RMLE Online
Research in Middle Level Education

Karen Weller Swanson, Ed.D., Editor
Mercer University
Atlanta, Georgia

2012 • Volume 35 • Number 9 ISSN 1940-4476

Effective Science Instruction: Impact on High-Stakes Assessment Performance

Carla C. Johnson
University of Cincinnati
Cincinnati, OH

Danhui Zhang
Utah State University
Logan, UT

Jane Butler Kahle
Miami University
Broomfield, CO

Abstract

This longitudinal prospective cohort study was conducted to determine the impact of effective science instruction on performance on high-stakes high school graduation assessments in science. This study provides powerful findings to support authentic science teaching to enhance long-term retention of learning and performance on state-mandated assessments. Students experienced some combination of zero to three effective teachers throughout their middle school experience. Findings revealed that all students who experienced effective science teachers who engaged students in inquiry-based science outperformed students who had less effective teachers. Additionally, those who had more effective teachers over time performed increasingly better. Implications for stakeholders will be discussed.

Introduction

Accountability is the driving force behind the focus of the educational system in the United States today. The No Child Left Behind Act of 2001 established that adequate yearly progress (AYP) for schools would be determined through annual growth in reading and mathematics. In 2007, states began mandated assessment of science, though science scores would not be included in the determination of AYP. High-stakes assessments have created a barrier to science education reform (Johnson, 2007; Settlage & Meadows, 2002). As a result, science instructional time, especially at the elementary and middle level, has been minimized, and little support has been provided to improve practice for teachers (Johnson, 2007; Barton, 2001; Shaver, Cuevas, Lee, & Avalos, 2007).

Students typically begin departmentalized study of science in middle school at Grade 6 or 7. At this level, science instruction is often focused on memorization of factual information in preparation for state-mandated multiple-choice science assessments (Shaver et al., 2007; Settlage & Meadows, 2002) and reading large selections from textbooks and conducting verification labs to prepare students for high school study (Anderson 2002). These strategies are not supported by research as effective practices. Additionally, as Keys and Bryan (2000) argued, "Teacher beliefs about students and learning, such as ability levels or the need for drill and practice, represent obstacles to inquiry-based instruction" (p. 635). Unfortunately, many secondary teachers feel students cannot gain content knowledge through standards-based strategies such as inquiry (Johnson, 2009, 2006).

Focus on Inquiry as the Central Strategy

In the United States, the National Science Education Standards (NSES) called for an instructional shift in science teaching (National Research Council, 1996), moving away from teacher-centered, less effective instruction and toward the use of more effective instructional practices, including inquiry and exploration driven by student interests in the context of the real world (Anderson, 2002; NRC, 1996). These strategies have been a focus of reform in many other countries, as a growing emphasis on science, technology, engineering, and mathematics STEM skills has emerged concurrently. Effective instruction, as detailed in the NSES (National Research Council, 1996), includes "inquiry into authentic questions generated from student experiences" (p. 31) and has been used as a guide for science instruction in the United States for more than a decade. According to the NSES, "learning science is something that students do, not something that is done to them" (p. 20). In an inquiry-based classroom, students should be engaged in inquiry through describing objects and events, posing questions, gathering data to construct explanations, conducting investigations, and considering alternative explanations. Students "actively develop their understanding of science by combining scientific knowledge with reasoning and thinking skills" (p. 2). Science is for all students and is an active process grounded in how science contributes to culture (NRC, 1996).

Effective Instruction and Student Performance in Science

The science education research base continues to support the strategies called for in the NSES as effective ways to enhance student learning of science (Johnson, Kahle, & Fargo, 2007b; Wood, Lawrenz, Huffman, & Schultz, 2006). A recent meta-analysis of the science education literature revealed eight categories of effective science teaching strategies associated with increased student achievement, which included: enhanced context strategies, collaborative learning strategies, questioning strategies, inquiry strategies, manipulation strategies, assessment strategies, instructional technology strategies, and enhanced material strategies (Schroeder, Scott, Tolson, Huang, & Lee, 2007). Enhanced context, collaborative learning, questioning, and inquiry were the top four of the eight strategies, though Schroeder and colleagues (2007) noted that effective instruction such as called for in the NSES requires a combination of multiple strategies.

Teachers who create effective science instructional environments contextualize instruction to appeal to student interests and present new science concepts (Rivet & Krajick, 2008). Contextualized science instruction uses students' prior knowledge and experiences to foster understanding (Rivet & Krajick, 2008). Guided inquiry, problem-based learning, and project-based learning are some strategies science teachers use to bring state-mandated science content alive for students through their own personal lenses. According to Rivet and Krajcik, "Contextualizing instruction is believed to promote transfer of science ideas to other contexts because students learn to relate content ideas to problems and situations meaningful in their lives and [in] the real world" (p. 82). Traditional, teacher-centered instruction relies heavily on text, lecture, and other structured activities that provide few opportunities for students to feel a connection to what they are learning in science (Johnson, 2007; Anderson 2002).

