International Journal of Science Education
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tsed20

The Impact of High School Science Teachers' Beliefs, Curricular Enactments and Experience on Student Learning During an Inquiry-based Urban Ecology Curriculum

Katherine L. McNeill [a] , Diane Silva Pimentel [a] & Eric G. Strauss [b]
[a] Boston College , Lynch School of Education , MA , USA
[b] Biology Department , Loyola Marymount University , Los Angeles , CA , USA
Published online: 19 Sep 2011.

International Journal of Science Education, 2013
Vol. 35, No. 15, 2608–2644, http://dx.doi.org/10.1080/09500693.2011.618193

The Impact of High School Science Teachers' Beliefs, Curricular Enactments and Experience on Student Learning During an Inquiry-based Urban Ecology Curriculum

Katherine L. McNeill[a*], Diane Silva Pimentel[a] and Eric G. Strauss[b]
[a]*Boston College, Lynch School of Education, MA, USA;* [b]*Biology Department, Loyola Marymount University, Los Angeles, CA, USA*

Inquiry-based curricula are an essential tool for reforming science education yet the role of the teacher is often overlooked in terms of the impact of the curriculum on student achievement. Our research focuses on 22 teachers' use of a year-long high school urban ecology curriculum and how teachers' self-efficacy, instructional practices, curricular enactments and previous experience impacted student learning. Data sources included teacher belief surveys, teacher enactment surveys, a student multiple-choice assessment focused on defining and identifying science concepts and a student open-ended assessment focused on scientific inquiry. Results from the two hierarchical linear models indicate that there was significant variation between teachers in terms of student achievement. For the multiple-choice assessment, teachers who spent a larger percentage of time on group work and a smaller percentage of time lecturing had greater student learning. For the open-ended assessment, teachers who reported a higher frequency of students engaging in argument and sharing ideas had greater student learning while teachers who adapted the curriculum more had lower student learning. These results suggest the importance of supporting the active role of students in instruction, emphasising argumentation, and considering the types of adaptations teachers make to curriculum.

Keywords: *Teacher beliefs; Curriculum enactment; Scientific inquiry*

*Corresponding author: Boston College, Lynch School of Education, 140 Commonwealth Avenue, Chestnut Hill, MA 02467, USA. Email: kmcneill@bc.edu

Historically, the role of the teacher has been overlooked when examining the impact of curriculum on student learning (Ball & Cohen, 1996). Yet the teacher is essential in enacting curriculum materials in terms of the culture of inquiry that is fostered in the classroom, which ultimately impacts students' development of rich conceptual understandings (Puntambekar, Stylianou, & Goldstein, 2007). Although teachers largely rely on curriculum materials for science instruction, their enactments often include modifications of those materials (Forbes & Davis, 2010a). Underlying teacher characteristics, such as teachers' views of science, may shape the ways teachers adapt curriculum for their classrooms (Forbes & Davis, 2010b). For example, teachers' understanding of curriculum as well as their beliefs about teaching and learning influence their enactments (Ball & Cohen, 1996). Furthermore, research suggests that teachers' background, such as degrees and years experience, may also impact student learning (Rice, 2003). Consequently, in this study we examined the role of the teacher on student learning during a high school urban ecology curriculum. Specifically, we examined how teachers' self-efficacy, instructional practices, curricular enactments and backgrounds impacted both student learning on a multiple-choice assessment focused on defining and identifying science concepts and on an open-ended assessment in which students applied science concepts in problem-solving or scientific inquiry practices.

Theoretical Framework

Teacher Use and Enactment of Curriculum

The design and dissemination of curriculum materials has historically been one strategy to inform and reform classroom instruction (Ball & Cohen, 1996). Specifically in science education, inquiry-based science curricula have been found to support greater student learning (Geier et al., 2008). Yet studies in science education suggest that teachers' enactments of inquiry-based curriculum vary and can potentially impact student achievement (Fogleman, McNeill, & Krajcik, 2011; Kanter & Konstantopoulos, 2010; Lawrenz, Wood, Kirchhoff, Kim, & Eisenkraft, 2009; Puntambekar et al., 2007; Schneider, Krajcik, & Blumenfeld, 2005). For example, Kanter and Konstantopoulos (2010) found that students' science learning was impacted by the role of the teacher during a 10- to 12-week, project-based science curriculum that included a focus on project-based and inquiry-based elements. Consequently, it is important to investigate not only if reform-based curriculum impact student achievement, but how teachers' use of the curriculum impact student learning.

One perspective on curriculum use focuses on fidelity of implementation. Fidelity of implementation can be defined as, 'the determination of how well an intervention is implemented in comparison with the original program design during an efficacy and/or effectiveness study' (O'Donnell, 2008, p. 33). Fidelity of implementation studies examine how closely teachers' practices align with the original intention of the curriculum. Often these studies are interested in examining the effectiveness of the curriculum. Fidelity can include both the *fidelity of structure* in terms of the amount of the curriculum and the time spent on the curriculum and the *fidelity to process* in terms of the quality of delivery and programme differentiation (O'Donnell, 2008). Specifically, we were interested in fidelity of structure and as such investigated the impact of how much of the curriculum and the time spent on the curriculum impacted student learning.

In terms of process, we were less interested in whether the teachers' practices aligned with those intended in the curriculum; rather, we were interested in what practices the teachers chose to use (regardless of alignment) and how those practices impacted student learning. Consequently, our perspective aligns more closely with the view of teacher curriculum use as a design activity (Brown & Edelson, 2001) in that the teacher is an active designer of classroom instruction rather than just a transmitter of the written curriculum (Remillard, 2005). Learning environments consist of a variety of different social and material supports that can interact synergistically to support students in developing rich disciplinary knowledge (Tabak, 2004). Curriculum use is a design activity in that teachers actively use all available resources in the classroom environment and their interactions with students in order to achieve their goals (Brown & Edelson, 2001). As such, a teacher's pedagogical design capacity is influenced by the teacher's personal characteristics, the science curriculum and the features of the instructional context (Forbes & Davis, 2010a). Curriculum is not used in a vacuum, but rather in a complex learning environment. Teachers need to design, fit and adapt curriculum in order to meet the needs of their local circumstances (Barab & Luehmann, 2003). Consequently, our perspective on designing the urban ecology curriculum was that it was a tool, which teachers would use in a myriad of ways. Yet, we also assumed that some of those adaptations would be more successful than others. Therefore, we were interested in how teachers' curricular enactments impacted student learning.

Teacher Instructional Practices and Scientific Inquiry

Recent research emphasises the importance of moving away from a traditional model of instruction in which the teacher's main role is that of transmission of information to a new model of a community of learners in which students actively construct their own conceptual understandings (Sawyer, 2006). Specifically, science education reform documents and national standards from a variety of countries, including Australia (Goodrum, Rennie, & Commonwealth of Australia, 2007), Taiwan (Ministry of Education, 1999) and the USA (National Research Council [NRC], 1996, 2000), emphasise the importance of scientific inquiry. Although there are a variety of definitions of inquiry instruction, common components include engaging students in asking questions, designing investigations, analysing data, drawing conclusions and communicating findings (Minner, Levy, & Century, 2010). In the USA, the *National Science Education Standards* (NRC, 1996) describe inquiry as when, '…students describe objects and events, ask questions, construct explanations, test those explanations against current scientific knowledge, and communicate their ideas to others' (p. 2). Inquiry-based science curriculum incorporates a range of scientific phenomena and scaffold student experiences to support them in successfully engaging in investigations by providing guidance and structure during the planning, conducting and analysing phases of inquiry as well as provide support to encourage scientific discourse within the learning community (Singer, Marx, Krajcik, & Chambers, 2000). The curricula include experiences to explicitly facilitate students' successful engagement in scientific inquiry practices such as analysing data and justifying claims with evidence. However, many teachers have difficulty integrating scientific inquiry into their instruction such as supporting students in designing research questions, conducting investigations and communicating evidence-based explanations (Marx, Blumenfeld, Krajcik, & Soloway, 1997). These challenges can impact the instructional practices teachers use in their classrooms when enacting reform-oriented curriculum materials.

Teachers' enactment of inquiry-based curriculum can vary in terms of a variety of instructional practices such as the cognitive demand of the activities, their use of different activity structures and the types of classroom discussions. Teachers are more likely to adapt curriculum materials when the challenges are the greatest such as when supporting students in scientific inquiry investigations and guiding small group discussions (Schneider et al., 2005). Stein, Grover, and Henningsen (1996) found that when teachers implemented reform-oriented mathematics instruction with students that the teachers decreased the cognitive demands of high-level tasks such as problem solving often reducing task ambiguity or complexity by providing specific steps to follow or by telling students how to complete the problem. In terms of science curriculum, teachers can change the activity structure of the curriculum such as adapting student-led science investigations to instead be teacher-led demonstrations. In a study examining a middle school science inquiry-oriented curriculum, researchers found that students who completed the investigations themselves rather than observing their teachers conduct investigations had greater student learning gains (Fogleman et al., 2011). Teacher-led discussions during inquiry curriculum can also vary between teachers. Puntambekar and her colleagues (2007) found that teacher-led discussions that helped connect activities and enabled deeper conceptual understandings were associated with greater student learning. This work emphasises the important role of the teacher in the enactment of inquiry-oriented curriculum.

In addition to studies examining different teacher instructional practices in the context of the same curriculum enactment, there have been a variety of studies that more generally examine inquiry-oriented teacher instructional practices and their impact on student learning. For example, Minner et al. (2010) conducted a review of 138 inquiry studies and found a positive trend indicating that inquiry-based instructional strategies support greater student learning. In particular, they found that inquiry-based instruction that emphasised students engaging in active thinking about drawing conclusions from data resulted in greater student achievement. Wilson, Taylor, Kowalski, and Carlson (2010) compared the effect of students receiving two weeks of instruction using inquiry-based teaching strategies compared with more commonplace science teaching strategies. The two instructional contexts included a variety of differences such as the amount of time spent on: lecture, writing work, small group discussion, receipt of knowledge and knowledge construction. They found that students receiving the inquiry-based instruction had significantly higher levels of achievement in terms of factual knowledge, model-based reasoning and argumentation.

