MONTGOMERY COUNTY PUBLIC SCHOOLS

Evaluation of the High School Biology Curriculum Pilot

Office of Shared Accountability

August 2015

Elizabeth Cooper-Martin, Ph.D.
Natalie Wolanin

Montgomery County Public Schools

OFFICE OF SHARED ACCOUNTABILITY

850 Hungerford Drive
Rockville, Maryland 20850
301-279-3553

Mr. Larry A. Bowers
Interim Superintendent of Schools

Dr. Maria V. Navarro
Chief Academic Officer

Table of Contents

List of Tables

List of Figures

Executive Summary

The Maryland State Department of Education has a goal for all Maryland school systems to adopt the scope and sequence of the Next Generation Science Standards and to align their curriculum documents to these new, rigorous standards by 2019. To incorporate the new starndards, Montgomery County Public Schools (MCPS) is revising the high school biology curriculum. The MCPS revision utilizes a problem-based learning (PBL) approach because it was thought to be an engaging and effective way for students to learn. A group of 16 MCPS high schools voluntarily piloted one unit of the revised biology curriculum in 2014–2015. In this unit on ecology, students examined interdependent relationships in ecosystems. The problem related to controlling an invasive species; for their unit project, students devised a culminating management plan through a series of ongoing investigations throughout the unit.

This evaluation of the new curriculum for high school biology focused on implementation of the new invasive species unit and its impact on students, by addressing the following questions:

1. To what extent do teachers in pilot schools implement PBL practices with fidelity?
2. To what extent do teachers in pilot schools implement with fidelity the lesson components and activities as presented in the lesson plans for the invasive species unit?
3. Does students' level of engagement with the ecology unit depend on which curriculum the teacher used? Is the result the same for each student subgroup (i.e., Black or African American, Hispanic/Latino, and Two or More Races; Free and Reduced-price Meals Systems recipients; limited English proficiency; special education services recipients)?
4. Does student achievement in biology depend on which curriculum the teacher used? Is the result the same for each student subgroup?

To answer questions 1 and 2, this evaluation utilized classroom observations at all pilot schools (i.e., only the high schools piloting the invasive species unit). To answer questions 3 and 4, student surveys and unit assessments were collected from pilot high schools and also from all nonpilot high schools, to allow for comparisons between the PBL curriculum used at pilot schools and the traditional (i.e., non-PBL) curriculum used at the nonpilot schools.

Summary of Findings

In this study, a moderate level of implementation means one third or more, but less than two thirds, of the classroom observation sample. A high level of implementation means more than two thirds of the classroom observation sample.

Question 1. Observers collected data on seven PBL practices during 61 classes at pilot schools. Two PBL practices, central problem (i.e., the teacher connects the lesson to the unit project) and student choice (i.e., the teacher gives students opportunities to make meaningful choices about the unit project), were critical. Evidence of either of these practices indicated adequate implementation of PBL. Also, if a teacher used at least three PBL practices and one of them was either constructive investigation (i.e., an indication that multiple solutions to a problem are possible) or student-centered learning (i.e., teacher functions as a facilitator), implementation of PBL was considered adequate. Based on these criteria, teachers in more than than three quarters of the observed classes (79%) had adequate implementation of PBL, indicating a high level of implementation of PBL practices in pilot schools.

Question 2. For each of the 61 observations at pilot schools, observers rated the extent to which the observed lesson matched the components and activities described in the lesson plan. The fidelity of implementation to the lesson plan was exact or with slight variations for the majority of classes (57%). This moderate level of implementation of lesson plans for the invasive species unit in pilot schools is reasonable given that the evaluation occurred during the first year of implementation.

Question 3. There was some evidence that student engagement was higher at the pilot schools, as expected. Three types of engagement were examined, with the following results.

The level of cognitive engagement (i.e., commitment to learning and understanding the content) was significantly higher among students at pilot schools than at nonpilot schools; this difference was large enough to be useful in an educational setting. The relationship was similar for most subgroups; the mean level of cognitive engagement was higher in pilot schools, but the difference was not statistically significant.

The level of emotional engagement (i.e., excitement and enjoyment of class), was higher at pilot schools than at nonpilot schools for students in on-level biology, but was lower at pilot than at nonpilot schools for those in honors biology. This pattern of differences was statistically significant and large enough to be useful in an educational setting. For most student subgroups, this pattern of differences was the same, but some of the differences were not significant. However, the level of emotional engagement for recipients of special education services (in on-level and honors classes) was significantly higher at pilot schools.

The level of psychological engagement (i.e., connection to team and project solution) did not differ between students at pilot and nonpilot schools. The result was the same for each student subgroup.

Question 4. There was some evidence that assessment scores were higher at the pilot schools than at nonpilot schools for all students and most student subgroups. The largest effect was for students with limited English proficiency. Although the difference for this subgroup was large enough to be useful in an educational setting, the differences for other groups were too small to be useful.

Recommendations

As program staff support the second year of implementation of the new biology curriculum, they should focus training, guidance, and other supports on the following suggestions:

- Place more emphasis on teacher implementation of those PBL practices that are more critical and also have the lowest levels of implementation: connection to the central problem/unit project, student choice related to the unit project, and constructive investigation. Communication of expectations in lesson plans, along with professional learning opportunities, both may be helpful in increasing implementation.
- Encourage higher levels of fidelity to the invasive species lesson plans by overcoming some of teachers' concerns with pacing, specific lesson content, and lesson clarifications.
- Explore ways to increase the level of emotional engagement (i.e., excitement and enjoyment of class) among honors level students; the lesson plans were primarily written for on-level students, with some honors distinctions/extensions, so additional honors distinctions/extensions may be helpful.
- Explore ways to increase the level of psychological engagement (i.e., connection to team and unit project solution) among all students.

Evaluation of the High School Biology Curriculum Pilot

The Office of Curriculum and Instructional Programs asked the Office of Shared Accountability (OSA) to conduct an evaluation of an updated curriculum for high school biology. The evaluation focused on implementation of the updated curriculum and its impact on students.

Background

In spring 2013, the Maryland State Department of Education (MSDE) officially adopted the Next Generation Science Standards (NGSS), a new set of rigorous and internationally benchmarked standards for K–12 science education (MSDE, 2013). These standards were developed to identify science and engineering practices and content that all K–12 students should master to be fully prepared for college and careers. MSDE has a goal for all Maryland school systems to adopt the NGSS scope and sequence and to align their curriculum documents to NGSS by 2019.

The NGSS outlines a curriculum sequence from kindergarten to Grade 12 focused on four disciplinary core ideas: physical science; life science; earth and space sciences; and engineering, technology, and application of science (NGSS, 2013). The NGSS core ideas specify what information should be taught and at what level; the focus is on depth of coverage rather than the amount of information covered. Along with what students need to learn, the NGSS emphasizes how students apply their content knowledge by outlining scientific and engineering practices and crosscutting concepts. The practices are those that scientists and engineers use when they investigate, create theories, design, build, and communicate. The crosscutting concepts link ideas and practices that are common across science and engineering courses.

To incorporate NGSS, the Science, Technology, and Engineering (STE) team within Montgomery County Public Schools (MCPS), along with a group of school-based writers, is revising the curriculum for high school biology. The MCPS revision utilizes a problem-based learning (PBL) approach because research indicates that PBL is an engaging and effective way for students to learn (Buck Institute for Education, 2014a). PBL is a teaching method in which students gain content knowledge and skills, such as problem solving, critical thinking, collaboration, communication, and creativity, by investigating and responding to a complex question, problem, or challenge over time (Buck Institute for Education, 2014b). Students realize that they need to gain knowledge and understand concepts to answer the driving problem. With the teacher as a facilitator, students are allowed to make choices about the products they create and how they complete the project tasks. Additionally, they give and receive feedback about their product so that they can investigate further and make revisions.

A group of 16 MCPS high schools voluntarily piloted one unit of updated biology curriculum in 2014–2015; all schools are expected to implement a minimum of two units (one per semester) of the revised biology curriculum in 2015–2016. To support development of instructional resources and curriculum materials and the roll-out of the updated curriculum, this study examined the implementation of the revised unit and its impact on students during the 2014–2015 school year.

Program Description

Background and Goals

Next Generation Science Standards

The updating of the MCPS high school biology curriculum will reflect the Life Science student performance expectations of the NGSS. The goal of revising the biology curriculum is to provide students with explicit opportunities to demonstrate the eight scientific and engineering practices outlined in NGSS. Figure 1 lists all the practices, along with a brief definition for each one.

Asking Questions and Defining Problems A basic practice of the scientist is the ability to formulate empirically answerable questions about phenomena to establish what is already known, and to determine what questions have yet to be satisfactorily answered.
Developing and Using Models Science often involves the construction and use of models and simulations to help develop explanations about natural phenomena.
Planning and Carrying Out Investigations A major practice of scientists is planning and carrying out systematic scientific investigations that require identifying variables and clarifying what counts as data.
Analyzing and Interpreting Data Scientific investigations produce data that must be analyzed to derive meaning. Scientists use a range of tools to identify significant features and patterns in the data.
Using Mathematics, Information & Computer Technology, and Computations Thinking In science, mathematics and computation are fundamental tools for representing physical variables and their relationships.
Constructing Explanations and Designing Solutions The goal of science is the construction of theories that provide explanatory accounts of the material world.
Engaging in Argument from Evidence In science, reasoning and argument are essential for clarifying strengths and weaknesses of a line of evidence and for identifying the best explanation for a natural phenomenon.
Obtaining, Evaluating, and Communicating Information Science cannot advance if scientists are unable to communicate their findings clearly and persuasively or learn about the findings of others.

Figure 1. Next Generation Science Standards scientific practices.

Additionally, the revised biology curriculum includes crosscutting concepts. An example is cause and effect; as high school students progress through the biology course, they will build on their earlier exposure to the concept that empirical evidence is necessary to distinguish between causality and correlation and to make claims about specific causes and effects.

