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

E.M. Granger and T.H. Bevis
Office of Science Teaching Activities
College of Arts & Sciences
Florida State University

Y. Saka and S.A. Southerland
School of Teacher Education, Science Education
Florida State University

ABSTRACT

This research explores the relationship between reform-based curriculum and the development of students' knowledge of and attitudes toward space science. Using a quasi-experimental 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). The sample included 32 treatment and 29 control teachers working with 1178 4th and 5th grade students. 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. Implications of the study are discussed.

E.M. Granger
granger@bio.fsu.edu

Paper presented at the annual meeting of the National Association for Research in Science Teaching, Garden Grove, CA, April 17-20, 2009.

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

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 quasi-experimental 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 the *National Science Education Standards* (NRC, 1996) and *Inquiry and the National Science Education Standards* (NRC, 2000), *Taking Science to School* (Duschl, Schweingruber, & Shouse, 2007), and *Ready, Set, Science* (NRC, 2008). Additionally, it exemplifies curricula designed to best enable teachers to provide reform-minded instruction for their students, including detailed classroom lesson plans complete with exact enactment instructions, questioning and responding outlines, detailed content background appropriate for elementary-school teachers' understanding, pedagogical professional development notes for the teacher about the instruction (e.g., learning cycle, leading discussions, nature of science, etc.), timeframes, critical junctures, and assessments. 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.

Recent studies on learning science emphasize three key areas critical to student success: (1) understanding the discipline's core concepts, theories, and models; (2) understanding, in general, how scientific knowledge is generated, tested, and accepted; and (3) using the first two to extend understanding into new areas. Research findings suggest that in order for students to come to understand scientific concepts, to be able to apply the concepts, and to understand how science is done, their learning of science should in some ways echo the way science is conducted by scientists (Bybee, 1997; Duschl et al., 2007; Chinn & Maholtra, 2002; Flick, 2003).

The findings of science education research and the central features of national reform efforts have been applied to produce promising, well-designed instructional materials that have undergone years of development, field-testing, and revision (e.g., *Great Explorations in Math and Science, Insights and Outcomes, Full Option Science Systems, Math Connections, Connected Mathematics Project, BSCS).* However, despite extensive efforts to precipitate educational change using such materials, classroom practices remain largely unaffected. One of the reasons for this lack of use of these materials has to do with teachers' reticence to use reform-based methods in the face of accountability pressures (Abrams et al., 2007) as teachers (and their administrators) remain unconvinced that reform based practices (e.g., levels of inquiry, learning cycle, writing to learn) are successful in helping students learn science concepts (Settlage & Blanchard, 2007; Settlage & Meadows, 2002).

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). In examining these and other studies that comprise the research literature, one finds that not only are the results mixed, but also the research methodologies employed by the studies are mixed. Explaining one aspect of this mixed methodology is the form of inquiry employed. Indeed, there does not appear to be shared terminology about what constitutes inquiry as one feature of reform-based instruction, or for that matter what constitutes traditional instruction. Recently, in recognition of the nebulous nature of inquiry (Settlage, 2007), there have been ongoing efforts to systematically describe the kind of inquiry employed (Blanchard, et al., 2009; Abrams et al., 2007) with researchers and teacher educators describing inquiry in terms of Schwab's levels (1, 2, or 3) and the efforts of Colburn (2000) and others.

Studies also vary according to the nature of their research design. Even when examining only those studies employing quasi-randomized control studies, mixed results are apparent (e.g., Shymansky, et al. 2004; Lynch, et al, 2006; Blanchard, et al, in review). As reported in *Education Week* (Viadero, 2009), "Of the eight such [randomized control trials] studies released by the federal institute [Federal Institute of Education Sciences] this academic year, six have produced mixed results pointing to few, or no, significant positive effects on student achievement." 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 discussed by Lynch et al. (2006), 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 (e.g., Blanchard et al., in review; Colburn, 2000; Klahr & Nigam, 2004; Lederman et al., 2007; Lynch et al., 2006; Marx et al., 2000; Shymansky et al., 2004) 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. This paper will focus on the effects of professional development and teaching with the GEMS curriculum compared to the effects of teaching with a more traditional curriculum on two student outcomes: content knowledge and attitude. Overall, our work is designed to begin to systematically address several 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? For whom does the use of such curricula plus professional development work, that is, what student moderator variables can we identify from the data collected? What are the characteristics of teachers who gain the most in terms of their own development from such curricula in combination with professional development? This paper will report on some preliminary answers to the first two of these questions with respect to student outcomes. The preliminary results about the fourth question, that focusing on teacher outcomes, are reported elsewhere (Granger et al., 2009). The remainder of the questions await the conclusion of the second year of the study.