Inquiry strategies are student centered, with students answering scientific questions through investigation. Bateman (1990) argued, "The problem with the method of inquiry is that not enough teachers use it, not enough understand the power inherent in inquiry, not enough see their job as other than transmitting information" (p. XV). Collaborative learning strategies include grouping students around meaningful projects or tasks. Questioning strategies include engaging students in more highly cognitive, probing questions. Teacher content knowledge is a factor in the level of questioning and engagement in the science classroom (Anderson, 2002; Anderson & Helms, 2001; Kennedy, 1998). Some authors suggest that science teachers, in general, are inadequately prepared in both content knowledge and scientific inquiry experiences (Berns & Swanson, 2000).

Few research studies have linked the use of effective science instruction, including inquiry and contextualized science instruction, to school level achievement on state standardized assessments (Johnson et al., 2007a; Rivet & Krajcik, 2008; Wood et al., 2006). Overall, results have been mixed, as some have found little to no difference longitudinally (Shymansky, Yore, & Anderson, 2004), and others have demonstrated gains on pre/post assessments and annual state assessments during the duration of professional development programs (Czerniak, Beltyukova, Struble, Haney, & Lumpe, 2005; Johnson, Kahle, & Fargo, 2007a; Kahle, Meece, & Scantlebury, 2000; Ruby, 2006). Currently, a dearth remains in the science education literature on long-term performance for students who experience effective science learning environments including the use of contextualized instruction, inquiry, questioning, and cooperative learning as main foci.

Since most state assessments focus on recall of content knowledge through de-contextualized facts, teachers are apprehensive to use strategies other than basic recall as primary methods for teaching science to children (Keys & Bryan, 2000; Rivet & Krajcik, 2008). This study is a purposeful examination of effective science instructional strategies including contextualized instruction, inquiry, questioning, and cooperative learning and the impact on student learning and retention of learning of science longitudinally. Performance on a state-mandated graduation assessment, as well as a widely used middle school assessment modeled after National Assessment of Educational Progress (NAEP), will be used to compare students of teachers who use more effective practices (NRC, 1996; Schroeder et al., 2007) versus teachers who use more traditional practices (Anderson, 2002) to teach science to determine which learning environments are more conducive to learning and retention.

The following research questions guided this study:

1. Does student participation in effective science instructional environments in middle school influence performance on a high-stakes state graduation assessment in science?
2. Does student participation in effective science instructional environments have the same impact on Caucasian and Non-Caucasian student performance on state graduation assessment in science?

Effective instruction was defined using the Local Systemic Change Classroom Observation Protocol (Horizon Research, 2001) rubric score of a four or five. A score of four describes accomplished, effective instruction as,

> Instruction is purposeful and engaging for most students. Students actively participate in meaningful work (e.g., investigations, teacher presentations, discussions with each other or the teacher, reading). The instruction is well-designed and the teacher implements it well, but adaptation of content or pedagogy in response to student needs and interests is limited. Instruction is quite likely to enhance most students' understanding of the discipline and to develop their capacity to successfully do science. (p.11)

A score of five on the rubric describes exemplary instruction as,

> Instruction is purposeful and all students are highly engaged most or all of the time in meaningful work (e.g., investigations, teacher presentations, discussions with each other or the teacher, reading). The lesson is well designed and artfully implemented, with flexibility and responsiveness to students' needs and interests. Instruction is highly likely to enhance most students' understanding of the discipline and to develop their capacity to successfully do science. (p.11)

All teachers were observed and scored according to this protocol to determine teacher effectiveness for this study.

Theoretical Framework

Situated learning was used as the theoretical foundation for this study (Brown, Collins, & Duguid, 1989; Lave & Wenger, 1988; Resnick, 1987). The fundamental basis of situated learning theory is that knowledge should not be the transmission of de-contextualized facts from teacher to student. Rather, learning is a social process where knowledge is co-constructed within a learning environment mirroring the real world (Lave & Wenger, 1988). This framework provides an authentic approach for science instruction in which enhanced context, collaborative learning, questioning, and inquiry are included in a real-world situated approach to teaching important concepts (Resnick, 1987; Schroeder et al., 2007).

Methodology

Participants

This study was conducted from 2002 to 2007. Pseudonyms will be used for all teachers and schools referred to in this study. The study included all 176 students who began sixth grade at Star Middle School in 2002. Star Middle School is located in a Midwestern state and is one of two middle schools that feed into Star High School. Six students whobegan study at Star Middle School did not continue to Star High School and were dropped from this study. The study group cohort for this research was comprised of all sixth grade students who were in sixth grade in 2002. The demographics for the student group are found in Table 1.

There were 11 science teachers at Star Middle School, five sixth grade (including two half-time job share teachers who taught only science and one teacher who taught a science and social studies split), three seventh grade, and three eighth grade teachers.

This study examines the performance of tenth graders on the Ohio Graduation Test (OGT) with respect to their middle school instructional environments (i.e., exposure to effective teaching of science or not). All students in this study were from the same middle school, which was one of two feeder schools for the only high school in the district. All freshmen in this district take physical science in the freshman year from the same teacher and are placed randomly into these classes. Sophomore year students take biology, again from the same teacher and are randomly placed. Advanced placement courses and science options were not open to students in this school until junior year. Therefore, all classes were deemed the same, as they all were taught by the same teacher and the instructional mode was the same in all sections of each class (verified by building and district administration). Both courses used standard textbooks and completed accompanying verification laboratories. Inquiry projects were limited to one per semester (two per year) in each class. The remainder of the year, students worked with a lab partner to complete weekly labs, and the remainder of the class time was devoted to lecture and note taking. Students were tracked over time using their student identification numbers, and data were provided on the state assessment by the district administration.