Problem-Based Learning Overview

The revised biology curriculum uses the PBL approach; this teaching method allows students to gain content knowledge and content-relevant skills by working on a complex project, problem, or challenge for an extended period of time (Buck Institute for Education, 2014b). Some of the essential elements of PBL are the following:

- **Central problem**. There is a "driving question" or central problem that focuses the project work; students must understand the problem and find it intriguing (Larmer & Mergendoller, 2012). Further, it frames their study and exploration.
- **Collaboration**. Students work together or collaborate in small cooperative teams to complete their project, with guidance from the teacher on roles within the team and instruction on collaboration and communication (Larmer & Mergendoller, 2012).
- **Constructive investigations**. Projects involve students in a constructive investigation that includes inquiry, knowledge building, and resolution (Thomas, 2000). Students must develop new understandings and new skills, not simply apply already-learned information or skills. The central problem is open-ended and complex, without a single solution.
- **Need to know**. Students are motivated to learn new concepts, understand content, and master skills so that they can answer the central problem and create their projects (Larmer & Mergendoller, 2012).
- **Student choice**. Students can choose how they work and also have some choices about the products they create. The teacher guides them, as appropriate, for their age level and PBL experience (Larmer & Mergendoller, 2012).
- **In-depth inquiry**. Students engage in an extended and rigorous process of generating questions, using resources, and creating answers (Larmer & Mergendoller, 2012). Each team identifies what resources to consult, finds answers to the team's questions, and constructs its own response or solution to the central problem.
- **Student-driven**. The student projects or problems are not teacher-led (Thomas, 2000). Instead they are student-driven; the teacher functions as a facilitator. Students take responsibility for their learning in a variety of ways: use resources other than the teacher to get information; make choices on topics, products, or resources; collaborate with other students to solve the problem; pose questions; and give and receive feedback on the quality of their solutions to the problem (Larmer & Mergendoller, 2012).

Problem-Based Learning in MCPS

MCPS began using PBL in science classes in 2008 at the middle school level; Investigations in Science 6, Investigations in Science 7, and Earth Space Systems (for Grade 8) courses are each problem based. Instruction was interwoven around a relevant problem to provide a focus for student learning and to allow students to engage in science, technology, engineering, and mathematics in order to propose solutions to identified problems.

MCPS has incorporated the PBL approach in several high school science and technology courses: Foundations of Horticulture, Foundations of Technology, Geoscience Explorations, and Integrated & Applied Physical Science. Currently the high school biology course is under revision using the PBL approach. Biology is a required course for high school graduation in Maryland; further, graduation requirements include passing the Maryland High School Assessments, one of which is biology (MSDE, 2013). Most MCPS students take biology in Grade 10. However, schools may choose to have the majority of their students take this class in Grade 9 or 11.

In the revised MCPS biology curriculum, which will be available 2015–2016, each unit will begin with a "real-world" problem that is complex and has several possible solutions. The students are presented with this problem in the form of a Request for Proposal (RFP). Each lesson sequence is aligned to the problem; as the students learn science content, the content helps them to solve the problem. At the end of each lesson sequence, teams of students produce part of the solution to the problem. At the end of the unit, the teams communicate their full solution to others.

The pilot unit for the biology curriculum was a five week unit on ecology in which students examined interdependent relationships in ecosystems. The lesson plans for the pilot were primarily for on-level students but included some distinctions/extensions for honors students. The problem related to controlling an invasive species; for their unit project, students devised a culminating management plan through a series of ongoing investigations throughout the unit.

Participating Schools and Students

High school staff volunteered to teach the pilot unit of biology during the first semester of the 2014–2015 school year. The pilot schools included the following group of 16 high schools:

- Montgomery Blair
- Bethesda-Chevy Chase
- Clarksburg
- Albert Einstein
- Gaithersburg
- John F. Kennedy
- Col. Zadok Magruder
- Richard Montgomery
- Northwest
- Quince Orchard
- Seneca Valley
- Sherwood
- Springbrook
- Walter Johnson
- Wheaton
- Walt Whitman

The expectation was that the entire biology team at each pilot school would deliver the revised unit, including common assessments, in on-level as well as honors biology classes. The biologty team sent data to the STE team for two identified formative assessments and also completed feedback documents on what worked and what needed changed. At most pilot schools, the five-week invasive species unit was either the first one of the semester (i.e., August–September 2014) or the final unit (i.e., December 2014–January 2015). At nonpilot schools, teachers taught the ecology unit from the traditional curriculum, instead of the new invasive species unit.

Supports Provided

During summer 2014, a week-long (five-day) professional learning session about the revised biology curriculum was offered to teachers at all high schools including pilot and nonpilot. One day of this session was devoted to the invasive species unit; content for the remaining days included the NGSS standards, the PBL vision for biology, and a glimpse of the other PBL units for the revised biology curriculum. At least two teachers from each pilot school were expected to attend the session. Using a trainer of trainers model, these representatives were expected to bring the professional learning related to the invasive species unit back to the rest of the biology team at each pilot school.

In addition to the summer 2014 session, pilot teachers may have had experience with PBL or NGSS standards due to teaching middle school or high school science courses with a PBL curriculum, writing the PBL units for biology, training other teachers for the PBL unit, attending professional development on PBL for biology resource teachers during summer 2013, attending professional learning sessions on the scientific practices during 2013–2014, or coaching from resource teachers on the scientific practices.

The entire curriculum for the invasive species unit, including lesson plans for each topic, how-to strategies (e.g., descriptions of ways to create various types of student groups), and several teacher resources (e.g., pacing guide), was available online to teachers. Further, each biology teacher at pilot schools received two full-day professional leave substitutes to support the team with pacing, planning, data analysis, and unit feedback with respect to the invasive species unit.

Because the items on the county semester exam for biology are not aligned with NGSS, teachers at the pilot schools could create a final exam from an approved list of questions developed by the STE program staff.

Expected Student Outcomes

The revised biology curriculum should result in several benefits to students. The expected student outcome of incorporating NGSS is that students have explicit opportunities throughout the course to demonstrate the eight NGSS scientific and engineering practices (see Figure 1).

With PBL, students are engaged in in-depth inquiry, and the real-world problem (e.g., unit project) is focused on an open-ended question that students find intriguing and relevant. With the teacher as a facilitator, they are allowed to make choices about the products they create and how they complete the projects' tasks. Thus, one of the expected student outcomes is higher engagement. In particular, students should be interested and excited about the PBL class, active as participants and contributors during the problem-solving activities in class, and committed to learning and understanding the content.

With PBL, students realize that they need to gain knowledge and understand concepts to solve the real-world problem. Because students understand content more deeply and are more engaged in their learning, the PBL approach is expected to improve students' mastery of the biology content.

Literature Review

Problem-Based Learning

The evaluation examines both fidelity of implementation of PBL and its impact on two student outcomes: student engagement and achievement. Therefore, this review includes both implementation and outcome evaluations of PBL curriculum, starting with MCPS studies.

MCPS Research

Investigations in Science 6. Wang and Cooper-Martin (2010) evaluated the extent and quality of implementation of three advanced middle school courses, including Investigations in Science 6 which utilized a PBL approach. This implementation evaluation focused on instructional practices that were key to rigorous instruction, including content-specific practices. For Investigations in Science 6, this practice was the engineering design process, which reflected the PBL approach. Data from classroom observations indicated a low level of implementation of the PBL practices in the Investigations in Science 6 classes.

SCALE-uP. Scaling up Curriculum for Achievement, Learning, and Equity Project (SCALE-uP), was a collaborative effort between The George Washington University and MCPS. SCALE-uP was designed to implement and scale up three different middle school science curriculum units over a six-year period (Merchlinsky & Hansen-Grafton, 2007). The curriculum units incorporated a PBL approach. Classroom observations were used to examine fidelity of implementation; at least 80% of the teachers mostly or fully implemented 10 of the 11 indicators (Lastica & O'Donnell, 2007). With respect to outcomes, students in both SCALE-uP and comparison classrooms completed pretests, posttests, and delayed posttests on content assessment and on measures of motivation and engagement (Lynch, Kuipers, Pyke, & Szesze, 2005). Assessment scores, basic learning engagement, and goal orientation were all significantly higher for students in the SCALE-uP classes than those in the comparison classes who did not receive PBL instruction.

External Research

Research on PBL has identified several positive effects for students (see summary in Center of Excellence in Leadership Learning, 2009). Given the context and scope for the evaluation, the following sections are limited to studies that examined the impact of PBL on student achievement or on student engagement and concerned science or high school classes. The following section includes only outcome studies, because no external implementation evaluations of PBL were identified.

PBL and student achievement. A pair of studies examined middle school students who participated in PBL science courses that used a PBL approach. These seventh and eighth graders gained more science content understanding, as measured by performance on assessments aligned to the curriculum (Marx et al., 2004), and had higher pass rates on their state standardized test in science (Geier et al., 2008), than their peers in a non-PBL version of the course. Positive results also were found in a study of a high school economics course; students in the PBL classes outperformed the students in non-PBL classes on a content based assessment (Finkelstein, Hanson, Huang, Hirschman, & Huang, 2010).

PBL and student engagement. Brush and Saye (2008) studied a PBL unit in a high school history class; they defined engagement as “the student’s psychological investment in and effort directed toward learning, understanding, and mastering the knowledge, skills, or crafts that academic work is intended to promote” (Newmann, Whelage, & Lamborn, 1992, p. 12). Brush and Saye determined the level of engagement by examining students’ levels of participation in classroom activities, their enthusiasm and interest expressed in tasks, and the degree of care demonstrated in completing their assignments, as revealed in videotaped observations of all class activities and student presentations. Based on these data, the authors concluded that students had high levels of engagement.

Student Engagement

To guide the evaluation’s study of the impact of PBL instruction on student engagement, the investigators reviewed literature on this topic as follows.

In a review of research on student engagement, Fredericks, Blumenfeld, and Paris (2004) identified and defined three types of engagement: behavioral, cognitive, and emotional. Common definitions for behavioral engagement were positive conduct, involvement in learning, involvement in academic tasks, or participation in school-related activities. The authors identified two sets of definitions for cognitive engagement: 1) investment in learning; and 2) being strategic or self-regulating. The former includes perseverance in the face of challenges, and use of deep rather than superficial strategies. Lastly, emotional engagement has been defined as affective reactions, emotional reactions to school or the teacher, and as feelings of belonging. The authors proposed that engagement is a multidimensional construct which includes the three types of engagement.