Methods and Data Sources

Overview: This study employed a randomized quasi-experimental design. The experimental treatment group was comprised of 32 teachers who 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 either personal issues or non-compliance with instructional and/or data collection procedures). The control group teachers 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). Students' learning of space science concepts, inquiry skills, and affective dimensions (pre, post, and delayed post) were compared across groups.

Participants: This study was conducted in a county in central Florida in 4th and 5th grade classrooms during the 2007-08 school year. Randomization occurred at the level of the teacher assignment to treatment group (not at the student level). The sample was drawn from students from 32 experimental treatment group teachers and 29 control group teachers. The year one sample originally targeted 36 experimental treatment group teachers and 36 control group teachers, but by the completion of year one there were 32 treatment and 29 control teachers remaining in the study. 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. Each class consisted of approximately 14-20 students for a total of 1178 students—696 from classes of the GEMS group teachers and 482 from the classes of control group teachers. As for grade level of these students, 348 were 4th grade students and 830 were 5th grade students. Complete background information on each student (ethnicity, SES, etc.) was not available for all classes at the time of the statistical analysis for this presentation, but will be available for future reports.

Curricula: The curriculum employed by the experimental treatment group was the GEMS *Space Science Curriculum Sequence* (hereafter referred to as GEMS *SSCS*) (2007), which is a curriculum for teaching space science concepts for grades 3 through 5. Through experiential learning, discussions, and reflections centered upon the pedagogies that underpin the GEMS *SSCS,* teachers were prepared for teaching using this very "teacher friendly" curriculum. We label this curriculum as "teacher friendly" in that it 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 *SSCS* 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). The activities in the curriculum specifically target these core concepts and misconceptions in the attempt to change students' knowledge. 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 appropriate for the grades for which it is written, the GEMS *SSCS* 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 GEMS curriculum contrasted with the district adopted science text for grades 4 and 5 that 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 follow the text's presentation for the space science unit.

The structure of the experiment in which they were participating (i.e., quasi experimental 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 study. In this context, teachers in both groups were instructed 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.)

Professional Development: Because we realize that teachers' use of any curriculum is an act of interpretation, 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. This occurred in a 4-day professional development workshop two weeks before the beginning of school, a three-hour follow up immediately prior to the beginning of the teaching of the space science unit, and a three-hour session midway through the teaching of the unit to discuss questions that had arisen. A "science coach" was available to help them with logistical or pedagogical issues. In actuality the support requested was wholly logistical.

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.

All teachers attended a session for their group, either treatment or control, approximately two weeks following the teaching of the unit to discuss the curriculum, their experiences with it, and to complete the teacher post assessments.

Data Collection: 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.) 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 variance resulted from dates of completion of the unit that varied somewhat between classes given typical uncontrollable occurrences in individual teacher's schedules such as assemblies, field trips, amount of scheduled science time per day, etc.). The same pre-, post-, and delayed post-test administration schedule was used with both treatment and control groups. Semi-structured in-depth content interviews were also conducted with a subsample of 36 students (18 GEMS/18 control) using interview questions from Barnett and Morran (2002) 2.5 months ± 2 weeks post instruction (which will described in a later manuscript). Finally, at the time of the preparation of this paper, the VOSI-E scoring for student tests is not yet complete.

Data Analysis: 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 (GEMS = 1 for the GEMS treatment group and GEMS =0 for the 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. The resulting final models for student Achievement were the following:

Student-level model:

$$ACH_{ij} = \beta_{0j} + \beta_{1j} \,\mathrm{Pr}e\_ACH_{ij} + r_{ij} \qquad (1)$$

where the subscript ij indicates the ith student within the jth teacher. Thus, ACH_{ij} and Pre_ACH_{ij} are the post and pre measures, respectively, of Achievement for the ith student in the class of the jth teacher. The r_{ij} term is the model residual. Model (1) is identical to a simple linear regression, except that the equation coefficients with subscript j (β_{0j} and β_{1j}) allow the linear student-level relationship to vary over teachers. Modeling of this variation is provided by the following:

Teacher-level model:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} GEMS_j + \gamma_{02} \,\mathrm{Pr}e\_OE_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \end{aligned} \qquad (2)$$

where *Pre_OE* is the teacher pre_Outcome Expectancy variable and the *u* terms are residual terms. This portion of the complete model states that teacher-to-teacher variation of the intercept, β_{0j}, is explained by the teacher-level variables of *GEMS* and *Pre_OE*. In this case, no explanatory variables have been specified for the within-teacher effect of *Pre_ACH* (β_{1j}), but it has been allowed to vary randomly by inclusion of the random residual, u_{1j}.

