Secondary Analyses of the National Environmental Literacy Assessment: Phase One & Phase Two Student, Teacher, Program, and School Surveys

~ Final Research Report ~

Prepared by:
Bill McBeth, University of Wisconsin-Platteville, Principal Investigator
Tom Marcinkowski, Florida Institute of Technology
Christos Giannoulis, Florida Institute of Technology
Harold Hungerford, Center for Instruction Staff Development and Evaluation
Trudi Volk, Center for Instruction Staff Development and Evaluation

With:
Jennifer Howell, Florida Institute of Technology

Submitted to:
Sarah Schoedinger, Senior Program Manager, Office of Education,
National Oceanic and Atmospheric Administration,
U.S. Department of Commerce
and
Judy Braus, Executive Director,
North American Association for Environmental Education

Grant # NA12SEC0080018
July 29, 2014

ACKNOWLEDGEMENTS

We send our most sincere gratitude to the National Oceanic and Atmospheric Administration who provided funding for this project, the United States Environmental Protection Agency for its continued support, and to The North American Association for Environmental Education, who administered the grant.

And especially, we would like to recognize the following individuals, groups and organizations for their assistance and support:

- Sarah Shoedinger from NOAA's Office of Education;
- Ginger Potter from the US EPA's Office of Environmental Education;
- Judy Braus, Drew Price, and other staff at NAAEE;
- The Center for Instruction Staff Development and Evaluation;
- University of Wisconsin-Platteville: School of Education; and
- Florida Institute of Technology, including the Department of Education and Interdisciplinary Studies, Dean Hamid Rassoul, and Dr. Kastro Hamed for their support for this research.

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

EXECUTIVE SUMMARY

This report is the culmination of Phase Three of the National Environmental Literacy Assessment (NELA) project. Funding was provided by the National Oceanic and Atmospheric Administration (NOAA), and the project was administered by the North American Association for Environmental Education (NAAEE). Key partners included researchers from the University of Wisconsin Platteville, the Center for Instruction, Staff Development and Evaluation in Illinois, and the Florida Institute of Technology.

During Phase One of NELA the research team generated a baseline for environmental literacy by assessing the environmental literacy of a randomly selected sample of middle school students in grades six and eight. Phase Two utilized a sample of schools that were purposefully selected for their efforts in the development of environmental literacy among middle school students and included six-, seventh-, and eighth-grade students.

While the Middle School Environmental Literacy Survey was used to assess the environmental literacy of these middle grades students, its first section also contained student demographic questions. These questions were utilized as a source of analysis in this phase. In addition, both of these phases used three additional surveys: the Teacher Information Form gathered demographic and general information on the classroom teachers; the Program Information Form gathered information on the programs to which students were exposed; and the School Information Form gathered information on school demographics.

Phase Three was designed to determine which, if any, of the data from this instruments could be used to predict aspects of environmental literacy or had significant impacts on the development of environmental literacy and/or environmental behavior. For this purpose the following research questions were created.

1. What is the relative contribution of knowledge, affect, and skill variables to actual commitment or behavior within the Phase One sample and within the Phase Two sample?
2. To what extent does the sequence of grade-level programming in Phase Two schools appear to have contributed to or influenced students’ environmental literacy scores?

3. To what extent do the student, teacher, program, and school variables measured during Phases One and Two appear to have contributed to or influence students' environmental literacy composite scores?
4. To what extent do the student, teacher, program, and school variables measured during Phase Two appear to differentiate between high- and low- performing schools as determined from students' environmental literacy scores?

Research Question One

What is the relative contribution of knowledge, affect, and skill variables to actual commitment or behavior within the Phase One sample and within the Phase Two sample?

Findings

The affective characteristics of verbal commitment and environmental sensitivity appear to be significant predictors of actual commitment (or environmental behavior) for both sixth- and eighth-grade students from a representative sample of middle schools in the U.S (the representative group; Phase One).

Verbal commitment, environmental sensitivity, and environmental feeling all appear to be significant predictors of actual commitment (or environmental behavior) for six-, seventh-, and eighth-grade students in schools with environmental education programs (the environmental education group; Phase Two). In addition, cognitive skills emerged as important predictors in this group, with issue analysis skills a significant predictor for grades six and seven, and issue identification skill a significant predictor for grade eight.

Students in the representative group and students in the environmental education group were similar in that verbal commitment, environmental sensitivity, and environmental feeling (respectively) were the strongest predictors of actual commitment or environmental behavior for all grades

There was a noteworthy difference between the two groups of students regarding cognitive skills as significant predictors. Cognitive skills appear to be significant predictors of environmental behavior within the environmental education group (issue analysis skills for grades six and seven, and issue identification skills for grade eight), but were not present as significant predictors within the representative group.

Discussion

Affective components (i.e., Verbal Commitment, Environmental Sensitivity, and Environmental Feeling) appear to be significant predictors of environmental literacy and merit attention in curriculum development and educational programming.

The emergence of cognitive skills as significant predictors of environmental behavior among students in schools with environmental education programming also merits considerable attention. Two aspects of cognitive skills surfaced as significant predictors of environmental behavior (issue analysis for sixth- and seventh-grade students and issue identification for eighth-grade students).

While the relationship between affect and behavior appears similar in middle school students engaged in school-based environmental education programs and other middle school students, there are striking differences between those two groups regarding the relationship between cognitive skills and environmental behavior.

Recommendations for Further Study

Questions such as the following might shed light on many aspects of environmental literacy and environmental behavior.

1) How is the development of environmental sensitivity, and other important affective dimension of environmental literacy related to students' educational experiences and how is it related to environmental influences outside the school, such as home and community?
2) What educational practices lead to a stronger relationship between aspects of environmental affect and environmental behavior?
3) What educational practices lead to a stronger relationship between cognitive skills and environmental behavior?

4) Why do cognitive skills appear to be inconsistent as predictors of environmental literacy?
5) What kinds of interactions, if any, appear to exist among environmental knowledge, affect, and skill variables as they relate to environmental behavior?

Research Question Two

To what extent does the sequence of grade-level programming in Phase Two schools appear to have contributed to or influenced students' environmental literacy scores?

Findings

There appeared to be no statistically significant difference between the composite scores of seventh-grade students in schools that offered environmental education programming only for Grade Seven and those of seventh-grade students in schools that offered environmental education programming for Grades Six and Seven.

A similar comparison was made among eighth-grade students in schools with environmental education programming for Grade Eight only, for Grades Seven and Eight, and for Grades Six, Seven, and Eight. There were significant differences in composite scores between and among these three groups, favoring schools offering programs across two or three grades. Further, significant differences in knowledge, skill, and behavior component scores followed this pattern. However, this did not hold for the affective component scores.

Discussion

From developmental and educational standpoints, the results for both seventh- and eighth-grade students might be explained by the likely effects of additional opportunities for experience and learning among students who have been involved in environmental education programming at multiple grade levels. We do not understand why significant differences in composite scores were evident when comparing schools with programming in eighth grade only to those with programming in seventh through eighth and those with programming in sixth through eighth grades and not between the two groups of younger seventh-grade students.

Recommendations for Further Study

Rather than focus on groups of students from different grade levels at a single point in time, it might be more productive to focus on one group of students, over time, as they move through several grades of school. Such research might include the following.

1) What are the major features of the program in each grade (e.g., goals, curricula, other instructional resources and sites, institutional and network affiliations, teacher subject areas and responsibilities)?
2) What is the scope and sequence of environmental education programming across grades, if any?
3) In what ways, if any, are steps taken to ensure both fidelity of programming within and continuity of programming across grades?

Research Question Three

To what extent do the student, teacher, program, and school variables measured during Phases One and Two appear to have contributed to or influence students' environmental literacy composite scores?

Findings

The only variable found to have a significant influence on student composite scores in at least one grade in Phase One and in Phase Two were general environmental program characteristics reflected in the various EE Program Types (e.g., environmental curricula, outdoor labs, residential camp programs, school clubs).

A number of other Program variables were found to have a positive or a negative influence on student composite scores. These included variables associated with Program Goals (i.e., issue investigation skills), Instructional Sites (i.e., science labs and libraries/media centers), and Instructional Methods (i.e., cooperative learning, projects, discussion, and service learning).

Several Teacher variables were found to have an influence on environmental literacy composite scores. These included variables associated with each teacher's professional development: their Highest Degrees Earned (i.e., the level of schooling featured in that degree program); their Teacher Certification(s) (i.e., whether they had earned or were working on this; the level of schooling reflected in each certification); their Years Teaching (i.e., in total, and in middle schools); and the number of EE inservices they had attended (i.e., those lasting 1-2 days). The results on these variables were mixed (i.e., some had a positive influence on student composite scores and others a negative influence) and spotty (i.e., only the use of science labs as an instructional site was a significant influence in more than one grade).

Discussion

The findings reported here are constrained by the nature of the Phase One and Two samples and reflect major characteristics of those samples. The representative group in Phase One included only one program and one teacher for each grade (i.e., one sixth-grade and one eighth-grade class per school). From a multilevel perspective, there were only multiple students per school and multiple schools (i.e., two levels) that could be included within the analyses. The environmental education group in Phase Two included only one program per grade, although there were as many as nine (9) teachers per grade. From a multilevel perspective, there were multiple students per teacher and multiple teachers per school, as well as multiple schools. However, this number of teachers was reduced by about half due to missing teacher data. Whether or not this missing data had an influence on the results of the multilevel analyses reported here is unclear. However, what is clear is that, as in Phase One, multiple students and multiple schools had to be accounted for in the multilevel analyses. These findings are also constrained by the size of and variability within the Phase One and Two samples. In many instances, variables were found to have a significantly positive or significantly negative influence on student composite scores, but this held true for students associated with a small number of schools and/or teachers.

Recommendations for Future Research

These analyses were conducted on data sets collected during two prior studies and were an attempt to explore relationships among Students, Teacher, Program, and School variables and environmental literacy scores. The findings provide direction for those interested in continuing to probe these relationships. In terms of future research, we recommend replication of these analyses, using larger and more diverse samples of schools, programs and teachers that would be sufficient for multilevel analyses of this kind. This would help to overcome the limitations associated with significant findings associated with only a few schools, programs, and/or teachers. Additional questions to pursue might include the following clusters of questions. This research could be guided by the following questions.

1) Which program characteristics are apparent in each school's environmental education programs?
2) Which of these characteristics have a stable and prominent role in each program? Further, what are these roles?
3) What kinds of evidence exist regarding the effects/impacts of each school's environmental education program as a whole, and of each of these general program characteristics on environmental literacy and/or environmental behavior?
4) Which characteristics associated with the professional development of teachers, both in general and in environmental education, are apparent in the background and/or practices of teachers in each school's environmental education programs?
5) In what ways, if any, do state, district, and school policies appear to be related to or influential on these professional development variables? Further, in what ways, if any, do the perspectives of state, district and school administrators appear to be related or influential on these variables?
6) Which of these professional development variables have a visible, stable, and prominent role in each program? In addition, what are those roles? Further, in what ways do these variables appear to influence each program?

Research Question Four

To what extent do the student, teacher, program, and school variables measured during Phase Two appear to differentiate between high- and low- performing schools as determined from students' environmental literacy scores?

Findings

This question sought to determine the student, teacher, program, or school characteristics that could differentiate between high-performing schools and low-performing schools. In these results, we have searched for patterns that span multiple grade levels rather than individual grade-level findings. These findings include the key characteristics that show a pattern of repetition across at least two grade levels.

School and Student characteristics did not show up as significant predictors of school performance on levels of environmental literacy. Of the remaining variables, more Teacher characteristics were predictive of high-levels of performance at multiple grade levels (five predictors) than Program characteristics (two predictors).

Discussion

One may hypothesize that in middle schools with environmental programming across consecutive grades, neither the location of the school, its socioeconomic situation nor the make-up of the student body may be as important as who is teaching the students and how they are taught. Further, it appears to be much easier to predict characteristics of high-level performance (21 predictors) than low-level performance (5 predictors). This may mean that it would be easier to identify, based on characteristics, high- rather than low-performing schools. That being said, we must add that care must be taken when comparing and/or combining variables from the analysis of school/program/student variables with the analysis of teacher variables due to sample differences (e.g., due to data loss, the sample used in the analysis of teacher variables was smaller than the sample used in the analysis of student, program, and school variables). We do not know how this loss of data may have influenced these results.

Recommendations for Further Study

These analyses yielded few clear-cut results. Among the program characteristics, however, teaming appears to be a significant ingredient of high-performing schools, particularly at the eighth-grade level. This raises questions such as the following.

1) What is the duration (in weeks) and intensity (hour per week) of each environmental program?
2) What roles and responsibilities do teachers in each subject have in planning and implementing this program? Further, are there noticeable differences in the amount of instruction in this program associated with each subject?

A key teacher characteristic in high performing schools appears to be the self-reported perceived level of activity in environmental protection. The apparent 'ordinal' differences among teachers with slight/moderate and considerable/extreme perceived levels of participation raise the following questions:

3) How do teachers understand the domain of citizenship participation/environmental action and interpret their level of participation?
4) What kinds of issues are these modes of participation intended to target?
5) How frequently are these activities undertaken (per week, month, or year) and for how long (in years) have these teachers been active?

INTRODUCTION AND BACKGROUND

The National Environmental Literacy Assessment Project

The National Environmental Literacy Assessment research team undertook the NELA project with a broad goal of assessing and studying environmental literacy in U.S. middle schools. We look at environmental literacy as the capacity to act on a broad understanding of how people and societies relate to each other and to natural systems. Environmental literacy also includes the ability and willingness to incorporate environmental considerations into daily decisions about consumption, lifestyle, career, and civic participation, through individual and collective action. (Definition adapted from Campaign for Environmental Literacy, 2007).

We envisioned the NELA project as a series of connected studies. Our first step was to conduct a national assessment to establish a baseline measure of environmental literacy among middle school students in the U.S (McBeth, et al, 2008). Our next step was an assessment of environmental literacy among U.S. middle school students engaged in ongoing environmental education programs and a comparison of those levels to the baseline level (McBeth, et al, 2011). Following that, we hoped to broaden our efforts to include both quantitative and qualitative methodologies, and to oversee a collection of studies (conducted within selected middle schools and middle school classrooms), designed to explore factors that appear to influence the nurturing and development of environmental literacy. Ultimately, we hoped to identify educational practices and conditions that might enhance the development of environmental literacy in our young people.

The series of related yet separate studies were designed around the following leading questions.

> What is the state of environmental literacy among U.S. middle school students?
>
> How does environmental literacy among U.S. middle school students, in general, compare to that of their counterparts who participate in school-based environmental education programs?
>
> How do the knowledge, affect, and skill components of environmental literacy impact responsible environmental behavior in adolescents and what are the relative contributions of the components to responsible environmental behavior?

What individual variables or combinations of variables (i.e., student, family, teacher, program, and school) appear to influence the development of environmental literacy?

How do the characteristics and practices of students, families, teachers, administrators, programs and schools differ in schools that measure high on the environmental literacy continuum from those that measure low on the continuum?

Earlier Phases of NELA

The NELA Phase One study addressed question one, above, and established a baseline measure of environmental literacy among U.S. middle school students. Phase One utilized a nationally representative sample of 48 middle schools that included 93 teachers and 2,004 sixth- and eighth-students, and provided baseline data on environmental literacy among U.S. middle school students.

The NELA Phase Two study measured environmental literacy among U.S. middle school students engaged in environmental education programs and compared those results to those of the general population of U.S. middle school students. Phase Two utilized a purposeful sample of 64 middle schools with ongoing environmental education programs that included 214 teachers and 7,965 sixth-, 7^{th}-, and 8^{th}-grade students, and that provided comparison data related to environmental literacy among middle school students engaged in environmental education.

These two phases were important in that they provided the team with a measure of the general state of environmental literacy in U.S. middle schools, a measure of environmental literacy in U.S. middle schools where environmental education was an intentional component within the curriculum, and a comparison of those two measures of environmental literacy.

We used descriptive and inferential statistics (t-test, z-test, and Cohan's d) to compare sixth- and eighth-grade students from the representative sample of U.S. middle schools to sixth- and eighth-grade students from the purposeful sample of middle schools with environmental education in place, and found small, but significant, differences between the two samples for both sixth- and eighth-grade levels. These differences were evident in comparisons of grade-level students scores and grade-level school scores.

Comparing the results from the two samples also allowed us to identify Phase Two schools in which students' grade level environmental literacy scores were significantly higher than those of grade-level cohorts from schools in Phase One. However, these comparisons provide little insight into why these elevated literacy levels exist.

In exploratory fashion, we created quartiles among the Phase Two purposeful sample of schools to examine how the top and bottom performing schools might differ on several contextual and demographic variables (McBeth et al., 2011). Among the variables that appeared to differ across the top and bottom quartiles of schools were school size, curricular organization and organization for teaching, programming across grades, ethnic composition of schools, and percent of students on free and reduced lunch, etc.

These descriptive comparisons allowed us to make preliminary observations about differences between the two samples. The preliminary nature of these observations did not enable us to draw conclusions, or to make educational recommendations related to these differences. Nonetheless, the results of the comparisons suggested that further study into the relationships between demographic and contextual variables and the development of environmental literacy might be productive.

Phase Three Overview

Research Problem

If we hope to positively impact the development of environmental literacy within a general population of students within U.S. schools, is appears to us that it is extremely important to delve into and explore the factors and conditions that exist in those school settings with higher environmental literacy levels. These factors and conditions might include demographic characteristics of students, teachers, and schools, teacher attributes and attitudes, instructional methods and practices, and administrative, parental and community features. In so doing, perhaps we might begin to understand conditions conducive for fostering the development of environmental literacy among U.S. middle school students.

Research Purposes

A major purpose of the Phase Three study was to conduct secondary analyses of school, program, teacher, and student data collected during the Phase One and Phase Two studies. We intend these analyses to be an initial attempt to address to the fourth leading question: What individual variables or combinations of variables (i.e., student, family, teacher, program, and school) appear to influence the development of environmental literacy?

In general, we wanted to learn how differences in environmental literacy (both within and across the Phase One and Phase Two data sets) might be explained by student, teacher, program, and school variables. We hoped that by applying methods of statistical analysis to the Phase One and Phase Two data sets, we might discover relationships between these contextual variables and environmental literacy. In turn, we expect that findings from these analyses might provide guidance as we approach the subsequent NELA effort, the collection of studies (conducted within selected middle schools and middle school classrooms), designed to explore factors that appear to influence the nurturing and development of environmental literacy.

A second purpose of the Phase Three study was to utilize the information we had already gathered via the MSELS to explore the relationship between environmental behavior and the other measured environmental literacy variables (i.e., ecological knowledge, verbal commitment, environmental feelings, environmental sensitivity, issue identification skills, issue analysis skills, and action planning skills). We were also interested in discovering how, if at all, these relationships differed across the two samples of U.S. middle school students, those from the nationally-representative baseline sample of schools and those from the purposeful sample of schools with ongoing environmental education programs.

Thus, Phase Three addressed questions concerning the magnitude and influence of contextual and demographic variables on environmental literacy, as well as the relationships between selected environmental literacy variables and environmental behavior. Phase Three also explored differences in students and schools between the Phase One and Phase two samples and differences in students and schools within the Phase Two sample.

Research Questions

The research team designed four major research questions to guide Phase Three. For each research question, one or more specific research questions were also crafted. The major research questions and their specific research questions are presented below.

1. What is the relative contribution of knowledge, affect, and skill variables to actual commitment or behavior within the Phase One sample and within the Phase Two sample? Further, what are the similarities in and differences between these two samples?

    1a. Within the Phase One sample, what is the relative contribution of MSELS scale scores to an explanation of the variance in student MSELS actual commitment (behavior) scores for sixth-grade students and for eighth-grade students?

    1b. Within the Phase Two sample, what is the relative contribution of MSELS scale scores to an explanation of the variance in student MSELS actual commitment (behavior) scores for sixth-grade students, for seventh-grade students, and for eighth-grade students?

    1c. What are the similarities in and differences in the relative contribution of MSELS scale scores to an explanation of the variance in student MSELS actual commitment (behavior) scores for sixth-grade, seventh-grade, and eighth-grade students in the Phase One and Phase Two samples?

2. To what extent does the sequence of grade-level programming in Phase Two schools appear to have contributed to or influenced students' environmental literacy scores?

    2a. Within the Phase Two sample of schools (n=64), what differences, if any, exist between MSELS *component* and *composite scores* of seventh-grade students who attended schools with environmental programming at Grade Seven only, and seventh-grade students in schools with environmental programming at Grades Six and Seven?

    2b. Within the Phase Two sample of schools (n=64), what differences, if any, exist between MSELS *component* and *composite scores* of eighth-grade students who attended schools with environmental programming at Grade 8 only, schools with environmental programming at Grades Seven and Eight only, and schools with environmental programming at Grades Six, Seven, and Eight?

3. To what extent do the student, teacher, program, and school variables measured during Phases One and Two appear to have contributed to or influence students' environmental literacy composite scores?

   3a. To what extent do Student, Teacher, Program, and School variables measured during Phase One appear to appear to have contributed to or influenced sixth-grade and eighth- grade students' environmental literacy composite scores?

   3b. To what extent do Student, Teacher, Program, and School variables measured during Phase Two appear to appear to have contributed to or influenced sixth-grade, seventh-grade, and eighth-grade students' environmental literacy composite scores?

   3c. What are the similarities and differences in the extent to which Student, Teacher, Program, and School variables appeared to have contributed to or influenced students' environmental literacy composite scores for sixth-grade, seventh-grade, and eighth-grade students in the Phase One and Phase Two samples?

4. To what extent do the student, teacher, program, and school variables measured during Phase Two appear to differentiate between high and low performing schools as determined from students' environmental literacy scores?

   4a. Which school, program, teacher, and student characteristics appear to differentiate between the high- and low-performing schools, as determined by the distribution of *composite scores*, at the 6^{th} grade level, at the 7^{th} grade level, and at the 8^{th} grade level?

Delimitations and Limitations of the Study

The research team found this Phase Three study, a secondary analysis of Phase One and Phase Two data, to be challenging in many ways. Two of the these challenges are inherent in the exploratory nature of this study: (1) there were no applicable theories to guide the selection of variables for inclusion in Research Question Three or Four analyses; and (2) there was little, if any, prior research in or closely related to environmental education to inform our methods and procedures, or against which to compare these results. As a result, the research team has had to rely on literature in other fields. In these and other ways, the methods described in this report, along with the delimitations and limitations described below, reflect the historical context in which this study was undertaken.

Delimitations. There are three delimitations applicable to this study:

1. This study is restricted to the use and analysis of data collected during Phase One and Phase Two of this National Environmental Literacy Assessment (NELA) Project.

2. The kinds of data analysis undertaken in this study have focused on the four Research Questions presented in this document. The only exceptions to this were the exploratory analyses described in the Methodology section, two of which can be viewed as preparatory or preliminary steps in analyses for several research questions (i.e., the analysis of bivariate relationships and of School, Program, School, and Student variables as separate sets of variables). Beyond this, in only two sets of analyses did we use anything other than composite scores as the dependent variable: (a) for Research Question One where behavior component scores served as the dependent variable, and (b) for the exploratory analysis in which skill component scores served as the dependent variable.

3. This project was bound by time. This was proposed as a one-year project and has been granted a one-year no-cost extension, so all work on this project had to be completed in this two-year period.

Limitations. Five limitations influenced the conduct and reporting of this study.

1, Missing Student Data: Steps were taken during Phase One and Phase Two to address the problem of missing data within the student data set (i.e., responses on the MSELS). Students who left 25% or more of the items for any MSELS scale blank had all responses for that scale deleted as invalid and unusable responses. Further, hot-deck imputation methods were used to fill in as many of the remaining blanks as possible. Finally, when preparing data sets for Phase Three analyses, any student who was missing one or more component scores had her/his entire record deleted from the data set. These addressed most, but not all, problems with missing student data.

Perhaps more importantly, these steps did not address problems of missing data in School, Program, and Teacher data files. During Phases One and Two, attempts were made to fill in missing School data using data available on the U.S. Department of Education's National Center for Educational Statistics database. Again, this addressed some, but not all, problems associated with missing School data. However, there are few methods available to address problems associated with missing Program and Teacher data, and none of these were considered feasible (e.g., the time interval between the collection of data in Phases One and Two and the start of this study; changes in schools, programs, and school personnel; and the absence of contact information for teachers).

Thus, the research team acknowledged these missing data problems, and conducted an analysis to determine if these were missing at random, which they were. To accommodate missing data in these analyses, we used the *listwise deletion* option in the statistical analysis packages in an effort to reduce the effects of missing data on these results.

2. <u>Missing Teacher Data</u>: The problem of missing teacher data is substantially larger than missing responses (blanks) in Teacher Information Forms (TIFs). Here's why that is so. The research team attempted to match the names of teachers on TIFs to the names of teachers on Scantron forms completed by students. As discussed in this report, we were unable to do so or were able to do so on a very limited basis (e.g., establish a match for 5-10 Scantron forms). As summarized in Appendix D, this led to the loss of about one half of the Phase Two data set. Although this loss had no influence on analyses for Research Questions One and Two, it did have an influence on analyses for Research Questions Three and Four.

For Research Question Three, because student composite scores served as the unit of analysis and the size of student samples was sufficiently large, the research team decided to accept this loss of data when merging School, Program, Teacher, and Student data for each Phase and grade level into one large data file. The resulting sample sizes are reported for this research question. However, this could not be done for Research Question Four because we lost about half of the data for each grade in Phase Two due to the use of quartiles (i.e., keeping schools in Quartiles 1 and 4, but dropping schools in Quartiles 2 and 3). Had we allowing for any further loss of data due to these missing teacher data, these Phase Two samples would have been reduced to approximately one quarter of their original size. The research team did not think this was acceptable, so for Research Question Four, the analysis of Teacher data was conducted separately from the analysis of School, Program, and Student data.

3. Assumptions Associated with Linear Regression and Other Modeling Techniques: Shortly after starting data analysis, the research team quickly discovered that the Phase One and Phase Two student data sets did not meet several of the assumptions associated with linear regression and other modeling techniques. Initial attempts were made to identify and eliminate student records in an effort to create data sets that would meet these assumptions. However, even when records were deleted to meet one assumption, the reduced data sets failed to meet other assumptions, and therefore would have led to the deletion of additional student records. It quickly became apparent that this would lead to the loss of hundreds of student records in each data set, and this was deemed unacceptable. Thus, using each original data set, the research team used resampling techniques and other robust methods to ensure that analysis results such as coefficients and confidence intervals would accurately represent that data set.

4. Use of Multilevel Analysis Methods: It is fair to point out that the Phase One study and, to a lesser extent, the Phase Two study were not designed with multilevel analyses in mind (i.e., where each grade-level sample would include different programs, multiple teachers, and multiple classes for each teacher). On the other hand, for Research Question Three, multilevel analysis methods were used in variable selection (Step 3), to confirm the need for multilevel analysis (Step 4), and in the final HML analyses (Step 5). Multilevel analysis methods were not used in any of the analyses for Research Questions One, Two, or Four. In light of the confirmation of the need to use (Step 4) and the benefits of using (Step 5) multilevel analysis for Research Question Three, the research team will need to explore whether additional forms of multilevel analysis could be used and could be added any value for any other step in analyses for Research Questions One, Two and/or Four, once Phase Three has been completed.

5. Use of Non-linear Analysis Methods: Members of the research team were aware of advances in non-linear analysis, and of the potential value of non-linear analysis in light of the data sets we were working with. However, non-linear analysis methods used on a very limited basis, notably in the use of MLP analyses during Step 3 of variable selection for Research Question Three.

Whether additional forms of non-linear analysis could have been used and could have added any value for any other step in analyses for these four research questions also is a question that the research team will need to explore after Phase Three has been completed.

METHODOLOGY

Sources of Data

The earlier phases of NELA were similar in a number of ways. Both studies addressed general levels of environmental literacy on eight variables associated with environmental literacy, and the developmental level of the subjects was similar in both studies (Phase One: sixth- and eighth-grade students; Phase Two: sixth- seventh- and eighth-grade students). Additionally, in both studies, data were collected late in the spring semester to allow for maximum effect from both student maturation and environmental programming. Finally, and of importance to this current study, both of the earlier studies utilized similar data collection instruments and protocols. The four data collection instruments used in the two studies are described below.

The Middle School Environmental Literacy Survey (MSELS) was used to gather information on knowledge, affect, skill, and behavior variables associated with environmental literacy. It was designed to be used with sixth- through eighth-grade students. A description of the development of and research on the MSELS is reported by McBeth et al., (2008).

The School Information Form (SIF) was completed by a school official, and was designed to gather descriptive information on the participating schools (e.g., ethnic make-up of the student body, class size, number of students on free and reduced lunch).

Also, at least one **Program Information Form (PIF)** was required from each participating grade level in each school that participated. The program form was designed to gather information about the environmental programs to which participating students were exposed (e.g., length of program, type of curriculum, organization for instruction, and aspects of curriculum implementation).

In addition, a **Teacher Information Form (TIF)** was completed by each teacher of a participating class. The teacher form contained personal questions (e.g., age and ethnicity), professional questions (e.g., years in teaching, years at different developmental levels, and types of educational license), as well as questions that asked teachers about their views on the environment and environmental education (e.g., importance of EE to students, sensitivity to the environment, and active involvement in environmental protection efforts). (McBeth et al., 2008).

Table 1 presents a complete list of contextual and demographic variables included on the school, program, and teacher forms, organized by type of variable. We collected data related to those variables because we considered them ones that might influence the development of environmental literacy.

Table 1. *Variables Included in NELA Data Collection*

School Information	**Student Information**
School grade level configuration	Student age
School enrollment	Student grade
School type	Student gender
School locale	Student ethnicity
Student/teacher ratio	
Number reduced/free lunches	**Teacher Information**
Ethnic composition of school	Teaching years - total
Other social characteristics of school	Teaching years – middle grades
School-wide theme (e.g., environmental)	Current grade level(s)
Other school information	Current subject(s)
	Previous grade level(s)
EE Program Information	Previous subject(s)
Grade Level involvement in EE	Certification status
Nature of EE program by grade	Certification type
Years of Program existence by grade level	Additional certification/endorsements
Length of program by grade	Educational degrees earned
Program affiliation with EE network	EE college courses (number)
Program use of EE curricula	EE college courses (type)
Program use of EE approach	EE training – inservices /workshops (number)
Purpose or direction of EE program	EE training – inservices/workshops (length)
Major Program goals and objectives	Characteristics of relevant EE courses
Curricular/Instructional organization	Teacher gender
Organization for teaching	Teacher age
Organization of students for learning	Teacher ethnicity
Teaching/learning settings	Teacher perception of Importance of EE to students
Classroom teaching methods	Teacher perception of Importance of EE to self
Classroom assessment techniques	Teacher sensitivity toward the environment
Additional important program features	Teacher concern on environmental problems/issues
Extent of EE participation w/in grade level	Teacher level of environmental action

Preparation of Data Files

Each specific research question required the preparation of unique data files. Three steps were involved in the preparation of these data files. In the **first step,** we prepared initial MicroSoft Excel data files for items on the School Information Form (SIF), Program Information Form (PIF), Teacher Information Form (TIF), and student demographic items on the MSELS. In the **second step,** we prepared MicroSoft Excel data files required to conduct statistical analyses pertinent to each research question, including appropriate data from these initial data files and appropriate Environmental Literacy Scale, Component, and Composite scores. In the **third step**, we modified these Excel data files for use in statistical software programs (e.g., JMP, SPSS, R language scripts).

Research Methods and Statistical Analyses

In a similar fashion, each specific research question demanded distinct methods and statistical analyses. Analytical methods of analyses included multiple linear regression analyses, univariate and multivariate analyses of variance, post-hoc analyses, chi-square tests, and discriminant analysis. In addition, resampling techniques were employed, where appropriate.

Organization of This Report

Reports for Each Research Question

We will present individual reports of the methods and results for each research question. Each report will present the intent of the research question and its rationale. This will be followed by a description of each related specific research questions, along with its variables, research methodology and statistical analyses, and results. Finally, we will close each report with findings, discussion, and recommendations, tying the results from the specific questions back to their research question. It is important to note that the four research reports that follow are simplified descriptions of the methods and statistics employed in answering each research question. Those who are interested in detailed descriptions and explanations of methods and results for the research questions should refer to Appendices A, B, C, and D.

Exploratory Analyses

In addition to the analyses related to the four research questions, the research team conducted two sets of exploratory analyses to better understand relationships between pairs of variables and sets of variables, and with other dependent variables in mind (e.g., skill component scores instead of composite scores). Information about the methods and results of these analyses are presented in Appendices F and G.

One set of exploratory analyses focused on the extent to which School, Program, Teacher, and Student characteristics may differentiate between two groups of Phase Two schools: schools with high skill component scores and schools with low skill component scores (Appendix F). In these analyses, all School, Program, Teacher, and Student characteristics were included in one master set, rather than analyzed as separate sets. Using this master set, we sought to determine which characteristics best explained (predicted) the differences in environmental literacy skill component scores in these two groups of schools.

A second set of exploratory analyses focused on the extent to which sets of School, Program, Teacher, and Student characteristics may help to explain differences (variability) in Environmental Literacy Composite scores (Appendix G). In these analyses, we sought: (a) to determine how well each of these four sets of characteristics, considered individually, might predict these scores; and (b) to identify characteristics within each set that may be the best predictors of these scores.

REPORTS FOR EACH RESEARCH QUESTION

Research Question One

This research question was designed to explore the contribution of knowledge, affect, and skill to actual commitment (environmental behavior). In these analyses we sought to determine the extent to which the MSELS scores for knowledge, affective, and skill variables contribute to an explanation of (predict) actual commitment scores. Detailed descriptions and explanations of methods and results for this research question can be found in Appendix A.

Rationale

There are a number of theoretical and research-oriented writings related to environmental behavior and its development (Aizjen & Fishbein, 1985; Bamberg & Moser, 2007; Fishbein & Ajzen, 1975; Hines, Hungerford & Tomera, 1986/87; Hungerford & Volk, 1990; Kollmuss and Agyeman, 2002). Similarly, a number of researchers have studied predictors of responsible behavior in adult populations (Marcinkowski, 1989, 2005; Sia, Hungerford & Tomera, 1985/86; Sivek & Hungerford, 1989/90). However, we know of no studies of environmental behavior and its predictors among middle school-aged individuals. The large sample sizes from the NELA Phase One and Phase Two studies (n = 2,004 and n = 7,965, respectively) provided us with opportunities to explore the extent to which MSELS knowledge, affect, and skill variables might contribute to (or predict) actual commitment (environmental behavior) scores. Thus, this research question was designed to explore the contribution of ecological knowledge, verbal commitment, environmental feeling, environmental sensitivity, issue identification and analysis skills, and action planning skills to actual commitment (environmental behavior) among middle school-aged individuals.

Specific Research Questions and Variables

Initially, we explored how the environmental behavior of sixth- and eighth-grade students in the Phase One sample might be explained by their environmental knowledge, affect, and skill characteristics (Specific Research Question 1a). Subsequently, we explored how the environmental behavior of sixth-, seventh-, and eighth-grade students in the Phase Two sample might be explained by their environmental knowledge, affect, and skill characteristics (Specific

Research Question 1b). Finally (in Specific Research Question 1c), we compared the two samples with regard to the outcomes of the earlier analyses.

Overview of Methods

For specific research questions 1a and 1b, multiple linear regression analyses were undertaken for each grade level within each sample. We used student scores from the MSELS scales that measured their knowledge, affective, and skill characteristics as potential predictor variables. For these research questions, the individual student was the unit of analysis. For specific research question 1c, the results of these regression analyses were compared descriptively.

Cohen et al. (2003), Tabachnik and Fidell (2013), and others recommend that the data must meet a set of necessary assumptions in order to obtain valid and accurate multiple linear regression results. The research team found that Phase One and Phase Two data sets did not meet several of these assumptions (e.g., the influence of extreme scores referred to as outliers). Rather than drastically reduce each data file in order to meet all of these assumptions, the research team used resampling techniques called bootstrapping (Efron & Tibshirani, 1994; Field & Miles, 2012). This allowed the team to identify the relative contribution of student MSELS scale scores to the prediction of actual commitment scores.

Results for Specific Research Question 1a

These results include the relative contribution of scores on MSELS measures of students' knowledge, affect, and skills as predictors of their Actual Commitment (behavior) scores for Phase One samples.

For the **Phase One sixth-grade sample** (Table 2), 54% of the variance in Actual Commitment scores was explained by MSELS scale scores ($F=133.51$, $p < 0.001$). Two affective scales (characteristics) were statistically significant predictors of Actual Commitment scores for this sample: Verbal Commitment ($\beta = .51$, $p <.001$); and Environmental Sensitivity ($\beta = .33$, $p <.001$).

