Comparing the Efficacy of Reform-Based and Traditional/Verification Curricula to Support Student Learning about Space Science

Ellen M. Granger,1 Todd H. Bevis,1 Yavuz Saka,2 and Sherry A. Southerland2

1Florida State University, Biological Science, 319 Stadium Drive, 1062 King Life Sciences Bldg., Tallahassee, Florida 32306-4295, USA
2Florida State University, College of Education, 215 Stone Building, Tallahassee, Florida 32306-4490, USA

Abstract. This research explores the relationship between reform-based curriculum and the development of students’ knowledge of and attitudes toward space science. Using a randomized cluster design, the effectiveness of Great Exploration in Math and Science (GEMS) Space Science Curriculum Sequence was compared with the effectiveness of more traditional curriculum in supporting 4th and 5th grade students’ learning of and attitudes toward space science. GEMS employed an inductive approach to content (learning cycle), explicit use of evidence, and attention to scientific inquiry. The comparison group experienced traditional, verification means of teaching. Randomization occurred at the level of the teacher assignment to treatment group (not at the student level). Students in the classrooms in which GEMS was employed demonstrated a statistically significant increase in content knowledge and attitudes toward space science: Students in classrooms in which the traditional curriculum was employed did not show these increases. The GEMS effect on student achievement was greater for students in classrooms in which the teacher experienced a greater increase in content knowledge.

1. Introduction

This research explores the influence of the enactment of reform-based curricula in science on fourth- and fifth-grade students’ learning of space science content. Using a randomized cluster design, the effectiveness of the Great Exploration in Math and Science (GEMS) Space Science Curriculum Sequence (Lawrence Hall of Science, 2007) for grades 3–5 was compared to the effectiveness of more traditional curriculum in supporting student content knowledge and attitudes about science. GEMS was chosen for this study because it employs an inductive approach to teaching and embodies some of the qualities of the reform-based approach to science teaching as described in national reform documents (e.g., NRC, 1996; 2000; 2008; Duschl, Schweingruber, & Shouse, 2007. Furthermore, the space science content, science process, and nature of science content embodied in this curriculum are aligned with the national standards and the science standards of the state in which the study occurred.

Previous studies on the effectiveness of inquiry- and reform-based instruction in science have produced mixed results, with some studies finding reform-based instruction to be superior to traditional instruction, some finding no difference, and some finding reform-based instruction to be inferior (e.g., Blanchard, et al., in review; Colburn,
2000; Dean, et al., 2006; Hall et al., 1990; Klahr, et al., 2004; Kirschner et al., 2006; Lederman et al., 2007; Leonard, 1983; Leonard et al., 1981; Lynch, et al., 2006; Marx, et al., 2004; Shymansky et al, 2004). Importantly, not only are the results of these and other studies mixed, but the research methodologies employed by the studies also are mixed.

There are many factors that may be contributing to the mixed nature of the results of these studies. As those who have attempted randomized control trials in educational research know, it is difficult, if not impossible to tightly control for every possible variable. Another important potential factor is that the comparison condition (i.e., the control group) is “too often overlooked in educational research. Frequently the comparison condition is undefined or assumed to be ‘traditional’ NRC, 2004” (Lynch et al., 2006). Likewise, Shymansky et al. (2004) discuss possible “contamination of the untreated [comparison group] teachers” and lack of attempt to “vigorously guard” against special resource materials as possibly influencing their own results. In fact, many, perhaps most, studies reported to date do not discuss how fidelity to the curriculum and/or instructional methodology was measured, if it was assessed at all.

This study follows others that investigate the effectiveness of reform-minded practices—in this case as embodied by a curriculum project in combination with teacher professional development—to foster student science learning. We report here some preliminary results from the data collected during the first year (cohort one) of a two year (two cohort) study to begin to systematically address two student outcome questions surrounding enactment of reform-based curriculum following professional development: Does the use of such reform-based curricula in combination with professional development work, that is, does it support student science learning (both cognitively and affectively)? How do teacher characteristics influence the learning their students experience as a result of the use of such curricula and professional development, that is, in this case what teacher moderating variables can we identify from the data collected?