Research Design

A prospective cohort study was conducted, during which teacher classrooms were observed each year while students participated in their classes (using the Local Systemic Change [LSC] Classroom Observation Protocol). During the year in which teachers were observed (i.e., sixth grade teachers were evaluated in year one, seventh grade teachers in year two), classrooms were visited four times via random visits by two raters. Raters collected data on instructional environment using the LSC protocol. Teachers were informed of the day of the visit but not the particular class period. The visits were conducted in September, November, February, and April. Training was provided for raters (both veteran middle school science teachers with advanced degrees in science education) according to the guidelines established by Horizon Research, Inc., which developed the instrument and conducted the evaluation of the LSC programs for the National Science Foundation. Inter-rater reliability ranged from .83 to .88 across the years of this study. All 11 participating teachers were interviewed following the fourth and final observation for them to reflect on their science instructional environments by the primary author of this paper. Interview data analysis (the majority of which has been reported in other manuscripts) was conducted by two science educators and (Johnson, 2009, 2007).

Student participants from Star Middle School were followed from sixth grade to tenth grade (n = 176). Student participants completed the Discovery Inquiry Test (DIT) in sixth, seventh, and eighth grades. There was no significant difference at baseline

Table 1
Demographics for Students in Study (n = 176) Compared to Entire School Population

	Socioeconomic Status	**Caucasian %**	**African American %**	**Latino %**	**Other %**	**Gender %**
Overall School	39	65	30	3	2	48
Study Group	38	73	22	3	2	48

between students who had been placed with effective teachers and those placed with non-effective sixth grade teachers. Analyses of data from sixth, seventh, and eighth grades indicated a significant difference in student science achievement for students who participated in effective classrooms compared to those who had less or no exposure (Johnson, et al., 2007a).

Instruments

The LSC Classroom Observation Protocol (Horizon Research Inc., 2002) was used as the instrument to determine degree of standards-based instruction in environments in four areas: (a) design of lesson, (b) implementation of lesson, (c) science content of lesson, and (d) classroom culture. Teachers were observed during the year that students were placed with them to get an overall teaching quality measure for that particular year. For each observation, a teacher received a score from 1 (worst) to 5 (best) on each of the four subscales of the protocol (design, implementation, science content, and classroom culture) Appendix A. The LSC Classroom Observation Protocol has been used widely in National Science Foundation (NSF) funded State Systemic Initiatives (SSI) and in other science education research over the past ten years (Johnson, et al 2007a; Johnson, Kahle, & Fargo, 2007b). Inter-rater reliability for the current study was ranged from .83 to .88 across the years of the study.

The Discovery Inquiry Test (DIT) in Science assessment was developed in 1994 by members of Ohio's SSI academic leadership teams, university science faculty, and other Ohio teachers for use in the SSI funded program. The DIT included 29 items: 11 focusing on life science, 8 on physical science, 6 on earth and space science, and 4 on the nature of science. It was validated by a national and international expert panel of science educators. A total of 20 points were possible on the DIT score. It has revealed a high internal consistency reliability; Cronbach's alpha coefficient = .94. Items from the National Assessment of Educational Progress 1990 and 1992 public release version focused on measuring student ability to analyze and interpret data, to extrapolate from one situation to another, and to use conceptual understanding were included (Kahle, et al., 2000).

The Ohio Graduation Test (OGT) is considered to be a high-stakes assessment because all students in the state are required to successfully pass all components of the test to graduate high school in Ohio. The OGT aligns with state academic content standards in science, mathematics, reading, writing, and social studies and is administered first in tenth grade, then in subsequent years until the student has successfully passed the assessment. There are approximately 35 multiple-choice and eight open-response questions on each section (content area) of the assessment. There are four categories in which students are placed, depending on their performance. The categories and score ranges are *advanced* (432–508), *accelerated* (415–431), *proficient* (400–414), *basic* (385–399), and *limited* (251–384). Students who score advanced, accelerated, and proficient "pass" the assessment. Basic and limited scores are considered unsuccessful. OGT scores by student identification number were provided by the participating school districts in this study. Appendix B has sample items from the DIT and OGT.

Data Analysis

The first step in the data analysis was to examine the teacher observational data to determine which teachers were effective and which were not categorized as effective. The four observation scores for each teacher were compiled into a mean score. Teachers who had a score of three or higher in all four domains of the LSC protocol were deemed to be effective. Teacher interview data were used, along with field notes from the observations, to provide evidence of effective teaching components including enhanced context, collaborative learning, questioning, and inquiry (Schroeder et al., 2007) couched in the situated learning theory (Lave &Wenger, 1988).

Next, student data were analyzed, and four variables were created to categorize student exposure to effective science teaching:

1. Whether the student had at least one effective teacher in sixth, seventh, or eighth grade.
2. Whether the student had effective teachers in both seventh and eighth grades.
3. Whether the student had at least one effective teacher in seventh or eighth grade.
4. The total number of effective science teachers in sixth through eighth grades.