Table 2. *Summary of Results for MSELS Scale Scores as Predictors of Actual Commitment (Behavior) for the Phase One Sixth-Grade Sample*

	R^2	*B*	Confidence Intervals	*SE B*	*ß*	*p*
Model	0.54					
(Constant)		1.02	-1.56, 3.56	1.31		.43
Ecological Knowledge		0.12	-0.03, 0.29	0.08	.05	.11
Verbal Commitment		0.52	0.45, 0.58	0.03	.51*	<.001
Environmental Sensitivity		0.41	0.33, 0.49	0.04	.33*	<.001
Environmental Feeling		-0.10	-0.38, 0.17	0.26	-.02*	.50
Issue Identification		0.07	-0.44, 0.60	0.26	.01	.79
Issue Analysis		0.14	-0.13, 0.42	0.14	.03	.29
Action Planning		-0.01	-0.08, 0.07	0.00	.00	.87

Note: * Statistically significant at $p < .05$ ** Statistically significant at $p < .01$

For the **Phase One eighth-grade sample** (Table 3), 52.1% of the variance in Actual Commitment scores was explained by MSELS scale scores ($F=118.81$, $p < 0.001$). All three affective scales (characteristics) were statistically significant predictors of Actual Commitment scores: Verbal Commitment ($\beta = .50$, $p <.001$); Environmental Sensitivity ($\beta = .28$, $p <.001$); and Environmental Feeling ($\beta = .07$, $p <.05$).

In summary, MSELS scale scores explained 54% and 52% of the variance in Actual Commitment scores for the Phase One sixth- and eighth grade samples, respectively. Verbal Commitment and Environmental Sensitivity were found to be significant predictors of Actual Commitment for both grades in Phase One. In addition, Environmental Feeling was found to be a significant predictor for grade eight.

Table 3. *Summary of Results for MSELS Scale Scores as Predictors of Actual Commitment (Behavior) for the Phase One Eighth-Grade Sample*

	R^2	*B*	Confidence Intervals	*SE B*	*ß*	*p*
Model	0.52					
(Constant)		-0.13	-2.87, 2.58	1.44		.93
Ecological Knowledge		-0.13	-0.31, 0.05	0.09	-.05	.14
Verbal Commitment		0.53	0.46, 0.61	0.04	.51*	<.001
Environmental Sensitivity		0.36	0.28, 0.44	0.04	.28*	<.001
Environmental Feeling		0.31	0.02, 0.61	0.15	.07*	<.05
Issue Identification		0.52	-0.03, 1.08	0.28	.05	.06
Issue Analysis		0.18	-0.10, 0.45	0.14	.04	.20
Action Planning		0.03	-0.05, 0.12	0.04	.02	.43

Note: * Statistically significant at p < .05 ** Statistically significant at p < .01

Results for Specific Research Question 1b

These results include the relative contribution of scores on MSELS measures of students' knowledge, affect, and skills as predictors of their Actual Commitment (behavior) scores for Phase Two samples.

For the **Phase Two sixth-grade sample** (Table 4), 53.9% of the variance in Actual Commitment scores was explained by MSELS scale scores ($F=376.35$, $p < 0.001$). Four scales were found to be statistically significant predictors of Actual Commitment scores: all three affective scales, namely Verbal Commitment ($\beta = .44$, $p <.001$), Environmental Sensitivity ($\beta = .33$, $p <.001$), and Environmental Feeling ($\beta = .11$, $p <.001$); and one skill scale, namely Issue Analysis ($\beta = .08$, $p <.001$).

Table 4. *Summary of Results for MSELS Scale Scores as Predictors of Actual Commitment (Behavior) for the Phase Two Sixth-Grade Sample*

	R^2	*B*	Confidence Intervals	*SE B*	*ß*	*p*
Model	0.54					
(Constant)		0.54	-1.19, 2.23	0.87		.53
Ecological Knowledge		0.07	-0.02, 0.17	0.05	.03	.12
Verbal Commitment		0.46	0.42, 0.50	0.02	.44*	<.001
Environmental Sensitivity		0.40	0.36, 0.43	0.02	.33*	<.001
Environmental Feeling		0.52	0.36, 0.69	0.08	.11*	<.001
Issue Identification		0.17	-0.12, 0.46	0.15	.02	.25
Issue Analysis		0.38	0.23, 0.53	0.08	.08*	<.001
Action Planning		-0.03	-0.08, 0.02	0.03	-.02	.27

Note: * Statistically significant at p < .05 ** Statistically significant at p < .01

For the **Phase Two seventh-grade sample** (Table 5), 55.7% of the variance in Actual Commitment scores was explained by MSELS scale scores (F=367.42, p < 0.001). Four scales were found to be statistically significant predictors of Actual Commitment scores: all three affective scales, Verbal Commitment (β = .54, p <.001), Environmental Sensitivity (β = .28, p <.001), and Environmental Feeling (β = .06, p <.001); and one skill scale: Issue Analysis (β = .04, p <.05).

Table 5. *Summary of Results for MSELS Scale Scores as Predictors of Actual Commitment (Behavior) for the Phase Two Seventh-Grade Sample*

	R^2	*B*	Confidence Intervals	*SE B*	*ß*	*p*
Model	0.56					
(Constant)		0.44	-1.30, 2.15	0.92		.63
Ecological Knowledge		0.02	-0.08, 0.12	0.05	.01	.71
Verbal Commitment		0.54	0.50, 0.58	0.02	.54*	<.001
Environmental Sensitivity		0.35	0.31, 0.40	0.02	.28*	<.001
Environmental Feeling		0.28	0.11, 0.44	0.09	.06*	<.001
Issue Identification		0.27	-0.03, 0.56	0.15	.03	.07
Issue Analysis		0.19	0.04, 0.34	0.08	.04*	<.05
Action Planning		0.00	-0.06, 0.05	-.03	.00	.88

Note: * Statistically significant at p < .05 ** Statistically significant at p < .01

For the **Phase Two eighth-grade sample** (Table 6), 53.9% of the variance in Actual Commitment scores was explained by MSELS scale scores (F=216.26, $p < 0.001$). Four scales were found to be statistically significant predictors of Actual Commitment scores: all three affective scales, Verbal Commitment ($\beta = .52$, $p < .001$), Environmental Sensitivity ($\beta = .27$, $p < .001$), and Environmental Feeling ($\beta = .05$, $p < .05$); and one skill scale: Issue Identification ($\beta = .05$, $p < .05$).

In summary, MSELS scale scores explained 54%, 56%, and 54% of the variance in Actual Commitment scores for the Phase Two sixth-, seventh-, and eighth grade samples, respectively. All three MSELS affective scales, Verbal Commitment, Environmental Sensitivity, and Environmental Feeling, were found to be significant predictors across these three grades in Phase Two. In addition, an MSELS skill scale was found to be a significant predictor in each of the three grades in Phase Two, although the scale varied somewhat by grade (i.e., grades six and seven: Issue Analysis; and grade eight: Issue Identification).

Table 6. *Summary of Results for MSELS Scale Scores as Predictors of Actual Commitment (Behavior) for the Phase Two Eighth-Grade Sample*

	R^2	*B*	Confidence Intervals	*SE B*	*ß*	*p*
Model	0.54					
(Constant)		0.91	-1.22, 3.05	1.08		.41
Ecological Knowledge		0.09	-0.03, 0.20	0.06	.03	.13
Verbal Commitment		0.53	0.48, 0.58	0.03	.52*	<.001
Environmental Sensitivity		0.34	0.29, 0.39	0.03	.27*	<.001
Environmental Feeling		0.21	0.01, 0.42	0.20	.05*	<.05
Issue Identification		0.43	0.04, 0.82	0.20	.05*	<.05
Issue Analysis		0.15	-0.05, 0.34	0.10	.03	.15
Action Planning		0.02	-0.05, 0.09	0.03	.01	.61

Note: * Statistically significant at $p < .05$ ** Statistically significant at $p < .01$

Results for Specific Research Question 1c

These results include a descriptive comparison of results for the sixth-grade samples in Phase One and Phase Two, and for the eighth-grade samples in Phase One and Phase Two. Table 7 presents a summary of those results.

Variances within the Phase One sample ranged from 52% to 54%.; variances within the Phase Two sample ranged from 54% to 56%. The results for Phase One and Phase Two were similar in that two affective variables, Verbal Commitment and Environmental Sensitivity were predictors of Actual Commitment scores for all grade in both phases. Again, for all grades, Verbal Commitment was the strongest predictor, and Environmental Sensitivity was the second strongest predictor. In addition, a third affective variable, Environmental Feeling, was a predictor of Actual Commitment scores for the Eight-grade sample in Phase One and for all grades in Phase Two.

Table 7. *Explained Variance and Significant Predictors of Actual Commitment (REB) for Phase One and Phase Two, 6th, 7th, and 8th Grade Student Data.*

	Phase One 6th Grade		**Phase One 8th Grade**		**Phase Two 6th Grade**		**Phase Two 7th Grade**		**Phase Two 8th Grade**	
Model R^2	0.54		0.52		0.54		0.56		0.54	
	β	ρ	β	ρ	β	ρ	β	ρ	β	ρ
Ecological Knowledge	.05	.11	-.05	.14	.03	.12	.01	.71	.03	.13
Verbal Commitment	**.51***	**<.001**	**.51***	**<.001**	**.44***	**<.001**	**.54***	**<.001**	**.52***	**<.001**
Environmental Sensitivity	**.33***	**<.001**	**.28***	**<.001**	**.33***	**<.001**	**.28***	**<.001**	**.27***	**<.001**
Environmental Feeling	-.02	.50	**.07***	**<.05**	**.11***	**<.001**	**.06***	**<.001**	**.05***	**<.05**
Issue Identification	.01	.79	.05	.06	.02	.25	.03	.07	**.05***	**<.05**
Issue Analysis	.03	.29	.04	.20	**.08***	**<.001**	**.04***	**<.05**	.03	.15
Action Planning	.00	.87	.02	.43	-.02	.27	.00	.88	.01	.61

Note: *Statistically significant at $p < .05$ or $p < .001$

Variances within the Phase One sample ranged from 52% to 54%.; variances within the Phase Two sample ranged from 54% to 56%. The results for Phase One and Phase Two were <u>similar</u> in that two affective variables, Verbal Commitment and Environmental Sensitivity were predictors of Actual Commitment scores for all grade in both phases. Again, for all grades, Verbal Commitment was the strongest predictor, and Environmental Sensitivity was the second strongest predictor. In addition, a third affective variable, Environmental Feeling, was a predictor of Actual Commitment scores for the Eight-grade sample in Phase One and for all grades in Phase Two.

There also was a noteworthy <u>difference</u> between Phase One and Phase Two samples. In the Phase Two samples, skill scores were found to be a significant predictor of Actual Commitment scores, while this was not found in the Phase One samples. In specific, Issue Analysis scores were found to be a significant predictor for the Phase Two sixth- and seventh-grade samples, and Issue Identification scores were found to be a significant predictor for the Phase Two eighth-grade sample.

Findings, Discussion, and Recommendations - Research Question One

Research Question One: What is the relative contribution of knowledge, affect, and skill variables to actual commitment or behavior within the Phase One sample and within the Phase Two sample?

Findings

1. The affective characteristics of verbal commitment and environmental sensitivity appear to be significant predictors of actual commitment (or environmental behavior) for both sixth- and eighth-grade students from a representative sample of middle schools in the U.S (the representative group). An additional affective characteristic, environmental feeling, was a significant predictor for the representative eighth-grade students.

2. Verbal commitment, environmental sensitivity, and environmental feeling all appear to be significant predictors of actual commitment (or environmental behavior) for six-, seventh-, and eighth-grade students in schools with environmental education programs (the environmental education group). In addition, cognitive skills emerged as important predictors in this group, with issue analysis skills a significant predictor for grades six and seven, and issue identification skill a significant predictor for grade eight.

3. Students in the representative group and students in the environmental education group were similar in that verbal commitment was the strongest predictor, and that environmental sensitivity was the second strongest predictor of actual commitment or environmental behavior for all grades. Environmental feeling was the third strongest predictor for the eighth-grade sample in the representative group and for all grades in the environmental education group.

4. There was a noteworthy difference between the two groups of students regarding cognitive skills as significant predictors. Cognitive skills appear to be significant predictors of environmental behavior within the environmental education group (issue analysis skills for

grades six and seven, and issue identification skills for grade eight), but were not present as significant predictors within the representative group.

Discussion

For the most part, affective components (i.e., Verbal Commitment, Environmental Sensitivity, and Environmental Feeling) appear to be significant predictors of environmental literacy and merit attention in curriculum development and educational programming. The predictive relationship between verbal commitment (a measure of intention) and actual commitment or environmental behavior are consistent with findings reported in reviews of research (Hines et al., 1986/87, Table 1, p. 3; Bamberg & Moser, 2007, Table 3, p. 20, and Table 4, p. 22), and in theory and research regarding the intention – behavior relationship (e.g., the *Theory of Reasoned Action* and *Theory of Planned Behavior* in Fishbein & Ajzen, 1975; Ajzen, 1985). The predictive relationship between environmental sensitivity and environmental behavior are consistent with prior studies in which this variable was found to be a predictor of behavior in selected adult populations (Marcinkowski, 1989, 2005; Sia, 1985/86; Sivek & Hungerford, 1989/90). One might hypothesize that the relationship between environmental sensitivity and environmental behavior may well operate within young people as it does within adults.

The emergence of cognitive skills as significant predictors of environmental behavior among students in schools with environmental education programming also merits considerable attention. This cognitive skill dimension garnered the lowest scores for both students in the representative sample of U.S. middle schools and for students in schools with environmental education programs, with performance in this dimension falling well below both knowledge and affect (McBeth, et al, 2008; McBeth, et al, 2011). Additionally, there appeared to be no significant differences in cognitive skill scores between these two groups. However, two aspects of cognitive skills surfaced as significant predictors of environmental behavior in the latter sample of students (issue analysis for sixth- and seventh-grade students and issue identification for eighth-grade students). This might imply that one or more facets of environmental education might promote a positive relationship between cognition and behavior.

It also is noteworthy that ecological knowledge was not a significant predictor of environmental behavior for students from the representative sample of schools or students from schools with environmental education programs. While this does not comment on the importance of ecological knowledge, it may suggest that middle-grade students receive comparable educational experiences in ecology or environmental science, regardless of the nature of environmental education emphasis in their schools.

Recommendations

The findings related to Research Question One leave us with more questions than answers. While the relationship between affect and behavior appears similar in middle school students engaged in school-based environmental education programs and other middle school students, there are striking differences between those two groups regarding the relationship between cognitive skills and environmental behavior. Questions such as the following might shed light on many aspects of environmental literacy and environmental behavior:

1. How is the development of environmental sensitivity, and other important affective dimension of environmental literacy related to students' educational experiences and how is it related to environmental influences outside the school, such as home and community?
2. To what extent, if any, do perceptions of administrators and community members toward the environment and environmental education influence environmental literacy within the school?
3. What educational practices lead to a stronger relationship between aspects of environmental affect and environmental behavior?
4. What educational practices lead to a stronger relationship between cognitive skills and environmental behavior?
5. Why do cognitive skills appear to be inconsistent as predictors of environmental literacy?
6. What kinds of interactions, if any, appear to exist among environmental knowledge, affect, and skill variables as they relate to environmental behavior?
7. What kind of relationships, if any, appear to exist among environmental knowledge, affect, skill, and behavior variables as components of environmental literacy?

Research Question Two

This research question was designed to explore whether there would be any difference in environmental literacy composite scores among Phase Two schools that offered an environmental program to students in only one grade, in two consecutive grades, or in three consecutive grades. The grade-level sequences of interest are identified in the specific research questions, below. Detailed descriptions and explanations of methods and results for this research question can be found in Appendix B.

Rationale

One might assume that environmental literacy among students would be heightened by involvement in environmental education programming across multiple consecutive grade levels. Despite this assumption, the impact of sequenced programming has not been an emphasis of research in environmental education, as it relates to the development of environmental literacy. In the absence of longitudinal data, this research question emulates a cross-sectional study, using data from students in schools with environmental education programs at one grade level, two grade levels, and three grade levels. We planned to use these findings to hypothesize about the impact of progressive exposure to environmental education**.**

Specific Research Questions and Variables

For Specific Research Question 2a, we explored whether there were differences environmental literacy composite and component scores for Phase Two seventh-grade students in schools that offer environmental education programs in one grade (Grade Seven only), as compared to environmental literacy composite and component scores for Phase Two seventh grade students in schools with environmental education programs in two consecutive grades (Grades Six and Seven). In a similar manner, for Specific Research Question 2b, we explored whether there were differences in environmental literacy composite and component scores for Phase Two eighth-grade students in schools that offer environmental education programs in one grade (Grade Eight only), as compared to environmental literacy composite and component scores for Phase Two eighth grade students in schools with environmental education programs in two consecutive grades (Grades Seven and Eight), and as compared to environmental literacy composite and component scores for Phase Two eighth grade students in schools with environmental education programs in three consecutive grades (Grades Six, Seven and Eight),

Overview of Methods

The research team followed a four-step procedure to prepare for and conduct the analyses for these two specific research questions. The first step involved identifying schools that fell into each of the grade-related subsamples in these two research questions. Table 8 identifies the number of schools in each of these subsamples, as well as the total number of students in each subsample.

The second step involved the preparation of a data file for each Phase Two subsample. We found that these data files did not meet the necessary assumptions recommended by Cohen et al. (2003), Tabachnik and Fidell (2013), and others to obtain valid and accurate univariate analyses of variance (ANOVA) and multivariate analyses (MANOVA) results. As a result, we used resampling methods to guide the removal of outliers; outliers removed were less than 5% of the data set for each specific research question.

Table 8. *Number of Phase Two Schools and Students in Each Subsample in Research Questions 2a and 2b*

Research Question	Grade-Related Subsamples	N of Schools	N of Students
RQ 2a	Grade Seven only	9	1,194
	Grades Six and Seven	5	417
RQ 2b	Grade Eight only	6	935
	Grades Seven and Eight	6	143
	Grades Six, Seven and Eight	20	764

In the third step, we conducted several types of robust statistical analyses (Wilcox, 2013). We used robust ANOVA tests to determine if there was any difference in *composite score* means. Further, we used robust MANOVA tests to determine if there was an overall difference across the *four component score* means. Finally, when we found an overall difference in component scores means using MANOVA, we used robust ANOVA tests to determine if there was a difference in *each component score* mean.

In the fourth step, post-hoc analyses were undertaken only when the results of the above robust ANOVAs indicated there was a statistically significant difference between or among *composite score* means. When such a difference was found, we used robust methods to remove 5% of the outliers in each subsample. For these modified subsamples, we calculated new *composite score* means, and used resampling techniques (bootstrapping) to run robust t-tests to compare these modified means (Wilcox, 2013).

Results for Specific Research Question 2a

These results indicate whether there were differences in composite and component scores between seventh-grade students attending schools with environmental programming for Grade Seven only and seventh-grade students attending schools with environmental programming for both Grades Six and Seven.

Table 9 summarizes the results of these analyses. The ANOVA results indicated there was not a statistically significant difference in environmental literacy *composite scores* between seventh-grade students attending schools with environmental programming only for Grade Seven and seventh-grade students attending schools with environmental programming for Grades Six and Seven ($Ft = 2.7986$, $p = 0.0997$). Further, the MANOVA results indicate there was not a statistically significant difference in environmental literacy *component scores* for these two subsamples ($Ft = 7.181$, $p = 0.1266$).

Due to the exploratory nature of this study, we conducted robust ANOVA tests for each environmental literacy *component score* even though there was not a statistically significant difference in these MANOVA results. Based on these ANOVA tests, we did not find statistically significant differences in *knowledge* or *skill component scores*, but did find statistically significant differences for *affect component scores* ($Ft = 5.361$, $p < .05$), and for *behavior component scores* ($Ft = 4.191$, $p < .05$) for these subsamples.

Results for Specific Research Question 2b

These results indicate whether there were differences in composite and component scores among eighth-grade students attending schools with environmental programming only for Grade Eight,

Table 9. *Major Results of Robust ANOVA and MANOVA Analyses for Research Question 2a*

Scores Compared	Test Used	F_t value	p value
Composite Scores	ANOVA	2.7986	0.0997
All Component Scores	MANOVA	7.1810	0.1266
Each Component Score	ANOVA		
Knowledge		1.007	0.3133
Skills		0.079	0.7806
Affect		5.361	0.0224*
Behavior		4.192	0.0441*

Note: * = statistically significant at p < .05 ** = statistically significant at p < .01
*** = statistically significant at p < .001

eighth-grade students attending schools with environmental programming for Grades Seven and Eight, and eighth-grade students attending schools with environmental programming for Grades Six, Seven and Eight.

In Table 10, the ANOVA results indicate there was a statistically significant difference in environmental literacy *composite scores* among eighth-grade students who attended schools with environmental programming in Grade Eight only, in Grades Seven and Eight, and in Grades Six, Seven and Eight (Ft = 24.624, p < 0.0001). Further, MANOVA results indicate there also was a statistically significant difference in *component scores* for these three sub-samples (Ft = 58.752, p < 0.0001).

These MANOVA results indicated that it was acceptable to undertake an ANOVA test for each environmental literacy component score. When these individual ANOVA tests were run, we found statistically significant differences in subsample scores for all four components: *knowledge* (Ft = 7.495, p < 0.001); *skill* (Ft = 13,595, p < .0001); *affect* (Ft = 8.815, p < .0002); and *behavior* (Ft = 6.693, p < .001).

In light of these differences, post-hoc tests were conducted. The modified *composite score* means for these subsamples were:

Grade Eight, x = 140.15

Grades Seven and Eight, x = 150.99

Grades Six, Seven and Eight, x = 148.56.

Table 10. *Major Results of Robust ANOVA and MANOVA Analyses for Research Question 2b*

Scores Compared	Test Used	F_t value	p value
Composite Scores	ANOVA	24.624	< 0.0001***
All Component Scores	MANOVA	58.752	< 0.0001***
Each Component Score	ANOVA		
Knowledge		7.495	0.0007***
Skills		13.595	0.0002***
Affect		8.815	< 0.0001***
Behavior		6.693	0.0015**

Note: * = statistically significant at p < .05 ** = statistically significant at p < .01
*** = statistically significant at p < .001

Results of robust t-test comparisons of these composite score means were:

For Grade Eight only vs. Grades Seven and Eight, t = 12.95 (p < .001);

For Grade Eight only vs. Grades Six, Seven and Eight, t = 25.05 (p < .0001);

For Grades Seven and Eight vs. Grades Six, Seven and Eight, t = 0.57 (p < 0.45).

These results indicated that eighth-grade students in the Grades Seven and Eight subsample and in the Grades Six, Seven and Eight subsample had significantly higher composite score means than eighth-grade students in the Grade Eight only subsample. However, there was not a significant difference in composite score means between the eighth-grade students in the Grade Sevens and Eight subsample and those in the Grades Six, Seven and Eight subsample.

Findings, Discussion, and Recommendations - Research Question Two

Research Question Two: To what extent does the sequence of grade-level programming in Phase Two schools appear to have contributed to or influenced students' environmental literacy scores?

Findings.

1. There was no statistically significant difference between the composite scores of seventh-grade students in schools that offered environmental education programming only for Grade Seven and those of seventh-grade students in schools that offered environmental education programming for Grades Six and Seven. However, despite this, and due to the exploratory nature of these analyses, further analyses were run to compare the component scores of these groups. Here, significant differences were found in these groups' affective and behavior component scores, favoring students in schools with programming in Grades Six and Seven.

2. A similar comparison was made among eighth-grade students in schools with environmental education programming for Grade Eight only, for Grades Seven and Eight, and for Grades Six, Seven, and Eight. There were significant differences in composite scores between and among these three groups, favoring schools offering programs across two or three grades. Further, significant differences in knowledge, skill, and behavior component scores followed this pattern. However, this did not hold for the affective component scores.

Discussion

The questions related to grade-level sequencing are at the root questions of student experience in environmental education. We are not familiar with any environmental education research that presents evidence related to these questions, in part because longitudinal studies in environmental education require substantial time and cost, and are therefore rare.

The results for both seventh- and eighth-grade students might be explained by the likely effects of additional opportunities for experience and learning among students who have been involved in environmental education programming at multiple grade levels. However, we do not understand why significant differences in composite scores were evident among the three groups of eighth-grade students and not between the two groups of younger seventh-grade students.

Within the comparison of seventh-grade students (i.e., one grade vs. two grades of environmental education programming), affective and behavior component scores differed, but knowledge and skill component scores did not. Again, the factors that may be associated with or that may contribute to this are not well understood. Similarly, within the second comparison (i.e., one vs. two vs. three grades of environmental education programming), it is not clear why affective scores were not consistent in differentiating between groups (sub-samples) when knowledge, skill, and behavior scores did so.

Recommendations

These inconclusive results suggest questions that will require careful, in-depth data collection and analysis that might permit us to begin to shift from this type of cross-sectional analysis toward a longitudinal analysis. That is, rather than focus on groups of students from different grade levels at a single point in time, it would be more productive to focus on students, over time, as they move through several grades of school. Such questions might include the following.

1. In which grades is some form of environmental education programming offered?

2. What are the major features of the program in each grade (e.g., goals, curricula, other instructional resources and sites, institutional and network affiliations, teacher subject areas and responsibilities)?

3. What is the scope and sequence of environmental education programming across grades, if any?

4. In what ways do students exposed to environmental education programming in one grade (e.g., Grade Six) gain access to environmental education programming in subsequent grades (e.g., Grades Seven and/or Eight)? In other words, what percentage of students who are involved in environmental education in Grade Six continue to receive environmental education in Grade Seven?

5. In what ways, if any, are steps taken to ensure both fidelity of programming within grades and continuity of programming across grades?

6. What other kinds of evidence exist regarding the effects/impacts of environmental educational programming within and across grades (e.g., evidence from the community, school district, and/or state)?

Research Question Three

This research question focuses on the extent to which Student, Teacher, Program, and School characteristics appeared to influence students' environmental literacy composite scores within the Phase One and the Phase Two samples. In these analyses, all Student, Program, Teacher, and School characteristics were included in one master set, rather than analyzed as separate sets. Using this master set, we sought to determine which characteristics best explained (or predicted) the differences in environmental literacy scores of students in each Phase and grade. Detailed descriptions and explanations of methods and results for this research question can be found in Appendix C.

Rationale

Our earlier research collected data from students from a representative sample of U.S. middle schools and from students in a purposeful sample of schools with ongoing environmental education programs. Although additional data were collected at the time of those studies (school, program, and teacher information), we did not attempt to explore relationships between environmental literacy and the variables represented in those data sets. Moreover, to date, no research studies have had access to such broad and diverse data sets as those provided by the earlier NELA studies. With these data sets in hand, we surmised that it might prove beneficial to identify relationships between environmental literacy and social, demographic and contextual variables. Ultimately, we hope we that these analyses might lead to a better understanding of conditions conducive for fostering the development of environmental literacy among U.S. middle school students.

Specific Research Questions and Variables.

Initially, we explored which Student, Teacher, Program, and School variables might help to explain differences in students' environmental literacy composite scores in the Phase One sample (Specific Research Question 3a). Subsequently, we explored which Student, Teacher, Program, and School variables may help to explain differences in students' environmental literacy composite scores in the Phase Two sample (Specific Research Question 3b). Finally (in Specific Research Question 3c), we compared how well Student, Teacher, Program, and School variables helped to explain Phase One sixth-grade and eighth-grade students' environmental literacy composite scores to how well Student, Teacher, Program, and School variables helped to explain Phase Two sixth-grade, seventh-grade, and eighth-grade students' environmental literacy composite scores.

Overview of methods

The research team followed a six-step procedure to prepare for and conduct the analyses for these specific research questions. In the first step, we prepared a data file for schools in each grade within each Phase. We included all Student, Teacher, Program, and School variables in each of these data files. For each of the School, Program, Teacher, and Students items, there were as many as eight possible responses, each of which could be coded separately in these data sets. The statistical software program we used was able to treat each possible response as a separate variable, so the number of possible predictor variables was very large (e.g., number of items multiplied by the number of possible responses), even when using individual students as the unit of analysis.

Therefore, in the second step, we began the process of selecting a smaller number of variables to be included in the final analyses. By using this selection process, we hoped to identify the most promising variables in include in the analysis. In this step, the analyses conducted for each data file involved the use of analysis of variance (ANOVA) tests to identify variables as likely predictors when individual student scores were used as the dependent variable: (a) the 15 variables with the largest F value, and (b) the 15 variables with the smallest probability or p value in each grade. We used the results of these ANOVAs to reduce the large number of possible predictor variables in each data file to between 19 variables (Phase Two, Grade 8) and 23 variables (Phase Two, Grade 7). It is noteworthy that the only Student variable to meet these selection criteria was Student Age, which was included in the set of selected variables for the Phase One sixth-grade, and Phase Two sixth- and seventh-grade samples (see Appendix C).

We found that this second step was insufficient because the number of possible predictor variables (i.e., 19-23) was still too large for use in the final analyses. Therefore, in the third step, additional analyses were undertaken to further reduce the number of variables in each set. For this, we used a Neural Network approach within the Statistica software program.

Due to the requirements of multilevel analysis, the research team took a fourth step for two reasons: (a) to ensure that each final set of selected variables met the assumptions of multilevel analysis to obtain valid and accurate results [as recommended by Tabachnik and Fidell (2013) and others]; and (b) to determine which of the levels apparent in these data sets should be included in multilevel analysis for each Phase and grade (i.e., based on student ID, teacher ID, and school ID information). For example, due to the design of the Phase One study, we knew there was only one teacher per school, and that teacher delivered the same program to all participating students. Thus, in the absence of multiple teachers and programs in each school, there were only two levels of variability within the Phase One sample (i.e., students and schools). However, due to the design of the Phase Two study, there could be more than one teacher in each grade within each school. Thus, there could be two or three viable levels in Phase Two (i.e., students, teachers, and/or schools).

In this fourth step, Intraclass Correlation Coefficients (ICCs) were calculated to determine the influence of relatedness of students by schools and by teachers, both separately and combined, on student composite scores for Phase Two samples. The results of this ICC analysis of students multiplied by school multiplied by teachers indicated that schools had a larger ρ value than teachers, which supported the use of schools, but not teachers, as an additional level in the final multilevel analyses. As a further check, Design Effects were calculated to determine if it was necessary to include schools as well as students in the analyses to determine the influence of selected predictor variables on student composite scores. A resulting Design Effect value for any Phase and grade that was greater than two would indicate that it was necessary to do so. All Design Effect values ranged from a low of 4.402 (Phase Two, Grade 8) to a high of 7.644 (Phase Two, Grade 6). Thus, the results of these ICC and Design Effect analyses indicated that it was necessary to include students and schools as levels in the final multilevel analyses of selected variables.

The fifth step was to conduct multilevel analyses of the final 10 variables selected for each Phase and grade (Singer, 1998). In summary, for these analyses: (a) students and schools served as the levels of analysis in all models (i.e., using Student ID and School ID numbers); (b) the multilevel analysis of data for each Phase and grade began with the calculation of the effect of different schools on student composite scores (i.e., Model 1); and (c) one by one, the final 10 selected variables were added to the models being analyzed (i.e., Models 2 through 11). Thus, in these analyses, Model 11 included all of the final 10 variables for each sample. The detailed results of these multilevel analyses are summarized by Phase and by grade in Appendix C.

For the sixth and final step, we did not conduct any further statistical analyses. Rather, we compared the results from the analyses of Phase One data (Research Question 3a) to the results from the analyses of Phase Two data (Research Question 3b) on a descriptive basis. Thus, for Research Question 3c, we will present and highlight apparent similarities and differences in these two sets of results.

Results for Specific Research Question 3a.

These results indicate which Student, Teacher, Program, and School variables were most powerful in explaining differences among Phase One sixth-grade and eighth-grade environmental literacy composite scores.

There were statistically significant results among the results for all Models 2-11 (see Appendix C). However for ease of reporting, we are including in the body of the report a summary of only the significant results from the analysis of Model 11 for the sixth-grade and for the eighth-grade sample. We chose Model 11 because this model included all 10 of the final variables selected for each data set, and therefore serves as a common frame of reference for reporting and interpreting

Table 11 summarizes which of the 10 final variables were found to have a significant influence on student composite scores. On that table, variables that have a positive influence on student composite scores have a positive coefficient (i.e., they are related to an increase in the average composite score), while variables that have a negative influence on student composite scores have a negative coefficient (i.e., they are related to a decrease in the average composite scores).

For the Phase One sixth-grade sample, two variables had a significantly positive influence on sixth-grade student composite scores (i.e., under Program Goals, *issue investigation skills*: coefficient = 14.269; and under Years Teaching, *middle school*: coefficient = 1.230). Further, four variables were found to have a significantly negative influence on sixth-grade student composite scores: under Teacher Certification, *working on*: coefficient = - 34.230; under Years

Table 11 *Summary of Multilevel Analysis Results for the Phase One Grade 6 and Grade 8 Samples*

	Grade 6 Model 11 Coefficients[1]	**Grade 8** Model 11 Coefficients[1]
Fixed Effects		
Intercept – Students	147.614***	146.690***
% Black Students	- 0.420***	
% ESOL Students	- 0.491**	- 0.757***
EE Program Type (vs. No Env. Program):		
• Env. Curriculum		24.969*
• Env. Curriculum + Env. Club		27.684***
Program Goal (vs. Not Checked):		
• Investigation Skills	14.269***	
Instructional Method (vs. Not Checked):		
• Service Learning		- 23.440***
Teacher Age (vs. > 60):		
• 21-30		- 17.151***
• 31-40		- 8.288*
• 41-50		- 16.730***
• 51-60		- 12.615*
Years Teaching, Total	- 1.413***	
Years Teaching, Middle School	1.230**	
Teacher Certification (vs. Yes):		
• Working On	- 34.230***	
Random Effects		
Intercept – Schools	79.367**	14.724

Note: (1) * = <.05; ** = <.01; *** = <.001

Teaching, *total*: coefficient = - 1.413; and under School Composition, the *percent of black students*: coefficient = - 0.420, and *the percent of ESOL students*: coefficient = - 0.491.

For the Phase One eighth-grade sample, two variables had a significantly positive influence on eighth-grade student composite scores: under EE Program Type, having *an environmental curriculum and an environmental club*: coefficient = 24.969; and having *an environmental curriculum, an environmental club, and a residential program:* coefficient = 27.684. Of the six variables that had a significantly negative influence on eighth-grade student composite scores, four were related to Teacher Age: when teacher age was *21-30*: coefficient = - 17.151, when it was *31-40*: coefficient = - 8.288, when it was *41-50*: coefficient = - 16.730, and when it was *51-*

60: coefficient = - 12.615. To help make sense of these results, the highest average composite score for the eighth-grade sample was found for students of teachers who were *older than 60*. Thus, these negative coefficients indicated that students of teachers in these younger age groups had lower average composite scores than did students of teachers older than 60. The two additional variables that had a significantly negative influence on eighth-grade student composite scores were: under Instructional Methods, use of *service-learning*: coefficient = - 23.440; and under School Composition, the *percent of ESOL students*: coefficient = - 0.757.

When the results for the Phase One sixth-grade sample were compared to those of the eighth-grade sample, only one variable was found to have a significant influence on average composite scores in both grades; i.e., under School Composition, the *percent of ESOL students*. In both cases, this variable had a negative influence on those scores.

Results for Specific Research Question 3b

These results indicate which Student, Teacher, Program, and School variables were most powerful in explaining differences among Phase Two sixth-grade, seventh-grade, and eighth-grade environmental literacy composite scores. As in Table 11, we are including a summary of only the significant results from the analysis of Model 11 for these samples (Table 12). The results from the analyses of Models 2-10 are reported in Appendix C.

For the Phase Two sixth-grade sample, six variables had a significantly positive influence on sixth-grade student composite scores. These were:

- under EE Program Type, having *an environmental curriculum and an environmental club*: coefficient = 9.630;
- under EE Program Type, having *an environmental curriculum, an environmental club, and a residential program*: coefficient = 5.640;
- under Instructional Methods, use of *projects*: coefficient = 4.131;
- under Instructional Methods, use of *projects*: coefficient = 4.131;
- Program Duration (in weeks): coefficient = 0.258;
- under School Composition, the *percent of Asian students*: coefficient = 0.890; and
- under School Composition, the *percent of white students*: coefficient = 0.111.
-
- In addition, seven variables were found to have a significantly negative influence on sixth-grade student composite scores. These were:

Table 12. *Summary of Multilevel Analysis Results for the Phase 2 Grade 6, Grade 7, and Grade 8 Samples*

	Grade 6: Model 11 Coefficient[1]	**Grade 7:** Model 11 Coefficient[1]	**Grade 8:** Model 11 Coefficient[1]
Fixed Effects	141.570***	158.699***	195.735***
Intercept – Students			
% Asian Students	0.890**		
% White Students	0.111*		
EE Program Type (vs. Env. Curriculum):			
· Env. Curriculum + Club	9.630*		
· Env. Curriculum + Outdoor Lab.	- 9.738**		
· Env. Curriculum + Resident Program	- 6.624*		
· Env. Curric. + Club + Resident Program	5.640*		
Program Duration	0.258*		
Instructional Methods (vs. Not Checked):			
· Discussion		- 11.488***	
· Cooperative Learning		12.373***	
· Projects	4.131*		
Instructional Sites (vs. Not Checked):			
· Library/Media Center		- 11.785***	
· Science Labs	- 13.251***	- 10.800***	
Highest Degree Earned, Ed. Level			- 5.183*
Teacher Certification, Ed. Level (vs. Multiple):			
· Elementary	- 5.408**		
· Secondary	- 11.090**		
# EE Inservices, 1-2 Days		- 0.608*	
Student Age (vs. ≤ 11 yrs.):			
· 13	- 17.144***		
· 15	- 21.659***		
Random Effects			
Intercept – Schools	7.422	96.827***	0

Note: (1) * = < .05 ** = < .01 *** = < .001

- under EE Program Type, having *an environmental curriculum and outdoor lab*: coefficient = - 9.738;
- under EE Program Type, having *an environmental curriculum and a residential program:* coefficient = - 6.624;
- under Instructional Sites, use of *science labs*: coefficient = - 13.251;
- under Teacher Certification, Educational Level, *elementary*: coefficient = - 5.408;
- under Teacher Certification, Educational Level, *secondary*: coefficient = - 11.090;
- under Student Age, *13 years old*: coefficient = - 17.114; and
- under Student Age, *15 years old*: coefficient = - 21.659.

