# The Effects of a Statewide Evaluation Initiative in Gifted Education on Practitioner Knowledge, Concerns, and Program Documentation

**Article** · November 2014

DOI: 10.1177/1932202X14549356



**4 authors**, including:

Ann Robinson
University of Arkansas at Little Rock



Alicia Cotabish
University of Central Arkansas

# The Effects of a STEM Intervention on Elementary Students' Science Knowledge and Skills

**Alicia Cotabish**
*University of Central Arkansas*

**Ann Robinson**
*University of Arkansas at Little Rock*

**Debbie Dailey**
*University of Arkansas at Little Rock*

**Gail Hughes**
*University of Arkansas at Little Rock*

*The purpose of the study was to assess elementary students' science process skills, content knowledge, and concept knowledge after one year of participation in an elementary Science, Technology, Engineering, and Mathematics (STEM) program. This study documented the effects of the combination of intensive professional development and the use of inquiry-based science instruction in the elementary classroom, including the benefits of using rigorous science curriculum with general education students. The results of the study revealed a statistically significant gain in science process skills, science concepts, and science-content knowledge by general education students in the experimental group when compared with students in the comparison group. Moreover, teacher participation in the STEM program had a statistically significant impact on students' variability in posttest scores. These interim student performance data support the implementation of rigorous differentiated science curriculum focused on improving science concept, content knowledge, and process skills.*

Our nation's Science, Technology, Engineering, and Mathematics (STEM) innovators are dependent on our educational system to cultivate, excite, and promote their STEM learning to influence their future career decisions. Unfortunately, the numbers of U.S. high school graduates choosing to major in a STEM-related field has declined steadily (National Science Board [NSB], 2010). Recently, the NSB presented recommendations to support the identification and development of talented young men and women who have the potential to become the next generation of STEM innovators (NSB, 2010). The suggestions included providing research-based STEM preparation for general education (elementary) teachers in the area of preservice training and professional development. Also, in efforts to encourage the STEM-related careers, the NSB recommended early exposure to STEM opportunities for all students and the opportunity for students to engage in inquiry-based learning, peer collaboration, and open-ended, real-world problem solving.

Twenty years ago, Brandwein (1995) recommended that science talent development begin in the early grades and include "evocative instruction, stimulating idea-enactive, inquiry-oriented behavior consistently in the classroom" (p. 41) to increase science proneness in children. More recently, Keeley (2009) stressed the importance of science in the early grades to maximize the cumulative learning processes involved in developing science talent. Both Keeley (2009) and Goldston (2005) argued that science achievement and conceptual understanding are affected when students are not exposed to science instruction until the middle grades. Moreover, the curiosity and enthusiasm for science may continually diminish if not fostered in the early grades (Pratt, 2007). Either those students choose to pursue another interest, or they lose the desire to take an advanced course in science. In the interest of developing our nation's future STEM innovators, changes in our science programs are necessary, especially in the elementary grades.

## Perspectives and Theoretical Framework

The theoretical framework of this study was guided by the principle that sustained and embedded teacher professional development together with the implementation of an inquiry-based science curriculum can positively influence student achievement. Brandwein (1995) stressed the importance of both the teacher and the curriculum in developing the science talent in young students. He further stated the greatest barriers to developing science talent include inadequately prepared teachers and an outdated curriculum that neglects the needs and interests of the child. More recently, Bitan-Friedlander, Drefus, and Milgrom (2004) suggested that the lack of follow-up support received by teachers constitutes a major barrier in implementing a new innovation. Bitan-Friedlander et al. recommended that professional development be lengthy enough to internalize the innovation and extended into the *real world* of the classroom. Furthermore, Johnson, Kahle, and Fargo (2006) reported greater increases in science achievement when students had effective teachers who were able to engage students in purposeful learning and active participation as characterized by an inquiry-based approach. Unfortunately, very few studies have been able to link student achievement to teacher professional development. For example, in a systematic review, Yoon, Duncan, Lee, and Shapley (2008) examined more than 1,300 studies to assess the effects of teacher professional development on student achievement, but only nine met their design criteria for inclusion in the review. Across the nine studies, researchers found an average effect size of .54 indicating professional development had a moderate effect on student achievement. The single science study examined revealed an effect size of .51 (Yoon et al., 2008).

**Elementary Science Teacher**

Elementary teachers are the gatekeepers to fostering the gifts and talents of future STEM innovators. Brandwein (1995) suggested that a significant improvement in science teaching can increase the talent pool of science learners. In order for teachers to cultivate science talent in their students, they must know how to teach science and how students learn science. Teachers need to be familiar with the language of science, and they need to have a deep understanding of the processes of science. Specifically, Michaels, Shouse, and Schweingruber (2008) maintained that teachers must be able to create science learning opportunities for their students and be able to relate science language and learning to real-life events.

The National Academy of Science report, *Taking Science to School*, indicated that a teacher's quality of understanding science affects instruction and, thereby, student achievement even more than the teacher's quantity of knowledge developed (Duschl, Schweingruber, & Shouse, 2007). The quality of understanding science is not stressed in many undergraduate preparation programs where science is often seen as a body of facts and investigation is following sequential steps in the scientific method with no real emphasis on inquiry or evidence-based inferences and conclusions (Duschl et al., 2007). Unfortunately, few elementary teachers will engage in professional development activities to improve their science teaching after receiving their undergraduate degree (Rice, 2005). Fulp (2002) reported that a majority of elementary science teachers have had fewer than 15 hours of science-specific professional development.

**Professional development.** As recommended by the National Research Council (NRC, 1996), all grade K-8 teachers should receive sustained science-specific professional development in teacher-preparation programs and in teacher in-service components (Duschl et al., 2007). These programs should focus on developing teachers' content knowledge in their classroom-specific science curriculum, providing examples of how students learn science and developing a plan to integrate support methods, such as technology, into the science curriculum. The teacher preparation should mirror what the teacher will actually teach in the classroom. The professional development opportunities should extend into the real world of the classroom, taking into account teachers concerns, their need for resources, and their current knowledge and skills.

