THE PARADOX OF PREFERENCE VS. PERFORMANCE: TOWARDS A UNIFIED VIEW OF SIMULATION EXPERIENCE

Katelynn A. Kapalo, Kevin P. Pfeil, Pamela Wisniewski, and Joseph J. LaViola Jr.
University of Central Florida

Interactive simulation is one of the most prominent methods used to train and measure learning outcomes across multiple disciplines. Despite the ubiquity of simulation-based training in a variety of domains including nursing, serious games, military operations, etc., there is a paucity of research on how simulation experience is defined and how individual differences impact user experience. Towards this end, this paper provides a critical review of the existing literature. We describe how we can leverage existing findings and emergent themes to better understand and define simulation experience, and we outline areas for further investigation of the role of individual differences in user experience to enhance not only training outcomes, but also perception of simulation.

INTRODUCTION

Simulation-based training provides end users with the opportunity to learn and train important skills within virtual or mixed reality environments, reducing cost and risk, without compromising skill acquisition (Salas, Rosen, Held, & Weissmuller, 2009). For example, interactive simulations have been employed in high-risk training scenarios, such as military operations and medical procedures (Moroney & Lilienthal, 2008). They have also been used in pedagogical settings for teaching a diverse set of end users, including children, individuals with disabilities, and adults (Kincaid, Hamilton, Tarr, & Sangani, 2003; Nojavanasghari, Hughes, & Morency, 2017; Sitzmann, 2011). Towards this end, as technology advances, simulation-based solutions are more accessible to a wider audience.

As Human-Computer Interaction (HCI) researchers, we can contribute to the simulation and training domain by bridging the gap between user experience and successful skill acquisition. However, to do this, we need to be well-informed about the current state-of-the-art in simulation and training. Therefore, in this paper we examine the current state of how simulation experience is *evaluated and measured* across multiple domains. We provide a systematic review of simulation for training literature by examining multiple training domains, dimensions of simulation experience, and current measures and methods for assessing simulation experience. Our key research questions include:

- How is simulation experience defined across the diverse set of research domains in simulation and training?
- What methods are currently used to evaluate user experience of interactive simulation?
- How can user-centered evaluations be improved in simulation training and research?

To answer these questions, we critically evaluated 106 simulation and training peer-reviewed articles published between 2006 and 2017. Overall, simulation and training research within three primary domains emerged: healthcare education, serious gaming, and military operations. We argue, based upon our findings, that simulation and training research could be enhanced by applying a user-centered lens to account for individual differences, such as gender, age, and prior experience with technology. In this paper, we summarize the state-of-the-art in simulation experience research within training environments, synthesize literature across multiple domains, and recommend future research directions.

BACKGROUND

To understand simulation and training research, a discussion about several constructs is necessary. In the following sections, we define simulation experience and we discuss how studies on user experience have historically demonstrated that preference and performance are not always complementary. We then discuss the relationship between simulation fidelity and training as a way of understanding user experience, and finally, we define individual differences and how they potentially impact user experience with relation to simulation.

Defining Simulation Experience

The introduction of the term *simulation experience* evolved from need to assess and measure the impact of serious games (Raybourn, 2006). Predominately, simulation experience is defined as the perceived level of engagement end users have when interacting with the simulation. This would be considered the closest construct related to HCI research that has been widely studied in the context of simulation and training. Yet, it is important to note that simulation experience is often evaluated in the literature without a clear definition. For example, research in healthcare has defined simulation experience as the "authenticity of the experience," expressing authenticity on a continuum (Hotchkiss, Biddle, Fallacro, 2002). Other researchers have made a similar argument, using cue comparisons to evaluate simulation experiences (Dieckmann, Manser, Wehner, & Rall, 2007). Additionally, research has introduced the idea that self-reflection and debriefing are both an integral part of the simulation experience when a simulation is designed for educational purposes (Lasater, 2007). Therefore, it is unclear in many cases whether simulation experience refers to: the actual *act of training using a simulation*, level of *engagement*, level of *authenticity*, or whether simulation experience refers to more *metacognitive actions such as reflection or debriefing*. This conflation of simulation experience with other related constructs presents an issue to the HCI community and invokes a need to better understand the factors that underpin simulation-based training to provide a positive end user experience. As such, it is important to evaluate user perception of enjoyment for the purposes of better understanding how the psychological aspects of simulation impact the end user (Bargas-Avila, & Hornbæk, 2011). Motivation is one key component in understanding performance, and is in some cases, correlated with enjoyment; however, military simulations typically do not measure subjective user experience from the lens of enjoyment.