Like Fredericks et al. (2004), Appleton, Christenson, Kim, and Reschly (2006) proposed that student engagement is multidimensional, but included four types: behavioral, cognitive, academic, and psychological. They identified indicators of each type as follows:

- Academic—time on task, homework completion
- Behavioral—attendance, extra credit options, voluntary classroom participation
- Cognitive—self-regulation, value of learning, personal goals, strategizing, relevance of school to future aspirations
- Psychological—feelings of belonging or identification with school, relationships with students or teachers

With respect to measuring student engagement, Fredricks et al. (2011) identified and described 21 instruments for measuring this construct in upper elementary through high school. They identified 14 student self-reports, 4 observational reports, and 3 teacher surveys. The instruments varied on which dimensions of engagement (i.e., behavioral, cognitive, or emotional) were assessed and also by focus (i.e., school or class).

A recent MCPS evaluation, like the evaluation of the biology curriculum, examined both fidelity of implementation by teachers and the impact of that implementation on student engagement (Cooper-Martin & Wolanin, 2014). It concerned Universal Design for Learning (UDL), an instructional framework that seeks to give all students equal opportunities to learn by providing multiple means of representation, of action and expression, and of engagement. Cooper-Martin and Wolanin used classroom observations to measure the extent of implementation of UDL instructional strategies and used student self-reports to measure three types of student engagement: academic, affective, and cognitive. For Grades 3–5 students, they found that only affective engagement was significantly higher for students in UDL classes than for students in comparison (i.e., non-UDL) classes. For students in Grades 6–8, all three types of engagement were significantly higher for students in UDL classes than for those in comparison classes.

Scope of Evaluation

This evaluation examined the fidelity and extent of implementation of the curriculum for the invasive species unit of the MCPS high school biology course and its impact on student engagement and achievement for that unit. Findings will provide feedback to the program staff to improve instructional resources, curriculum materials, and implementation of the updated biology curriculum.

Evaluation Questions

1. To what extent do teachers in pilot schools implement PBL practices with fidelity? Possible practices to investigate include the following:

    - Student centered learning with the teacher as a facilitator
    - Small, cooperative work groups
    - An overarching, central problem to motivate learning
    - Constructive investigations
    - Student choice
    - In-depth inquiry

2. To what extent do teachers in pilot schools implement with fidelity the lesson components and activities as presented in the lesson plans for the invasive species unit?

3. Does students' level of engagement with the ecology unit depend on which curriculum (i.e., invasive species vs. traditional) the teacher used? Is the result the same for each student subgroup, such as racial/ethnic groups?

4. Does student achievement in biology depend on which curriculum (i.e., invasive species vs. traditional) the teacher used? Is the result the same for each student subgroup, such as racial/ethnic groups?

Methodology

Design and Participating Schools

To answer evaluation questions 1 and 2 on classroom implementation of PBL practices and of lesson plans for the invasive species unit, this study used classroom observations. The sample for observations included only the 16 high schools piloting the invasive species unit. Due to resource limitations, there were no observations at nonpilot high schools.

To answer evaluation questions 3 and 4 on the impact of the revised curriculum on student engagement and achievement, this study utilized student surveys for engagement and unit tests for achievement. To provide a comparison group of students, the samples for these data collection activities included the nine nonpilot MCPS high schools along with all the high schools piloting the new unit.

Based on program materials and previous research, the evaluators, in collaboration with STE team staff members, developed instruments for the data collection activities.

Data Collection Activities

Classroom Observations

Sample. The goal was to observe four teachers of biology at each pilot school for one class period, with approximately even numbers of on-level and honors-level classes at each school. If there were more than four biology teachers at a school, the first priority was teachers who attended training about the pilot unit, and the second priority was teachers who taught the most sections of biology. Observers did not visit classes restricted to English language learners or to students with disabilities. At two schools, there were fewer than four observations, either because teachers were out on maternity leave, teachers skipped some of the lessons and finished the unit earlier than scheduled, or the school had fewer than four biology teachers. A total of 61 observations were completed.

Instrument. To answer evaluation question 1 on fidelity of teacher implementation of PBL practices, the evaluators worked with the program staff to develop a protocol for classroom observations. The observers included four OSA staff members who piloted the observation protocol in two classrooms at two separate schools. (Data from the pilot observations were not used in the report.) Based on the pilot observations, the protocol was refined prior to final data collection to insure valid reporting of the practices and reliable reporting across observers.

Each indicator had multiple options for observers to document the extent of evidence (see detail in Table 1). Options for extent of evidence varied by indicator and were defined as follows. *Evident* meant that the teacher implemented the practice as expected. *Evident with emphasis* meant that the teacher implemented the practice to a degree that was distinctly better than the expected level, as defined by program staff. The final observation protocol included seven PBL practices, as listed and defined in Table 1. Note that students were presented with the unit project in the form of an RFP.

Table 1
Problem-Based Learing Practices, Definitions, and Options for Extent of Evidence Used for Classroom Observations

PBL practice	Definitions	Options for extent of evidence
Central problem (i.e., unit project/RFP)	Is there evidence of a connection to the unit project/RFP or a component of the RFP (i.e., poster, management plan, town hall presentation)?	Not evident, evident, evident with emphasis
Collaboration	Does the teacher give student teams or groups opportunities to collaborate (i.e., work together toward a common goal)?	Not evident, evident
Constructive investigation: Multiple solutions	Is there an indication that multiple solutions to a problem or for the project are possible? (Not the same as multiple results.)	Not evident, evident
Inquiry	Does the teacher give students opportunities to justify or defend their thinking or reasoning?	Not evident, evident, evident with emphasis
Student choice	Does the teacher give students opportunities to make meaningful choices about the unit project/RFP?	Not evident, evident, not applicable
Student centered learning	Does the teacher function as a facilitator (e.g., keeps students on task without giving the answer, not leading or directive)?	Not evident, evident, evident with emphasis
Students working in small groups or pairs	What is the length of time during which students are supposed to work in small groups or pairs during class?	Evident: $\geq$ ⅓ class time Evident with emphasis: $\geq$ ½ class time

To provide descriptive data on student engagement in classes at pilot schools, observers collected data on student engagement when students were working in groups. Focusing on one group at a time over the course of one minute, the observer recorded the minimum number of members who displayed each of the following forms of engagement:

- On task—As individuals, students do the assigned work and do not engage in off-task behavior (e.g., get out of seat without permission, have an off-topic conversation, gaze out window).
- Actively contribute/participate with the group, nonverbally (e.g., make an effort toward the task at hand, conduct task, manipulate, create).
- Actively contribute/participate with the group, verbally (e.g., discourse, talk, ask questions, discuss; must relate to group's task or topic).

To answer evaluation question 2 on fidelity of teacher implementation of the lesson plans, observers noted additions, changes, and deletions to the components and activities in the lesson plan for that class. The observer summarized the extent to which the delivered instruction matched the lesson plan on the following scale: exactly, slight variations, some variations, lot of variations, or completely different. Matching meant the teacher closely followed the spirit of the lesson plan; it did not have to be verbatim or exactly match the recommended time for each step of the lesson. If the teacher started with part of a lesson from the previous day, or if the teacher did not finish the entire lesson, or if the teacher added components, these differences were not considered. But observers did monitor the sequence of components and activities and considered the activity's length when determining how much it contributed to the total amount of variation.

Observers contacted each teacher after the visit with two questions about the observed class: 1) Did you make any changes to the lesson plan as written in the curriculum guide? Please describe those changes. 2) Did the lesson today go as you planned? Please explain your answer. More than 80% of the observed teachers replied to at least one of the questions (52 of 61, 85%). The evaluators analyzed the responses to both questions to identify changes, problems, or suggestions for the lessons and categorized comments on similar topics.

Student Surveys

Instrument. A paper and pencil survey was used to collect data from students. The survey included items about the extent to which teachers implemented PBL practices in that classroom; responses were used to verify that there was a higher level of PBL implementation in the pilot schools than in the nonpilot schools. Items from the observation protocol were used to develop items for this section of the student survey.

The survey also included items on students' level of engagement with the unit. All items referred to the invasive species unit in the pilot schools or the ecology unit in the nonpilot schools. The survey covered four types of engagement:

- Behavioral—Actively involved in certain skills and processes, participating or actively contributing to activities when working in group
- Cognitive—Committed to learning the content and to understanding the content
- Emotional—Enthusiasm, excitement, and enjoyment of class
- Psychological—Connection to the team and to the solution for the unit project

For emotional engagement, the authors modified items from a previous survey used with middle school students (Cooper-Martin & Wolanin, 2014). Some items on other types of engagement were based on previous surveys (Fredricks et al., 2011). The full survey was pilot-tested with a small sample of students; items were revised, as necessary, to improve validity and reliability.

A scale of multiple items, rather than a single item, was used to measure each type of engagement, for better reliability. The range for each scale was 1–4. The internal consistency or reliability of each scale was examined using Cronbach's alpha (Nunnaly, 1978). The proposed scale for behavioral engagement was not used, because the Cronbach's alpha for this scale did not reach a satisfactory level of .70. The items for behavioral engagement were combined with the cognitive or psychological scale (Table 2). The internal reliability was relatively high for the three remaining scales, because Cronbach's alpha was at least .80 for each one (Table 2).

Table 2
Items, Response Options, and Cronbach's Alpha for Three Scales of Student Engagement

Scale and items	Response options (value for scale)	Cronbach's alpha
Cognitive	Very true (4)	.80
• During this unit, I worked as hard as I could.[a]	Sort of true (3)	
• During this unit, I actively participated in class most of the time.[a]	Not very true (2)	
• I was very committed to learning the topics for the [] unit.	Not at all true (1)	
• I worked hard to understand the material covered for this unit.		
• How much effort did you put into learning the material for the [] unit?[b]		
Emotional	Very true (4)	.90
• When I was in class during the [] unit, I felt good.	Sort of true (3)	
• When we worked on something in the [] unit, I felt interested.	Not very true (2)	
• I enjoyed the work I did in this unit.	Not at all true (1)	
• I felt excited by the work during the [] unit.		
• The [] unit was engaging.		
• I enjoyed learning new things during this unit.		
Psychological	Very true (4)	.85
• I was a frequent contributor to my group.[a]	Sort of true (3)	
• When I was with my group, I was an active participant.[a]	Not very true (2)	
• When working in my group, I felt like a part of a team.	Not at all true (1)	
• I felt valued as a contributor, when working together with my group.		
• I was committed to my group reaching the goals for our assignment(s).		
• My group's success was important to me.		