Three brief explanations may help make sense of the results above pertaining to EE Program Type, Teacher Certification, and Student Age. First, with respect to EE Program Type, several program types had a positive influence and several had a negative influence on average composite scores for this sample. All program types were compared to the average composite score for those schools that had only *an environmental curriculum* of some kind (average = 144.261). The program types that had significantly positive coefficients were found to have average scores greater than this, and those that had negative coefficients had average scores less than this.

Second, with respect to Teacher Certification, the highest average composite score for the Phase Two sixth-grade sample was found among students whose teachers held Teacher Certifications for *multiple grade levels* (average = 138.337). As a result, average composite scores were slightly lower for teachers whose only Certification was at the middle level, even lower for teachers whose only Teacher Certification was at the elementary level, and lowest for teachers whose only Teacher Certification was at the secondary level. Finally, with respect to Student Age, the highest average composite score for this sample was found for students who were 11 years old or younger (average = 144.737). Here, average composite scores were significantly lower than this for 13 and for 15 year olds (see Appendix C).

For the Phase Two seventh-grade sample, only one variable had a significantly <u>positive</u> influence on student composite scores: under Instructional Methods, the use of *cooperative learning*: coefficient = 12.373. However, four variables had a significantly <u>negative</u> influence on these student composite scores. They were:

- under Instructional Methods, the use of *discussion*: coefficient = - 11.488;
- under Instructional Sites, the use of *libraries and media centers*: coefficient = - 11.785;
- under Instructional Sites, the use of *science labs*: coefficient = - 10.800; and
- under the # of EE Inservices completed by Teachers lasting *1-2 days*: coefficient = - 0.608.

For the Phase Two eighth-grade sample, no variable was found to have a significantly <u>positive</u> influence on these student composite scores. Further, only one variable was found to have a significantly <u>negative</u> influence on their composite scores; i.e., under Highest Degree Earned, the *educational level* associate with that degree (e.g., a Bachelor’s in Elementary Education, or a Master’s in Secondary Science Education). The coefficient for this variable was – 5.183.

When the results for the Phase Two sixth-grade, seventh-grade, and eighth-grade sample were compared, only one variable appeared in the significant results for more than one of these grades. This variable fell under Instructional Sites: *science labs*. It was found to have a significantly negative influence on average composite scores for the Phase Two sixth-grade and seventh-grade samples.

Results for Specific Research Question 3c.

These results follow from a basic descriptive comparison of the results of the analyses of Phase One data (Table 11) to the results from the analyses of Phase Two data (Table 12). For this research question, we highlight the apparent similarities and differences in these two sets of results.

A careful comparison of the results in Tables 11 (Phase 1) and 12 (Phase 2) indicated that there were few similarities and few differences in these two sets of results. First, only two School variables were included among the statistically significant variables in Phase One (i.e., under School Composition, the *percent of black* and the *percent of ESOL students*) and in Phase Two (i.e., under School Composition, the *percent of Asian* and the *percent of white students*). For all four of these variables, Model 11 coefficients were less than $\pm$ 1.0. Due to the small number of variables and the magnitude of these coefficients, School variables did not have as much influence on student composite scores as did Program and Teacher variables.

Second, the only Student variable included among these statistically significant variables was Student Age, and it was found among the results for Phase Two, but not for Phase One. Third, EE Program Type was included in the results for Phase One (Table 11) and for Phase Two (Table 12). For the Phase One eighth-grade sample, two specific EE Program Types were found to have a significantly positive influence on student composite scores (i.e., having *an environmental curriculum*, and having both *an environmental curriculum and an environmental club*). This was an interesting finding because the Phase One sample was a national baseline sample generated on a stratified random basis, and only 18 of the 48 schools in this sample reported having any type of EE program. On the other hand, the Phase Two sample was a nationally purposive sample of 64 schools with a stable environmental program in the middle grades. For the Phase Two sixth-grade sample, specific EE Program Types were found to have a positive or a negative influence on student composite scores. Thus, there were differences between the Phase One and Phase Two samples, as well as in the manner in which the results pertaining to EE Program Type were generated by the statistical program for each sample (i.e., Phase One: comparison against schools with *no program*; and Phase Two: comparison against schools with only *an environmental curriculum*). Despite this, it is interesting to note that schools with both *an environmental curriculum and an environmental club* had a significantly positive influence on average student composite scores in the Phase One eighth-grade and the Phase Two sixth-grade sample.

Finally, Teacher Certification variables were included in the results for Phase One (Table 11) and for Phase Two (Table 12). For the Phase One sixth-grade sample, *working on* a teacher certification was found to have a significantly negative influence on this sample's average student composite scores. A more careful review of those data indicated that all of the teachers who were working on teacher certification were teaching in private schools. For the Phase Two sixth-grade sample, the Teacher Certification variables focused on the kind of certificate these teachers held rather than on the status of their efforts to earn a certificate. In specific, it was found that the students of Phase Two sixth-grade teachers who held only an *elementary*, a *middle*, or a *secondary* certificate had lower average composite scores than did students of teachers who held teacher certifications at two or more of these levels (i.e., *multiple*).

Findings, Discussion, and Recommendations - Research Question Three

Research Question Three: To what extent do the student, teacher, program, and school variables measured during Phases One and Two appear to have contributed to or influence students' environmental literacy composite scores?

Findings

1. Student Variables

Within the representative group (Phase One), none of the Student variables were included in the final multilevel analyses for this group.

Within the environmental education group (Phase Two), the only Student variable included in any of the final multilevel analysis for this group was Student Age, which was included in the analyses for Grade Six. When student age was 13 or 15 (i.e., older than 11 or 12), it had a significantly negative influence on these student composite scores.

2. School Variables

Within the representative group (Phase One), only two School variables were found to have a significant influence on these student composite scores. In both cases, this was a negative influence (i.e., under School Composition, the percent of Black and ESOL students). Of these, the percent of ESOL students had a significantly negative influence on student composite scores in both Grades Six and Eight.

Within the environmental education group (Phase Two), only two School variables were found to have a significant influence on these student composite scores. In both cases, this was a negative influence (i.e., under School Composition, the percent of Asian and While students). However, both of those School variables had a significantly positive influence on student composite scores, but only in Grade Six.

In summary, although School variables were included in the results of these analyses, they were relatively few in number, and fewer in the environmental education group than in the representative group. Further, even though several of these variables did have a statistical influence on student composite scores, in only one case was the coefficient for these variables in Model 11 greater than +/- 1.0. Therefore, their practical significance was noticeably lower than for Program and Teacher variables.

3. Program Variables

Within the representative group (Phase One), three Program variables were found to have a significant influence on student composite scores. In Grade Six, the selection of Issue Investigation Skills as a major program goal had a significant, positive and pervasive influence on composite scores. In Grade Eight, schools that had both an environmental curriculum, and schools that had an environmental curriculum and an environmental club, had a significant and positive, but not a pervasive influence on composite scores. In addition, for this grade, the use of Service-Learning as a teaching method had a significant and negative, but again not a pervasive influence on these scores.

Within Grade Six of the environmental education group (Phase Two), schools with different types of environmental programs also had a significant influence on composite scores. In several cases, this influence was significantly positive (i.e., for schools with an environmental curriculum and an environmental club, both with and without a residential program), and in several cases it was significantly negative (i.e., schools with an environmental curriculum plus either an outdoor lab or a residential program), although none of these program types was pervasive.

In Grade Six of the environmental group, two additional Program variables had significant and positive influences on student composite scores: Program Duration, and under Instructional Methods, the use of projects.

In Grade Seven of the environmental education group, four Program variables had a significant influence on student composite scores. Only one of these had a positive and pervasive influence on these scores: the use of cooperative learning as an instructional method. Another instructional method had a negative and pervasive influence on these scores: discussion. Also, the use of libraries and media centers as an instructional site had a negative, and pervasive influence on these scores.

In the environmental education group, only one Program variable had a significant, negative and pervasive influence on student composite scores in both Grades Six and Seven: the use of science labs as an instructional site.

3. Teacher Variables

Within the representative group (Phase One), three Teacher variables had a significant influence on student composite scores in Grade Six, but none of them did in Grade Eight. Interestingly, the number of years teaching in middle schools had positive influence on composite scores, while the total number of years teaching had a negative influence on composite scores. The third significant Teacher variable pertained to the status of teachers' certification. When teachers were working on their certification, this had a significant and negative, but not pervasive influence on composite scores (i.e., all four of these teachers were employed in private schools).

Within the environmental education group (Phase Two), only one Teacher variable had a significant influence on student composite scores in Grade Six, in Grade Seven, and in Grade Eight. In Grade Six, when the educational level associated with a teacher's certification was only elementary or only secondary, this had a negative influence on student composite scores. In Grade Seven, the number of 1-2 day EE inservices completed by a teacher also had a significant and negative influence on student composite scores. Finally, in Grade Eight, the educational level (level of schooling) associated with a teacher's highest degree earned had a negative influence on these scores.

Discussion

The findings reported here are constrained by the nature of the Phase One and Two samples and reflect major characteristics of those samples. The representative group in Phase One included only one program and one teacher for each grade (i.e., one sixth-grade and one eighth-grade class per school). From a multilevel perspective, there were only multiple students per school and multiple schools (i.e., two levels) that could be included within the analyses. The environmental education group in Phase Two included only one program per grade, although there were as many as nine (9) teachers per grade. From a multilevel perspective, there were multiple students per teacher and multiple teachers per school, as well as multiple schools. However, this number of teachers was reduced by about half due to missing teacher data.

Whether or not this missing data had an influence on the results of multilevel analysis reported here is unclear. However, what is clear is that, as in Phase One, multiple students and multiple schools had to be accounted for in the multilevel analyses. These findings are also constrained by the size of and variability within the Phase One and Two samples. On occasion, when variables were found to have a significantly positive or significantly negative influence on student composite scores, this held true for students associated with a small number of schools and/or teachers.

Recommendations

These analyses were conducted on data sets collected during two prior studies and were an attempt to explore relationships among Students, Teacher, Program, and School variables and environmental literacy scores. The findings provide direction for those interested in continuing to probe these relationships. In terms of future research, we recommend replication of these analyses, using larger and more diverse samples of schools, programs and teachers that would be sufficient for multilevel analyses of this kind. This would help to overcome the limitations associated with significant findings associated with only a few schools, programs, and/or teachers. Additional questions to pursue might include the following clusters of questions.

The only variables found to have a significant influence on student composite scores in at least one grade in Phase One and in Phase Two were the general environmental program characteristics reflected in the various EE Program Types (e.g., environmental curricula, outdoor labs, residential camp programs, school clubs).

1. Which program characteristics are apparent in each school's environmental education programs?
2. Which of these characteristics have a stable and prominent role in each program? Further, what are these roles?
3. To which grade(s) and which students are these program characteristics available?
4. What kinds of evidence exist regarding the effects/impacts of each school's environmental education program as a whole, and of each of these general program characteristics on environmental literacy and/or environmental behavior?

A number of other Program variables were found to have a positive or a negative influence on student composite scores. These included variables associated with Program Goals (i.e., issue investigation skills), Instructional Sites (i.e., science labs and libraries/media centers), and Instructional Methods (i.e., cooperative learning, projects, discussion, and service learning). The four questions presented above regarding general program characteristics are equally relevant to these specific program characteristics.

5. Which specific program characteristics receive prominent attention and/or use in each school's environmental education programs?

6. Which of these characteristics have a stable and prominent role in each program? Further, what are these roles?
7. In which grade(s) and with which students are these specific program characteristics in use?
8. What kinds of evidence exist regarding the effects/impacts of each school's environmental education program as a whole, and of each of these specific program characteristics on environmental literacy and/or environmental behavior?

In addition to these Program variables, several Teacher variables were found to have an influence on environmental literacy composite scores. These included variables associated with each teacher's professional development: their Highest Degrees Earned (i.e., the level of schooling featured in that degree program); their Teacher Certification(s) (i.e., whether they had earned or were working on this; the level of schooling reflected in each certification); their Years Teaching (i.e., in total, and in middle schools); and the number of EE inservices they had attended (i.e., those lasting 1-2 days). The results on these variables were mixed (i.e., some had a positive influence on student composite scores and others a negative influence) and spotty (i.e., only the use of science labs as an instructional site was a significant influence in more than one grade), and therefore deserve more careful and thorough attention in future research efforts. Questions that seem to follow from these results include the following.

9. Which characteristics associated with the professional development of teachers, both in general and in environmental education, are apparent in the background and/or practices of teachers in each school's environmental education programs?
10. In what ways, if any, do time-related variables appear to be related to these professional development variables (e.g., how long since the last degree or certification was earned, or how long since the last EE courses and inservice was completed)?
11. In what ways, if any, do state, district, and school policies appear to be related to or influential on these professional development variables? Further, in what ways, if any, do the perspectives of state, district and school administrators appear to be related or influential on these variables?

12. Which of these professional development variables have a visible, stable, and positive role in each program? In addition, what are those roles? Further, in what ways do these variables appear to influence each program?

13. In which grade(s) and with which students are these professional development variables most apparent and influential?

14. What kinds of evidence exist regarding the effects/impacts of these general and these environmental education-specific professional development characteristics on their students' environmental literacy and/or environmental behavior?

Research Question Four

This research question focused on the extent to which School, Program, Teacher, and Student characteristics might differentiate between two groups of Phase Two schools: schools with high environmental literacy scores (high-performing schools) and schools with low environmental literacy scores (low-performing schools). Detailed descriptions and explanations of methods and results for this research question can be found in Appendix D.

Rationale

The NELA Phase Two sample included only schools with ongoing environmental education programs. Using the school composite scores, we were able to identify high-performing schools and low-performing schools. At the time of the study, we were limited to a descriptive comparison between those two groups of schools. This Phase Three study afforded us the opportunity to conduct statistical analyses of those two groups, in order to ascertain whether there might be students, teacher, program, and school characteristics (measured in the earlier study) that distinguish between the high and low performing schools.

Specific Research Question and Variables

There was a single specific research question related to Research Question Four. In Specific Research Question 4a, we tried to identify the School, Program, Teacher, and Student characteristics that might best differentiate between Phase two schools with high environmental literacy composite scores and those with low environmental literacy composite scores. In these analyses, all School, Program, Teacher, and Student characteristics were included in one master set. Using this master set, we sought to determine which characteristics best explained (predicted) the differences in environmental literacy scores in the two groups of schools.

Overview of Methods

The research team developed and followed a four-step procedure to prepare for and conduct the analyses for this specific research question. In the first step, we identified high-performing and low-performing schools for each specific research question. To accomplish this, we used mean student composite scores to derive school composite scores and assigned schools to quartiles based on their average composite scores. For Research Question 4a, quartiles were prepared using average environmental literacy composite scores, while for Research Question 4b, quartiles were prepared using average skill component scores. In both cases, quartiling was done following procedures presented by Setek and Gallo (2009, Figure 8.43). The results of quartiling for this specific research question are summarized in Table 13.

Table 13. *Summary of Results of Quartiling of Phase Two Schools on the Basis of Composite Scores (range, 0 – 240), by Grade*

	Number of Schools		Range of Mean Composite Scores, by Quartile	
Grade	Total	# per Quartile	Quartile 4 (Q4)	Quartile 1 (Q1)
6	43	10	124.01 – 139.82	162.13 – 177.57
7	40	10	119.13 – 142.96	162.16 – 183.52
8	33	8	124.80 – 141.36	165.64 – 185.25

In the second step, we prepared data files for high-performing (top quartile, or Q1) and for low-performing (bottom quartile, or Q4) schools in each grade. While we had intended to include all school, program, teacher, and student variables in a single data file, this was not possible due to missing teacher data in Phase Two (see Appendix D). For this reason, two data files were prepared for each grade in Phase Two: (a) one file contained school, program, and student data, along with school composite scores; and (b) a second file contained teacher data, along with the school composite scores.

These data files were large because they included data from more than 50 items from the School, Program, and Teacher Information Forms, and the first section of the MSELS (see Table 1, p. x). For each of these items, there were as many as eight possible responses, each of which had to be coded separately in these data sets. The statistical software program treats each possible response as a separate variable, so the number of possible predictor variables was very large (number of items X multiplied by number of responses), far greater than the number of schools to be included in each analysis.

Therefore, in the third step, we had to select a smaller number of variables to be included in the final analyses. In order to narrow the selection of variables, we conducted two preliminary analyses for each data file, one using individual student composite scores and the other using school composite scores.

When individual student scores were used as the dependent variable, we conducted ANOVA tests to select important variables as likely predictors (i.e., the 10 variables with the largest F value, and the 10 variables with the smallest probability or p value in each grade). When school quartiles were used as the dependent variable, we conducted chi-square tests to select important variables as likely predictors (i.e., the 10 variables with the largest χ^2 value, and the 10 variables with the smallest p value in each grade). Thus, we considered both magnitude (size of F values and size of χ^2 values) and significance (p value associated with F value and χ^2 values) in selecting variables as likely predictors. We used the results of these two analyses to substantially reduce the large number of possible predictor variables in each data file. It is noteworthy that none of the Student variables met these selection criteria for this research question. In other words, the characteristics of age, grade, gender, and ethnicity do not appear to be likely predictor variables when considering student composite scores or when considering school composite scores.

The fourth step was to conduct Partial Least Squares – Discriminant Analysis (PLS-DA) for each of these 12 narrowed data sets (Vinzi et al., 2010). The purpose of PLS-DA is to identify the variables in each data set that are most powerful in discriminating between high- and low-performing schools. As in the preliminary analyses, significance and magnitude were used to identify which variables in each data set did so. For significance, *Variables Important to Projection* (VIP) Eigenvalues were used to estimate the importance of each variable in predicting whether schools would be classified as high or low performing. For magnitude, variable coefficients in each discrimination model were used to estimate the size of the effect of each variable in predicting whether schools would be classified as high or low performing (Appendix D). Only the results of this fourth step (i.e., the most conclusive step) will be summarized and discussed here.

Results for Specific Research Question 4a

One of the noticeable results of the preliminary analyses for Grades Six, Seven, and Eight was that none of the Student variables were found to discriminate between high- and low-performing schools. Therefore, Table 14 presents the final results of the PLS-DA analyses for Program and Teacher variables which were subjected to the Partial Least Squares - Discriminant Analysis and which were significant. For ease of presentation, the variables are grouped by type (program or teacher), and only the Discriminant Analysis Coefficients are included. The positive coefficients are indicators of high-performing schools (Quartile One) and the negative coefficients are indicators for low-performing schools (Quartile Four). These results indicate which Program, and Teacher variables were most powerful in explaining the difference between schools with high and low environmental literacy composite scores.

Phase Two Sixth-Grade Sample

For the Phase Two sixth-grade sample, four Program variables and nine Teacher Variables had VIP Eigenvalues greater than 1.0 and discrimination model coefficients greater than 0.10.

The variables with positive model coefficients were (in order of magnitude):

- Teacher Certification - Subject, *social studies/history* (coefficient = + 0.313);
- Highest Degree Earned – Subject, *social studies/history* (coefficient = + 0.292);
- Level of participation in environmental protection, *slightly* (coefficient = + 0.245);
- Organization of Teachers, teaching in *self-contained* classrooms (coefficient = + 0.211);
- Teacher Certification, *no* (coefficient = + 0.186);
- Teacher Age, *41-50* (coefficient = + 0.159);
- Highest Degree Earned, *Master's + 30* (coefficient = + 0.143); and
  Assessment Methods, ranking *alternative assessment* second (coefficient = + 0.102).

These were the best Program and Teacher indicators of high-performing schools in the Phase Two sixth-grade sample.

The variables with negative model coefficients were (in order of magnitude):

- Teacher Certification, *yes* (coefficient = - 0.205);
- Instructional Settings, use of science labs (coefficient = - 0.171);
- Level of participation in environmental protection, *extremely* (coefficient = - 0.149);
- Instructional Settings, use of school grounds (coefficient = - 0.118); and
- Highest Degree Earned – Subject, *science* (coefficient = - 0.102).

These were the best Program and Teacher indicators of low-performing schools in the Phase Two sixth-grade sample.

Table 14. *Discriminant Analysis Coefficients of Program and Teacher Variables That Differentiate Phase Two Quartile One Schools from Quartile Four Schools, by Grade**

Program and Teacher Variables	**Discriminant Analysis Coefficients*** Grade 6 n = 20	Grade 7 n = 20	Grade 8 n = 16
PROGRAM VARIABLES			
Organization of teachers			
Self-contained teaching	**+0.211**		
Cross-disciplinary team-teaching			**+0.120**
Other		**+0.247**	
Instructional Organization, Ranked			
Groups/teams Rank 1		**+0.135**	
Teaching/learning settings			
Science labs	**-0.171**		
School grounds	**-0.118**		
Assessment, Ranked			
Alternative assessment, Rank 2	**+0.102**		
Alternative assessment Rank 3		**-0.107**	
Informal assessment Rank 3		**+0.290**	
TEACHER VARIABLES			
Teacher Age			
21-30			**+0.162**
31-40		**+0.116**	**+0.106**
51-60	**+0.159**	**-0.189**	
>60		**+0.247**	**-0.200**
Education – Highest degree earned			
Bachelors			**-0.125**
Masters			**+0.125**
Masters + 30	**+0.143**		
Highest degree, Educational level			
Multiple			**-0.159**
Education – Highest degree earned, Subject			
Science	**-0.102**		
Social studies/History	**+0.292**		
Multiple		**+0.456**	
Teacher Certification			
Has certificate	**-0.205**		
No certificate	**+0.186**		
Teacher certification – Educational Level			
Multiple levels		**+0.117**	
Teacher certification – Subject			
Social studies/History	**+0.313**	**+0.358**	
English/Language Arts/Reading		**+0.295**	
Length of EE Inservices			
3-7 days		**+0.101**	
Level of participation in environmental protection			
Slight	**+0.245**	**+0.502**	
Moderate			**+0.142**
Extreme	**-0.149**		

*** NOTE: Positive coefficients are indicators for high-performing schools (Quartile One); Negative coefficients are indicators for low-performing schools (Quartile Four).**

Phase Two Seventh-Grade Sample

For the Phase Two seventh-grade sample, four Program variables and nine Teacher variables had VIP Eigenvalues greater than 1.0 and discrimination model coefficients greater than 0.10.

The variables with positive model coefficients were (in order of magnitude):

- Level of participation in environmental protection, *slightly* (coefficient = + 0.502);
- Highest Degree Earned – Subject, *multiple* (coefficient = + 0.456);
- Teacher Certification - Subject, *social studies/history* (coefficient = + 0.358);
- Teacher Certification - Subject, *English/Language Arts/Reading* (coefficient = + 0.295);
- Assessment Methods, *informal assessment* ranked third (coefficient = + 0.290);
- Teacher Age, > *60* (coefficient = + 0.247);
- Organization of Teachers, the selection of *other* as a response (coefficient = + 0.247);
- Instructional Groups, *groups/teams* ranked first (coefficient = + 0.135);
- Teacher Certification – Education Level, *multiple* (coefficient = + 0.117);
- Teacher Age, *31-40* (coefficient = + 0.116); and
- Number of EE Inservices, 3-7 days (coefficient = + 0.101).

These were the three best Program and Teacher indicators of high-performing schools in the Phase Two seventh-grade sample.

Phase Two Eighth-grade Sample

For the Phase Two eighth-grade sample, one Program variable and seven Teacher variables had VIP Eigenvalues greater than 1.0 and discrimination model coefficients greater than 0.10.

The variables with positive model coefficients were (in order of magnitude):

- Teacher Age, *21-30* (coefficient = + 0.162);
- Level of Participation in environmental protection, *moderately* (coefficient = + 0.142);
- Highest Degree Earned, *Master's* (coefficient = + 0.125);
- Organization of Teachers, *teaming* (coefficient = + 0.120); and
- Teacher Age, *31-40* (coefficient = + 0.106).

These were the best Program and Teacher indicators of high-performing schools in the Phase Two eighth-grade sample.

The variables with negative model coefficients were (in order of magnitude):

- Teacher Age, >*60* (coefficient = - 0.200);
- Highest Degree Earned – Education Level, *multiple* (coefficient = - 0.159); and
- Highest Degree Earned, *Bachelor's* (coefficient = - 0.125).

These were all Teacher indicators and were the best indicators of low-performing schools in the Phase Two eighth-grade sample.

Findings, Discussion, and Recommendations - Research Question Four

Research Question Four: To what extent do the student, teacher, program, and school variables measured during Phase Two appear to differentiate between high and low performing schools as determined from students' environmental literacy scores?

Findings

This question sought to determine the student, teacher, program, or school characteristics that could differentiate between high-performing schools and low-performing schools. In these results, we have searched for patterns that span multiple grade levels rather than individual grade-level findings. Therefore, even though distinct quartiles were created for each grade level, we will discuss the key characteristics that show a pattern of repetition across at least two grade levels.

1. No Student or School characteristics emerged as significant predictors of high- or low-performing schools in the sample of schools with ongoing environmental education programs.

2. Two Program Characteristics differentiated between high- and low- performing schools over two or more grade levels. High-performing schools tended to organize sixth-grade teachers into self-contained classrooms and eighth-grade teachers into cross-disciplinary teaching teams, while seventh-grade teachers were organized in some other fashion, possibly a mix of departmentalized and teaming. When offered five assessment methods, ranking alternative and informal assessment as second and third, respectively was predictive of high-performing schools.

3. Five Teacher Characteristics differentiated between high- and low- performing schools over two or more grade levels. Teacher certification in the Social Studies/History content area was predictive of high performing sixth- and seventh-grade students, and teacher certification in the English/Language Arts/Reading was predictive in the seventh grade. In terms of subject areas in

which their highest degree was earned, Social Studies/History was predictive of high-performance at the sixth grade and preparation in two or more subjects was predictive of high-performing schools at the seventh grade. Highest degrees earned in Science was predictive of low-performing (bottom quartile) schools at the sixth grade.

A Masters degree and a Masters +30 were predictive of high-performing schools at the eighth and sixth grades, respectively. Completing more then one degree was predictive of high-performance at the seventh grade, and conversely, a Bachelor's degree only was a predictor of low-performing schools at the eighth grade.

Of the self-reported teacher perceptions elicited in the data collection, only their perceived level of participation in environmental protection (not at all – slightly – moderately – considerably – extremely) emerged as predictive of high- or low-performing schools. A slight level of teacher participation in environmental protection was predictive of high-performing schools at the sixth and seventh grade and a moderate level of participation was predictive at the eighth grade. An extreme level of teacher participation in environmental protection was predictive of low-performing schools at the sixth grade.

Finally, Teacher Age was predictive of high-performing schools when 41-50 was selected at the sixth grade; 31-40 and >60 was selected at the seventh grade; and 21-30 or 31-40 were selected at the eighth grade. Teacher selected ages of 41-50 and >60 were predictive of low- performing schools at the seventh and eighth grades, respectively.

Discussion

School and Student characteristics did not show up as significant predictors of school performance on levels of environmental literacy. Of the remaining variables, more Teacher characteristics were predictive of high-levels of performance at multiple grade levels (five predictors) than Program characteristics (two predictors). One may hypothesize that in middle schools with environmental programming across consecutive grades, neither the location of the school, its socioeconomic situation nor the make-up of the student body may be as important as who is teaching the students and how they are taught. Further, it appears to be much easier to predict characteristics of high-level performance (21 predictors) than low-level performance (5 predictors). This may mean that it would be easier to identify, based on characteristics, high- rather than low-performing schools. That being said, we must add that care must be taken when comparing and/or combining variables from the analysis of school/program/student variables with the analysis of teacher variables due to sample differences (e.g., due to data loss, the sample used in the analysis of teacher variables was smaller than the sample used in the analysis of student, program, and school variables). We do not know how this loss of data may have influenced these results.

Recommendations

These analyses yielded few clear-cut results. Among the program characteristics, however, teaming appears to be a significant ingredient of high-performing schools, particularly at the eighth-grade level. This raises questions such as:

1) What is the duration (in weeks) and intensity (hour per week) of each environmental program?
2) Which subjects are represented in these teaming situations?
3) What are the major features of teaming in each school?
4) What roles and responsibilities do teachers in each subject have in planning and implementing this program? Further, are there noticeable differences in the amount of instruction in this program associated with each subject?
5) What kind of background and experience does each teacher have who participate this kind of teaming? Are there strengths associated with particular backgrounds?

A key teacher characteristic in high performing schools appears to be the self-reported perceived level of activity in environmental protection. The apparent 'ordinal' differences among teachers with slight/moderate and considerable/extreme perceived levels of participation raise the follwig questions:

6) How do teachers understand the domain of citizenship participation/environmental action and interpret their level of participation?
7) Using Hungerford and Peyton's (1980) framework, what are the reported levels of activity in the various action categories?

8) Using Stern's (2000) framework, are there apparent differences in the 'sphere' in which this reported participation occurs (public activism, public non-activist, private sphere)?

9) What kinds of issues are these modes of participation intended to target?

10) How frequently are these activities undertaken (per week, month, or year) and for how long (in years) have these teachers been active?

11) To what extent are these teachers affiliated with national, state, and/or local organizations that are in some way associated with their modes of participation? (What is the social context of the participation)?

List of References

Ajzen, I. (1985). From intentions to action: A theory of planned behavior. In J. Kuhl and J. Beckman, (Eds.), Action – Control: From Cognition to Behavior (pp. 11-139). Heidelberg, Germany: Springer.

Bamberg, S., & Moser, G. (2007). Twenty years after Hines, Hungerford and Tomera: A new meta-analysis of psycho-social determinants of pro-environmental behavior. *Journal of Environmental Psychology, 27*(2007), 14-25.

Bishop, C. (1995). *Neural networks for pattern recognition*. Oxford, UK: Oxford University Press.

Breiman, L. (1996). Bagging predictors. *Machine learning, 24*(2), 123-140.

Bryk, A., & Raudenbush, S. (1992). *Hierarchical linear models: Applications and data analysis methods*. Newbury Park, CA: Sage.

Campaign for Environmental Literacy. (2007). What is Environmental Literacy? Retrieved from http://www.fundee.org/facts/envlit/whatisenvlit.htm

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*(1), 155-159.

Cohen, J., Cohen, P., West, S., & Aiken, L. (2003). *Applied multiple regression correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.

Efron, B., & Tibshirani, R. (1994). *An introduction to the bootstrap.* Boca Raton, FL: *Chapman & Hall/CRC*

Field, A. & Miles, J. (2012). *Discovering statistics using R.* Thousand Oaks, CA: Sage Publications.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior*. Reading, MA: Addison-Wesley.

Ganesan, R., Dhanavanthan, P., Kiruthika, C., Kumarasamy, P., & Balasubramanayam, D. (2014). Comparative study of linear mixed-effects and artificial neural network models for longitudinal unbalanced growth data of Madras red sheep. *Veterinary World, 7*(2), 52-58.

Gelman, A. (2005). Analysis of variance -- Why it is more important than ever. *The Annals of Statistics, 33*(1), 1-53. Retrieved from http://www.unt.edu/rss/class/Jon/MiscDocs/Gelman_2005.pdf

Hines, J., Hungerford, H., & Tomera, A. (1986/87). Analysis and synthesis of research on responsible environmental behavior: A meta-analysis. *The Journal of Environmental Education, 18* (2), 1-8.

Hungerford, H. & Peyton, R. (1980). A paradigm for citizen responsibility: Environmental action. In A. Sacks, et al. (Eds.), *Current Issues VI: The Yearbook of Environmental Education and Environmental Studies* (pp. 146-154), Columbus OH: ERIC/SMEAC.

Hungerford, H. & Volk, T. (1990). Changing learner behavior through environmental education. *The Journal of Environmental Education. 21*(3), 8-21.

Iozzi, L. (1984). III: Environmental education research related to the affective domain. In L. Iozzi (Ed.), *A Summary of Research in Environmental Education, 1971-1982; The Second Report of the National Commission on EE Research. Monographs in Environmental Education and Environmental Studies, Volume II* (pp. 45-60). Columbus, Ohio: ERIC/SMEAC.

Kollmuss, A. & Agyeman, J. (2002). Mind the gap: Why do people act environmentally and what are the barriers to pro-environmental behavior? *Environmental Education Research, 8*(3), 239-260.

Marcinkowski, T. (1989). An analysis of correlates and predictors of responsible environmental behavior. (Doctoral dissertation, Southern Illinois University at Carbondale, 1988). *Dissertation Abstracts International: Section A, Social Sciences and Education 49* (12), 3677.

Marcinkowski, T. (2005). Predictors of responsible environmental behavior: A review of three dissertation studies. In H. Hungerford, W. Bluhm, T. Volk, and J. Ramsey (Eds.), *Essential Readings in Environmental Education* (pp. 265-294). Champaign, IL: Stipes Publishing, L.L.C.

McBeth, W., Hungerford, H., Marcinkowski, T., Volk, & Cifranick, K. (2011). *National Environmental Literacy Assessment, Phase Two: Measuring the effectiveness of North American environmental education programs with respect to the parameters of environmental literacy. Final research report*. (Report to the National Oceanic and Atmospheric Administration, and North American Association for Environmental Education under Grant #NA08SEC4690026). Carbondale, IL: CISDE. Retrieved from http://www.oesd.noaa.gov/outreach/reports/NELA_Phase_Two_Report_020711.pdf

McBeth, W., Hungerford, H., Marcinkowski, T., Volk, T., & Meyers, R. (2008). *National Environmental Literacy Assessment Project: Year 1, National baseline study of middle grades students. Final report*. (Report to the U.S. Environmental Protection Agency, National Oceanic and Atmospheric Administration, and North American Association for Environmental Education under Grant #NA06SEC4690009). Retrieved from http://www.oesd.noaa.gov/outreach/reports/Final_NELA_minus_MSELS_8-12-08.pdf

Muthén, B. & Satorra, A. (1995). Complex sample data in structural equation modeling. In P. Marsden (Ed.), *Sociological Methodology,* pp. 216-316. Washington, DC: American Sociological Association.

Rickinson, M. (2001). Special Issue: Learners and learning in environmental education:

A critical review of the evidence. *Environmental Education Research*, *7*(3), 208-320.

Setek, W., & Gallo, M. (2009). *Fundamentals of mathematics* (11th ed.). Upper Saddle River, NJ: Pearson Custom Publishing.

Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association*, *88*(422), 486-494.

Sia, A., Hungerford, H., & Tomera, A. (1985/86). Selected predictors of responsible environmental behavior. *The Journal of Environmental Education*, *17*(2), 31-40.

Singer, J. (1998). Using SAS PROC Mixed to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, *23*(4), 323-355.

Sivek, D., & Hungerford, H. (1989/90). Predictors of responsible behavior in members of three Wisconsin conservation organizations. *The Journal of Environmental Education*, *21*(2), 35-40.

StatSoft, Inc. (2013). *Electronic statistics textbook*. Tulsa, OK: StatSoft. Retrieved from http://www.statsoft.com/textbook/.

Stern, P.C. (2000). Toward a coherent theory of environmentall significant behavior. In *Journal of Social Issues*, *56(*3), 407-424.

Tabachnik, B. G., & Fidell, S. L. (2013). *Using multivariate statistics*. Boston: Pearson Education Inc.

Vinzi V., Chin, W., Henseler, J., & Wang H. (eds.) (2010). *Handbook of partial least squares: Concepts, methods and applications in marketing and related fields*. Berlin, Germany: Springer.

Wilxox, R (2012). *Introduction to robust estimation and hypothesis testing*. San Diego CA: Elsevier Publishing.

Wilcox R., & Tian T. (2011). Measuring effect size: A robust heteroscedastic approach for two or more groups. *Journal of Applied Statistics*. *38*, 1359–1368.

Zelezny, L. (1999). Educational interventions that improve environmental behaviors: A meta-analysis. *The Journal of Environmental Education*, *31*(1), 5-14.

Zhang, P. (1993). Model selection via multifold CV. *The Annals of Statistics*, *2*(1), 299-311.

Appendix A.

Detailed Description of Methods and Results for Research Question One

Appendix A.

Detailed Description of
Methods and Results for Research Question One

Research Question One focused on the relative contribution of student's *environmental knowledge*, *environmental sensitivity*, *environmental feeling*, *issue identification skills*, *issue analysis skills*, *action planning skills*, and *verbal commitment* (intention) to an explanation of their *actual commitment* (behavior) scores. There were three specific research questions in this cluster.

1a. Within the Phase One sample, what is the relative contribution of MSELS scale scores to an explanation of the variance in student MSELS actual commitment (behavior) scores for sixth-grade students and for eighth-grade students

1b. Within the Phase Two sample, what is the relative contribution of MSELS scale scores to an explanation of the variance in student MSELS actual commitment (behavior) scores for sixth-grade students, for seventh-grade students, and for eighth-grade students?