One high-profile report, *Rising Above the Gathering Storm*, maintained that high quality, content driven professional development can have a significant impact on student performance (Committee on Science, Engineering, and Public Policy, 2007). Professional development has the power to affect teacher instruction and student learning if it is of high quality. High-quality professional development must be empirically validated, be able to promote and extend effective curricula and instructional models, and be intensive, sustained, content focused, well defined, and strongly implemented (Yoon et al., 2008). Yoon et al. (2008) further explained that professional development can improve student achievement by enhancing teacher knowledge, skills, and motivation, which lead to better teacher instruction; thereby, increasing student success. In examining science-specific professional development, Kennedy (1998) reported increased student achievement in science when teacher in-service programs modeled scientific processes to develop content knowledge. Kennedy noted that professional development programs focused on specific teacher instructional strategies, such as cooperative learning, had smaller influences on student achievement than those programs focused on content knowledge, curriculum implementation, or student science-learning behavior. In a comparison of three professional development programs targeting earth science, Penuel, Gallagher, and Moorthy (2011) reported increased student learning when teachers were provided with explicit instructions on models of teaching (effect size = .34). Teachers were better able to use specific learning strategies when provided ample time and support to practice and develop these strategies. Moreover, in a study of science professional development to improve teacher content knowledge and inquiry practices, Buczynski and Hansen (2010) found an increase in science achievement among students whose teachers participated in a targeted professional development program. Furthermore, teachers believed their increased use of inquiry practices contributed to student success. Borman, Gamoran, and Bowdon (2008) reported mixed results when analyzing student achievement tied to teacher professional development in science. Borman et al. found a statistically significant negative effect associated with experienced teachers' participation in the professional development. In contrast, the intervention produced positive effects on student achievement for novice teachers with three or fewer years of experience.

***Duration.*** Contact time with teachers is also reported to have an effect on teacher instruction and thereby, student achievement. In an analysis of teacher professional development programs in math and science, the Council of Chief State School Officers (CCSSO, 2008) reported that most effects were reported in programs providing a minimum of 45 hours of professional development annually. These were usually in the form of summer institutes plus job-embedded activities throughout the year. Furthermore, Supovitz and Turner (2000) reported that teacher adoption of inquiry-based instructional practices were not evident until after 40 to 79 hours of science-specific professional development. In a review of literature, Gerard, Varma, Corliss, and Linn (2011) reported that teachers who participated in sustained professional development for over a year were more likely to incorporate the curriculum that increased the students' inquiry-based learning experiences.

***Peer coaching.*** Sustained, embedded professional development can involve a collaborative program utilizing a mentor or a peer coach. This type of professional development allows the teacher to apply learning in the real world of the classroom while being supported by a peer coach. Showers and Joyce (1996) defined peer coaching as a means of helping teachers transfer newly acquired learning and skills from the "workshop" to the classroom. The purpose of peer coaching is to support teacher learning and change and thereby, ultimately, transfer benefits to students. Little (2005) argued that peer coaching can provide teachers a natural support system that can enhance teacher performance by the privileged sharing of knowledge and expertise through collaboration.

When quality professional development is combined with a coaching or mentoring program, optimal changes in teacher instruction occur (Neuman & Cunningham, 2009). Neuman and Cunningham reported a large effect size ($d$ = .77) when peer coaching was paired with professional development in comparison with implementation of a professional development component only. When measuring the effect of peer coaching on student learning, Showers (1984) found coaching significantly contributed to higher student achievement scores on a concept attainment measure, $F(2, 138) = 4.34$, $p = .01$. Taken together, results of these studies indicated peer coaching enhances the ability of teachers to transfer their learning to the classroom leading to greater student achievement. Drawing from a qualitative case study, Appleton (2008) suggested that peer coaching in science needs to focus on classroom support to enhance pedagogical content knowledge in order to generate teacher change. Providing support to teachers is vital in an elementary classroom where science materials and content knowledge are limited.

## Science Curriculum

**Inquiry-based science.** As defined by the National Research Council (NRC, 1996), inquiry-based science involves an investigatory instructional approach that allows students to develop an understanding of science knowledge through questioning, designing, and conducting experiments, basing conclusions on the analysis of experimental data, and reporting their findings. Research in science education suggests that curriculum based on in-depth understanding of science concepts and instruction focused on investigatory rather than a traditional approach best develop the talents and motivation of students to do science in the real world (Robinson, Shore, & Enersen, 2007; Yoon, 2009).

Metz (2008) argued that typical science curriculum found in most elementary classrooms fails to encourage and cultivate a child's natural curiosity in science. In comparison to a traditional elementary science class, Granger, Bevis, Saka, and Southerland (2009) found a statistically significant increase in elementary student content knowledge when a reform-based science curriculum utilizing an inquiry approach, Great Explorations in Math and Science (GEMS), was implemented. Teachers received professional development on the curriculum focused on the use of models and evidence in science and had access to a science coach throughout the year. Finally, in a longitudinal study, Banilower, Fulp, and Warren (2010) reported increased science achievement scores from students whose teachers implemented a module-based curriculum which utilized a hands-on, inquiry-based approach. Student growth was more pronounced when teachers attended multiple professional development sessions on the curriculum.

**Science knowledge and skill development.**

***Science content and process skill development.*** Science knowledge cannot be comprehended without understanding the process through which that knowledge was developed (Duschl et al., 2007). In a typical science classroom, the construction of science knowledge is often separated from the development of science process skills. In contrast, science process should be a means by which science knowledge is gained. Students who participate in scientific practices, such as, questioning, hypothesizing, investigating, developing models, interpreting and analyzing data, linking conclusions to data, and sharing findings with others, are able to develop and provide meaning to knowledge and explanations of the natural world (Michaels et al., 2008). Science knowledge is being acquired through doing science as opposed to reading about science. Science instruction focused on active learning also promotes student interest in science. Metz (2008) reported increased student attention spans among first graders when they were allowed to engage in their own investigations. Kim et al. (2012) reported a statistically significant increase ($\eta^2$ = .13, medium effect size) in science achievement scores over a three-year period among students who were exposed to an inquiry-based science curriculum when compared to students who received traditional science instruction.