The Paradox of User Preference vs. Performance

Research in human factors points to inconsistencies between what the user wants and what is necessary to improve or support performance (Andre, & Wickens, 1995; Frøkjær, Hertzum, & Hornbæk, 2000; Nielsen & Levy, 1994). Frøkjær et al. (2000) found weak correlations between *satisfaction, efficiency, and effectiveness*, suggesting that research in user experience largely relies on a subset of these measures to evaluate systems. This study raised the question regarding *how* to effectively measure these three key components of user experience within more specific domains and applications. More recent work suggests we still do not fully understand the relationship between user satisfaction and effectiveness due to the lack of consistency in user preference and satisfaction measures (Hornbæk, 2006). Simulation-based training provides an ideal environment for such evaluations since they are largely domain-dependent, assess specific tasks or aspects of training, and require input from a human user. Thus, simulation provides a reliable way to evaluate product effectiveness (aspects of simulation fidelity) and user satisfaction (simulation experience). As highlighted in Bargas-Avila and Hornbaek's (2011) comprehensive review on user experience, it is critical to consider not only product qualities, but the quality of the interaction experience between the technology and the user. For this reason, we consider how simulation experience intersects with simulation fidelity and individual differences to provide a more holistic way of assessing simulation-based training.

The Intersection of Fidelity and Simulation Experience

Simulation fidelity is the degree to which a simulation or model represents the realistic characteristics of a specific environment, in this case, the environment in which users would train skills (Hays & Singer, 1989; 2012). The relationship between simulation fidelity, user experience, learning outcomes, and evaluation is unclear. For instance, recent research has demonstrated the disconnect between the level of fidelity required for transfer of training (ToT) and the perception of users (Brydges, Carnahan, Rose, Rose, & Dubrowski, 2010; Dahlstrom, Dekker, Van Winsen, & Nyce, 2009; Hamstra, Brydges, Hatala, Zendejas, & Cook, 2014; Norman, Dore, & Grierson, 2012). While there are inconsistencies in the literature regarding the degree to which physical fidelity is necessary for learning outcomes to be optimal, it has previously been demonstrated that users tend to prefer simulations that mimic the physical environment as opposed to simulations that emphasize the cognitive fidelity of training tasks (Hays & Singer, 2012). Dahlstrom et al. (2009) demonstrated that there are also differences between the perception of novice end users and experts, although we argue that this inclusion of differences falls outside the scope of this paper, since there are multiple factors that underpin expertise. The lack of consensus in the literature regarding the degree to which fidelity influences user perception provides a rich area for further investigation, since there are cognitive and user experience implications if a user does not perceive the simulation environment as aligned with their expectations. The paucity of literature in this area seems to point out that there is room for exploring the degree to which fidelity and simulation experience should be linked and how these two concepts inform other factors, such as immersion and presence. Finally, few studies have demonstrated the role of individual differences in understanding the degree to which simulation fidelity matters, and more importantly, whether these differences can be implicated in terms of simulation experience. Despite a relatively large body of research on simulation for training from the perspectives of learning outcomes, fidelity, and transfer of training, there is a lack of literature taking a user-centered approach to simulation and training research. While we do not refute that fidelity is a critical factor to understanding simulation experience, we want to further extend this body of knowledge to add that individual differences are also a key component to understanding why fidelity matters in simulation experience.

