

# Simulations for STEM Learning: Systematic Review and Meta-Analysis

## Report Overview

Cynthia D'Angelo, Daisy Rutstein, Christopher Harris, Geneva Haertel,  
Robert Bernard, Evgueni Borokhovski

March 2014

## Purpose

The rise in computing has been accompanied by a decrease in the costs of computers and an increase in the use of simulations for learning. In the fields of science, technology, engineering and mathematics (STEM) field in particular, real equipment can be difficult to obtain, so simulations let students experience scientific phenomena they normally would not be able to experience firsthand. The potential benefits of simulations, such as allowing for phenomenon to be investigated at different physical scales and time periods, have led many experts to suggest that using simulations in the classroom can help improve learning (NRC, 2011a).

Several recent literature reviews (e.g., Smetana & Bell, 2012; Scalise, et al., 2011) have focused on whether and in what ways simulations aid the improvement of student learning. Some of these reviews are focused on a very narrow range of simulation studies while others are focused only on overall trends of the findings of these studies, but none conducted a comprehensive quantitative meta-analysis. To date, the simulation literature has not been systematically reviewed and quantitatively summarized to determine if simulations do in fact have an effect on student learning.

This review addresses this gap in the current research literature on computer-based simulations for STEM learning. It was conducted by a team of researchers at SRI International under a contract with the Bill & Melinda Gates Foundation. The review looked at effects of and the role of computer-based simulations for learning in K-12 education in the areas of STEM education.

A simulation, for the purposes of this review, was defined as a computer-based, interactive environment with an underlying model. This review focused on computer-based simulations that are neither simple visualizations nor games, recognizing that a continuum exists among these types of digital tools. For the purposes of this review, simulations needed to be constructed with an underlying model that was based on some real-world behavior or natural/scientific phenomena (such as models of the ecosystem, or simulated animal dissections). In addition, all studies included in the review are characterized by simulations that include some amount of user interactivity, centered on inputs and outputs of the model.

This review examined studies that compared simulation-based instruction to non-simulation-based instruction as well as studies that compared two versions of simulation-based instruction to each other.

## SRI Education

Developed by SRI International with funding from the Bill & Melinda Gates Foundation.

This report is based on research funded by the Bill & Melinda Gates Foundation. The findings and conclusions contained within are those of the authors and do not necessarily reflect positions or policies of the Bill & Melinda Gates Foundation.

# Meta-Analysis Approach

A meta-analysis is the systematic synthesis of quantitative results from a collection of studies (Borenstein, et al., 2009) focused on a given topic. Part of the systematic approach in a meta-analysis is to document the decisions that are made regarding the collection of the articles and the steps of the analysis. In a meta-analysis, articles are included based on pre-defined criteria and not due to favorable results found in the article or familiarity with certain authors. This can help to remove some of the bias and subjectivity that would result from a less systematic review.

Meta-analysis quantifies results by using effect sizes. Effect sizes are a measure of the difference between two groups. In the case of an intervention, an effect size can be thought of as a measure of the (standardized) difference between the control group and the treatment group, thereby providing a measure of the effect of the intervention. Effect sizes are not the same as statistically significant differences that are typically reported and determined through the use of inferential statistics, such

as t-tests or analysis of variance (ANOVAs). A research study, for example, could have a statistically significant finding, but the effect of that difference could be minimal. Thus the effect size allows researchers to determine the magnitude of the impact of an intervention, not just whether or not the intervention made a difference. For example, an effect size of 1.00 would be interpreted as a difference of one standard deviation between the two groups being compared. Another way of interpreting a one standard deviation effect size would be moving a student at the 50th percentile before the intervention to the 84th percentile after the intervention.

## Search Criteria

The initial search criteria for this meta-analysis study were peer-reviewed journal articles published between 1991 and 2012, using three databases (ERIC, PsychINFO, and Scopus) and the search terms “simulation”, “linked representation”, and “dynamic representation” paired with various STEM terms (such as “science education” and “mathematics education”). A total of 2,722 articles were identified as part of this initial search. A team of researchers screened the abstracts of these articles to ensure a match with study inclusion criteria. From this screening process, we identified 260 articles that met our initial criteria. We examined these 260 studies and found

a total of 59 unique studies that were either experimental (i.e., random assignment with treatment and control groups) or quasi-experimental (i.e., not randomized but with treatment and control groups).