***Science content development and concept connections.*** As recommended by Duschl et al. (2007), science curriculum should be organized around a focused set of core ideas as opposed the typical science curriculum where many disconnected topics are covered. The inclusion of overarching concepts is recommended to increase the number content connections students are able to make throughout the science curriculum (Metz, 2008; VanTassel-Baska, 1998). These big or core ideas should be explored extensively and increased in complexity across succeeding grade levels. Students will be more successful if allowed to explore fully a few scientific concepts throughout the year instead of superficially covering many concepts as in a typical science textbook. Roth et al. (2011) reported a large effect size ($ES$ = .32) on student achievement when students were able to make content connections using big ideas. Roth et al. described this strategy as “linking content ideas to other science ideas” (p. 137), by creating a science content storyline to form connections across the curriculum. Typical science concepts such as change, scale, evolution, and systems can be used to connect many varied science topics and can increase in complexity as students progress through grade levels. Kim et al. (2012) suggested curriculum development include the use of overarching concepts as a means to increase science understanding and skills; thereby, enhancing both science achievement and critical thinking skills. In summary, research supports emphasizing overarching concepts and developing scientific process skills to build science-content knowledge.

### Purpose of the Study

The purpose of the study was to assess elementary students’ science process skills, content knowledge, and concept knowledge after one year of participation in an elementary STEM program. STEM Starters is an intervention that provided teacher professional development in efforts to increase student and teacher science-content knowledge, process skills, and concept development. Professional development components included week-long summer institutes and peer-coaching classroom support. These components focused on training and implementation of an inquiry-based science curriculum, which included the use of problem-based learning units and a STEM biography series. Professional development also promoted the use of technology to engage learners in the classroom. Specifically, the research questions were:

**1.** Are elementary students’ understandings of science process skills significantly different after one year of participation in STEM Starters compared with their peers who did not participate?

**2.** Is the science-content knowledge of elementary students significantly different after one year of participation in STEM Starters compared with their peers who did not participate?

**3.** Is the science-concept knowledge of elementary students significantly different after one year of participation in STEM Starters compared with their peers who did not participate?

## Method

### Design

The current study was part of a larger randomized field study of teacher professional development and student learning in science. Only student data are reported in the current study. The data represent the first year of a two-year intervention consisting of science-focused curriculum units implemented in the grade-level classrooms and the gifted pull-out classroom. As part of the intervention, experimental teachers received 30 hours of professional development in the form of a week-long summer institute and 30 hours of one-to-one peer coaching focused on increasing teacher knowledge and skills in science. Data were organized and reported for both experimental and comparison group students.

**Participants.** Randomly selected from two districts in a southern state, 70 teachers from Grades 2 through 5 were assigned to the treatment and control groups. Although random selection was utilized, only two teachers were males. Years in teaching ranged from 0 to 34 with an average of 12.8 years. With the exception of two teachers, all reported less than one year of experience in teaching science, and no participant held a teaching degree or licensure in science. This finding was expected because literacy

Table 1
*Number of Students by Grade Level*

| Grade | Experimental Students | Comparison Students |
|---|---|---|
| 2 | 197 | 220 |
| 3 | 206 | 256 |
| 4 | 194 | 235 |
| 5 | 221 | 221 |
| Total | 818 | 932 |

and math are the foci of instruction; science is often ignored in the elementary grades. Arkansas requires 60 hours of professional development annually, but there are no requirements for science-focused professional development. All teachers were White with the exception of one African American female assigned to the control group.

Students assigned to experimental teachers were designated as experimental STEM Starters students, and those assigned to control teachers were designated as comparison students. In year 1 of the project, 818 students were served including 179 gifted students and 185 special education students. Table 1 highlights the number of experimental and comparison students presented in this study. Because complete data were required (no missing data) for the analysis of each research question and cell sizes of at least 5 were needed, the reduced sample sizes for each research question were: Fowler = 713, content = 965, and concept = 739.

### Intervention

**Project components.** The main components of the intervention were: (a) inquiry or problem-based science units of nine weeks in duration (one unit for general education students; two units for gifted students), and (b) two types of teacher professional development including: (a) summer institutes which train directly on curriculum materials, technology use in the classroom, and (b) Blueprints for Biography. In this particular study, teachers participated in a total of 60 hours of professional development (30 hours of summer institute and 30 hours of peer coaching).

**Curriculum.** The nationally award-winning William and Mary science curriculum utilized by STEM Starters focuses on specific overarching concepts that were integrated throughout the unit. The overarching concepts emphasized in these units were change (Grades 2 and 3) and systems (Grades 4 and 5). Students were asked to brainstorm examples and non-examples; as well as categorize, and make generalizations about these concepts. Comparison students received science instruction using the school-adopted science curriculum, and their teachers conducted science as usual.

Students assigned to an experimental teacher received instruction in one William and Mary inquiry-based curriculum unit. The following William and Mary curriculum units were implemented over the course of a semester, for the minimum of nine weeks: *Weather Reporter* (Grade 2), *What's the Matter* (Grade 3), *Electricity City* (Grade 4), and *Acid, Acid Everywhere* (Grade 5).

The Grade 2 and 3 units were considered inquiry-based learning units focused on exposing young students to science concepts and scientific processes. These units engaged students in creative and critical thinking opportunities through investigations and problem solving (Bracken et al., 2008). Both *Weather Reporter* and *What's the Matte*r featured age-appropriate scientific content.

The Grade 4 and 5 units used by STEM Starters were inquiry-based science units that involve students in real-world problem solving. The Grade 4 unit (Electricity City) allowed students to design a model city complete with electrical lighting. The Grade 5 unit (Acid, Acid Everywhere) exposed students to the problems of a chemical spill. Both units incorporate scientific thinking and reasoning as well as advanced science content.

**Teacher professional development.** STEM Starters teachers participated in a one week-long summer institute focused on science content and delivery, specific curriculum units, technological applications, and differentiation of instruction over the course of a summer. The summer institute provided 30 hours of professional development necessary for the implementation of the science curriculum units. Professional development involved teachers taking the role of students in the implementation of the curriculum units. An expert science instructor led the teachers through the problem-solving units by modeling effective science instruction for high-ability learners. Emphasis was placed on overarching concepts, higher order thinking skills, inquiry-based instruction, experimental design, and the use of technology as recommended by VanTassel-Baska (1998).

Professional development also involved the use of technology in the classroom to enhance the learning processes. Teachers were exposed to multiple Internet resources that correlate with their specific units. These resources provided content information for the student and teacher, as well as offered multiple activities and games to motivate student learning.