Note. [] was "invasive species" in the pilot schools and was "ecology" in the nonpilot schools.
[a] Item from behavioral engagement scale.
[b] Response options for this item differed from the other items; see text for details.

The response option for the last item on cognitive engagement was a five-point scale as follows:

○ 5	○ 4	○ 3	○ 2	○ 1
Extremely high (Probably as much effort as I've ever put into a class)				Extremely low (Probably the least amount of effort I've ever put into a class)

Because this item on cognitive engagement was combined with other items that had a four-point scale, the scale value was set at 4 for extremely high, at 3.25 if the student checked 4, at 2.5 if the student checked 3, at 1.75 if the student checked 2, and at 1 for extremely low. Thus the range of values for the cognitive engagement scale, like the others, was 1–4.

Sample. At each pilot school, the sample of students invited to complete the survey was limited to three classrooms; preference was given to teachers who, based on the classroom observations, had implemented more PBL practices and more closely followed the lesson plans. Because the level of student engagement was expected to be higher only if the teacher implemented PBL with fidelity, this approach was an appropriate way to limit the student sample. To provide a comparison group, the sample also included four classrooms from each nonpilot school. In the majority of schools, the four classrooms consisted of two on-level and two honors-level classes. Surveys were offered within five days of the completion of the invasive species unit at pilot schools and within five days of the completion of the ecology unit at nonpilot schools. A total of 2,030 surveys were completed, from 1,190 respondents in all 16 pilot schools (95% response rate) and 840 respondents in all 9 nonpilot schools (97% response rate).

Survey administration. To ensure confidentiality and avoid burdening school staff, an OSA staff member administered and collected the student survey in each classroom. The survey asked for the student's ID to allow the evaluators to match data from other sources. Directions to students promised that their responses were confidential, would be reported only at the aggregate level, and would not be shared with staff at their school or with their parents.

Unit Tests and Student Data

To answer evaluation question 4 on the impact of the curriculum on student achievement in biology, students in both pilot and nonpilot schools completed an end-of-unit test on ecology, which is the content area for the invasive species unit. All unit tests included 10 common items (9 multiple choice plus 1 brief constructed response) to allow for comparisons between pilot and nonpilot schools. Based on the same scoring rubric used by teachers, each of the nine multiple choice questions was worth .5 and the brief constructed response was worth 4 points, for a maximum score of 8.5 points on the common items. Scores used in the analysis equaled the number of points attained divided by 8.5.

Two versions of the unit test were available: 1) the pilot unit test contained only the 10 common items; and 2) the nonpilot unit test contained 21 items (the 10 common items plus 11 additional ones on ecology). Some or all of the unit tests from four pilot schools were the nonpilot version; all unit tests from one nonpilot school were the pilot version.

Although all 10,064 students enrolled in biology during the fall semester of the 2014–2015 school year were expected to complete a unit test, scores were available for only 6,855 students. However, the sample used for analysis was smaller, for the following reasons.

Tests from less than 1% of the biology students were available for two schools (one pilot and one nonpilot); therefore, the results for these two schools were deleted. Because the unit was implemented in semester 1, unit tests taken in semester 2 were deleted from the analytical sample, as there was too much time between the end of classes and the unit test. These deletions included all tests from one nonpilot school. Lastly, with respect to the nonpilot school that submitted the pilot version of the unit test, anecdotal data revealed that at least some of the teachers at this school told students that this unit test "did not count" towards their course grade. To prevent bias, all tests from this nonpilot school were deleted. The final analytical sample totaled 6,065 tests (4,745 tests from 15 pilot schools and 1,320 from 6 nonpilot schools). This was a completion rate of 60% for all biology students, 68% for biology students at pilot schools, and 43% for biology students at nonpilot schools.

Student characteristics for all students who completed a unit test or the student survey were downloaded from MCPS databases and used for analysis, as described below.

Analytical Procedures

Descriptive statistics were used to analyze the findings for all questions. As appropriate, chi-square (χ^2) tests were used to determine whether differences in percentages (e.g., of students who agree on a survey item) between two groups were statistically significant. For each χ^2 test, the degrees of freedom, which equals the number of groups minus one, is reported along with the number of students, represented as *N*. For example, in an analysis of differences between on-level and honors classes, the number of groups would be two and so the degrees of freedom equal one (i.e., 2–1), as in the following: $\chi^2(1, N = 3{,}351) = 6.98$. In analyses for all questions, independent sample t-tests were used to determine whether differences in mean values for specific measures between two groups were statistically significant. For each t-test, the degrees of freedom, which typically equal the number of students minus two, are reported.

Also, for evaluation question 4, regression analysis was used to examine the relationship between the curriculum used and student achievement, while controlling for differences in student characteristics. The dependent variable was scores on the 10 items that every student completed on a unit test. Student characteristics included as independent variables in the analysis were grade level, gender, race/ethnicity, receipt of Free and Reduced-price Meals System (FARMS), receipt of special education services, limited English proficiency, and scores on the Grade 8 Maryland School Assessment (MSA) in science. As described above, there were about 3,000 more tests from pilot schools than from nonpilot schools. Therefore, a second regression analysis was conducted using a random sample of 2,000 (in total) students from the pilot schools, along with all students in the analytical sample from nonpilot schools.

Additional analyses examined whether the results for evaluation question 4 were the same for student subgroups by conducting analyses with only the students in the subgroups of interest; these subgroups were Black or African American, Hispanic/Latino, and Two or More Races; FARMS recipients; limited English proficiency; special education services recipients.

For all analyses, tests of statistical significance were calculated to judge whether the observed relationship between student outcomes (i.e., level of student engagement, scores on 10 common items) and the curriculum used occurred by chance. Also, tests of practical significance were calculated to judge whether statistically significant relationships were large enough to be useful to program staff (American Psychological Association, 2001). Effect sizes were used as tests of practical significance. For t-tests, Cohen's *d* was calculated as an effect size measure; Cohen's *d* was calculated as the difference in means divided by the pooled standard deviation. For regression analysis, standardized regression coefficients (β values) were used as an effect size measure (Kline, 2005).

As argued by Cooper-Martin and McNary (2007), previous research provides a context for interpreting the magnitude of effect sizes that is more useful than the conventional benchmarks suggested by Cohen (1988). Lipsey et al. (2012) provide average effect sizes from a collection of randomized studies on various types of educational interventions; in their studies of curriculum programs, the median effect size was .08 and the mean effect size was .13. Therefore, this report set .10 as the minimum value for a meaningful effect size.

Strengths and Limitations of Methodology

Strengths

The observation protocol was developed by working closely with STE program staff who were the curriculum writers for the invasive species unit and experts in the MCPS biology curriculum as a whole. The observation protocol was piloted and revised to improve construct validity and reliability of the evaluation findings.

Classroom observations included all 16 pilot schools. Additionally, all 25 MCPS high schools, including the 16 pilot schools, were included in the student survey. These samples included a mix of MCPS classes (i.e., variation of class period, class level, and teacher) and a variety of student demographics, thus enhancing the external validity of the study.

The three engagement scales used in the student survey had a relatively high level of internal reliability (Cronbach's alpha > .80 for each scale). Also, surveys were administered within five school days after completion of the unit, whether pilot or nonpilot, which helped ensure that the unit was still in recent memory for the students taking the survey and thereby decreasing the measurement error. To ensure the validity of the findings, the analytical sample excluded tests taken in the second semester and thus more than five school days after completion of the unit.

Finally, there were 10 common items among the two versions of unit tests, which allowed for comparisons between pilot and nonpilot schools.

Limitations

A limitation to this study is that the data came from the first year of implementation for the invasive species unit and of a PBL approach to a high school biology unit. As with any new lesson or curriculum, full implementation during the first year is unlikely.

The unit test was another limitation; there were two versions and almost all pilot schools used the pilot version and almost all nonpilot schools used the nonpilot one. Thus, test version was very highly correlated with school group and so was not appropriate for use in regression analysis. Therefore, the analysis could not determine whether differences in student achievement between pilot and nonpilot schools were due to differences in the test version or differences in the curriculum. Furthermore, at one school, teachers announced that the unit test would not count towards course grades; therefore, we eliminated this school from the analysis. If teachers at other schools made similar announcements, this could have affected performance on the tests and the study's findings about this performance.

The final sample for the unit test included 21 of the 25 high schools and 60% of students enrolled in biology. Although there was a sufficient sample size for analyses, the results may not generalize to all MCPS students enrolled in biology because the sample size was less than 100% and not from a random sample.

Results

Findings for Question 1: To what extent do teachers in pilot schools implement PBL practices with fidelity?

Data from 61 classroom observations in pilot schools were used to evaluate implementation of PBL practices. Due to resource constraints, there were no observations in nonpilot schools.

Level of Implementation

The observation protocol included seven PBL instructional practices. Two PBL practices, central problem (i.e., the teacher connects the lesson to the unit project) and student choice (i.e., the teacher gives students opportunities to make meaningful choices about the unit project) were critical. Evidence of either of these practices indicated adequate implementation of PBL. Also, if a teacher used at least three PBL practices and one of them was either constructive investigation (i.e., an indication that multiple solutions to a problem are possible) or student-centered learning (i.e., teacher functions as a facilitator), implementation of PBL was considered adequate. Each practice had to be at the evident or evident with emphasis level. (Evident meant that the teacher implemented the practice as expected. Evident with emphasis meant that the teacher implemented the practice to a degree that was distinctly better than the expected level, as defined by program staff.)

Based on these criteria, teachers in more than three quarters of the observed classes (79%) had adequate implementation of PBL (Table 3). In this study, implementation by more than two thirds of the sample indicated a high level of implementation, implementation by one third or more of the sample but less than two thirds indicated a moderate level, and implementation by less than one third of the sample indicated a low level of implementation. Therefore, this finding indicates an overall high level of implementation of PBL practices in pilot schools.

