academic Medicine*. 1990;65(9):S55-6.
32. Ladyshevsky R, Baker R, Jones M, Nelson L. Reliability and validity of an extended simulated patient case: a tool for evaluation and research in physiotherapy. *Physiotherapy Theory and Practice*. 2000;16(1):15-25.
33. Hill AE, Davidson BJ, Theodoros DG. The performance of standardized patients in portraying clinical scenarios in speech-language therapy. *Int J Lang Commun Disord*. 2013;48(6):613-24.
34. Morrow SL, Smith ML. Constructions of survival and coping by women who have survived childhood sexual abuse. *Journal of counseling psychology*. 1995;42(1):24.
35. Hill AE, Davidson BJ, Theodoros DG. The performance of standardized patients in portraying clinical scenarios in speech–language therapy. *International journal of language & communication disorders*. 2013;48(6):613-24.
36. Paige JB, Morin KH. Simulation fidelity and cueing: A systematic review of the literature. *Clinical Simulation in Nursing*. 2013;9(11):e481-e9.
37. Dieckmann P, Gaba D, Rall M. Deepening the theoretical foundations of patient simulation as social practice. *Simulation in Healthcare*. 2007;2(3):183-93.
38. Johnson E. *Situating simulators: The integration of simulations in medical practice*: Arkiv förlag; 2004.
39. Rystedt H, Lindwall O. The interactive construction of learning foci in simulation-based learning environments: a case study of an anaesthesia course. *PsychNology Journal*. 2004;2(2):168-88.
40. Dieckmann P. *Using simulations for education, training and research*: Pabst Science Publ.; 2009.

CHAPTER 3: DISCUSSION

Findings from this definitional review reveal that the concept of reliability in a healthcare simulation setting is conceptually similarly to reliability in health professions education (HPE) where the goal is to cancel out any interference. In HPE, observers aim to identify the primary construct being assessed by developing psychometric properties that generate a quantifiable measure. However, healthcare simulation is used as a tool in HPE to assess for education achievement, but the social dynamics add another layer of errors not considered in assessment. The social nature of simulation subjects participants to random error, thus not providing them fair and trustworthy conditions to demonstrate the competency criteria set by faculty that determine whether or not they can progress in their education or licensure. Normally social factors are not considered when evaluating participants for education achievement in HPE. However, more recent research seeks to achieve reliability in a healthcare simulation setting by addressing the social properties of standardized or simulated patient (SP) performance.

Nuances in SP performance authors attend to (i.e., performance over time, interaction between participants, inaccuracies in standardized patient portrayal) offer clues as to how authors define the primary challenges to achieve reliability. In other words, reliability is tailored to the construct researchers want to assess. To achieve simulation setting reliability, the goal is to refine the signals provided to participants by SPs during interactions to reduce interference that influences participants' behavior and potentially prevents them from achieving activity goals. Behaviorism posits that learning occurs when an external stimulus causes a response behavior; within this view, training

aims to refine SP performance (the external stimulus) to achieve simulation setting reliability.

From a social constructivist perspective, all authors perceived the social nature of the simulation setting and understood that errors stem from the social interactions between the SP and the participant. Researchers also acknowledge attributes that influence SP performance. These attributes vary depending on the design of the setting. In some cases, SPs are also trained to rate participant performance, an additional element of the social context. Additional attributes that influence SP performance reliability include simulation encounter time frames, and other SPs involved in the setting.(1, 2) Baig's study reveals SP fatigue is also possible and can be caused by the heavy socio-emotional workload (i.e., facial expression, body language and emotions).(1) Other attributes interfering with simulation setting reliability include the SPs' physical presentation and content delivery.(3) Authors aim to mitigate interference by training for SP performance consistency.

In healthcare simulation, SPs become the intermediary for observers to assess participant performance achievement. This concept of reliability in healthcare simulation differs from the concept of reliability in HPE whereby designed psychometric properties assist observers with determining participant performance achievement. In HPE, reproduced scores of an individual across numerous repeated measures will cancel out random error (4-8). This process results in an individual's "true score." When an individual's true score or collective true scores across groups are in agreement, the connections can be translated as a quantifiable measure to estimate reliability. Conversely, in healthcare simulation, the idea is to refine SP performance to make it

more reliable. Achieving SP performance reliability diminishes the amount of error introduced into a simulation setting, thus providing fair and trustworthy conditions for participants to achieve performance competency.