STEM Starters provided peer coaching on a weekly basis to the participant teachers. The peer coach is a former secondary chemistry/physics and a gifted and talented teacher. Once school began, the peer coach was in each of the schools two to three times per week for a total of 30 hours total across the school year. The peer coach visited each class at least twice a month to provide support to the teacher. In the classroom, the peer coach modeled the lesson, assisted the teacher with instruction, and monitored and encouraged student involvement. Outside of the classroom, the peer coach made certain that all necessary science activity materials were in the schools and maintained contact with all teachers by phone or e-mail to ensure their needs are being met.

### Instrumentation

**Raters.** Education majors and teachers were recruited to assist the research team with scoring student assessments. All raters participated in six hours of training, and only raters with an inter-rater reliability of .90 or greater ($n$ = 7) were selected to assist with the scoring. During scoring which was supervised, researchers randomly selected assessments to monitor consistency and accuracy in scoring and to ensure that inter-rater reliability remained above .90.

**Student science process skills.** To address research question one, the Scoring Rubric for Scientific Processes—Adapted Fowler Test (Adapted from Fowler, 1990) was used to assess students' understanding of the design of science experiments. The Adapted Fowler Test is an open-ended assessment that required students to design a simulated controlled experiment to address a scientific question. For example, in one form of the measure, students are asked to design an experiment to answer the question, "Are bees attracted to diet cola?" Responses describing a students' proposed design were scored across five criteria with ratings of (0) no evidence (3) to strong evidence, with two additional points possible on one criterion resulting in 17 points possible. According to Adams and Callahan (1995), the general approach of the Adapted Fowler Test for assessing science process skills is requiring students to mimic intellectual processes used in real-world science. The same researchers examined the convergent and discriminate validity and found weak patterns of correlation that limit the use of scores for making decisions for individual students; however, the results did support use for decisions about groups of students. Reports of alternate forms, inter-rater, and intra-rater reliabilities ranged from .76 to .95 (Adams & Callahan, 1995).

**Student knowledge of science content.** To address research question two, students' science-content knowledge was assessed using pre–post embedded curriculum-based assessments for each unit to capture student learning gains. In Grades 2–3, the curriculum-based content assessments utilized open-ended concept mapping to ascertain student understanding of knowledge of content topics in science (i.e., weather and matter). To score the concept map, raters followed a scoring rubric that assigned point totals to clear and accurate levels, propositions, cross connections, and examples that were representative of the selected content. In Grades 4–5, the curriculum-based content assessments utilized short-answer questions to assess the student understanding of the science content. The Grades 2–4 versions had 15 points possible, and the Grade 5 version had 20 points possible.

**Student knowledge of science concepts.** To address research question 3, students' science concept knowledge was assessed using pre–post embedded curriculum-based assessments for each unit to capture student learning gains. Specifically, the curriculum-based concept assessments encompassed: (a) concepts in science and mathematics; (b) overarching concepts that unify STEM disciplines (i.e., systems, change, patterns, cause and effect); and (c) interrelated science process skills; (d) critical thinking skills; (e) creative thinking; and (f) curiosity and interest in the world of science. The Grades 2–3 versions had 20 points possible, and the Grades 4–5 versions had 35 points possible.

## Results

### Analytic Strategy

Because students in the study were nested within classrooms, the researchers used multilevel modeling for the analysis to avoid the aggregation bias and misestimation of standard errors associated with ordinary least squares regression (Bickel, 2007; Heck, Thomas, & Tabata, 2010; Raudenbush & Bryk, 2002). Value-added models were used to analyze the treatment effects with students (level 1) nested within teachers/classrooms (level 2) on the dependent variable for each research question (Fowler, content, and concept-posttest scores, respectively). The level 1 student variables, grade level, and pretest score were entered into the model as controls. The level 2 variable was a dummy-coded variable representing teachers' participation in the experimental or comparison group. For each research question, three models were estimated the null or unconditional model, Model 2 with the student covariates, and Model 3 reflecting the treatment groupings. Researchers conducted a likelihood ratio test (LRT) to examine the model fit with the additional parameters of each new model using the –2 log likelihood (-2LL) statistics. The LRT test uses a $\chi^2$ distribution with the degrees of freedom reflecting the number of parameters added to the model. If a model provided a better fit than the previous model, then the null hypothesis was rejected. All analyses hierarchical linear modeling analyses were conducted using the Statistical Package for the Social Sciences (SPSS) version 16 (Chicago, IL: SPSS Inc.).

**Fowler**

The descriptive statistics for scores on the Fowler for both the experimental and comparison groups are presented in Table 2 for each grade level. Both the raw scores and the percentage correct scores (i.e., raw score/17 points possible) for the test are displayed. Additionally, the gains between pretest and posttest are presented.

First, an unconditional, or null model with no predictors, was estimated to compute the intra-class correlation (ICC). The ICC represents the proportion of variance at the classroom level in relation to the total variance. The ICC for the Fowler was .49; thus, 49% of the variance in Fowler posttest scores was between classrooms. The results of the unconditional model (see Table 3, Model 1) were statistically significant (Wald $Z = 4.55$, $p < .001$) and suggested that the development of a multilevel model was warranted.

Model 2 included students' grade level and Fowler pretest score as covariates, and both variables provided statistically significant contributions to the model, and the classroom variance component remained statistically