In terms of large-scale studies using hierarchical linear modelling (HLM), previous research has found mixed results for the effect of inquiry-based instructional practices on student achievement. The studies that have been conducted about inquiry teaching practices have found a positive effect (Blanchard et al., 2010; Kahle, Meece, & Scantlebury, 2000), both positive and negative effects (Von Secker & Lissitz, 1999), and no significant effects (Lee, Penfield, & Maerten-Rivera, 2009) on student learning. Specifically, Kahle et al. (2000) found that middle school science teachers who frequently used standard-based teaching practices, such as having students solve problems, asking students questions with more than one answer and requiring students to support their answers with information and reasons, positively impacted students' science achievement and attitudes. Blanchard et al. (2010) also found that students who received guided inquiry-based instruction had higher learning gains compared with students who received more traditional, verification laboratory instruction. Von Secker and Lissitz (1999) found that high schools that placed a greater emphasis on laboratory inquiry (i.e. more experiments and more time on lab periods) and less emphasis on teacher-directed instruction (i.e. more small group and individual work) resulted in greater science learning for all students. Yet schools in this study that placed a greater emphasis on critical thinking (i.e. emphasis on scientific methods, problem-solving skills and scientific writing) resulted in an increased achievement gap between minority and majority students with White and Asian students outperforming Blacks, Hispanics and Native Americans. This suggests that inquiry-based teacher instructional practices had both positive and negative effects on student learning. These mixed results could in part be influenced by the teacher survey items used in the study and the fact that because of data limitations they had to run their analysis at the school level instead of the teacher level. Lee et al. (2009) found that neither teacher self-report on a questionnaire nor classroom observations focused on teaching practices for scientific inquiry significantly impacted third graders' learning of science concepts.

Because of the limited number of studies and mixed results, more large-scale quantitative studies still need to be conducted to link specific teaching practices to student outcomes in inquiry-based learning environments (Thadani, Stevens, & Tao, 2009). The majority of studies conducted thus far describe how a suite of inquiry-based practices impact student learning and do not provide information about individual practices. Identifying specific practices will enable curriculum developers to identify essential components of curriculum that teachers should include in their enactments to support greater student learning (Remillard, 2005). Consequently, this study addresses this critical gap in the research by investigating how teachers' different enactments and instructional practices during an inquiry-based curriculum impacted student learning in a large-scale quantitative study.

Teacher Belief

Teachers' perceptions and beliefs impact what the implemented curriculum looks like in their classrooms (Barab & Luehman, 2003; Remillard, 2005). This includes beliefs about teaching and learning and views about curriculum use. For example, Johnson

(2009) found that teacher beliefs about instructional strategies and their effectiveness impact teachers' willingness to use more inquiry-based instruction in their classroom. Teachers with more traditional views such as teacher-directed instruction, teachers as the giver of knowledge, and the need for drill and practice for state assessments may limit student opportunities for scientific inquiry. Lotter, Harwood, and Bonnor (2007) investigated high school teachers' conceptions and use of inquiry-based instruction. They found that teachers' beliefs about science, their students, effective teaching practices and the purpose of education can all influence the amount and type of inquiry instruction in teachers' classrooms. For example, one teacher in their study limited the inquiry instruction in his classroom, because he believed that students need to receive content knowledge from the teacher before engaging in inquiry. The teacher's beliefs influenced his more teacher-directed instruction. The instructional approaches that teachers use in the classroom are influenced by their beliefs about themselves, their students and other factors in their environment (Jones & Carter, 2007).

Teachers' beliefs include a large system of intertwined ideas such as beliefs about science, science teaching, motivation, curriculum, schools and students (Jones & Carter, 2007). Specifically, in this study we focused on two types of teacher beliefs: self-efficacy and beliefs about curriculum use. We define teacher self-efficacy as a teacher's belief in his or her own capacities as a teacher to successfully implement an instructional strategy (Jones & Carter, 2007) or influence student learning (Tschannen-Moran, Hoy, & Hoy, 1998). This definition builds on the work of Bandura (1986) who discusses self-efficacy in terms of both perceived self-efficacy and outcome expectations in that, '[p]erceived self-efficacy is a judgment of one's capability to accomplish a certain level of performance, whereas an outcome expectation is a judgment of the likely consequence such behaviour will produce' (p. 391). As mentioned previously, teachers' beliefs can impact their use of inquiry-based practices in their science classrooms (Johnson, 2009; Lotter et al., 2007). Consequently, we were interested in teachers' self-efficacy for science inquiry. Specifically, we were interested in whether teachers' comfort with supporting students in scientific inquiry, such as designing and conducting investigations or constructing scientific arguments, impacted student learning. Furthermore, teachers who lack confidence in teaching science content are less likely to teach that content (Jones & Carter, 2007). Because of the innovative and interdisciplinary nature of the content in the urban ecology curriculum, we were also interested in teachers' self-efficacy with the science content. Previous research suggests that teachers' sense of efficacy can have significant impacts on student outcomes such as achievement (Tschannen-Moran et al., 1998).

Teachers' beliefs about curriculum influence how they interpret the material and how they choose to frame the material for their students (Ball & Cohen, 1996). For example, Remillard and Bryans (2004) found that teachers' orientations towards curriculum impacted how they used the curriculum in their classroom. Teachers studied and read the curriculum materials to varying degrees with some teachers closely reading the materials and other teachers primarily using the curriculum to look for activities for students. Teachers' self-efficacy for science inquiry, self-efficacy for science content, and beliefs about curriculum use can impact what their enactment looks like in their classroom. Consequently, we were interested in investigating whether or not these beliefs influenced student learning.

Teacher Background and Experience

Recent federal and state policies in the USA that focus on highly qualified teachers include a consideration of qualifications such as years of teaching experience, degrees in science and degrees in education. Yet, over the years researchers examining links between teacher qualifications and student learning have found mixed results (Marx & Harris, 2006). For example, Rice (2003) completed a literature review that examined the impact of different teacher attributes on teacher quality. Several studies in the review found a positive effect between teacher experience and effectiveness. Furthermore, having an advanced degree in science had a positive effect on student achievement. On the other hand, Monk (1994) found that high school biology teacher academic degree and experience were unrelated to student achievement. Rather, other measures such as number of content-specific courses had a positive effect on student learning with coursework playing a more positive role in less experienced teachers. Monk suggests that preparation loses its impact with time and what is most important is how much the teacher knows about what is being taught. Although Kanter and Konstantopoulos (2010) did not specifically examine teacher experience, they did investigate the effects of teacher disciplinary and pedagogical content knowledge on student learning during a project-based science curriculum. They also found a correlation between increases in teachers' content knowledge and pedagogical content knowledge and student achievement. In Wilson, Floden, and Ferrini-Mundy's (2002) literature review on teacher preparation, they found that research on the impact of majors and coursework on teacher effectiveness was limited. Both subject matter coursework and education coursework were at times associated with higher student achievement, yet the results were not consistent. Consequently, we were interested in investigating whether teachers' qualifications (e.g. years of experience, science degrees and education degrees) had a significant impact on student learning in the context of the inquiry-based urban ecology curriculum.

In the context of enacting an inquiry-based science curriculum, the role of the teacher is essential. Previous research suggests that a variety of different factors could potentially impact student learning. However, there have been relatively few large-scale studies about inquiry-based science curriculum examining the relationship between the teacher and student outcomes and the studies that have occurred offer mixed results (Thadani et al., 2009). Findings about the impact of teacher characteristics and curricular adaptations on student learning would provide important insights for the design of teacher education, professional development and inquiry-based curriculum. In order to address this gap in the existing literature, we investigated the following research questions: (1) What variation in student achievement exists between teachers during the enactment of the *Urban EcoLab* curriculum?; (2) What relationships exist between student demographics and pre-test scores with post-test achievement?; and (3) How do teachers' curricular enactments, instructional practices, self-efficacy for science inquiry, self-efficacy for science content, beliefs about curriculum use and backgrounds impact high school students' science learning?

Methods

Curricular Context

This study took place in the context of a year-long high school urban ecology curriculum entitled *Urban EcoLab: How can we develop healthy cities?* (Strauss, McNeill, Barnett, & Reece, 2008). The curriculum was funded in part by a grant from the United States National Science Foundation and is freely available to educators online (http://urbanecolabcurriculum.com). This curriculum is a collaborative response to the need for an environmental course that is urban focused, inquiry-based and trans-disciplinary. Traditional environmental science books have often lacked a critical central theme. This perspective is bolstered by findings that suggest the current suite of high school environmental textbooks reviewed by the Environmental Literacy Council (2004) suffered from several central weaknesses. The Council found that each book varied greatly in quality from chapter to chapter because there was not a coherent theme that unified the materials. Using urban ecology in a social justice context and an inquiry learning environment provides the connecting thread that weaves the narrative and lessons together.

Urbanisation, as a process, represents a tremendous challenge to global sustainability with predictions that include a doubling of the size of US cities before the end of the 21st century (UNFPA, 2007). Urban ecology as a science captures the multi-disciplinary scholarship required to fully understand the dynamic processes that drive the changes observed in urbanised landscapes. These drivers may be physical, such as climate, topography and biodiversity. Equally important are the human social factors, such as governance, information flow and ethical values (Hollweg, Pea, & Berkowitz, 2003). In addition, the historical narrative of the land use and land cover change is central to developing sustainable urban management practices. These interactive forces are at the core of the discipline and are the focus of our curriculum.

The curriculum consists of eight modules covering different topics including patterns of land use, climate change, hazardous waste, public health and biodiversity. Each module contains approximately 10 lessons that include different activity structures such as inquiry investigations, development of models, role-play, computer simulations, field investigations of their city and environmental action plans. Table 1 provides an overview of the eight modules in terms of the topic, time, driving question and big idea.

The organisation of the curriculum reflects the trans-disciplinary nature of urban ecology. While the core ideas reflect the critical tenets of the ecological sciences, the curriculum engages public health, engineering, climate science and environmental justice. This approach is unorthodox, but from our perspective, a necessary structure for engaging urban youth in the environmental sciences.

Table 1. Urban ecology curriculum—How do we develop healthy and sustainable cities?