This study uncovered two training design approaches and strategies that aim to standardize patient performance. Russell aims to achieve SP performance standardization through a *calibration* process. (9) This process is a unique training strategy to standardize patients who portray the *same* case, thus Russell conceptualizes reliability as a process whereby simulated participants perform consistently in a standardized fashion.(9) To achieve consistency, delivery of the opening statement, the demeanor of the ‘patient’ and cueing (what information to give, how and when) must be agreed upon by all parties involved in the meeting and include a quality assurance observer. The quality assurance observer is the final test of reliability.(9) The purpose of the scripted opening statement aims to present the participant with the opportunity to take the lead in the interaction. Cueing can be viewed as a form of feedback where SPs are trained to cue participants to help navigate or clarify any gaps in the fidelity of the setting. (10) In Russell’s study, the calibration process is the determining factor for SP performance standardization. (2)

The second design strategy to achieve standardization is training for accuracy. Baig trains for performance accuracy and standardizes simulated patient to achieve “concordance between the SP’s appearance, symptom representation, adequacy of SP preparation and appropriateness of the case.”(1, 11) Baig’s standardization properties resonate with achieving performance reliability and consistency where SPs

have a sound understanding of the purpose of the simulation and the skills being assessed.(1, 12)

Although Russell (9) emphasizes SP standardization as the means to achieve reliability, he does not address achieving SP performance accuracy, even though SP performance reliability pertains to how authors address reducing errors that influence participant behaviors. As a measurement criterion, accuracy dictates that SPs consistently achieve performance reliability, thus “accuracy is another dimension of validity”.(13) Accuracy is represented by evidence collected through *observation* of SP performance. Performance is measured by the “proportion of essential clinical features presented correctly in each encounter” (11). In other words, observed performance operationalizes accuracy where activities are the unit of measurement to ensure performance consistency (14). To ensure SPs maintain a designed level of accuracy, it is recommended to observe them in multiple consecutive encounters (15).

In addition, to achieving SP performance reliability through accuracy or standardization, SP performance may be assessed for replicability or reproducibility. Reproducibility and replicability are two unique terms that describe the *accuracy* of a particular type of performance. Although these terms are often considered synonyms, they have somewhat different meanings. Reproducibility is defined as “consistent performance of each SP across student interviews” and replicability is defined as “consistent performance of a *number* of SPs across each scenario”.(16) In HPE, Downing (5) refers to reproducibility of data or scores on the assessment. He states, “high score reliability indicated that if the test were to be repeated over time, participants would

receive about the same scores on retesting as they received the first time”.(17) Although mentioned in the article in this study, this measurement criteria is hardly discussed.

MILITARY RELEVANCE

The relevance of this study extends beyond civilian healthcare simulation settings. Healthcare professionals preparing for deployment face critical skills testing that must be reliable in order to ensure their safety and the safety of the patients they treat on tour. The skills they are tested on can be separated into two categories: medical and non-medical.

Regardless of their chosen field of medicine, deployed healthcare professionals must demonstrate mastery of a variety of medical skills that may be relevant in a deployment situation. These skills primarily consist of competencies in trauma medicine including assessing traumatic brain injury patients, responding to emergency procedures, and communicating with foreign nationals in health disparity cases. This study also extends to other response readiness needs to respond to threat from chemical, biological, radiological, nuclear and explosives. All of these skills require social interactions and thus trainees would benefit from the use of simulations to both learn and test their abilities. The information provided in this study can be used to develop accuracy-focused instructional programs for simulation performers and reliable simulation settings.