1c. What are the similarities in and differences in the relative contribution of MSELS scale scores to an explanation of the variance in student MSELS actual commitment (behavior) scores for sixth-grade, seventh-grade, and eighth-grade students in the Phase One and Phase Two samples?

Methods for Research Questions 1a, 1b, and 1c

For research questions 1a and 1b, multiple linear regression analyses were undertaken in each grade (Phase One: Grades 6 & 8; Phase Two: Grades 6, 7, & 8). For research question 1c, the results of these regression analyses were compared descriptively. To enable these analyses, data files were prepared for each Phase and grade, for a total of 5 data files.

For Research Questions 1a and 1b, the individual student served as the unit of analysis. Each data file included MSELS scale scores for each student in each school (e.g., the Phase One Grade 6 data file included scale scores for each student in these 48 schools, each as a separate row). Further, these data files included the behavior (actual commitment) scale score for each participating students within that grade in that school as the dependent variable.

The research team encountered a problem in the use of these data files for these multiple linear regression analyses. Specifically, the research team found that these samples did not meet the necessary assumptions recommended by Cohen et al. (2003), Tabachnik and Fidell (2013), and others to obtain valid and accurate linear regression results (i.e., a ratio of the number of records in each data set to the number of predictor variables; singularity and multicolinearity among predictor variables; the normality, heteroscedasticity, and independence of residuals; and the presence and influence of outliers). While the research team explored ways to remove outliers to meet each of these assumptions, they found that the resulting data file without those outliers did not meet other assumptions. Rather than drastically reduce each data file in order to meet all assumptions, the research team used bootstrapping with replacement as a random resampling method (i.e., 10,000 analyses of samples drawn from a given data set) (Efron & Tibshirani, 1994; Field & Miles, 2012). This procedure aggregates results across analyses of these samples ("bagging"), yielding one set of robust results for parameter estimates and associated confidence intervals (Brieman, 1996). The *Split Selected Column* option in JMP (Version 10.02) was used to conduct these analyses. This allowed the team to identify the relative contribution of each MSELS scale score to the prediction of behavior scale scores. It should be noted that the primary regression assumption that bootstrapping did not address was the independence of residuals (i.e., nesting of data can occur when students are organized at multiple levels).

Results for Research Question 1a

The results of this bootstrapping approach to multiple linear regression analysis for Research Question 1a, the contribution of MSELS scale scores to the prediction of composite scores, for the **Phase One sixth-grade sample** are summarized in Table A.1.

The results of these multiple linear regression analyses indicated that two predictor variables explained 54% of the variance (R^2= .540, F (7, 784)=133.51, p < .001). In specific, *Verbal Commitment* (β = .51, p < .001) and *Environmental Sensitivity* (β = .33, p < .001) were found to be statistically significant predictors of *Actual Commitment* scores. In the presence of these two significant predictors, the other five independent variables appeared to make an indirect contribution to the prediction (explained variance) of sixth graders' *Actual Commitment* scores.

Table A.1.
Predictors of Actual Commitment (Behavior) for the Phase 1 Sixth-Grade Sample

	Adjusted R^2	B^+	$SE\ B^+$	β	*p*
Model	0.54				
(Constant)		1.02 (-1.56, 3.56)	1.31		.43
Environmental Knowledge		0.12 (-0.03, 0.29)	0.08	.05	.11
Verbal Commitment		0.52 (0.45, 0.58)	0.03	.51**	<.001
Environmental Sensitivity		0.41 (0.33, 0.49)	0.04	.33**	<.001
Environmental Feeling		-0.10 (-0.38, 0.17)	0.14	-.02	.50
Issue Identification Skills		0.07 (-0.44, 0.60)	0.26	.01	.79
Issue Analysis Skills		0.14 (-0.13, 0.42)	0.14	.03	.29
Actual Planning Skills		-0.01 (-0.08, 0.07)	0.04	.00	.87

Notes: + These *B* figures reflect a 95% Bias Corrected and Accelerated Confidence Intervals reported in parentheses. These Confidence Intervals (*B)* and Standard Errors (*SE B)* are based on 10000 bootstrap samples.
* Statistically significant at $p < .05$ ** Statistically significant at $p < .001$

The results of this bootstrapping approach to multiple linear regression analysis for Research Question 1a, the contribution of MSELS scale scores to the prediction of composite scores, for the **Phase One eighth-grade sample** are summarized in Table A.2.

The results of these multiple linear regression analyses indicated that three predictor variables explained 52.1% of the variance (R^2= .521, $F(7, 750)$= 118.81, p <. 001). In specific, *Verbal Commitment* (β = .50, p < .001), *Environmental Sensitivity* (β = .28, p < .001), and *Environmental Feeling* (β = .07, p < .05) were found to be statistically significant predictors of *Actual Commitment* scores. In the presence of these three significant predictors, the other four independent variables appeared to make an indirect contribution to the prediction (explained variance) of eighth graders' *Actual Commitment* scores. When these results for Phase One sixth- and eighth-grade students were compared,
what stood out was that *Verbal Commitment* and *Environmental Sensitivity* were found to be highly significant predictors of Actual Commitment scores.

Table A.2.
Predictors of Actual Commitment (Behavior) for the Phase 1 Eighth-Grade Sample

	Adjusted R^2	*B^+*	*SE B^+*	*β*	*p*
Model	0.52				
(Constant)		-0.13 (-2.87, 2.58)	1.44		.93
Environmental Knowledge		-0.13 (-0.31, 0.05)	0.09	-.05	.14
Verbal Commitment		0.53 (0.46, 0.61)	0.04	.50**	<.001
Environmental Sensitivity		0.36 (0.28, 0.44)	0.04	.28**	<.001
Environmental Feeling		0.31 (0.02, 0.61)	0.15	.07*	<.05
Issue Identification Skills		0.52 (-0.03, 1.08)	0.28	.05	.06
Issue Analysis Skills		0.18 (-0.10, 0.45)	0.14	.04	.20
Actual Planning Skills		0.03 (-0.05, 0.12)	0.04	.02	.43

Notes:
+ These *B* figures reflect a 95% Bias Corrected and Accelerated Confidence Intervals reported in parentheses. These Confidence Intervals (*B)* and Standard Errors (*SE B)* are based on 10000 bootstrap samples.
* Statistically significant at p < .05 ** Statistically significant at p < .001

Results for Research Question 1b

The results of this bootstrapping approach to multiple linear regression analysis for Research Question 1b, the contribution of MSELS scale scores to the prediction of composite scores, for the **Phase Two sixth-grade sample** are summarized in Table A.3.

The results of these multiple linear regression analyses indicated that four predictor variables explained 53.9% of the variance (R^2= .539, F(7, 2241) = 376.35, p < .001). In specific, *Verbal Commitment* (β = .44, p < .001), *Environmental Sensitivity* (β = .33, p < .001), *Environmental Feeling* (β = .11, p < .001), and *Issue Analysis Skills* (β = .08, p < .001) were found to be statistically significant predictors of *Actual Commitment* scores. In the presence of these four significant predictors, the other independent variables appeared to make an indirect contribution to the prediction (explained variance) of sixth graders' *Actual Commitment* scores.

Table A.3.
Predictors of Actual Commitment (Behavior) for the Phase 2 Sixth-Grade Sample

	Adjusted R^2	*B^{+}*	*$SE\ B^{+}$*	*β*	*P*
Model	.54				
(Constant)		0.54 (-1.19, 2.23)	0.87		0.53
Environmental Knowledge		0.07 (-0.02, 0.17)	0.05	.03	0.12
Verbal Commitment		0.46 (0.42, 0.50)	0.02	.44**	<0.001
Environmental Sensitivity		0.40 (0.36, 0.43)	0.02	.33**	<0.001
Environmental Feeling		0.52 (0.36, 0.69)	0.08	.11**	<0.001
Issue Identification Skills		0.17 (-0.12, 0.46)	0.15	.02	0.25
Issue Analysis Skills		0.38 (0.23, 0.53)	0.08	.08**	<0.001
Actual Planning Skills		-0.03 (-0.08, 0.02)	0.03	-.02	0.27

Notes:+ These *B* figures reflect a 95% Bias Corrected and Accelerated Confidence Intervals reported in parentheses. These Confidence Intervals (*B)* and Standard Errors (*SE B)* are based on 10000 bootstrap samples.
* Statistically significant at p < .05 ** Statistically significant at p < .001

The results of this bootstrapping approach to multiple linear regression analysis for Research Question 1b, the contribution of MSELS scale scores to the prediction of composite scores, for the **Phase Two seventh-grade sample** are summarized in Table A.4.

The results of these multiple linear regression analyses are similar to those for the seventh grade sample. They indicated that four predictor variables explained 55.7% of the variance (R^2= .557, F(7, 2029) = 367.42, p < .001). In specific, *Verbal Commitment* (β = .54, p < .001), *Environmental Sensitivity* (β = .28, p < .001), *Environmental Feeling* (β = .06, p < .001), and *Issue Analysis Skills* (β = .04, p < .05) were found to be statistically significant predictors of *Actual Commitment* scores. In the presence of these four significant predictors, the other independent variables appeared to make an indirect contribution to the prediction (explained variance) of seventh graders' *Actual Commitment* scores.

Table A.4.
Predictors of Actual Commitment (Behavior) for the Phase 2 Seventh-Grade Sample

	Adjusted R^2	B^+	$SE\ B^+$	*β*	*p*
Model	0.56				
(Constant)		0.44 (-1.30, 2.15)	0.92		.63
Environmental Knowledge		0.02 (-0.08, 0.12)	0.05	.01	.71
Verbal Commitment		0.54 (0.50, 0.58)	0.02	.54**	<.001
Environmental Sensitivity		0.35 (0.31, 0.40)	0.02	.28**	<.001
Environmental Feeling		0.28 (0.11, 0.44)	0.09	.06**	<.001
Issue Identification		0.27 (-0.03, 0.56)	0.15	.03	.07
Issue Analysis		0.19 (0.04, 0.34)	0.08	.04*	<.05
Actual Planning		0.00 (-0.06, 0.05)	0.03	.00	.88

Notes:+ These *B* figures reflect a 95% Bias Corrected and Accelerated Confidence Intervals reported in parentheses. These Confidence Intervals (*B)* and Standard Errors (*SE B)* are based on 10000 bootstrap samples.
* Statistically significant at p < .05 ** Statistically significant at p < .001

The results of this bootstrapping approach to multiple linear regression analysis for Research Question 1b, the contribution of MSELS scale scores to the prediction of composite scores, for the **Phase Two eighth-grade sample** are summarized in Table A.5.

The results of these multiple linear regression analyses are similar to those for the seventh grade sample. They indicated that four predictor variables explained 53.9% of the variance (R^2= .539, $F(7, 1281) = 216.26$, $p < .001$). In specific, *Verbal Commitment* (β = .52, p < .001), *Environmental Sensitivity* (β = .27, p < .001), *Environmental Feeling* (β = .048, p < .05), and *Issue Identification Skills* (β = .046, p < .05) and were found to be statistically significant predictors of *Actual Commitment* scores. In the presence of these four significant predictors, the other independent variables appeared to make an indirect contribution to the prediction (explained variance) of eighth graders' *Actual Commitment* scores.

Table A.5.
Predictors of Actual Commitment (Behavior) for the Phase 2 Eighth-Grade Sample

	Adjusted R^2	*B^+*	*SE B^+*	*β*	*p*
Model	0.54				
(Constant)		0.91 (-1.22, 3.05)	1.08		.41
Environmental Knowledge		0.09 (-0.03, 0.20)	0.06	.03	.13
Verbal Commitment		0.53 (0.48, 0.58)	0.03	.52**	<.001
Environmental Sensitivity		0.34 (0.29, 0.39)	0.03	.27**	<.001
Environmental Feeling		0.21 (0.01, 0.42)	0.11	.048*	.046
Issue Identification		0.43 (0.04, 0.82)	0.20	.046*	.031
Issue Analysis		0.15 (-0.05, 0.34)	0.10	.03	.15
Actual Planning		0.02 (-0.05, 0.09)	0.03	.01	.61

Notes:+ These *B* figures reflect a 95% Bias Corrected and Accelerated Confidence Intervals reported in parentheses. These Confidence Intervals (*B)* and Standard Errors (*SE B)* are based on 10000 bootstrap samples.
* Statistically significant at p < .05   ** Statistically significant at p < .001

When these results for Phase Two sixth-, seventh-, and eighth-grade students were compared, what stood out was that *Verbal Commitment* and *Environmental Sensitivity* were found to be highly significant predictors of *Actual Commitment* scores, and *Environmental Feeling* was found to be a moderately significant predictor of those scores.

Results for Research Question 1c

Research Question 1c focused on differences in the contribution of MSELS scale scores to the prediction of composite scores in these Phase One and Phase Two samples. The results of the regression analyses reported for Research Questions 1a and 1b were charted and compared descriptively. Table A.6 presents a summary of results for each grade in the Phase One and Phase Two sample.

Table A.6.

Explained Variance and Significant Predictors of Actual Commitment (REB) for Phase 1 and Phase 2 Data for 6th, 7th, and 8th Grade Students

	Phase One 6th Grade		*Phase One 8th Grade*		*Phase Two 6th Grade*		*Phase Two 7th Grade*		*Phase Two 8th Grade*	
Model R^2	0.54		0.52		0.54		0.56		0.54	
	β	ρ	β	ρ	β	ρ	β	ρ	β	ρ
Ecological Knowledge	.05	.11	-.05	.14	.03	.12	.01	.71	.03	.13
Verbal Commitment	**.51****	**<.001**	**.50****	**<.001**	**.44****	**<.001**	**.54****	**<.001**	**.52****	**<.001**
Environmental Sensitivity	**.33****	**<.001**	**.28****	**<.001**	**.33****	**<.001**	**.28****	**<.001**	**.27****	**<.001**
Environmental Feeling	-.02	.50	**.07***	**<.05**	**.11****	**<.001**	**.06****	**<.001**	**.05***	**<.05**
Issue Identification	.01	.79	.05	.06	.02	.25	.03	.07	**.05***	**<.05**
Issue Analysis	.03	.29	.04	.20	**.08****	**<.001**	**.04***	**<.05**	.03	.15
Action Planning	.00	.87	.02	.43	-.02	.27	.00	.88	.01	.61

Notes: For each variable with a statistically significant β, that β and associated p value have been bolded.
* Statistically significant at $p < .05$ ** Statistically significant at $p < .001$

For the **Phase One sixth- and eighth-grade sample**, two MSELS scale scores were found to be significant predictors of these students' Actual Commitment (behavior) scores: Verbal Commitment, and Environmental Sensitivity. The only other MSELS scale score that was found to be a significant predictor of Actual Commitment scores in either grade was Environmental Feeling (i.e., for eighth-grade only).

For the **Phase Two sixth-, seventh-, and eighth-grade sample**, Verbal Commitment and Environmental Sensitivity scale scores were the two strongest predictors of these students' Actual Commitment (behavior) scores. Further, Environmental Feeling was found to be a significant, but less prominent, predictor of these students' Actual Commitment scores for all three grades.

The results summarized in Table A.6 indicate that there were several similarities between the Phase One and Phase Two samples. In specific, Verbal Commitment was the strongest predictor (β values = .44 - .54, $p < .001$), and Environmental Sensitivity was the second strongest predictor for all grades in Phases One and Two (β values = .27 - .33, $p < .001$). The findings regarding verbal commitment, a measure of intention, are consistent with findings reported in reviews of research (Hines et al., 1986/87, Table 1, p. 3; Bamberg & Moser, 2007, Table 3, p. 20, and Table 4, p. 22), as well as other theory and research regarding the intention – behavior relationship (e.g., the *Theory of Reasoned Action* and *Theory of Planned Behavior* in Fishbein & Ajzen, 1975; Ajzen, 1985). The findings regarding environmental sensitivity are consistent with prior studies in which this variable was found to be a predictor of behavior in selected adult populations (Marcinkowski, 1989, 2005; Sia, 1985/86; Sivek, 1989/90). In addition, Environmental Feeling was found to be the third strongest predictor for all grades in Phase Two and for the Phase One eighth-grade sample (β values = .05 - .11, $p < .05$), but not for the Phase One sixth-grade sample.

The results summarized in Table A.6 also indicate that there was a noteworthy difference between the Phase One and Phase Two samples. In the Phase Two samples, skill scores were found to be a significant predictor of these students' Actual Commitment scores, while this was not found in the Phase One samples. In specific, Issue Analysis scores were found to be a significant predictor for the Phase Two sixth- and seventh-grade samples, and Issue Identification scores were found to be a significant predictor for the Phase Two eighth-grade sample (β values = .04 - .08 $p < .05$). These results are consistent with prior research findings on the effects of environmental education on the development of these and other cognitive skills, as well as on the relationship of these and other cognitive skills on behavior presented in reviews of research (Iozzi, 1984; Hines et al., 1986/87; Zelezny, 1999; Rickinson, 2001; Bamberg & Moser, 2007).

Appendix B

Detailed Description of Methods and Results for Research Question Two

Appendix B.

Detailed Description of
Methods and Results for Research Question Two

Research Question Two focused on what the research team referred to as *sequential differences*. This research question explores whether schools that offered their environmental program to students in only one, in two, or in three grades would exhibit any differences in student composite scores. The research team chose to delimit this analysis to schools in the Phase Two sample because all of these schools had some type of environmental program, while in the Phase One sample, only 18 of the 48 schools did, rendering this Phase One sample too small for analysis purposes (McBeth et al., 2008, Table 27, pp. 74-75). Further, the manner in which this will be presented and discussed is important because in Phase Two, data were collected from students who were in the sixth, seventh, <u>or</u> eighth grade in the spring of 2009 (McBeth et al., 2011). In other words, all data were collected in the same school year, and no attempt was made to track these students as they may have progressed from one year to another in their school's environmental program. For this reason, the manner in which these data were collected is apparent in the phrasing of the two research questions, below.

> **2a**. Within the Phase Two sample of schools (n=64), what differences, if any, exist between MSELS *component* and *composite scores* of seventh-grade students who attended schools with environmental programming at Grade 7 only, and seventh-grade students in schools with environmental programming at Grades 6 and 7?
>
> **2b**. Within the Phase Two sample of schools (n=64), what differences, if any, exist between MSELS *component* and *composite scores* of eighth-grade students who attended schools with environmental programming at Grade 8 only, schools with environmental programming at Grades 7 and 8 only, and schools with environmental programming at Grades 6, 7, and 8?

Methods for Specific Research Questions 2a and 2b

The first step in preparing for these analyses involved the identification of Phase Two schools that fell into each of the grade-related subsamples in these two specific research questions. Table B.1 identifies the number of schools in each of these subsamples, as well as the total number of students in each subsample. While not included in Table B.1, these counts reflect substantial differences in the number of classes and students in each Phase Two participating school.

Table B.1
Number of Phase 2 Schools and Students in Each Subsample in Research Questions 2a and 2b

Research Question	Grade-Related Subsample	No. of Schools	No. of Students
RQ 2a	Grade 7 only	9	1,194
	Grades 6 and 7 only	5	417
RQ 2b	Grade 8 only	6	935
	Grades 7 and 8 only	6	143
	Grades 6, 7, and 8	20	764

The second step involved the preparation of a data file for each Phase Two subsample. One data file was prepared for each research question and analysis. Only those schools with subsamples relevant to Research Question 2a were included in that data file. Schools with Grade 7 only were coded with a 2, and schools with Grade 6 and 7 only were coded with a 4. Because individual students served as the unit of analysis, this file included the following coding: the School ID, the ID of each student in these grades in each school, and environmental literacy component and composite scores for each student. This same procedure was used to prepare the data file for Research Question 2b, with one exception (i.e., schools with Grade 8 only were coded with a 3, schools with Grade 7 and 8 only were coded with a 6, and schools with Grades 6, 7, and 8 were coded with a 7).

In the third step, robust univariate and multivariate analyses were undertaken for each of these research questions. As occurred in analyses for Overarching Research Question 1, the research team encountered problems in the use of these data files for these multivariate analyses. Specifically, the research team found that these samples did not meet the necessary assumptions recommended by Cohen et al. (2003), Tabachnik and Fidell (2013), and others to obtain valid and accurate univariate analysis of variance (ANOVA) and multivariate analysis of variance (MANOVA) results (i.e., even n sizes, normality, heteroscedasticity, and independence of residuals; and the presence and influence of outliers). As a result, the research team chose to use the *Percentile t* bootstrapping methods (i.e., 10,000 analyses of samples drawn from a given data set) using scripts embedded in R language (Wilcox, 2013). Features of this method include the testing of the hypothesis of equal trimmed means, and when these means are unequal, the removal of outliers. In the data files for Research Questions 2a and 2b, the outliers removed constituted 5% of the data set. Using these refined data files, three sets of statistical analyses were undertaken: (a) robust ANOVA tests were used to determine if there was any difference in subsample composite score means; (b) robust MANOVAs were used to determine if there was any difference in subsample component score means; and (c) robust ANOVAs were used to determine if there was any difference in subsample means for each component score.

Results for Research Question 2a

The results of this bootstrapping approach to ANOVA and MANOVA analyses for Research Question 2a, differences in environmental literacy composite and components scores for these two subsamples, are summarized in Table B.2.

Table B.2
Major Results of Robust ANOVA and MANOVA Analyses for Research Question 2a

Scores Compared	Test Used	F_t value	p value
Composite Scores	ANOVA	2.7986	0.0997
All Component Scores	MANOVA	7.1810	0.1266
Each Component Score	ANOVA		
Knowledge		1.007	0.3133
Skills		0.079	0.7806
Affect		5.361	0.0224*
Behavior		4.192	0.0441*

Note: * = statistically significant at $p < .05$

The results in Table B.2 indicate that there was not a statistically significant difference in student MSELS composite scores between seventh-grade students who attended schools with environmental programming only in Grade 7 and seventh-grade students who attended schools with environmental programming in Grades 6 and 7 ($Ft = 2.7986$, $p = 0.0997$). Further, there was no statistically significant difference in student MSELS component scores for these two subsamples ($Ft = 7.181$, $p = 0.1266$).

In most research settings, the absence of significant MANOVA results would preclude further analysis. However, due to the seminal and exploratory nature of this study, the research team chose to conduct the individual robust ANOVA tests to determine if there may be any significant difference between the MSELS component scores for these subsamples. When these individual ANOVA tests were run, no statistically significant differences were found in either the *Knowledge* or *Skill* component score means for these subsamples. However, statistically significant differences were found for *Affect* component score means for these subsamples ($Ft = 5.361$, $p < .05$), and for *Behavior* component score means for these subsamples ($Ft = 4.191$, $p < .05$).

Results for Specific Research Question 2b

The results of this bootstrapping approach to ANOVA and MANOVA analyses for Research Question 2b, differences in environmental literacy composite and components scores for these two subsamples, are summarized in Table B.3.

Table B.3
Major Results of Robust ANOVA and MANOVA Analyses for Research Question 2b

Scores Compared	Test Used	F_t value	p value
Composite Scores	ANOVA	24.624	< 0.0001***
All Component Scores	MANOVA	58,752	< 0.0001***
Each Component Score	ANOVA		
Knowledge		7.495	0.0007***
Skills		13.595	0.0002***
Affect		8.815	< 0.0001***
Behavior		6.693	0.0015**

Note: * = statistically significant at p < .05 ** = statistically significant at p < .01
*** = statistically significant at p < .001

The results in Table B.3 indicate that there was a statistically significant difference in student MSELS composite scores between eighth-grade students who attended schools with environmental programming in Grade 8 only, in Grades 7 and 8 only, and in Grades 6, 7, and 8 (Ft = 24.624, p < 0.0001). Further, there also was a statistically significant difference in student MSELS component scores for these three sub-samples (Ft = 58.752, p < 0.0001).

Unlike MANOVA results for Research Question 2a, these MANOVA results indicated that it was acceptable to undertake ANOVA analyses for each MSELS component score. When these individual ANOVA tests were run, statistically significant differences were found in subsample mean scores for all four components: *Knowledge* (Ft = 7.495, p < 0.001); *Skill* (Ft = 13,595, p < .0001); *Affect* (Ft = 8.815, p < .0002); and *Behavior* (Ft = 6.693, p < .001).

To determine which subsample(s) were responsible for these results, it was necessary to calculate subsample *component* and *composite score* means, as well as to conduct post-hoc robust t-tests. The results of these robust post-hoc analyses are presented in Table B.4.

Table B.4
Summary of Post Hoc Analyses for Specific Research Question 2b

	Subsample Mean Scores				
Score	Gr. 8 only	Grs. 7 & 8	Grs. 6, 7 & 8	t	p *
Composite	140.15	150.99		12.950	.000***
	140.15		148.56	25.053	.000***
		150.99	148.56	0.574	.446
Components					
Knowledge	42.13	43.61		1.586	.211
	42.13		45.09	18.676	.000***
		43.61	45.09	2.203	.137
Affect	39.15	41.27		10.776	.001**
	39.15		39.14	0.002	.964
		41.27	39.14	10.558	.001**
Skill	21.91	26.80		11.896	.000***
	21.91		25.62	22.272	.000***
		26.80	25.62	0.628	.425
Behavior	37.33	39.94		10.223	.002**
	37.33		38.93	11.697	.000**
		39.94	38.93	1.686	.195

Note: The alpha levels used to determine statistical significance were:
* = p < .05; ** = p < .01; and *** = p < .001.

The results in Table B.4 indicate that there were noticeable differences in modified *composite score* means for these subsamples: (Grade 8 only) x = 140.15; (Grades 7 & 8) x = 150.99; and (Grade 6, 7, & 8) x = 148.56. Results of robust t-test comparisons revealed statistically significant differences between these *composite score* means for the Grade 8 only vs. Grades 7 and 8 subsample comparison (t = 12.95, p < .000), and for the Grade 8 vs. Grades 6, 7, and 8 subsample comparison (t = 25.05, p < .000). Although the mean *composite score* for the Grades 7 and 8 subsample was slightly larger than for the Grades 6, 7, and 8 subsample, this difference was not statistically significant (t = 0.57, p = 0.446).

The results in Table B.4 also indicate that there were noticeable differences in modified *knowledge component score* means for these subsamples: (Grade 8 only) x = 42.13; (Grades

7 & 8) x = 43.61; and (Grade 6, 7, & 8) x = 45.09. These differences were orderly in that the mean score for the Grade 6, 7, and 8 subsample was largest, and the mean score for the Grade 8 only subsample was smallest. Results of robust t-test comparisons revealed that the only statistically significant differences between modified *knowledge component score* means was for the Grade 8 only vs. Grades 6, 7 and 8 subsample comparison (t = 18.676, p < .000).

In addition, the results in Table B.4 indicate that there were differences in modified *affect component score* means for these subsamples: (Grade 8 only) x = 39.15; (Grades 7 & 8) x = 41.27; and (Grades 6, 7, & 8) x = 39.14. Unlike the results for the other three component scores, these differences were not orderly (i.e., the mean scores for the Grade 8 only subsample and the Grades 6, 7, and 8 subsample were nearly equivalent). Results of robust t-test comparisons revealed that the only statistically significant differences between these *affect component score* means were found in the Grades 7 and 8 vs. Grade 8 only subsample comparison (t = 10.77, p < .001), and in the Grades 7 and 8 vs. Grades 6, 7, and 8 subsample comparison (t = 10.558, p < .001).

Further, the results in Table B.4 indicate that there were pronounced differences in modified *skill component score* means for these subsamples: (Grade 8 only) x = 21.91; (Grades 7 & 8) x = 26.80; and (Grade 6, 7, & 8) x = 25.62. These differences were somewhat orderly in that the mean score for the Grade 8 only subsample was smallest. Results of robust t-test comparisons revealed statistically significant differences between these *skill component score* means for the Grade 8 only vs. the Grades 7 and 8 subsample comparison (t = 11.896, p < .000), and for the Grade 8 only vs. Grades 6, 7, and 8 subsample comparison (t = 22.272, p < .000). Although the *skill component score* mean for the Grades 7 and 8 subsample was slightly larger than for the Grades 6, 7, and 8 subsample, this difference was not statistically significant (t = 0.63, p = 0.425).

Finally, the results in Table B.4 indicate that there were differences in modified *behavior component score* means for these subsamples: (Grade 8 only) x = 37.33; (Grades 7 & 8) x = 39.94; and (Grade 6, 7, & 8) x = 38.93. These differences were somewhat orderly in that the mean score for the Grade 8 only subsample was smallest. Results of robust t-test comparisons revealed statistically significant differences between these *behavior component score* means for the Grade 8 only vs. the Grades 7 and 8 subsample comparison (t = 10.223, p < .002), and for the Grade 8 only vs. Grades 6, 7, and 8 subsample comparison (t = 11.697, p < .000). Again, although the *component score* mean for the Grades 7 and 8 subsample was slightly larger than for the Grades 6, 7, and 8 subsample, this difference was not statistically significant (t = 1.686, p = 0.195).

In summary, modified mean scores for the Grades 6, 7, and 8 subsample were significantly larger than for the Grade 8 only sample for *composite scores* and for *knowledge*, *skill*, and *behavior component scores*. Similarly, modified mean scores for the Grades 7 and 8 subsample were significantly larger than for the Grade 8 only sample for *composite scores* and for *affect*, *skill*, and *behavior component scores*. The only instance in which the Grade 8 only subsample scored higher than either of these other subsamples occurred on *affect component scores*.

Appendix C

Detailed Description of Methods and Results for Research Question Two

Appendix C.

More Detailed Description of Methods and Results for Overarching Research Question Three

Research Question 3 focuses on the extent to which Student, Teacher, Program, and School characteristics appeared to influence students' environmental literacy composite scores within the Phase One and the Phase Two sample. In these analyses, all Student, Program, Teacher, and School characteristics were included in one master set, rather than analyzed as separate sets. Using this master set, we sought to determine which characteristics best explained (predicted) the differences in environmental literacy scores of students in each Phase and grade. There were three specific research questions for Research Question Three.

> **3a**. To what extent do Student, Teacher, Program, and School variables measured during Phase One appear to have contributed to or influenced sixth-grade and eighth- grade students' environmental literacy composite scores?
>
> **3b**. To what extent do Student, Teacher, Program, and School variables measured during Phase Two appear to have contributed to or influenced sixth-grade, seventh-grade, and eighth-grade students' environmental literacy composite scores?
>
> **3c**. What are the similarities and differences in the extent to which Student, Teacher, Program, and School variables appeared to contribute to or influence student environmental literacy composite scores for sixth-grade, seventh-grade, and eighth-grade students in the Phase One and Phase Two samples

Methods for Research Questions 3a, 3b, and 3c

The research team followed a six-step procedure to prepare for and conduct the analyses for these specific research questions. In the <u>first</u> step, we prepared a data file for schools in each grade within each Phase. We included all Student, Teacher, Program, and School variables in each of these data files. These data files were large because they included data from more than 50 items in the School, Program, and Teacher Information Forms, the Screening Survey used in Phase 2, and the first section of the MSELS. However, because the team decided to include Teacher variables in the same file as School, Program, and Student variables, the size of the PH 2 sample for each grade was reduced substantially due to missing teacher data (Table C.1).

Table C.1
The Effect of Missing Teacher Data on the Size of the Phase 2 Sample, by Grade

	Original Sample Size			Reduced Sample Size		
Grade	Schools	Teachers	Students	Schools	Teachers	Students
6	43	112	2,849	35	57	1,791
7	40	86	2,488	30	42	1,574
8	33	83	1,678	23	32	690

For each of the School, Program, Teacher, and Students items, there were as many as 8 possible responses, each of which could be coded separately in these data sets. The statistical software program can treat each possible response as a separate variable, so the number of possible predictor variables was very large (e.g., number of items multiplied by the number of possible responses), even when using individual students as the unit of analysis.

Therefore, in the second step, we began the process of selecting a smaller number of variables to be included in the final analyses. The analyses conducted for each data file involved the use of analysis of variance (ANOVA) tests to identify variables as likely predictors when individual student scores were used as the dependent variable: (a) the 15 variables with the largest F value, and (b) the 15 variables with the smallest probability or p value in each grade. We used the results of these ANOVAs to reduce the large number of possible predictor variables in each data file to between 19 variables (Phase Two, Grade 8) and 23 variables (Phase Two, Grade 7). It is noteworthy that the only Student variable to meet these selection criteria was *Student Age*, which was included in the set of selected variables for the Phase One sixth-grade, and Phase Two sixth- and seventh-grade samples.

We found that this second step was insufficient because the number of predictor variables (i.e., 19-23) was still too large for use in the final analyses. Therefore, in the third step, additional analyses were used to further reduce the number of variables in each set. We used a Neural Network approach within the Statistica software program because it allows for "sophisticated modeling and prediction techniques capable of modeling extremely complex functions and data relationships" between independent and dependent variables (StatSoft, 2013), especially the kind of multilevel relationships to be explored in the final analyses. The specific Neural Network architecture was constructed using the Multilayer Perceptron (MLP) method. We used MLP to identify and select the 10 most sensitive or important predictors of student composite scores in each Phase and grade (Bishop, 1995; Ganesan, Dhanavanthan, Kiruthika, Kumarasamy, & Balasubramanayam, 2014). More specifically, in these Neural Network analyses, we randomly split the data set for each grade-level sample into three separate subsamples: a training sample (70%), which was used to train the network (i.e., develop 1,000 prediction models); (b) a testing sample (15%), which was used to test the performance of all of the prediction models developed by the network; and (c) a validation sample (15%), which was used to identify the five best prediction models for this 'new' sample. Due to the relatively small size of the Phase Two eighth-grade sample, we ran this sequence 10,000 times (bootstrapping) to improve the accuracy of the results for that sample. The results from the five best models for each Phase and grade were summarized and used to select these 10 variables for the final analysis.

Due to the requirements of multilevel analysis, the research team took a fourth step for two reasons: (a) to ensure that each final set of selected variables met the assumptions of multilevel analysis as recommended by Tabachnik and Fidell (2013) and others to obtain valid and accurate results; and (b) to determine which of the levels apparent in these data sets should be included in multilevel analysis for each Phase and grade (i.e., based on student ID, teacher ID, and school ID information). For example, due to the design of the Phase One study, we knew there was only one teacher per school, and that teacher delivered the same program to all participating students. Thus, in the absence of multiple teachers and programs in each school, there were only two levels of variability in each Phase One sample (i.e., students and schools). However, due to the design of the Phase Two study, there could be more than one teacher in each grade within each school. Thus, there could be two or three viable levels in Phase Two (i.e., students, teachers, and/or schools).

In this fourth step, Intraclass Correlation Coefficients (ICCs) were calculated to determine the influence of relatedness of students by schools and by teachers, both separately and combined, on student composite scores for Phase Two samples. For these samples, the results of the ICC analysis of students x school x teachers indicated that schools had a larger ρ value than teachers. In addition, in all three of these analyses, the 95% confidence interval associated with ρ_τ crossed zero (Table C.2). These results supported the use of schools, but not teachers, as an additional level in the final multilevel analyses. As a further check, Design Effects were calculated to determine if it was necessary to include schools as well as students in these analyses to determine the influence of selected predictor variables on student composite scores. Design Effect is a function of the ICC and average class size (Table C.2, Note 4). A resulting Design Effect value for any Phase and grade that was greater than 2 would indicate that it was necessary to do so (Muthen & Satorra 1995). As indicated in Table C.2, all Design Effect values ranged from a low of 4.402 (Phase Two, Grade 8) to a high of 7.644 (Phase Two, Grade 6). Thus, the results of these ICC and Design Effect analyses indicated that it was necessary to include students and schools as levels in the final multilevel analysis of selected variables.