Model specification in the HLM 6 software uses the multilevel formulation shown in (1) and (2). For purpose of interpretation of results, an alternative "mixed model" formulation is often useful. This form is obtained by simply substituting the two equations in (2) into the appropriate locations in (1), resulting in the following:

Mixed model formulation:

$$ACH_{ij} = \gamma_{00} + \gamma_{01} GEMS_j + \gamma_{02} Pre\_OE_j + \gamma_{10} Pre\_ACH_{ij} + (error\ term) \qquad (3)$$

This expression is identical to a multiple regression model with three independent variables, except for the presence of a more complex residual or error term. (If one were to use regression analysis for [3], the complex error term would result in violations of the assumptions associated with regression, which in turn would produce incorrect statistical inference.) The resulting analysis can be viewed as a multilevel version of analysis of covariance (ANCOVA) allowing the estimate of an "adjusted" effect of GEMS on Achievement, represented by γ_{01} in (3), controlling for student- and teacher-level covariates.

A search for possible moderators of the GEMS effect can be accomplished with simple extensions of the above "main effects" model. For example, assuming that we wish to determine if the *student-level* Pre_Achievment variable moderates the GEMS effect, the student-level model is identical to that in (1) and the teacher-level model is specified as:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} GEMS_j + \gamma_{02} Pre\_OE_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} GEMS_j + u_{1j} \end{aligned} \qquad (4)$$

Note that the second equation in (4) states that a variation in GEMS causes a variation in the effect of Pre_Achievment on the Achievement outcome. This implies an interaction between GEMS and Pre_Achievment. This interaction is better seen in the mixed model formulation of (1) and (4), which is

$$\begin{aligned} ACH_{ij} &= \gamma_{00} + \gamma_{01} GEMS_j + \gamma_{02} Pre\_OE_j + \gamma_{10} Pre\_ACH_{ij} \\ &+ \gamma_{11} GEMS_j \; Pre - ACH_{ij} + (error\ term) \end{aligned} \qquad (5)$$

This form clearly shows the product term that represents an interaction, with the corresponding coefficient of γ_{11} representing the interaction effect.

When one wishes to search for possible *teacher-level* moderators, a different kind of extension of the additive model is required. For example, we might want to determine if the teacher Pre_Achievement (*Pre_TACH*) moderates the effect of GEMS. Also, suppose (in order to reflect an actual analysis reported later) that the earlier teacher-level variable of *Pre_OE* is eliminated. The student-level model is still that shown in (1). In the teacher-level model, a product term involving the two interacting variables is added, as shown below.

$$\beta_{0j} = \gamma_{00} + \gamma_{01}GEMS_j + \gamma_{02}\,Pre\_TACH_j + \gamma_{03}GEMS_j\,Pre\_TACH_j + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j} \tag{6}$$

Substituting (6) into (1), the following mixed model formulation is obtained.

$$ACH_{ij} = \gamma_{00} + \gamma_{01}GEMS_j + \gamma_{02}\,Pre\_TACH_j$$
$$+ \gamma_{03}GEMS_j\,Pre\_TACH_j + \gamma_{10}\,Pre\_ACH_{ij} + (error\ term) \tag{7}$$

In this expression, the interaction effect is represented by the coefficient γ_{03}.

When the interaction term in either (5) or (7) is statistically significant, it is concluded that the GEMS effect varies over different levels of the moderator, and it is then necessary to describe the nature of this variation. Such a description, involving estimation of "simple effects" of GEMS for different moderator values, will be illustrated below in the Results section.

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 was viewed as exploratory, there was no attempt to control family-wise error.

Results

The results reported herein will 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 ($p<.001$). Likewise, teacher confidence with their content knowledge on the post assessment for the GEMS group compared to the control group was positive and statistically significant ($p=.089$).