REFERENCES

1. Baig LA, Beran TN, Vallevand A, Baig ZA, Monroy-Cuadros M. Accuracy of portrayal by standardized patients: results from four OSCE stations conducted for high stakes examinations. *BMC medical education*. 2014;14(1):97.
2. Russell D, Simpson R, Eden M, Foreman P. The use of child actors to simulate paediatric patients in the MRCGP Clinical Skills Assessment (CSA): developing an effective model. *Educ Prim Care*. 2015;26(4):242-7.
3. Hill AE, Davidson BJ, Theodoros DG. The performance of standardized patients in portraying clinical scenarios in speech-language therapy. *Int J Lang Commun Disord*. 2013;48(6):613-24.
4. Downing SM. Reliability: on the reproducibility of assessment data. *Medical education*. 2004;38(9):1006-12.
5. Downing SM, Yudkowsky R. *Assessment in health professions education*: Routledge; 2009 2009.
6. Brannick MT, Erol-Korkmaz HT, Prewett M. A systematic review of the reliability of objective structured clinical examination scores. *Medical education*. 2011;45(12):1181-9.
7. Norman GR, Vleuten C, Newble D. *International Handbook of Research in Medical Education*: Springer Science & Business Media; 2002.
8. Roberts J, Norman G. Reliability and learning from the objective structured clinical examination. *Medical education*. 1990;24(3):219-23.
9. Russell D, Simpson R, Rendel S. Standardisation of role players for the Clinical Skills Assessment of the MRCGP. *Educ Prim Care*. 2011;22(3):166-70.
10. Paige JB, Morin KH. Simulation fidelity and cueing: A systematic review of the literature. *Clinical Simulation in Nursing*. 2013;9(11):e481-e9.
11. Tamblyn RM, Klass D, Schanbl G, Kopelow M. Factors associated with the accuracy of standardized patient presentation. *Academic Medicine*. 1990;65(9):S55-6.
12. MacLean S, Geddes F, Kelly M, Della P. Simulated patient training: Using inter-rater reliability to evaluate simulated patient consistency in nursing education. *Nurse Educ Today*. 2018;62:85-90.
13. Ladyshevsky R, Baker R, Jones M, Nelson L. Reliability and validity of an extended simulated patient case: a tool for evaluation and research in physiotherapy. *Physiotherapy Theory and Practice*. 2000;16(1):15-25.

14. Battista A. Activity theory and analyzing learning in simulations. *Simulation & Gaming*. 2015;46(2):187-96.
15. Adamo G. Simulated and standardized patients in OSCEs: achievements and challenges 1992-2003. *Medical teacher*. 2003;25(3):262-70.
16. Hill AE, Davidson BJ, Theodoros DG. The performance of standardized patients in portraying clinical scenarios in speech–language therapy. *International journal of language & communication disorders*. 2013;48(6):613-24.
17. Downing SM, Yudkowsky R. *Assessment in Health Professions Education*: Routledge; 2009 2009/5/7. 344 p.

APPENDIX I: PUBMED ARTICLE SEARCH STRATEGY

To reveal the initial results of healthcare simulation-users, copy and paste the full search string into PubMed.

((("Simulation Training"[Majr] OR "simulation design"[tiab] OR "simulation setting"[tiab] OR "simulation development"[tiab]) AND (Reliable[tiab] OR reliability[tiab]) AND ("2009/01/01"[PDat] : "3000/12/31"[PDat]) AND English[lang]) NOT (Comment[sb] OR Case Reports[ptyp] OR Editorial[ptyp] OR Letter[ptyp] OR "Weightlessness Simulation"[Mesh] OR "Space Simulation"[Mesh] OR "Molecular Dynamics Simulation"[Mesh] OR "Molecular Docking Simulation"[Mesh] OR "Computer Simulation"[Mesh] OR "Malingering"[Mesh]))

APPENDIX II: CONSTRUCTS OF SIMULATION SETTING RELIABILITY

	MacLean (2018)⁽¹⁾	Russell (2015)⁽²⁾
Research purpose	Investigate if a framework provides a reliable and consistent approach to train "simulated" patient raters	Describe the action research-based cycle through which we developed our training procedures for child actors for the Medical Royal Colleges of General Practitioner Clinical Skills Exam
Reliability conceptualized	Reliability relates to the consistency of the SP's rating performance using a measurement tool overtime ⁽¹⁾	"An important component of the reliability of the OSCE format is the consistency of the simulated patient's performance between candidates. ⁽²⁾ Role-player consistency between candidates is a crucial component of the reliability of high-stakes assessments such as the Membership of the Royal College of General Practitioner Clinical Skills Assessment ⁽²⁾
Code	Reliability relates to the consistency of "simulated" patient performance over time	Setting reliability as consistency of "simulated" patient performance between candidates