2. Methods and Data Sources

This study employed a randomized cluster design. The experimental treatment group was comprised of 32 teachers who had participated in a four-day professional development experience focusing on the GEMS Space Science Curriculum Sequence and the pedagogy underpinning it (i.e., learning cycle, evidence circles, cooperative learning, discussion techniques, space science misconceptions, guided inquiry) and who then enacted this curriculum in their classrooms. At the same time, the control group was comprised of 29 science teachers (3 dropped out due to personal or experiment non-compliance issues) who used the district adopted text to address the same space science content, science process, and nature of science standards through traditional, transmission mode approaches to instruction (lecture, reading from the text, and “hands-on” activities that are related to the topic but do not extend the depth of the student learning about core concepts nor address misconceptions). The learning of space science concepts, inquiry skills, and affective dimensions (pre, post, and delayed post) of the teachers and their students were compared across groups. Randomization occurred at the level of the teacher assignment in which teacher volunteers were randomly assigned to treatment or control group, with control/experimental group matching according to grade level, SES, school grade, and ethnic diversity based on their students’ demographics. There was a total of 1178 students—696 from classes of the GEMS group teachers
The experimental treatment group curriculum was the GEMS Space Science Curriculum Sequence (GEMS SSSCS) (2007) for teaching space science concepts for grades 3 through 5. This curriculum was easily implementable by teachers inexperienced with reform methodologies given its detailed materials preparation and classroom enactment instructions including time frames for preparation and teaching, scripting of lessons and discussions, embedded formative and summative assessments, etc. The GEMS SSSCS is designed to address age-appropriate core concepts in space science (NSES, 1996) and common misconceptions that students harbor about them (Kavanagh et al., 2007a,b). In addition, the curriculum has an explicit focus on the role of models and evidence in science. Students are encouraged throughout the unit to evaluate alternative explanations, to use evidence to support explanations, and to critique the merits of an explanation in a scaffolded, age-appropriate way. In general, all lessons are structured around a learning cycle format (e.g., Bybee, 1997). Thus, in many ways, the GEMS SSSCS reflects the science education reforms emphasized above (Duschl et al., 2008; NRC, 1996, 2000, 2008). Teachers in the treatment group were further instructed by the research team to adjust their normal classroom practice to closely follow the instructions described in the curriculum.

The district adopted science text for grades 4 and 5 served as the basis of the control group classroom instruction. The district curriculum was centered on more didactic presentation of space science concepts including direct instruction, reading of text, students answering very focused questions. The activities included in the text served as a verification of the content already presented in the text or, more commonly, activities were peripherally associated with the topic but did not address the core concepts. Control teachers were further instructed by the research team to adjust their normal classroom practice, if different from this model, in order to adhere to this traditional teaching format and to the text’s presentation for the space science unit.

The structure of the experiment in which they were participating (i.e., randomized-cluster, treatment/control-group design) was discussed with both groups of teachers. The importance of such studies to their profession was stressed as was the importance of their contributions to their profession through their participation. In this context, teachers in both groups were instructed to adhere to the assigned curriculum and instructional methodology and to refrain from adding any additional activities to those present in their assigned curriculum. Teachers from GEMS treatment and control groups in the same school (and this was done for a matched control whenever possible) were instructed not to discuss their curriculum with the other group until after the administration of the delayed post testing five months after the teaching of the unit. Teacher fidelity to the assigned methodology was assessed through direct observation and/or videotaped observation at least twice during the unit using the RTOP instrument (Sawada et al., 2002). (Analysis of all videotapes for cohort one is not yet complete at this time.)

The GEMS treatment group was involved in professional development in which the teachers experienced the curriculum as learners, then learned about the pedagogies that underpin it (learning cycle approach to science instruction, questioning/discussion strategies, evidence circles, assessment strategies, nature of science teaching strategies, etc.) through an explicit/reflective experiential approach two weeks before the beginning of school, a three-hour follow up immediately prior to the beginning of the teach-
ing of the space science unit, and a three-hour session midway through the teaching of the unit to discuss questions that had arisen. Teachers in the control group participated in a meeting 3 weeks prior to the teaching of the control space science unit to review their part in the project, to discuss the need for employing traditional teaching approaches through fidelity to the district-adopted textbook curriculum and the traditional approach to teaching asked of them (see above), and to complete the teacher pre-assessments.

Four instruments were used to assess treatment and control group student conceptual development around space science, models, and scientific inquiry, and affective dimensions (pre-, post-, and delayed post-instruction). These instruments were: 1) Space science content test (from Sadler et al., 2007), 2) The Homerton Science Attitudes survey (Warrington, Younger, & Williams, 2000), 3) Models and Evidence survey (Granger, Saka, & Bevis, 2007), and 4) Views of Scientific Inquiry–Elementary (VOSI-E, Schwartz et al., 2008). Each assessment was administered to students prior to space-science unit instruction, immediately following completion of teaching the space science unit, and 5 months ± 2 weeks following the completion of the teaching of the space science unit. This paper reports results from the first two instruments, the only ones for which analysis was complete at the time of submission.