Table 3
Adequate Implementation of Problem-Based Learning Instructional Practices in Pilot Schools for All Classes and by Course Level

	All (*N* = 61)		On-level (*N* = 28)		Honors[a] (*N* = 33)	
Adequate implementation	*n*	%	*n*	%	*n*	%
Yes[b]	48	78.7	21	75.0	27	81.8
No	13	21.3	7	25.0	6	18.2

[a] Includes pre-International Baccalaureate biology and mixed (honors and on-level) classes.
[b] All practices were at the evident or evident with emphasis level.

Although slightly fewer on-level classes (75%) than honors classes (82%) had adequate implementation of PBL practices, implementation was at a high level for both course levels.

Implementation of Specific Practices

All classes. Table 4a shows the extent of implementation for each PBL practice across all 61 observed classes. (Full definitions for each practice are in Table 1 above.) The most frequent practice was collaboration; in almost all classes (85%), the teacher gave the students an opportunity to collaborate (i.e., work together toward a common goal).

Table 4a
Implementation of Specific Problem-Based Learning Practices in Pilot Schools ($N = 61$)

PBL practice	n	%
Collaboration: Working together toward a common goal		
Evident	52	85.2
Students in small groups or pairs: Time		
Evident or evident with emphasis	39	63.9
Evident with emphasis	21	34.4
Student centered learning: Teacher functions as facilitator [a]		
Evident or evident with emphasis	38	63.3
Evident with emphasis	20	33.3
Inquiry: Opportunities to justify or defend thinking		
Evident or evident with emphasis	37	60.6
Evident with emphasis	25	41.0
Central problem: Connection of content to unit project		
Evident or evident with emphasis	28	45.9
Evident with emphasis	12	19.7
Student choice: Opportunities to make choices about unit project [b]		
Evident	13	36.1
Constructive investigation: Indication that multiple solutions are possible		
Evident	20	32.8

[a] $N = 60$; data not collected for one class.
[b] $N = 36$; student choice applicable only for observations during which students worked on unit project.

Three PBL practices occurred in at least 60% of observed classes. Students worked in small groups (or pairs) for at least one third of the class time (64%), the teacher encouraged student centered learning by acting as a facilitator at least once (63%), and the teacher encouraged inquiry by giving students an opportunity to justify or defend their thinking or reasoning at least once (61%). These three practices were less frequent at the evident with emphasis level. In about one third of classes, students worked in small groups for at least one half of class time (34%) and the teacher acted as a facilitator more than one time (33%). The teacher encouraged inquiry more than one time in 41% of classes.

Finally, three PBL practices were observed in at least one third of the classes. Two of these practices were deemed the most critical of all the practices in a PBL classroom: central problem and student choice. Central problem (i.e., connection to the unit project/RFP or a component of the RFP) was evident (i.e., the RFP was posted in the room or the teacher mentioned the unit project/RFP at least briefly) in close to one half of the classes (46%). However, a connection to the central problem was evident with emphasis (i.e., the teacher invited students to apply knowledge and skills to the unit project/RFP or encouraged students to evaluate or refine their solutions to the unit project/RFP) in only one fifth of the observed classes (20%). Student choice (i.e., the teacher gave students an opportunity to make meaningful choices about the unit project/RFP) occurred in just over one third of classes that included the unit project (36%). Lastly, there was evidence of constructive investigation (i.e., the teacher indicated that multiple solutions to a problem or for the project are possible) in one third of observed classes (33%).

On-level vs. honors classes. Table 4b shows the extent of implementation of PBL practices for on-level and honors classes. Most practices, including collaboration, students in small groups, student centered learning, inquiry, and central problem, were more frequently observed in honors classes. The biggest difference between on-level and honors classes was for inquiry; this practice was evident in one half (50%) of on-level classes but in 7 out of 10 (70%) honors classes. Both student choice and constructive investigation were more frequently implemented in on-level classes. However, none of the differences between on-level and honors classes was statistically significant, based on χ^2 tests.

Table 4b
Frequency of Implementation of Specific Problem-Based Learning Practices by Course Level in Pilot Schools

	Practice evident or evident with emphasis			
	On-level (*N*=28)		Honors[a] (*N*=33)	
PBL practice	*n*	%	*n*	%
Collaboration: Working together toward a common goal	23	82.1	29	87.9
Students in small groups or pairs	16	57.1	23	69.7
Student centered learning: Teacher functions as facilitator[b]	17	60.7	21	65.6
Inquiry: Opportunities to justify or defend thinking	14	50.0	23	69.7
Central problem: Connection of content to unit project	12	42.9	16	48.6
Student choice: Opportunities to make choices about unit project[c]	7	43.8	6	30.0
Constructive investigation: Indication that multiple solutions are possible	10	35.7	10	30.3

[a] Includes pre-International Baccalaureate biology and mixed (honors and on-level) classes.
[b] Data not collected for one honors class.
[c] Student choice applicable only for 16 on-level and 20 honors classes in which students worked on unit project.

Findings for Question 2: To what extent do teachers in pilot schools implement with fidelity the lesson components and activities as presented in the lesson plans for the invasive species unit?

Data from 61 classroom observations in pilot schools were used to evaluate teachers' fidelity of implementation of the lesson plans for the invasive species unit. Due to resource constraints, no observations were conducted in nonpilot schools.

Ratings

For each observation, observers rated the extent to which the observed lesson matched the lesson components and activities as described in the lesson plan using one of the following categories: exactly, slight variations, some variations, a lot of variations. Table 5 presents the results on fidelity of implementation of the lesson for all 61 classes and separately for on-level and honors classes.

Table 5
Fidelity of Implementation to Lesson Components and Activities
in the Lesson Plans Among Classes in Pilot Schools and by Course Level

	All (N = 61)		On-level (N = 28)		Honors[a] (N = 33)	
Extent to which class followed the lesson plan	*n*	%	*n*	%	*n*	%
Four categories of variations						
No variations	11	18.0	6	21.4	5	15.2
Slight variations	24	39.3	8	28.6	16	48.5
Some variations	15	24.6	9	32.1	6	18.2
A lot of variations	11	18.0	5	17.9	6	18.2
Two categories of variations						
No variations or Slight variations	35	57.4	14	50.0	21	63.6
Some variations or a lot of variations	26	42.6	14	50.0	12	36.4

[a] Includes pre-International Baccalaureate biology and mixed (honors and on-level) classes.

About one fifth of all classes (18%) exactly matched the lesson plans; a similar number (18%) had a lot of variations from the lesson plans. There were slight variations from the lesson plan for 4 out of 10 classes (39%) and some variations for one quarter of the classes (25%). Across all 61 observed classes, the fidelity of implementation to the lesson plan was exact or with slight variations for more than one half of the classes (57%). In this study, implementation by more than two thirds of the sample indicated a high level of implementation, implementation by one third or more of the sample but less than two thirds indicated a moderate level, and implementation by less than one third of the sample indicated a low level of implementation. Therefore, this finding indicates an overall moderate level for fidelity of implementation of the lesson plans for the invasive species unit. This moderate level of implementation is reasonable given that the evaluation occurred during the first year of implementation.

In on-level and honors classes, as for all classes, about one fifth of each group exactly matched the curriculum (21% for on-level and 15% for honors) and about one fifth had a lot of variations (18% for both on-level and honors) (Table 5). While less than one third of on-level classes (29%) showed slight variations from the lesson plans, one half of honors classes (49%) were in this category. Lastly, one third of on-level classes (32%) but only about one fifth of honors classes (18%) had some variations from the questions 1 and 2 on classroom implementation of PBL practices and of the lesson plans. The fidelity of implementation to the lesson plans was exact or with slight variations for one half of on-level classes (50%) and for almost two thirds of honors classes (64%). However, this difference in percentages between on-level and honors classes was not statistically significant, based on a χ^2 test.

Descriptions of Variations

In the 24 classes rated as slight variations, the teachers followed the lesson plans very closely. The most common variations related to groupings (e.g., independently vs. with a partner) or to a video (e.g., did not show all of it). Very few of these classes (4 of 24) had steps out of order. Among the 15 classes rated as having some variations, almost all (14) had at least three of the following: change, deletion, out of order. In more than one half of the classes with some variations, the teacher deleted at least one step (13), did a task themselves that the students were supposed to do (9), or varied the student groupings (8). In almost all of the 11 classes with a lot of variations, teachers deleted multiple steps or did multiple steps themselves that students were supposed to do. In 5 of these 11 classes, steps were out of order. All three observations of the yeast lab class for Day 5 had a lot of variation.

Teacher Comments about Variations

After each observation, teachers received a follow-up e-mail with questions about the observed lesson. Of the 61 observed teachers, 51 (84%) replied to at least one of these questions. The evaluators analyzed the responses to both questions to identify changes, problems, or suggestions about the lessons and categorized comments on similar topics. Table 6 shows the categories with frequencies and examples. The most frequent comments related to not enough time for the lesson (27, 53%) and a desire to change some of the lesson implementation (21, 41%). About one third of respondents said the lesson needed more detail, instructions, clarification, or examples (15, 29%). At least one tenth of teachers mentioned the following: groups arranged differently than what was in the lesson plan (6, 12%), technical problems with video links (6, 12%), vocabulary created or needed to accompany lesson (6, 12%), students worked individually instead of in groups (5, 10%), and worksheets were reformatted (5, 10%). Finally, four teachers wrote about other student needs.

Table 6
Teacher Comments About Changes and Problems for Invasive Species Lessons (*N* = 51)

Category	*n*	%	Examples
Not enough time for lesson	27	53	• Not enough time for lab/discussion/review • Tough to get through entire lesson in 45 minutes • Set up mixture/lab before class to save time
Changed, or would like to change, some of the lesson implementation	21	41	• Continued to collect data after period to get better data • Did not make cards because we've done so many • Each group had a different question to read • Changed poster size and type
Need more detail, instructions, clarification, or examples	15	29	• The brown snake video not informative enough • Directions needed to be simplified/clarified • Added examples
Groups arranged differently	6	12	• Did town hall all together (vs. two groups) • Used table arrangement group (vs. project group)
Technical problems with video links	6	12	• Trouble getting link to work • Audio didn't work the first time
Vocabulary created or needed	6	12	• A vocabulary list has been prepared to go with each lesson • Students have not been exposed to lots of words in article
Students worked individually, not in groups	5	10	• Answered questions individually to save time • Read parts individually instead of in group
Worksheets reformatted	5	10	• Took out extension • Added more writing space • Made two documents from graph and BCR [brief constructed response]
Other student needs	4	8	• Students have not bought into team concept • Graphing difficult for some students • Students asked about slow pace of unit

Findings for Question 3: Does students' level of engagement with the ecology unit depend on which curriculum (i.e., invasive species vs. traditional) the teacher used? Is the result the same for each student subgroup, such as racial/ethnic groups?