Module	Time	Module driving questions	*The Big Idea* for the module
Module 1: Introduction to urban ecology	3 weeks	What do we need to know to develop a healthy and sustainable city?	Urban ecology is the study of cities as the interactions among biological, chemical, physical and social forces Large human populations, living in technologically advanced society, have the least impact on the global ecosystem if they live in healthy and sustainable cities
Module 2: Patterns of urban land use	3 weeks	How and why has my neighbourhood changed over the years?	Humans have transformed urban landscapes over time and have left a legacy for us to discover and understand
Module 3: Energy and climate change	3 weeks	How do we develop cities that minimise their impact on the climate?	Climate change is a systematic change in the long-term characteristics of weather patterns (i.e. temperature, precipitation or winds) sustained over several decades or longer Humans can modify their behaviour (e.g. transportation, food choices, reproductive decisions, etc.) to reduce their consumption of energy
Module 4: Garbage and hazardous waste	3 weeks	How can we minimise the environmental impact of our city's garbage and hazardous waste?	Solid waste and chemicals (including e-waste) enter the ecosystem through various transport mechanisms but a large percentage of this pollution, particularly heavy metals, is carried by storm water into waterways and can become trapped in the soil
Module 5: Public health	3 weeks	How can I make my community a healthier place to live?	Healthy urban neighbourhoods have green space, robust social networks, effective water delivery and waste removal, clean air and healthy foods Cities across the world have different infrastructure and systems to support healthy urban neighbourhoods
Module 6: Biodiversity	4 weeks	How do we develop cities that sustain biodiversity?	Urban systems and natural systems transform each other in complex ways that help us understand both in a new way. A

(*Continued*)

Table 1. (*Continued*)

Module	Time	Module driving questions	*The Big Idea* for the module
			city's wildlife and natural areas are integral parts of a healthy ecosystem Emerging urban ecosystems can alter biodiversity through a variety of ways
Module 7: Animal behaviour	4 weeks	How can we use our understanding of animal behaviour to develop healthy cities?	Pace of change in the urban landscape exceeds the rate of evolution Organisms living in urban ecosystems have modified their behaviour in response to the impacts of the humans that live there
Module 8: Taking action	3 weeks	How can I improve my neighbourhood's environment?	You can have a positive impact on your city and local neighborhood

Participants

The participants in this study included 22 teachers from 21 different schools piloting the urban ecology curriculum and their 935 students. Their schools were located in three regions of the USA: Northeast, Midwest and Southwest. The teachers were recruited by a university partner in each of the three locations who sent out a solicitation about piloting the curriculum to state and school district listservs. Teachers who participated in the pilot received the curriculum materials as well as a stipend upon completion of the data collection requirements. All teachers participating in the pilot also attended professional development focused on the content and activities (e.g. inquiry investigations, field studies, etc.) in the urban ecology curriculum. The professional development varied in the three locations, but included approximately three days of professional development during the summer and four professional development days that occurred bimonthly during the school year.

The schools and participants in the study included a range of backgrounds. School size varied such that 8 could be classified as small (<400 students), while the other 13 were large (>1,000 students). Student demographics among schools varied on a spectrum with some schools serving predominantly one race (African-American, Hispanic or White) of students and others serving a more diverse group of learners with similar representations of two or more races. Over 50% of the students in most schools participated in a free or reduced lunch plan with only five schools falling below 20% participation. The student mobility also varied greatly among schools. Mobility rates could only be acquired for 14 schools. Although the reported mobility rate for 1 school of the 14 was 5%, the majority of schools had a rate greater than twice this amount, with two schools reporting a mobility rate of 42%. In most cases, there was a relationship between mobility rate and dropout rate such that those schools with a higher mobility rate tended to also have a relatively higher dropout rate. Teacher experience ranged from 1 to 29 years with an average of 10 years for all teachers. Although there was some variation in the degrees held by teachers, most teachers had a Bachelors degree in science and a Masters degree in education (Table 2).

Study Design

In order to investigate the relationship between the role of the teacher and student outcomes during an inquiry-oriented science curriculum, we conducted a large-scale quantitative study utilising a quasi-experimental design with student pre- and post-tests, but without a control group that capitalised on the natural variation in classroom instruction (Shadish, Cook, & Campbell, 2002). We did not include a control group, because urban ecology is a relatively new domain (Hollweg et al., 2003) as such there was not a suitable curricular comparison. Furthermore, we were specifically interested in the teacher variation within the use of the curriculum in order to identify key characteristics of curricular use and the teachers' backgrounds that impacted student learning. We purposively included teachers with a range of backgrounds and previous experiences, because we were interested in the naturally occurring contrasts (Shadish et al., 2002) in the teachers' curricular enactments, instructional practices, beliefs and backgrounds. One limitation of the study is that the teachers volunteered to participate in the research and cannot be considered a random sample of teachers. Consequently, the teachers' beliefs, instructional practices and backgrounds may not be typical of other high school science teachers in the USA. We utilised HLM, because the students could not be randomly assigned to teachers and students are nested within teachers' classrooms as such multi-level modelling provides a more accurate estimation of effects and variance (Raudenbush & Bryk, 2002).

We collected and analysed a variety of data sources to address our research question including: teacher pre-survey, teacher enactment surveys and student pre- and post-assessments. Another limitation of this study is the reliance on teacher surveys instead of observational data to measure the teachers' enactment of the urban ecology curriculum. Previous research suggests that teachers' beliefs about instruction can be inconsistent with their actual instruction (Jones & Carter, 2007). Unfortunately, because of limitations in the project budget we were unable to observe all of teachers'

Table 2. Teacher descriptive statistics

Years of experience	1–5	6–10	11–15	16–20	21–25	26–30
# of teachers	8	6	3	3	1	1
Highest degree obtained in science			NA	Bachelors	Masters	Doctorate
# of teachers			1	14	6	1
Highest degree obtained in education			NA/Associate	Bachelors	Masters	Doctorate
# of teachers			3	5	14	0

enactment of the curriculum. As Lee et al. (2009) discuss, large-scale programmes require extensive resources and personnel in order to obtain reliable observation measures of classroom practice. Consequently, the results from this analysis can only provide insights into the relationship between teachers' self-report of their classroom instruction and student outcomes, not the actual classroom instruction. In the following sections, we describe each of the three data sources in more detail.

Teacher pre-survey. Teachers completed a pre-survey, which asked questions about their instructional practices, beliefs about curriculum use, self-efficacy for science inquiry and self-efficacy for science content. We designed the survey to include at least nine items each that aligned with the following constructs of interest: (1) Frequency of instructional practices; (2) Frequency of curriculum use; (3) Comfort designing and conducting investigations with students; (4) Preparedness to teach students about scientific explanations and arguments; and (5) Preparedness to teach urban ecology content. We developed the teacher survey by adapting items from the existing measures, creating items that aligned with the *National Science Education Standards* (NRC, 1996), and constructing new items that aligned with the content goals in the *Urban EcoLab* curriculum. We looked at two existing measures. The first measure was the Horizon teacher questionnaire for Local Systemic Change through Teacher enactment (Banilower, Heck, & Weiss, 2007). The majority of the items in the frequency of instructional practices construct stemmed from this measure with the addition of a few items that aligned with the instructional practices advocated for in the *National Science Education Standards* (NRC, 1996). The second measure we adapted was an interview protocol addressing math teachers' use of curriculum, which we used to develop the items on the frequency of different types of curriculum use (Kennedy, Ball, & McDiarmid, 1993). The items asking about comfort with designing and conducting investigations and preparedness to teach students' about scientific explanation and argument were designed to align with the science as inquiry content standards in the *National Science Education Standards* (NRC, 1996). Finally, the last set of items which focused on preparedness to teach urban ecology content were designed specifically for this curriculum to align with the content learning goals in each module. In the actual teacher survey, the items were randomised and not organised by the five constructs. The final items used in this analysis are included in the data analysis section in which we describe the creation of the final teacher belief factors (see Table 5).

In addition, the teacher pre-survey included questions about teachers' backgrounds. Specifically, teachers were asked how many years they had been teaching, their highest degree in science and their highest degree in education. These items were included in the analysis to examine the impact of teachers' background on student achievement in science.

Teacher enactment surveys. During the enactment, teachers completed module surveys that asked about the level of completion, the level of adaptation and the amount of time using different activity structures (e.g. individual work, group work, full class discussion and lecture) for each activity within a lesson. Table 3 provides a sample of the enactment survey from Lesson 1 in Module 1: Introduction to Urban Ecology.

In addition, teachers were asked how long their class periods were since some teachers saw their students every day for shorter periods (e.g. 45 min) and other teachers were on a block schedule and saw their students every other day for longer periods (e.g. 90 min). The information from the enactment survey was used to create variables that provided a synthesis of the teachers' enactment of the curriculum including number of modules used, amount of time spent on the curriculum, level of curriculum completion, level of adaptation and the frequency of different activity structures (e.g. group work or lecture). Table 4 provides a summary of the different variables investigated in the analysis and how they were calculated. The teacher enactment variables allowed us to investigate if level of curriculum completion and the teacher enactment choices impacted student learning.

Student assessment. Similar to other research (Blanchard et al., 2010; Wilson et al., 2010), we considered what type of student assessment to use to evaluate the impact of teachers' enactments, instructional practices, beliefs and backgrounds on student learning. Previous research suggests that teacher instructional practices in inquiry science instruction can have different impacts on students' conceptual understanding when measured by multiple-choice items compared with open-ended questions (Puntambekar et al., 2007). Consequently, we designed an identical two-day pre- and post-test that consisted of 21 multiple-choice items and 4 open-ended questions. The more traditional multiple-choice items focused on knowledge of facts and the open-ended items focused on the application of science concepts in problem solving and scientific inquiry. Each day of the assessment consisted of 10 or 11 multiple-choice items and two open-ended items. The multiple-choice items were

Table 3. Teacher enactment survey

	# Class periods (e.g. How many days?)	Level of completion 2 = completed; 1 = partially;0 = not used	Level of adaptation 3 = major changes; 2 = minor changes; 1 = no changes; 0 = not used	Main activity structure 4 = student individual work; 3 = student group work; 2 = full class discussion; 1 = lecture; 0 = not used
Lesson 1: Cities as systems				
1.1 Brainstorm about cities	0 $\frac{1}{2}$ 1 2 3 4	0 1 2	0 1 2 3	0 1 2 3 4
1.2 Narrative-opening story	0 $\frac{1}{2}$ 1 2 3 4	0 1 2	0 1 2 3	0 1 2 3 4
1.3 Cities as systems	0 $\frac{1}{2}$ 1 2 3 4	0 1 2	0 1 2 3	0 1 2 3 4

Table 4. Teacher enactment variables

Variable	Description
Number of modules completed	Number of modules completed ranged from 4 to 8
Total number of periods spent on the curriculum	Added together the number of periods for each activity within each module and then added across all modules to calculate a total
Total number of minutes spent on the curriculum	Multiplied the total number of periods times the length of the period
Average level of completion of curriculum	Calculated the average level of completion for each module by adding the level of completion (2, 1 or 0) and then dividing by the number of activities in the module. Then added together the averages for each module and divided by the number of modules
Average level of adaptation	Calculated the average level of adaptation for each module by adding the level of adaptation for any used activity (3, 2, or 1) and then dividing by the number of used activities in the module. If a teacher selected 0, that means they did not use the activity and it was not included in calculating the average. Then added together the averages for each completed module and divided by the number of completed modules
Average percentage of time students spent on individual work	Counted the number of times a teacher recorded 4 for activity structure within a module. Divided this by the total number of activities they completed within a module to create a percentage. Then added together the percentages for each completed module and divided by the total number of completed modules
Average percentage of time students spent on group work	Counted the number of times a teacher recorded 3 for activity structure within a module. Divided this by the total number of activities they completed within a module to create a percentage. Then added together the percentages for each completed module and divided by the total number of completed modules
Average percentage of time class engaged in full class discussion	Counted the number of times a teacher recorded 2 for activity structure within a module. Divided this by the total number of activities they completed within a module to create a percentage. Then added together the percentages for each completed module and divided by the total number of completed modules
Average percentage of time teacher lectured	Counted the number of times a teacher recorded 1 for activity structure within a module. Divided this by the total number of activities they completed within a module to create a percentage. Then added together the percentages for each completed module and divided by the total number of completed modules

combined to create a measure of science facts and vocabulary. The open-ended items were combined to create a measure of scientific inquiry and problem solving. The student assessment also included three questions about students' backgrounds (i.e. grade level in school, gender and race) to determine whether these variables were associated with student achievement. Unfortunately, we were unable to include background questions about the students' socioeconomic status or family income because of our agreement with the schools.

