# Integrating Scientifically Based and Design Experiment Research

Steven McGee
Jennifer Kirby

Center for Educational Technologies®
Wheeling Jesuit University
Wheeling, WV

**Abstract**

This study explored tradeoffs between two research methodologies – scientifically based research and design experiments. We conducted this exploration through a case study of one teacher's four-year implementation of a multimedia learning environment called *Astronomy Village*®. During the first year of her implementation, she participated in the scientifically based summative evaluation of *Astronomy Village*. Her students successfully learned complex solar system concepts. In subsequent years, the teacher customized her implementation to meet local needs. An analysis of her design decisions and resulting learning outcomes reveals that the subsequent implementations were more successful than the first-year implementation. Results indicate the importance of tracking long-term implementation. They also reveal tradeoffs between scientifically based and design experiment research.

## 1 Objectives and Significance

The long-term success of any educational program depends on the extent to which teachers can implement the program without direct support from the program developers. A successful scientifically based evaluation should not be the stopping point for a project. It is important to monitor the program after the developers no longer provide direct support to see if the participating teachers can continue to be successful. The summative evaluation of the National Science Foundation-funded *Astronomy Village: Investigating the Solar System*® indicated that students who used the program significantly improved in their understanding of complex solar system concepts. During the summative evaluation teachers were constrained in their implementation based on the quasi-experimental design of the summative evaluation. It was important for comparability to exist across the participating classrooms. After participating in the summative evaluation effort, teachers were free to implement the program in ways that integrated with their local curriculum. Teacher K has implemented the program over four consecutive school years. We have analyzed both the yearly curriculum adjustments as well as the resulting learning outcomes. The goal of this study is to examine the extent to which K was able to adapt the program to her local context in a manner congruent with the goals of the software and to examine the extent to which her adaptations led to increased learning outcomes.

## 2 Theoretical Framework

The design experiment approach, a recent advancement in educational research, is a powerful means to conduct ongoing research and evaluation of educational multimedia (Brown, 1992; Cobb et. al., 2003, McGee & Howard, 1998). Using this approach, researchers work with teachers to implement a program and evaluate the impact of the program on student learning. By reflecting on student performance, teachers and researchers can identify areas of weakness and make adjustments for the next implementation of the program. A new design experiment cycle begins as teachers implement the program using the new adjustments and then once again evaluate student performance.

The design experiment approach stands in contrast to scientifically based research in which researchers attempt to isolate the effects on a given phenomenon in order to generate a causal explanation. In the randomized experimental approach researchers can manipulate only one variable at a time to determine the effect of that one factor. In contrast, the primary goal of design experiments is to create classroom conditions that will lead to increased learning outcomes. It is often necessary to manipulate several variables at the same time, making it difficult to determine the effects due to any one variable.

For studies of *Astronomy Village*® implementation, we have combined the design experiment approach with the scientifically based research approach. During the summative evaluation of the program, K recruited a colleague at her school to serve as the matched no treatment comparison group. The comparison students were administered the same pre- and posttest measures at the same times as K's students. The summative evaluation served as a baseline for future implementations. In subsequent years, K was able to use the design experiment approach to make principled changes to her instruction in an attempt to improve instruction. Using qualitative analysis techniques, we characterized the nature of the changes from one year to the next. With the quasi-experimental approach we are able to remove any nontreatment effects within the first implementation, such as maturation or history effects. In subsequent years, a time series design allows for investigation of changes from one year to the next. The integration of both methodologies allows for an exploration of the tradeoffs of each approach.

## *3 Astronomy Village: Investigating the Solar System®*

Through Astronomy Village students are transported to a virtual village in Hawaii where they investigate one of two core research topics: what the surface of Pluto might look like when the first NASA mission arrives in 2015, or the search for life in the solar system (McGee & Howard, 1999). The program is designed such that a virtual mentor guides students in completing multiple investigation cycles that mirror the phases of scientific inquiry. In the first investigation cycle students are introduced to the core research question concerning either the surface of Pluto or the core requirements for life. The exploration phase of the investigation prepares students for data collection and analysis by exposing them to the types of data they will be using later in the investigation. In the background research phase students read library articles and listen to lectures to help them understand key background concepts. During the main part of the module, the data collection and analysis phases, students use the results of their analyses to draw conclusions about the research question. This core investigation cycle lasts about one week. Students then follow the same sequence of phases as they did in the core investigation when they undertake a focused investigation on a narrower topic. For example, students may examine temperature/pressure relationships on a variety of planets and moons to determine where the conditions are right to support liquid water. Students complete the investigation by hosting a virtual press conference in front of a virtual press corps that asks the students questions about the investigation they just completed.