When estimating the GEMS effects on the student outcomes, hierarchical linear modeling (HLM) was required to account for the interdependencies of student outcomes within teachers (e.g., Raudenbush et al., 2002). All computations were accomplished with HLM 6 software (Raudenbush et al., 2004). For initial estimation of the GEMS main effects for each of the student outcomes of interest a two-level model was assumed. The single student-level independent variable was the student pre-measure of the outcome variable and the teacher-level variables were GEMS/Control group, teacher experience, grade, and grade by GEMS product term. Possible additional teacher-level pre-measures were identified with the Exploratory Analysis option in the HLM 6 software. In every case, the teacher experience, grade and grade by GEMS variables were deleted due to non-significance. Given the different scales of the outcomes of interest, standardized effects were included in the results in order to allow direct comparison of GEMS effect sizes for the various outcomes. These were obtained by dividing the estimated raw score GEMS effect by the standard deviation of the outcome variable. The strengths of the effects were characterized based on the following definitions for standardized effects (Cohen, 1977): 0.2 is “small,” 0.5 is “medium,” and 0.8 is “large.” The results herein contain many statistical hypothesis tests, resulting in substantial inflation of family-wise error rate. Since this study is exploratory, there was no attempt to control family-wise error.

3. Results

The results reported herein focus on some of the data collected from the student portion of the larger study. Nevertheless, it is important to situate these results in the context of the overall teacher outcomes of the study (Granger et al., 2009). Briefly, teacher achievement on the post content test (questions from Sadler et al., 2007) for the GEMS group teachers compared to the control group teachers was positive and statistically significant (p < 0.001).

When the focus is turned to student learning, the results indicate that the effects of the GEMS curriculum plus professional development on student outcomes were posi-
tive for students in the GEMS group compared with those for the control group. Table 1 shows that the estimated main effects of GEMS on the post tests for achievement (space science content) and attitude (Homerton survey) were positive and statistically significant in the GEMS treatment group ($p = 0.004$ and $p = 0.067$ respectively). The delayed-post test for both science content and attitude 5 months after instruction were positive, but not statistically significant ($p = 0.116$ and $p = 0.239$, respectively).

Table 1. GEMS Main Effects for Student Outcomes

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Unstandardized GEMS Effect</th>
<th>Standard Error</th>
<th>p-value</th>
<th>Standardized GEMS Effect$^a$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post Measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Achievement</td>
<td>0.629$^*$</td>
<td>0.206</td>
<td>0.004</td>
<td>0.22$^*$</td>
</tr>
<tr>
<td>Attitude</td>
<td>1.862$^*$</td>
<td>0.995</td>
<td>0.067</td>
<td>0.12$^*$</td>
</tr>
<tr>
<td>Delayed Post Measures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Achievement</td>
<td>0.340</td>
<td>0.213</td>
<td>0.116</td>
<td>0.12</td>
</tr>
<tr>
<td>Attitude</td>
<td>1.282</td>
<td>1.075</td>
<td>0.239</td>
<td>0.08</td>
</tr>
</tbody>
</table>

$^a$Statistically significant at the 0.10 level.

$^*$Standardized effects were obtained by dividing the raw score GEMS coefficient by the outcome standard deviation.

Thus, it was concluded that the estimated outcomes for the GEMS group were greater than those for the control group, adjusting for any initial differences on the other variables. For example, it was estimated that the post test measure of student Achievement for the GEMS group was 0.629 higher than that for the control group, controlling for other variables. Considering the standardized effects, the magnitude of the effect on Achievement is characterized as “small” by Cohen’s rule, while that for Attitude is about half that size. Nevertheless, in large scale studies these effects are not trivial. The estimated magnitudes of the standardized forms of the delayed post measures were roughly half of the corresponding post test measures.

An exploratory analysis searching for student-level moderators of the GEMS effects was conducted by adding the GEMS variable to the model for the slope of the student-level pre-measure. There was no statistical support for any of the student-level moderators. Another exploratory search, this time for teacher-level moderators, was accomplished by adding product terms involving the corresponding teacher-level pre-test measures. This search resulted in evidence for one moderator: teacher pre-Achievement was a moderator for student post test Achievement. Figure 1 shows the standardized form of two models obtained by dividing by the standard deviation of post test Achievement. It is seen from Figure 1 that the GEMS group is estimated to be superior to the control group for teacher content pre-test score values up to about 9. For higher values of the moderator, the direction of the difference reverses, with the control group appearing to be superior.

A 90% confidence band indicates that the GEMS effect estimated was statistically significant for values of teacher Pre-Achievement up to approximately 6.30. For higher moderator values these effects were not statistically significant. In the region of significance, the largest standardized GEMS effects are roughly 0.5, a magnitude judged here to be medium in strength.
Thus, for the student post test outcomes, the effect of GEMS on student space science content knowledge was positive and strongest for students in the classrooms of teachers with limited content knowledge of space science at the outset of the study (i.e., on their pre test); the effects decreased and then disappeared for students in the classrooms of teachers with greater content knowledge of space science at the outset of the study (i.e., on their pre test). There was no evidence of an interaction involving the corresponding teacher knowledge pre-instruction measure on student attitude results.

