Assessing the Barriers to Mechanical Simulator Development

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Ryan DeLoach

Spring 2025

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

Pedro A. P. Francisco, Department of Engineering and Society

STS Research Paper

Introduction

In the medical world, all doctors and physicians take the hippocratic oath. Acting only in the best interests of the patient is at the core of a doctor's duties. They should take the best course of action regarding the treatment of a patient in any given case, providing the best possible outcome for the patient. Unfortunately, a dilemma is created when students are training, or even when an experienced physician who is learning a new skill. Training for a medical procedure is often directly in conflict with the patient's best interests, especially if there is someone more experienced available. A less experienced person would need to, for example, perform a surgery in order to gan experience in doing that surgery. In most cases, the trainee is closely supervised by an experienced physician who can take the lead if any complications arise. Still, the risk to the patient is inherently greater. In a case where an orthopedic resident was performing a total hip replacement for the first time, a femur was fractured, requiring a more complex procedure to be performed. This caused the patient to be "subjected to greater risk, discomfort, and potential complications" (Capozzi & Rhodes, 2015). In a systematic review, van der Leeuw et al. (2012) found that while in most situations, the risk was well mitigated through proper training and supervision, but "a minority of results pointed to some negative patient outcomes from the involvement of residents." The review concluded that training was one of the factors that could lead to improved patient outcomes.

In the past decades, training residents using simulators has become more prevalent in the education programs at many universities and hospitals (Elendu et al., 2024). Simulators have the potential to improve the skill of students without the need of patients; they allow practice through repetition on any specific procedure, training through the means of muscle memory, much like a professional sports player, chef, or any skilled laborer. This could prove to be

invaluable in some procedures, like closed reductions (a procedure in which a broken bone is set and often splinted), where the motion performed by the physician is predictable, but failure to perform correctly could result in damage to nerves or blood vessels (Ma, 2024).

The design, manufacture, and use of medical simulators is a complex process that will be unique for each one. With such a large number of possible procedures to simulate, it is difficult to determine which procedures should be simulated. Furthermore, simulation is a broad category, which can be narrowed to subcategories; narrowing will make it easier to define the general process of simulator design and testing. For the purposes of this paper, I will be reviewing the design, use, and validation of mechanical orthopedic simulators. Mechanical simulators are meant to simulate the physical feel of a procedure using a model, which students and physicians can practice on in place of patients. This leads to my research question; what barriers exist to making more accurate mechanical simulators, and when they are made, how do we know if they are actually a useful training tool?

The design of mechanical simulators is often the most difficult part of the process. There are many variables at play, and the variable nature of patients pose challenges to designing a "one-fits-all" solution. Bone density, muscle mass, body fat percentage, and overall size can be different for any given person. As an example, in a closed fracture reduction simulator, one of the more important variables is the aggregate force acting on the bone fragments (A. Freilich, personal communication, September 16, 2024). Every broken bone is slightly different, even in when the same bone is broken in the same location, and every person's muscles provide varying levels of force, so it is very difficult to make a model that is "accurate." No single simulator for this problem can ever be fully representative of its target procedure..

Another challenge in making simulators is the evaluation and testing methods used after they are created. There are many forms of validation, which will be discussed at length in the literature review of this paper. All of these involve the input of experts in their field, which can create some logistical issues, but generally there are doctors and medical students willing to help with testing simulators. A real problem comes when examining existing commercially available simulators, and figuring out that only a small portion of them have been validated using sound methods. In a study by Zendejas et al. (2013), only 14 out of 50 simulator studies examined used appropriate units for analyzing their design. Another study by Stunt et al. (2014) found that only 8% of commercially available studies were not known to be tested for validity. The say that predictive methods and methods that focus on patient outcomes, which are harder to do and require more time, tend to be better methods of analysis. Both studies posited that the effectiveness of a simulator is only able to be understood after proper analysis has been done.

I propose two main things that must be done in order to improve mechanical simulators: a design process which prioritizes the inclusion of expert feedback, and a testing process that adequately validates designs. Mechanical simulators are an important asset to medical education, and improvements in their design and functionality will have broader improvements for all patients.

Methodology

The first part of the research question is “what barriers exist to making accurate simulators?” In order to discover what barriers exist to making accurate simulators, a literature review will be conducted that will first examine two existing simulators, their design processes, manufacturing methods, and validation techniques. The simulators examined will be orthopedic

models, as this is the field where mechanical simulation the most common, and poorly performed operations have decent risks associated.