The twenty-one, multiple-choice items were designed to align with the key science content learning goals in each of the first seven modules in the curriculum. Specifically, there were three items for each module. The last module in the curriculum, Module 8, focused on having students design and enact an action plan to have a positive environmental impact on the students' neighbourhood or city. This module required that students apply the science concepts learned in the previous seven modules, but did not introduce any new science content. Consequently, separate content items were not designed to align with this module. The multiple-choice items focused predominantly on knowledge of science facts and vocabulary in that students were asked to define or identify concepts. Appendix 1 provides one sample multiple-choice item that aligns with each module. For each multiple-choice item, students' responses were scored as either correct (1) or incorrect (0) and then summed for a total possible score of 21 points. In order to check the reliability, we calculated Cronbach's alpha for the post-test multiple-choice items which was 0.904 suggesting that they reliably measured the same phenomenon.

The four open-ended items focused on students' ability to apply the science concepts in the unit to engage in problem solving. Specifically, the items focused on two aspects of problem solving or scientific inquiry that were highlighted in the curriculum: the construction of scientific arguments and environmental action planning. Two items were designed to focus on argumentation and two items were designed to focus on environmental action planning. Appendix 2 includes one sample item for environmental action planning and one sample item for argumentation. We developed four specific rubrics, one for each item, that combined both the accuracy of the science content with the appropriate structure for the problem-solving task. For example, the rubric for the sample environmental action planning item evaluated students' abilities to provide scientifically appropriate and accurate goals, stakeholders, resources and actions for their action plan to support public health. The rubric for the sample argument assessment item evaluated students' ability to provide scientifically appropriate and accurate evidence and reasoning for their claim (either Location A or B) to support greater biodiversity of rabbits. The number of possible points varied for each of the four items because the rubrics were specific to the content and task. For the four questions, the possible scores were 8, 14, 16 and 9, respectively, for a total possible score of 47 points on the open-ended assessment.

The open-ended responses were scored by one rater using the rubric. Twenty percent of these open-ended test items were randomly sampled and scored by a second independent rater. Estimates of inter-rater reliability were calculated by percent agreement. Interrater reliabilities for the four questions were 94%, 88%, 96%, and 95%, respectively. As a second reliability check, we calculated Cronbach's alpha for students' scores on the open-ended items which was 0.856 suggesting that the items reliably measured the same phenomenon.

As mentioned previously, the test required two days at the beginning of the school year and two days at the end of the school year. Only students who completed all parts of the test were included in the analysis. Due to high absenteeism, the high mobility rate of students, the high drop out rate in the urban high schools (particularly for 11th and 12th graders who were the target population of the curriculum) and the necessity of students being in class for all four days of testing, only 366 students took all parts of the pre- and posttest assessments. Consequently, the results of this study are limited to those students who remained in the same science class all school year and who attended the science class on the required assessment days (suggesting they may have higher attendance rates than their peers).

Data Analysis

Dummy coding. To prepare the data for analysis, we coded the categorical data as dummy variables for inclusion in the model. For the student data, this included three variables—grade, gender and race. The urban ecology curriculum was specifically designed to be a capstone science course for the 11th and 12th grade students. The majority of the students in the analysis were in either the 11th grade ($n = 119$) or 12th grade ($n = 182$) with a smaller number of students in the 10th grade ($n = 34$) and 9th grade ($n = 31$). Consequently, we selected the 12th grade as the reference group for grade-level dummy codes. We were interested in whether students in lower grade levels had significantly different learning gains over the course of the curriculum. For gender, male was selected as the reference group for the dummy code to compare achievement of females to males. For race, students were provided with the following choices on the student assessment: (1) Black/African American ($n = 69$); (2) Asian American or Pacific Islander ($n = 25$); (3) White ($n = 159$); (4) Latino/Latina ($n = 86$); (5) Native American or American Indian ($n = 3$); and (6) Other ($n = 20$). In addition, four students chose not to respond to this question. In creating the dummy code, white was included as the reference group and all other groups were compared with White. Because only three students selected Native American or American Indian and only four students opted not to respond to this question, these categories were combined with Other to create a new variable that included all of these options. Consequently, the final race codes used in the analysis were: (1) Black/African American ($n = 69$); (2) Asian American or Pacific Islander ($n = 25$); (3) White ($n = 159$); (4) Latino/Latina ($n = 86$); and (5) Other—Native American, Other Race and No Response ($n = 27$).

For the teacher data, dummy codes were developed for two variables—degree in science and degree in education. For degree in science, teachers' degrees originally included the following: (1) No degree ($n = 1$); (2) Bachelors degree ($n = 14$); (3) Masters degree ($n = 6$); and (4) PhD ($n = 1$). For degree in education, teachers' degrees originally included the following: (1) No degree ($n = 1$); (2) Associate degree ($n = 2$); (3) Bachelors degree ($n = 5$); and (4) Masters degree ($n = 14$). Since the majority of teachers in the study either had a Bachelors degree or Masters degree in both areas, a dummy code was created to split the teachers into two groups for each degree—graduate degree or no graduate degree. The no graduate degree included the teachers with no degree, an Associate degree or a Bachelors degree. The graduate degree included teachers with a Masters or PhD. Although theoretically we would have preferred to model the teachers with no degrees and associate degrees separately, empirically there were not enough teachers in these two categories for the statistical analysis. For both science and education degrees, we used graduate degree as the reference group.

Factor analysis. We used principle component factor analysis using Varimax rotation to combine multiple items from the teacher pre-survey into constructs to increase the reliability of our measures and to create more manageable constructs for analysis. We conducted five-factor analyses based on the original theoretical design of the items into categories: (1) Frequency of instructional practices; (2) Frequency of curriculum use; (3) Comfort designing and conducting investigations with students; (4) Preparedness to teach students about scientific explanations and arguments; and (5) Preparedness to teach urban ecology content. All factors with an eigenvalue greater than 1 were then checked for reliability using Cronbach's alpha. All factors with a reliability greater than 0.7 were included in this analysis. This resulted in the creation of six teacher instructional practices and belief factors, which included two factors from 'Frequency of instructional practices' and one factor from each of the other four theoretical constructs. Factors were created by summing the individual items and dividing by the total number of items. Table 5 includes the final name, Cronbach's alpha and the items from the survey for each factor.

Hierarchical linear modelling. Determining the impact of the teacher on student achievement is complex because students in the same class are not independent. Multi-level modelling recognises the dependence and grouping of data, which leads to a more correct estimation of effects and variance. We used HLM in a two-level format to investigate the effect of teachers' enactments, instructional practices, beliefs and backgrounds on student learning (Raudenbush & Bryk, 2002). We developed two models. One model used the multiple-choice items as the outcome, which focused on students' ability to define and identify the key science concepts in the curriculum. The second model used the student open-ended items as the outcome, which focused on students' ability to apply the science concepts in the construction of scientific arguments and environmental action planning. The creation of both models consisted of three steps.

In the first step, we created a fully unconditional model (FUM). This model provides an estimate of the mean and confidence interval for the outcome measure (γ_{00}). It also provides the results of partitioning the outcome variance into within-group (σ^2) and between-group (τ_{00}) components, testing whether the between-group component is significantly different from zero. From these measures, we computed the intraclass correlation coefficient (ICC), ρ, which is the proportion of variation in the outcome measure that is due to differences between groups.

Table 5. Teacher instructional practices and belief factors

Factor name	Cronbach's alpha	Survey items in factor
Frequency students engage in argument and share ideas[a]	0.9320	Have students provide their reasoning for why their evidence supports their claim Have students provide evidence to support their claims Have students justify their claims Have students communicate and defend a scientific argument Have students critique alternative arguments using evidence and reasoning Have students investigate science issues in their community Have students share ideas with each other
Frequency students engage in traditional classroom practices[a]	0.7385	Have students take notes from a textbook Introduce content through formal presentations or lecture Have students answering textbook and/or worksheet questions Have students read from a textbook Engage students in open-ended inquiry where students develop their own questions (reversed)
Frequency use curriculum[a]	0.9424	Read the curriculum materials to identify common student misconceptions Read the curriculum materials to identify teaching strategies for helping students engage in science Read the curriculum materials to identify teaching strategies for helping students engage in scientific inquiry Read the curriculum materials to refresh/reinforce my understanding of the science content Use the curriculum materials to find new activities or labs to use with my students Read the curriculum materials to learn new strategies for addressing diverse learners Use the curriculum materials to provide background/supplementary readings for my students Use the curriculum materials to find suggestions of probing questions I could ask my students in discussion Read the curriculum materials to provide the overall sequence of activities/tasks
Comfortable teaching students to	0.9032	I am comfortable teaching students how to

(*Continued*)

Table 5. (*Continued*)

Factor name	Cronbach's alpha	Survey items in factor
design and conduct investigations[b]		accurately observe and measure science phenomena. I am confident in my ability to teach students how to identify questions that can be answered through scientific investigations. I am comfortable teaching students how to design a scientific investigation including appropriate techniques to gather and record data I am comfortable teaching students how to ask a question about objects, organisms and events in the environment I am comfortable teaching students how to design a scientific investigation. I am comfortable teaching students how to conduct a scientific investigation.
Prepared to teach students to construct arguments[b]	0.9054	I am prepared to teach students how to justify their claims I am prepared to teach students how to critique alternative arguments I am prepared to teach students how to provide their reasoning for why their evidence supports their claim I am prepared to teach students how to use scientific principles to determine and support their claims I am prepared to teach students how to communicate and defend a scientific argument
Prepared to teach urban ecology content[b]	0.9439	I feel prepared to teach my students about the impact of technology on waste production I feel prepared to teach my students about the forces that interact to determine land use policies over time I feel prepared to teach my students about the disposal and fate of wastes produced in urban ecosystems I feel prepared to teach my students about the elements of cities that influence human health I feel prepared to teach my students about the function of behavioural plasticity in the survival of animals I feel prepared to teach my students about the effect of green space on human health I feel prepared to teach my students about the

(*Continued*)

Table 5. (*Continued*)

Factor name	Cronbach's alpha	Survey items in factor
		effect of air quality and water quality on human health I feel prepared to teach my students about the unique characteristics of urban ecosystems I feel prepared to teach my students about the biophysical and social drivers in the growth of cities I feel prepared to teach my students about the land use patterns associated with healthy ecosystems

[a]Teachers' choices were: Never, A few times a year, Once a month, Once a week, Two to three times a week or Every science lesson.