The fifth step was to conduct multilevel analyses of the final 10 variables selected for each Phase and grade (Singer, 1998). Multilevel analysis takes into account both fixed effects and random effects when predicting an outcome. However, the distinction between fixed and random effects is not entirely clear (Gelman, 2005). As a result, Gelman provided his own definition of fixed and random effects: "We define effects (or coefficients) in a

Table C.2

Results for Step 4, Intraclass Correlation and Design Effect Analyses Pertaining to Multilevel Nesting, by Phase and Grade

Phase	Grade	Levels			ICC (ρ)	95% Confidence Intervals[1,2]		Sample Size[3]			Design Effect[4]
		Students	Teachers	Schools		Lower	Upper	Schools	Students	Median	
1	6	X	X		$\rho_{\tau} = 0.278$	99.078	283.817				
		X		X	$\rho_{sc} = 0.278$	99.078	283.817	47	964	20.5	6.42
		X	X	X	$\rho_{\tau} = 0.103$ $\rho_{sc} = 0.114$	0.0 99.287	0.0 283.817				
1	8	X	X		$\rho_{\tau} = 0.216$	104.244	306.552				
		X		X	$\rho_{sc} = 0.216$	104.244	306.552	47	936	19	4.89
		X	X	X	$\rho_{\tau} = 0.000$ $\rho_{sc} = 0.216$	0.0 306.552	0.0 306.552				
2	6	X	X		$\rho_{\tau} = 0.164$	81.943	215.191				
		X		X	$\rho_{sc} = 0.151$	60.548	215.543	34	1,791	45	7.644
		X	X	X	$\rho_{\tau} = 0.046$ $\rho_{sc} = 0.119$	- 0.424 26.293	85.494 190.593				
2	7	X	X		$\rho_{\tau} = 0.206$	97.103	317.113				
		X		X	$\rho_{sc} = 0.179$	67.332	284.639	30	1,574	38	7.623
		X	X	X	$\rho_{\tau} = 0.057$ $\rho_{sc} = 0.140$	- 10.052 24.868	123.603 251.443				
2	8	X	X		$\rho_{\tau} = 0.172$	63.063	264.422				
		X		X	$\rho_{sc} = 0.189$	54.728	308.572	23	690	19	4.402
		X	X	X	$\rho_{\tau} = 0.047$ $\rho_{sc} = 0.143$	- 10.154 -10.172	126.971 286.900				

Notes: (1) For Phase 1, Grades 6 and 8, Lower and Upper Confidence Intervals for Student/Teacher and Student/School ICC analyses are identical due to the fact that there was one class and therefore one teacher per school. (2) For Phase 2, Grades 6, 7, and 8, when the confidence interval for Student/Teacher or Student/School includes 0, there is no effect of that level on student composite scores. (3) For Sample Size, because the number of student varied widely from school to school (imbalanced), the median was used in place of the mean to calculate the Design Effect. (4) Design Effect = 1 + [(median size - 1) * ICC (ρ)]. Design Effects > 2 indicate that it was necessary to take into consideration the nesting of students in schools when estimating the predictor of student composite scores.

multilevel model as constant if they are identical for all groups in a population, and varying if they are allowed to differ from group to group" (2005, p. 21). In other words, fixed effects assume scores are independent, while random effects assume some type of relationship exists between variables (e.g., students are nested in schools). In summary, for these analyses: (a) students and schools served as the levels of analysis in all models (i.e., using Student ID and School ID numbers); (b) the multilevel analysis of data for each Phase and grade began with the calculation of the effect of different schools on student composite scores (i.e., for Model 1); and (c) using Hierarchical Linear Mixed-Effects Modeling (HLM) procedures (Bryk & Raudenbush, 1992), the final 10 selected variables were added to the models being analyzed one by one (i.e., for Models 2 through 11). In all HLM analyses we used a bootstrapping approach (10,000 samples) to improve the accuracy of coefficients and confidence intervals because we knew from prior analyses that these data sets did not meet all of the assumptions required for linear and linear-mixed regression analyses (Tabachnik & Fidell, 2013). In these analyses, Model 11 included all of the final 10 variables selected for each sample. The results from the HLM analysis of each model include model coefficients for each variable, as well as overall measures of the goodness of fit of that model to the data such as Log Likelihood (Bryk & Raudenbush, 1992). The detailed results of these multilevel analyses are summarized by Phase and by grade below.

For the <u>sixth</u> and final step, we did not conduct any further statistical analyses. Rather, we compared the results from the analyses of Phase One data (Research Question 3a) to the results from the analyses of Phase Two data (Research Question 3b) on a descriptive basis. Thus, for Research Question 3c, we will highlight any apparent similarities and differences in these Phase One and Phase Two results.

Results for Research Question 3a

The results for this research question are presented by grade; i.e., the results for the Phase One sixth-grade sample are presented first, and are followed by the results for the Phase One eighth-grade sample. This was done to allow us to present the results of the variable selection process (Steps 2 and 3) and, immediately after that, to present the results of HLM analyses for the final set of selected variables for each grade (Step 5).

Results for the Phase One sixth-grade sample. The results of the variable selection process are summarized in Table C.3. In **Step 2**, ANOVAs were used to select the 15 variables with the largest F values and the 15 variables with the smallest probability (p) values. The F values for these 15 variables ranged from a high of 21.19 (under Program Goal, *issue investigation*) to a low of 9.443. Similarly, the p values for these 15 variables ranged from $p < .000000$ (under School Composition, *% Asian students*) to $p = .000007$. When these two sets of selected variables were combined, a total of 20 variables had been selected. Of these, 7 were School variables, 4 were Program variables, 8 were Teacher variables, and 1 was a Student variable (i.e., Student Age).

In **Step 3**, Neural Network and MLP procedures were used to reduce this set of 20 variables to the final set of 10 variables to be included in HLM analyses. This set of final variables

Table C.3
*Results of Analysis Using the Full Variable Set: Selection of Variables from Phase 1 Grade 6 Related to Individual Student Composite Scores (n= 963)**

Selected Variables	**Type of Variable**				**Step 2: Composite Scores as DV**: 15 Most Significant Predictors		**Step 3: Neural Network Analysis, Multilayer Perceptron Method:** 10 Most Important Variables	
	School	Program	Teacher	Student	F Value	p Value	Ranking	Sensitivity Index
Student : Teacher Ratio	X					(9) .000000		
% Native American Students	X					(6) .000000		
% Asian Students	X				(6) 17.745	(1) .000000	10	1.041
% Black Students	X				(4) 18.499	(2) .000000	8	1.084
% Reduced Lunch Students	X				(14) 9.547	(5) .000000		
% Free Lunch Students	X				(13) 9.981	(4) .000000		
% ESOL Students	X				(10) 12.679	(10) .000000	7	1.120
Program Goal, Investigation Skills		X			(1) 21.119	(14) .000006	2	2.160
Program Goal, Service/Action Skills		X			(7) 15.975		3	1.361
Instr. Grouping, Individualized		X			(12) 10.747	(12) .000001		
Instr. Sites: Other Community Sites		X			(15) 9.443		4	1.295
Highest Degree Earned			X		(2) 19.172	(8) .000000		
Highest Degree Earned, (Ed.) Level			X		(5) 17.933	(3) .000000		
Teacher Certification: Y/N			X		(9) 12.693		1	2.934
Year Teaching, Total			X			(11) .000000	5	1.197
Years Teaching, Middle School			X			(7) .000000	6	1.195
# EE Inservices, Total			X			(15) .000007		
# EE Inservices, < 1 Day			X		(3) 19.057		9	1.082
Perceived Level of Env. Sensitivity			X		(11) 11.481			
Student Age				X	(8) 13.571	(13) .000002		

Notes: (Step 2) ANOVAs were used because Student Composite Score is a continuous variable. All p values were < .000007
(Step 3) In Statistica, a bootstrapping approach to Multilayer Perceptron (MLP) Methods was used to identify the most important predictors among those identified in Step 2 (1,000 bootstrapped samples).

Table C.4

Results of HLM Analyses of the Influence of Selected Variables on Student Composite Scores for the Phase 1, Grade 6 Sample (n=962)

Fixed Effects	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Intercept - Students	142.924***	143.619***	138.450***	138.320***	137.525***	137.759***	136.454***	137.199***	142.562***	147.791***	147.614***
Teacher Certification:											
• No (vs. Yes)		4.894	5.679	5.322	1.229	1.278	9.667*				
• Working On (vs. Yes)		- 9.495**	- 12.518***	- 12 .237***	-15.238***	- 15.374***	- 17.042***	- 23.066***	- 27.725***	- 33.967***	- 34.230***
Prog. Goal, Investigation Skills			9.221***	7.732***	8.443***	8.504***	9.766***	13.273***	13.324***	13.816***	14.269***
Prog. Goal, Service/Action Skills				2.537	- 0.642	- 0.738	1.092	- 1.147	1.201	3.877	3.689
Instr. Site, Other Comm. Sites					12.858***	12.852***	14.556***	12.741***	9.330**	4.007	3.654
Years Teaching, Total						- 0.018	- 1.221***	- 1.688***	- 1.386***	- 1.432***	- 1.413 ***
Years Teaching, Middle School							1.638***	2.010***	1.567***	1.265***	1.230 **
% ESOL Students								- 0. 161	- 0.244**	- 0.469**	- 0.491**
% Black Students									- 0.290***	0.401***	- 0.420***
# EE Inservices, < 1 Day										5.082	4.905
% Asian Students											0.215
Random Effects											
Intercept – Schools	186.689***	184.258***	168.122***	167.313***	149.126***	149.034***	102.656***	103.463***	86.845**	79.946***	79.367**
Residuals	687.771	675.791	675.945	675.917	675.828	675.841	680.079	674.503	674.142	687.852	687.839
Model Fit Statistics											
Degrees of Freedom	3	5	6	7	8	9	10	10	11	12	13
Log Likelihood	9120.853	8877.578	8685.409	8685.193	8680.850	8680.845	8493.926	6112.676	6108.420	5270.651	5270.508
Difference of Deviances		243.275***	192.169***	0.216	5.697	0.005	186.919***	2381.250***	4.256	837.769***	0.143
AIC	9126.853	8887.578	8697.409	8699.193	8696.850	8698.845	8513.926	6132.676	6130.420	5294.651	5296.508
BIC	9141.467	8911.808	8726.355	8732.964	8735.445	8742.264	8561.961	6177.430	6179.650	5346.565	5352.748`

Note: Levels of statistical significance: * p < .05 ** p < .01 *** p < .001

included 3 School variables, 3 Program variables, and 4 Teacher variables. Of these, the variables found to have the largest influence on student composite scores were Teacher Certification (ranked first), and the Program Goal, *issue investigation* (ranked second).

In **Step 5**, these 10 variables were included in HLM analyses (Table C.4). Each variable was entered into a separate model for analysis on the basis of variable rankings included in Table C.3. Thus, the variable ranked first was included in Model 2, the variables ranked first and second were included in Model 3, and so on, until all 10 variables had been included in Model 11. As described in the Methods section for **Step 4**, all of these HLM analyses included two levels: students (fixed effects) and schools (random effects).

In Table C.4, two types of results served as a general indicator of the statistical significance of each model. The first type appears under the heading Random Effects, and is the intercept for schools as a second and separate level of analysis in each model. The influence of schools on student composite scores was significant in Models 1 – 11 (Table C.4). The second type appears under the heading Model Fit Statistics, with particular attention to Log Likelihood and Difference of Deviances. Log Likelihood is a deviance statistic that reflects how well a given model fits the data set (e.g., the larger this deviance statistic, the poorer the fit of a given model to the data). We use the term Difference of Deviances to refer to the difference between the Log Likelihood value for a given model and the Log Likelihood value for the previous model (e.g., Log Likelihood for Model 9 – Log Likelihood for Model 10). Thus, a large Difference of Deviances value indicates that the unique variable added to that model helped to explain or predict student composite scores in a statistically significant manner. In Table C.4, the Difference of Deviances values were largest for Model 8 (2381.250, $p < .001$), where the School Composition variable *% of ESOL students* was added as a unique variable to this model, and Model 10 (837.769, $p < .001$), where the Number of EE Inservices *≤ 1 day* was added as a unique variable to this model. In other words, the addition of each of these variables helped to significantly improve the ability of these two models to explain or predict student composite scores.

Although many of the unique variables added to Models 2 – 10 were found to be significant predictors of student composite scores (Table C.4), only the results for Model 11 are discussed here. To aid in the interpretation of the results for Model 11, we have summarized the results for variables in Model 11 that were statistically significant and, for any nominal variables, included sample size and average composite scores data for each (Table C.5). In Table C.5, variables that had a positive influence on student composite scores have a positive coefficient (i.e., they are related to an increase in the average composite score), while variables that had a negative influence on student composite scores have a negative coefficient (i.e., they are related to a decrease in the average composite scores).

For the Phase One sixth-grade sample, two variables had a significantly positive influence on student composite scores: (1) under Program Goals, *issue investigation skills*: coefficient = 14.269 ($p < .001$); and (2) under Years Teaching, *middle school*: coefficient = 1.230 ($p < .01$). As indicated in Table C.5, for (1) the Program Goal *issue investigation skills*, the average composite score for students in schools for which this goal was checked was

Table C.5

Summary of Significant HLM Analysis Results for the Phase 1 Grade 6 and Grade 8 Samples

	Variables[1]		Results for Phase 1, Grade 6			Results for Phase 1, Grade 8		
Fixed Effects	Set	Type	Model 11 Coefficient[2]	Freq.	Composite Average[3]	Model 11 Coefficient[2]	Freq.	Composite Average[3]
Intercept – Students			147.614***			146.690***		
% Black Students	School	Inter.	**- 0.420*** **			- 0.162		
% ESOL Students	School	Inter.	**- 0.491** **			**- 0.757*** **		
EE Program Type (vs. No Env. Program):	Program	Nom.						
• Environmental Club						- 0.349	n = 1	118.408
• Env. Curriculum						**24.969***	n = 6	144.988
• Env. Curriculum + Env. Club						**27.684*** **	n = 1	148.377
• Outdoor Lab						5.080	n = 3	125.385
• Env. Curriculum + Outdoor Lab						- 12.547	n = 3	119.031
• No EE Program							n = 33	132.360
Program Goal:	Program	Nom.						
• Investigation Skills (checked)			**14.269*** **	n = 26	137.760			
Not checked				n = 20	123.491			
Instructional Method (n = 48)	Program	Nom.						
• Service Learning						**- 23.440*** **	n = 4	114.085
Teacher Age (vs. > 60):	Teacher	Nom.						
• 21-30						**- 17.151*** **	n = 7	122.012
• 31-40						**- 8.288***	n = 12	133.353
• 41-50						**- 16.730*** **	n = 11	128.510
• 51-60						**- 12.615***	n = 10	130.314
• > 60							n = 3	142.935
Years Teaching, Total	Teacher	Inter.	**- 1.413*** **					
Years Teaching, Middle School	Teacher	Inter.	**1.230** **					
Teacher Certification (vs. Yes):	Teacher	Nom.						
• No				n = 2	148.513			
• Working On			**- 34.230*** **	n = 4	134.125			
• Yes				n = 41	143.620			

Notes: (1) Nom. = nominal, Ord. = Ordinal, and Inter. = Interval; (2) * = < .05 ** = < .01 *** = < .001;
(2) Sample Sizes: Phase 1, GR 6: Schools, Programs & Teachers = 48; Phase 1, GR 8: Schools, Programs & Teachers = 48
(3) Composite Score Means: Phase 1, GR 6 = 143.99; Phase 1, GR 8 = 140.19

137.760, which is more than 14 points higher than the average composite score for when this goal was not checked. In addition, this was checked for 26 school programs and not checked for 20, which indicates that this pattern was reasonably pervasive in this sample. However, because (2) Years Teaching in *middle school* was not a nominal variable, the additional of sample size and average composite score data was not possible.

Further, four variables were found to have a significantly <u>negative</u> influence on sixth-grade student composite scores: (1) under Teacher Certification, *working on*: coefficient = - 34.230 (p < .001); (2) under Years Teaching, *total*: coefficient = - 1.413 (p < .001); and under School Composition, (3) the *percent of black students*: coefficient = - 0.420 (p < .001), and (4) *the percent of ESOL students*: coefficient = - .491 (p < .01). Of these, (1) Teacher Certification, *working on*, was the only nominal variable, and therefore the only variable with additional sample size and average score data in Table C.5. These data indicated that only four teachers were working on their Teacher Certification, and further review of the data set revealed that all four of these teachers held positions in private schools. Further, the data in Table C.5 indicate that, in the presence of all 10 variables in Model 11, the average composite score for students associated with teachers *working on* certification was about 9.5 points lower than for students associated with the 41 teachers in this sample who were certified (i.e., *yes*). Due to the small number of teachers who were working on Teacher Certification, this was not a pervasive pattern in this sample.

Results for the Phase One eighth-grade sample. The results of the variable selection process are summarized in Table C.6. In **Step 2**, ANOVAs were used to select the 15 variables with the largest F values and the 15 variables with the smallest probability (p) values. The F values for these 15 variables ranged from a high of 24.75 (under Instructional Methods, *cooperative learning*) to a low of 9.911. Similarly, the p values for these 15 variables ranged from p < .000000 (under Teacher Certification, *educational level*) to p = .000004. When these two sets of selected variables were combined, a total of 22 variables had been selected. Of these, 8 were School variables, 8 were Program variables, and 6 were Teacher variables (i.e., none were Student variables).

In **Step 3**, Neural Network and MLP procedures were used to reduce this list of 22 variables to the final set of 10 variables to be included in HLM analyses. This set of final variables included 3 School variables, 6 Program variables, and 1 Teacher variable. Of these, the variables found to have the largest influence on student composite scores were Assessment Methods, *informal* (ranked first), and EE Program Type (ranked second).

In **Step 5**, these 10 variables were included in HLM analyses (Table C.7). Each variable was entered into a separate model for analysis on the basis of variable rankings included in Table C.6. Thus, the variable ranked first was included in Model 2, the variables ranked first and second were included in Model 3, and so on, until all 10 variables had been included in Model 11. As described in the Methods section for **Step 4**, all of these HLM analyses included two levels: students (fixed effects) and schools (random effects).

In Table C.7, two types of results serve as a general indicator of the significance of each model. The <u>first type</u> appears under the heading Random Effects, and is the intercept for

Table C.6
*Results of Analysis Using the Full Variable Set: Selection of Variables from Phase 1 Grade 8 Related to Individual Student Composite Scores (n= 916)**

Selected Variables	**Type of Variable**			**Step 2: Composite Scores as DV**: 15 Most Significant Predictors		**Step 3: Neural Network Analysis, Multilayer Perceptron Method:** 10 Most Important Variables	
	School	Program	Teacher	F Value	p Value	Ranking	Sensitivity Index
Total Enrollment	X				(5) .000000		
Student : Teacher Ratio	X				(8) .000000		
% Asian Students	X				(4) .000000		
% Hispanic Students	X				(14) .000001	8	1.018
% Black Students	X			(9) 12.345	(3) .000000	7	1.025
% ESOL Students	X			(10) 10.836		10	1.011
% Federal IDEA Students	X				(9) .000000		
% Special Needs Students	X				(6) .000000		
Env. Program, Type		X		(7) 14.141	(7) .000000	2	1.189
Instr. Grouping, Whole Class		X		(11) 19.628	(12) .000001		
Instr. Grouping, Individualized		X		(5) 17.324	(2) .000000		
Instr. Methods, Labs		X		(12) 1.197		5	1.062
Instr. Methods, Cooperative Learning		X		(1) 24.751	(13) .000001	4	1.092
Instr. Methods, Service Learning		X		(3) 21.557		3	1.161
Instr. Sites, Other Community Sites		X		(6) 16.983		6	1.048
Assessment, Informal		X		(4) 18.088		1	1.318
Teacher Age			X	(14) 10.017	(11) .000000	9	1.017
Highest Degree Earned			X	(15) 9.911			
Teacher Certification, Ed. Level			X	(2) 22.568	(1) .000000		
# EE Inservices, Total			X	(13) 10.031	(10) .000000		
# EE Inservices, 1 Week			X	(8) 13.453			
Perceived Level of Participation			X		(15) .000004		

Notes: (Step 2) ANOVAs were used because Student Composite Score is a continuous variable. All p values were < .000005
(Step 3) In Statistica, a bootstrapping approach to Multilayer Perceptron (MLP) Methods was used to identify the most important predictors among those identified in Step 2 (1,000 bootstrapped samples).

Table C.7

Results of HLM Analyses of the Influence of Selected Variables on Student Composite Scores for the Phase 1, Grade 8 Sample (n=918)

Fixed Effects	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Intercept – Students	138.724***	134.511***	132.378***	132.390***	134.554***	134.415***	135.531***	137.960***	137.928***	152.788***	146.690***
Assessment, Informal		5.704**	7.326**	7.320***	8.933***	8.663***	10.108***	8.802***	9.360***	9.062***	4.822
EE Program Type:											
• Env. Club only			- 19.782**	- 19.788**	- 23.575***	- 23.155**	- 25.717**	-19.468**	- 18.905**	- 13.952	- 0.349
• Env. Curriculum or Project			14.311***	14.314***	13.298***	13.045***	12.609***	12.004***	12.933***	12.628***	24.969*
• Env. Curriculum + Env. Club			21.232***	21.220***	19.046**	19.195**	18.079**	15.871*	16.865**	16.017*	27.684***
• Outdoor Lab			- 7.120**	- 1.130	- 2.660	- 2.240	- 5.312	- 5.192	- 6.200	- 6.975	5.080
• Env. Curriculum + Outdoor Lab			- 6.951*	- 6.945	- 9.670*	- 9.427*	-10.356**	- 11.186**	- 11.135**	- 13.328***	-12.547
Instr. Method, Service Learning				- 23.621***	- 23.617***	-23.617***	- 16.361*	-12.753*	-13.320*	3.316	- 23.440*
Instr. Method, Coop. Learning					- 4.946*	- 4.527*	- 5.988*	- 4.361	- 2.907	- 6.390**	3.764
Instr. Method, Labs						2.293	- 0.050	- 0.283	- 0.181	- 1.581	2.710
Instr. Site, Other Comm. Sites							- 6.746	- 6.651*	- 6.147	- 16.270***	4.143
% Black Students								- 0.240***	- 0.240***	- 0.208**	- 0.162
% Hispanic Students									- 0.128**	- 0.065	0.085
Teacher Age:											
• 21-30 (vs. > 60)										- 20.924***	- 17.151***
• 31-40 (vs. > 60)										- 9.512*	- 8.288*
• 41-50 (vs. > 60)										- 14.426***	- 16.730***
• 51-60 (vs. > 60)										-12.621**	-12.615*
% ESOL Students											- 0.757***
Random Effects											
Intercept - Schools	200.122***	1 99.043***	147.471***	137.650***	133.157***	133.020***	130.690***	118.949***	115.068***	55.389***	14.724
Residuals	744.252	743.631	744.053**	744.303	744.425	744.415	744.589	744.703	744.679	744.567	740.803
Model Fit Statistics											
Degrees of Freedom	3	4	9	10	11	12	13	14	15	19	20
Log Likelihood	8751.491	8588.190	8577.355	8575.141	8574.090	8574.042	8573.613	8570.428	8569.250	7853.346	5847.641
Difference in Deviances		163.301***	10.835	2.214	1.051	0.048	0.429	3.185	1.178	715.904***	2005.705***
AIC	8757.491	8596.190	8595.355	8595.141	8596.090	8598.042	8599.613	8598.428	8599.250	7891.346	5887.641
BIC	8771.954	8615.400	8638.577	8643.165	8648.916	8655.670	8662.044	8665.662	8671.286	7980.984	5976.171

Note: Levels of statistical significance: * p < .05 ** p < .01 *** p < .001

schools as a second and separate level of analysis in each model. The influence of schools on student composite scores was significant in Models 1 – 10, but not in Model 11 (Table C.7). The second type appears under the heading Model Fit Statistics, with particular attention to Log Likelihood and Difference of Deviances. In Table C.7, the Difference of Deviances values were largest for Model 11 (2005.705, $p < .001$), where the School Composition variable *% of ESOL students* was added as a unique variable to this model, and Model 10 (715.904, $p < .001$), where Teacher Age was added as a unique variable to this model. In other words, the addition of each of these variables helped to significantly improve the ability of these two models to explain or predict student composite scores.

Although many of the unique variables added to Models 2 – 10 were found to be significant predictors of student composite scores (Table C.7), only the results for Model 11 are discussed here. To aid in the interpretation of the results for Model 11, we have summarized the results for variables in Model 11 that were statistically significant and, for any nominal variables, included sample size and average composite scores data for each (Table C.5). In Table C.5, variables that had a positive influence on student composite scores have a positive coefficient (i.e., they are related to an increase in the average composite score), while variables that had a negative influence on student composite scores have a negative coefficient (i.e., they are related to a decrease in the average composite scores).

For the Phase One eighth-grade sample, two variables had a significantly positive influence on student composite scores: under EE Program Type, (1) programs with only an *environmental curriculum*: coefficient = 24.965 ($p < .05$); and (2) programs with *an environmental curriculum and an environmental club*: coefficient = 27.684 ($p < .001$). As indicated in Table C.5, in the presence of all 10 variables in Model 11: (1) for programs with only an *environmental curriculum*, the average composite score for students in schools with this EE Program Type was 144.988, which is more than 12.5 points higher than the average score for schools without any type of environmental program (average = 132.360). In addition, this EE Program Type was checked for 6 school programs, as compared to 33 schools that had no environmental program, which indicates that this pattern was apparent, but not pervasive, in this sample. Again, in the presence of all 10 variables in Model 11, (2) for programs with *an environmental curriculum and an environmental club,* the average composite score for students in schools with this EE Program Type was 148.377, which is more than 16 points higher than the average score for schools without any type of environmental program (average = 132.360). However, this EE Program Type was checked for only 1 school program, as compared to 33 schools that had no environmental program, indicating that this pattern was not pervasive in this sample.

Of the six variables that had a significantly negative influence on eighth-grade student composite scores, four were related to Teacher Age. These were: when teacher age was *21-30*: coefficient = - 17.151 ($p < 001$), when it was *31-40*: coefficient = - 8.288 ($p < .05$), when it was *41-50*: coefficient = - 16.730 ($p < .001$), and when it was *51-60*: coefficient = - 12.615 ($p < .05$). As indicated in Table C.5, the highest average composite score for the eighth-grade sample was found for students of teachers who were *older than 60* (average = 142.935), although this only applied to students of three teachers. The negative coefficients noted above indicate that students of teachers in these younger age groups had lower average composite scores than did students of teachers older than 60. More specifically, when teacher age was: (*21-30,* n = 7*)* the average composite score was 20.9 points lower; (*31-40,* n = 12) the average composite score was about 9.5 points lower; (*41-50,* n = 11) the average composite score was about 14.4 points lower; and (*51-60*, n = 10) the average composite score was more than 12.5 points lower. Due to the small sample size for teachers older than 60 and to fluctuations in these average composite scores, this pattern is apparent, but not pervasive, in this sample.

The two other variables that had a significantly <u>negative</u> influence on eighth-grade student composite scores were: under Instructional Methods, use of *service-learning*: coefficient = - 23.440 ($p < .001$); and under School Composition, the *percent of ESOL students*: coefficient = - 0.757 ($p < .001$). Of these, only the former was a nominal variable. In the presence of all 10 variables in Model 11, the average composite score for this variable was 114.085. However, Table C.5 indicates that only four of 48 teachers reported using *service-learning* as an Instructional Method, so this was not a pervasive pattern in this sample.

Comparison across grades. When results for the Phase One sixth-grade and eighth-grade sample were compared (Table C.5), only one variable was found to have a significant influence on student composite scores in both grades; i.e., under School Composition, the *percent of ESOL students*. In both cases, this variable had a negative influence on student composite scores.

Results for Research Question 3b

As for Research Question 3a, the results for this research question are presented by grade; i.e., the results for the Phase Two sixth-grade sample are presented first, then the results for the Phase Two seventh-grade sample, and finally the results for the Phase Two eighth-grade sample. This was done to allow us to present the results of the variable selection process (Steps 2 and 3) and, immediately after that, the presentation of results of HLM analyses for the final set of selected variables for each grade (Step 5).

Results for the Phase Two sixth-grade sample. The results of the variable selection process are summarized in Table C.8. In **Step 2**, ANOVAs were used to select the 15 variables with the largest F values and the 15 variables with the smallest probability (p) values. The F values for these 15 variables ranged from a high of 26.69 (Student Age) to a low of 10.923. Similarly, all p values for these 15 variables were $p < .000000$, although the smallest of these p values was for the variable Years Teaching, *middle school*. When these two sets of selected variables were combined, a total of 20 variables had been selected. Of these, 4 were School variables, 8 were Program variables, 7 were Teacher variables, and 1 was a Student variable (i.e., Student Age).

In **Step 3**, Neural Network and MLP procedures were used to reduce this set of 20 variables to the final set of 10 variables to be included in HLM analyses. This set of final variables included 2 School variables, 5 Program variables, 2 Teacher variables and 1 Student variable. Of these, the variables found to have the largest influence on student composite

Table C.8
*Results of Analysis Using the Full Variable Set: Selection of Variables from Phase 2 Grade 6 Related to Individual Student Composite Scores (n= 1,790)**

Variables	**Type of Variable**				**Step 2: Mean Composite Scores** as DV: 15 Most Significant Predictors		**Step 3: Neural Network Analysis, Multilayer Perceptron Method:** 10 Most Important Variables	
	School	Program	Teacher	Student	F Value	p Value	Ranking	Sensitivity Index
Total Enrollment	X					(9) .000000		
% Asian Students	X				(10) 14.341	(4) .000000	6	1.012
% White Students	X					(15) .000000	8	1.005
% Special Needs Students	X					(14) .000000		
Env. Program, Type		X				(10) .000000	4	1.041
Program Duration		X			(12) 13.654	(6) .000000	9	1.004
Program Goal, Science		X			(14) 11.247		7	1.007
Program Goal, Affective Disp.		X			(15) 10.923			
Instr. Method, Projects		X			(5) 19.905		2	1.080
Instr. Method, Service Learning		X			(7) 16.397			
Instr. Sites, Science Labs		X			(2) 26.150	(11) .000000	1	1.249
Instr. Sites, Other Comm. Sites		X			(8) 16.103			
Teacher Age			X		(11) 14.226	(8) .000000		
Teacher Ethnicity			X		(3) 25.674	(12) .000000	3	1.066
Teacher Certification, Ed. Level			X		(4) 22.973	(3) .000000	5	1.017
Year Teaching, Total			X		(9) 14.854	(2) .000000		
Years Teaching, Middle School			X		(6) 18.595	(1) .000000		
# EE Inservices, Total			X			(7) .000000		
# EE Inservices, 3-7 Days			X		(13) 11.247	(13) .000000		
Student Age				X	(1) 26.689	(5) .000000	10	1.004

Notes: (Step 2) ANOVAs were used because Student Composite Score is a continuous variable. All p values were < .000000.
(Step 3) In Statistica, a bootstrapping approach to Multilayer Perceptron (MLP) Methods was used to identify the most important predictors among those identified in Step 2 (1,000 bootstrapped samples).

Table C.9

Results of HLM Analyses of the Influence of Selected Variables on Student Composite Scores for the Phase 2, Grade 6 Sample (n=1,791)

Fixed Effects	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Intercept - Students	149.542***	155.291***	153.006***	153.002***	159.966***	159.137***	152.085***	150.928***	140.507***	138.933***	141.570***
Instr. Site, Science Labs		- 10.907***	- 11.955***	- 11.393***	- 14.049***	- 11.182***	- 10.824***	-11.491***	-13.021***	-13.799***	- 13.251***
Instr. Method, Projects			5.601***	5.597***	0.609	2.913	5.146**	4.970**	5.729**	4.683**	4.131*
Teacher Ethnicity				- 5.892	- 5.893	- 9.840**	- 9.058*	- 9.041*	- 6.816	- 7.217	- 6.572
EE Prog. Type: (vs. curr. only)											
• env. club only					- 6.882	- 0.541	0.187	- 0.147	- 0.431	0.922	- 0.646
• curr. + club					15.611***	12.850***	12.917***	12.735***	14.923**	10.638*	9.630*
• curr. + outdoor lab					- 7.258**	- 6.209*	- 5.418*	- 5.110*	- 3.552	- 9.821**	- 9.738**
• curr. + club + out. lab.					1.539	7.127	- 5.448	- 3.846	- 5.313	- 7.683	- 6.659
• curr. + resident. program					- 17.151***	- 11.340***	- 6.048*	- 6.723*	- 6.876**	- 6.071*	- 6.624*
• curr. + club + res. prog.					2.106	1.386	2.707	3.063	4.016	5.944**	5.640*
• curr. + out. lab + res. prog.					- 6.847*	- 0.509	- 2.587	- 3.314	- 6.276	- 6.044	- 6.349
Teacher Certificate, Ed Level:											
• elementary (vs. multiple)						- 6.963***	- 6.282**	- 6.315**	- 5.935**	- 5.482**	- 5.408**
• middle (vs. multiple)						- 7.129*	- 5.081	- 5.084	0.049	- 1.115	- 1.382
• secondary (vs. multiple)						- 12.521**	- 12.143**	- 12.075**	-11.440**	- 11.569**	- 11.090**
% Asian Students							1.212***	1.193***	1.275***	0.973***	0.890**
Program Goal, Science								2.077	4.215	3.485	3.968
% White Students									0.115***	0.121***	0.111*
Program Duration										0.254**	0.258*
Student Age:											
• 12 (vs. ≤ 11 yrs.)											- 1.719
• 13 (vs. ≤ 11 yrs.)											- 17.144***
• 14 (vs. ≤ 11 yrs.)											- 13.826
• 15 (vs. ≤ 11 yrs.)											- 21.659***
Random Effects											
Intercept - Schools	133.354***	110.071***	103.310***	102.215***	61.651***	41.027***	22.959***	22.149**	12.994	10.168	7.442
Residuals	778.270	777.557	777.466	774.084	775.672	779.048	778.529	778.699	779.998	779.253	764.723
Model Fit Statistics											
Degrees of Freedom	3	4	5	6	13	16	17	18	19	20	24
Log Likelihood	17,076.573	17,069.455	17,065,468	16,840.830	16,716.444	15,971.702	15,959.711	15,959.475	15,954.393	15,949.834	15.895.802
Difference in Deviances		7.118	3.987	224.638***	124.386***	744.742***	11.991	0.236	5.082	4.559	54.320**
AIC	17,082.573	17,077.455	17.077.468	16,852.830	16,742,444	16,003.702	15,993.711	15,995.475	15,992.393	15,989.834	15,943.802
BIC	17,099.045	17,099.415	17,104.921	16,885.696	16.813.565	16,090.508	16,085.942	16,093.131	16.095.475	16.098.341	16.073.982

Note: Levels of statistical significance: * p < .05, ** p < .01 *** p < .001

scores were: under Instructional Sites, *science labs* (ranked first), and under Instructional Methods, *projects* (ranked second).

In **Step 5**, these 10 variables were included in HLM analyses (Table C.9). Each variable was entered into a separate model for analysis on the basis of variable rankings included in Table C.8. Thus, the variable ranked first was included in Model 2, the variables ranked first and second were included in Model 3, and so on, until all 10 variables had been included in Model 11. As described in the Methods section for **Step 4**, all of these HLM analyses included two levels: students (fixed effects) and schools (random effects).

In Table C.9, two types of results serve as a general indicator of the significance of each model. The first type appears under the heading Random Effects, and is the intercept for schools as a second and separate level of analysis in each model. The influence of schools on student composite scores was significant in Models 1 – 8, but not for Models 9 – 11 (Table C.9). The second type appears under the heading Model Fit Statistics, with particular attention to Log Likelihood and Difference of Deviances. In Table C.9, the Difference of Deviances values were largest for Model 6 (744.742, $p < .001$), where Teacher Certification, Educational Level was added as a unique variable to this model, and Model 4 (224.638, $p < .001$), where Teacher Ethnicity was added as a unique variable to this model. In other words, the addition of each of these variables helped to significantly improve the ability of these two models to explain or predict student composite scores, even when the added variable itself was not found to be a significant predictor (e.g., Teacher Ethnicity).

Although many of the unique variables added to Models 2 – 10 were found to be significant predictors of student composite scores (Table C.9,), only the results for Model 11 will be discussed here. To aid in the interpretation of the results for Model 11, we have summarized the results for variables in Model 11 that were statistically significant and, for any nominal variables, included sample size and average composite scores data for each (Table C.10). In Table C.10, variables that had a positive influence on student composite scores have a positive coefficient (i.e., they are related to an increase in the average composite score), while variables that had a negative influence on student composite scores have a negative coefficient (i.e., they are related to a decrease in the average composite scores).

For the Phase Two sixth-grade sample, six variables had a significantly positive influence on sixth-grade student composite scores. These were:

- under EE Program Type, having *an environmental curriculum and an environmental club*: coefficient = 9.630 ($p < .05$);
- under EE Program Type, having *an environmental curriculum, an environmental club, and a residential program*: coefficient = 5.640 ($p < .01$);
- under Instructional Methods, use of *projects*: coefficient = 4.131 ($p < .05$);
- Program Duration (in weeks): coefficient = 0.258 ($p < .05$);
- under School Composition, the *percent of Asian students*: coefficient = 0.890 ($p < .01$); and
- under School Composition, the *percent of white students*: coefficient = 0.111 ($p < .05$).

Table C.10
Summary of HLM Analysis Results for the Phase 2 Grade 6, Grade 7, and Grade 8 Samples

	Variables[1]	Results for Phase 2, Grade 6			Results for Phase 2, Grade 7			Results for Phase 2, Grade 8		
Fixed Effects	**Set, Type**	**Model 11 Coefficient[2]**	**Freq.[3]**	**Composite Average[4]**	**Model 11 Coefficient[2]**	**Freq.[3]**	**Composite Average[3]**	**Model 11 Coefficient[2]**	**Freq.[3]**	**Composite Average[4]**
Intercept – Students										
% Asian Students	S, I	**0.890****								
% While Students		**0.111***								
EE Program Type (vs. Env. Curriculum):	P, N									
Environmental Club		- 0.646	n = 2	143.417						
Env. Curriculum + Club		**9.630***	n = 2	165.910						
Env. Curriculum + Outdoor Lab		**- 9.738****	n = 7	143.041						
Env. Curriculum + Resident Program		**- 6.624***	n = 4	133.148						
Env. Curric. + Club + Outdoor Lab		- 6.659	n = 1	129.051						
Env. Curric. + Club + Resident Program		**5.640***	n = 3	152.045						
Env. Curric. + Out. Lab + Res. Program		- 6.349	n = 2	143.453						
Environmental Curriculum			n = 13	144.261						
Program Duration	P, I	**0.258***								
Instructional Methods (vs. Not Checked):	P, N									
Discussion					**- 11.488*****	n = 12	142.076			
Not Checked						n = 15	153.564			
Cooperative Learning					**12.373*****	n = 11	154.006			
Not Checked						n = 17	141.633			
Projects		**4.131***	n = 18	149.957						
Not Checked			n = 17	144.360						
Instructional Sites (vs. Not Checked):	P, N									
Library/Media Center					**- 11.785*****	n = 9	141.927			
Not Checked						n = 19	153.713			
Science Labs		**- 13.251*****	n = 18	144.384	**- 10.800*****	n = 20	142.420			
Not Checked			n = 17	155.291		n = 9	153.220			

Table C.10 (cont.)