Individual Differences

Individual differences are defined as psychological traits or characteristics that vary from person to person (Eysenck & Eyesenck, 1987). The field of human factors has studied the influence of individual differences on performance for decades (Matthews, 2000; Motowildo, Borman & Schmit, 1997). Typically, these differences are framed in terms of cognitive, social, or behavioral differences, but may also include biological differences, such as gender. Due to the scope and nature of this work, we will be focusing exclusively on what has traditionally been studied in virtual reality contexts (Chen, Czerwinski, & Macredie, 2000). This is because individual differences are considered variables of interest for advancing performance, training, and/or practice and consequently have relevance to aspects of simulation-based training. The extant literature is sparse when it comes to providing recommendations for simulations that satisfy both the intended outcome and provide the end-user with a fulfilling or rewarding experience. Towards this end, the purpose of this paper is to provide an overview on the existing literature in this area, and to further extend these findings to encompass some recommendations for future research that will point toward a more comprehensive understanding of how individual differences may influence simulation experience.

METHODS

We conducted a literature review on peer-reviewed publications related to simulation and training between the years of 2006 and 2017. The year 2006 was chosen as a starting point due to the introduction of the term "simulation experience," in the literature (Raybourn, 2006). The process involved three iterations and was conducted between August 2016 and January 2018. We conducted our search via the ISI Web of Knowledge and Science Direct. The search terms were derived from key words used in combination such as*: "user experience of simulation," "simulation and individual differences," "immersion* and *individual differences,"* and *"simulation experience."* For the purposes of our literature review, articles must have met three criteria to warrant inclusion:

1. Peer reviewed and published work
2. Published between the years 2006 and 2017
3. Involved measuring or quantifying user experience in the context of simulation-based training or research

We synthesized the relevant topics related to behavioral, personality-related, or demographical individual differences. We excluded articles and works that covered biological differences at a micro-level (e.g., neurotransmitters, cellular differences, etc.). Additionally, papers related to generating or modeling responses to motivate behavioral change were also excluded. Our search terms identified over 400 papers that were initially reviewed for relevance based on the criteria above. Of these articles, 106 papers were included in this review. Because of space limitations, a limited, but representative selection of articles is included for discussion in this paper. We aggregated common thematic findings to communicate general trends in domains leveraging simulation.

RESULTS

Three main fields emerged as leading the research efforts in simulation experience: healthcare education, game-based training, and military training. Of these domains, healthcare contributed the most to the literature in simulation experience (62%), while serious games and military training contributed 22% and 16%, respectively. Within these domains, we outline the important conclusions from the literature and focus on an overview of the state-of-the-art training employed within these domains.

How is Simulation Experience Currently Measured?

Efforts to measure simulation experience, or user experience specifically related to simulation-based training, are primarily post-questionnaires administered after the training scenario. In the literature we evaluated, most questionnaires focused on perceptions of the system. For example, studies will ask students to rate their overall perception of the scenario using Likert scale ratings or questions about aspects of the training. Subjective student ratings of simulations used in the classroom or clinicals as part of the educational process has been an important addition to current nursing curriculums (Burns, O'Donnell, & Artman, 2010; Cant & Cooper, 2010; Harder, 2010; McCaughey & Traynor 2010; Rosen, 2008). Levett Jones et al. (2011) developed the Satisfaction with Simulation Experience Scale (SSE), an instrument specifically designed to measure students' perceptions of simulators varying in level of fidelity. This instrument, along with others, demonstrate the connection healthcare practitioners and educators identified as important for better design of educational and training programs.

How Often are Individual Differences Reported?