One of the expected student outcomes from a PBL curriculum is a higher level of engagement. Analyses of student surveys from both pilot and nonpilot schools were conducted to test whether students at pilot schools were more engaged with the ecology unit than their peers at nonpilot schools.

Problem-Based Learning in Pilot vs. Nonpilot Schools

Prior to analyzing the relationship between student engagement and the curriculum used, it was important to confirm that PBL practices were more frequent at the pilot schools. The student survey included items about problem-based learning. Compared to students at nonpilot schools, more students at pilot schools strongly agreed or agreed with items 1, 2, 4, and 5 (Table 7a). These differences were statistically significant: item 1, $\chi^2(1, N = 2{,}027) = 120.4, p < .001$; item 2, $\chi^2(1, N = 2{,}024) = 16.4, p < .001$; item 4, $\chi^2(1, N = 2{,}017) = 46.7, p < .001$; and item 5, $\chi^2(1, N = 1{,}538) = 8.3, p < .01$.

Table 7a
Student Responses to All Survey Items about Problem-Based Learning, by Pilot and Nonpilot Schools

Item	School group	N	% Strongly agree or agree	χ^2
1. In the [] unit, we often worked together in groups toward a common goal.	Pilot	1,189	95.3	
	Nonpilot	838	79.7	120.4***
2. The teacher frequently asked students to defend their thinking, during this unit.	Pilot	1,186	85.8	
	Nonpilot	838	78.9	16.4***
3. During the [] unit, the teacher usually did not give students the answer, but gave hints to find the answer.	Pilot	1,188	86.2	
	Nonpilot	839	85.6	0.16
4. In this unit, I learned that often there are multiple solutions to a problem.	Pilot	1,185	87.3	
	Nonpilot	832	75.5	46.7***
5. The topics I studied in the [] unit were to help me complete my unit project.	Pilot	1,175	90.0	
	Nonpilot [a]	363	84.6	8.3**
6. I got to make choices about my [] unit project.	Pilot	1,187	78.8	
	Nonpilot [a]	362	76.5	0.83

Note. [] was "invasive species" in the pilot schools and was "ecology" in the nonpilot schools.
[a] Excludes students who did not have a unit project.
**p* < .05, ***p* < .01, ****p* < .001.

Although all students at pilot schools had unit projects because of the PBL curriculum, most students in nonpilot schools did not. The above analyses of items 5 and 6, which both concern the unit project, excluded those students at nonpilot schools who did not have such a project. In a second analysis of items 5 and 6, students at nonpilot schools who did not have a unit project were included with the group of students who chose disagree or strongly disagree (Table 7b). In this analysis, differences between school groups were significant, in favor of pilot schools, for both item 5, $\chi^2(1, N = 2{,}012) = 638.0, p < .001$ and item 6, $\chi^2(1, N = 2{,}023) = 425.3, p < .001$.

Table 7b
Student Responses to Two Survey Items about Problem-Based Learning, by Pilot and Nonpilot Schools

Item	School group	N	% Strongly agree or Agree	χ^2
5. The topics I studied in the [] unit were to help	Pilot	837	90.0	
me complete my unit project.*	Nonpilot [a]	1,175	36.7	638.0***
6. I got to make choices about my [] unit project.*	Pilot	1,187	78.8	
	Nonpilot [a]	836	33.1	425.3***

Note. [] was "invasive species" in the pilot schools and was "ecology" in the nonpilot schools.
[a] Includes students who did not have a unit project with those who chose disagree or strongly disagree.
*$p < .05$, **$p < .01$, ***$p < .001$.

Thus, the student responses confirmed that PBL practices were more frequent at the pilot schools than the nonpilot schools, as expected based on the invasive species curriculum. For further analysis, the six items about PBL were combined into a scale.

In an analysis by school, two nonpilot schools had higher mean values on the PBL scale than the other nonpilot schools; their levels (3.04, 3.45) were closer to the mean values at pilot schools, which ranged from 3.05 to 3.48. Therefore, the following analyses excluded these 2 schools and are based on the remaining 7 nonpilot high schools, along with the 16 pilot schools.

All Students

The analysis tested whether the mean level for each engagement scale differed between the curricula used, as represented by pilot vs. nonpilot schools. As expected, the mean level of cognitive engagement was significantly higher at the pilot schools with the PBL curriculum ($t(1,857) = 2.39$, $p < .05$) (Table 8). The effect size (Cohen's $d = 0.11$) is large enough that this difference is useful in an educational setting. However, mean levels of emotional and psychological engagement did not differ significantly between the two groups of schools.

Table 8
Mean Level and Standard Deviation of Student Engagement Scales
by Pilot and Selected Nonpilot Schools

Type of engagement (range)	School group	N	Mean	Standard deviation	t	df
	Pilot	1,189	3.10	0.56		
Cognitive (1.0-4.0)	Nonpilot	670	3.04	0.58	2.39*	1,857
	Pilot	1,188	2.80	0.69		
Emotional (1.0-4.0)	Nonpilot	669	2.85	0.69	-1.29	1,855
	Pilot	1,189	3.41	0.55		
Psychological (1.0-4.0)	Nonpilot	594	3.39	0.58	0.83	1,781

*$p < .05$, **$p < .01$, ***$p < .001$.

Additional analysis indicated that the level of emotional engagement varied with whether students were in on-level or honors classes, along with whether students were in pilot or nonpilot schools (Figure 2 and Table 9). In on-level classes, the level of emotional engagement was higher at pilot schools than at nonpilot schools, as expected. This difference was statistically significant ($t(727) = 1.95$, $p \leq .05$) and practically significant (Cohen's $d = .15$), meaning that the difference was large enough to be useful to program staff. In honors classes, the relationship was reversed: the level of emotional engagement was lower at pilot schools than at nonpilot schools. This difference was statistically significant ($t(1,125) = -3.27$, $p \leq .001$) and practically significant (Cohen's $d = -.20$), meaning that the difference was large enough to be useful to program staff.

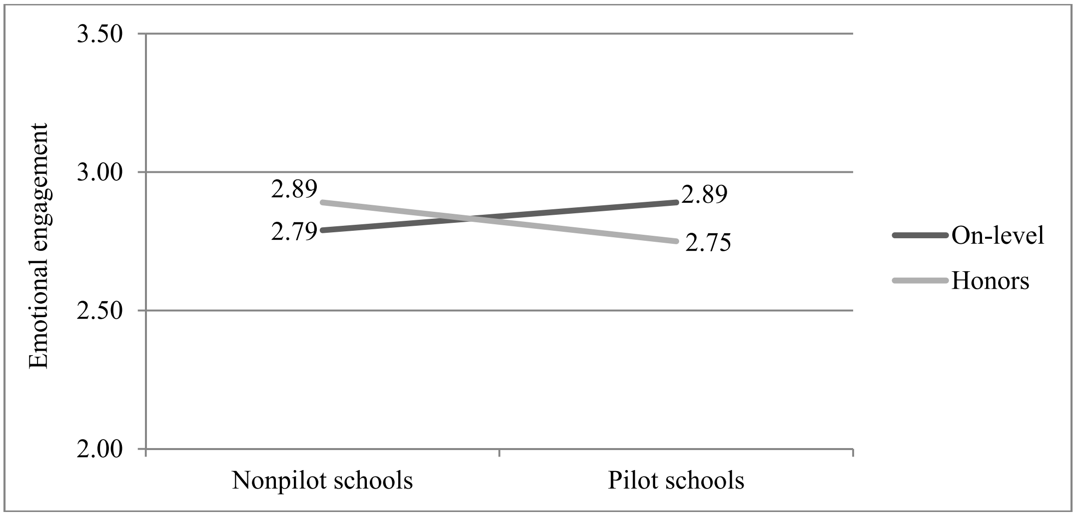

Figure 2. Mean level of emotional engagement by school group and course level.

Table 9
Mean Level and Standard Deviation of Student Emotional Engagement Scale by Pilot and Selected Nonpilot Schools and by Course Level

Course level	Pilot schools *N*	Pilot schools Mean	Pilot schools Standard deviation	Nonpilot schools *N*	Nonpilot schools Mean	Nonpilot schools Standard deviation	*t*	df
On-level	460	2.89	0.70	269	2.79	0.70	1.95*	727
Honors	727	2.75	0.67	400	2.89	0.68	-3.27***	1,125

*$p \leq .05$, **$p < .01$, ***$p \leq .001$.

In summary, there was some evidence that students' level of engagement with the ecology unit depended on which curriculum the teacher used. As expected, students at pilot schools had higher levels of cognitive engagement than those at nonpilot schools. The level of emotional engagement was higher at pilot than at nonpilot schools for those in on-level biology, but was lower at pilot than at nonpilot schools for students in honors biology.

Student Subgroups

To determine whether the results for the relationships between student engagement and the curriculum for all students were the same for student subgroups, additional analyses were conducted using only students in the subgroups of interest.

Cognitive. For all students, the level of cognitive engagement was significantly higher in pilot schools than in nonpilot ones. Likewise, for most subgroups, the level of cognitive engagement was higher in pilot schools (Table 10), but the difference was not significant for every subgroup, as follows. For students who were Black or African American, Hispanic/Latino, or identified as Two or More Races, the level of cognitive engagement was higher at pilot schools. This difference was statistically significant ($t(993) = 1.94$, $p \leq .05$) and large enough to be useful in an educational setting (Cohen's $d = .33$).

Table 10
Mean Level and Standard Deviation of the Cognitive Engagement Scale for Student Subgroups, by Pilot and Selected Nonpilot Schools

Student subgroup	School group	*N*	Mean	Standard deviation	*t*	df
Black or African American,	Pilot	671	3.11	0.55		
Hispanic/Latino, and Two or More Races	Nonpilot	324	3.04	0.56	1.94*	993
	Pilot	343	3.13	0.53		
FARMS recipients	Nonpilot	173	3.09	0.55	0.74	514
	Pilot	100	3.03	0.57		
Limited English proficiency	Nonpilot	36	3.04	0.54	-0.13	134
	Pilot	103	3.18	0.49		
Special education services recipients	Nonpilot	59	3.01	0.58	2.02*	160

$^*p \leq .05$, $^{**}p < .01$, $^{***}p < .001$.