	Variables[1]	Results for Phase 2, Grade 6			Results for Phase 2, Grade 7			Results for Phase 2, Grade 8		
Fixed Effects	**Set, Type**	**Model 11 Coefficient[2]**	**Freq.[3]**	**Composite Average[4]**	**Model 11 Coefficient[2]**	**Freq.[3]**	**Composite Average[3]**	**Model 11 Coefficient[2]**	**Freq.[3]**	**Composite Average[4]**
Highest Degree Earned, Ed. Level	T, O							**- 5.183***		
Teacher Certification, Ed. Level:	T, N									
Elementary (vs. Multiple)		**- 5.408****	n = 26	132.930						
Middle (vs. Multiple)		- 1.382	n = 4	136.955						
Secondary (vs. Multiple)		**- 11.090****	n = 2	127.247						
Multiple			n = 23	138.337						
# EE Inservices, 1-2 Days	T, I				**- 0.608***					
Student Age	S, N									
12 (vs. ≤ 11 yrs.)		- 1.719	n = 1,202	143.018						
13 (vs. ≤ 11 yrs.)		**- 17.144*****	n = 99	127.593						
14 (vs. ≤ 11 yrs.)		- 13.826	n = 5	130.911						
15 (vs. ≤ 11 yrs.)		**- 21.659*****	n = 2	123.078						
11 yrs.			n = 48	144.737						

Notes: (1) Nom. = nominal, Ord. = Ordinal, and Inter. = Interval;

(2) * = < .05 ** = < .01 *** = < .001;

(3) Sample Sizes: Phase 2, GR 6: Schools & Programs = 35; Teachers = 57;
Phase 2, GR 7: Schools & Programs = 30; Teachers = 42;
Phase 2, GR 8: Schools & Programs = 23; Teachers = 32

(4) Average Composite Scores for the Samples in These Analyses: Phase 2, GR 6 = 147.21; Phase 2, GR 7 = 144.76; Phase 2, GR 8 = 150.57

In addition, seven variables were found to have a significantly negative influence on sixth-grade student composite scores. These were:

- under EE Program Type, having *an environmental curriculum and outdoor lab*: coefficient = - 9.738 ($p < .01$);
- under EE Program Type, having *an environmental curriculum and a residential program:* coefficient = - 6.624 ($p < .05$);
- under Instructional Sites, use of *science labs*: coefficient = - 13.251 ($p < .001$);
- under Teacher Certification, Educational Level, *elementary*: coefficient = - 5.408 ($p < .01$);
- under Teacher Certification, Educational Level, *secondary*: coefficient = - 11.090 ($p < .01$);
- under Student Age, *13 years old*: coefficient = - 17.114 ($p < .001$); and
- under Student Age, *15 years old*: coefficient = - 21.659 ($p < .001$).

Using data presented in Table C.10, three brief explanations will help to further the interpretation of results above pertaining to EE Program Type, Teacher Certification, and Student Age. First, with respect to EE Program Type, two EE Program Types had a positive influence and two had a negative influence on average composite scores for this sample. All EE Program Types were compared to the average composite score for students in those schools that had only *an environmental curriculum* of some kind (average = 144.261). In the presence of all 10 variables in Model 11, the program types that had significantly positive coefficients had an average scores greater than this: programs with *an environmental curriculum and an environmental club* (about 21.6 points greater), and those with *an environmental curriculum, an environmental club and a residential program* (about 7.8 points greater). However, the former was found in only two schools, and the latter in only three schools. The program types that had significantly negative coefficients had an average scores lower than the average noted above: programs with *an environmental curriculum and outdoor lab* (about 1.2 points lower), and those with *an environmental curriculum and a residential program* (about 11.1 points lower). The former was found in seven schools, and the latter in only four schools. Although these positive and negative patterns are apparent in the data, these small sample sizes suggest that these were not pervasive within this sample.

Second, with respect to Teacher Certification, the highest average composite score for the Phase Two sixth-grade sample was found among students whose teachers held Teacher Certifications for *multiple grade levels* (average = 138.337). Thus, in the presence of all 10 variables in Model 11, average composite scores were slightly lower for teachers whose only Certification was at the *middle level* (about 1.4 points lower), even lower for teachers whose only Teacher Certification was at the *elementary level* (about 5.4 points lower), and lowest for teachers whose only Teacher Certification was at the *secondary level* (about 11.1 points lower). Finally, with respect to Student Age, the highest average composite score for this sample was found for students who were *11 years old or younger* (average = 144.737). Thus, in the presence of all 10 variables in Model 11, average composite scores were significantly lower than this for *13 year olds* (about 17.1 points lower) and for *15 year olds* (about 21.7 points lower).

Table C.11
*Results of Analysis Using the Full Variable Set: Selection of Variables from Phase 2 Grade 7 Related to Individual Student C Composite Scores (n= 1,574)**

Variables	Type of Variable				Step 2: **Mean Composite Scores** as DV: 15 Most Significant Predictors		Step 3: **Neural Network Analysis, Multilayer Perceptron Method:** 10 Most Important Variables	
	School	Program	Teacher	Student	F Value	p Value	Ranking	Sensitivity Index
Student : Teacher Ratio	X					(10) .000000		
% Free Lunch Students	X					(5) .000000		
% Red Lunch Students	X					(11) .000012		
Program Duration		X			(8) 14.465	(2) .000000		
Program Intensity		X				(6) .000001		
Program Goal, Science		X			(12) 12.850			
Program Goal, Affect. Dispositions		X			(10) 13.312		5	1.400
Program Goal, Service/Action Skills		X			(11) 13.312		6	1.211
Instr. Grouping, Rank 1st		X			(13) 10.930			
Instr. Grouping, Rank 3rd		X			(15) 9.180			
Instr. Sites, Lib./Media Center		X			(3) 23.871	(8) .000002	3	1.454
Instr. Site, Science Labs		X			(6) 18.800	(15) .00002	8	1.130
Instr. Sites, Field Sites		X			(7) 14.891		1	1.507
Instr. Method, Coop. Learning		X			(4) 19.670	(13) .000015	7	1.184
Instr. Method, Service Learning		X			(5) 19.670	(12) .000015	10	1.102
Instr. Method, Discussion		X			(9) 13.746		4	1.423
Assess. Methods, Ranked 3rd		X			(14) 9.605			
Teacher Certification, Subject			X			(14) .000018	2	1.473
Year Teaching, Middle School			X			(9) .000007		
# EE Inservices, Total			X			(3) .000000		
# EE Inservices, < 1 Day			X		(2) 24.495	(7) .000001		
# EE Inservices, 1-2 Days			X			(4) .000000	9	1.102
Student Age				X	(1) 26.639	(1) .000000		

Notes: (Step 2) ANOVAs were used because Student Composite Score is a continuous variable. All p values were < .00002.
(Step 3) In Statistica, a bootstrapping approach to Multilayer Perceptron (MLP) Methods was used to identify the most important predictors among those identified in Step 2 (1,000 bootstrapped samples).

Results for the Phase Two seventh-grade sample. The results of the variable selection process are summarized in Table C.11. In **Step 2**, ANOVAs were used to select the 15 variables with the largest F values and the 15 variables with the smallest probability (p) values. The F values for these 15 variables ranged from a high of 26.64 (Student Age) to a low of 9.180. Similarly, the p values for these 15 variables ranged from a low of $p < .000000$ (Student Age) to .0002. When these two sets of selected variables were combined, a total of 23 variables had been selected. Of these, 3 were School variables, 14 were Program variables, 5 were Teacher variables, and 1 was a Student variable (i.e., Student Age).

In **Step 3**, Neural Network and MLP procedures were used to reduce this set of 20 variables to the final set of 10 variables to be included in HLM analyses. This set of final variables included 8 Program variables and 2 Teacher variables (i.e., no School or Student variables). Of these, the variables found to have the largest influence on student composite scores were: under Instructional Sites, *field sites* (ranked first), and under Teacher Certification, Subject (ranked second).

In **Step 5**, these 10 variables were included in HLM analyses (Table C.12). Each variable was entered into a separate model for analysis on the basis of variable rankings included in Table C.11. Thus, the variable ranked first was included in Model 2, the variables ranked first and second were included in Model 3, and so on, until all 10 variables had been included in Model 11. As described in the Methods section for **Step 4**, all of these HLM analyses included two levels: students (fixed effects) and schools (random effects).

In Table C.12, two types of results serve as a general indicator of the significance of each model. The <u>first type</u> appears under the heading Random Effects, and is the intercept for schools as a second and separate level of analysis in each model. The influence of schools on student composite scores was significant in Models 1 – 11 (i.e., all models; Table C.12). The <u>second type</u> appears under the heading Model Fit Statistics, with particular attention to Log Likelihood and Difference of Deviances. In Table C.12, the Difference of Deviances values were largest for Model 10 (2,754.630, $p < .001$), where the Number of EE Inservices lasting *1-2 days*, was added as a unique variable to this model, and Model 4 (1,102.949, $p < .001$), where Instructional Sites, *library/media center* was added as a unique variable to this model. In other words, the addition of each of these variables helped to significantly improve the ability of these two models to explain or predict student composite scores.

Although many of the unique variables added to Models 2 – 10 were found to be significant predictors of student composite scores (Table C.12,), only the results for Model 11 will be discussed here. To aid in the interpretation of the results for Model 11, we have summarized the results for variables in Model 11 that were statistically significant and, for any nominal variables, included sample size and average composite scores data for each (Table C.10). In Table C.10, variables that had a positive influence on student composite scores have a <u>positive coefficient</u> (i.e., they are related to an increase in average composite score), while variables that had a negative influence on student composite scores have a <u>negative coefficient</u> (i.e., they are related to a decrease in these scores).

Table C.12

Results of HLM Analyses of the Influence of Selected Variables on Student Composite Scores for the Phase 2, Grade 7 Sample (n = 1,574)

Fixed Effects	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Intercept - Students	150.092***	141.042***	143.404***	147.959***	146.409***	147.789***	147.578***	146.983***	156.236***	158.722***	158.699***
Instr. Sites, Field Sites		11.663***	9.550	8.015***	8.527***	7.777***	8.041***	7.589***	3.446	4.433	4.722
Teacher Certification, Subject:											
• Multiple (vs. None)			- 0.069	0.909	1.121	0.347	0.444	0.121	- 0.457	- 1.392	- 0.120
• Other (vs. None)			14.857	11.840	10.612	9.302	9.034	4.018	- 5.255	- 5.017	- 0.093
• English/Lang. Arts (vs. None)			10.915	17.479*	16.849*	15.469*	15.390*	11.566	10.695	4.369	2.655
• Social St./History (vs. None)			12.690	12.152	12.553	12.378	12.292	13.190	11.615	9.377	9.704
• Science (vs. None)			- 6.338	- 6.760	- 6.555	- 6.983	- 6.899	- 6.684	- 6.857	- 8.173*	- 7.976
Instr. Sites., Library/Media Ctr.				- 9.581***	- 10.180***	- 10.109***	- 10.028***	- 11.490***	- 8.396***	- 9.780***	-11.785***
Instr. Method, Discussion					2.263	- 3.111	- 2.570	- 4.053	- 7.894**	- 10.217***	-11.488***
Prog. Goal, Affective Dispositions						2.944	1.261	1.987	10.097*	9.543*	9.029
Prog. Goal, Service/Action Skills							1.627	1.408	- 3.119	- 1.726	- 0.416
Instr. Method, Coop. Learning								5.827**	6.407**	8.987	12.373***
Instr. Site, Science Labs									- 11.496***	- 9.965***	-10.800***
# EE Inservices, 1-2 Days										- 0.713**	- 0.608*
Instr. Method, Service Learning											- 10.422
Random Effects											
Intercept - Schools	168.663***	125.894**	107.386**	91.063***	91.091***	90.957***	89.811***	85.610***	74.639***	104.854***	96.827***
Residuals	806.803	813.889	814.922	809.322	809.183	809.104	809.212	808.980	808.584	759.118	759.618
Model Fit Statistics											
Degrees of Freedom	3	4	9	10	11	12	13	14	15	16	17
Log Likelihood	15,067.334	14,460.695	14,122.333	13,019.384	13,019.160	13,019.000	13,018.932	13,017.623	13,014.363	10,259.400	10,258.657
Difference in Deviances		606.639***	338.362***	1,102.949***	0.224	0.160	0.068	1.309	3.260	2,754.630***	0.743
AIC	15,073.334	14,468.695	14,140.333	13,039.384	13,041.160	13,043.000	13,044.932	13,045.623	13,044.363	10,291.400	10,292.657
BIC	15,089.418	14,489.975	14,188.000	13,091.544	13,098.536	13.105.592	13,112,740	13,118.646	13,122.603	10,371.141	10,377.382

Note: Levels of statistical significance: * p < .05, ** p < .01 *** p < .001

For the Phase Two seventh-grade sample, only one variable had a significantly positive influence on average composite scores: under Instructional Methods, the use of *cooperative learning*: coefficient = 12.373 ($p < .001$). In the presence of all 10 variables in Model 11, when cooperative learning was checked (n=11), the average composite score was 154.006, and when it was not checked (n=17), the average composite score was almost 12.4 points lower (Table C.10). These results and sample sizes indicates that this pattern was reasonably pervasive in this sample.

However, four variables had a significantly negative influence on student composite scores. They were:

- under Instructional Methods, the use of *discussion*: coefficient = - 11.488 ($p < .001$);
- under Instructional Sites, the use of *libraries and media centers*: coefficient = - 11.785 ($p < .001$);
- under Instructional Sites, the use of *science labs*: coefficient = - 10.800 ($p < .001$); and
- under the # of EE Inservices completed by Teachers lasting *1-2 days*: coefficient = - 0.608 ($p < .001$).

Data in Table C.10 also can be used to further interpret the results for the first three variables listed above. First, in the presence of all 10 variables in Model 11, when the use of *discussion* was checked as an Instructional Method (n=12), the average composite score was 142.076, and when it was not checked (n=15), the average composite score was almost 11.5 points higher. Second, when the use of *libraries and media centers* was checked as an Instructional Site (n=9), the average composite score was 141.927, and when it was not checked (n=19), the average composite scores was 11.8 points higher. Finally, when the use of *science labs* was checked as an Instructional Site (n=20), the average composite score was 142.420, and when it was not checked (n=9), the average composite score was 10.8 points higher. These results and sample sizes indicate that the pattern for each of these three variables was reasonably pervasive in this sample.

Results for the Phase Two eighth-grade sample. The results of the variable selection process are summarized in Table C.13. In **Step 2**, ANOVAs were used to select the 15 variables with the largest F values and the 15 variables with the smallest probability (p) values. The F values for these 15 variables ranged from a high of 8.19 (Instructional Site, *library/media center*) to a low of 4.345. Similarly, the p values for these 15 variables ranged from a low of $p < .000032$ (Total Enrollment), to .012. When these two sets of selected variables were combined, a total of 19 variables had been selected. Of these, 5 were School variables, 7 were Program variables, and 7 were Teacher variables, (i.e., none were Student variables).

In **Step 3**, Neural Network and MLP procedures were used to reduce this set of 20 variables to the final set of 10 variables to be included in HLM analyses. This set of final variables included 2 School variables, 4 Program variables, and 4 Teacher variables. Of these, the variables found to have the largest influence on student composite scores were: under # EE Higher Education Courses, *methods* (ranked first), and, Instructional Sites, *school grounds* (ranked second).

Table C.13
*Results of Analyses Using the Full Variable Set: Selection of Variables from Phase 2 Grade 8 Related to Individual Student Composite Scores (n= 690)**

Variables	Type of Variable			Step 2: **Mean Composite Scores** as DV: 15 Most Significant Predictors		**Step 3: Neural Network Analysis, Multilayer Perceptron Method:** 10 Most Important Variables	
	School	Program	Teacher	F Value	p Value	Ranking	Sensitivity Index
Total Enrollment	X			(13) 4.807	(1) .000032		
% Native American Students	X			(12) 4.843	(8) .000800	7	1.015
% Hispanic Students	X			(9) 5.121	(4) .000145		
% Black Students	X			(7) 5.747	(2) .000039		
% Federal IDEA Students	X			(15) 4.345	(12) .005538	10	1.000
Program Intensity		X			(7) .000545		
Instr. Method, Projects		X		(5) 6.255		3	1.030
Instr. Method, Service Learning		X		(4) 6.255		5	1.022
Instr. Site, Lib./Media Center		X		(1) 8.191	(11) .004709	4	1.024
Instr. Site, School Grounds		X		(8) 5.391		2	1.040
Assess. Methods, Ranked 2nd		X		(11) 4.926	(13) .007689		
Assess. Methods, Ranked 4th		X		(14) 4.644	(14) .010121		
Highest Degree, Ed. Level			X	(10) 5.028	(5) .000163	6	1.018
Teacher Age			X	(3) 6.831	(6) .000166		
Years Teaching, Total			X		(10) .001286	9	1.005
Years Teaching, Middle School			X		(9) .001116	8	1.006
# EE Higher Ed. Courses, Methods			X	(6) 5.907		1	1.071
# EE Inservices, 1-2 Days			X	(2) 7.349	(3) .000082		
# EE Inservices, 3-7 Days			X		(15) .012232		

Notes: (Step 2) ANOVAs were used because Student Composite Score is a continuous variable. All p values were < .02
(Step 3) In Statistica, a bootstrapping approach to Multilayer Perceptron (MLP) Methods was used to identify the most important predictors among those identified in Step 2 (1,000 bootstrapped samples).

Table C.14

Results of HLM Analyses of the Influence of Selected Variables on Student Composite Scores for the Phase 2, Grade 8 Sample (n= 690)

Fixed Effects	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7	Model 8	Model 9	Model 10	Model 11
Intercept - Students	154.720***	152.382***	155.069***	155.070***	154.996***	176.718***	187.293***	190.278***	190.406***	188.476***	195.735***
# EE HIED Courses, Methods		- 2.973	- 2.987	- 3.001	- 2.576	- 2.408	- 1.323	- 1.725	- 1.672	- 3.194	- 7.074
Instr. Site, School Grounds			- 3.083	- 2.537	- 1.322	- 15.946*	- 20.128*	- 20.751**	- 20.678*	- 8.611	- 9.283
Instr. Method, Projects				- 0.829	- 2.556	- 2.512	1.657	- 1.153	- 1.246	5.786	11.997
Instr. Site, Library/Media Ctr.					- 5.385	- 6.969	- 4.021	- 3.891	- 4.030	- 9.869*	- 11.890
Instr. Method, Service Learning						- 22.003**	- 24.448**	- 25.253**	- 25.329**	- 5.420	- 2.678
Highest Degree, Ed. Level							- 2.277	- 1.999	- 1.968	- 3.238*	- 5.183*
% Native American Students								- 1.763*	- 1.746*	- 7.998***	- 9.928
Years Teaching, Middle School									- 0.018	0.736	1.079
Year Teaching, Total										- 0.843*	- 0.840
% Federal IDEA Students											- 0.225
Random Effects											
Intercept - Schools	172.700***	174.868***	173.714***	173.388***	168.110***	92.883**	119.272	106.150	106.150	2.112	0.0
Residuals	798.853	798.131	798.141	798.164	798.125	801.159	793.390	794.363	794.363	834.986	841.355
Model Fit Statistics											
Degrees of Freedom	3	4	5	6	7	8	9	10	11	12	13
Log Likelihood	6609.233	5458.249	5458.164	5458.153	5457.695	5452.086	5449.780	5448.985	5448.979	4726.416	2575.108
Deviance		1150.984***	0.085	0.011	0.458	5.609	2.306	0.795	0. 006	722.563***	2151.308***
AIC	6615.233	5466.249	5468.164	5470.153	5471.695	5468.086	5467.780	5468.985	5470.979	4750.416	2601.108
BIC	6628.843	5483.631	5489.892	5496.277	5502.115	5502.851	5506.891	5512.441	5518.781	4800.847	2647.839

Note: Levels of statistical significance: * p < .05, ** p < .01 *** p < .001

In **Step 5**, these 10 variables were included in HLM analyses (Table C.14). Each variable was entered into a separate model for analysis on the basis of variable rankings included in Table C.13. Thus, the variable ranked first was included in Model 2, the variables ranked first and second were included in Model 3, and so on, until all 10 variables had been included in Model 11. As described in the Methods section for **Step 4**, all of these HLM analyses included two levels: students (fixed effects) and schools (random effects).

For the Phase Two eighth-grade sample (Table C.14), no variable was found to have a significantly positive influence on these student composite scores. Further, only one variable was found to have a significantly negative influence on their composite scores; i.e., under Highest Degree Earned, the *educational level* associated with that degree (e.g., a Bachelor's in Elementary Education, or a Master's in Secondary Science Education). The coefficient for this variable was – 5.183 ($p < .05$).

Comparison across grades. When the results for the Phase Two sixth-grade, seventh-grade, and eighth-grade sample were compared, only one variable appeared in the significant results for more than one of these grades (Table C.10). This variable fell under Instructional Sites: *science labs*. It was found to have a significantly negative influence on average composite scores for the Phase Two sixth-grade and seventh-grade samples.

Results for Research Question 3c

These results follow from a basic descriptive comparison of the results of the analyses of Phase One data (Table C.5) to the results from the analyses of Phase Two data (Table C.10). For this research question, we highlight the apparent similarities and differences in these two sets of results.

A careful comparison of the results in Table C.5 (Phase 1) and Table C.10 (Phase 2) indicated that there were few similarities and few differences in these two sets of results. First, only two School variables were included among the statistically significant variables in Phase One (i.e., under School Composition, the *percent of black* and the *percent of ESOL students*) and in Phase Two (i.e., under School Composition, the *percent of Asian* and the *percent of white students*). For all four of these variables, Model 11 coefficients were less than $\pm$ 1.0. Due to the small number of variables and the magnitude of these coefficients, School variables did not have as much influence on student composite scores as did Program and Teacher variables.

Second, the only Student variable included among these statistically significant variables was Student Age, and it was found among the results for Phase Two, but not for Phase One.

Third, EE Program Type was included in the results for Phase One (Table C.5) and for Phase Two (Table C.10). For the Phase One eighth-grade sample, two specific EE Program Types were found to have a significantly positive influence on student composite scores (i.e., having *an environmental curriculum*, and having both *an environmental curriculum and an environmental club*). This was an interesting finding because the Phase One sample was a national baseline sample generated on a stratified random basis, and only 18 of the 48 schools in this sample reported having any type of EE program. On the other hand, the Phase Two sample was a nationally purposive sample of 64 schools with a stable environmental program in the middle grades. For the Phase Two sixth-grade sample, specific EE Program Types were found to have a positive or a negative influence on student composite scores. Thus, there were differences between the Phase One and Phase Two samples, as well as in the manner in which the results pertaining to EE Program Type were generated by the statistical program for each sample (i.e., Phase One: comparison against schools with *no program*; and Phase Two: comparison against schools with only *an environmental curriculum*). <u>Despite this, it is interesting to note that schools with both *an environmental curriculum and an environmental club* had a significantly positive influence on student composite scores in the Phase One eighth-grade and the Phase Two sixth-grade sample.</u>

<u>Finally</u>, Teacher Certification variables were included in the results for Phase One (Table C.5) and for Phase Two (Table C.10). <u>For the Phase One sixth-grade sample, *working on* a teacher certification was found to have a significantly negative influence on this sample's student composite scores. A more careful review of those data indicated that all of the teachers who were working on teacher certification were teaching in private schools</u>. For the Phase Two sixth-grade sample, the Teacher Certification variables focus on the kind of certificate these teachers held rather than on the status of their efforts to earn a certificate. In specific<u>, it was found that the students of Phase Two sixth-grade teachers who held only an *elementary*, a *middle*, or a *secondary* certificate had lower average composite scores than did students of teachers who held teacher certifications at two or more of these levels (i.e., *multiple*).</u>

Appendix D

Detailed Description of Methods and Results for Research Question Four

Appendix D.

Detailed Description of
Methods and Results for Research Question Four

Research Question 4 focused on the extent to which School, Program, Teacher, and Student characteristics may differentiate between high- and low-performing schools in the Phase Two sample. For this research question, high-performing schools were those whose mean *composite score* fell in the top quartile (Quartile 1, or Q1) for that grade, and low-performing schools were those whose mean *composite score* fell in the bottom quartile (Quartile 4, or Q4) for that grade. One specific research question guided these analyses.

> **4a.** Which school, program, teacher, and student characteristics appear to differentiate between the high- and low-performing schools, as determined by the distribution of *composite scores*, at the 6th grade level, at the 7th grade level, and at the 8th grade level?

Methods for Specific Research Question 4a

When Research Question 4a was written, the research team planned to include all relevant School (SIF), Program (PIF), Teacher (TIF) and Student (MSELS) variables in one analysis for each grade in the Phase Two sample. These analyses are different than those undertaken for other research questions in three important ways. First, unlike the exploratory analysis of separate sets of variables for research questions in Appendix G, the intent here was to determine which of the independent variables from all four variable sets would remain as significant predictors. Second, unlike the exploratory analyses for the research question in Appendix F in which student *skill component scores* served as the dependent variable, here student *component scores* were used as the dependent variable. Third, unlike the analyses for Overarching Research Question 3, which included all students in each grade in the Phase Two sample, here all data files and analyses were restricted to include only high- and low-performing schools (Q1 and Q4) in these grades and Phases.

The first step in preparing data files for these analyses involved the use of mean student scores to assign schools to quartiles. This was done following procedures presented by Setek and Gallo (2009, Figure 8.43). A summary of the results of this quartiling for both research questions is presented in Table D.1.

Once the schools that fell within Quartile 1 and Quartile 4 had been identified, the data files needed to conduct these analyses could be prepared. While the research team had intended to include all School, Program, Teacher, and Student variables in a single data file, this was not possible due to the loss of teacher data in Phase Two. As a result, two separate data files were prepared for each grade in Phase Two: (a) one file that contained school, program, and student data, along with dependent variable scores; and (b) a second file that,

Table D.1

Summary of Results of Quartiling of Phase 2 Schools on the Basis of Composite Scores, by Grade

Research Question/ Dependent Variable	Grade	Number of Schools Total	# per Quartile	Range of Mean Composite Scores, by Quartile Quartile 4 (Q4)	Quartile 1 (Q1)
Composite Scores	6	43	10	124.01 – 139.82	162.13 – 177.57
(range 0 – 240)	7	40	10	119.13 - 142.96	162.16 – 183.52
	8	33	8	124.80 – 141.36	165.64 – 185.25

contained only teacher data, along with those same scores.

Once these files had been prepared, the research team faced two concerns. The first concern had to do with the number of variables. These data files were very large because they included data from more than 50 items in the School, Program, and Teacher Information Forms, and the first section of the MSELS. Further, for each item, there were as many as 8 possible responses, each of which had to be coded separately in these data sets. The statistical software program treated each possible response as a separate variable, so the number of possible predictor variables in each data file was very large (e.g., the number of items X the number of responses), and far greater than the number of schools to be included in each analysis. The second concern had to do with the unit of analysis. For these research questions, the research team had planned to conduct analyses using individual student scores as the unit of analysis. However, as presented in Table D.1 school mean scores were used to select high- and low- performing schools in each grade.

To address these concerns, preliminary analyses were conducted (a) for the purpose of substantially reducing the number of potential predictor variables, and (b) using both individual student scores and school quartiles as dependent variables. In specific, two preliminary analyses were conducted for each data file (i.e., the team ran two analyses of the file containing School, Program, and Student variables, and two analyses of the file containing Teacher variables). The first type of analysis involved the use of ANOVA tests to select important independent variables as likely predictors of individual student scores as the dependent variable. The second type of analysis involved the use of chi-square tests to select important independent variables as likely predictors of each set of school quartiles.

The results of these analyses were used to narrow the very large number of independent variables by removing redundant and less relevant independent variables. These results reflect different dependent variable measures (quartiles based on school mean scores and individual student scores) and for each measure, different methods for ranking and selecting predictor variables (magnitude and significance). In specific, we used ANOVA results to select the 10 variables with the largest F values and the 10 variables with the smallest p value, as well as Chi-square results to select the 10 variables with the largest χ^2 values and the 10 variables with the smallest p value. These predictor variables were combined into one list (i.e., there was one narrowed set of School, Program and Student variables, and one narrowed set of Teacher variables).

The final step in data analysis involved the use of XLSTAT (2013) to conduct Partial Least Squares – Discriminant Analysis (PLS-DA) for each of these 12 narrowed data sets (Vinzi et al., 2010). The purpose of PLS-DA is to identify the variables in each data set that are most powerful in discriminating between high- and low-performing schools. As in the preliminary analyses, significance and magnitude were used to identify which variables in each data set did so. For significance, *Variables Important to Projection* (VIP) Eigenvalues were used to estimate the importance of each variable in predicting whether schools would be classified as high or low performing. Further, for magnitude, variable coefficients in each discrimination model were used to estimate the size of the effect of each independent variable on predicting whether schools would be classified as high or low performing.

Results for Research Question 4a

In the preliminary analyses, both quartiles, based on school mean *composite scores*, and individual student *composite scores*, were used as dependent variables. The narrowed set of independent variables derived from these analyses for use in these PLS-DA analyses are summarized in Table D.2.

Table D.2

Variables Selected by Preliminary Analyses for Inclusion in PLS-DA Analyses, by Grade

			Grade		
Variable Set	Source	Variable	6	7	8
School, Program, and Student	School Info. Form	Census Region X	X		
		Total Enrollment	X	X	X
		Student : Teacher Ratio	X	X	X
		% Asian Students	X	X	X
		% Native American Sts.	X		X
		% Black Students	X	X	X
		% White Students		X	X
		% Free Lunch Students	X	X	X
		% Reduced Lunch Sts. X	X	X	
		% Special Needs Students	X	X	X
	Screening Survey	Program Duration		X	
		Program Intensity		X	X
	Program Info. Form	Curriculum Organization	X	X	X
		Goal: Prob./Issue Aware.			X
		Goal: Investigation Skills			X
		Organization of Teachers		X	X
		Instructional Grouping	X	X	X
		Instr. Method: InquiryX			
		Instr. Method: Labs	X		
		Instr. Method: Coop. Lrn.		X	
		Instr. Method: Hands-On			X
		Instr. Method: Service-Lrn.			X
		Instr. Sites: Sci. Labs	X		
		Instr. Sites: School Grounds	X	X	
		Instr. Sites: Library/Media		X	

		Instr. Sites: Other Comm.	X		X
		Assessment Methods	X	X	

Table D.2 (cont.)

Variable Set	Source	Variable	Grade 6	Grade 7	Grade 8
Teacher	Teacher Info. Form	Teacher Age	X	X	X
		Teacher Gender	X	X	X
		Highest Degree: Earned	X	X	X
		Highest Degree: Ed. Level	X		X
		Highest Degree: Subject	X	X	
		Teacher Certification: Y/N	X	X	
		Teacher Cert.: Ed. Level	X	X	
		Teacher Cert.: SubjectX	X	X	
		Years Teaching: Total X	X	X	
		Years Teach. : Middle Sch.	X	X	X
		# EE Inservices: Total X	X	X	
		# EE Inservices: < 1 day	X	X	
		# EE Inservices: 1-2 days	X	X	X
		# EE Inservices: 3-7 days	X	X	X
		Importance of EE for Sts.	X		
		Importance of EE to Tch.	X	X	
		Perceived Level of Particip.	X	X	X

One of the noticeable results of the preliminary analyses for Grades 6, 7, and 8 was that none of the Student variables were found to discriminate between high- and low-performing schools; only School, Program, and Teacher variables did so (Table D.2).

Results for school and program variables. Table D.3 presents results for the PLS-DA analysis of these variables found to be significant (VIP Eigenvalues = or > 1.0) and variables with a substantial magnitude (model coefficients = or > 0.10). For these analyses, schools served as the unit of analysis.

In the PLS-DA analyses of the **sixth-grade sample**, Table D.3 indicates that variables associated with six items in the School Information Form (SIF) and with three items in the Program Information Form (PIF) were included as significant VIP variables. A total of 11 variables associated with these nine items were found to have VIP Eigenvalues greater than 1 for both Q1 and Q4 schools. The variables with the greatest VIP Eigenvalues were: under Instructional Settings, *science labs*; under School SES, the percent of students in *Free Lunch* and in *Reduced*

Lunch Programs; and under the rankings for Assessment Methods, ranking *alternative assessment* as the second most commonly used method.

Table D.3 also indicates that a total of seven variables associated with one item on the SIF and three items on the PIF were included as variables with a substantial discrimination model coefficient (SIF: Region; PIF: Organization of Teachers, Instructional Settings, and Assessment Methods). Of these seven variables, four had positive coefficients and were more characteristic of high-performing (Q1) schools, while three had negative

Table D.3

*Selected School and Program Variables that Differentiate Composite Score Quartile 1 from Quartile 4 Schools in the Phase 2 Sample, By Grade**

		Discriminant Analysis Result: Variable Importance (VIP) Eigenvalues and Model Coefficients*								
		Grade 6 (n =20 schools)			**Grade 7 (n =20 schools)**			**Grade 8 (n = 16 schools)**		
Items and Variables	**Source of Data**	Variable Importance		DA Coefficient	Variable Importance		DA Coefficient	Variable Importance		DA Coefficien t
		Q1 (n)	Q4 (n)	Top Q/Q1	Q1 (n)	Q4 (n)	Top Q/Q1	Q1 (n)	Q4 (n)	Top Q/Q1
Census Region	School Form									
* Midwest		1.217 (2)	1.348 (0)							
* South		1.220 (6)	1.159 (7)							
* West				+ 0.142		1.422 (3)	- 0.184			
Total Enrollment	School Form	1.244 (11)	1.203 (10)					1.718 (9)	1.589 (8)	
School Ethnicity	School Form									
* % Black Students		1.024 (10)			1.279 (8)	1.194 (10)		1.431 (8)	1.391 (7)	
* % White Students					1.030(10)			1.436 (9)	1.318 (8)	
School SES	School Form									
* % Free Lunch Students		1.931 (7)	1.850 (9)		1.798 (5)	1.511 (10)		2.245 (6)	2.057 (8)	
* % Reduce Lunch Students		1.751 (6)	1.658 (10)		1.803 (3)	1.502 (10)		1.869 (4)	1.859 (8)	
School Special Populations	School Form									
* % Special Needs Students		1.190 (6)	1.171 (6)							
Student : Teacher Ratio	School Form							1.194 (8)	1.414 (7)	
Curriculum Organization:	Program Form									
* Separate Subjects								1.093 (1)	1.035 (2)	
***** Common Themes/Sep. Subj.						1.595 (3)	- 0.129			
***** Common Themes/Integrated						1.042 (3)				
* Other							+ 0.194			+ 0.165
Organization of Teachers:	Program Form									
* Self-Contained		1.045 (2)	1.039 (1)	**+ 0.211**						
* Departmentalized			1.109 (6)		1.662 (1)	1.489 (5)		1.079 (2)	1.022 (3)	
* Teaming						1.023 (2)		1.628 (4)	1.492 (1)	**+ 0.120**
* Other					1.096 (1)	1.032 (0)	**+ 0.247**	1.049 (1)		+ 0.149

* Combination (SC, D, T, O)				- 0.111						- 0.114
Table D.3 (cont.)										
		Discriminant Analysis Result: Variable Importance (VIP) Eigenvalues and Model Coefficients								
		Grade 6 (n =20 schools)			**Grade 7 (n =20 schools)**			**Grade 8 (n = 16 schools)**		
Variables & Values	**Source of Data**	Variable Importance		DA Coefficient	Variable Importance		DA Coefficient	Variable Importance		DA Coefficient
Instructional Groups, Ranked:	Program Form									
* Ranked #1: Whole Class					1.335 (2)	1.111 (6)				
* Ranked #2: Whole Class					1.372 (7)	1.142 (2)				
* Ranked #1: Groups/Teams					2.296 (7)	1.914 (2)	**+ .135**			
* Ranked #2: Groups/Teams					1.294 (2)	1.078 (5)				
* Ranked #2: Not Checked										- 0.100
Instructional Settings:	Program Form									
* Science Labs		2.145 (2)	1.995 (7)	**- 0.171**						
* School Grounds		1.313 (9)	1.228 (9)	**- 0.118**						
* Library/Media Center					1.436 (2)	1.197 (6)				
Instructional Methods:										
* Cooperative Learning					1.488 (7)	1.374 (3)				
* Service-Learning								1.780 (0)	1.637 (3)	
Assessment Methods Ranked:	Program Form									
* Ranked #2: Traditional			1.031 (0)	+ 0.223	1.404 (0)	1.205 (3)				
* Ranked #3: Traditional					1.285 (2)	1.084 (1)				
* Ranked #2: Alternative		1.598 (5)	1.519 (1)	**+ 0.102**						
* Ranked #3: Alternative					1.321 (0)	1.318 (3)	**- 0.107**			
* Ranked #2: Informal		1.342 (2)	1.269 (5)							
* Ranked #3: Informal					1.026 (1)	1.094 (0)	**+ 0.290**			
* Ranked #3: Not Checked							+ 0.100			

* **Notes**: Only variables with Variable Importance (VIP) Eigenvalues equal to or greater than 1.0, and a model coefficient equal to or greater than 0.10 were included in this table. For each variable with a VIP Eigenvalue and a model coefficient at or above this level, the model coefficient has been bolded.

coefficients and were more characteristic of low-performing (Q4) schools.