[b]Teachers' choices were: Strongly agree, Agree, Neutral, Disagree or Strongly disagree.

Specifically, this told us what proportion of the variation in student achievement on the post-test was between teachers.

Next, we created a level 1 or within-teacher model to examine the effect of student-level variables (e.g. demographics and pre-test) on student achievement. We can include variables in this model as either fixed or random. A fixed effect is one that has a constant effect on the outcome across all groups, whereas a random effect has a random component and is allowed to differ across groups. Because of the relatively small number of groups or teachers in this study ($n = 22$), we included all level 1 variables as fixed effects in the model and centred each variable around the grand mean. After running the within-school model, we determined how much of the total unexplained individual-level variance for our outcome was explained by the addition of our level 1 variables. Specifically, this told us what proportion of the variation in post-test achievement we were able to explain for each teacher by including the students' pre-test scores, race, gender and grade level.

Finally, we developed a level 2 or between-teacher model to examine the effect of teacher-level variables on student achievement. Specifically, we tested 18 variables including the 3 teacher background variables (i.e. years teaching experience, science degree and education degree), the 9 teacher enactment variables (see Table 4), and the 6 teacher instructional practices and belief factors (see Table 5). All continuous variables were added as grand mean centred and all categorical variables were added as uncentred. As a general rule, 10 cases are required at a level (either level 1 or level 2) for each significant variable included in a model (Raudenbush & Bryk, 2002). Consequently, since we only had 22 teachers in our study, our expectation was that we would only have two significant teacher-level variables in each of our final models. Raudenbush and Bryk (2002) recommend adding in variables one at a time, rather than in large groups and then deleting, because of issues of multicollinearity.[1] Consequently, we tested each teacher variable individually to test for significance. We then tested all combinations of two significant variables building up the final model. The final model is the model with the greatest number of significant variables (i.e. one teacher variable for the multiple-choice assessment and two teacher variables for the open-ended assessment) with the lowest significant levels. As with the within-teacher model, we determined how much of the total unexplained teacher-level variance of our outcome was explained by the addition of our level 2 variables. In other words, we determined what proportion of the variation in student achievement on the post-test between teachers was explained by the addition of the teacher variables in the model.

Results

The analyses address the following research questions: (1) What variation in student achievement exists between teachers during the enactment of the *Urban EcoLab* curriculum?; (2) What relationships exist between student demographics and pre-test scores with post-test achievement?; and (3) How do teachers' curricular enactments, instructional practices, self-efficacy for science inquiry, self-efficacy for science content, beliefs about curriculum use and backgrounds impact high school students' science learning? For each research question, we examine separately the impact on student learning in terms of defining and identifying science concepts (as measured by the multiple-choice assessment) and in terms of applying science concepts in argumentation and environmental action planning (as measured by the open-ended items). We describe our results from the three steps of the hierarchical linear model, the FUM, the within-teacher model and the between-teacher model, which address each of the research questions. Then we present some exploratory analyses to further investigate possible reasons for the results in the between-teacher model.

Fully Unconditional Model

The FUM suggests that there was a significant difference in student achievement between teachers for both the multiple-choice assessment, $\chi^2 = 182.48$ (df $= 21$), $p < 0.001$ and the open-ended assessment, $\chi^2 = 256.69$ (df $= 21$), $p < 0.001$. The reliability of both models is above 0.7, which is considered high. Consequently, it is appropriate to use the adjusted ICC to estimate the percentage of variance in student achievement that exists between teachers. Table 6 presents the results from both FUMs.

These results suggest that 34.5% of the variance in student achievement on the multiple-choice assessment existed between teachers and 42.5% of the variance in student achievement on the open-ended assessment existed between teachers. This suggests that even when all of the students are receiving the same curriculum that a large percentage of student achievement is impacted by the role of the teacher.

Within-Teacher HLM Model

The within-teacher model explored which student-level characteristics were associated with science achievement. We include these characteristics in our model in order to

Table 6. Fully unconditional model

	Multiple-choice assessment	Open-ended assessment
Tau (τ_{FUM})	2.575	12.168
Sigma-squared ($\sigma^2{}_{\mathrm{FUM}}$)	5.973	18.959
Lambda-reliability (λ)	0.820	0.868
ICC[a]	0.301	0.391
Adjusted-ICC[b]	0.345	0.425

[a]ICC $= \tau/(\tau + \sigma^2{}_{\mathrm{FUM}})$.

[b]Adjusted ICC $= \tau/(\tau + (\lambda\sigma^2{}_{\mathrm{FUM}})$.

control for them and to see if there were any relationships with science learning. Table 7 presents the results from this analysis. For both the multiple-choice and open-ended assessments, only two of the student-level variables significantly impacted student learning: pre-test and race. Students who performed higher on the pre-test were more likely to perform higher on the post-test for both the multiple-choice and open-ended assessments suggesting that incoming knowledge influences student learning. The results for race differed for the multiple-choice assessment compared with the open-ended assessment. On the multiple-choice assessment, students who identified their race as Black or African American on average scored significantly lower compared with the performance of White students. On the open-ended assessment, students

Table 7. Within-teacher model

	Multiple-choice assessment	Open-ended assessment
Random effects		
Intercept (β_0)	10.403***	23.574***
Fixed effects		
Pre-test	0.395***	0.302***
Black/African American	−0.812*	−0.792
Asian/Pacific Islander	−0.070	−1.674***
Latino/a	−0.375	0.188
Other race	−0.997	−0.833
Female	−0.225	0.530
9th Grade	−0.0407	1.910
10th Grade	0.413	0.288
11th Grade	−0.662	−0.154
Variance components for random effects		
Intercept variance (τ_{within})	2.242***	11.070***
Sigma-squared ($\sigma^2{}_{\mathrm{within}}$)	5.320	17.516
Proportion of variance explained[a]	0.109	0.076

$^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Comparison groups: White, male, 12th grader.

[a]Within proportion of variance explained $= (\sigma^2{}_{\mathrm{FUM}} - \sigma^2{}_{\mathrm{within}})/\sigma^2{}_{\mathrm{FUM}}$.

who identified their race as Asian or Pacific Islander scored on average significantly lower compared with the performance of White students. There was not a significant difference in performance for students who selected other races. Furthermore, gender and grade level did not have a significant impact on student learning as measured by either the multiple-choice or open-ended assessments.

Our within-teacher model explains 10.9% of the individual-level variance in science learning for the multiple-choice assessment and 7.6% of the individual-level variance for the open-ended assessment. Furthermore, the intercept variance remains significant in both models suggesting there is still a significant amount of unexplained variance between students. Since this is a relatively small percentage, it suggests that there are other variables besides pretest, race, gender and grade level that influence the variation in student achievement within a teacher's classroom.

Between-Teacher HLM Model

The between-teacher model investigated the impact of teacher instructional practices, curricular enactments, beliefs and backgrounds on student achievement. The final models for both the multiple-choice assessment and open-ended assessment are given in Table 8. As described in the methods, we initially ran separate models with each of the teacher-level variables independently. Then we tested all combinations of two significant variables to develop the final model.

Table 8. Between-teacher HLM model

	Multiple-choice assessment	Open-ended assessment
Random effects		
Intercept (β_0)		
Percentage of time teacher lectured	−5.146*	–
Level of adaptation	–	−5.994*
Belief—frequency of argument	–	0.947∼
Fixed effects		
Pre-test	0.391***	0.302***
Black/African American	−0.796*	−1.138
Asian/Pacific Islander	−0.033	−1.744*
Latino/a	−0.366	0.059
Other race	−0.984	−1.040
Female	−0.203	0.530
9th Grade	−0.057	3.137
10th Grade	0.413	0.716
11th Grade	−0.654	0.013
Variance components for random effects		
Intercept variance (τ_{between})	1.987***	7.438***
Proportion of variance explained[a]	0.255	0.328

$\sim p < 0.10$; $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$. Comparison groups: White, male, 12th grader.

[a]Between proportion of variance explained = $(\tau_{\text{within model}} - \tau_{\text{between model}})/\tau_{\text{within model}}$.

For the multiple-choice assessment, only 2 of the 18 variables had a significant impact on student learning. Both variables were enactment variables: *average percentage of time spent on group work* and the *average percentage of time spent lecturing*. None of the teacher background or teacher belief variables significantly impacted student learning on the multiple-choice assessment. Including more group work had a positive impact on student learning while including more lecture had a negative impact on student learning. We then tested a model that included both variables, but with both variables included neither was significant most likely due to issues of multicollinearity, because the two variables are highly negatively correlated.[2] The final model in Table 8 only includes average percentage of time the teacher spent lecturing, since that variable had a lower significance level compared with group work. The greater the percentage of time the teacher spent lecturing had a negative effect on student achievement on the multiple-choice assessment, $t = -2.614$ (df $= 20$), $p < 0.05$. The bottom of Table 8 shows that the between-teacher variance for the intercept was still significant meaning that we have not explained away all of the between-teacher variance for science learning. Yet the inclusion of one variable, the percentage of time a teacher lectures, explained 25.5% of the variance in student achievement on the multiple-choice assessment between teachers.

For the open-ended assessment, again only 2 of the 18 variables had a significant impact; however, this model included different variables. One of the variables was an enactment variable: *average level of adaptation*. The more a teacher adapted the curriculum had a negative impact on student achievement as measured by the open-ended items. The other variable focused on teacher instructional practices: *frequency students engage in argument and share ideas*. The more frequently a teacher reported that they engaged their students in argumentation and sharing ideas at the beginning of the school year, the greater the student learning on the open-ended assessment. We then tested both variables in the same model. When both variables were included in the model, level of adaptation had a significant impact on student learning and frequency of argument and sharing ideas had a marginally significant impact. Because of the small sample size at level two (i.e. 22 teachers), we decided to leave in the marginally significant effect. This final model is included in Table 8. The more a teacher adapted the curriculum had a negative effect on student achievement on the open-ended assessment, $t = -2.186$ (df $= 19$), $p < 0.05$, and the more frequently a teacher reported that they spent time on argument and sharing of ideas had a marginally positive significant effect on student achievement on the open-ended assessment, $t = 1.871$ (df $= 19$), $p = 0.076$. The intercept variance is still significant suggesting that we have not explained all of the variance between teachers in terms of student learning on the open-ended responses. Yet with the inclusion of two teacher-level variables we were able to explain 32.8% of the variation between teachers in student learning in terms of the open-ended assessment, which included a greater focus on problem solving and scientific inquiry.

The student-level variables are also included in the final models in Table 8 (i.e. Fixed Effects). The significance and direction of these effects remained the same as the within-teacher models. The pre-test still had a positive significant effect in both models, with students who performed higher on the pre-test also scoring higher on the post-test. In terms of race, on average Black/African American students scored significantly lower on the multiple-choice assessment compared with White students and on average Asian/Pacific Islander students scored significantly lower on the open-ended assessment compared with White students.