These variables were included in separate analyses. The first variable was important to classify students if they had received an effective teacher during the three-year span of middle school. The second variable examined whether students had effective teachers in seventh and eighth grades and, in particular, looked for progress between two groups—those who had an effective sixth grade teacher or those who did not to see if similar gains could be made regardless of year of exposure to effective teaching. The third variable looked at whether the student only had an effective teacher in sixth grade or if there was an additional effective teacher afterward. The final variable examined the total number of effective teachers. Ethnicity/race was a dichotomous variable (Caucasian versus non-Caucasian).

The process of analyzing data consisted of two steps. First, descriptive statistics were computed on students' eighth and tenth grade science scores and other demographic variables. Second, sequential (or hierarchical) ordinary least-squares regression analyses were conducted to investigate what factors influenced student tenth science scores. A sequential approach was used to adjust for the effect of demographic characteristics and middle school science achievement on the relationship between exposure to effective teachers in the middle school and long-term science achievement in high school. In the first model, student race and gender were entered as predictors. In n the second model, mean-centered eighth grade science scores were added. The third model differed from models one and two by the inclusion of each of the teaching effectiveness variables in turn or one at a time. Substantively meaningful interactions among the predictors were tested (e.g., teacher effectiveness and student ethnicity/race).

Diagnostic analyses were performed on model three and indicated problems with influential observations and nonlinearity but no collinearity problems. Specifically, eight influential observations were identified. There were no substantial differences when the final model was computed with and without those influential observations included. A quadratic trend was observed between tenth grade students' science scores and number of effective teachers, so a quadratic (centered) version of this predictor was added in a fourth sequential regression model. Therefore, the final or fourth regression model included student ethnicity/race, gender, eighth-grade scores, the total number of effective teachers, and the total number of effective teachers squared as predictors of students' tenth grade science scores. All statistical analyses were conducted using the R language for statistical computing (R Development Core Team, 2006).

Results

Identification of effective vs. non-effective teachers

All teachers in this school who were included in this study were full-time science teachers, responsible for teaching five sections of science each day, with 60-minute class periods. All teachers had their own classrooms and had been provided with resources to teach science, including textbooks, equipment, consumables, technology, and had administrative support to teach science effectively. Teachers were provided with individual planning time (one class period) each day. All teachers in this study were certified to teach science through either a seventh to twelfth grade content specific certification or first to eighth grade generalist certification with science endorsement.

Sixth grade teachers. At the sixth grade level, two teachers, Mrs. Bingham and Mr. Adams, were found to be effective—both at the Level 4 range on the overall LSC rubric. Mrs. Bingham had four years' experience teaching science and a first to eighth grade certification. During all observations of Mrs.Bingham's instruction, students were actively engaged in guided inquiries in which the teacher provided the question to explore in the context of the real world and the materials needed. Her room was set up with individual student desks pushed together in collaborative clusters of four, and each group member had assigned roles (recorder, materials, facilitator, or researcher). Each science lesson was at least 60 minutes long, and the last ten minutes was spent in whole-class discourse with sharing of findings and challenging of ideas or in a "spot check" for longer investigations to share the day's progress. Mrs. Bingham stated that having students "doing" science on a daily basis "enables them to develop background knowledge and experiences to tie the new concepts to." She would introduce vocabulary prior to the guided inquiry so that students could use it in their ongoing discourse during the investigation.

Mr. Adams, a second year teacher with a first to eighth grade certification, said he had a structured classroom in which students would read excerpts from National Geographic Science Xpeditions books to engage them in a concept. Then students would have a few minutes of discourse to reveal prior knowledge and situate the new concepts in the context of student interests. Adams had students pose questions that they would like to learn more about on the topic and give them time on the Internet to explore and come back and share with their group, then with the whole class. Adams then used a problem-based learning scenario grounded in a real-world context for students to explore. One lesson observed included building bridges from Popsicle sticks with various constraints and testing them to see how much weight they could support. In another lesson, the class was broken into eight groups, each with the challenge of designing and creating a mode of transportation. One group constructed a mini hot air balloon from scratch during a weeklong exploration and competition. Adams said that students "must learn how to problem solve and use critical thinking skills like we do in the real world, and the best way to do this is through doing science."

In contrast, Mr. Moore, another sixth grade teacher with eight years experience and a first to eighth grade license, who was not found to be effective (in the Level 1 range on the LSC rubric), spent each of the four class periods going over assigned homework questions and calling on students to give the answers. Then, students took turns reading aloud and began working in class on the assigned homework questions.

Seventh grade teachers. Mrs. Hamilton, a teacher with 18 years experience and first to eighth grade certification, was the only effective seventh grade teacher (at the Level 4 range on the LSC Rubric). Mrs. Hamilton, when asked what the most effective science environment looked like, responded, "Inquiry. The kids are actively participating. They want to learn. They are not bored. They want to be part of it. They want to learn it. And that's the way to learn." Her classroom was set up to enable guided inquiry ona daily basis, as she provided students with multiple questions to explore and various materials for students to choose from, including those brought from home. Students could choose questions of interest to work on. One of the inquiries observed was a problem-based learning unit in which students chose from a heaping pile of butter tubs, oatmeal cans, plastic cups, cardboard tubes, pins, and so forth to create their own "rock band." They had to make instruments from these household materials that would deliver the right pitch and sound to accompany their song of choice. Students played a concert for their peers with their instruments and learned that music can come from multiple resources. Mrs. Hamilton followed this activity with a study of instruments used by Native Americans and others before instruments became mass produced.