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 positive 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 and delayed-post tests for achievement (space science content) and attitude (Homerton survey) were positive for students in the GEMS treatment group, though only the post test results were statistically significant ($p=0.004$ 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

Outcome	Unstandardized GEMS Effect	Standard Error	p-value	Standardized GEMS Effect[a]
Post Measures				
Achievement	0.629*	0.206	0.004	0.22*
Attitude	1.862*	0.995	0.067	0.12*
Delayed Post Measures				
Achievement	0.340	0.213	0.116	0.12
Attitude	1.282	1.075	0.239	0.08

* Statistically significant at the 0.10 level.
[a] Standardized effects have been obtained by dividing the raw score GEMS coefficient by the outcome standard deviation.

The HLM models required to obtain the GEMS main effects for the post test measures of student Achievement and Attitude were, as described in the Methods section, derived by a backward elimination approach starting with a HLM model having the corresponding student pre-measure at the student level and the GEMS, teacher Experience, Grade, and GEMS by Grade product terms at the teacher level. The Exploratory Analysis option in the HLM 6 software was used to add any teacher-level pre-measures that would significantly contribute. The result was that the final model for post Achievement, as illustrated in (3) in the Methods section, contained the corresponding student pre-measure at the student level and GEMS (GEMS = 1 for the GEMS treatment group and GEMS = 0 for the control group) and the teacher pre_Outcome Expectancy variables at the teacher level. The final model for post Attitude was identical except that teacher pre_Outcome Expectancy was not included. There was no evidence of important model differences between Grades 4 and 5 for either outcome. The models derived for the post test outcome measures were also used for the delayed-post test outcomes.

The estimated main effects of GEMS on the post and delayed-post test measures of the student Achievement and Attitude outcomes are summarized in Table 1. For the post test measures, both main effects were positive and statistically significant. 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 estimated 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. Although the main effects for the two delayed-post test measures were statistically nonsignificant, both were still positive in direction. The estimated magnitudes of the standardized forms of the delayed post measures were roughly half of the corresponding post test measures.

As explained in the Methods section, 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 (see [4]). 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 (see [6] or [7]). This search resulted in evidence for one moderator; teacher pre-Achievement was a moderator for student post test Achievement. Substituting estimates into (7), the interactive model for the predicted outcome was

$$\begin{aligned} ACH_{ij} &= 3.32 + 1.61 GEMS_j + 0.144 Pre\_TACH_j \\ &- 0.176 GEMS_j \, Pre\_TACH_j + 0.594 Pre\_ACH_{ij} \end{aligned} \quad (8)$$

One approach to describing the nature of this interaction is based on decomposing (8) into two models, one for each of the two study groups, by substituting 1 or 0 into the *GEMS* variable. The result is

$$\begin{aligned} GEMS\ group: ACH_{ij} &= 6.44 - 0.0315 Pre\_TACH_j \\ Control\ group: ACH_{ij} &= 4.83 + 0.0500 Pre\_TACH_j \end{aligned} \quad (9)$$

Figure 1 shows the standardized form of these two models that is obtained by dividing (9) by the standard deviation of post test Achievement of 2.88. It is seen from Figure 1 that the GEMS group is estimated to be superior to the control group for *Pre_TACH* 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.