Table 2
*Means and Standard Deviations of Scores on the Fowler*

| | Grade 2 | Grade 3 | Grade 4 | Grade 5 | Overall |
|---|---|---|---|---|---|
| Experimental raw scores | | | | | |
| Pretest | 4.22 (2.28) | 3.47 (2.36) | 1.34 (1.56) | 1.85 (1.90) | 2.97 (2.40) |
| Posttest | 7.67 (2.61) | 7.19 (2.73) | 3.17 (2.23) | 6.62 (3.29) | 6.44 (3.18) |
| Gain | 3.45 | 3.72 | 1.83 | 4.77 | 3.47 |
| *n* | 126 | 134 | 82 | 81 | 423 |
| Comparison raw scores | | | | | |
| Pretest | 4.85 (2.02) | 4.99 (2.84) | .95 (1.21) | 1.96 (1.93) | 3.42 (2.73) |
| Posttest | 6.68 (1.96) | 7.61 (2.69) | 2.29 (1.96) | 3.89 (2.28) | 5.33 (3.07) |
| Gain | 1.83 | 2.62 | 1.34 | 1.93 | 1.91 |
| *n* | 101 | 69 | 75 | 45 | 290 |
| Experimental % correct | | | | | |
| Pretest | 24.84 (13.39) | 20.41 (13.85) | 7.89 (9.16) | 10.89 (11.17) | 17.48 (14.09) |
| Posttest | 45.10 (15.36) | 42.27 (16.08) | 18.65 (13.10) | 38.93 (19.32) | 37.89 (18.68) |
| Gain | 20.26 | 21.86 | 10.76 | 28.04 | 20.41 |
| Comparison % correct | | | | | |
| Pretest | 28.54 (11.86) | 29.33 (16.83) | 5.57 (7.10) | 11.50 (11.35) | 20.14 (16.07) |
| Posttest | 39.31 (11.55) | 44.76 (15.83) | 13.49 (11.56) | 22.88 (13.40) | 31.38 (18.05) |
| Gain | 10.77 | 15.43 | 7.92 | 11.38 | 11.24 |

*Note.* 17 points possible.

Table 3
*Coefficients and Variance Components on Fowler Posttest Scores (713 Students and 49 Teachers)*

| | (1) Null Model | (2) Student Covariates | (3) Treatment Group |
|---|---|---|---|
| Coefficient estimates | | | |
| Intercept | 6.00*** | 5.69*** | 4.94*** |
| Grade level | | −.60** | −.54* |
| Fowler pretest | | .62*** | .63*** |
| Treatment group | | | 1.13*** |
| Variance (% of total) | | | |
| Teacher | 4.94 (48.72%) | 1.93 (32.94%) | 1.49 (27.49%) |
| Student | 5.20 (51.28%) | 3.93 (67.06%) | 3.93 (72.51%) |
| Total | 10.14 | 5.86 | 5.42 |
| Model fit statistics | | | |
| −2 Log likelihood | 3,326.85 | 3,097.71† | 3,086.79† |

* $p < .05$. ** $p < .01$. *** $p < .001$. † Provided a better model fit than the preceding model.

Table 4
*Means and Standard Deviations of Scores on the Content Test*

| | Grade 2 | Grade 3 | Grade 4 | Grade 5 | Overall |
|---|---|---|---|---|---|
| Experimental raw scores | | | | | |
| Pretest | .66 (1.25) | .80 (2.47) | 5.74 (3.72) | .10 (.10) | |
| Posttest | 4.46 (3.44) | 4.54 (3.47) | 7.80 (4.37) | 14.57 (7.06) | |
| Gain | 3.80 | 3.74 | 2.06 | 14.47 | |
| *n* | 151 | 138 | 130 | 101 | |
| Comparison raw scores | | | | | |
| Pretest | .00 (.00) | .37 (.93) | 6.23 (3.77) | .00 (.00) | |
| Posttest | .50 (1.14) | .75 (1.73) | 7.85 (4.13) | 2.66 (5.09) | |
| Gain | .50 | .38 | 1.62 | 2.66 | |
| *n* | 158 | 109 | 109 | 68 | |
| Experimental % correct | | | | | |
| Pretest | 4.37 (8.36) | 5.31 (16.47) | 38.26 (24.79) | .05 (.50) | 12.25 (21.77) |
| Posttest | 29.76 (22.95) | 30.29 (23.11) | 53.30 (29.15) | 72.87 (35.28) | 44.16 (32.20) |
| Gain | 25.39 | 24.98 | 15.04 | 72.82 | 31.91 |
| Comparison % correct | | | | | |
| Pretest | .00 (.00) | 2.45 (6.20) | 41.56 (25.13) | .00 (.00) | 10.80 (21.74) |
| Posttest | 3.33 (7.63) | 5.02 (11.55) | 52.32 (27.50) | 13.31 (25.46) | 17.30 (27.33) |
| Gain | 3.33 | 2.57 | 10.76 | 13.31 | 6.50 |

*Note.* 15 points possible for Grades 2–4 and 20 points possible for Grade 5.

significant (Wald $Z = 4.34$, $p < .001$). Moreover, the inclusion of the covariates reduced the variability in students' Fowler posttest scores from 10.14 to 5.86 points, a substantial reduction of 42.21%. The results of the LRT also indicated a statistically significantly better fit of Model 2 than Model 1 ($\Delta\chi_{df=2}^2 = 229.14$, $p < .001$).

In Model 3, teachers' group membership (experimental or control) was statistically significant as was the classroom variance component (Wald $Z = 4.18$, $p < .001$). Group membership reduced the variability in students' Fowler posttest scores by an additional 7.51% from 5.86 to 5.42. The results of the LRT also indicated a statistically significantly better fit of Model 3 than Model 2 ($\Delta\chi_{df=1}^2 = 10.92$, $p < .001$). Thus, after controlling for students' grade level and pretest score and accounting for the nested nature of the data; teacher participation in the treatment program appeared to account for approximately 7.5% of the variability in students' posttest scores.

**Content**

The descriptive statistics for scores on the Content Test for both the experimental and comparison groups are presented in Table 4 for each grade level. Both the raw scores and the percentage correct scores for the test are displayed. Additionally, the gains between pretest and posttest are presented.

Again, analysis began with the estimation of an unconditional or null model with no predictors to allow for computation of the ICC. The ICC for the Content Test was .67. The ICC value indicated that 67% of the variance in Content posttest scores was between classrooms. The results of the unconditional model (see Table 5, Model 1) were statistically significant (Wald $Z = 5.06$, $p < .001$) and suggested that the development of a multilevel model was warranted.

Model 2 included students' grade level and content pretest score as covariates, and both variables provided statistically significant contributions to the model, and the classroom variance component remained statistically significant (Wald $Z = 4.50$, $p < .001$). Moreover, the inclusion of the covariates reduced the variability in students' content posttest scores from 25.32 to 18.64 points or 26.38%. The results of the LRT also indicated a statistically significantly better fit of Model 2 than Model 1 ($\Delta\chi_{df=2}^2 = 68.46$, $p < .001$).