Mrs. Davis and Mr. Brown were the other two seventh grade science teachers. Mr. Davis had four years experience and a seventh to twelfth grade biology certification (at the Level 2 range on the LSC rubric). Mr. Brown had 16 years of experience and a first to eighth elementary certification (at the Level 2 range on the LSC rubric). Observations of their classes revealed a majority of teacher- and textbook-centered instruction. One day a week, each teacher had students do a step-by-step verification lab from the book, with little opportunity for student questioning or discourse about the one real answer. Mr. Davis said that effective science instruction was, "In my mind, trying to make concepts that seem to be concrete into, you know, so that they can understand what they've got."

Eighth grade teachers. Mrs. Green was the eighth grade teacher with a first to eighth grade certification who had an effective instruction (at the Level 4 range on the LSC rubric). Each day, Mrs. Green had students involved in exploring science phenomena. Mrs. Green used multiple inquiries in a station setting, which, she explained, provided the context of what scientists do. At first glance, it seemed to be a more traditional science laboratory, but a closer look revealed student groups spending multiple days at each station completing an inquiry, generating their own questions, and engaging in contextualized learning. Mrs. Green explained that she wanted to model how scientists worked for many days on one problem, and she also wanted her classroom to have a real-world science feel to it. Open stations enabled groups to work at their own pace and to choose their investigations from a set of options based on their interests. Mrs. Green would spend the entire class period "working the room" and supporting the various groups in their inquiries. She said, "In the real world, you don't always have a room full of 30 people working on the same thing. Scientists have multiple projects going on in the same room at the same time. That is how I want my room to feel." At the end of each inquiry cycle, groups reported back and engaged in discourse with the whole class on their findings, similarities, and differences.

Mrs. Hillard ,who holds a seventh to twelfth grade biology certification, and Mrs. Brown, who holds a seventh to twelfth grade physical science certification, were the other two eighth grade science teachers. Both teacher registered at the Level 2 range of the LSC protocol. Their classrooms were much like Mrs. Davis's seventh grade classroom: textbook centered, drill and practice with focus on vocabulary memorization, and the occasional cookbook-style lab. Mrs. Hillard said, "We have to get them ready for high school, where they are going to have to do a lot of reading and completing lab reports. I think that in eighth grade that is our biggest job, to get them ready for that." Mr. Brown was the science fair coordinator for the school and also shared a similar view of the teaching of science: "Our students need experiences with the scientific method and proposing hypothesis and testing them. They get to do that through our science fair here each year." Mr. Brown seemed to believe that once a year was often enough for students to engage in inquiry.

Student Performance Based on Middle School Experiences

Based on the results from the LSC protocol, it was determined that 32% of the 176 participating students did not have an effective teacher (*n* = 57), while 47% had one effective teacher (*n* = 82); 18% had two effective teachers (*n* = 31), and 3% had three effective teachers (*n* = 6).

The mean tenth grade science score on the Ohio Graduation Test (OGT) for the entire sample (students of effective and non-effective teachers) was 429.35 (*SD* = 24.75) of a maximum score of 608. All students who experienced one or more effective teachers in middle school passed the science portion of the OGT. Conversely, students who did not have an effective science teacher and failed the science portion of the test equated to 11% of the entire study group (*M* score = 382.40, *SD* = 11.16), while 89% of the entire study group of students (including all students with an effective teacher) passed the OGT (*M* score = 435.40, *SD* = 18.79). The results of the sequential regression analyses are summarized in Table 2. There was no effect of gender or ethnicity/race in any of the models for OGT scores.

In eighth grade, the overall study group of students took the DIT in science and scored a mean of 13.28 (*SD* = 3.40) on a 19-point scale. Students' eighth grade science scores significantly predicted the outcome in Model 2 (*b* = 1.95, p < .0001). However, when accounting for effective teachers (added in Model 3), students' eighth grade science scores no longer significantly predicted tenth grade science scores. Results of the four versions of Model 3 were similar (a different form of the teacher effectiveness variable was included in each version). For instance, the R^2 for the four regression models in which effective teacher was coded differently were as follows:

Table 2
Results of Sequential Regression Analysis of Tenth Grade Students' Science Performance

Predictors	**Model 1** ***b*** **(t-statistic)**	**Model 2** ***b*** **(t-statistic)**	**Model 3** ***b*** **(t-statistic)**	**Model 4** ***b*** **(t-statistic)**
Constant	437.73	439.62 (44.14***)	434.81	436.14 (113.95***)
Ethnicity/Race	(63.33***) (-1.00)	(-1.38)	(110.87***) (-1.03)	(-0.68)
Gender	-4.34 (-1.16)	-3.82 (-1.06)	-1.82 (-0.87)	-1.67 (-0.83)
Eighth grade science score		1.95 (3.66**)	-0.09 (-0.29)	-0.12 (-0.39)
Number of SBIE teachers			22.95 (16.44***)	24.89 (16.87***)
Number of SBIE teachers^2				-4.52 (-3.30**)
R^2	.00	.07	.64	.66

** *p* < .01, *** *p* < .001, **** *p* < .0001

1. Any versus no effective teacher: .59
2. Effective teacher in both seventh and eighth grade: .24
3. Effective teacher in either seventh or eighth grade: .57
4. Total number of effective teachers: .64

The number of effective teachers a student had was chosen as a criterion for interpretation because it contained more informative data about exposure to effective teaching than the other variables. No interactions among predictors in this area were found to be statistically significant, indicating that the relationship between exposure to effective teachers and student outcomes in tenth grade did not vary as a function of student race/ethnicity or gender.