SP definition	Undefined	"A simulated patient simply presents the symptoms and problems of an actual patient." ⁽²⁾ A standardized patient is s trained to present those symptoms and problems in a consistent manner which does not vary from candidate to candidate, nor from standardised patient to standardised patient." ⁽²⁾
Attributes interfering with setting reliability	"SP performance as raters, task involved, rating scale, timeframe." ⁽¹⁾	Child fatigue, SP parent-role facilitation, cognitive level of understanding, performance embellishment, appearance
Approach to mitigate setting errors	Training process for standardize patient to consistently validate participant performance	Training process to develop children simulated patients as standardized patients
SP implementation purpose	Simulated patient portrayed a patient with diabetes Assess participant communication interview skills	Details about the pediatric case was not provided

Participants' skills assessed by SP	Participants were assessed on communication skills	Participant skills were not assessed
SP's performance assessment	Trainers assessed simulated patient performance and ability to accurately and consistency complete measurement instrument	Simulated patient children skills were assessed from 3 training phases Phase I, simulated patients' performance were used to refine cases development Phase II, performance without training assessment Phase III, trainers assess the calibration process that standardizes SP performance
Instructional design to refine SP performance	Eight and a half hours using the NHET- Sim Training Program, experiential activities and role modeling video for simulated patient to gain knowledge on how to use the assessment tool ⁽³⁾	Three phase action-based approach incorporating strategic calibration sessions and simulation environment familiarization

	Hill (2013) ⁽⁴⁾	Russell (2011) ⁽⁵⁾
Research purpose	Determine the content accuracy, reproducibility and replicability of standardized patient portrayal	Describe the processes used to prepare role players who simulate ‘patients’ in the Clinical Skills Assessment (CSA)
Reliability conceptualized	"Errors in SP presentation of history details may not negatively affect student performance in the clinical encounter as much as errors in physical presentation (Tamblyn et al. 1991), but even minor inaccuracies in SP portrayal can impact the student’s performance. ^(6, 7) The need for reliability of SP portrayal may be more reduced in situations where SPs are used for formative teaching rather than assessment situations.” ⁽⁸⁾	"We know from experience that the reliability of role player performance suffers when these elements (i.e., case run-through, body language, demeanor, opening statement review) are not covered efficiently at the calibration meeting.” ⁽⁵⁾

Code	Reliability as inaccuracies in "standardized" patient portrayal	Reliability as rehearsal inefficiency to standardize performance
SP definition	Standardized patient undefined. Simulated patient undefined.	"A simulated patient simply presents the symptoms and problems of an actual patient."⁽⁵⁾ "A standardized patient is s trained to present those symptoms and problems in a consistent manner which does not vary from candidate to candidate, nor from standardised patient to standardised patient."⁽⁵⁾
Attributes interfering with setting reliability	Content delivery, physical presentation, portrayal, adherence to training protocol	Case run-through, body language, demeanor, opening statement review
Approach to mitigate setting errors	Training standardized patients towards performance accuracy, reproducibility, and replicability	Training process development for "standardizing" patient performance

SP implementation purpose	Portray a parent or grandparent of a child with a speech delay Standardized patient did provide participant feedback	Standardized patient to portray a surgical patient for OSCE implementation Standardized patient not instructed to assess or provide participant feedback
Participants' skills assessed by SP	Participants were not assessed by standardize patient The simulation was a learning opportunity for participants to gain interpersonal and clinical communication skills	Participants were not assessed
SP's performance assessment	Trainer assessed standardized patient performance to deliver accurate clinical history content	Trainers assessed the calibration process of standardizing simulated patient performance
Instructional design to refine SP performance	Training occurred one-week prior to simulation implementation. Nine-hour training program with a speech pathologist, video footage observation of children with speech delays. Standardized patients memorized key feature of children with speech delays. Standardized patient withheld information when not prompted by the participant.	Half-day training program, 1-hour calibration session directly before setting implementation, standardized opening statements, demeanor, tone