Although the mean level of cognitive engagement for FARMS recipients was higher at pilot schools than at nonpilot schools, this difference was not statistically significant ($t(514) = 0.74$, $p >.05$). For students with limited English proficiency, the level of cognitive engagement did not differ significantly between pilot and nonpilot schools ($t(134) = -0.13$, $p >.05$). For students who received special education services, the level of cognitive engagement was higher at pilot schools than at nonpilot ones; this difference was statistically significant ($t(160) = 2.02$, $p < .05$) and large enough to be useful in an educational setting (Cohen's $d = .13$).

Emotional. When analyzing all students, the mean level of emotional engagement was higher at pilot schools than at nonpilot schools for those in on-level biology, but was lower at pilot than at nonpilot schools for students in honors biology. This pattern was the same for most but not all subgroups (Table 11). For students who were Black or African American, Hispanic/Latino, or identified as Two or More Races, the pattern of differences for emotional engagement matched that for all students. The difference in emotional engagement for on-level courses was statistically significant ($t(429) = 2.12$, $p <.05$) and practically significant with an effect size greater than .10 (Cohen's $d = .22$). However, the difference for honors courses was not significant ($t(433) = -0.93$, $p > .05$). For FARMS recipients, the pattern of differences matched that for all students, but the differences were not significant for on-level ($t(303) = 0.50$, $p > .05$) or honors courses ($t(209) = -1.31$, $p > .05$).

Table 11
Mean Level and Standard Deviation of the Emotional Engagement Scale for Student Subgroups, by Pilot and Selected Nonpilot Schools and by Course Level

		Pilot schools			Nonpilot schools				
Student subgroup	Course level	*N*	Mean	Standard deviation	*N*	Mean	Standard deviation	*t*	df
Black or African American,	On-level	291	2.90	0.58	140	2.75	0.71	2.12*	429
Hispanic/Latino, and Two or More Races	Honors	300	2.76	0.70	135	2.83	0.66	-0.93	433
	On-level	204	2.86	0.72	101	2.82	0.67	0.50	303
FARMS recipients	Honors	139	2.81	0.70	72	2.93	0.64	-1.31	209
	On-level	64	2.94	0.73	23	2.75	0.69	1.12	85
Limited English proficiency	Honors	36	2.85	0.77	13	3.20	0.38	-2.13*	42
	On-level	92	2.91	0.66	59	2.70	0.78	1.86	149
Special education services recipients	Honors	22	2.86	0.79	9	2.76	0.81	0.30	29

$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

For students with limited English proficiency, the mean level of emotional engagement showed the same pattern of differences as those for all students. However, the difference in emotional engagement for on-level courses was not significant ($t(85) = 1.12$, $p > .05$). The difference for honor courses was statistically significant ($t(42) = -2.13$, $p < .05$) and large enough to be useful in an education setting, with an effect size greater than .10 (Cohen's $d = .53$).

For students who received special education services, the findings differed from those for all students. For these students, the level of emotional engagement was higher at pilot schools than at nonpilot ones for both on-level and honors courses. Additional analysis for students who received special education combined students in honors and on-level courses to compare the level of emotional engagement between pilot and nonpilot schools. This analysis revealed that, for this subgroup, the difference in the level of emotional engagement between pilot and nonpilot schools was statistically significant ($t(160) = 2.03$, $p < .05$) and large enough to be useful in an educational setting (Cohen's $d = .33$).

Psychological. In analyzing all students, the level of psychological engagement did not differ between students at pilot and nonpilot schools. The result was the same for every subgroup (Table 12). There was no difference in the level of psychological engagement between pilot and nonpilot schools for students who were Black or African American, Hispanic/Latino, or identified as Two or More Races ($t(968) = 1.17$, $p > .05$) or for FARMS recipients ($t(498) = 1.12$, $p > .05$). Further, the mean level of psychological engagement did not differ between pilot and nonpilot schools for students with limited English proficiency ($t(130) = 0.34$, $p > .05$) or for students who received special education services ($t(154) = 1.23$, $p > .05$).

Table 12
Mean Level and Standard Deviation of Psychological Engagement Scale for Student Subgroups, by Pilot and Selected Nonpilot Schools

Student subgroup	School group	*N*	Mean	Standard deviation	*t*	df
Black or African American,	Pilot	672	3.39	0.57		
Hispanic/Latino, and Two or More Races	Nonpilot	298	3.34	0.55	1.17	968
	Pilot	343	3.38	0.58		
FARMS recipients	Nonpilot	157	3.32	0.57	1.12	498
	Pilot	100	3.29	0.59		
Limited English proficiency	Nonpilot	32	3.24	0.58	0.34	130
	Pilot	104	3.28	0.60		
Special education services recipients	Nonpilot	52	3.15	0.63	1.23	154

$^*p < .05$, $^{**}p < .01$, $^{***}p < .001$.

Engagement in Small Groups

To provide descriptive data on student engagement during classes with the PBL curriculum, observers collected data on engagement when students were working in groups. Data was available only for pilot schools. Focusing on each group for one minute, the observer recorded the minimum number of members who displayed each of the following indicators of behavioral engagement:

- Nonverbal—Actively contribute/participate with the group, nonverbally (e.g., make an effort toward the task at hand, conduct task, manipulate, create).
- On task—As individuals, students do assigned work and do not engage in off-task behavior (e.g., get out of seat without permission, have off-topic conversations, gaze out window).

- Verbal—Actively contribute/participate with the group verbally (e.g., discourse, talk, ask questions, discuss; must relate to group's task or topic).

Observers collected data on small groups in 54 of 61 classes; the remaining classes had no small group work or groups met for less than one minute. In classes with small groups, two thirds of students (68%) showed nonverbal engagement in their small group (Table 13). (There is data on fewer students for the nonverbal indicator than others because some small group tasks did not include opportunities for it.) Further, almost all students (93%) displayed on task behaviors while in their small group and over one half (57%) were verbally engaged.

Table 13
Behavioral Engagement During Small Groups in Classes at Pilot Schools

Indicator	# in groups	# engaged	% engaged
Nonverbal	947	640	67.6
On task	1,238	1,145	92.5
Verbal	1,229	698	56.8

Table 14 displays behavioral engagement during small group activities for classes that followed the lesson plan with no or slight variations and for classes with more variation from the lesson plan. As expected, significantly more students were engaged during classes with fewer variations from the lesson plans for each indicator: nonverbal, 71% vs. 62% ($z = 2.68$, $p < .01$); on task, 95% vs. 88% ($z = 4.10$, $p < .001$); and verbal: 59% vs. 53% ($z = 2.19$, $p < .05$).

Table 14
Behavioral Engagement During Small Groups in Classes at Pilot Schools, by Variation from Lesson Plans

	Classes by variation from lesson plans						*z* score for group 1 vs. group 2
	No or slight variation (group 1)			Some or a lot of variations (group 2)			
Indicator	# in groups	# engaged	% engaged	# in groups	# engaged	% engaged	
Nonverbal	619	437	70.6	328	203	61.9	2.68**
On task	772	734	95.1	466	411	88.2	4.10***
Verbal	763	452	59.2	466	246	52.8	2.19*

*$p < .05$, **$p < .01$, ***$p < .001$.

Table 15 displays behavioral engagement during small group activities separately for classes with adequate implementation of PBL practices and for classes with inadequate implementation of PBL practices. On each indicator of behavioral engagement, more students were engaged in classes with adequate implementation. These differences were significant for two indicators: nonverbal, 70% vs. 58% ($z = 2.82$, $p < .01$) and on task, 94% vs. 85% ($z = 3.53$, $p < .001$).

Table 15
Behavioral Engagement During Small Groups in Classes at Pilot Schools,
by Level of Implementation of Problem-Based Learning Instructional Practices

	Level of implementation of PBL practices						*z* score for group 1 vs. group 2
	Adequate (group 1)			Not adequate (group 2)			
Indicator	# in groups	# engaged	% engaged	# in groups	# engaged	% engaged	
Nonverbal	789	549	69.6	158	91	57.6	2.82**
On task	1,024	963	94.0	214	182	85.0	3.53***
Verbal	1,021	594	57.9	208	107	51.4	1.71

*$p < .05$, **$p < .01$, ***$p < .001$.

Findings for Question 4: Does student achievement depend on which curriculum (i.e., invasive species vs. traditional) the teacher used? Is the result the same for each student subgroup, such as racial/ethnic groups?

Student scores on 10 items that every student completed on a unit test were used as measures of student achievement. The scoring rubric from the teachers was used; each of the nine multiple choice questions was worth .5 each (for a maximum of 4.5 points) and the one brief constructed response was worth 4 points. Thus the common items were worth a total of 8.5 points. Scores used in the analysis equaled the number of points attained divided by 8.5.

Analysis of Means

The first analysis tested whether the mean level for scores on the common items differed between the curricula used, as represented by pilot and nonpilot schools (Table 16). Scores on the common items were significantly higher at the pilot schools ($t(2{,}042) = 5.06$, $p < .001$). The effect size (Cohen's $d = .16$) indicated the difference was large enough to be useful in an educational setting. Because there were about 3,000 more tests from pilot schools than from nonpilot ones, a second analysis was conducted using only 2,000 tests randomly selected from pilot schools, along with all tests from nonpilot schools; the findings were the same.

Table 16
Mean Level and Standard Deviation of the 10 Common Items for Two Samples, by Pilot and Selected Nonpilot Schools

Sample (N)	School group	*N*	Mean	Standard deviation	*t*	df
	Pilot	4,745	80.0	18.1		
Full analytic sample (5,081)	Nonpilot	1,320	77.1	18.9	5.06***	2,042
Sample with 2,000 students	Pilot	2,000	80.3	18.2		
at pilot schools (2,775)	Nonpilot	1,320	77.1	18.9	4.87***	2,754

*$p < .05$, **$p < .01$, ***$p < .001$.