Level of Teacher Adaptation

One interesting finding from the between-teacher model was that teachers who adapted the curriculum more had lower student learning in terms of the open-ended responses. This led us to question what types of adaptations the teachers made that led to a significantly negative impact on the assessment that focused on scientific inquiry and problem solving, but not on the multiple-choice assessment that focused on defining and identifying science concepts. The enactment survey only asked teachers about the level of adaptation and not specifically about what adaptations they made or why they made those changes (see Table 3). In order to explore this further, we determined the Pearson correlations between teachers' responses to the level of adaptation with their other responses on the enactment survey (see Table 9).

For the teachers who adapted more, they did not complete significantly more or less modules and the total number of periods and minutes they spent on the curriculum was not significantly different. Yet the level of completion of each module was significantly less. Since the amount of time was not different, but the teachers completed less, this suggests that they either spent more time on the activities in the curriculum or that they added other activities to the curriculum. Perhaps spending more time on activities or drawing them out or adding additional activities into the curriculum that did not align with the inquiry and problem-solving learning goals of the open-ended assessment can have a negative impact on student learning. In terms of activity structure, teachers who adapted the curriculum more were also more likely to spend a greater percentage of time lecturing. This suggests that the structure of these classrooms may have been more traditional or didactic and again did not align with the inquiry and problem-solving learning goals. This analysis is exploratory and cannot provide any causal claims. But it suggests the importance of future investigations

Table 9. Correlations between teacher adaptations and other enactment items

	Total # modules	Total # periods	Total # minutes	Level of completion	% Time individual work	% Time group work	% Time discussion	% Time lecture
Level of adaptation	−0.062	−0.064	0.041	−0.528*	−0.042	−0.247	−0.152	0.472*

*$p < 0.05$.

exploring not only the level of adaptation, but also the types of adaptations that teachers make to inquiry-oriented science curriculum materials.

Discussion

Curriculum is one important strategy to inform classroom instruction yet it is important to consider the role of the teacher in curriculum use (Ball & Cohen, 1996). Similar to previous research (Fogleman et al., 2011; Kanter & Konstantopoulos, 2010; Lawrenz et al., 2009; Puntambekar et al., 2007; Schneider et al., 2005), we found that there was significant variation in student achievement between teachers all of whom enacted the same inquiry-oriented science curriculum. Specifically, in terms of our first research question, we found that 34% of the variation on the multiple-choice assessment and 42.5% of the variation on the open-ended assessment was between teachers.

Interestingly, a greater proportion of the variation existed between teachers for the open-ended items, which included a focus on problem solving and scientific inquiry. Specifically, we designed these items to ask students to construct scientific arguments and environmental action plans, which were two key learning goals of the curriculum. This result suggests that there may be greater variation in how teachers support students in scientific inquiry learning goals compared with more traditional learning goals such as defining and identifying science concepts. This may be one reason that previous research has found significant differences in student achievement for more open-ended assessments, but not on more traditional multiple-choice assessments (Puntambekar et al., 2007). This also suggests that teachers may need greater support around instructional practices for scientific inquiry learning goals than traditional science content learning goals.

In terms of student-level variables, our second research question focused on the relationship between student demographics and pre-test scores with post-test achievement. We found that pre-test and race were associated with greater student learning. Students with higher incoming knowledge on the pre-test performed better on the post-test suggesting the importance of supporting science education in k-8 and not just in high school. Similar to previous research (Lawrenz et al., 2009; Von Secker & Lissitz, 1999), we found an achievement gap between minority students and their White peers. Specifically, Black or African American students scored lower on the multiple-choice assessment and Asian or Pacific Islander students scored lower on the open-ended items. One limitation of the study was that we were unable to collect data about each student's socioeconomic status or family income. If such a measure was included in our model, this may have impacted the effect of other student-level variables, particularly race, on student learning. Furthermore, our sample size of teachers was not large enough to model whether or not specific enactments or instructional practices narrowed or widened the achievement gap between students. Future research should include a measure of students' socioeconomic status as well as a larger number of teachers in order to more explicitly investigate a student achievement gap and the role of the teacher in narrowing this gap to better support all students in learning science.

Inquiry-based Teaching Practices

Our final research question examined the role of the teacher during the curriculum enactment on students' science learning. Inquiry-based teaching strategies can promote greater student learning compared with commonplace teaching strategies (Blanchard et al., 2010; Wilson et al., 2010). Specifically, we examined how different teaching strategies utilised during the enactment of the curriculum and based on teachers' self-reports about the frequency in which they used different strategies in their classroom impacted student learning. In terms of the multiple-choice assessment, which focused on defining and identifying science concepts, teachers' self-report of two teaching practices significantly impacted student learning—*average percentage of time spent on group work* and the *average percentage of time spent lecturing*. Specifically, teachers who reported that they had students spend a larger percentage of times engaged in group work had greater student learning and teachers who reported that they spent a greater percentage of time lecturing had less student learning. Of the two variables, the percentage of time lecturing had a greater impact on student learning and this one variable explained 25.5% of the student variation between teachers. These findings suggest that teachers who report having classrooms that align more with a community of learners perspective in which students actively construct their own knowledge (Sawyer, 2006), specifically through more group work and less lecture, results in greater student learning. Even when the learning goal is relatively simple, such as defining biodiversity, it is still more effective to actively engage students in constructing that understanding than simply disseminating the correct answer through lecture.

Although there is a range of acceptable enactments of curriculum, curriculum developers need to identify essential components that support student learning (Remillard, 2005). The results from this study suggest that key aspects of inquiry-based curriculum include: (1) engaging students in group work; and (2) limiting teacher-directed lecture. This focus on group work in high school science may be challenging for teachers and require a shift in practice and belief. Teachers who have more traditional beliefs about teaching science, such as a focus on teacher-directed instruction and teachers as the providers of knowledge, may struggle or resist integrating such strategies into their classroom practice (Johnson, 2009). Lotter et al. (2007) argue that in order for inquiry professional development to be successful that it must assess and address teachers' core beliefs about teaching and instruction. This also may be an important new avenue to explore in curriculum development. In addition to being educative, or supporting teacher learning, curriculum may also need to support changes in teacher beliefs in order to successfully impact classroom instruction.

For the open-ended items, teachers who reported that they spent more time having students engage in argument and share ideas had greater student learning on the open-ended assessment. These instructional practices align with the learning goals of the open-ended responses, which asked students to construct scientific arguments and develop environmental action plans. Recent science education reform documents

(Duschl, Schweingruber, & Shouse, 2006) and research (Berland & Reiser, 2009; Driver, Newton, & Osborne, 2000; Jiménez-Aleixandre & Erduran, 2008; McNeill, 2009; Sampson & Clark, 2008; Zembal-Saul, 2009) advocate for the importance of integrating scientific argumentation into k-12 classrooms in which students generate and evaluate scientific evidence and explanations, consider and debate alternative explanations and justify claims with appropriate evidence and reasoning. Science is fundamentally about argumentation in which scientists construct and debate multiple explanations for phenomena, not about memorising discrete facts (Osborne, Erduran, & Simon, 2004). A major goal of the urban ecology curriculum, and often of inquiry-based curricula more generally, is to support students in similar scientific practices in which they apply science concepts in the practice of argumentation.

Previous research suggests that the role of the teacher is essential for supporting students in argumentation in terms of both classroom discourse (McNeill & Pimentel, 2010; Simon, Erduran, & Osborne, 2006) and writing (Martin & Hand, 2009; McNeill & Krajcik, 2008). Yet there have been few large-scale studies linking teacher instructional practices for scientific argumentation to student outcomes. Specifically in this study, argumentation was not the sole focus of the study, but one of many teacher variables investigated to assess their impact on student learning. Yet teachers' self-report of the frequency of argumentation in their classroom significantly impacted student learning while a variety of other teacher variables, such as a teacher's comfort teaching students to design and conduct investigations or whether a teacher had a graduate degree in science or education, did not impact student learning. Similar to other research (Monk, 1994), we found that neither academic degrees nor experience significantly impacted student learning. Rather, the actual instruction that teachers chose to use in their classrooms had a greater effect on achievement than teacher background factors (Palardy & Rumberger, 2008). This highlights the importance of integrating scientific argumentation into science classrooms. If we want students to excel on student assessments that go beyond the memorisation of facts, teachers need to integrate instructional strategies such as having students justify their claims and critique alternative arguments using evidence and reasoning into their classroom practice.

A limitation of this study is that the teacher variables are dependent on teacher self-report. Self-report of teacher practices can be limited in part because of a lack of shared language between teachers and researchers (Thadani et al., 2009). The teachers could have interpreted the survey items in a different manner than we intended. Furthermore, teachers' beliefs about their instruction can differ from their actual instruction (Jones & Carter, 2007). Historically, teachers' self-report of curricular enactment has been questioned in terms of their validity compared with classroom instruction (Snyder, Bolin, & Zumwalt, 1992). The limited research that has occurred in this area suggests that self-report can be an accurate reflection of teachers' instructional practices in English classrooms (Koziol & Burns, 1986) and Mathematics classrooms (Herman, Klein, & Abedi, 2000; Mayer, 1999). However, a study in science education using a rubric adapted from the National Science Education Standards five essential features of inquiry (NRC, 2000) found a relatively low correlation ($r = 0.58$) between teacher self-report and external observers (Bodzin & Beerer, 2003). Since our survey was not validated with actual classroom observations, we can only make claims about the impact of teachers' self-report of classroom instruction on student achievement and not actual classroom instruction. Future research should use other measures such as observation protocols or video analysis to confirm the relationships between teacher instructional practices and student learning, which we found in this study. Furthermore, additional research needs to explicitly compare the validity of teacher self-report compared with classroom observations in science education.

Level of Teacher Adaptation

In terms of teacher enactment, the more teachers reported adapting the curriculum resulted in lower student learning in terms of the open-ended assessment. This raises the question of what types of adaptations were the teachers making that had a negative impact on student learning of problem solving and scientific inquiry. Consequently, we conducted an exploratory analysis examining the correlations between teacher self-report of the other enactment variables and the level of adaptation of the curriculum. Reporting a higher level of adaptation correlated with completing less of each module (though in the same amount of time) and spending a greater percentage of time on lecture. Teachers may have completed less of each module either because they were drawing out the activities to take more time or because they were adding their own additional activities, such as lecture, to the modules. These teachers may have adapted the curriculum to be more traditional or didactic; consequently, providing students with less support for the inquiry-oriented assessment items.