For each of the 59 included studies, we identified and recorded the simulation intervention, participants and settings, research questions, research methodology, assessment instrument information, implementation information, and results.

# Meta-analysis Results

From these included 59 studies, we extracted 128 effect sizes; 96 of these effects were in the achievement outcome category, 17 were in the scientific inquiry and reasoning skills category, and the remaining 15 were in the non-cognitive measures category.

## Achievement outcomes

For achievement outcomes, when computer-based interactive simulations were compared to similar instruction without simulations there was a moderate to strong effect in favor of simulations ( $g+ = 0.62, p < .001$ ). This category included 46 effect sizes. The improvement index (i.e., percentage of increase between the control mean and the treatment mean) was 23%. This means that the achievement of students at the median in the control group (no simulation) could have increased 23% if they had received the simulation treatment.

For achievement outcomes, when computer-based interactive simulations were modified to include further enhancements (such as additional learner scaffolding and certain kinds of feedback) and then compared to their original non-modified versions, the modified simulations had a moderate effect on student learning over the non-modified simulations ( $g+ = 0.49, p < .001$ ). This category included 50 effect sizes. The improvement index was 19%.

## Scientific Inquiry and Reasoning Skills

For scientific inquiry and reasoning skills outcomes, when computer-based interactive simulations were compared to similar instruction without simulations there was a modest effect in favor of simulations ( $g+ = 0.26, p = .17$ ). The improvement index was 10%. Of note is that only six effect sizes fell into this category, so this finding is not generalizable.

For scientific inquiry and reasoning skills outcomes, when computer-based interactive simulations were modified to include further enhancements and then compared to their original non-modified versions, the modified

simulations had a moderate effect on scientific inquiry and reasoning skills above the non-modified simulations ( $g+ = 0.41, p < .001$ ). This category included 11 effect sizes. The improvement index was 16%.

## Non-Cognitive Outcomes

For non-cognitive outcomes, when computer-based interactive simulations were compared to similar instruction without simulations there was a moderate to strong effect in favor of simulations ( $g+ = 0.69, p < .001$ ). This category included 12 effect sizes. The improvement index was 26%.

For non-cognitive outcomes, when computer-based interactive simulations were modified to include further enhancements and then compared to their original non-modified versions, the modified simulations had a moderate effect on non-cognitive outcomes above the non-modified simulations ( $g+ = 0.52, p < .001$ ). Of note is that only three effect sizes fell into this category, so this finding is not well suited for generalization.

These effect sizes were further categorized according to STEM areas and research questions. In the area of science achievement outcomes, there were a sufficient number of effect sizes to report moderator variable findings in addition to overall findings by category. When comparing instruction that used simulations to instruction that did not, results suggest that the number of sessions has an impact on student achievement outcomes. When comparing student performance on simulations to the same simulation with modifications, results suggest that the use of non-embedded, technology-based assessment reported significantly smaller learning outcomes than embedded technology-based assessments or non-technology-based assessments. More details on these moderator variable findings (such as types of simulations, types of modifications) can be found in the executive summary of this report.

## Assessment Information

There were 76 assessments used in the 59 articles. Some articles contained multiple outcome measures which each corresponded to unique assessments. In other articles the same assessment was used for multiple studies leading to the same assessment corresponding to multiple effect sizes.

Results from our analysis of the assessments indicate that, when investigating the effectiveness of computer-based simulations in the classroom, most studies use researcher-designed paper/pencil assessments. Moreover, we found that studies typically did not provide sufficient detail about the assessments. While nearly a third of the assessments were taken from other studies, in which case further information may be contained in other articles, most studies did not clearly describe the types of items or tasks that comprised their assessments, or their purposes, design method, and technical qualities.

Reliability measures were reported more often than validity measures, with the most common reliability measure being internal consistency. Validity was often established by having experts review the assessments to determine the content validity and representativeness of the assessment. Some studies also included inter-rater reliability, and psychometric analyses performed using data collected from pilot studies as evidence of the validity of the inferences drawn from the assessment.

Overall, within the collection of studies that reported sufficient information on their assessments, we found wide variation in the types of constructs addressed and in the number and types of items included on the assessments. The studies also varied on the length of time between the use of the simulation and the administration of the assessment.

