Three-Year Study of Astronomy Village® Implementation

Steven McGee[1], Dimiter M Dimitrov[2], Jennifer Kirby[1], Steven K. Croft[1]

[1]Center for Educational Technologies®
Wheeling Jesuit University

[2]Kent State University

Abstract

The purpose of this study was to track teachers using Astronomy Village® over three school years. Prior research indicated that teachers' adjustments form the first implementation year to the second implementation year resulted in significantly higher learning outcomes. In this study we sought to determine whether this trend would continue. We used a hybrid experimental design that combined quasi-experimentation and design experiments. The results indicate that teachers made slight adjustments to their implementation from the second to the third year. An analysis of the learning outcomes indicates that students in the third implementation year achieved about the same learning outcomes as students in the second implementation year.

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 summative 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 not only significantly improved in their understanding of complex solar system concepts, they also improved in their ability to engage in inquiry on the solar system (McGee et. al., 2001). McGee et. al. (2002) also conducted a case study of teachers who used Astronomy Village for a second consecutive year after the direct support of the program developers had terminated. The results indicated that student performance was greater in the second year. In this study we extended the case study to include teachers who used Astronomy Village during three consecutive school years. The goal of this study is to see whether teachers continued to achieve greater student performance in the third year or whether student performance reached a plateau.

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 randomized experiments 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.

In addition to the comparison from one year to the next, which is a hallmark of the design experiment approach, we have integrated a quasi-experiment design approach. Each participating teacher recruited a matched no treatment control group for the first implementation year. The control students were administered the same pre- and posttest measures at the same times as the Astronomy Village students. Through the design experiment approach teachers are free to make principled changes to their instruction in an attempt to improve instruction. Using qualitative analysis techniques, we characterized the nature of the changes from one year to the next and from one classroom to the next. With the quasi-experimental approach we are able to remove any nontreatment effects within a given implementation, such as maturation or history effects. Once the nontreatment effects are removed, it is reasonable to conclude that any remaining improvement is because of the teacher's implementation.

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

We identified two teachers who conducted the Search for Life core research modules during the 1999-2000 school year summative evaluation of Astronomy Village (year 1) and who conducted the Search for Life modules during the 2000-2001 (year 2) and the 2001-2002 (year 3) school years. During year 1 teachers were asked to spend 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 years 2 and 3 the teachers were free to implement the program in a way that fit their curriculum.

At the end of the year 2 and 3 implementations, we conducted phone interviews with each of the teachers to discuss the changes they made to the implementation and to gauge their ideas for how they would change the implementation next year. The phone interviews were audiotaped and transcribed. The following sections describe the years 2 and 3 implementations for the two case study teachers—teacher K and teachers J/D.

5.1 Teacher K

In the year 2 implementation of the program, K felt students needed more guidance compared to the year 1 implementation. She wanted to add more structure for these students. She accomplished this 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 Bishop 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

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 different 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 had 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 labs. (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.

K's students spent between six and seven weeks on the Search for Life investigation. 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. Each team leader had to inform K when they needed materials for the hands-on experiments. During the investigations K saw her role as helping when needed, providing needed equipment, and making sure students stayed 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 paying attention to 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 the individual science project.

For the year 3 implementation K maintained the same basic structure but integrated more of the successful lab activities she has used over the years. In the future she would like the students to have time to conduct the other Astronomy Village investigation, Mission to Pluto.

5.2 Teachers J/D

Astronomy Village was taught in the context of a connections course that combines science and language arts. The specific science class is called Searching for Extraterrestrial Life Intelligently.

> "Actually, connections is a category of classes that our seventh- and eighth-graders take. They are not just science; they are a combination of language arts in a subject matter. J and I both happen to be science teachers, so we do the science and language arts together." — D phone interview

In implementation year 3 J/D did not feel that students did as well as they did in year 2 for two reasons. They felt the students were rushed. They also felt the students didn't like some of their teams. J/D had assigned four-member teams at random. In contrast, the students had input on the team composition in year 2. J/D felt that random assignment did not work well and led to team problems.

The class spent between five and six weeks conducting the search for life investigation. The class did the core research together in a little more than a week. J/D were fortunate that they had four undergraduate teaching practicum students in their classes while they were conducting the search for life investigation. Each practicum student oversaw one of the four focused investigations. Each group spent a week on a focused investigation and then rotated to the next station. Although some groups needed more time on the investigation, the timeline had to be enforced so that all teams would be rotating at the same time. The teachers felt the students were rushed and that one week was too short. In year 2 J/D used videos and class discussions more, but in year 3 the main source of information was the CD. They thought the students grew tired of the CD in year 3.

For the course grade J/D graded the logbook. Each section of the notebook was graded out of 100 points. The grade was assigned to the team as a whole. Some students weren't happy with their team and were frustrated with the team grade. These students were able to continue to work on the logbook to improve their grade. J/D commented that some of the labs in Astronomy Village could be used for the students' science portfolios. At the time the state had plans to implement student portfolios as part of the state accountability system. At the time of year 3 implementation, they were not yet required. However, the teachers in the state were getting ready for the portfolios to be implemented.