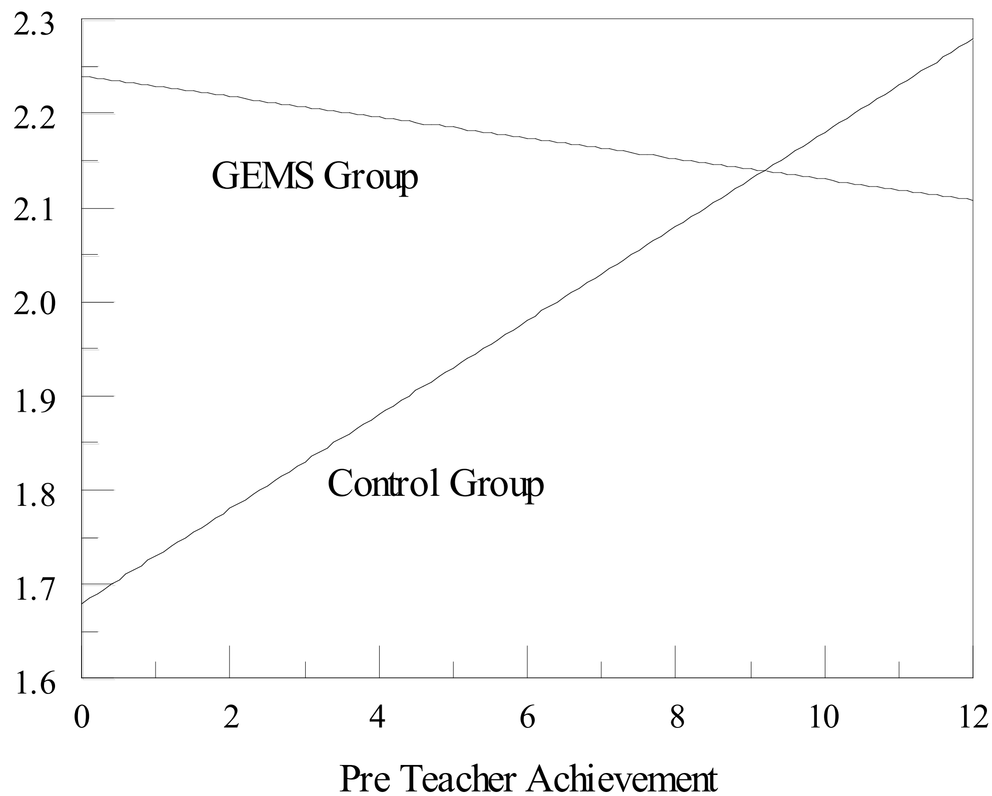

Figure 1: Predicted post student Achievement plotted as a function of the teacher moderator variable for the experimental and control groups. The GEMS standardized simple effect (i.e., the standardized difference between the predicted outcomes for the GEMS and control groups) is represented by the vertical separation of the two lines.

The values of the moderator for which the direction of the estimated GEMS effect can be trusted can be determined by the following expression for the simple GEMS effect.

$$Simple\ GEMS\ Effect = 1.61 - 0.176\,Pre\_TACH_j \quad (10)$$

This expression was obtained by collecting all terms involving *GEMS* in (8) and then dividing out the *GEMS* term (or, mathematically, by taking the partial derivative of [8] with respect to GEMS). This estimated simple effect is shown in standardized form as the dashed straight line in Figure 2. As expected from Figure 1, the GEMS contrast was positive (GEMS group superior to the control group) for values of *Pre_TACH* up to approximately 9 and then becomes negative for larger values of the moderator. The 90% confidence band in Figure 2 indicates that the GEMS effect estimated was statistically significant for values of teacher Pre_Achievment up to approximately 6.3 (i.e., the lower and upper values on the band do not contain zero). For higher moderator values the 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.