In Model 3, teachers' group membership (experimental or control) was statistically significant as was the classroom variance component (Wald $Z = 4.70$, $p < .001$). Group membership reduced the variability in students' content posttest scores by an additional 12.07% from 18.64 to 16.39. The results of the LRT also indicated a statistically significantly better fit of Model 3 than Model 2 ($\Delta\chi_{df=1}^2 = 7.93$, $p < .05$). Therefore, after controlling for students' grade level and pretest score and accounting for the nested nature of the data; teacher participation in the treatment program appeared to account for approximately 12% of the variability in students' posttest scores.

Table 5
*Coefficients and Variance Components on Content Posttest Scores (965 Students and 54 Teachers)*

| | (1) Null Model | (2) Student Covariates | (3) Treatment Group |
|---|---|---|---|
| Coefficient estimates | | | |
| Intercept | 4.21*** | 2.42 | 3.26*** |
| Grade level | | 2.01*** | 2.05*** |
| Content pretest | | .32*** | .32** |
| Treatment group | | | 1.58** |
| Variance (% of total) | | | |
| Teacher | 17.09 (67.50%) | 10.80 (57.94%) | 8.52 (51.98%) |
| Student | 8.23 (32.50%) | 7.84 (42.06%) | 7.87 (48.02%) |
| Total | 25.32 | 18.64 | 16.39 |
| Model fit statistics | | | |
| −2 Log likelihood | 4,957.53 | 4,889.07† | 4,881.14† |

* $p < .05$. ** $p < .01$. *** $p < .001$. † Provided a better model fit than the preceding model.

Table 6
*Means and Standard Deviations of Scores on the Concept Test*

| | Grade 2 | Grade 3 | Grade 4 | Grade 5 | Overall |
|---|---|---|---|---|---|
| Experimental raw scores | | | | | |
| Pretest | 7.52 (3.78) | 8.78 (3.44) | 3.13 (4.52) | 4.22 (6.25) | |
| Posttest | 11.68 (3.86) | 10.20 (2.73) | 11.21 (8.87) | 22.09 (7.80) | |
| Gain | 4.16 | 1.42 | 8.08 | 17.87 | |
| *n* | 130 | 45 | 126 | 95 | |
| Comparison raw scores | | | | | |
| Pretest | 8.45 (3.40) | 9.74 (2.79) | 4.58 (6.40) | 3.55 (5.43) | |
| Posttest | 9.27 (3.55) | 10.77 (2.01) | 2.44 (4.45) | 9.81 (8.67) | |
| Gain | .82 | 1.03 | –2.14 | 6.26 | |
| *n* | 139 | 31 | 106 | 67 | |
| Experimental % correct | | | | | |
| Pretest | 37.62 (18.90) | 43.89 (17.19) | 8.96 (12.92) | 12.06 (17.84) | 23.08 (22.13) |
| Posttest | 59.38 (19.30) | 51.00 (13.63) | 31.52 (25.16) | 63.13 (22.85) | 50.14 (25.31) |
| Gain | 21.76 | 7.11 | 22.56 | 51.07 | 27.06 |
| Comparison % correct | | | | | |
| Pretest | 42.27 (17.01) | 48.71 (13.96) | 13.07 (18.28) | 10.15 (15.51) | 27.55 (23.15) |
| Posttest | 46.33 (17.76) | 53.87 (10.06) | 5.77 (9.10) | 28.02 (24.77) | 30.90 (24.91) |
| Gain | 4.06 | 5.16 | –7.3 | 17.87 | 3.35 |

*Note.* 20 points possible for Grades 2–3 and 35 points possible for Grades 4–5.

## Concept

The descriptive statistics for scores on the Content Test for both the experimental and comparison groups are presented in Table 6 for each grade level. Both the raw scores and the percentage correct scores for the test are displayed. Additionally, the gains between pretest and posttest are presented.

Analysis began with the estimation of an unconditional or null model with no predictors to allow for computation of the ICC. The ICC for the Concept Test was .53. The ICC value indicated that 53% of the variance in Concept posttest scores was between classrooms. The results of the unconditional model (see Table 7, Model 1) were statistically significant (Wald $Z = 4.48$, $p < .001$) and suggested that the development of a multilevel model was warranted. Further, 47.40% of the variance in content posttest scores was between students within classrooms and 52.60% was between classrooms.

Model 2 included students' grade level and content pretest score as covariates, and both variables provided statistically significant contributions to the model, and the classroom variance component remained statistically significant (Wald $Z = 4.49$, $p < .001$). However, the inclusion of the covariates only reduced the variability in students'

Table 7

*Coefficients and Variance Components on Concept Posttest Scores 735 Students and 44 Teachers)*

| | (1) Null Model | (2) Student Covariates | (3) Treatment Group |
|---|---|---|---|
| Coefficient estimates | | | |
| Intercept | 9.34*** | 9.37 | 6.81** |
| Grade level | | −.22 | −.10 |
| Concept pretest | | .10* | .84* |
| Treatment group | | | 4.96** |
| Variance (% of total) | | | |
| Teacher | 27.44 (52.60%) | 27.20 (52.58%) | 20.41 (45.96%) |
| Student | 24.73 (47.40%) | 24.53 (47.41%) | 24.00 (54.04%) |
| Total | 52.17 | 51.73 | 44.41 |
| Model fit statistics | | | |
| −2 Log likelihood | 4,592.87 | 4,587.14 | 4,560.08† |

* $p < .05$. ** $p < .01$. *** $p < .001$. † Provided a better model fit than the preceding model.

concept posttest scores by 1%. The results of the LRT also indicated that Model 2 was not a statistically significantly better fit than Model 1 ($\Delta\chi_{df=2}^2 = 5.73$, $p > .05$).

In Model 3, teachers' group membership (experimental or control) was statistically significant as was the classroom variance component (Wald $Z = 4.41$, $p < .001$). Group membership reduced the variability in students' concept posttest scores by an additional 14.15% from 51.73 to 44.41. The results of the LRT indicated a statistically significantly better fit of Model 3 than Model 2 ($\Delta\chi_{df=1}^2 = 27.06$, $p < .001$). Therefore, after controlling for students' grade level and pretest score and accounting for the nested nature of the data, teacher participation in the treatment program appeared to account for approximately 14% of the variability in students' posttest scores.