To determine the best possible design process, well performing models will be examined and the most crucial elements in their design method will be extrapolated. Their process will be compared with a standard engineering design process, which includes four main parts: defining the problem, generating solutions, selecting and developing a solution, and evaluating the design. As mentioned throughout this paper, evaluation of simulators is complex and contains nuances that must be explored.

The first part of the design process should be relatively straightforward. The problem to solve is similar across all simulators: improving medical training. In the case studies conducted, the initial challenge should be in agreement with this overarching goal.

The second part of the design process is to generate solutions. This can also be thought of as the "brainstorming" period of the project. Especially close attention will be paid to this portion, as this is where much of the thought processes that influence into a design are formed.

The third part of the design process is an extension of the second. In most projects, a solution is determined, and then the specifications are narrowed to constrain the design. These constraints help the development of a solution stay within the scope of the project while giving testable goals to measure against. Once a design meets all required specifications, it can be evaluated and iterated upon. In the context of simulator case studies, I will examine what kinds of specifications design teams utilize and determine the driving functions and needs of their solution. Without proper specifications defined, it is difficult to make an effective solution, so simulators that perform well should have stringent requirements.

The final part of the design process is evaluation: testing the solution against the design requirements and determining the level of success. This is where the analysis will deviate from the standard engineering process. Part of the solution is an improved and standardized way to validate simulator designs, which will inevitably influence the requirements of the design. The engineering teams that create design specifications might not know these best practices, so a more thorough literature review will be conducted to consolidate them. This will help answer the second part of the research question, "how do we know if they are effective?"

Literature Review

Testing a simulator is paramount to ensure that it has measurable benefits to those who train with it, and subsequently the patients that are treated as a result. It is also important to test simulators through useful metrics, as some methods can be misleading. For example, Carter et al. (2016) conducted a study that measured students' levels of confidence before and after training on a simulator. When comparing their self-reported confidence scores with a more adequate validity test, there was little to no correlation. That is, students were unable to assess their own improvement. There are four main types of validity that a simulator can be tested on. The first is called face validity; this is the most basic type of validity, in which a participant using the simulator can judge whether it "looks like" it accurately simulates the corresponding medical procedure. This is a purely subjective test, so it is somewhat irrelevant for evaluating medical simulators compared to other metrics (i.e. patient outcomes aren't necessarily correlated with face validity) (Harris et al., 2020).

The next type of validity is content validity. This tests whether a simulator contains all of the elements necessary in a procedure in the correct order. For example, to pass a content validity

test, a reduction simulator would have to be set and then splinted. If a trainee could set the fracture after splinting, the simulator would fail this test. This type of test is useful, but only to ensure all aspects of a procedure are present and accounted for (Atesok et al., 2019).

Next is construct validity, which is based on the simulator's ability to differentiate between experienced and inexperienced users. If an experienced physician performs the simulator's procedure often and is able to perform better than a person with limited knowledge and experience, then the simulator passes the construct validity test. This type of validity can be used to show the simulator as a good testing tool, but it does not determine whether it works as a training device. It only definitively shows a retroactive use, not a proactive one, but can still be very useful in determining the skill level of a physician (Gavazzi et al. 2011).

The last, and most thorough type of test, is a concurrent validity test. In this test, the simulator would be tested against a known source of validation. For example, a study that includes many physicians of varying experience levels could use the simulator. When they use the simulator, their performance will be assessed in reference to their previously validated skill. Previous validation of physician skill can come in a few ways. One is their actual experience doing a procedure, and the patient outcomes in those instances. Another is the Objective Structured Assessment of Technical Skill (OSATS), which is a validated technical test for physicians. Even the use of another simulator that has already been tested for validation can be used. The concurrent validity test has the most objective foundation, and simulators with high scores on this test tend to show better patient outcomes than those with lower scores (Atesok et al., 2019).

When looking at simulators that score well in concurrent validity tests, a few common design traits tend to occur. In the development of a distal radius fracture (i.e. a broken wrist)

simulator, a team from Carnegie Mellon University used OSATS scores to validate their design. This fulfills the concurrent validity test, meaning that the design was successful. The design process of this team was not heavily documented, but something that sticks out is their heavy interaction with experienced surgeons who have performed the procedure many times. This team was dedicated to making a simulator that was indistinguishable from a real arm when put under an x-ray, which also gave their design a high face validity, content validity, and construct validity. Every possible action that would have been taken by a physician working on a distal radius fracture was accounted for, and so this simulator passed every test for validity. They remark that iteration was key to ensure the x-ray images would turn out correctly; iteration is a cornerstone of engineering development (Raeker-Jordan et al., 2014).