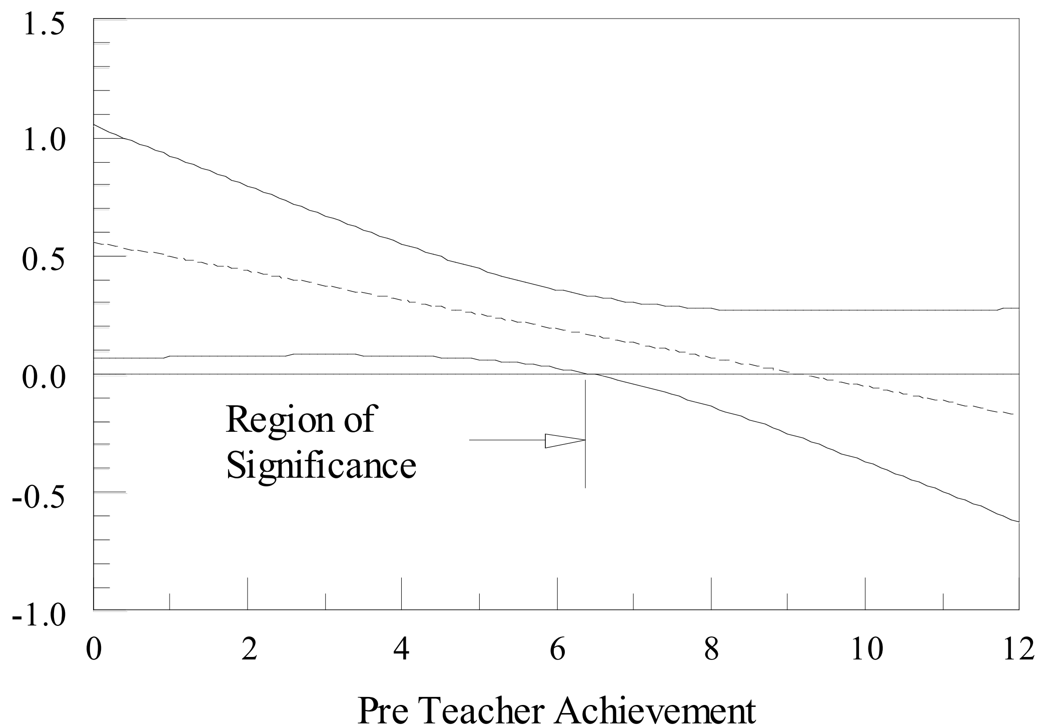

Figure 2: Estimate of the standardized simple effect of GEMS on the post student Achievement outcome as a function of teacher pre-Achievement (straight dashed line). The two solid curves represent the 90% confidence band on the effect estimate. The estimated effects are statistically significant at the 0.10 level for all values of Pre-Efficacy where the confidence interval does not capture zero.

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 also a positive GEMS effect on student attitude of space science, with no evidence of an interaction involving the corresponding teacher knowledge pre-instruction measure. The effects of GEMS on the delayed-post test measures were positive in direction but not statistically significant.

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 meeting). 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.

Discussion

It is important to note at the outset of this discussion that this is a preliminary analysis of the results from the first year of a two year study. The completed study will consist of two cohorts of teachers and students (year one plus year two) with teachers from both cohorts participating in the same regimen of professional development and teaching and students experiencing the same regimen of classroom instruction. The experimental methods were the same for each cohort.

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). It has been argued that teaching for accountability ***or*** teaching using reform-based practices such as inquiry is a false dichotomy (Rose, 2007; Shymansky et al., 1990). But given teachers' and school administrators' reticence to employ reform based practices in the classroom, the reasons behind the mixed results discussed above need to be elucidated. To provide the needed rationale to warrant the desired change of practice in the face of pressures exerted by required standardized testing accountability measures, 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 a systematic examination of the use of reform-based curricula in combination with teacher professional development focusing on the central question: Does the use of reform-based curricula in combination with professional development support student science learning (both cognitively and affectively)? Additional questions addressed include: How do teacher characteristics influence the learning their students experience as a result of the use of such curricula and professional development (i.e., teacher moderating variables)? What groups of students are most influenced by teachers' use of such curricula and professional development (i.e., student moderating variables)?

In answer to the 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 and 2).

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 due to the "newness" of their knowledge. 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 to learn. 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 we would argue that 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 that we employed. Further, in our research design, we 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 at least two classroom observations and/or videotapings during the unit. Fidelity to the instructional methodology is currently being quantified through RTOP analysis.

Implications

Our empirical analysis of the effectiveness of reform-based instruction using a specified curriculum in combination with professional development demonstrated positive effects on student cognitive and affective outcomes as compared with more traditional curricula and instruction. These results indicate that reform-based curricula and instructional methodologies are not in conflict with improved student performance on statewide assessment measures. The positive, though non-significant, trend that was maintained over 5 months is also encouraging for standardized assessments. This will be an important message to share with teachers and district administrators as curricula are selected on a district level and teachers make decisions about the kinds of teaching practices that are effective and can be employed in the teaching of science.

We further suggest that at least some of the mixed results found in an examination of the literature of such studies may be due to oversights in experimental design. We concur with Darling-Hammond as quoted in *Education Week* (Viadero, 2009) who argues that, though both difficult and expensive, attention to research design, including tight control of the comparison group, must become standard in studies employing experimental and quasi-experimental designs if we are to be able to compare and explain the results of educational research studies. Although laborious, in the long run, such attention to design will result in resource savings.

Further Research

The results reported herein give rise to several additional questions, some of which can be assessed with the data set from this study, and some of which will require additional research. Using the current data set and additional statistical analyses, we hope to examine, at least in part:

- For which student groups (i.e., SES, primary language) does this combination of curriculum and professional development work best? (That is, what student level moderators can we identify as influential?)

- What teacher characteristics influence the success of the application of combination of curriculum and professional development? (Note: some of our work on this to date are reported in Granger et al., 2009.)

Further studies will be needed to address these questions:

- What is the reason behind the result that teachers with low pre achievement test scores garnered such an increase in the test scores of their students?

- How many classroom samples (either observations or videotapes) are needed to reliably determine fidelity to the curriculum or instructional methodology?