## Discussion

The results of this study revealed a statistically significant gain in science process skills, science concepts, and science-content knowledge by general education students in the experimental group when compared with students in the comparison group. Teacher participation in STEM Starters had a statistically significant impact on students' variability in posttest scores. Overall, experimental students demonstrated a gain of nearly 32 percentage points in science-content knowledge as compared to a 6.5 percentage point increase in science-content knowledge by their counterparts. Furthermore, experimental students gained 20 percentage points in science process skills (ability to design and control variables) as compared to an 11 percentage point gain by the comparison students.

Teacher participation in the treatment program accounted for approximately 7.5% of the variability in student posttest scores on the Fowler (process skills) assessment. Teacher facilitation of inquiry-type skills, such as the ability to design a science experiment, can be increased with content-focused sustained professional development (Supovitz, Mayer, & Kahle, 2000), thereby leading to increased student results as reported in this study. Results indicated that students in the STEM Starters treatment classrooms were better able to design science experiments when presented with a real-world problem, a key feature of science competency (Klahr & Li, 2005). These findings are consistent with those reported by Michaels et al. (2008), who highlighted the benefits of student participation in scientific practices such as questioning, hypothesizing, investigating, developing models, interpreting and analyzing data, linking conclusions to data, and sharing findings with others.

Teacher participation in the treatment program accounted for approximately 12% of the variability in student posttest scores on the curriculum-embedded content assessments, a result consistent with previous findings (Banilower et al., 2010; Showers, 1984). The inquiry-based curriculum implemented in the STEM Starter study allowed students to fully explore the age-appropriate content in an investigatory manner as recommended by the NRC (Duschl et al., 2007). Teachers also received training on specific curriculum units which impacts student achievement, a finding consistent with results reported by Schoen, Cebulla, Finn, and Fi (2003).

Teacher participation in the treatment program appeared to account for approximately 14% of the variability in student posttest scores on the curriculum-embedded concept assessment. These results indicated that students in the treatment classrooms were better able to make scientific connections using overarching concepts such as change and systems. These findings are consistent with previous research (Roth et al., 2011), which demonstrated increased student achievement when teachers had the ability to lead students in making content connections using big ideas.

Interim student performance data support the first-year implementation of rigorous differentiated science curriculum focused on improving science concept and content knowledge, and process skills for a student cohort which included general, special, and gifted education students. Overall findings are consistent with previous research which indicated that treatment effects including teacher adoption of inquiry-based instructional practices were seen in programs providing a minimum of 45 hours of teacher professional development annually (CCSSO, 2008) and that the pairing of peer coaching and professional development created optimal changes in teacher instruction (Neuman & Cunningham, 2009) contributing to higher student achievement.

## Limitations of the Study

Although care was taken to limit diffusion effects and confounding variables, isolated incidences occurred. A control teacher coached her students to do well on the posttests, and another control teacher reportedly dismissed the importance of the student pre-and posttest assessments and verbally communicated this message to his students. No such reported events took place in the experimental group.

## Scholarly Significance of the Study

The results of this study documented the effects of the combination of intensive professional development and the use of inquiry-based science instruction in the elementary classroom, including the benefits of using rigorous science curriculum with general education students. The STEM Starters intervention resulted in a statistically significant gain in science process skills, science concepts, and science-content knowledge by general education students in the experimental group when compared with students in the comparison group. Teachers' participation in the professional development provided the best model fit, accounting for approximately 7.5 to 14.15% of the variability in students' posttest scores.

These interim student performance data support the implementation of rigorous differentiated science curriculum focused on improving science concept and content knowledge and process skills. Given the concerns about elementary students and their teachers' lack of science concept and content knowledge, their failure to understand science processes deeply, and the limited research focused on early STEM education interventions, the effects documented at the conclusion of a one-year content-specific intervention are noteworthy. Moreover, the findings contradict many beliefs about science teaching and learning at the elementary grades. Due to the importance placed on the STEM disciplines and the calls from policy makers to build a pipeline for science and mathematics talent, STEM Starters is a timely catalyst for developing such opportunities for elementary teachers and ultimately for their young students.

## References

Adams, C., & Callahan, C. M. (1995). The reliability and validity of a performance task for evaluating science process skills. *Gifted Child Quarterly*, *39*, 14–20.

Appleton, K. (2008). Developing science pedagogical content knowledge through mentoring elementary teachers. *Journal of Science Teacher Education*, *19*, 523–545. doi: 10.1007/s10972-008-9109-4.

Banilower, E. R., Fulp, S. L., & Warren, C. L. (2010). *Science: It's elementary. Year four evaluation report*. Chapel Hill, NC: Horizon Research, Inc. Retrieved from http://www.horizon-research.com/reports/2010/sieyr4.php

Bickel, R. (2007). *Multilevel analysis for applied research: It's just regression*. New York: Gulford.

Bitan-Friedlander, N., Drefus, A., & Milgrom, Z. (2004). Types of "teachers in training": The reactions of primary school science teachers when confronted with the task of implementing an innovation. *Teaching and Teacher Education*, *20*, 607–619. doi: 10.1016/j.tate.2004.06.007.

Borman, G. D., Gamoran, A., & Bowdon, J. (2008). A randomized trial of teacher development in elementary science: First-year effects. *Journal of Research on Educational Effectiveness*, *1*, 237–264. doi: 10.1080/19345740802328273.

Bracken, B. A., Holt, K. A., Lee, M. L., McCormick, C. J., Reintjes, C. L., Robbins, J. I., & Stambaugh, T. L. (2008). *Perceptual assessment of science teaching and learning preliminary examiner's manual*. Williamsburg, VA: The College of William and Mary.

Brandwein, P. F. (1995). *Science talent in the young expressed within ecologies of achievement* (RBDM 9510). Storrs, CT: The National Research Center on the Gifted and Talented, University of Connecticut.

Buczynski, S., & Hansen, C. B. (2010). Impact of professional development on teacher practice: Uncovering connections. *Teaching and Teacher Education*, *26*, 599–607.

Committee on Science, Engineering, and Public Policy (2007). *Rising above the gathering storm: Energizing and employing America for a brighter economic future*. Washington, DC: National Academy Press.