## 4 Assessment Instrument

As part of the summative evaluation effort, researchers at the Center for Educational Technologies® created an assessment instrument to measure student-learning outcomes related to solar system astronomy. There were three guiding principles for the design of the assessment instrument. First, the assessment instrument should reflect important thinking and problem-solving skills from the discipline of planetary science (Hickey, Wolfe, & Kindfield, 1999; Sheppard, 2000). Second, the instrument should measure the extent to which students transfer their thinking and problem-solving skills into new contexts (Bransford, Brown, & Cocking, 1999). And third, the assessment should be easy to administer and score for the target population.

We identified the key complex content ideas that were presented in each of the nine investigations within *Astronomy Village®* along with the key problem-solving skills related to drawing conclusions from data and inferring planetary processes from analyzing images of surface features. We identified publicly available NAEP and TIMSS assessment items that addressed those concepts. We also contracted with item writers to develop the assessment items related to the underlying concepts within the investigations. There were two resulting instruments—one for the Search for Life core research investigation and one for the Mission to Pluto core research investigation (Dimiter, McGee, & Howard, 2001). This study focused on the Search for Life test.

## 5 Design Experiment

K conducted the Search for Life core research modules during the 1999-2000 school year summative evaluation of *Astronomy Village®*. We labeled that year as Year 1. The three subsequent school years are labeled Year 2, Year 3, and Year 4 respectively. During Year 1 K followed the implementation guidelines of the summative evaluation. She spent approximately four weeks having students conduct as many of the modules as they could during that time period. In most cases the students completed the core module and 2-3 project modules related to the core topic. All of the students conducted the same modules. In subsequent years, K was free to adapt the program to fit with her curriculum. At the end of each implementation, we conducted interviews with K to discuss the changes made to the implementation and to gauge ideas for how she would change the implementation the subsequent year. The initial phone interview was audio taped and transcribed. Since subsequent interviews dealt with adjustments to the curriculum, they were documented through field notes. The following section details the progression of the implementations of Year 2 to Year 4.

### *5.1 Teacher K*

K's classroom has a lecture area with lab tables and chairs facing the front. Behind the lecture area is a lab area with two freestanding lab benches and lab space along the walls of the classroom. K has access to eight Macintosh iBook portable laptop computers. In Years 1 to 3, these laptops were shared with the social studies department. In Year 4, the social studies department secured their own computers so K's classroom no longer shared the computers. Keeping the computers in the science classroom has cut down on the number of computer-related problems. In addition, K feels that over the years students have gotten better at solving computer-related problems. With the combination of increased student proficiency and no longer sharing the computers, K has found she can spend more time interacting with students around content rather than computer problems.

In the Year 2 implementation of the program, K felt the students needed more guidance compared to students in the Year 1 implementation. She wanted to add more structure for these students. She accomplished this in three main ways:

> "There was a lot more guidance in terms of structure and telling the kids when they needed to move on and keeping them on task. Also, we had assigned reading in the book. This time I had it all selected for them so they knew specifically what pages and what parts of the two textbooks we were using were related to the program as well as important for understanding astronomy and Earth science. As well as, we had a special exhibition at [the local museum] called Extreme Science, which tied in perfectly with what we were doing in *Astronomy Village*®. So, I think those three factors—the extra guidance, the assigned reading in the textbook, and the Extreme Science exhibit having tied together so well—accounted for their better scores." — K phone interview

In the Year 3 and 4 implementations, K maintained the structured approach to the program, although there was no Extreme Science exhibit in those years.

K assigned students to three-member teams. In some of her classes, she had one four-member team. The team composition was based on student performance on projects completed earlier in the year. The goal was to make the groups heterogeneous in terms of abilities. This approach is what Cohen (1994) would call grouping for multiple abilities. This style of grouping when combined with a task like *Astronomy Village* that requires multiple abilities leads to the greatest level of student interaction.