There is a tension around teacher adaptations of science curriculum. As we mentioned previously, we view curriculum use as a design activity (Brown & Edelson, 2001) in that teachers need to adapt materials in consideration of the resources and tools in their classroom environments. The process of curriculum adoption requires the adaptation of the curricula to meet the needs of the local context and culture (Barab & Luehman, 2003). Yet in this study, we found that teachers who reported making major adaptations actually resulted in less student learning in terms of the inquiry-oriented assessment items. Stein et al. (1996) found that the higher the cognitive demands of a task, the more likely teachers were to adapt the task during the implementation phase. Teachers decreased the cognitive demand of high-level tasks such as problem solving through strategies such as providing specific steps, telling students how to solve the problem, focusing on the correctness of the answer instead of the process, and spending either too little or too much time on the task. Specifically, around scientific argumentation teachers who simplify this complex scientific inquiry practice to include a greater focus on the surface-level features such as providing an algorithm or formula to complete the task have students with lower achievement in terms of their ability to construct written arguments in which they support their claims with appropriate evidence and reasoning (McNeill, 2009).

Although we were unable to obtain data on what specific adaptations the teachers made, it is possible that they were also decreasing the cognitive demand of the scientific inquiry tasks, which is why these adaptations had a detrimental effect on the inquiry-oriented assessment, but not on the multiple-choice assessment. Teachers may have inadvertently made 'lethal mutations' to the curriculum (Brown & Campione, 1996, p. 292) in terms of the inquiry learning goals. Lethal mutations occur when innovations or strategies become disconnected from the underlying learning principles and instead become surface activities that no longer support the overarching learning goal. Lethal mutations are an issue from the perspective of curriculum use as a design activity, because design principles and learning goals are hard to script. Rather, they require alignment of teacher belief and teachers having sufficient knowledge to make appropriate decisions in their classroom contexts.

Although the way in which teachers enacted the curriculum impacted student learning, the number of modules they completed or the amount of time they spent on the curriculum did not impact learning. Consequently, we found that fidelity of structure in terms of the amount of the curriculum and the time spent on the curriculum (O'Donnell, 2008) did not significantly impact student learning. This suggests that *how* teachers enact the curriculum is more important than *how much* of the curriculum they use. The role of the teacher is essential in the enactment of the curriculum materials. Between 34% and 42.5% of the variation in student achievement was a result of how the teachers used the curriculum materials in their classrooms. Specifically, we found that high school science teachers who spent a greater percentage of time on group work, a greater percentage of time on having students engage in argument and share ideas and a smaller percentage of time on lecture had students with greater science learning gains. These three instructional practices may be essential components of inquiry-based instruction for high school science classrooms.

Implications

We recommend that future educative curriculum materials and professional development emphasise essential instructional practices, such as limiting the use of lecture and increasing group work and argumentation, in order to support teachers in adapting curriculum to meet the needs of their students yet maintaining the key aspects of the curriculum to support student learning. Davis and Krajcik (2005) suggest that educative curriculum make visible developers' pedagogical judgments by making rationales for particular curriculum decisions explicit to teachers. Furthermore, educative features should be both lesson-specific to scaffold concrete instructional choices, but also support teachers' generative learning (Forbes & Davis, 2010b). Providing educative supports that highlight general principles of practice may help expand teachers' knowledge of effective science teaching (Beyer & Davis, 2009). Specifically, the results from this study suggest that it may be important to support high school science teachers in understanding why the activity structure for a specific lesson is set up as group work instead of as a lecture. Furthermore, it may be important to provide a rationale for engaging students in argumentation and sharing ideas rather than more traditional science instruction. Focusing on the rationales behind the key elements of the curriculum may help teachers in making informed decisions and avoid potentially lethal mutations for that particular lesson, but also support them in developing a deeper understanding that they can apply to their future instruction. Consequently, future research needs to examine the impact of different educative features of curriculum, such as providing both lesson-specific support and general rationales, on both teachers' enactment of curriculum and teacher learning.

The findings have similar implications for the design of future professional development. Pinto (2005) argues for the importance of providing professional development for innovative science curriculum particularly when the curriculum requires the teachers to play a new role. Specifically, professional development should include a focus on the 'critical details' which are essential aspects of the innovation (Viennot, Chauvet, Colin, & Rebmann, 2005). The results from this study suggest that use of lecture, group work and argumentation may all be critical details of high school science inquiry-oriented curriculum, which should be focused on in future professional development. Yet simply stating these characteristics as essential during the professional development may not be sufficient to support teacher learning. Instead, it is important to ground teacher learning in authentic classroom practice, such as utilising cases (Putnam & Borko, 2000). Specifically, the results of this study suggest that it may be important to use cases of high school science teachers effectively integrating group work as well as supporting students in argumentation and the sharing of ideas. Future research should explore the impact of integrating these critical details into professional development using a variety of strategies (such as videos, transcripts and student work) in terms of the effect on teachers' adaptations to curriculum and teacher learning.

Acknowledgements

This research was conducted as part of the Urban EcoLab project, supported in part by the National Science Foundation grant ESI 0607010. Any opinions expressed in this work are those of the authors and do not necessarily represent either those of the funding agency, Boston College or Loyola Marymount University. We would like to thank Michael Barnett, Christopher Harris, Jeremy Price, Meredith Houle Vaughn and Lisa Yamagata-Lynch as well as our colleagues at the Urban Ecology Institute for their work on this project.

Notes

1. Multicollinearity is when two or more predictors are highly correlated. Multicollinearity does not impact the power or reliability of the model as a whole, but it may not give valid results for any one individual predictor.
2. In terms of the enactment variables, percentage of time engaged in group work and percentage of time spent lecturing were significantly negatively correlated with each other, $r(22) = -0.613$, $p < 0.01$. Consequently, including both variables in the model does not provide valid results for each individual predictor.

References

Ball, D.L., & Cohen, D.K. (1996). Reform by the book: What is—or might be—the role of curriculum materials in teacher learning and instructional reform? *Educational Researcher*, *25*(9), 6–8, 14.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Banilower, E.R., Heck, D.J., & Weiss, I.R. (2007). Can professional development make the vision of the *Standards* a reality? Impact of the National Science Foundation's local systemic change through teacher enactment initiative. *Journal of Research in Science Teaching*, *44*(3), 375–395.

Barab, S.A., & Luehmann, A.L. (2003). Building sustainable science curriculum: Acknowledging and accommodating local adaptation. *Science Education*, *87,* 454–467.

Berland, L.K., & Reiser, B.J. (2009). Making sense of argumentation and explanation. *Science Education*, *93,* 26–55.

Beyer, C., & Davis, E.A. (2009). Supporting preservice elementary teachers' critique and adaptation of science lesson plans using educative curriculum materials. *Journal of Science Teacher Education*, *20,* 517–536.

Blanchard, M.R., Sutherland, S.A., Osborne, J.W., Sampson, V.D., Annetta, L.A., & Graner, E.M. (2010). Is inquiry possible in light of accountability?: A quantitative comparison of the relative effectiveness of guided inquiry and verification laboratory instruction. *Science Education*, *94,* 577–616.

Bodzin, A.M., & Beerer, K.M. (2003). Promoting inquiry-based science instruction: The validation of the science teacher inquiry rubric (STIR). *Journal of Elementary Science Education*, *15*(2), 39–49.

Brown, A.L., & Campione, J.C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In R. Glaser (Ed.), *Innovations in learning: New environments for education* (pp. 289–325). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Brown, M., & Edelson, D.C. (2001, April). *Teaching by design: Curriculum design as a lens on instructional practice*. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.

Davis, E.A., & Krajcik, J.S. (2005). Designing educative curriculum materials to promote teacher learning. *Educational Researcher*, *34*(3), 2–14.

Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, *84*(3), 287–312.

Duschl, R.A., Schweingruber, H.A., & Shouse, A.W. (Eds.). (2006). *Taking science to school: Learning and teaching science in grades k-8*. Washington, D.C: National Academy Press.

Environmental Literacy Council. (2004). *2004 review of environmental science textbooks*. Retrieved September 12, 2011, from http://www.enviroliteracy.org/article.php/1126.html.

Fogleman, J., McNeill, K.L., & Krajcik, J. (2011). Examining the effect of teachers' adaptations of a middle school science inquiry-oriented curriculum unit on student learning. *Journal of Research in Science Teaching*, *48*(2), 149–169.

Forbes, C.T., & Davis, E.A. (2010a). Curriculum design for inquiry: Preservice teachers' mobilization and adaptation of science curriculum. *Journal of Research in Science Teaching*, *47*(7), 820–839.

Forbes, C.T., & Davis, E.A. (2010b). Beginning elementary teachers' curriculum design and development of pedagogical design capacity for science teaching: A longitudinal study. In L.E. Kattington (Ed.), *Handbook of curriculum development* (pp. 209–232). New York: Nova Science Publishers.

Geier, R., Blumenfeld, P.C., Marx, R.W., Krajcik, J.S., Fishman, B., Soloway, E., & Clay-Chambers, J. (2008). Standardized test outcomes for students engaged in inquiry-based science curricula in the context of urban reform. *Journal of Research in Science Teaching*, *45*(8), 922–939.

Goodrum, D., Rennie, L., & Commonwealth of Australia. (2007). *Australian school science education national action plan 2008–2012: Volume 1*. Retrieved from http://www.dest.gov.au/sectors/school_education/publications_resources/profiles/Australian_School_Education_Plan_2008_2012.htm.

Herman, J.L., Klein, D.C. D., & Abedi, J. (2000). Assessing students' opportunity to learn: Teacher and student perspectives. *Educational Measurement: Issues and Practice*, *19*(4), 16–24.

Hollweg, K., Pea, C.H., & Berkowitz, A.R. (2003). Why is understanding urban ecosystems an important frontier for education and educators? In A.R. Berkowitz, C.H. Nilon, & K. Hollweg (Eds.), *Understanding urban ecosystems: A new frontier for science and education* (pp. 19–38). New York: Springer.

Jiménez-Aleixandre, M.P., & Erduran, S. (2008). Argumentation in science education: An Overview. In S. Erduran & M.P. Jimenez-Aleixandre (Eds.), *Argumentation in science education: Perspectives from classroom-based research* (pp. 3–28). Dordrecht: Springer.

Johnson, C.C. (2009). An examination of effective practices: Moving toward elimination of achievement gaps in science. *Journal of Science Teacher Education*, *20,* 287–306.

Jones, M.G., & Carter, G. (2007). Science teacher attitudes and beliefs. In S. Abell & N. Lederman (Eds.), *Handbook of research on science education* (pp. 1067–1104). Mahwah, NJ: Lawrence Erlbaum Associates.

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.

Kanter, D.E., & Konstantopoulos, S. (2010). The impact of a project-based science curriculum on minority student achievement, attitudes and careers: The effects of teacher content and pedagogical content knowledge and inquiry-based practices. *Science Education*, *94*(5), 855–887.

Kennedy, M.M., Ball, D.L., & McDiarmid, G.W. (1993). *A study package for examining and tracking changes in teachers' knowledge*. East Lansing, MI: The National Center for Research on Teacher Education.