Individual differences are often ignored in the literature specifically relevant to simulation experience. While most studies in our review reported age or gender, under 10% reported analyses specifically evaluating gender or age differences. In the papers that did address gender differences, the majority was grounded in gaming literature (Hilgard, Engelhardt, & Bartholow, 2013; Jenson & De Castell, 2010; Kapalo, Dewar, Rupp & Szalma, 2015). Additionally, there were inconsistencies in the reporting of a participant or user's prior experience with technology. If it was measured, it was usually to point out that simulation-based training was a *novel* approach for that specific user group. Finally, there were very few (<10) papers that considered any motivational or self-regulation behaviors in their evaluations of the training or simulator. Despite the gap in the research specifically focused on investigating individual differences in simulation-based training, there are studies that have focused on factors such as gender, spatial ability, and age. Traditionally, individual differences have been studied in the context of the training task rather than from the lens of user experience (Motowildo et al., 1997). More recent research has emphasized the importance of including user feedback early in the design process (Vasalou, Ingram, & Khaled, 2012). However, we argue that similar importance should be placed upon understanding how these differences affect user experience for the purposes of enhancing simulation-based training. Basic psychological research has pointed to the cognitive factors that influence the effectiveness of training, but without accounting for how individual differences impact user experience, we are not adequately evaluating whether these simulations are most effective. By studying these individual differences, we not only improve simulation experience, but we potentially can develop more objective measures of simulation evaluation, leading to better performance measures and in turn, better outcomes for simulation-based training.

Overall Evaluation

Given the use of simulation in a variety of critical and complex domains such as nursing education, serious games, and military operations, it is necessary to examine the level of engagement the users experience when interacting with simulation. This feeling of engagement during the experience is what contributes to better understanding how to improve user experience to leverage the aspects of simulation that would allow the end user to gain more than just targeted learning outcomes. The assessment of simulation experience is an underdeveloped area of research. Efforts in healthcare education primarily compose the only comprehensive effort to create a measure evaluates simulation experience from the perspective of the end user (Cant & Cooper, 2010). These measures are primarily distributed via *self-report*. By recording students' performance in high fidelity, full-body simulators, it was discovered that there were several key components to understanding student learning and engagement (Levett-Jones et al., 2011). Simulation-based methods were perceived as favorable and satisfying to students, regardless of fidelity or realism. This finding has critical implications for developing cost effective training. In the case of the SSE, this measure is designed to assess simulation experience for nursing students (Levett-Jones et al., 2011). Although just one example, this represents the need to assess simulation experience in a variety of contexts. For logistical reasons and for limits on sampling in participant populations, it can be difficult to evaluate these individual differences. However, it does point to some valuable areas for further investigation: age, gender, and exposure to technology. Gender is largely an ignored individual difference, but it may be crucial to identifying some of the other individual differences in simulation experience. In the case of military simulation and training, particularly with infantry and pilot samples, due to the male gender bias and convenience sampling constraints, it is generally difficult to measure gender related differences since there are very few females in these roles. However, with the recent availability of combat positions to females, this demographic is changing (Rosenberg & Phillips, 2015).

To compensate and accurately provide a more holistic framework from which we draw conclusions about simulation fidelity and user experience, new data is necessary. Due to the focused nature of simulation-based training, age is also a factor that is not commonly examined. In the case of nursing students, data can be collected from a wider age range. Regarding military training, age limits are imposed on positions in the infantry; therefore, it is important to consider the age range of the user. In these cases, age is a consistent variable which can be helpful for understanding simulation experience from a specialized user group. This is important for evaluating emotional, motivational, or self-regulated behaviors that may correlate with age. In addition to gender and age, the exposure level to technology and to simulations must also be representative. In considering factors that would correlate with gender and age, acceptance of technology is a major individual difference. Students who have never experienced a simulation before may be evaluating this experience based on novelty rather than enjoyment or engagement (Levett-Jones et al., 2011). Novelty as a mediating factor may present challenges to designing simulators for experience professionals or experts. Thus, more holistic measures of simulation experience should be developed to better understand how prior experience influences perception. This approach could be augmented using physiological measures which may be able to better delineate which responses are based on novel stimuli as opposed to enjoyment or engagement (Mandryk, Inkpen, & Calvert, 2006).

LIMITATIONS

Although our literature search was exhaustive based on publicly accessible databases, we believe there is additional literature within the military domain that was not included in our review due to its restricted status. For this reason, we are not excluding the possibility that there may be more expansive research in this area; however, we are stating that to our knowledge, this is how the literature on simulation-based training appears within the scope of our search. Also, we have captured the theoretical underpinnings that drive research, but without subsequent empirical evidence, we cannot be sure that these theoretical foundations are applicable in all circumstances. Therefore, user studies and instruments for measuring simulation experience must continue to be refined and developed.