J/D saw their role as facilitating and keeping students on track. During the focused investigations they were free to roam the classroom and observe how the practicum students were interacting with the middle school students. Each student teacher played a slightly different role in terms of the directedness. For the most part the students followed the investigation cycle diagram. J/D felt that most students were able to follow the investigation cycle diagram.

J/D were not satisfied with their year 3 implementation. They discussed the idea of having students focus on one focused investigation and becoming an expert in that area. Then students would host a live press conference with their peers. The approach they are leaning toward is similar to the approach that K uses.

6 Results

Table 1 shows means of the pretest, posttest, and gain scores for each implementation year by teacher. An ANOVA with interactions of teacher by year by pre vs. post reveals that there was a statistically significant increase from pretest to posttest for each year for both K and J/D. In addition, the posttest scores and the gain scores for year 1 were statistically lower than the posttest and gain scores for years 2 and 3 for each teacher. There is no statistically significant difference between years 2 and 3 on either posttest score or the gain score for both teachers. It seems that the adjustments that teachers made from year 1 to year 2 had a significant effect, but the adjustments made from year 2 to year 3 had little effect on increasing learning outcomes. Year-by-year comparisons of gain scores by teacher shows that K's students had statistically larger gain scores than J/D's students.

Table 1: Pretest and Posttest Treatment Effects

Teacher	Year	Pretest	Posttest	Gain	N
K	1	50.3%	69.8%	19.5%	125
	2	54.5%	82.0%	27.4%	123
	3	54.7%	83.1%	28.4%	118
J/D	1	55.8%	59.9%	4.0%	31
	2	57.5%	65.8%	8.3%	57
	3	61.2%	70.1%	8.9%	47
Control		41.6%	43.2%	1.5%	387

To better support the analyses of the quasi-experimental design, we adopted the Linear Logistic Model for Change (LLMC) technique to analyze pretest to posttest changes in learning outcomes. Benefits of the LLMC include information about the magnitude of the changes on a ratio scale, which facilitates design experiment analyses and separation of changes because of treatment from changes due to natural trends across time points of measurements (e.g., pretest and posttest), which facilitates quasi-experimental analyses. The theoretical framework of the LLMC is not presented here because of its relative complexity and prerequisites of psychometric background for the reader (see, e.g., Fischer, 1995; Fischer & Ponocny-Seliger, 1998). Provided are basic concepts and interpretations of LLMC results reported in this study.

In the item response theory context, the term ability connotes a latent trait that underlies the student's performance on a test (e.g., Hambleton, Swaminathan, & Rogers, 1991). The ability score of a student relates to the probability for this student to answer any test item correctly. The units of the ability scale, called "logits," typically range from -4 to +4. The results of the analyses are subdivided into <u>treatment effect</u> and <u>trend effect</u> that measure ability changes due to treatment and natural trends, respectively. The trend effect accounts for factors such as natural biological maturation and cognitive development over the period of time between pretest and posttest measurements. The ratio of two effects indicates <u>how many times</u> one of the effects is greater (or smaller) than the other effect. In this study the LLMC calculations were performed using the computer program LPCM-WIN 1.0 (Fisher & Ponocny-Seliger, 1998).

Table 2: LLMC Pretest to Posttest Treatment Effects

Year	Gain Score (logit)
1	0.93
2	1.40
3	1.39

Table 2 shows the results of the LLMC analyses. For these analyses the results for both Astronomy Village teachers were combined. The results of the LLMC analysis are consistent with the results of the ANOVA. After controlling for the effects due to the control, the learning gains for each year were statistically significant. The gain scores for year 1 were statistically smaller than the gain scores for both years 2 and 3. There was no statistical difference of the gain scores between years 2 and 3.

6.1 Conclusion

The primary goal for this study was to examine student performance from three consecutive years of implementation of Astronomy Village. There seemed to be a large increase in student performance between the first and second year, whereas performance between the second and third implementation year seems to have reached a plateau.

Typical of the design experiment approach, these teachers changed a number of factors between the three implementation years. There were different students. The students spent more time on the investigations in the second and third years. The basic curriculum structure was altered significantly between the first and second implementation years and adjusted slightly between the second and third implementation years. The same test was used from one year to the next, so the teachers were more familiar with the specific questions asked on the test. (Prior research has indicated that as teachers become familiar with a test, their teaching gradually emphasizes the content on that specific test.) It is our belief that all of these factors played a role in the improved learning outcomes. The design experiment approach does not allow for analyses that tease apart the effects due to any one factor.

This research on the implementation of a specific program is in its nascent stage. The critical assumption underlying this research is that long-term reform involves much more than the design 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.

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.

Fischer, G. H.(1995). Linear logistic models for change. In G. H. Fischer and I.W. Molenaar (Eds.), *Rasch models. Foundations, recent developments, and applications* (pp. 157-180). New York: Springer-Verlag.

Fischer, G. H., & Ponocny-Seliger, E. (1998). *Structural Rasch modeling. Handbook of the usage of LPCM-WIN 1.0*. Netherlands: ProGamma.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: SAGE.

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 illstructured problem solving in an astronomy simulation. *Journal of Research in Science Teaching*.