Koziol, S.M. Jr., & Burns, P. (1986). Teachers' accuracy in self-reporting about instructional practices using a focused self-report inventory. *The Journal of Educational Research*, *79*(4), 205–209.

Lawrenz, F., Wood, N.B., Kirchhoff, A., Kim, N.K., & Eisenkraft, A. (2009). Variables affecting physics achievement. *Journal of Research in Science Teaching*, *46*(9), 961–976.

Lee, O., Penfield, R., & Maerten-Rivera, J. (2009). Effects of fidelity of implementation on science achievement gains among English language learners. *Journal of Research in Science Teaching*, *46*(7), 836–859.

Lotter, C., Harwood, W.S., & Bonner, J.J. (2007). The influence of core teaching conceptions on teachers' use of inquiry teaching practices. *Journal of Research in Science Teaching*, *44,* 1318–1347.

Martin, A.M., & Hand, B. (2009). Factors affecting the implementation of argument in the elementary science classroom. A longitudinal case study. *Research in Science Education*, *39*(1), 17–38.

Marx, R.W., Blumenfeld, P.C., Krajcik, J.S., & Soloway, E. (1997). Enacting project-based science. *The Elementary School Journal*, *97*(4), 341–358.

Marx, R.W., & Harris, C.J. (2006). No child left behind and science education: Opportunities, challenges and risks. *The Elementary School Journal*, *106*(5), 467–477.

Mayer, D.P. (1999). Measuring instructional practice: Can policymakers trust survey data? *Educational Evaluation and Policy Analysis*, *21*(1), 29–45.

McNeill, K.L. (2009). Teachers' use of curriculum to support students in writing scientific arguments to explain phenomena. *Science Education, 93*(2), 233–268.

McNeill, K.L., & Krajcik, J. (2008). Scientific explanations: Characterizing and evaluating the effects of teachers' instructional practices on student learning. *Journal of Research in Science Teaching, 45*(1), 53–78.

McNeill, K.L., & Pimentel, D.S. (2010). Scientific discourse in three urban classrooms: The role of the teacher in engaging high school students in argumentation. *Science Education, 94*(2), 203–229

Ministry of Education. (1999). *Curriculum outlines for nature of science and living technology.* Taipei: Ministry of Education.

Minner, D.D., Levy, A.J., & Century, J. (2010). Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984–2002. *Journal of Research in Science Teaching*, *47*(4), 474–496.

Monk, D.H. (1994). Subject area preparation of secondary mathematics and science teachers and student achievement. *Economics of Education Review*, *13*(2), 125–145.

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

National Research Council. (2000). *Inquiry and the national science education standards: A guide for teaching and learning.*. Washington, DC: National Academies Press.

O'Donnell, C.L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in k-12 curriculum intervention research. *Review of Educational Research*, *78*(1), 22–84.

Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, *41*(10), 994–1020.

Palardy, G.J., & Rumberger, R.W. (2008). Teacher effectiveness in first grade: The importance of background qualifications, attitudes, and instructional practices for student learning. *Educational Evaluation and Policy Analysis*, *30,* 111–140.

Pinto, R. (2005). Introducing curriculum innovations in science: Identifying teachers' transformations and the design of related teacher education. *Science Education*, *89,* 1–12.

Puntambekar, S., Stylianou, A., & Goldstein, J. (2007). Comparing classroom enactments of an inquiry curriculum: Lessons learned from two teachers. *The Journal of the Learning Sciences*, *16*(1), 81–130.

Putnam, R.T., & Borko, H. (2000). What do new views of knowledge and thinking have to say about research on teacher learning? *Educational Researcher*, *29*(1), 4–15.

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

Remillard, J.T. (2005). Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, *75*(2), 211–246.

Remillard, J.T., & Bryans, M.B. (2004). Teachers' orientations toward mathematics curriculum materials: Implications for teacher learning. *Journal for Research in Mathematics Education*, *35*(5), 352–388.

Rice, J.K. (2003). *Teacher quality: Understanding the effectiveness of teacher attributes*. Washington, DC: Economic Policy Institute.

Sampson, V., & Clark, D.B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, *92,* 447–472.

Sawyer, R.K. (2006). Analyzing collaborative discourse. In R.K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 187–204). New York, NY: Cambridge University Press.

Schneider, R.M., Krajcik, J., & Blumenfeld, P. (2005). Enacting reform-based science materials: The range of teacher enactments in reform classrooms. *Journal of Research in Science Teaching*, *42*(3), 283–312.

Shadish, W.R., Cook, T.D., & Campbell, D.T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton Mifflin Company.

Simon, S., Erduran, S., & Osborne, J. (2006). Learning to teach argumentation: Research and development in the science classroom. *International Journal of Science Education*, *28*(2–3), 235–260.

Singer, J., Marx, R.W., Krajcik, J., & Chambers, J.C. (2000). Constructing extended inquiry projects: Curriculum materials for science education reform. *Educational Psychologist*, *35*(3), 165–178.

Snyder, J., Bolin, F., & Zumwalt, K. (1992). Curriculum implementation. In P.W. Jackson (Ed.), *Handbook of research on curriculum* (pp. 402–435). New York: Macmillan.

Stein, M.K., Grover, B.W., & Henningsen, M. (1996). Building student capacity for mathematical thinking and reasoning: An analysis of mathematical tasks used in reform classrooms. *American Educational Research Journal*, *33*(2), 455–488.

Strauss, E.G., McNeill, K.L., Barnett, M., & Reece, F. (2008). *Urban EcoLab: How do we develop healthy and sustainable cities?* Chestnut Hill, MA: Boston College. Retrieved September 12, 2011, from http://urbanecolabcurriculum.com/.

Tabak, I. (2004). Synergy: A complement to emerging patterns in distributed scaffolding. *The Journal of the Learning Sciences*, *13*(3), 305–335.

Thadani, V., Stevens, R.H., & Tao, A. (2009). Measuring complex features of science instruction: Developing tools to investigate the link between teaching and learning. *The Journal of the Learning Sciences*, *18*(2), 285–322.

Tschannen-Moran, M., Hoy, A.W., & Hoy, W.K. (1998). Teacher efficacy: Its meaning and measure. *Review of Educational Research*, *68*(2), 202–248.

UNFPA. (2007). *UNFPA State of World Population 2007: Unleashing the Potential of Urban Growth*, United Nations Population Fund. Alexandria, VA: ACSD.

Viennot, L., Chauvet, F., Colin, P., & Rebmann, G. (2005). Designing strategies and tools for teacher training: The role of critical details examples in optics. *Science Education*, *89*, 13–27.

Von Secker, C.E., & Lissitz, R.W. (1999). Estimating the impact of instructional practices on student achievement in science. *Journal of Research in Science Teaching*, *36*(10), 1110–1126.

Wilson, C.D., Taylor, J.A., Kowalski, S.M., & Carlson, J. (2010). The relative effects and equity of inquiry-based and commonplace science teaching on students' knowledge, reasoning, and argumentation. *Journal of Research in Science Teaching*, *47*(3), 276–301.

Wilson, S.M., Floden, R.E., & Ferrini-Mundy, J. (2002). Teacher preparation research: An insider's view from the outside. *Journal of Teacher Education*, *53*(3), 190–204.

Zembal-Saul, C. (2009). Learning to teach elementary school science as argument. *Science Education*, *93*, 687–719.

Appendix 1. Student Assessment—Multiple-Choice Items

1. (Module 1) In a particular area, living organisms and the nonliving environment function together as:
   a. a population
   b. a community
   c. an ecosystem
   d. a species
2. (Module 2) Urban heat islands are mainly caused by increased:
   a. amount of vegetation
   b. population
   c. dark surfaces
   d. pollution
3. (Module 3) Climate change:
   a. is a shift in long-term weather patterns

b. is only caused by human activities
c. explains a previous summer's heat wave
d. explains the decrease in available fossil fuels

4. (Module 4) Bodies of water within our community are polluted most seriously by:
   a. People participating in water sports
   b. Storm water runoff after it rains
   c. Animals that live near the water
   d. Air pollution that dissolves into the water
5. (Module 5) The ozone in the air does **NOT**:
   a. change depending on temperature
   b. damage people's lungs
   c. cause global warming
   d. depend on the amount of green space
6. (Module 6) When is it particularly important for an ecosystem to contain many different species?:
   a. When ecosystems remain stable over long periods of time
   b. When significant changes occur in the ecosystem
   c. When natural selection does not occur
   d. When the finite resources of Earth increase
7. (Module 7) Which characteristic will be most advantageous to an animal species living in a rapidly changing environment?:
   a. Obtaining food from a variety of sources
   b. Using one specific material to make a shelter
   c. Seeing the environment in colour
   d. Having a long life span

Appendix 2. Student Assessment—Problem-Solving Items

2. (Action Plan) An urban neighborhood has asked you to develop an action plan to promote public health. Listed below are some of the characteristics of the neighborhood.

Characteristic	Description
Green space	Limited access to local green space.
Access to fresh fruits and vegetables	Low access to fresh fruits and vegetables
Access to fast food	High access to fast food restaurants
Air quality	High ground level ozone level
Bird biodiversity	Low levels of bird biodiversity
Measures of Health	High incidences of heart disease and asthma, and a low life expectancy.
Violent Crime	Incidence of violent crime is at about the same level as the rest of the city.

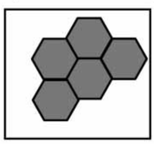
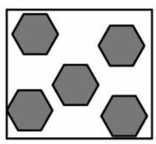
In your action plan, make sure to address the following:

- Define the goals and possibilities for improving public health in the neighborhood;
- Identify key stakeholders;
- Figure out what resources are necessary; and
- Describe actions that will allow the neighborhood to improve public health.

Explain how your action plan will improve the public health of the neighborhood.

4. (Scientific Argument) A developer is building houses in a new section of the city. The city is trying to preserve a section of land as urban wild (where houses cannot be built). As a part of this process, scientists collected data on a variety of rabbit species in each location. Each rabbit requires an area of approximately 15,000 square yards as a territory. In addition, the population of rabbits must be able to interact in order to breed and disperse their young. An isolated population of rabbits will likely go extinct.

The city has enough money to purchase, and preserve, all of the grey area in Location A or all of the grey area in Location B. The white areas have already been developed by humans (e.g. houses, roads, etc.) and are not suitable for the rabbits to survive.

	Location A	**Location B**
Data for the Rabbits	25,000 square yards (5 football fields) 	25,000 square yards (5 football fields) 
Species Richness	5	9
Species Abundance	143	98
Species Distribution	even	even
Shannon-Weaver Index	1.6	2.19
Simpson's Index	0.8	0.9

Would you preserve location A or location B?

Write a scientific argument justifying which location you would preserve. Use evidence from the table above about the rabbits, and provide reasoning for why the evidence supports your claim.