FUTURE WORK AND CONCLUSION

With the above discussion serving as a foundation, we offer the following research questions to apply and extend the findings from this thematic review:

1. How can we establish a consistent definition of simulation experience in the literature? More specifically, how can we explicitly link HCI methods to measure simulation-based training from a user's perspective?

2. HCI methods can bridge the gap in understanding how individual differences impact performance beyond self-reporting methods of data collection. For example, studies have successfully demonstrated the use of physiological measures to better understand user experience in the context of entertainment (Mandryk et al., 2006). Could a similar approach collect more objective data on engagement of the user?

3. How can we determine whether outcomes are a result of individual differences or novelty of experience within a simulation-based training context?

4. Based upon observations related to exposure to technology, can we apply and extend the literature on computer-based efficacy to simulation-based training (Marakas, Johnson, & Clay, 2007)? That is, do beliefs related to one's own ability to use the simulator competently and correctly mediate or moderate performance? If so, do these beliefs stabilize or change over time?

By posing the above research questions, we outline to extend this knowledge and evaluate simulation experience using a more holistic approach. By leveraging what we know from the existing literature, we hope to provide a foundation for researchers to further investigate user experience within interactive simulation.

ACKNOWLEDGMENTS

We would like to thank the members of the UCF Interactive Computing Experience Research Cluster. The authors would also like to acknowledge the anonymous reviewers and Hannah Kapalo for comments that improved the manuscript. These views are the authors alone and do not represent any governmental agency.

REFERENCES

Andre, A. D., & Wickens, C. D. (1995). When users want what's not best for them. *Ergonomics in design*, *3*(4), 10-14.

Bargas-Avila, J. A., & Hornbæk, K. (2011, May). Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2689-2698). ACM.

Brydges, R., Carnahan, H., Rose, D., Rose, L., & Dubrowski, A. (2010). Coordinating progressive levels of simulation fidelity to maximize educational benefit. *Academic Medicine*, *85*(5), 806-812.

Burns, H. K., O'Donnell, J., & Artman, J. (2010). High-fidelity simulation in teaching problem solving to 1st-year nursing students: a novel use of the nursing process. *Clinical Simulation in Nursing*, *6*(3), e87-e95.

Cant, R. P., & Cooper, S. J. (2010). Simulation□based learning in nurse education: systematic review. *Journal of advanced nursing*, *66*(1), 3-15.

Chen, C., Czerwinski, M., & Macredie, R. (2000). Individual differences in virtual environments—introduction and overview. *Journal of the Association for Information Science and Technology*, *51*(6), 499-507.

Dahlstrom, N., Dekker, S., Van Winsen, R., & Nyce, J. (2009). Fidelity and validity of simulator training. *Theoretical Issues in Ergonomics Science*, *10*(4), 305-314.

Davies, D. R., Matthews, G., Stammers, R. B., & Westerman, S. J. (2013). *Human performance: Cognition, stress and individual differences*. Psychology Press.

Dieckmann, P., Manser, T., Wehner, T., & Rall, M. (2007). Reality and fiction cues in medical patient simulation: an interview study with anesthesiologists. *Journal of Cognitive Engineering and Decision Making*, *1*(2), 148-168.

Eysenck, H. J., & Eysenck, M. W. (1987). *Personality and individual differences*. New York, NY: Plenum.

Frøkjær, E., Hertzum, M., & Hornbæk, K. (2000, April). Measuring usability: are effectiveness, efficiency, and satisfaction really correlated?. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 345-352). ACM.

Hamstra, S. J., Brydges, R., Hatala, R., Zendejas, B., & Cook, D. A. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine*, *89*(3), 387-392.

Hilgard, J., Engelhardt, C. R., & Bartholow, B. D. (2013). Individual differences in motives, preferences, and pathology in video games: the gaming attitudes, motives, and experiences scales (GAMES). *Frontiers in psychology*, *4*, 608.