## Implications

These findings will help GlassLab and other educational simulation designers make appropriate, research-based decisions in developing new educational materials for students. Teachers, researchers, and policy makers will also gain insight about how simulations can impact STEM learning. Knowing which factors relating to the implementation of simulation-based instruction have contributed to positive effects in past simulation studies can help educators and researchers plan what is likely to work in current projects (including the contexts and settings where those successes occurred).

Most of the included studies were in the area of science. Other STEM disciplines were underrepresented in this study because we found few research articles meeting our criteria in engineering, technology, and mathematics. Overall, the list of included studies indicates that there is a need for a more robust pool of high quality research studies across a range of disciplines in the K-12 grade range. The opportunity to carefully examine the research designs, assessments, and statistical methods used in the current corpus of simulation research studies can provide

guidance to researchers developing future research designs, assessments and statistical methods. Most of the assessments used to measure outcomes associated with simulations were pencil and paper-based. Few of the assessments used in existing studies took advantage of the affordances of the technology involved in the simulations that were studied.

These findings have many broad implications for the field of science education. They show that simulations can have a significant positive impact on student learning and are promising tools for improving student achievement in science. Simulations are a key way that students interact with models, which are an important focus in the new K-12 Science Framework (NRC, 2011b) and recently released Next Generation Science Standards (NGSS Lead States, 2013). This is especially relevant for models that are based on phenomena that are difficult to observe in the typical classroom setting (due to reasons such as scale, time, safety, or budget limitations), an area where simulations excel.

## Conclusion

In this report, we have described our systematic review and meta-analysis of the literature on computer simulations designed to support science, technology, engineering, and mathematics (STEM) learning in K-12 instructional settings. Both quantitative and qualitative research studies on the effects of simulation in STEM were reviewed. Studies that reported effect sizes or the data to calculate effect sizes were included in the meta-analysis. Results from the meta-analysis of 59 studies indicate that, overall, simulations have a beneficial effect over treatments in which there were no simulations. Also, simulations with modifications were shown to have a

beneficial effect over simulations without modifications. It is important to note that the studies included in the meta-analysis were predominately in science education, suggesting that an important need is a more robust pool of high quality research studies on simulations in other STEM domains at the K-12 level. Thus, while our work shows that simulations, in many different configurations or contexts within the classroom, can improve student learning, there is still much to be learned about the educational benefits of computer simulations across the STEM domains.

## References

- Borenstein, M., Hedges, L., Higgins, J. & Rothstein, H. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.
- Clark, D. B., Nelson, B., Sengupta, P., D'Angelo, C. M. (2009). Rethinking Science Learning Through Digital Games and Simulations: Genres, Examples, and Evidence. Invited Topic Paper in the *Proceedings of The National Academies Board on Science Education Workshop on Learning Science: Computer Games, Simulations, and Education*. Washington, D.C.
- National Research Council. (2011a). *Learning science through computer games and simulations*. Committee on Science Learning: Computer Games, Simulations, and Education, Margaret A. Honey and Margaret L. Hilton, Eds. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- National Research Council. (2011b). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. Committee on a Conceptual Framework for New K-12 Science Education Standards. Board on Science Education, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
- NGSS Lead States (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press
- Scalise, K., Timms, M., Moorjani, A., Clark, L., Holtermann, K., & Irvin, P.S. (2011). Student learning in science simulations: Design features that promote learning gains. *Journal of Research in Science Teaching*, 48(9), 1050-1078.
- Smetana, L. K., & Bell, R. L. (2012). Computer simulations to support science instruction and learning: A critical review of the literature. *International Journal of Science Education*, 34(9), 1337-1370.

Developed by SRI International with funding from the Bill & Melinda Gates Foundation.

**SRI** Education

SRI International • 333 Ravenswood Avenue  
Menlo Park, CA 94025 • [www.sri.com/education](http://www.sri.com/education)  
Phone: 650.859.2995 • Email: [education@sri.com](mailto:education@sri.com)

**BILL & MELINDA**  
GATES *foundation*

© 2014 Bill & Melinda Gates Foundation. All Rights Reserved.  
Bill & Melinda Gates Foundation is a registered trademark  
in the United States and other countries.