Another team from Wolters Kluwer Health created low cost simulators that were meant to train students on percutaneous pinning of a humerus fracture (inserting stabilizing pins into a broken upper arm). The simulators, with the main goal of being low-cost while still retaining a high level of effectiveness, would not pass face validity tests. In order to prove its effectiveness, the team used a concurrent validity test by comparing the participants' scores on a previously validated simulator to their model. They found that their model was able to allow a novice physician to reach 85% of an expert's level of proficiency through repetition using the simulator. The design process here was simple, as the task being emulated was fairly simple. The main constraint they faced was finding a suitable material to mimic the feeling of a bone while retaining the low-cost goal of the design (Urwin et al., 2020).

Discussion: Design Process

In order to determine what a simulator should do or include. The procedure being simulated is the most important determining factor. The main steps in the procedure should be

examined, and their level of difficulty should be deduced by interviews with experts on the procedure. After either simplifying or expanding the desired functions of a simulator, the four main testing categories should be considered.

The level at which face validity is important should determine the visual accuracy of the model. A design team would ask the question, “would this simulator perform well if it did not look exactly like the real thing?” For example, in the distal radius fracture example, face validity was important due to the x-ray functionality of their design.

Then, the team would consider the content validity, asking which elements of the procedure are the most important. In the humerus fracture example, the team determined that the percutaneous pinning was the most useful skill to simulate, even though there are other steps to the procedure, like x-ray imaging or splinting.

The construct and concurrent validity of a design are more important to the testing phase of the completed device, but should be thought about in the design phase. Designers should understand that their device needs to be tested in a thorough and validated manner, and should plan ahead to do so. For example, a design team should include a part in the design that would allow for simple evaluation of a completed procedure that mimics a similar real life evaluation.

Overall, the design process for a simulator should be done with testing in mind. The specifics in each design should be determined based on feedback from experienced medical personnel. Beyond the technical part of the design process, factors like cost and reproducibility should be considered as well. A simulator that can be easily reproduced, either through means like 3D printing or through larger scale manufacturing, or can be easily bought at a low cost, will be widely more accessible and useful than a simulator that is expensive or hard to make. This can create a dilemma, as a higher fidelity product may lead to a more complete educational

experience, and thus a better outcome. However, a more widely available model can be used by more people, and thus the outcomes of more patients are improved, albeit to a lesser degree. Both case studies reviewed teams that chose to pursue cheap and easily accessible designs, and this is likely base on a utilitarian ethical standpoint.

Another area that might create a dilemma while designing simulators is the demographic aspect. It makes sense that demographic data about an injury should be examined to determine who is most affected by the injury. This would allow design teams to adjust the size and forces present in the simulator to more accurately represent the constituent group. Hypothetically, if children between the ages of 8-13 make up 90% of the patients for a certain fracture, it follows that a simulator should be designed for practice with that group. This can create an issue with the 10% of the injured that are part of other age ranges. If a physician only knows how to tend to a child, the skills might not transfer well to an adult who has much stronger and stiffer ligaments and muscles.

Discussion: Testing Process

Testing a simulator is just as important as making one. As outlined throughout this paper, certain types of testing are more conclusive than others in terms of improving patient outcomes. Specifically, concurrent validity testing is the standard in which simulators should be evaluated. The first step in testing for this is to measure the proficiency of users against some known metric. OSATS scores provide a good way to do this, but as seen in the case with the humerus fracture simulator, testing participants on a previously validated simulator works as well. However, in cases where a simulator is based on a novel procedure, or at least a procedure with no validated simulators in existence, OSATS scores can be used.

When testing a simulator, teams should include physicians of varying expertise to form a more complete dataset. In comparing the performance of differing levels of experience, construct validity can also be tested for. Multiple trials should be done with each physician to see how performance changes with the simulator through repetition. Ideally, the experts should only improve marginally (i.e. getting better at the simulator, as opposed to gaining new skills), while novices should improve more substantially (i.e. actually gaining the relevant skills). If a simulator shows this, along with concurrent validity, it should be considered a successful simulator.

Conclusion

Simulators are an important step in the further development of medical education, as they help to adress a perennial ethical dilemma between training and patient wellbeing. This paper explored the barriers to the development and evaluation of orthopedic simulators and proposed a solution in the form of a design and testing process. The transfer of knowledge from one generation of physicians to the next can be enhanced using well designed and tested simulators. The design of simulators should heavily rely on the input of experienced physicians, as their expertise is invaluable. The testing of simulators should rely on objective measures that take into account patient outcomes, like a concurrent validity test. Testing simulators using poor metrics can be misleading and have the potential to provide false positive outcomes.

