

Millions of dollars and countless hours are invested in the development of serious games for education internationally every year. Developers and designers generally evaluate the learning systems that they develop through some objective measure. These measures are often internally designed or imposed upon the development team by the sponsoring agency. A gap exists in the comprehensiveness of these evaluations, as the user experience analysis does not necessarily quantify the effectiveness and efficacy of the specific game or experience being evaluated. "Serious games are mainly assessed in terms of the quality of their content, not in terms of their intention-based design" (Mitgutsch & Alvarado, 2012). Measurements of user experience conducted by designers are generally positive, indicating user engagement, enjoyment, and preference over other methods of instruction (Dunleavy, Dede, & Mitchell, 2009; Thomas, William John, & Delieu, 2010). While this is a valuable finding, which some may rely on to allude to engagement, the often subjective construct of engagement without demonstrated benefits to learning has a diminished generalizable meaning. As Frokjaer explains, the correlation between user satisfaction and effectiveness are often negligible and should be looked at separately (Frokjaer, Hertzum, & Hornbaek, 2000). Because outcome research results are specific to the samples (or populations from which they were drawn) and the outcomes measured, "it is essential that conclusions from the research be clear as to the population(s) and outcomes for which efficacy is claimed" (Flay, 2004). Flay goes on to explain that, "Effectiveness trials test whether interventions are effective under 'real-world' conditions or in 'natural' settings. Effectiveness trials may also establish for whom, and under what conditions of delivery, the intervention is effective" (Flay, 2004).

Some prevalent ways to evaluate the efficacy and effectiveness of learning tools include performance improvement assessments, blind coder ratings, qualitative and quantitative self-reports of social presence, questionnaires, and ultimately performance tests that measure improvement in desired knowledge, skills, or abilities (Bailenson, 2006; Botella, Bretón-López, Quero, Baños, & García-Palacios, 2010). Whether learning objectives in the serious games are explicit and didactic, or more discovery or inquiry based, there should be objectives for any serious game that are based on learning outcomes.

This article explores the intersections of the areas of efficacy, effectiveness, and user experience in assessing serious games and simulated experiences. The author builds the argument for a holistic approach to evaluating learning games and computer mediated experiences. Some examples are explored in which reasonably effective evaluations could have been improved by a more holistic approach to evaluation. The terms engagement, presence, efficacy, and learning are operationally defined based on research within the fields of education, learning theory, game design theory, and simulation. These constructs are then compiled to explain the need for a holistic approach to evaluating learning games in order to ensure usability, learning, and transfer of learning to the real world. The implications of industry embracing the holistic evaluation of games will include not only improvement in the output of development teams and consistency between evaluations of different systems, but also an approach to iterative evaluation driven by the constructs that contribute to effective learning through games.

- Bailenson, J. (2006). *The Effects of Fully Immersive Virtual Reality on the Learning of Physical Tasks*. Paper presented at the The 9th International Workshop on Presence Cleveland, Ohio.
- Botella, C., Bretón-López, J., Quero, S., Baños, R., & García-Palacios, A. (2010). Treating Cockroach Phobia With Augmented Reality. [Feature Article]. *Behavior Therapy, 41*(3), 401-413.
- Dunleavy, M., Dede, C., & Mitchell, R. (2009). Affordances and Limitations of Immersive Participatory Augmented Reality Simulations for Teaching and Learning. *Journal of Science Education and Technology, 18*(1), 7-22.
- Flay, B. (2004). Standards of Evidence book Retrieved from <http://www.preventionresearch.org/StandardsofEvidencebook.pdf>
- Frokjaer, E., Hertzum, M., & Hornbaek, K. (2000, 2000). *Measuring Usability: Are Effectiveness, Efficiency, and Satisfaction Really Correlated?*
- Mitgutsch, K., & Alvarado, N. (2012). *Purposeful by design?: a serious game design assessment framework*. Paper presented at the Proceedings of the International Conference on the Foundations of Digital Games, Raleigh, North Carolina.
- Thomas, R. G., William John, N., & Delieu, J. M. (2010). Augmented Reality for Anatomical Education. [Article]. *Journal of Visual Communication in Medicine, 33*(1), 6-15. doi: 10.3109/17453050903557359
- Witther, J., Tsai, Y.-T., & Azuma, R. (2011). Indirect augmented reality. *Computers & Graphics, 35*(4), 810-822. doi: 10.1016/j.cag.2011.04.010