K defined roles for the team members. (1) The team leader's role was to guide the team and serve as the liaison to the teacher. It was the team leader's job to schedule the necessary equipment for hands-on experiments. (2) The team recorder was responsible for completing the written materials for the investigations and keeping the team log. (3) The team navigator was responsible for operating the computer. The students assigned the roles within a group by writing rationales to each other, which were meant to increase team commitment. The teams had a little less than a week to get organized. The roles remained the same throughout the investigation. K noticed a great deal of substantive discussion within the teams throughout the investigation. The team members resolved any team conflicts.

Across the Year 2 through Year 4 implementations, K noticed that students had gotten better at group work in general. Starting in Year 2, K began having students grade each other on their group performance. She had gotten the idea from another group project that occurs later in the school year after *Astronomy Village*. The group grade is a small percentage, but it is still important.

 

It provides greater accountability among the team members. In some cases, K will adjust the group grade if she feels a student has been treated unfairly.

In the Years 2 and 3 implementations, K's students spent between six and seven weeks on the Search for Life investigation. In Year 4, students spent between seven and eight weeks. All teams began with the core research investigation. K had students conduct the investigation in a self-paced manner. When they completed the core research investigation, students could move on to a focused investigation. The students took about two weeks to complete the core research investigation. After completing the core research investigation, each team picked a focused investigation to complete. Each class had eight teams. Therefore, the selection of focused investigations was somewhat guided so that there were two groups conducting each of the four focused investigations. K gave the students just over one week to finish the focused investigation. During the Year 4 implementation, K inserted a modeling activity between the core and focused investigations. The modeling activity dealt with extra solar planets and extended the length of the Year 4 implementation by one week.

During the investigation, students maintained a paper-based log. In the log students stored copies of the printed articles from *Astronomy Village*®, notes on the investigation, and completed worksheets. In Years 3 and 4, K recycled the printouts of articles in order to save paper. The articles contained highlighting and notes from the previous year's implementation. In addition to the activities in *Astronomy Village*, K required students to read relevant sections from their textbook and write definitions of key words.

Each team leader had to inform K when they needed materials for the hands-on experiments. During the Year 3 and 4 implementations, K substituted her own hands on experiments for several of the hands on experiments in *Astronomy Village*. These substitutions included experiments on electromagnetic spectrum, condensation/dewpoint, and relative humidity. K also altered one of the experiments on solvents for life by having students conduct the experiments in baggies rather than beakers. Using baggies allowed students to feel the temperature change.

During the investigations K saw her role as helping when needed, providing necessary equipment, and making sure students stay on track.

> "I think I was a 'gopher.' If they needed something, I went to get it. I made sure they had the materials they needed. If they ran into glitches with the computer, I helped them through that … But other than that, I was making sure that they were on task and they knew what they were doing. If they got into a bind and they needed something, then they could come to me." — K phone interview

At the end of the focused investigation period, each team gave a presentation to the class on its investigation. Before the presentation K highlighted sections of each team's logbook and handouts, indicating what important topics should be covered during the presentation.

> "I highlighted sections that they should talk about. In other words, I just took the printout of the logbook, I highlighted that, and I said make sure you cover this for the other students. Then the other students could raise questions if they didn't understand.

 

> For example, if they were talking about the ‘It's Just a Phase’ library article and they couldn't explain it adequately, we had them put it up on the TV monitor to direct the students where on their own computer they could go and look at it later." — K phone interview

After each presentation the students could ask questions. The audience members had a vested interest in the presentations since they were responsible for understanding the content in each investigation. After the presentations were completed, students had four extra days to review material in preparation for the test. The posttest score for Search for Life was used as part of the course grade. The test was open notes, but the students had to use their own notes. In addition to the test grade, students were also graded on their group performance. The overall course grade also included other areas of study, such as an individual science project.

## 6 Results

Table 1 shows means of the pretest, posttest, and gain scores for the control group as well as each implementation year. An ANOVA with interactions of the scores by year and by pre vs. post reveals that there was a statistically significant increase from pretest to posttest for each year ($F(4, 729) = 48.69$, $p < 0.001$). In addition, an ANOVA of gain scores by year was statistically significant ($F(4, 364) = 110.67$, $p < 0.001$). The posttest scores and the gain scores for Year 1 were statistically lower than the posttest and gain scores for Years 2, 3, and 4. There is no statistically significant difference between Years 2, 3, and 4 on the gain score. It seems that the adjustments made from the summative evaluation year to Year 2 had a significant effect, but the adjustments made between Years 2, 3, and 4 had little effect on statistically increasing learning outcomes.