To appropriately model nonlinearity in the relationship between tenth grade students' scores and total number of effective teachers in middle school, the number of effective teachers was entered into a fourth model as a quadratic or squared term, in addition to its lower order (non-squared) term. The quadratic term accounted for a significant amount of unique variation in tenth grade science scores ($b = -4.52$, $p < .001$) above and beyond student gender, ethnicity/race, and eighth grade scores (which were non-significant). This result indicated that, although student science scores improved with increased exposure to effective teachers (more than one year), the rate of improvement flattened out as the number of effective teachers in the students" histories increased. Number of effective teachers (non-squared) remained a significant predictor ($b = 24.89$, $p < .0001$). Overall, in the final model, 66% of the variance in students' tenth grade science scores was explained by the combination of predictors. Table 3 presents descriptive statistics for tenth grade science achievement as a function of the number of effective teachers the student experienced during middle school.

These analyses address research questions one and two. Exposure to effective science teachers in middle school is positively associated with better student outcomes in the tenth grade; the more exposure, the better the performance. Additionally, there was no differential performance among students of different races/ethnicities, as indicated by the lack of a significant interaction between total number of effective teachers during middle school and race/ethnicity (see Table 4).

Discussion and Implications

Situated learning theory provided an interesting lens with which to conduct this study (Lave & Wenger, 1991). One of the fundamental issues in science education reform today is encouraging science teachers to use more effective instructional strategies including contextualized teaching, inquiry, questioning, and cooperative learning to teach science (Bateman, 1990; Keys & Bryan, 2000). This approach is supported by situated learning theory, which proposes learning as a social process, co-constructed and not focused on transmission of isolated facts (Brown et al., 1989). Many teachers fear the implications of devoting extended time for investigations may be lower performance on state-mandated assessments, due to less time for drill and practice (Anderson, 2002; Johnson, 2006; Lee & Luykx, 2006; Rivet & Krajcik, 2008). Further, we know little about the long-term influence of instructional environments on student learning and retention of science concepts (Shaver et al., 2007). Often, teachers struggle with students who have retention issues when they experience traditional

Table 3
Descriptive Statistics for Students' Tenth Grade Science Scores by Exposure to SBIE Teachers in Middle School

Number of SBIE Teachers	*M*	*SD*	*n*
0	402.42	17.49	57
1	436.66	13.96	82
2	452.06	9.58	31
3	467.83	10.48	6

Table 4
Tenth Grade Science Scores by Teacher Effectiveness and Student Race/Ethnicity

	Race/Ethnicity	
Number of SBIE Teachers	**White**	**Non-White**
None	*n* = 39 *M* = 402.71 *SD* =17.61	*n* = 18 *M* = 401.17 *SD* =17.39
One or More	*n* = 89 *M* = 442.06 *SD* = 16.48	*n* = 30 *M* = 443.17 *SD* =11.64

science instruction, which does not allow students opportunities to engage in "doing science" (Anderson, 2002).

Science teachers today face great challenges when deciding how to deliver science instruction. Large amounts of curriculum and standards are assigned to each grade level, and teachers often opt to cover more items in less depth, resulting in more teacher-centered instruction and fewer opportunities for inquiry (Barton, 2001). In this study, students who experienced effective science teachers in middle school outperformed their peer groups in eighth grade. (Interestingly, both groups were not significantly different at the sixth grade baseline).

On the tenth grade OGT, students who had experienced effective science teachers in middle school outperformed those who had no exposure. In fact, as students experienced more effective teachers, their performance level increased (maximum of three teachers, one each year), and all students who experienced at least one effective teacher passed the OGT on the first attempt. The misconception held by most science teachers is that students will do better on mandated state science assessments if they spend the majority of time on drilling vocabulary and practicing test items repeatedly throughout the school year. Findings in this study are to the contrary—as students who were exposed to more student-centered environments that contextualized science in the real world outperformed the other students. Additionally, students from all ethnic/racial backgrounds benefited from effective teachers and outperformed students who did not have this experience. Effective science instruction in middle school may serve as a foundation for future science learning in high school coursework.

Arguably, this study provides the basis for considering the impact of state-mandated assessments on the teaching of science as well as how teachers are preparing students for the test. Engaging students in science learning situated in the real-world is more effective than teaching scientific vocabulary and concepts out of a context of inquiry into real-world problems and questions. (Settlage & Meadows, 2002). Though many teachers feel compelled to "cover" curriculum to ensure that all topics are taught prior to the assessment, this spreading of peanut butter very thin on the slice of bread does not seem to benefit students. Retention is minimal and students fail to gain the ability to think critically, work collaboratively, solve problems, and ask questions about the world around them (Johnson, 2006; Lave & Wenger, 1988; Lee & Lukyx, 2007).