Hays, R. T., & Singer, M. J. (1989). Simulation fidelity as an organizing concept. In *Simulation Fidelity in Training System Design* (pp. 47-75). Springer, New York, NY.

Hays, R. T., & Singer, M. J. (2012). *Simulation fidelity in training system design: Bridging the gap between reality and training*. Springer Science & Business Media.

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International journal of human-computer studies*, *64*(2), 79-102.

Hotchkiss, M. A., Biddle, C., & Fallacaro, M. (2002). Assessing the authenticity of the human simulation experience in anesthesiology. *AANA journal*, *70*(6), 470-473.

Jenson, J., & De Castell, S. (2010). Gender, simulation, and gaming: Research review and redirections. *Simulation & Gaming*, *41*(1), 51-71.

Kapalo, K. A., Dewar, A. R., Rupp, M. A., & Szalma, J. L. (2015, September). Individual Differences in Video Gaming: Defining Hardcore Video Gamers. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 59, No. 1, pp. 878-881). Sage CA: Los Angeles, CA: SAGE Publications.

Kincaid, J. P., Hamilton, R., Tarr, R. W., & Sangani, H. (2003). Simulation in education and training. In *Applied system simulation* (pp. 437-456). Springer, Boston, MA.

Lasater, K. (2007). High-fidelity simulation and the development of clinical judgment: Students' experiences. *Journal of Nursing Education*, *46*(6), 269.

Levett-Jones, T., McCoy, M., Lapkin, S., Noble, D., Hoffman, K., Dempsey, J., Arthur, C. & Roche, J. (2011). The development and psychometric testing of the Satisfaction with Simulation Experience Scale. *Nurse Education Today*, *31*(7), 705-710.

Mandryk, R. L., Inkpen, K. M., & Calvert, T. W. (2006). Using psychophysiological techniques to measure user experience with entertainment technologies. *Behaviour & information technology*, *25*(2), 141-158.

Marakas, G. M., Johnson, R. D., & Clay, P. F. (2007). The evolving nature of the computer self-efficacy construct: An empirical investigation of measurement construction, validity, reliability and stability over time. *Journal of the Association for Information Systems*, *8*(1), 15.

Matthews, G. (2000). *Human performance: Cognition, stress, and individual differences*. Psychology Press.

McCaughey, C. S., & Traynor, M. K. (2010). The role of simulation in nurse education. *Nurse education today*, *30*(8), 827-832.

Moroney, W. F., & Lilienthal, M. G. (2008). Human factors in simulation and training. *Human Factors in Simulation and Training. CRC Press*, 3-38.

Nielsen, J., & Levy, J. (1994). Measuring usability: preference vs. performance. *Communications of the ACM*, *37*(4), 66-75.

Nojavanasghari, B., Hughes, C. E., & Morency, L. P. (2017, May). Exceptionally social: Design of an avatar-mediated interactive system for promoting social skills in children with autism. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 1932-1939). ACM.

Norman, G., Dore, K., & Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. *Medical education*, *46*(7), 636-647.

Raybourn, E. M. (2006). Applying simulation experience design methods to creating serious game-based adaptive training systems. *Interacting with computers*, *19*(2), 206-214.

Rosen, K. R. (2008). The history of medical simulation. *Journal of critical care*, *23*(2), 157-166.

Rosenberg, M., & Phillips, D. (2015). All combat roles now open to women, Defense Secretary says. *New York Times*, *3*.

Salas, E., Rosen, M. A., Held, J. D., & Weissmuller, J. J. (2009). Performance measurement in simulation-based training: A review and best practices. *Simulation & Gaming*, *40*(3), 328-376.

Sitzmann, T. (2011). A meta□analytic examination of the instructional effectiveness of computer□based simulation games. *Personnel psychology*, *64*(2), 489-528.

Vasalou, A., Ingram, G., & Khaled, R. (2012, June). User-centered research in the early stages of a learning game. In *Proceedings of the Designing Interactive Systems Conference* (pp. 116-125). ACM.