In summary, changes in student achievement and attitude for the GEMS group compared to the control group were positive and statistically significant for the post tests and in the positive direction but not statistically significant for the delayed-post tests (5 months ± 2 weeks). There was no support for student level moderators of the GEMS effect (at least for the student variables we were able to test prior to this submission). One teacher level moderator, teacher achievement on the (space science) content pre test for teachers, was supported as a moderator of the GEMS student post achievement (space science content test) effect.

4. Discussion

It is important to emphasize at the outset of this discussion that this is a preliminary analysis of the results from the first year of a two year study.

Empirical analyses and subsequent discussions of effectiveness of reform-based instruction in science are inconclusive in that the literature consists of mixed results (e.g., Blanchard et al., in review; Colburn, 2000; Dean et al., 2006; Hall et al., 1990; Kirschner et al., 2006; Klahr et al., 2004; Lederman et al., 2007; Leonard, 1983; Leonard et al., 1981; Lynch et al., 2006; Marx et al., 2004; Shymansky et al, 2004). To
understand the nature of these mixed findings, there is a need for research that attempts to more tightly control possible confounding variables so that a clearer portrait of the influence of reform-based instruction on student learning can be drawn. This research was such a systematic examination.

In answer to our central question, our work indicates that well-designed, reform-based curricula in combination with teacher professional development had positive and statistically significant influences on student achievement and student attitudes about science as a result of instruction. While the positive direction of this influence for both student content knowledge and interest in space science remained at 5 months post instruction, they were no longer statistically significant at that time.

In answer to our secondary question, “How do teacher characteristics influence the learning that their students experience as a result of the use of reform-based curricula and professional development?”, we found that one teacher level variable, teacher content knowledge (of space science) at the outset of the unit was a moderator of the GEMS student post-test achievement effect. That is, the students in the classrooms of teachers who were low on their content knowledge test score before beginning the study showed a significant increase in their post content test scores. However, this effect was statistically significant only for teachers who initially were categorized as low on their content pre-test score (Figs. 1).

What could be the explanation for this outcome? We know from our analysis of the teacher results of this study that teacher achievement on their post content tests and teacher confidence in their content knowledge was positive and statistically significant for the GEMS treatment group compared to the control group (Granger et al., 2009). Perhaps those teachers who had just learned the space science concepts during the professional development were better able to communicate with their students at the level of a novice learner. That is, they were much closer to being a novice learner in the content and so had recent personal experiences to draw upon to help their students learn these concepts in ways that they had just learned them. Another suggestion is that perhaps these space science “novice” teachers were more willing to rely heavily upon the curriculum’s scripting to help them lead the lessons and discussions. Or perhaps they had an increased interest in space science, given the recent nature of their learning and this evidenced itself in their teaching, and thus in motivating their students. Whatever the reason, these results indicate that when teachers employed reform-based curricula as embodied in the GEMS SSCS in their classrooms following professional development the space science content learning of their students was enhanced.

How then do we explain our results in light of the mixed results of previous studies? Certainly the GEMS SSCS is a well developed, well designed, and teacher “friendly” set of materials. Thus, part of the reason may be found in the strength of the curriculum. Further, our research design tightly controlled the curriculum and instructional practice employed not only of the treatment group, but also of the comparison group. That is, for both the experimental treatment and control groups the curriculum and instructional methodology was specified. Indeed, the need for fidelity to the assigned curriculum and instructional methodology in a controlled study was discussed in advance with both groups of teachers (experimental and control). Their role in this fidelity was identified to them explicitly as an important contribution that they were making to the profession of teaching. Fidelity to the curriculum and instructional methodology was monitored through observations and/or videotapings. As fidelity to
the instructional methodology and curricula has commonly not been monitored in many, if not most, previous studies, it may explain the aforementioned mixed results.

**Acknowledgments.** We are deeply indebted to Dr. Richard Tate, our statistician, for his talent, help, instruction, and advice with the statistical analyses in this paper and to our research assistants, Francesca Gerbino and Katrina Dial for their assistance with data management. We wish to acknowledge the significant contribution of our science coach, Tracie Ragland, to the project. Finally, the project is indebted to the teachers who agreed to constrain their teaching to the assigned methodology and curriculum in order to further knowledge about their profession.

This research was supported by a grant from the Florida Center for Research in Science, Technology, Engineering and Mathematics education, (FL DOE # 371-96700-7SF01, 371-96700-8SS01, 371-96700-9SF01).

**References**


Hall, D. A., & McCurdy, D. W. 1990, A Comparison of a Biological Sciences Curriculum Study (BSCS) Laboratory and a Traditional Laboratory on Student Achievement at Two Private Liberal Arts Colleges, Journal of Research in Science Teaching, 27, 625–636


HLM 6: Hierarchical Linear and Nonlinear Modeling, (Lincolnwood, IL: SSI Scientific Software International)