However, for the Phase Two sixth-grade sample, four Program variables had both VIP Eigenvalues greater than 1.0 and discrimination model coefficients greater than 0.10. Two variables had positive coefficients and two had negative coefficients. The two variables with positive model coefficients were: under Organization of Teachers, teaching in *self-contained* classrooms (coefficient = + 0.211); and under Assessment Methods, ranking *alternative assessment* second (coefficient = + 0.102). These are the two best School and Program indicators of high-performing schools in this sample. Further, the two variables with negative model coefficient were: under Instructional Settings, use of science labs (coefficient = - 0.171) and use of school grounds (coefficient = - 0.118). These are the two best School and Program indicators of low-performing schools in this sample.

Finally, these results from the PLS-DA analyses for the Phase Two sixth-grade sample can be viewed in a larger context, notably the ability of the discrimination model, which included the variables in Table D.2, in accurately classifying individual students into the Q1 and Q4 subsample of schools. The XLSTAT software used to conduct these PLS-DA analyses allowed the research team to review a summary of the reclassification of a 70% random subsample of individual students into the Q1 and Q4 sample of schools as well as a 30% random subsample used to validate these results (Table D.4).

Table D.4
Reclassification Analysis Results: Phase 2 Sixth Grade 70% and 30% Random Subsamples Using Composite Scores (DV) and the School and Program Variable Model (IVs)

Sample*	From/To	Highest	Lowest	Totals	% Accurate
R 70%	Highest	297	20	317	93.69%
	Lowest	1	534	535	99.81%
	Totals	298	554	852	**97.54%**
R 30%	Highest	96	26	122	78.69%
	Lowest	16	228	244	93.44%
	Totals	112	254	366	**88.52%**

*** Note:** Under Sample, R = Random

For the 70% random estimation subsample, 97.54% of individual students were accurately classified into the highest (Q1) and lowest (Q4) subsample of sixth-grade schools. Further, for the 30% random validation subsample, 88.52% of individual students were accurately classified.

The difference in the accurate reclassification of students in the 70% and 30% subsamples was 9.02%. Overall, these values fall within the range that allows this discrimination model to be deemed valid. In turn, this supports the results of the PLS-DA analyses presented in Table D.3 and summarized in prose, above.

In the PLS-DA analyses of the **seventh-grade sample**, Table D.3 indicates that variables associated with three items on the School Information Form (SIF) and with six items on the Program Information Form (PIF) were included as significant VIP variables. A total of 15 variables associated with these nine items were found to have VIP Eigenvalues greater than 1 for both Q1 and Q4 schools. The variables with the greatest VIP Eigenvalues were: under Instructional Groups, the ranking of *groups/teams* first; under School SES, the percent of students in *Reduced Lunch* and in *Free Lunch* Programs; and under the Organization of Teachers, *departmentalized*.

Table D.3 also indicates that a total of eight variables associated with one item on the SIF and four items on the PIF were included as variables with a substantial discrimination model coefficient (SIF: Region; PIF: Curriculum Organization, Organization of Teachers, Instructional Groups, and Assessment Methods). Of these eight variables, five had positive coefficients and were more characteristic of high-performing (Q1) schools, while three had negative coefficients and were more characteristic of low-performing (Q4) schools.

However, for the Phase Two seventh-grade sample, four Program variables had both VIP Eigenvalues greater than 1.0 and discrimination model coefficients greater than 0.10. Three of these variables had positive coefficients and only one had a negative coefficient. The three variables with positive model coefficients were: under Assessment Methods, ranking *informal assessment* third (coefficient = + 0.290); under Organization of Teachers, the selection of *other* as a response (coefficient = + 0.247); and under Instructional Groups, ranking *groups/teams* first (coefficient = + 0.135). These are the three best School and Program indicators of high-performing schools in this grade sample. Further, the only variable with a negative model coefficient was: under Assessment Methods, ranking *alternative assessment third* (coefficient = - 0.107). This is the best indicator of low-performing schools in this sample.

Finally, these results from the PLS-DA analyses for the Phase Two seventh-grade sample can be viewed within the context of the ability of the discrimination model, which included the variables in Table D.2, to accurately classify individual students into the Q1 and Q4 subsample of schools. The results of the PLS-DA reclassification of 70% and 30% random subsamples of individual students into the Q1 and Q4 sample of schools are summarized in Table D.5.

For the 70% random estimation subsample, 97.54% of individual students were accurately classified into the highest (Q1) and lowest (Q4) subsample of sixth-grade schools. Further, for the 30% random validation subsample, 88.52% of individual students were accurately classified. The difference in the accurate reclassification of students in the 70% and 30% subsamples was 20.52%. Overall, these values fall within the range that allows this discrimination model to be deemed valid. In turn, this supports the results of the PLS-DA analyses presented in Table D.3 and summarized in prose, above.

Table D.5
Reclassification Analysis Results: Phase 2 Seventh Grade 70% and 30% Random Subsamples Using Composite Scores (DV) and the School and Program Variable Model (IVs)

Sample*	From/To	Highest	Lowest	Totals	% Accurate
R 70%	Highest	137	2	139	98.56%
	Lowest	0	649	649	100.00%
	Totals	137	651	788	**99.75%**
R 30%	Highest	60	9	69	86.96%
	Lowest	61	207	268	77.24%
	Totals	121	216	337	**79.23%**

* **Note:** Under Sample, R = Random

In the PLS-DA analyses of the **eighth-grade sample**, Table D.3 indicates that variables associated with four items on the School Information Form (SIF) and with three items on the Program Information Form (PIF) were included as significant VIP variables. A total of 10 variables associated with these seven items were found to have VIP Eigenvalues greater than 1 for both Q1 and Q4 schools. The variables with the greatest VIP Eigenvalues were: under School SES, the percent of students in *Free Lunch* and in *Reduced Lunch* Programs; under Instructional Methods, *service-learning;* and *Total Enrollment*.

Table D.3 also indicates that five variables associated with three items on the PIF had substantial discrimination model coefficients (Curriculum Organization, Organization of Teachers, and Instructional Groups). Of these five variables, three had positive coefficients and were more characteristic of high-performing (Q1) schools, while two had negative coefficients and were more characteristic of low-performing (Q4) schools.

However, for the Phase Two eighth-grade sample, only one Program variable had both VIP Eigenvalues greater than 1.0 <u>and</u> discrimination model coefficients greater than 0.10. This variable fell under Organization of Teachers, *teaming,* and it had a positive model coefficient (coefficient = + 0.120). This was the best School and Program indicator of <u>high</u>-performing schools in this sample.

Finally, these results from the PLS-DA analyses for the Phase Two eighth-grade sample can be viewed within the context of the ability of the discrimination model, which included the variables in Table D.2, to accurately classify individual students into the Q1 and Q4 subsample of schools. The results of the PLS-DA reclassification of 70% and 30% random subsamples of individual students into the Q1 and Q4 sample of schools are summarized in Table D.6.

Table D.6

Reclassification Analysis Results: Phase 2 Eighth Grade 70% and 30% Random Subsamples Using Composite Scores (DV) and the School and Program Variable Model (IVs)

Sample*	From/To	Highest	Lowest	Totals	% Accurate
R 70%	Highest	127	16	143	88.81%
	Lowest	20	335	355	94.37%
	Totals	147	351	498	**92.77%**
R 30%	Highest	49	9	58	84.48%
	Lowest	12	144	156	92.31%
	Totals	61	153	214	**90.10%**

* **Note:** Under Sample, R = Random

For the 70% random estimation subsample, 92.77% of individual students were accurately classified into the highest (Q1) and lowest (Q4) subsample of eighth-grade schools. Further, for the 30% random validation subsample, 90.10% of individual students were accurately classified. The difference in the accurate reclassification of students in the 70% and 30% subsamples was 2.67%. Overall, these values fall within the range that allows this discrimination model to be deemed valid. In turn, this supports the results of the PLS-DA analyses presented in Table D.3 and summarized in prose, above.

Results for teacher variables. Table D.7 presents results for the PLS-DA analysis of Teacher variables found to be significant (VIP Eigenvalues = or > 1.0) and variables with a substantial magnitude (model coefficients = or > 0.10). For these analyses, individual teachers served as the unit of analysis.

In the PLS-DA analyses of the **sixth-grade sample**, Table D.7 indicates that variables associated with seven items in the Teacher Information Form (TIF) were included as significant VIP variables. Responses to Highest Degree Earned and Teaching Certification were rich in detail, so these responses to each were coded three ways and treated as if they were three sub-items.

A total of 16 variables associated with these items and sub-items were found to have VIP Eigenvalues greater than 1 for both Q1 and Q4 schools. The variables with the greatest VIP Eigenvalues were: under Highest Degree Earned - Subject, *science*; under teachers' Perceived Level of Participation in environmental protection, *slightly*; and under Teacher Certification - Subject, *social studies/history*.

Table D.7 also indicates that a total of 19 variables associated with four items on the TIF were included as variables with a substantial discrimination model coefficient (Teacher Age; Highest Degree Earned: all three sub-items; Teacher Certification: all three sub-items; Perceived Level of Participation). Of these 19 variables, 14 had positive coefficients and were more characteristic of high-performing (Q1) schools, while 5 had negative coefficients and were more characteristic of low-performing (Q4) schools.

Table D.7

*Selected Teacher Variables that Differentiate Composite Score Quartile 1 from Quartile 4 Schools in the Phase 2 Sample, By Grade**

		Discriminant Analysis Results: Variable Importance (VIP) Eigenvalues and Model Coefficients*								
		Grade 6 (n=24)			**Grade 7 (n=24)**			**Grade 8 (n=10)**		
Items and Variables	**Source of Data**	Variable Importance		DA Coefficients	Variable Importance		DA Coefficients	Variable Importance		DA Coefficients
		Q1 (n)	Q4 (n)	Top Q/Q1	Q1 (n)	Q4 (n)	Top Q/Q1	Q1 (n)	Q4 (n)	Top Q/Q1
Teacher Age	Teacher Form									
21-30							+ 0.117	1.102 (3)	1.079 (0)	**+ 0.162**
31-40				- 0.128	1.110 (4)	1.011 (1)	**+ 0.116**	1.027 (2)	1.005 (0)	**+ 0.106**
41-50										+ 0.160
51-60		1.436 (3)	1.214 (0)	**+ 0.159**	1.836 (2)	1.824 (3)	**- 0.189**			
> 60				+ 0.214	1.719 (2)	1.546 (0)	**+ 0.247**	1.925 (0)	1.849 (2)	**- 0.200**
Teacher Gender:										
Male			1.271 (4)							
Female			1.271 (7)							
Years Teaching:										
Middle School		1.353 (13)	1.232 (11)					1.938 (7)	1.914 (3)	
Total		1.544 (13)	1.361 (11)					1.021 (7)		
Highest Degree Earned	Teacher Form									
Bachelor’s		1.511 (4)	1.283 (8)					1.487 (3)	1.453 (3)	**- 0.125**
Master’s			1.031 (2)					1.487 (4)	1.453 (0)	**+ 0.125**
Master’s + 30		1.117 (5)	1.215 (1)	**+ 0.143**	1.073 (0)	1.000 (2)				
Doctorate							+ 0.205			
Highest Degree, Ed. Level	Teacher Form									
Elementary				+ 0.108						+ 0.163
Middle										
Multiple								1.611 (0)	1.544 (1)	**- 0.159**
Other			1.006 (4)						1.055 (1)	
Highest Degree, Subject:	Teacher Form									
Science		1.819 (1)	1.550 (5)	**- 0.102**		1.096 (6)	- 0.117			
Social Studies/History		1.321 (1)	1.398 (0)	**+ 0.292**						
English/LA/Reading				+ 0.161						
Multiple				+ 0.139	1.254 (1)	1.426 (0)	**+ 0.456**			
Other			1.242 (0)	+ 0.365						

None			1.039 (6)							

Table D.7 (cont.)

		Grade 6 (n=24)			**Grade 7 (n=24)**			**Grade 8 (n=10)**		
Items and Variables	**Source of Data**	Variable Importance		DA Coefficients	Variable Importance		DA Coefficients	Variable Importance		DA Coefficients
		Q1 (n)	Q4 (n)	Top Q/Q1	Q1 (n)	Q4 (n)	Top Q/Q1	Q1 (n)	Q4 (n)	Top Q/Q1
Teacher Certification:	Teacher Form									
Yes		1.632 (8)	1.409 (6)	**- 0.205**			- 0.205			
Working On				+ 0.254						
No		1.473 (1)	1.255 (0)	**+ 0.186**			+ 0.205			
Tch. Certification, Ed. Level:	Teacher Form									
Middle				+ 0.110			- 0.158			
Secondary		1.147 (0)	1.138 (2)		1.082 (0)	1.077 (3)				
Multiple					1.741 (9)	1.558 (5)	**+ 0.117**			
Tch. Certification, Subject:	Teacher Form									
Science		1.562 (2)	1.352 (5)		1.367 (3)	1.274 (8)				
Social Studies/History		1.604 (1)	1.595 (0)	**+ 0.313**	1.153 (1)	1.078 (0)	**+ 0.358**			
English/LA/Reading				- 0.189	1.216 (1)	1.089 (0)	**+ 0.295**			
Multiple		1.095 (1)		+ 0.110	1.156 (3)	1.104 (2)				- 0.157
Other							+ 0.177			
# EE Inservices:	Teacher Form									
Total					1.488 (9)	1.405 (11)				
< 1 Day		1.568 (7)	1.429 (1)		1.140 (7)	1.097 (4)				
1-2 Days					1.184 (8)	1.116 (6)				
3-7 Days		1.175 (7)	1.073 (1)		1.859 (7)	1.664 (2)	**+ 0.101**			
Perc. Level of Participation:	Teacher Form									
Slightly		1.700 (2)	1.559 (0)	**+ 0.245**	1.338 (1)	1.525 (0)	**+ 0.502**			
Moderately						1.095 (5)		1.224 (3)	1.172 (0)	**+ 0.142**
Considerably						1.051 (7)				
Extremely		1.069 (2)	1.229 (3)	**- 0.149**						- 0.167

* **Notes**: Only variables with Variable Importance (VIP) Eigenvalues equal to or greater than 1.0, and a model coefficient equal to or greater than 0.10 were included in this table. For each variable with a VIP Eigenvalue and a model coefficient at or above this level, the model coefficient has been bolded.

Further, Table E.7 (pp. 12-13) indicated that for the Phase Two sixth-grade sample, nine Teacher variables had both VIP Eigenvalues greater than 1.0 and discrimination model coefficients greater than 0.10. Six of these variables had positive coefficients and three had negative coefficients. The six variables with positive model coefficients were:

- under Teacher Certification - Subject, *social studies/history* (coefficient = + 0.313);
- under Highest Degree Earned – Subject, *social studies/history* (coefficient = + 0.292);
- under teachers' Perceived Level of Participation in environmental protection, *slightly* (coefficient = + 0.245);
- under Teacher Certification, *no* (coefficient = + 0.186);
- under Teacher Age, *41-50* (coefficient = + 0.159); and
- under Highest Degree Earned, *Master's + 30* (coefficient = + 0.143).

These are the best Teacher indicators of high-performing schools in this sample. Further, the three variables with negative model coefficients were: under Teacher Certification, *yes* (coefficient = - 0.205); under teachers' Perceived Level of Participation in environmental protection, *extremely* (coefficient = - 0.149); and under Highest Degree Earned – Subject, *science* (coefficient = - 0.102). These are the best Teacher indicators of low-performing schools in this sample.

Finally, these results from the PLS-DA analyses for the Phase Two sixth-grade sample can be viewed within the context of the ability of the discrimination model, which included the variables in Table D.2, to accurately classify individual students into the Q1 and Q4 subsample of schools. The XLSTAT software used to conduct these PLS-DA analyses allowed the research team to review a summary of the reclassification of a 70% random subsample of individual students into the Q1 and Q4 sample of schools as well as a 30% random subsample used to validate these results (Table D.8).

Table D.8

Reclassification Analysis Results: Phase 2 Sixth Grade 70% and 30% Random Subsamples Using Composite Scores (DV) and the Teacher Variable Model (IVs)

Sample*	From/To	Highest	Lowest	Totals	% Accurate
R 70%	Highest	167	32	199	83.92%
	Lowest	0	320	320	100.00%
	Totals	167	352	519	**93.83%**
R 30%	Highest	58	28	86	67.44%
	Lowest	0	137	137	100.00%
	Totals	58	165	223	**87.44%**

*** Note:** Under Sample, R = Random
For the 70% random estimation subsample, 93.83% of individual students were accurately classified into the highest (Q1) and lowest (Q4) subsample of sixth-grade schools. Further, for the 30% random validation subsample, 87.44% of individual students were accurately classified. The difference in the accurate reclassification of students in the 70% and 30% subsamples was 6.39%. Overall, these values fall within the range that allows this discrimination model to be deemed valid. In turn, this supports the results of the PLS-DA analyses presented in Table D.7 and summarized in prose, above.

In the PLS-DA analyses of the **seventh-grade sample**, Table D.7 indicates that variables associated with five items in the Teacher Information Form (TIF) were included as significant VIP variables. Responses to Highest Degree Earned and Teaching Certification were rich in detail, so these responses to each were coded three ways and treated as if they were three sub-items. A total of 16 variables associated with these items and sub-items were found to have VIP Eigenvalues greater than 1 for both Q1 and Q4 schools. The variables with the greatest VIP Eigenvalues were: under Number of EE Inservices, *3-7 days*; under Teacher Age, *51-60 years*, and *> 60 years*; and under Teacher Certification – Education Level, *multiple*.

Table D.7 indicates that a total of 16 variables associated with five items on the TIF were included as variables with a substantial discrimination model coefficient (Teacher Age; Highest Degree Earned: 2 of 3 sub-items; Teacher Certification: all 3 sub-items; Number of EE Inservices; and Perceived Level of Participation). Of these 16 variables, 12 had positive coefficients and were more characteristic of high-performing (Q1) schools, while 4 had negative coefficients and were more characteristic of low-performing (Q4) schools.

Further, for the Phase Two seventh-grade sample, nine Teacher variables had both VIP Eigenvalues greater than 1.0 <u>and</u> discrimination model coefficients greater than 0.10. Eight of these variables had positive coefficients and only one had a negative coefficient. The eight variables with positive model coefficients were:

- under teachers' Perceived Level of Participation in environmental protection, *slightly* (coefficient = + 0.502);
- under Highest Degree Earned – Subject, *multiple* (coefficient = + 0.456);
- under Teacher Certification - Subject, *social studies/history* (coefficient = + 0.358);
- under Teacher Certification - Subject, *English/language arts/reading* (coefficient = + 0.295);
- under Teacher Age, *> 60* (coefficient = + 0.247);
- under Teacher Certification – Education Level, *multiple* (coefficient = + 0.117);
- under Teacher Age, *31-40* (coefficient = + 0.116); and
- under Number of EE Inservices, *3-7 days* (coefficient = + 0.101).

These are the best Teacher indicators of high-performing schools in this sample. Further, the only variable with a negative model coefficient was: under Teacher Age, *51-60* (coefficient = -0.189). This is the best Teacher indicator of low-performing schools in this sample.
Finally, these results from the PLS-DA analyses for the Phase Two seventh-grade sample can be viewed within the context of the ability of the discrimination model, which included the variables in Table D.2, to accurately classify individual students into the Q1 and Q4 subsample of schools. The XLSTAT software used to conduct these PLS-DA analyses allowed the research team to review a summary of the reclassification of a 70% random subsample of individual students into the Q1 and Q4 sample of schools as well as a 30% random subsample used to validate these results (Table D.9).

Table D.9
Reclassification Analysis Results: Phase 2 Seventh Grade 70% and 30% Random Subsamples Using Composite Scores (DV) and the Teacher Variable Model (IVs)

Sample*	From/To	Highest	Lowest	Totals	% Accurate
R 70%	Highest	101	0	101	100.00%
	Lowest	0	328	328	100.00%
	Totals	101	328	429	**100.00%**
R 30%	Highest	31	8	39	79.49%
	Lowest	12	132	144	91.67%
	Totals	43	140	183	**89.07%**

*** Note:** Under Sample, R = Random

For the 70% random estimation subsample, 100% of individual students were accurately classified into the highest (Q1) and lowest (Q4) subsample of seventh-grade schools. Further, for the 30% random validation subsample, 89.07% of individual students were accurately classified. The difference in the accurate reclassification of students in the 70% and 30% subsamples was 10.93%. Overall, these values fall within the range that allows this discrimination model to be deemed valid. In turn, this supports the results of the PLS-DA analyses presented in Table D.7 and summarized in prose, above.

In the PLS-DA analyses of the **eighth-grade sample**, Table D.7 (pp. 12-13) indicates that variables associated with four items in the Teacher Information Form (TIF) were included as significant VIP variables, as well as two sub-items associated with Highest Degree Earned. A total of 8 variables associated with these items and sub-items were found to have VIP

Eigenvalues greater than 1 for both Q1 and Q4 schools. The variables with the greatest VIP Eigenvalues were: under Number of Years Teaching, at the *middle school* level; under Teacher Age, *21-30, 31-40*, and *> 60 years*; and under Highest Degree Earned – Education Level, *multiple*.

Table D.7 also indicates that a total of 11 variables associated with five items on the TIF were included as variables with a substantial discrimination model coefficient (Teacher Age; Highest Degree Earned: two of three sub-items; Teacher Certification: all three sub-items; Number of EE Inservices; and Level of Participation). Of these 11 variables, 6 had positive coefficients and were more characteristic of high-performing (Q1) schools, while 5 had negative coefficients and were more characteristic of low-performing (Q4) schools.

Further, for the Phase Two eighth-grade sample, seven Teacher variables had both VIP Eigenvalues greater than 1.0 and discrimination model coefficients greater than 0.10. Four of these variables had positive coefficients and three had negative coefficients. The four variables with positive model coefficients were: under Teacher Age, *21-30* (coefficient = + 0.162); under teachers' Perceived Level of Participation in environmental protection, *moderately* (coefficient = + 0.142); under Highest Degree Earned, *Master's* (coefficient = + 0.125); and under Teacher Age, *31-40* (coefficient = + 0.106). These are the best indicators of high-performing schools in this sample. Further, the three variables with negative model coefficients were: under Teacher Age, *>60* (coefficient = - 0.200); under Highest Degree Earned – Education Level, multiple (coefficient = - 0.159); and Highest Degree Earned, *Bachelor's* (coefficient = - 0.125). These are the best Teacher indicators of low-performing schools in this sample.

Finally, these results from the PLS-DA analyses for the Phase Two eighth-grade sample can be viewed within the context of the ability of the discrimination model, which included the variables in Table D.2, to accurately classify individual students into the Q1 and Q4 subsample of schools. The XLSTAT software used to conduct these PLS-DA analyses allowed the research team to review a summary of the reclassification of a 70% random subsample of individual students into the Q1 and Q4 sample of schools as well as a 30% random subsample used to validate these results (Table D.10).

Table D.10

Reclassification Analysis Results: Phase 2 Eighth Grade 70% and 30% Random Subsamples Using Component Scores (DV) and the Teacher Variable Model (IVs)

Sample*	**From/To**	**Highest**	**Lowest**	**Totals**	**% Accurate**
R 70%	Highest	70	2	72	97.22%
	Lowest	0	85	85	100.00%
	Totals	70	87	157	**98.73%**

R 30%	Highest	30	1	31	96.77%
	Lowest	0	36	36	100.00%
	Totals	30	37	67	**98.51%**

*** Note:** Under Sample, R = Random

For the 70% random estimation subsample, 98.73% of individual students were accurately classified into the highest (Q1) and lowest (Q4) subsample of eighth-grade schools. Further, for the 30% random validation subsample, 98.51% of individual students were accurately classified. The difference in the accurate reclassification of students in the 70% and 30% subsamples was 0.22%. Overall, these values fall within the range that allows this discrimination model to be deemed valid. In turn, this supports the results of the PLS-DA analyses presented in Table D.7 and summarized in prose, above.

Appendix E

Methods and Results for the First Set of Exploratory Analyses

Appendix E.

Methods and Results for the First Set of Exploratory Analyses

The first set of exploratory analyses focused on the extent to which School, Program, Teacher, and Student characteristics may differentiate between two groups of Phase Two schools: schools with high *skill component scores* (high-performing schools) and schools with low *skill component scores* (low-performing schools). In these analyses, the research team planned to include all School, Program, Teacher, and Student characteristics in one master set, rather than as separate sets. Using this master set, we sought to determine which characteristics best explained (predicted) the differences in skill composite scores in these two groups of schools.

One research question guided these analyses.

> Which school, program, teacher, and student characteristics appear to differentiate between the high- and low-performing schools, as determined by the distribution of *skill component scores*, at the 6th grade level, at the 7th grade level, and at the 8th grade level?

Methods for This Research Question

When this research question was written, the research team planned to include all relevant school (SIF), program (PIF), teacher (TIF) and student (MSELS) variables in one analysis for each grade in the Phase Two sample. These analyses are different than those undertaken for other overarching and exploratory research questions in several important ways. <u>First,</u> unlike the exploratory analysis of separate sets of variables for research questions in Appendix G, the intent here was to determine which of the independent variables from all four variable sets would remain as significant predictors. <u>Second</u>, unlike the analyses for Overarching Research Questions 3 and 4 in which student composite scores served as the dependent variable, here student skill component scores were used as the dependent variable. <u>Third</u>, unlike the analyses for Overarching Research Question 3, which included all students in each grade in the Phase Two sample, here all data files and analyses were restricted to include only high- and low-performing schools (Q1 and Q4) in these grades and Phases. While the use of quartiles is similar to what was done for Overarching Research Question 4, skill component scores rather than composite scores were used to generate these quartiles.

The first step in preparing data files for these analyses involved the use of mean student scores to assign schools to quartiles. This was done following procedures presented by Setek and Gallo (2009, Figure 8.43). A summary of the results of the quartiling for this research question is presented in Table E.1

Table E.1
Summary of Results of Quartiling of Phase 2 Schools on the Basis of Skill Component Score, by Grade

		Number of Schools		Range of Mean Composite Scores, by Quartile	
Research Question/ Dependent Variable	Grade	Total	# per Quartile	Quartile 4 (Q4)	Quartile 1 (Q1)
Skill Component Scores	6	43	10	12.86 – 20.46	29.28 – 39.15
(range 0 – 60)	7	40	10	15.47 – 23.46	29.39 – 40.78
	8	33	8	16.38 – 23.78	34.23 – 45.76

Once the schools that fell within Quartile 1 and Quartile 4 had been identified, the data files needed to conduct these analyses could be prepared. While the research team had intended to include all School, Program, Teacher, and Student variables in a single data file, we found this was not possible due to the loss of teacher data in Phase Two. Thus, as was done for Overarching Research Question 4, two separate data files were prepared for each grade in Phase Two: (a) one file that contained school, program, and student data, along with dependent variable scores; and (b) a second file that contained only teacher data, along with those same scores.

Once these files had been prepared, the research team faced two concerns. The first concern had to do with the number of variables. These data files were vary large because they included data from more than 50 items in the School, Program, and Teacher Information Forms, and the first section of the MSELS. Further, for each item, there were as many as 8 possible responses, each of which had to be coded separately in these data sets. The statistical software program treated each possible response as a separate variable, so the number of possible predictor variables in each data file was very large (number of items X number of responses), far greater than the number of schools to be included in each analysis. The second concern had to do with the unit of analysis. For this research question, the research team had planned to conduct analyses using individual student scores as the unit of analysis. However, as presented in Table E.1, school mean scores were used to select high- and low- performing schools in each grade.

To address these concerns, preliminary analyses were conducted (a) for the purpose of substantially reducing the number of potential predictor variables, and (b) using both individual student scores and school quartiles as dependent variables during the variable selection process. In specific, two preliminary analyses were conducted for each data file (i.e., the team ran two analyses of the file containing School, Program, and Student variables, and two analyses of the file containing Teacher variables). The first type of analysis involved the use of ANOVA tests to select important independent variables as likely predictors of individual student scores as the dependent variable. The second type of analysis involved the use of chi-square tests to select important independent variables as likely predictors of each set of school quartiles.

The results of these two types of analyses were used to narrow the very large number of independent variables by removing redundant and less relevant independent variables.
The results of these preliminary analyses reflect different dependent variable measures (quartiles based on school mean scores, and individual student scores) and for each measure, different methods for ranking and selecting predictor variables (magnitude and significance). In specific, we used ANOVA results to select the 10 variables with the largest F values and the 10 variables with the smallest p value, and Chi-square results to select the 10 variables with the largest χ^2 values and the 10 variables with the smallest p value. These predictor variables were combined into one list (i.e., there was one narrowed set of school, program and student variables, and one narrowed set of teacher variables).

The final step in data analysis involved the use of XLSTAT (2013) to conduct Partial Least Squares – Discriminant Analysis (PLS-DA) for each of these 12 narrowed data sets (Vinzi et al., 2010). The purpose of PLS-DA is to identify the variables in each data set that are most powerful in differentiating between high- and low-performing schools. As in the preliminary analyses, significance and magnitude were used to identify which variables in each data set did so. For significance, *Variables Important to Projection* (VIP) Eigenvalues were used to estimate the importance of each variable in predicting whether schools would be classified as high or low performing. Further, for magnitude, variable coefficients in each discrimination model were used to estimate the size of the effect of each independent variable on predicting whether schools would be classified as high or low performing.

Results for This Exploratory Research Question

In the preliminary analyses, both quartiles, based on school mean *skill component scores*, and individual student *skill component scores*, were used as dependent variables. The narrowed set of independent variables derived from these analyses for use in these PLS-DA analyses are summarized in Table E.2.

Table E.2
Variables Selected by Preliminary Analyses for Inclusion in PLS-DA Analyses, by Grade

Variable Set	Source	Variable	Grade 6	Grade 7	Grade 8
School, Program, and Student	School Info. Form	Census Region	X		X
		School Type: Public/Priv.		X	
		School Type: MS (y/n)	X		
		Total Enrollment	X	X	X
		Student : Teacher Ratio X	X	X	
		% Asian Students	X	X	X
		% Native American Sts.	X	X	
		% Hispanic Students	X	X	
		% Black Students		X	X
		% White Students		X	X
		% ESOL Students	X		
		% Free Lunch Students X	X	X	
		% Reduced Lunch Students	X	X	X
		% Special Needs Students	X	X	X
		% Federal IDEA Students			X
	Screening Survey	Program Duration	X	X	
		Program Intensity	X	X	X

Table E.2 (cont.)

Variable Set	Source	Variable	Grade 6	Grade 7	Grade 8
	Program Info. Form	Curriculum Organization	X		X
		Goal: Science		X	
		Goal: Social Studies	X		
		Organization of Teachers	X		X
		Instructional Grouping	X	X	X
		Instr. Method: Discussion	X		
		Instr. Method: Inquiry	X	X	
		Instr. Method: Labs	X		
		Instr. Method: Projects		X	X
		Instr. Method: Coop. Learn.		X	X
		Instr. Sites: Science Labs	X	X	
		Instr. Sites: Library/Media		X	
		Instr. Sites: School Grounds	X	X	X
		Instr. Sites: Field Sites		X	
		Instr. Sites: Other Comm.		X	X
		Assessment Methods	X	X	X
Teacher	Teacher Info. Form	Teacher Age	X	X	X
		Teacher Ethnicity		X	
		Highest Degree: Earned	X	X	
		Highest Degree: Ed. Level			X
		Highest Degree: SubjectX	X	X	
		Teacher Certification: Y/N	X		
		Teacher Cert.: Ed. Level		X	
		Teacher Cert.: Subject	X	X	X
		Years Teaching: Total	X	X	X
		Years Teaching: Middle Sch.	X	X	X
		# EE Inservices: Total	X	X	X
		# EE Inservices: < 1 day X	X	X	
		# EE Inservices: 1-2 days	X	X	X
		# EE Inservices: 3-7 days	X	X	X
		# EE Inservices: > 1 week			X
		Importance of EE to Teachers		X	
		Perceived Level of Sensitivity	X		
		Perceived Level of Particip.	X	X	X

One of the noticeable results of these preliminary analyses for Grades 6, 7, and 8 was that none of the Student variables were found to discriminate between high- and low-performing schools; only School, Program, and Teacher variables did so (Table E.2).

Results for student and program variables. Table E.3 presents results for the PLS-DA analysis of these variables found to be significant (VIP Eigenvalues = or > 1.0) as well as variables with a substantial magnitude (model coefficients = or > 0.10). For these analyses, schools served as the unit of analysis.

Table E.3

*Selected School and Program Variables that Differentiate Skill Component Score Quartile 1 from Quartile 4 Schools in the Phase 2 Sample, By Grade**

		Discriminant Analysis Result: VIP Eigenvalues and Model Coefficients								
		Grade 6 (n = 20 schools)			**Grade 7 (n = 20 schools)**			**Grade 8 (n = 16 schools)**		
Items and Variables	**Source of Data**	Variable Importance		DA Coefficient	Variable Importance		DA Coefficient	Variable Importance		DA Coefficient
		Q1 (n)	Q4 (n)	Top Q/Q1	Q1 (n)	Q4 (n)	Top Q/Q1	Q1 (n)	Q4 (n)	Top Q/Q1
Census Region	School Form									
* Midwest								1.979 (3)	1.860 (0)	**+ 0.183**
* South		1.023 (5)	1.009 (5)					1.663 (2)	1.587 (4)	
* West		1.368 (2)	1.263 (1)	**+ 0.130**						- 0.162
School Type:										
* 1: Private					1.437 (3)	1.334 (0)	**+ 0.155**			
Total Enrollment	School Form	1.670 (10)	1.531 (10)		1.354 (10)	1.345 (10)		1.172 (8)	1.096 (8)	
School Ethnicity	School Form									
* % Black Students								1.277 (8)	1.199 (8)	
* % White Students								1.765 (8)	1.659 (8)	
School SES	School Form									
* % Free Lunch Students		2.214 (6)	2.020 (10)		2.000 (5)	1.924 (10)		1.342 (6)	1.287 (8)	
* % Reduce Lunch Student		1.667 (5)	1.506 (9)		1.601 (4)	1.540 (10)		1.973 (5)	1.858 (8)	
Student : Teacher Ratio								1.151 (7)	1.427 (7)	
Program Duration	Program Form	1.611 (10)	1.550 (6)		1.018 (10)	1.069 (10)				
Program Intensity	Program Form	1.786 (8)	1.626 (8)					1.199 (7)	1.134 (7)	
Educational Goals:	Program Form									
* Social Studies		1.204 (5)	1.171 (6)							
Curriculum Organization:	Program Form									
* Separate Subjects		1.002 (2)								
* Other		2.301 (4)	2.092 (0)	**+ 0.181**						+ 0.252
Organization of Teachers:	Program Form									
* Sell-Contained		1.384(2)	1.332 (1)	**+ 0.192**						
* Departmentalized		1.614 (2)	1.563 (6)	**- 0.111**				1.370 (1)	1.270 (4)	
* Teaming		1.270 (5)	1.181 (2)					1.476 (4)	1.370 (1)	**+ 0.105**
* Other		1.048 (1)	1.075 (0)	**+ 0.224**						+ 0.156
* Combination of 1-4			1.006 (1)							
Instructional Groups, Ranked:	Program Form									
* Ranked #3: Whole Class			1.286 (1)	- 0.134			- 0.102			
* Ranked #3: Groups/Teams		1.203 (2)	1.134 (1)	**+ 0.106**						
* Ranked #4: Individualized			1.006 (1)							
* Ranked #4: Other		1.362 (1)	1.283 (0)	**+ 0.128**						

Table F.3 (cont.)

Items and Variables	Source of Data	Discriminant Analysis Result: VIP Eigenvalues and Model Coefficients								
		Grade 6 (n = 20 schools)			Grade 7 (n = 20 schools)			Grade 8 (n = 16 schools)		
		Variable Importance		DA Coefficient	Variable Importance		DA Coefficient	Variable Importance		DA Coefficient
		Q1 (n)	Q4 (n)	Top Q/Q1	Q1 (n)	Q4 (n)	Top Q/Q1	Q1 (n)	Q4 (n)	Top Q/Q1
Instructional Settings:	Program Form									
* Science Labs		1.415 (2)	1.309 (6)		1.316 (6)	1.226 (7)				
* School Grounds								1.520 (4)	1.409 (7)	
* Library/Media Center					1.091 (1)	1.202 (4)				
* Other Comm. Sites					1.022 (8)					
Instructional Methods:										
* Labs			1.006 (1)							
* Cooperative Learning					1.309 (5)	1.224 (2)				
* Projects					1.201 (5)	1.135 (1)				
Assessment Methods Ranked:	Program Form									
* Ranked #3: Standardized							+ 0.281			
* Ranked #1: Alternative					2.159 (3)	2.209 (0)	**+ 0.257**			
* Ranked #3: Alternative					1.062 (0)	1.065 (2)				
* Ranked #1: Informal						1.063 (5)	- 0.101			
* Ranked #3: Informal							+ 0.139			
* Ranked #4: Other				+ 0.140						

* **Notes**: Only variables with Variable Importance (VIP) Eigenvalues equal to or greater than 1.0, and a model coefficient equal to or greater than 0.10 were included in this table. For each variable with a VIP Eigenvalue and a model coefficient at or above this level, the model coefficient has been bolded.