- Are these same increases in student learning using similar curricula and professional development of teachers found in secondary-school classrooms where teachers usually have an increased background in science content (over that of the elementary-school teachers in this study)?

Acknowledgements: 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. The staff of the Florida Center for Research in STEM Education was essential in helping us to administer this project. 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

Abrams, E., Southerland, S.A., & Evans, C. (2007). Inquiry in the classroom: The necessary components of a useful definition. In E. Abrams, S.A. Southerland, & P Silva (Eds.), *Inquiry in the classroom: Realities and Opportunities*. Greenwich, CT: Information Age Publishing.

Barnett, M., & Morran, J. (2002). Addressing Children's Alternative Frameworks of the Moon's Phases and Eclipses. *International Journal of Science Education*. *24*, 859-879.

Blanchard, M., Southerland, S.A., Sampson, V., Annetta, L., & Granger, D.E. (in review). Student learning in a laboratory setting: Comparing the effectiveness of inquiry-based versus deductive science teaching methods. *Science Education.*

Blanchard, M., Southerland, S.A., & Granger, D.E. (2009). No silver bullet for inquiry: Making sense of teacher change following an inquiry-based research experience for teachers. *Science Education, 93*(2), 322-360.

Bybee, R. W. (1997). *Achieving science literacy*. Portsmouth, NH: Heinemann.

Chinn, C. & Maholtra, B. (2002) Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86*, 175-218.

Cohen, J. (1977). Statistical Power Analysis for the Behavioral Sciences (Revised Edition), New York: Academic Press.

Colburn, A. (2000). An inquiry primer. *Science Scope*, *23 (6),* 42-44.

Collaboration for Astronomy Education Research (CAER). (1999) Astronomy Diagnostic Test. http://solar.physics.montana.edu/aae/adt.

Dean, D., & Kuhn, D. (2006). Direct instruction vs. discovery: The long view. *Science Education, 91*, 384-397.

Dira-Smolleck, L.A. (2004). *The development and validation of an instrument to measure preservice teachers' self-efficacy in regards to the teaching of science as inquiry*. Unpublished doctoral dissertation. The Pennsylvania State University. (Creative Commons Attribution 2.5 License).

Duschl, R., Schwingruber, H.A., & Shouse, A.W. (2007). *Taking Science to School*: *Learning and Teaching Science in Grades K-8*. Committee on Science Learning, Washington, D.C.: National Academy Press.

Flick, F. B. (2003, March). *Inquiry as cognitive process*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Philadelphia.

Fraser, B.J. (1978). *Test of Science Related Attitudes*. Australian Council for Educational Research. Camberwell, Victoria, Australia: Allenby Press.

Granger, E.M., Bevis, T.H., Saka, Y., & Southerland, S. (2009). Learning About Space

Science: Comparing the Efficacy of Reform Based Teaching with a Traditional/Verifications Approach, Paper presented at the American Association for Educational Research, San Diego, CA.

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.

Howard, Michael and Associates (2007). MSP Teacher Feedback on Science Instructional Materials. (Instrument developed based on others used in National Science Foundation projects).

Karplus, R., & Their, H. D. (1967). *A new look at elementary school scienc*e. Chicago: Rand McNally.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: Effects of direct instruction and discovery learning. *Psychological Science, 15,* 661–667.

Kavanagh, C., & Sneider, C. (2007) Learning about gravity I. Free Fall: A guide for teachers and curriculum developers. *Astronomy Education Review, 5* (2).

Kavanagh, C. & Sneider, C. (2007) Learning about gravity II. Trajectories and orbits: A guide for teachers and curriculum developers, *Astronomy Education Review, 5* (2).

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*, 75–86.

Lawrence Hall of Science. (2007). *Space Science Curriculum Sequence*. Great Explorations in Math and Science. Berkeley, CA: University of California-Berkeley Press.

Lederman, J. & Lederman, N. (2005*). Developing and Assessing Elementary Teachers' and Students' Understandings of Nature of Science and Scientific Inquiry*. Paper Presented at the Annual Meeting of the National Association for Research in Science Teaching, Dallas, Texas, April, 2005.

Lederman, N., Lederman, J., Wickman, P. - O. & Lager-Nyqvist, L. (2007, April). *Development of a Valid and Reliable Protocol for the Assessment of Early Childhood Students' Conceptions of Nature of Science and Scientific Inquiry.* Paper presented at the annual conference of the National Association for Research in Science Teaching, New Orleans, LA.