Many of the teachers who are not using an effective approach to teaching science likely need professional development support to retool. Wojnowski, Bellamy, and Cooke (2003) suggested, "The only way to combat the problem of low student achievement is to improve the quality of science teaching in America's schools" (p. 24). Possibly, research such as this study will shed more light on the critical need to reform science classrooms and support teachers through the process (Desimone, Porter, Garet, Yoon, & Birman, 2002) as well as educate science teachers on the benefits of moving away from a test-preparation mode of teaching to a teaching-for-conceptual-understanding focus (Rivet & Krajcik, 2008). Many secondary science teachers are content experts, but those who have been in the field for more than ten years may not have had the opportunity to learn about the pedagogical strategies included in the NSES, as their teacher education training was completed prior to the enactment of the NSES. It could be concluded from this study that teacher experience and content preparation have little to do with whether or not a teacher is effective. Motivation to engage students in authentic science was a key factor in effective teachers' pedagogy (Johnson, 2009, 2006; Keys & Bryan, 2000). It is clear from the literature, and also supported by findings from this study, that science teacher professional development is needed to change teacher beliefs about pedagogy and to support teachers as they make the transition from teacher-centered to student-centered, contextualized, reform-based classrooms (Johnson, 2006, 2009; Keys & Bryan, 2000; Ruby, 2006).

The transformation of science instructional environments to a more inquiry-based, student-centered approach could enable many more students from all backgrounds to develop the rich science background that is often lacking (Ruby, 2006; Kahle et al., 2000). This transformation may translate into not only enhanced science performance for students but also an increased interest in STEM fields and more students who are better prepared to lead the 21st-century workforce. Basu and Barton (2007) found that student interest in science can be enhanced when science aligns with their own futures, supports social relationships, and supports students' sense of agency on the purpose of science. Contextualizing science instruction, as enabled through effective teaching of science, engaged students in focusing on these (Rivet & Krajcik, 2008).

References

Anderson, R. D. (2002). Reforming science teaching: What research says about inquiry. *Journal of Science Teacher Education, 13,* 1–12.

Anderson, R. D., & Helms, J. V. (2001). The ideal of standards and the reality of schools: Needed research. *Journal of Research in Science Teaching, 38,* 3–16.

Basu, S. J., & Barton, A. C. (2007). Developing a sustained interest in science among urban minority youth. *Journal of Research in Science Teaching, 44,* 466–489.

Barton, A. C. (2001). Science education in urban settings: Seeking new ways of praxis through critical ethnography. *Journal of Research in Science Teaching, 38*(8), 899–917.

Bateman, W. L. (1990). *Open to question: The art of teaching and learning by inquiry.* San Francisco, CA: Jossey-Bass.

Berns, B. B., & Swanson, J. (2000). *Middle school science: Working in a confused context.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED444944)

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18,* 32–42.

Czerniak, C. M., Beltyukova, S., Struble, J., Haney, J. J., & Lumpe, A. T. (2005). Do you see what I see? The relationship between a professional development model and student achievement. In R. E. Yager (Ed.), *Exemplary science in grades 5–8: Standards-based success stories* (pp. 13–43). Arlington, VA: NSTA Press.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of professional development on teachers' instruction: Results from a three-year longitudinal study. *Educational Evaluation and Policy Analysis, 24,* 81–112.

Horizon Research, Inc. (2002). *Local systemic change classroom observation protocol.* Retrieved from http://www.horizon-research.com

Johnson, C. C. (2006). Effective professional development and change in practice: Barriers teachers encounter and implications for reform. *School Science and Mathematics, 106*(3), 1–12.

Johnson, C. C. (2007). Effective science teaching, professional development, and No Child Left Behind: Barriers, dilemmas, and reality. *Journal of Science Teacher Education, 18*(2), 133–136.

Johnson, C. C. (2009). An examination of effective practice: Elimination of achievement gaps in science. *Journal of Science Teacher Education, 20*(3) 287–306.

Johnson, C. C., Kahle, J. B., & Fargo, J. (2007a). A study of sustained, whole-school, professional development on student achievement in science. *Journal of Research in Science Teaching, 44*(6), 775–786.

Johnson, C. C., Kahle, J. B., & Fargo, J. (2007b). Effective teaching results in increased science achievement for all students. *Science Education, 91*(3), 371–383.

Kahle, J. B., Meece, J., & Scantlebury, K. (2000). Urban African-American middle school science students: Does standards-based teaching make a difference. *Journal of Research in Science Teaching, 37*(9), 1019–1041.

Kennedy, M. M. (1998). Education reform and subject matter knowledge. *Journal of Research in Science Teaching, 35,* 249–263.

Keys, C. W., & Bryan, L. A. (2000). Co-constructing inquiry-based science with teachers: Essential research for lasting reform. *Journal of Research in Science Teaching, 38,* 631–645.

Lee, O., & Luykx, A. (2006). *Science education and student diversity: Synthesis and research agenda*. New York, NY: Cambridge University Press.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge, UK: Cambridge University Press.

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

R Development Core Team. (2006). R: A language and environment for statistical computing. Retrieved from http://www.R-project.org

Resnick, L. B. (1987). Learning in school and out. *Educational Researcher, 16,* 13–20.

Rivet, A. E., & Krajcik, J. S. (2008). Contextualizing Instruction: Leveraging students' prior knowledge and experiences to foster understanding of middle school science. *Journal of Research in Science Teaching, 45,* 79–100.

Ruby, A. (2006). Improving science achievement at high-poverty urban middle schools. *Science Education, 48,* 1005–1027.