Regression Analysis

Regression analysis was used to further analyze the relationship between students' achievement and the curriculum used, while controlling for the following differences in student characteristics: grade level, gender, race/ethnicity, limited English proficiency, receipt of special services (i.e., FARMS, special education), and scores on the Grade 8 MSA in science. The regression excluded the format of the unit test (pilot vs. nonpilot) because it was very highly correlated with school group (pilot vs. nonpilot) and excluded enrollment in on-level or honors biology because it was very highly correlated with scores on the MSA science test. Students in Grade 12 were excluded from the regression because of the likelihood they were repeating the course and because their MSA science scores were not current.

Based on regression analysis, the relationship between curriculum used and student achievement on the 10 common items was not statistically significant ($\beta = .02$, $p > .05$) (Table 17). In other words, scores on the 10 common items did not differ between pilot and nonpilot schools, when controlling for other differences in student characteristics. As with the analysis of means, a second analysis was conducted using only 2,000 students randomly selected from pilot schools, along with all students from nonpilot schools; the findings were the same (Table 17).

Table 17
The Relationship between Curriculum Used and Student Achievement for Two Samples

	Full analytic sample (N = 5,081)			Sample with 2,000 students at pilot schools (N = 2,775)		
Variable	B	SE B	β	B	SE B	β
Pilot or nonpilot school	0.83	0.49	.02	1.05	0.55	0.03
R^2		.37			.39	
F		590.55***			351.17***	

*p < .05, **p < .01, ***p < .001.

Student Subgroups

To determine whether the result for the relationship between student achievement and the curriculum for all students was the same for student subgroups, analyses were conducted using only students in the subgroups of interest. Because at least one of the subgroups within a school group had 150 or fewer members, t-tests were used to test for differences between groups of students with limited English proficiency. Regression analysis was used for the remaining subgroups.

Table 18 presents mean scores on the 10 common items for student subgroups. For students with limited English proficiency, scores at pilot schools were higher than at nonpilot schools. This difference was statistically significant ($t(702) = 3.47$, $p < .001$) and practically significant (Cohen's $d = .33$). This finding is similar to that from the analysis of means using all students.

Table 18
Mean Level and Standard Deviation of Scores on 10 Common Test Items
for Student Subgroups, by Pilot and Selected Nonpilot Schools

Student subgroup	School group	N	Mean	Standard deviation
	Pilot	565	71.2	20.6
Limited English proficiency	Nonpilot	139	64.4	20.3
	Pilot	419	65.2	22.5
Special education services recipients	Nonpilot	179	63.8	22.0
Black or African American,	Pilot	2,968	76.2	19.3
Hispanic/Latino, and Two or More Races	Nonpilot	867	72.9	19.2
	Pilot	1874	74.3	20.2
FARMS recipients	Nonpilot	575	71.7	18.8

For students who received special education services, the relationship between curriculum used and student achievement on the 10 common items was the same as that for the regression using all students. Scores on the common 10 items did not differ between pilot and nonpilot schools ($\beta = .01$, $p > .05$) (Table 19). For students who were Black or African American, Hispanic/Latino, or Two or More Races, the findings differed from those for all students; there was a relationship between curriculum used and student achievement on the 10 common items (Table 19). Scores on common items were higher at pilot schools than at nonpilot schools for this group. This difference was statistically significant but too small to be practically significant ($\beta = .04$, $p < .05$). Similarly, for students who were FARMS recipients, scores were higher at pilot schools than at nonpilot schools. This difference was statistically significant but too small to be useful in an educational setting ($\beta = .04$, $p < .05$).

Table 19
The Relationship between Curriculum Used and Student Achievement, for Student Subgroups

	Special education services recipients (N = 504)	Black or African American, Hispanic/Latino, and Two or More Races (N = 3,159)	FARMS recipients (N = 2,002)
Pilot or nonpilot school: B (SE)	0.44 (1.83)	1.64 (0.65)	1.79 (0.85)
Pilot or nonpilot school: β	0.01	0.04*	0.04*
Model fit: F (df)	75.84***(3)	415.61***(4)	147.10***(6)
Model fit: adjusted R^2	.31	.34	.31

*p < .05, **p < .01, ***p < .001.

Recommendations

As program staff members support the second year of implementation of the new biology curriculum, they should focus training, guidance, and other supports on the following suggestions for improvement:

- Place more emphasis on teacher implementation of those PBL practices that are more critical and also have the lowest levels of implementation: connection to the central problem/unit project, student choice related to the unit project, and constructive investigation. Communication of expectations in lesson plans, along with professional learning opportunities, both may be helpful in increasing implementation.

- Encourage higher levels of fidelity to the invasive species lesson plans by overcoming some of teachers' concerns with pacing, specific lesson content, and lesson clarifications.

- Explore ways to increase the level of emotional engagement (i.e., excitement and enjoyment of class) among honors level students; the lesson plans were primarily written for on-level students, with some honors distinctions/extensions, so additional honors distinctions/extensions may be helpful.

- Explore ways to increase the level of psychological engagement (i.e., connection to team and unit project solution) among all students.

Acknowledgements

The authors are grateful to Mrs. Amy N. Gensemer, supervisor, Pre-K–12 Science; and Dr. Rhonda C. Dillard, coordinator, Science and Engineering, in the Science, Technology and Engineering program for their assistance with all facets of this study.

The authors also thank Ms. Rachel A. Hickson, evaluation specialist, Testing Unit, and Mrs. Trisha A. McGaughey, evaluation support specialist, Program Evaluation Unit (PEU), for assistance with data collection; Mr. Seong (Jeff) Jang, evaluation support specialist, PEU, for assistance with analysis; and Dr. Nyambura Maina, evaluation specialist, PEU, for her helpful suggestions on an earlier draft of this report.

Lastly, the authors thank Dr. Shahpar Modarresi, supervisor, PEU, for her guidance and support throughout this study.

References

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Washington, DC: Author.

Appleton, J.J., Christenson, S.L., Kim, D., and Reschly, A.L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, *44*(5), 427–445.

Brush, T., and Saye, J. (2008). The effects of multimedia-supported problem-based inquiry on student engagement, empathy, and assumptions about history. *Interdisciplinary Journal of Problem-based Learning, 2*(1), 21–56.

Buck Institute for Education (2014a), *Why project based learning (PBL)?,* Retrieved August 1, 2014, from http://bie.org/about/why_pbl.

Buck Institute for Education (2014b). *What is project based learning (PBL)?,* Retrieved August 1, 2014, from http://bie.org/about/what_pbl.

Center of Excellence in Leadership Learning (2009). *Summary of research on project-based learning*. Indianapolis, IN: University of Indianapolis.

Cohen, J. (1988). *Statistical power analysis for the behavioral science* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Cooper-Martin, E. & McNary, S. (2007). *Outcomes of middle school extended year courses, summer 2005*. Rockville, MD: Montgomery County Public Schools.

Cooper-Martin, E. and Wolanin, N. (2014). *Evaluation for the Universal Design for Learning Projects*. Rockville, MD: Montgomery County Public Schools.

Finkelstein, N., Hanson, T., Huang, C.W., Hirschman, B., and Huang, M. (2010). *Effects of Problem-Based Economics on high school economics instruction.* (NCEE 2010-4002). Washington, DC: National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.

Fredricks, J., Blumenfeld, P., and Paris, A. (2004). School Engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109.

Fredricks, J., McColskey, W., Meli, J., Mordica, J., Montrosse, B., and Mooney, K. (2011). *Measuring student engagement in upper elementary through high school: a description of 21 instruments*. (Issues & Answers Report, REL 2011–No. 098). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Southeast. Retrieved December 27, 2012, from http://ies.ed.gov/ncee/edlabs.

Geier, R., Blumenfeld, P.C., Marx, R.W., Krajcik, J.S., Fishman, B., Soloway, E., and 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.

Kline, R.B. (2005). Beyond significance testing: Reforming data analysis methods in behavioral research (2nd ed.). Washington, DC: American Psychological Association.

Larmer, J., & Mergendoller, J. (2012). 8 Essentials for Project-Based Learning. bie.org.

Lastica, J.R. and O'Donnell, C.L. (2007). *Considering the role of fidelity of implementation in science education research: Fidelity as teacher and student adherence to structure*. Paper presented at the Annual Meeting of the American Educational Research Association Chicago, IL.

Lipsey, M.W., Puzio, K., Yun, C., Hebert, M.A., Steinka-Fry, K., Cole, M.W., Roberts, M., Anthony, K.S., and Busick, M.D. (2012). Translating the statistical representation of the effects of education interventions into more readily interpretable forms. (NCSER, 2013–3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education.

Lynch, S., Kuipers, J.U., Pyke, C., and Szesze, M. (2005). Examining the effects of a highly rated science curriculum unit on diverse students: Results from a planning grant. *Journal of Research in Science Teaching*, *42*, 921–946.

Maryland State Department of Education (2013). Bridge plan for academic validation administrative manual. Baltimore MD.: Maryland State Department of Education.

Marx, R., Blumenfeld, P., Krajcik, J., Fishman, B., Soloway, E., Geier, R. and Tal, T. (2004) Inquiry based science in the middle grades: Assessment of learning in urban systemic reform. *Journal of Research in Science Education*, *41*(10), 1063–1080.

Merchlinksy, S. and Hansen-Grafton, B. (2007). Considering the role of fidelity of implementation (FOI) in science education: Evaluation and science specialists' role in collecting FOI data in a large school district. Paper presented at the Annual Meeting of the American Educational Research Association Chicago, IL.

Newmann, F.M., Wehlage, G.G., and Lamborn, S.D. (1992). The significance and sources of student engagement. In F.Newmann (Ed.), Student engagement and achievement in American secondary schools (pp. 11–39). New York: Teachers College Press.

Next Generation Science Standards (2013). Next Generation Science Standards front matter. Retrieved July 31, 2014, from http://www.nextgenscience.org/print/121.

Nunnaly, J. (1978). *Psychometric theory.* New York: McGraw-Hill.

Thomas, J.W. (2000). A review of research on project-based learning. Report prepared for The Autodesk Foundation. Retrieved August 1, 2014, from http://www.bie.org/index.php/site/RE/pbl_research/29.

Wang, H. and Cooper-Martin, E. (2010). Evaluation of implementation of middle school advanced courses. Unpublished paper, Montgomery County Public Schools.