	Baig (2014) ⁽⁹⁾	Shirazi (2011) ⁽¹⁰⁾
Research purpose	This aim is to valuate a tool that SPs can use to rate a participant's educational achievement in attitude, interactions skills and management (affective skills and clinical reasoning skills).	Create consistent (reliable) and accurate (valid) unannounced "standardized" patients as assessors
Reliability conceptualized	"The reliability in Objective Structured Clinical Exams (OSCEs) is based on variance introduced due to examiners, stations, items, standardized patients (SP), and the interaction of one or more of these sources of error." ⁽⁹⁾	Not described
Code	Reliability as setting variance due the dynamic interactions of setting attributes (i.e., examiners, stations, items, standardized patients)	Reliability as consistency of "standardized" patient performance and tool assessment
SP definition	Standardized patient undefined. Simulated patient term not used	A standardized patient is defined as an individual who has been well-trained to replicate the patient presentation in a clinical encounter setting in a consistent and realistic way.

Attributes interfering with setting reliability	Variance introduced due to examiners, stations, checklist items, standardized patients (SP), and the interaction of one or more of these sources of error, fatigue resulting from a heavy socio-emotional workload	Unspecified
Approach to mitigate setting errors	Training SP towards performance accuracy	Training process development for unannounced adult standardized patients to assess practitioner performance
SP implementation purpose	Standardized patients: Case A: portray chest pain, Case B: portray distended stomach Case C: portray concerned family member Case D: portray patient for oral contraceptive consultation	Standardized patient trained to portray a depressed patient Assess general practitioner performance of depression management
Participant skills being assessed by SP	Participants were not assessed by standardized patient	Unannounced standardized patients assessed general practitioner management of depression

SP's performance assessment	Trainers assess standardized patient portrayal accuracy and realism of emotions, facial expressions and body language	Trainers assessed accuracy (validity) and consistency (reliability) of standardized patient performance and assessment tool
Instructional design to refine SP performance	Training timeframe not provided. Standardized patient trained by the same rater for one case, and two raters across two location for this multisite study Scripted opening statement Standardized patient withheld information when not prompted by the participant.	Thirteen-hour training program, in-clinic participation, test-retest

REFERENCES

1. MacLean S, Geddes F, Kelly M, Della P. Simulated patient training: Using inter-rater reliability to evaluate simulated patient consistency in nursing education. *Nurse Educ Today*. 2018;62:85-90.
2. Russell D, Simpson R, Eden M, Foreman P. The use of child actors to simulate paediatric patients in the MRCGP Clinical Skills Assessment (CSA): developing an effective model. *Educ Prim Care*. 2015;26(4):242-7.
3. Nestel D, Mobley BL, Hunt EA, Eppich WJ. Confederates in health care simulations: not as simple as it seems. *Clinical Simulation in Nursing*. 2014;10(12):611-6.
4. Hill AE, Davidson BJ, Theodoros DG. The performance of standardized patients in portraying clinical scenarios in speech-language therapy. *Int J Lang Commun Disord*. 2013;48(6):613-24.
5. Russell D, Simpson R, Rendel S. Standardisation of role players for the Clinical Skills Assessment of the MRCGP. *Educ Prim Care*. 2011;22(3):166-70.
6. Keen A, Klein S, Alexander DA. Assessing the communication skills of doctors in training: reliability and sources of error. *Advances in Health Sciences Education*. 2003;8(1):5-16.
7. Tamblyn RM, Klass D, Schanbl G, Kopelow M. Factors associated with the accuracy of standardized patient presentation. *Academic Medicine*. 1990;65(9):S55-6.
8. Zubin Austin PG, Tabak D. Simulated patients vs. standardized patients in objective structured clinical examinations. *American Journal of Pharmaceutical Education*. 2006;70(5).
9. Baig LA, Beran TN, Vallevand A, Baig ZA, Monroy-Cuadros M. Accuracy of portrayal by standardized patients: results from four OSCE stations conducted for high stakes examinations. *BMC medical education*. 2014;14(1):97.
10. Shirazi M, Sadeghi M, Emami A, Kashani AS, Parikh S, Alaeddini F, et al. Training and validation of standardized patients for unannounced assessment of physicians' management of depression. *Academic Psychiatry*. 2011;35(6):382-7.