**Table 1: Pretest and Posttest Treatment Effects**

| | Pretest | Posttest | Gain | N |
|---|---|---|---|---|
| **Control** | 50% | 53% | 3% | 48 |
| | | | | |
| **Year 1** | 53% | 72% | 19% | 64 |
| **Year 2** | 55% | 82% | 27% | 119 |
| **Year 3** | 57% | 84% | 27% | 78 |
| **Year 4** | 60% | 86% | 26% | 56 |

### *6.1 Conclusion*

The results of this investigation reveal important tradeoffs between scientifically based research and design experiments. The Year 1 implementation followed the scientifically based research methodology. The comparative nature of the evaluation allows us to make inferences about the impact of the program on student learning outcomes. In order to foster a comparison across all of the participating teachers it was necessary to place constraints on teachers' implementation of the program. The implementations reflect what the designers believed to be important in general. This approach neglects considerations of the local context.

Typical of the design experiment approach, K attempted to maximize the benefit of the program for her students during the Year 2 through Year 4 implementations. She changed a number of factors across all four years. For example, the students spent more time on the investigations after the first year. The basic curriculum structure was altered significantly between the first and second implementation years and adjusted slightly between the second, third, and fourth implementation years. The design experiment approach does not allow for analyses that tease apart the effects due to any one factor. However, it is a beneficial approach for systematically adapting the program to the local context.

Combining scientifically based research with design experiments provides an important lens for tracking long-term implementation of innovations. The critical assumption underlying this combination is that long-term reform involves much more than the design and evaluation of curriculum materials. Not enough is known about how innovations evolve over the course of time in the face of strong market forces. The track record for innovations in science education reform has been one of early success, followed by gradual obsolescence as the original designers eventually fade away from the project (McGee, 1996). A long-term perspective will provide developers and reformers with a better understanding of how to achieve long-term success for new innovations. The combination of scientific evidence from scientifically based research and professional judgment from design experiments provides the best hope for innovations to have lasting impact.

## 7 References

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Ed.). (1999). *How people learn: Brain, mind, experience, and school*. Washington, DC: National Research Council.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *The Journal of the Learning Sciences, 2*(2), 141-178.

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher 32(*1), 9-13.

Cohen, E. G. (1994). Restructuring the classroom: Conditions for productive small groups. *Review of Educational Research, 64*(1), 1-15.

Dimitrov, D. M., McGee, S., & Howard, B. C. (2002). Changes in students' science ability produced by multimedia learning environments: Application of the Linear Logistic Model for Change. *School Science and Mathematics Journal, 102*(1), 15-22.

Hickey, D. T., Wolfe, E. W., & Kindfield, A. C. H. (1999). Assessing learning in a technology-supported genetics environment: Evidential and systemic validity issues. *Educational Assessment, 6*(3), 155-196.

McGee, S. (1996). *Designing curricula based on science communities of practice*. Unpublished doctoral dissertation, Northwestern University, Evanston, IL. [Online] http://www.cet.edu/research/papers.html

 

McGee, S. & Howard, B. C. (1998). Evaluating multimedia in the context of use. *Journal of Universal Computer Science*, *4*(3), 273-291.

McGee, S., & Howard, B. C. (1999, April). Generalizing activity structures from high school to middle school science. Symposium presented at the meeting of the American Educational Research Association, Montreal. [Online] http://www.cet.edu/research/papers.html

McGee, S., Howard, B. C., Dimitrov, D. M., Hong, N. S., & Shia, R. (2001, April). Addressing the complexities of evaluating interdisciplinary multimedia learning environments. Presented at the annual meeting of the American Educational Research Association, Seattle. [Online] http://www.cet.edu/research/papers.html

McGee, S., Dimitrov, D. M., Kirby, J., & Croft, S. K. (2002, April). Using design experiments to investigate long-term program success. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. [Online] http://www.cet.edu/research/papers.html

Sheppard, L. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.

Shin, N., Jonassen, H. D., & McGee, S. (in press). Predictors of well-structured and ill-structured problem solving in an astronomy simulation. *Journal of Research in Science Teaching*.

 