Leonard, W. H. (1983). An experimental study of a BSCS-style laboratory approach for university general biology, *Journal of Research in Science Teaching, 20*(9), 807-813.

Leonard, W. H., Cavana, G. R., & Lowery, L. F. (1981). An experimental test of an extended discretion approach for high school biology laboratory investigations. *Journal of Research in Science Teaching, 18*, 497-504.

Lynch, S., O'Donnell, C.L., Hatchuel, E.E., Rethinam, V., Merchlinsky, S., & Watson, W.A. (2006) What's up with the Comparison Group?: How large quasi-experimental study of highly rated science curriculum units came to grips with unexpected results, Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA, April 7-11, 2006.

Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E., Geier, R., & Tal, R. T. (2004). Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform, *Journal of Research in Science Teaching, 41*, 1063-1080.

Metz, K. E. (2004). Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations. *Cognition and Instruction*, *22*, 219-290.

National Research Council. (1996). *National Science Education Standards*. Washington, D.C., National Academy Press.

National Research Council. (2000). *Inquiry and the National Science Education Standards*. Washington, D.C.: National Academy Press.

National Research Council. (2008) *Ready, Set, Science: Putting Research to Work in K-8 Science Classrooms.* Washington, D.C.: National Academy Press.

Pringle, R. M., & Carrier, S. M. (2005). The potential impacts of upcoming high-stakes testing on the teaching of science in elementary classrooms. *Research in Science Education, 35,* 347-361.

Raudenbush, S.W., & Bryk, A.S. (2002) *Hierarchical Linear Models: Applications and Data Analysis Methods* (second edition), Thousand Oaks, CA: Sage Publications.

Raudenbush, S., Bryk, A., Cheong, Y.F., Congdon, R., & du Toit, M. (2004) *HLM 6: Hierarchical Linear and Nonlinear Modeling,* Lincolnwood, IL: SSI Scientific Software International.

Rose, K. (2007). Doing Inquiry in the Classroom *Because* of Standardized Testing. In J. Settlage & S.A. Southerland, *Teaching science to all children: Using culture as a starting point.* New York: Routledge.

Sadler, P., Coyle, H., Cook-Smith, N. & Miller, J. (2007) Misconceptions-oriented Standards-based Assessment Resources for Teachers (MOSART), Cambridge, MA: Harvard College. http://www.cfa.harvard.edu/smgphp/mosart/site_map.html.

Sawada, D., Piburn, M.D., Kyle, W.C. & Alport, J.M. (2002) Measuring reform practices in science and mathematics classrooms: The reformed teaching observation protocol, *School Science and Mathematics, 102*, 245-253.

Schwartz, R. S., Lederman, N., & Lederman, N. (2008, March). An Instrument To Assess Views Of Scientific Inquiry: The VOSI Questionnaire. Paper presented at the international conference of the National Association for Research in Science Teaching. Baltimore, MD. March 30-April 2, 2008.

Settlage, J. (2007). Demythologizing science teacher education: Conquering the false ideal of open inquiry. *Journal of Science Teacher Education,* 18(4), 461-467.

Settlage, J., & Blanchard, M. (2007). Requisite Teacher Knowledge about Inquiry: Resources for Thinking about Teaching for Inquiry. In E. Abrams, S.A. Southerland, & P Silva (Eds.), *Inquiry in the classroom: Realities and Opportunities*. Greenwich, CT: Information Age

Publishing.

Settlage, J., & Meadows, L. (2002). Standards-based reform and its unintended consequences: Implications for science education within America's urban schools. *Journal of Research in Science Teaching*, 39(2), 114-127.

Shymansky, J. A., Hedges, L. V., & Woodworth, G. (1990). A reassessment of the effects of inquiry-based science curricula of the 60's on student performance. *Journal of Research in Science Teaching, 27*, 127-144.

Shymansky, J.A., Yore, L.D., & Anderson, J.O. (2004) Impact of a school district's science reform effort on the achievement and attitudes of third- and fourth-grade students, *Journal of Research in Science Teaching, 41*, 771-790.

Tate, R. (2004). Interpreting hierarchical linear and hierarchical generalized linear models with slopes as outcomes. *Journal of Experimental Education, 73*, 71-95.

Warrington, M., Younger, M., & Williams, J. (2000) Student attitude, image and the gender gap, *British Educational Research Journal, 26* (1), 393-407.