In the process of designing a simulator, ethical dilemmas can arise. The decision whether to make a more expensive, but more techincally accurate simulator versus a cheaper, less accurate one can be difficult. Both have their place, and the decision should be made on a case by case basis. Another ethical dilemma is based on the demographics affected by different

injuries. Ideally, a simulator would be made for every injury and every demographic, but that might not always be possible.

In spite of the complex nature of medical simulator design and validation. They are and will be a mainstay in medical training programs. The broader goal of improving patient wellbeing should come first, and to do that, training practices must be improved. As simulators become more widespread and well studied, their effectiveness should only continue to grow.

References

Atesok, Kivanc, Hurwitz, Shepard, Anderson, Donald D., Satava, Richard, Thomas, Geb W., Tufescu, Ted, Heffernan, Michael J., Papavassiliou, Efstathios, Theiss, Steven, Marsh, J. Lawrence, Advancing Simulation-Based Orthopaedic Surgical Skills Training: An Analysis of the Challenges to Implementation, *Advances in Orthopedics*, 2019, 2586034, 7 pages, 2019. https://doi.org/10.1155/2019/2586034

Capozzi, J.D., Rhodes, R. Ethical challenges in orthopedic surgery. Curr Rev Musculoskelet Med 8, 139–144 (2015). https://doi.org/10.1007/s12178-015-9274-y

Carter, O. B. J., Mills, B. W., Ross, N. P., Miles, A. K., Mould, J. M., & O'Brien, R. P. (2016). Assessing simulation-based clinical training: comparing the concurrent validity of students' self-reported satisfaction and confidence measures against objective clinical examinations. *BMJ Simulation and Technology Enhanced Learning*, *2*(2), 30–34. https://doi.org/10.1136/bmjstel-2015-000089

Elendu, C., Amaechi, D. C., Okatta, A. U., Amaechi, E. C., Elendu, T. C., Ezeh, C. P., & Elendu, I. D. (2024). The Impact of simulation-based Training in Medical education: a Review. Medicine, 103(27), 1–14. https://doi.org/10.1097/MD.0000000000038813

Gavazzi, A., Bahsoun, A. N., Van Haute, W., Ahmed, K., Elhage, O., Jaye, P., Khan, M. S., & Dasgupta, P. (2011). Face, content and construct validity of a virtual reality simulator for robotic surgery (SEP Robot). The Annals of the Royal College of Surgeons of England, 93(2), 152–156. https://doi.org/10.1308/003588411x12851639108358

Harris, D. J., Bird, J. M., Smart, P. A., Wilson, M. R., & Vine, S. J. (2020). A Framework for the Testing and Validation of Simulated Environments in Experimentation and Training. Frontiers in Psychology, 11. https://doi.org/10.3389/fpsyg.2020.00605

Ma, B. (2024, June 17). Closed Reduction of a fractured bone [Review of Closed Reduction of a fractured bone]. MedlinePlus; National Library of Medicine. https://medlineplus.gov/ency/patientinstructions/000521.htm

Raeker-Jordan, Elisha A. MS; Martinez, Miguel; Aziz, Keith T. MD; Miles, Megan R. MD; Means, Kenneth R. Jr MD; LaPorte, Dawn M. MD; Giladi, Aviram M. MD, MS; Shimada, Kenji PhD. High-Fidelity Wrist Fracture Phantom as a Training Tool to Develop Competency in Orthopaedic Surgical Trainees. JAAOS: Global Research and Reviews 5(5):p e20.00224-8, May 2021. | DOI: 10.5435/JAAOSGlobal-D-20-00224

Stunt, J., Wulms, P., Kerkhoffs, G., Dankelman, J., van Dijk, C., & Tuijthof, G. (2014). How valid are commercially available medical simulators? Advances in Medical Education and Practice, 5, 385–395. https://doi.org/10.2147/AMEP.S63435

Urwin, J. , Trionfo, A. , Storey, E. , Brusalis, C. , Gajewski, C. & Lawrence, J. (2020). A Novel Low-Cost Surgical Simulation Tool for Pinning Supracondylar Humerus Fractures. *Journal of Pediatric Orthopaedics, 40* (5), e317-e321. doi: 10.1097/BPO.0000000000001460.

van der Leeuw, R. M., Lombarts, K. M., Arah, O. A., & Heineman, M. J. (2012). A systematic review of the effects of residency training on patient outcomes. BMC Medicine, 10(1). https://doi.org/10.1186/1741-7015-10-65

Zendejas, B., Brydges, R., Wang, A.T. et al. Patient Outcomes in Simulation-Based Medical Education: A Systematic Review. J GEN INTERN MED 28, 1078–1089 (2013). https://doi.org/10.1007/s11606-012-2264-5