Council of Chief State School Officers (2008). *Does teacher professional development have effects on teaching and learning* (Grant No. REC 0438359). Retrieved from http://www.ccsso.org/Documents/2008/Does_Teacher_Professional_Development_2008.pdf

Duschl, R., Schweingruber, H. A., & Shouse, A. (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington, DC: The National Academies Press.

Fowler, M. (1990). The diet cola test. *Science Scope*, *13*(4), 32–34.

Fulp, S. L. (2002). *Status of elementary school science teaching*. Chapel Hill, NC: Horizon Research, Inc. Retrieved from http://2000survey.horizon-research.com/reports/elem_science/elem_science.pdf

Gerard, L. F., Varma, K., Corliss, S. B., & Linn, M. C. (2011). Professional development for technology-enhanced inquiry science. *Review of Educational Research*, *81*(3), 408–448.

Goldston, D. (2005). Elementary science: Left behind? *Journal of Science Teacher Education*, *16*, 185–187.

Granger, E. M., Bevis, T. H., Saka, Y., & Southerland, S. A. (2009 April). *Comparing the efficacy of reform-based and traditional/verification curricula to support student learning about space science*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, CA: Garden Grove.

Heck, R. H., Thomas, S. L., & Tabata, L. N. (2010). *Multilevel and longitudinal modeling with IBM SPSS*. New York: Routledge.

Johnson, C. J., Kahle, J. B., & Fargo, J. D. (2006). Effective teaching results in increased science achievement for all students. *Science Education*, *91*, 371–383. doi: 10.1002/sce.20195.

Keeley, P. (2009). Elementary science education in the K-12 System. *NSTA WebNews Digest*. Retrieved from http://www.nsta.org/publications/news/story.aspx?id=55954

Kennedy, M. (1998). *Form and substance of inservice teacher education* (Research Monograph No. 13). Madison, WI: National Institute for Science Education, University of Wisconsin–Madison.

Kim, K. H., VanTassel-Baska, J., Bracken, B. A., Feng, A., Stambaugh, T., & Bland, L. (2012). Project Clarion: Three years of science instruction in Title 1 schools among K-Third grade students. *Research in Science Education*, *42*, 813–829. doi: 10.1007/s11165-011-9218-5.

Klahr, D., & Li, J. (2005). Cognitive research and elementary science instruction: From the laboratory, to the classroom, and back. *Journal of Science Education and Technology*, *14*(2), 217–238. doi: 10.1007/s10956-005-4423-5.

Little, P. F. B. (2005). Peer coaching as a support to collaborative teaching. *Mentoring and Tutoring*, *13*(1), 83–94.

Metz, K. E. (2008). Narrowing the gulf between the practices of science and the elementary school science classroom. *The Elementary School Journal*, *109*(2), 138–161.

Michaels, S., Shouse, A. W., & Schweingruber, H. A. (2008). *Ready, set science! Putting research to work in K-8 science classrooms*. Washington, DC: The National Academies Press.

National Research Council (1996). *National science education standards*. Washington, DC: National Academy Press.

National Science Board (2010). *Preparing the next generation of STEM innovators*: *Identifying and developing our nation's human capital*. NSB-10-33. Retrieved from http://www.nsf.gov/nsb/publications/2010/nsb1033.pdf

Neuman, S. B., & Cunningham, L. (2009). The impact of professional development and coaching on early language and literacy instructional practices. *American Educational Research Journal*, *46*(2), 532–566.

Penuel, W. R., Gallagher, L., & Moorthy, S. (2011). Preparing teachers to design sequences of instruction in earth systems science: A comparison of three professional development programs. *American Educational Research Journal*, *48*(4), 996–1025. doi: 10.3102/0002831211410864.

Pratt, H. (2007). *Science education's "overlooked ingredient": Why the path to global competitiveness begins in elementary school*. Retrieved from http://science.nsta.org/nstaexpress/nstaexpress_2007_10_29_pratt.htm

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rice, D. C. (2005). I didn't know oxygen could boil! What preservice and inservice elementary teachers' answers to "simple" science questions reveals about their subject matter knowledge. *International Journal of Science Education*, *27*(9), 1059–1082.

Robinson, A., Shore, B. M., & Enersen, D. L. (2007). *Best practices in gifted education: An evidence-based guide*. Waco, TX: Prufrock Press.

Roth, K. L., Garnier, H. E., Chen, C., Lemmens, M., Schwille, K., & Wickler, N. I. Z. (2011). Videobased lesson analysis: Effective science PD for teacher and student learning. *Journal of Research in Science Teaching*, *48*(2), 117–148.

Schoen, H. L., Cebulla, K. J., Finn, K. F., & Fi, C. (2003). Teacher variables that relate to student achievement when using a standards-based curriculum. *Journal for Research in Mathematics Education*, *34*(3), 228–259.

Showers, B. (1984). *Peer coaching: A strategy for facilitating transfer of training* (Report No. ED 271 849). Eugene, OR: Center for Educational Policy and Management, College of Education, University of Oregon.

Showers, B., & Joyce, B. (1996). The evolution of peer coaching. *Educational Leadership*, *53*(6), 12–16.

Supovitz, J. A., Mayer, D. P., & Kahle, J. B. (2000). Promoting inquiry-based instructional practice: The longitudinal impact of professional development in the context of systemic reform. *Educational Policy*, *14*(3), 331–356.

Supovitz, J. A., & Turner, H. M. (2000). The effects of professional development on science teaching practices and classroom culture. *Journal of Research in Science Teaching*, *37*, 963–980.

VanTassel-Baska, J. (1998). Planning science programs for high ability learners. *ERIC Clearinghouse on Disabilities and Gifted Education*. Retrieved from http://www.ericdigests.org/1999-3/science.htm

Yoon, C. H. (2009). Self-regulated learning and instructional factors in the scientific inquiry of scientifically gifted Korean middle school students. *Gifted Child Quarterly*, *53*, 203–216.

Yoon, K. S., Duncan, T., Lee, S., & Shapley, K. (2008). *The effects of teachers' professional development on student achievement: Findings from a systemic review of evidence*. Paper presented at the annual meeting of the American Educational Research Association, New York, NY.