Schroeder, C. M., Scott, T. P., Tolson, H., Huang, T. Y., & Lee, Y. H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the USA. *Journal of Research in Science Teaching, 44,* 1436–1460.

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,* 114–127.

Shaver, A., Cuevas, P., Lee, O., & Avalos, M. (2007). Teachers perceptions of policy influences on science instruction with culturally and linguistically diverse elementary students. *Journal of Research in Science Teaching, 44,* 725–746.

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

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.

Wojnowski, B. S., Bellamy, M. L., & Cooke, S. (2003). A review of literature on the mentoring and induction of beginning teachers with an emphasis on the retention and renewal of science teachers. In J. Rhoton & P. Bowers (Eds.), *Science teacher retention: Mentoring and renewal* (pp. 23–33). Arlington, VA: NSTA Press.

Wood, N. B., Lawrenz, F., Huffman, D., & Schultz, M. (2006). Viewing the school environment through multiple lenses: In search of school-level variables tied to student achievement. *Journal of Research in Science Teaching, 43,* 237–254.

Appendix A

Local Systemic Change Classroom Observation Protocol Capsule Description of the Quality of the Lesson

Level 1: Ineffective Instruction

There is little or no evidence of student thinking or engagement with important ideas of science. Instruction is *highly unlikely* to enhance students' understanding of the discipline or to develop the capacity to successfully do science. Lesson was characterized by either

a. **Passive learning**
   Instruction is pedantic and uninspiring. Students are passive recipients of information from the teacher or textbook; material is presented in a way that is inaccessible to many of the students.

b. **Activity for activity's sake**
   Students are involved in hands-on activities or other individual or group work, but it appears to be activity for activity's sake. Instruction lacks a clear sense of purpose and/or a clear link to conceptual development.

Level 2: Elements of Effective Instruction

Instruction contains some elements of effective practice, but there are *serious problems* in the design, implementation, content, and/or appropriateness for many students in the class. For example, the content may lack importance and/or appropriateness; instruction may not successfully address the difficulties that many students are experiencing. Overall, the instruction is *very limited* in its likelihood to enhance students' understanding of the discipline or to develop their capacity to successfully do science.

Level 3: Beginning Stages of Effective Instruction

Instruction is purposeful and characterized by quite a few elements of effective practice. Students are, at times, engaged in meaningful work, but there are *weaknesses*, ranging from substantial to fairly minor, in the design, implementation, or content of instruction. For example, the teacher may short-circuit a planned exploration by telling students what they "should have found"; instruction may not adequately address the needs of a number of students; or the classroom culture may limit the accessibility or effectiveness of the instruction. Overall, the instruction is *somewhat limited* in its likelihood to enhance students' understanding of the discipline or to develop their capacity to successfully do science.

Level 4: Accomplished, Effective Instruction

Instruction is purposeful and engaging for most students. Students actively participate in meaningful work (e.g., investigations, teacher presentations, discussions with each other or the teacher, reading).The instruction is well-designed and the teacher implements it well, but adaptation of content or pedagogy in response to student needs and interests is limited. Instruction is *quite likely* to enhance most students' understanding of the discipline and to develop their capacity to successfully do science.

Level 5: Exemplary Instruction

Instruction is purposeful, and all students are highly engaged most or all of the time in meaningful work (e.g., investigations, teacher presentations, discussions with each other or the teacher, reading). The lesson is well designed and artfully implemented, with flexibility and responsiveness to students' needs and interests. Instruction is *highly likely* to enhance most students' understanding of the discipline and to develop their capacity to successfully do science.

Appendix B
Sample DIT and OGT Assessment Items

DIT Item #4
Which of the following objects has the most inertia?
a. a 50-kilogram rock
b. a 100-kilogram football player
c. an automobile
d. an oil tanker

DIT Item #5
You want to find out which would empty from a can the fastest—water, alcohol, cooking oil, syrup, or soda pop. To answer this question you will need equal amounts of the liquids as well as which of the following?
a. a can with a hole in the bottom and a stopwatch
b. a stopwatch only
c. cans with different sized holes
d. cans of different sizes, one for each liquid

OGT – Science, 2007 Items #8 & 9
On a humid summer day, Franklin put six ice cubes into each of two cups and then poured the same amount of lemonade into each cup. Cup A was made of plastic, and Cup B was made of glass. He left the cups for about 20 minutes and then came back. He found a small puddle of water around Cup A and a larger puddle around Cup B. Franklin determined that the cups were not leaking.

OGT Item #8
Which is the best explanation for the small puddle around Cup A and the larger puddle around Cup B?
a. Cup A contained more ice than Cup B.
b. Cup A was a better insulator than Cup B.
c. Cup A had a greater volume of lemonade than Cup B.
d. The contents of Cup A were initially colder than the contents of Cup B.

OGT Item #9
Suppose Franklin had a third cup, made of Styrofoam®, to which he added the same number of ice cubes and the same amount of lemonade. What would Franklin expect to observe after 20 minutes?
a. The Styrofoam cup would have a smaller puddle than either the glass or plastic cup.
b. The Styrofoam cup would have a larger puddle than either the glass or plastic cup.
c. The Styrofoam cup would have a puddle exactly the same size as the plastic cup.
d. The Styrofoam cup would have a larger puddle than the plastic cup but a smaller puddle than the glass cup.