In the PLS-DA analyses of the **sixth-grade sample**, Table E.3 indicates that variables associated with three items in the School Information Form (SIF), two items in the Screening Survey, and five items in the Program Information Form (PIF) were included as significant on the basis of VIP Eigenvalues. A total of 16 variables associated with these 10 items were found to have VIP Eigenvalues greater than 1 for both Q1 and Q4 schools. The variables with the greatest VIP Eigenvalues were: under Curriculum Organization, *Other*; under School SES, the percent of students in *Free Lunch* and in *Reduced Lunch* Programs; *Program Intensity*; *Total Enrollment*; under Organization of Teachers, *Departmentalized*; and *Program Duration*.

Table E.3 also indicates that a total of nine variables associated with one item on the SIF and four items on the PIF were included as variables with a discrimination model coefficient greater than 0.1 (SIF: Region; PIF: Curriculum Organization, Organization of Teachers, Instructional Groups, and Assessment Methods). Of these nine variables, seven had positive coefficients and were more characteristic of schools with high skill component scores (Q1), while two had negative coefficients and were more characteristic of schools with low skill component scores (Q4).

Further, for the Phase Two sixth-grade sample, seven variables had both VIP Eigenvalues greater than 1.0 <u>and</u> discrimination model coefficients greater than 0.10 (i.e., one School and six Program variables). Six of these variables had positive coefficients and one had a negative coefficient. The six variables with positive model coefficients were:

- under Organization of Teachers, *Other* (coefficient = + 0.224);
- under Organization of Teachers, *Self-contained* (coefficient = + 0.192);
- under Curriculum Organization, *Other* (coefficient = + 0.181);
- under Census Region, West (coefficient = + 0.130);
- under Instructional Groups, ranking *Other* fourth (coefficient = + 0.128); and
- under Instructional Groups, ranking *Groups/Teams* third (coefficient = + 0.106).

While most of these variables (response choices) are reasonably clear, the selection of *Other* as a response are not. Under Organization of Teachers, two teachers selected *Other*, only one of whom wrote in a description of what this meant: "Self-contained classrooms, with team teaching for certain subjects." Under Curriculum Organization, four teachers selected *Other*. Of these, two referred to incorporating lessons in after-school clubs and a third in after-school EE. The fourth teacher indicated that environmental science was a separate class, but in the sixth grade curriculum it is "often integrated across subjects." Lastly, under Instructional Groups, one teacher selected *Other*, and offered the following explanation: "Independent – students on their own."

These are the best School and Program indicators of schools with <u>high</u> skill component scores in the Phase Two sixth-grade sample. Further, the only variable with a negative model coefficient was: under Organization of Teachers, *Departmentalized* (coefficient =
- 0.111). This was the best School and Program indicator of schools with <u>low</u> skill component scores in this sample.

Finally, these results from the PLS-DA analyses for the Phase Two sixth-grade sample can be viewed within the context of the ability of the discrimination model, which included the variables in Table E.2, to accurately classify individual students into the Q1 and Q4 subsample of schools. The XLSTAT software used to conduct these PLS-DA analyses allowed the research team to review a summary of the reclassification of a 70% random subsample of individual students into the Q1 and Q4 sample of schools as well as a 30% random subsample used to validate these results (Table E.4).

Table E.4

Reclassification Analysis Results: Phase 2 Sixth Grade 70% and 30% Random Subsamples Using Skill Component Scores (DV) and the School and Program Variable Model (IVs)

Sample*	From/To	Highest	Lowest	Totals	% Accurate
R 70%	Highest	214	23	237	90.30%
	Lowest	0	594	594	100.00%
	Totals	214	617	831	**97.23%**
R 30%	Highest	79	16	95	83.16%
	Lowest	0	261	261	100.00%
	Totals	79	277	356	**95.51%**

*** Note:** Under Sample, R = Random

For the 70% random estimation subsample, 97.23% of individual students were accurately classified into the highest (Q1) and lowest (Q4) subsample of sixth-grade schools. Further, for the 30% random validation subsample, 95.51% of individual students were accurately classified. The difference in the accurate reclassification of students in the 70% and 30% subsamples was 1.72%. Overall, these values fall within the range that allows this discrimination model to be deemed valid. In turn, this supports the results of the PLS-DA analyses presented in Table E.3 and summarized in the prose, above.

In the PLS-DA analyses of the **seventh-grade sample**, Table E.3 indicates that variables associated with three items in the School Information Form (SIF), one item in the Screening Survey, and three items in the Program Information Form (PIF) were included as significant VIP variables. A total of 12 variables associated with these seven items were found to have VIP Eigenvalues greater than 1 for both Q1 and Q4 schools. The variables with the greatest VIP Eigenvalues were: under Assessment Methods, ranking *Alternative Assessment* first; and under School SES, the percent of students in *Free Lunch* and in *Reduced Lunch* Programs.

Table E.3 also indicates that a total of six variables associated with one item on the SIF and two items on the PIF were included as variables with a substantial discrimination model coefficient (SIF: School Type, *private*; PIF: Instructional Groups, and Assessment Methods). Of these six variables, four had positive coefficients and were more characteristic of schools with high skill component scores (Q1), while two had negative coefficients and were more characteristic of schools with low skill component scores (Q4).

Further, for the Phase Two seventh-grade sample, two School and Program variables had VIP Eigenvalues greater than 1.0 and discrimination model coefficients greater than 0.10. Both of these variables had positive coefficients: under Assessment Methods, ranking *alternative assessment* first (coefficient = + 0.257); and under School Type, *private schools* (coefficient = + 0.155). These are the best School and Program indicators of schools with high skill component scores in this sample.

Finally, these results from the PLS-DA analyses for the Phase Two seventh-grade sample can be viewed within the context of the ability of the discrimination model, which included the variables in Table E.2, to accurately classify individual students into the Q1 and Q4 subsample of schools. The XLSTAT software used to conduct these PLS-DA analyses allowed the research team to review a summary of the reclassification of a 70% random subsample of individual students into the Q1 and Q4 sample of schools as well as a 30% random subsample used to validate these results (Table E.5).

Table E.5

Reclassification Analysis Results: Phase 2 Seventh Grade 70% and 30% Random Subsamples Using Skill Component Scores (DV) and the School and Program Variable Model (IVs)

Sample*	From/To	Highest	Lowest	Totals	% Accurate
R 70%	Highest	161	4	165	97.58%
	Lowest	0	633	633	100.00%
	Totals	161	637	798	**99.50%**
R 30%	Highest	59	6	65	90.77%
	Lowest	0	277	277	100.00%
	Totals	59	283	342	**98.25%**

*** Note:** Under Sample, R = Random

For the 70% random estimation subsample, 99.5% of individual students were accurately classified into the highest (Q1) and lowest (Q4) subsample of seventh-grade schools. Further, for the 30% random validation subsample, 98.25% of individual students were accurately classified. The difference in the accurate reclassification of students in the 70% and 30% subsamples was 1.25%. Overall, these values fall within the range that allows this discrimination model to be deemed valid. In turn, this supports the results of the PLS-DA analyses presented in Table E3 and summarized in the prose, above.
In the PLS-DA analyses of the **eighth-grade sample**, Table E.3 indicates that variables associated with five items in the School Information Form (SIF), one item in the Screening Survey, and two items in the Program Information Form (PIF) were included as significant VIP variables. A total of 11 variables associated with these eight items were found to have VIP Eigenvalues greater than 1 for both Q1 and Q4 schools. The variables with the greatest VIP Eigenvalues were: under Census Region, *Midwest* and *South*; under School SES, the percent of students in *Reduced Lunch* Programs; and under School Ethnicity, the percent of *White* students.

Table E.3 also indicates that a total of five variables associated with one item on the SIF and two items on the PIF were included as variables with a substantial discrimination model coefficient (SIF: Census Region; PIF: Curriculum Organization, and Organization of Teachers). Of these five variables, four had positive coefficients and were more characteristic of schools with high skill component scores (Q1), while one had a negative coefficient and was more characteristic of schools with low skill component scores (Q4).

However, for the Phase Two eighth-grade sample, two variables had both VIP Eigenvalues greater than 1.0 and discrimination model coefficients greater than 0.10. Both of these variables had positive coefficients: under Census Region, *Midwest* (coefficient = + 0.183); and under Organization of Teachers, *teaming* (coefficient = + 0.105). These are the best School and Program indicators of schools with high skill component scores in this sample.

Finally, these results from the PLS-DA analyses for the Phase Two eighth-grade sample can be viewed within the context of the ability of the discrimination model, which included the variables in Table E.2, to accurately classify individual students into the Q1 and Q4 subsample of schools. The XLSTAT software used to conduct these PLS-DA analyses allowed the research team to review a summary of the reclassification of a 70% random subsample of individual students into the Q1 and Q4 sample of schools as well as a 30% random subsample used to validate these results (Table E.6).

For the 70% random estimation subsample, 99.37% of individual students were accurately classified into the highest (Q1) and lowest (Q4) subsample of eighth-grade schools. Further, for the 30% random validation subsample, 91.6% of individual students were accurately classified. The difference in the accurate reclassification of students in the 70% and 30% subsamples was 7.77%. Overall, these values fall within the range that allows this discrimination model to be deemed valid. In turn, this supports the results of the PLS-DA analyses presented in Table E.3 and summarized in the prose, above.

Table E.6

Reclassification Analysis Results: Phase 2 Eighth Grade 70% and 30% Random Subsamples Using Skill Component Scores (DV) and the School and Program Variable Model (IVs)

Sample*	From/To	Highest	Lowest	Totals	% Accurate
R 70%	Highest	140	4	144	97.22%
	Lowest	0	468	468	100.00%
	Totals	140	472	612	**99.37%**
R 30%	Highest	52	9	61	85.25%
	Lowest	13	188	201	93.53%
	Totals	65	197	262	**91.60%**

*** Note:** Under Sample, R = Random

Results for teacher variables, Table E.7 presents results for the PLS-DA analysis of these variables found to be significant (VIP Eigenvalues = or > 1.0) and variables with a substantial magnitude (model coefficients = or > 0.10). For these analyses, individual teachers served as the unit of analysis.

In the PLS-DA analyses of the **sixth-grade sample**, Table E.7 indicates that variables associated with five items in the Teacher Information Form (TIF) were included as significant VIP variables. A total of 12 variables associated with these five items were found to have VIP Eigenvalues greater than 1 for both Q1 and Q4 schools. The variables with the greatest VIP Eigenvalues were: under Teacher Certification – Subject, *social studies/history*; and under Teacher Certification, *yes*.

Table E.7 also indicates that a total of ten Teacher variables associated with three items on the TIF were included as variables with a substantial discrimination model coefficient (Highest Degree Earned: 2 of 3 sub-items; Teacher Certification: 2 of 3 sub-items; and Perceived Level of Participation). Of these 10 variables, six had positive coefficients and were more characteristic of schools with high skill component scores (Q1), while four had negative coefficients and were more characteristic of schools with low skill component scores (Q4).

Further, for the Phase Two sixth-grade sample, seven variables had both VIP Eigenvalues greater than 1.0 and discrimination model coefficients greater than 0.10. Four of these variables had positive coefficients and three had negative coefficients. The four variables with positive model coefficients were:

Table E.7

*Selected Teacher Variables that Differentiate Skill Component Score Quartile 1 from Quartile 4 Schools in the Phase 2 Sample, By Grade**

		Discriminant Analysis Results: VIP Eigenvalues and Model Coefficients								
		Grade 6 (n= 18)			**Grade 7 (n = 22)**			**Grade 8 (n = 11)**		
Variables	**Source of Data**	Variable Importance		DA Coefficient	Variable Importance		DA Coefficient	Variable Importance		DA Coefficient
		Q1 (n)	Q4 (n)	Top Q/Q1	Q1 (n)	Q4 (n)	Top Q/Q1	Q1 (n)	Q4 (n)	Top Q/Q1
Teacher Age:										
* 21-30										+ 0.204
* 31-40								1.141 (3)	1.059 (0)	**+ 0.121**
* 51-60					1.148 (1)	1.022 (6)				
* > 60					1.489 (1)	1.414 (0)	**+ 0.233**	1.471 (0)	1.432 (2)	**- 0.196**
Teacher Ethnicity:										
* Other					1.146 (1)	1.036 (0)	**+ 0.355**			
Years Teaching:										
* Middle School								1.605 (7)	1.618(4)	
* Total								1.011 (7)		
Highest Degree Earned:	Teacher Form									
* Bachelor's		1.317 (3)	1.231 (7)							
* Master's			1.166 (1)							
* Master's + 30				+ 0.114						
Highest Degree, Ed. Level:	Teacher Form									
* Elementary								1.209 (3)	1.123 (0)	**+ 0.131**
* Multiple								1.155 (1)	1.099 (1)	
* None										- 0.150
Highest Degree Earned, Subject	Teacher Form									
* Science		1.255 (1)	1.176 (5)	**- 0.101**		1.045 (6)	- 0.112	1.755 (1)	1.629 (3)	**- 0.149**
* Social Studies/History		1.465 (1)	1.487 (0)	**+ 0.288**						
* Other					1.787 (2)	1.603 (0)	**+ 0.341**			
* Multiple					1.126 (1)	1.248 (0)	**+ 0.471**			
* None		1.00 (4)	1.165 (3)					1.755 (6)	1.629 (1)	**+ 0.149**
Teacher Certification:	Teacher Form									
* Yes		1.539 (6)	1.440 (10)	**- 0.160**						
* Working On				+ 0.340						
* No		1.427 (1)	1.355 (0)	**+ 0.127**						

Table E.7 (cont.)											
		Discriminant Analysis Results: VIP Eigenvalues and Model Coefficients									
		Grade 6 (n= 18)			**Grade 7 (n = 22)**			**Grade 8 (n = 11)**			
Variables	**Source of Data**	Variable Importance		DA Coefficient	Variable Importance		DA Coefficient	Variable Importance		DA Coefficient	
		Q1 (n)	Q4 (n)	Top Q/Q1	Q1 (n)	Q4 (n)	Top Q/Q1	Q1 (n)	Q4 (n)	Top Q/Q1	
Teacher Certification, Subject:	Teacher Form										
* Science									1.099 (3)		
* Social Studies/History		1.952 (1)	1.864 (0)	**+ 0.297**	1.313 (1)	1.184 (0)	**+ 0.413**			+ 0.149	
* English/Lang. Arts/ Read.				- 0.246			- 0.111				
* Multiple					1.050 (4)	1.051 (2)	**+ 0.125**				
* None							- 0.112		1.056 (1)	**- 0.129**	
# of EE Inservices:	Teacher Form										
* Total					1.813 (7)	1.675 (8)					
* < 1 Day		1.364 (5)	1.377 (1)		1.967 (5)	1.787 (3)					
* 1-2 Days					2.182 (6)	1.948 (3)					
Perc. Level of Env. Sensitivity:											
* Considerably		1.079 (4)	1.043 (5)								
* Extremely		1.079 (4)	1.043 (5)								
Perc. Level of Participation:	Teacher Form										
* Slightly		1.427 (1)	1.355 (0)	**+ 0.127**	1.504 (1)	1.641 (0)	**+ 0.512**				
* Considerably		1.356 (1)	1.267 (4)	**- 0.102**							

* **Notes**: Only those variables with Variable Importance (VIP) Eigenvalues equal to or greater than 1.0, and a model coefficient equal to or greater than 0.10 were included in this table. For each variable with a VIP Eigenvalue and a model coefficient at or above these levels, the model coefficient has been bolded.

- under Teacher Certification – Subject, *Social Studies/History* (coefficient = + 0.297);
- under Highest Degree Earned – Subject, *Social Studies/History* (coefficient = + 0.288);
- under Perceived Level of Participation, *Slightly* (coefficient = + 0.127); and
- under Teacher Certification, *No* (coefficient = + 0.127).

These are the best Teacher indicators of schools with high skill component scores in the Phase Two sixth-grade sample. Further, the variables with negative model coefficients were: under Teacher Certification, *yes* (coefficient = - 0.160); under teachers' Perceived Level of Participation in environmental protection, *Considerably* (coefficient = – 0.102); and under Highest Degree Earned – Subject, *Science* (coefficient = – 0.101). These were the best Teacher indicators of schools with low skill component scores in this sample.

Finally, these results from the PLS-DA analyses for the Phase Two sixth-grade sample can be viewed within the context of the ability of the discrimination model, which included the variables in Table E.2, to accurately classify individual students into the Q1 and Q4 subsample of schools. The XLSTAT software used to conduct these PLS-DA analyses allowed the research team to review a summary of the reclassification of a 70% random subsample of individual students into the Q1 and Q4 sample of schools as well as a 30% random subsample used to validate these results (Table E.8).

Table E.8

Reclassification Analysis Results: Phase 2 Sixth Grade 70% and 30% Random Subsamples Using Skill Component Scores (DV) and the Teacher Variable Model (IVs)

Sample*	From/To	Highest	Lowest	Totals	% Accurate
R 70%	Highest	111	26	137	81.02%
	Lowest	0	319	319	100.00%
	Totals	111	345	456	**94.30%**
R 30%	Highest	44	13	57	77.19%
	Lowest	12	126	138	91.30%
	Totals	56	139	195	**87.18%**

*** Note:** Under Sample, R = Random

For the 70% random estimation subsample, 94.30% of individual students were accurately classified into the highest (Q1) and lowest (Q4) subsample of sixth-grade schools. Further, for the 30% random validation subsample, 87.18% of individual students were accurately classified.

The difference in the accurate reclassification of students in the 70% and 30% subsamples was 7.12%. Overall, these values fall within the range that allows this discrimination model to be deemed valid. In turn, this supports the results of the PLS-DA analyses presented in Table E.7 and summarized in the prose, above.
In the PLS-DA analyses of the **seventh-grade sample**, Table E.7 indicates that variables associated with six items in the Teacher Information Form (TIF) were included as significant VIP variables. A total of 10 variables associated with these five items were found to have VIP Eigenvalues greater than 1 for both Q1 and Q4 schools. The variables with the greatest VIP Eigenvalues were: under Number of EE Inservices, *1-2 Days*, *< 1 Day*, and *Total*; and Highest Degree Earned – Subject, *Other*.

Table E.7 also indicates that a total of ten variables associated with five items on the TIF were included as variables with a substantial discrimination model coefficient (Teacher Age; Teacher Ethnicity; Highest Degree Earned: Subject; Teacher Certification: Subject; and Perceived Level of Participation). Of these 10 variables, seven had positive coefficients and were more characteristic of schools with high skill component scores (Q1), while three had negative coefficients and were more characteristic of schools with low skill component scores (Q4).

Further, for the Phase Two seventh-grade sample, seven variables had both VIP Eigenvalues greater than 1.0 and discrimination model coefficients greater than 0.10. All seven of these variables had positive coefficients:

- under Perceived Level of Participation, *Slightly* (coefficient = + 0.512);
- under Highest Degree Earned – Subject, *Multiple* (coefficient = + 0.471);
- under Teacher Certification – Subject, *Social Studies/History* (coefficient = + 0.413);
- under Teacher Ethnicity, *Other* (coefficient = + 0.355);
- under Highest Degree Earned – Subject, *Other* (coefficient = + 0.341);
- under Teacher Age, *> 60* (coefficient = + 0.233); and
- under Teacher Certification – Subject, *Multiple* (coefficient = + 0.125).

While most of these variables (response choices) are reasonably clear, the *Other* and *Multiple* responses above are not. Under Highest Degree Earned - Subject, two teachers selected *Other*. One of these teachers wrote in "environmental education" and the other wrote in "liberal arts." Under this heading, one teacher selected *Multiple*, and this teacher wrote in "science, environmental education." Under Teacher Certification – Subject, six teachers selected *Multiple*. Four of these teachers were in high-performing (Q1) schools, and they wrote in the following: "environmental education"; "integrated"; "science, health/PE"; and "science, math." Also, two of these teachers were in low-performing (Q4) schools, and they wrote in: "English/language arts, math"; and "science, social studies/history." Lastly, under Teacher Ethnicity, one teacher selected *Other*, and offered the following: "Arabic."

The seven variables listed and described above are the best Teacher indicators of schools with high skill component scores in this sample.

Finally, these results from the PLS-DA analyses for the Phase Two seventh-grade sample can be viewed within the context of the ability of the discrimination model, which included the variables in Table E.2, to accurately classify individual students into the Q1 and Q4 subsample of schools. The XLSTAT software used to conduct these PLS-DA analyses allowed the research team to review a summary of the reclassification of a 70% random subsample of individual students into the Q1 and Q4 sample of schools as well as a 30% random subsample used to validate these results (Table E.9).

Table E.9

Reclassification Analysis Results: Phase 2 Seventh Grade 70% and 30% Random Subsamples Using Skill Component Scores (DV) and the Teacher Variable Model (IVs)

Sample*	From/To	Highest	Lowest	Totals	% Accurate
R 70%	Highest	68	22	90	75.56%
	Lowest	0	457	457	100.00%
	Totals	68	479	547	**95.98%**
R 30%	Highest	19	20	39	48.72%
	Lowest	0	186	186	100.00%
	Totals	19	206	225	**91.11%**

* **Note:** Under Sample, R = Random

For the 70% random estimation subsample, 95.98% of individual students were accurately classified into the highest (Q1) and lowest (Q4) subsample of seventh-grade schools. Further, for the 30% random validation subsample, 91.11% of individual students were accurately classified. The difference in the accurate reclassification of students in the 70% and 30% subsamples was 4.87%. While one of these % accurate values fell below 50%, overall, these values fall within the range that allows this discrimination model to be deemed valid. In turn, this supports the results of the PLS-DA analyses presented in Table E.7 and summarized in the prose, above.

In the PLS-DA analyses of the **eighth-grade sample**, Table E.7 indicates that variables associated with three items in the Teacher Information Form (TIF) were included as significant VIP variables. A total of seven variables associated with these three items were found to have VIP Eigenvalues greater than 1 for both Q1 and Q4 schools. The variables with the greatest VIP Eigenvalues were: under Highest Degree Earned – Subject, *science* and *none* (e.g., *none* was used for elementary education, and when no subject was listed or apparent*);* and under Years Teaching*,* years teaching *middle school*.

Table E.7 also indicates that a total of nine variables associated with three items on the TIF were included as variables with a substantial discrimination model coefficient (Teacher Age; Highest Degree Earned: Subject; and Teacher Certification: Subject). Of these nine variables, five had positive coefficients and were more characteristic of high-performing (Q1) schools, while four had negative coefficients and were more characteristic of low-performing (Q4) schools.

Further, for the Phase Two eighth-grade sample, five variables had both VIP Eigenvalues greater than 1.0 and discrimination model coefficients greater than 0.10. Three of these variables had positive coefficients: under Highest Degree Earned – Subject, *none* (coefficient = + 0.149); under Highest Degree Earned – Education Level, *elementary* (coefficient = + 0.131); and under Teacher Age, *31-40* (coefficient = + 0.121). These three variables are the best Teacher indicators of schools with high skill component scores in this sample. The two variables with negative model coefficients were: under Teacher Age, *> 60* (coefficient = – 0.196); and under Highest Degree Earned – Subject, *science* (coefficient = – 0.149). These were the best Teacher indicators of schools with low skill component scores in this sample.

Finally, these results from the PLS-DA analyses for the Phase Two eighth-grade sample can be viewed within the context of the ability of the discrimination model, which included the variables in Table E.2, to accurately classify individual students into the Q1 and Q4 subsample of schools. The XLSTAT software used to conduct these PLS-DA analyses allowed the research team to review a summary of the reclassification of a 70% random subsample of individual students into the Q1 and Q4 sample of schools as well as a 30% random subsample used to validate these results (Table E.10).

Table E.10

Reclassification Analysis Results: Phase 2 Eighth Grade 70% and 30% Random Subsamples Using Skill Component Scores (DV) and the Teacher Variable Model (IVs)

Sample*	From/To	Highest	Lowest	Totals	% Accurate
R 70%	Highest	70	2	72	97.22%
	Lowest	0	85	85	100.00%
	Totals	70	87	157	**98.73%**
R 30%	Highest	30	1	31	96.77%
	Lowest	0	36	36	100.00%
	Totals	30	37	67	**98.51%**

*** Note:** Under Sample, R = Random

For the 70% random estimation subsample, 98.73% of individual students were accurately classified into the highest (Q1) and lowest (Q4) subsample of eighth-grade schools. Further, for the 30% random validation subsample, 98.51% of individual students were accurately classified. The difference in the accurate reclassification of students in the 70% and 30% subsamples was 0.22%. Overall, these values fall within the range that allows this discrimination model to be deemed valid. In turn, this supports the results of the PLS-DA analyses presented in Table E.7 and summarized in prose, above.

Appendix F

Methods and Results for the Second Set of Exploratory Analyses

Appendix F

Description of Methods and Results for the Second Set of Exploratory Analyses

This second set of exploratory analyses focused on the extent to which variables within each of the School, Program, Teacher, and Student data sets would explain variability in Environmental Literacy Composite scores. There were four research questions in this set.

a. Within the Phase One sample of schools (n=47), which school characteristics and which program characteristics make statistically significant contributions to explanations of the variance in student MSELS composite scores?

b. Within the Phase Two sample of schools (n=64), which school characteristics and which program characteristics make statistically significant contributions to explanations of the variance in student MSELS composite scores?

c. Within the Phase One sample of schools (n=47), which teacher characteristics and which student characteristics make statistically significant contributions to explanations of the variance in student MSELS composite scores?

d. Within the Phase Two sample of schools (n=64), which teacher characteristics and which student characteristics make statistically significant contributions to explanations of the variance in student MSELS composite scores?

Methods for Research Questions a – d

For each of these research questions, multiple linear regression analyses were undertaken in each grade (Phase One: Grades 6 & 8; Phase Two: Grades 6, 7, & 8). To enable these analyses, data files were prepared for each Phase, grade, and set of variables, for a total of 20 data files.

For Research Questions a and b, the school served as the unit of analysis. Thus, each data file included data for each school (e.g., the Phase One, Grade 6 data file for School variables included the data for each of those 48 schools, each as a separate row). Further, the data for each school included the average environmental literacy composite score for all participating students within that grade in that school (e.g., all students in 1, 2, 5, 10, or more classes in that grade in that school) as the dependent variable. Finally, the data for each school also included responses to items on either the School Information Form (SIF) or the Program Information Form (PIF).

For the first part of Research Questions c and d, the individual teacher served as the unit of analysis. The data files for these research questions included data for each teacher from their Teacher Information Form (TIF), as well as the average environmental literacy composite score for all participating students within that teacher's class or classes. In the

Phase One baseline study, the number of schools and the number of teachers was nearly equal (i.e., in the Grade 6 and in the Grade 8 data file, only two schools had two teachers each). However, in the Phase Two study of schools with environmental programs, the number of schools and teachers varied considerably from one grade to another (i.e., Grade 6: 57 teachers in 35 schools; Grade 7: 42 teachers in 30 schools; and Grade 8: 32 teachers in 23 schools).

For the second part of Research Questions c and d, the individual student served as the unit of analysis. The data files for these research questions included data for each student from Section 1 of the MSELS (Items 1-4), as well as each student's environmental literacy composite score.

The research team encountered several problems in the use of these data files for these multiple linear regression analyses. First, the Phase One sample (48 schools), and the Phase Two sample (64 schools) are considered small for regression analysis. The use of small sample sizes can lead to instability in regression results, particularly when a large number of independent (predictor) variables have been included in each analysis (School Variables: 24; Program Variables: 38; Teacher Variables: 51; and Student Variables: 12). This instability makes it difficult, if not impossible, to identify and select predictors that were both stable and strong. The research team experienced this problem during repeated runs of stepwise regression using Phase One and Phase Two School data for Research Questions a and b. One way to address this problem is to conduct a K-fold cross-validation procedure within stepwise regression on a repeated basis for each data set (i.e., 100 times) (Burman, 1989; Shao, 1993; Zhang, 1993). This procedure will generate results across repeated analyses that can be aggregated, yielding one set of results for the variables included in the regression model for each full data set. Because the statistical program JMP did not have analysis options to conduct this type of analysis, Dr. Giannoulis prepared a code that would allow JMP to do so. This became the first step in all linear regression analyses for Research Questions a through d, and allowed the team to select only those predictor variables that were both stable and strong with a high degree of confidence. The second step in all linear regression analyses was to use these selected predictors as a refined stepwise regression model.

Second, often, small samples do not meet the necessary assumptions recommended by Cohen et al. (2003), Tabachnik and Fidell (2013), and others to obtain valid and accurate linear regression results (i.e., a ratio of the number of records in each data set to the number of predictor variables; singularity and multicolinearity among predictor variables; the normality, heteroscedasticity, and independence of residuals; and the presence and influence of outliers). The research team was unable to meet some of these assumptions. One way to largely overcome this problem is to use bootstrapping as a random resampling method (Field & Miles, 2012). For this reason, the research team used bootstrapping with replacement (i.e., 10,000 analyses of samples drawn from a given data set) (Efron & Tibshirani, 1994). This procedure aggregates results across analyses of these samples ("bagging"), yielding one set of robust results for parameter estimates and associated confidence intervals (Brieman, 1996). The *Split Selected Column* option in JMP (Version 10.02) was used to conduct these analyses. In cases where the resulting linear regression model was found to be significant but none of the selected variables in it were significant ($p < .05$), steps were taken to rerun that analysis after dropping the less significant variables. This became the third step in all linear regression analyses for Research Questions a through d, and allowed the team to identify and validate the a final regression model that contained only the most significant predictors of composite scores within each of these refined sets of predictor variables.

Results for Research Question a

As described above, three steps were taken in these linear regression analyses. The results of all three steps for Research Question a, the effects of school variables on composite scores for the Phase One sixth- and eighth-grade samples, are summarized in Table F.1.

For the first step, K-fold cross-validation (KFCV) methods were used to determine the influence of all school variables on student composite scores. Table F.1 includes results for those school variables found to be significant predictors of composite scores ($p < .05$) in at least 20% (20/100) of the repeated analyses. The resulting significant predictors were included in the second step. In the second step, these selected school variables were analyzed using stepwise linear multiple regression. Table F.1 includes p values for those school variables found to be significant predictors of composite scores ($p < .05$). Non-significant variables are identified as *N.S.* and were dropped from further analysis. Only these significant predictors were included in the third step. In the third step, bootstrapping resampling techniques were used. Table 1a includes p values for those school variables that met the following criteria: (a) the variable's confidence interval did not include zero (0); and (b) the variable's contribution to the explained variance in composite scores was significant ($p < .05$). School variables that did not meet these criteria were again identified as *N.S*. In addition, Table F.1 includes the positive (+) or negative (-) direction of the regression coefficient associated with each significant p value.

For the first step in the analysis of the **Phase One sixth-grade data for Research Question a**, Table F.1 indicates that seven school variables were significant in 20% or more of the repeated KFCV analyses ($p < .05$): three variables pertaining to the ethnic composition of school (the percent of Black, of White, and of Native American students); two that pertained to the socio-economic composition of schools (the percent of student in Free and in Reduced Lunch Programs); and two that pertained to the special education composition of schools (the percent of students in Federal IDEA and local Special Education programs). When these seven school variables were analyzed in the second step, only four remained as significant predictors of these students' composite scores: the percent of Black, White, Free Lunch, and Federal IDEA students. Finally, when these four variables were analyzed in the third step, three remained as significant predictors of student composite scores (i.e., only the percent of Black students did not). Of these three, only the percent of students in Federal IDEA programs (i.e., students who receive services in schools under the 1990 *Individuals with Disabilities Act* [Public Law 101-476]) was a positive predictor of student composite scores (i.e., it had a positive regression coefficient).

Table F.1
Major Results of Regression Analyses for Research Question a: School Variables by Composite Scores for the Phase 1 Sample

Step in Regression Analyses	**First Step**		**Second Step**		**Third Step**	
Analysis Method	Repeated K-Fold Cross Validation*		Basic Stepwise Regression*		Regression Bootstrapping*	
Grade	6	8	6	8	6	8
% of Significant Models (100 runs)	100%	100%				
F Ratio			8.226	6.393	6.576	6.609
p Values			0.0002	0.0007	0.001	0.0019
R-Square Value			0.772	0.582	0.523	0.442
Adj. R-Square Value			0.678	0.491	0.443	0.375
School Variables						
NCES Indicator:						
* City		100%		0.0002		+ 0.0056
* Suburban						
* Town		81%		N.S.		
* Rural						
Ethnic Composition of School						
* % Native American	37%		N.S.			
* % Asian						
* % Black	100%	98%	0.0001	0.003	N.S.	- 0.0006
* % Hispanic						
* % White	100%		0.0001		- 0.001	
SES Composition of School						
* % Reduced Lunch	42%		N.S.			
* % Free Lunch	100%		0.0021		- 0.0035	
Special Ed. Composition of School						
* % Federal IDEA (NCES)	100%	20%	0.0002	0.007	+ 0.0001	+ 0.039
* % Special Education (SIF)	35%		N.S.			

* **Notes**: NR = no response; and N.S. = not statistically significant at p < .05

For the first step in the analysis of the **Phase One eighth-grade data**, Table F.1 indicates that four school variables were significant in 20% or more of the repeated KFCV analyses ($p < .05$): two National Center for Education Statistics (NCES) indicators of the geographic area in which schools were located (i.e., in City and in Town areas), the percent of Black students in schools, and the percent of students in the Federal IDEA program. When these four variables were analyzed in the second step, three of the four remained as significant predictors of student composite scores (i.e., only the percent of schools in Town areas did not). Finally, when these three variables were analyzed in the third step, all three remained significant predictors of student composite scores. Of these, two were positive predictors of student composite scores: the location of schools in City areas; and the percent of students in the Federal IDEA program.

When these results for Phase One sixth- and eighth-grade students are compared, only two school variables were found to be significant predictors of student composite scores in the second step for both sixth- and eighth-grade samples (i.e., under Ethnic Composition of Schools: the percent of Black students; and under the Special Education Composition of Schools: the percent of Federal IDEA students). Of these two variables, only the percent of students in the Federal IDEA program was found to be a significant and positive predictor for both grades in the third and final step.

The results of all three steps for Research Question a, the effects of program variables on composite scores for the Phase One sixth- and eighth-grade samples are summarized in Table F.2. The same three steps used in the analysis of school variables were used to determine which, if any, of the program variables served as significant predictors of student composite scores. In most cases, program variables were nominal variables in which each possible response option was coded by the research team in a dichotomous form (e.g., yes = 1, no = 0). However, the organization [of students] for instruction was an ordinal variable in which each possible response could have several values (e.g., whole class instruction could be ranked 1, 2, or 3, or 0 if not ranked). For this ranking item, the JMP program treated each possible response as a separate dichotomous variable (e.g., ranked 1st: yes = 1, and no = 0). This increased the number of independent (predictor) variables and added a layer of complexity to the reporting of these results.

For the first step in the analysis of the **Phase One sixth-grade data for Research Question a**, Table F.2 indicates that variables associated with five items in the Program Information Form (PIF) were significant in 20% or more of the repeated KFCV analyses ($p < .05$). These were:

- Organization of Teachers: school programs for which *self-contained* and *departmentalized* were selected;
- Organization [of Students] for Instruction: school programs for which *individualized* was ranked 1, 2 or not at all;
- Instructional Sites: school programs for which *computer labs* was not selected; Instructional Methods school programs for which *lecture* and *discussion* were not selected; and

Table F.2
Major Results of Regression Analyses for Research Question a: Program Variables by Composite Scores for the Phase 1 Sample

Step in Regression Analyses	**First Step**		**Second Step**		**Third Step**	
Analysis Method	Repeated K-Fold Cross Validation*		Basic Stepwise Regression*		Regression Bootstrapping*	
Grade	6	8	6	8	6	8
% Significant Models (100 runs)	96%	100%				
F-Ratio			4.007	5.228	2.770	4.6703
Significance			0.0012	0.0001	0.018	0.0010
R-Square Value			0.548	0.758	0.388	0.498
Adj. R-Square Value			0.412	0.613	0.248	0.391
Program Variables						
Major Goals and Objectives (Y/N):						
* Science		45% (NR)		N.S.		
* Social Studies						
* Communication Skills						
* HO Thinking Skills						
* Affective Dispositions		62% (NR)		N.S.		
* Problem/Issue Awareness						
* Investigation Skills		62% (NR)		N.S		
* Service/Action Skills						
Curriculum Organization (Y/N):						
* Separate Subjects						
* Common Themes/Sep. Subjects		99%		0.0015		+ 0.0106
* Common Themes/Integrated						
Organization of Teachers:						
* Self-Contained	42%		0.0020		N.S.	
* Departmentalized	42%		0.0009		- 0.0121	
* Team Teaching						
Organization for Instruction (Rank):						
* Whole Class						
* Groups/Teams		100% (Rank 1)		0.0003		- 0.0232
		62% (Rank 2)		0.0047		
		100% (NR)		0.0047		N.S.

Table F.2 (cont.)

Step in Regression Analyses	**First Step**		**Second Step**		**Third Step**	
Analysis Method	Repeated K-Fold Cross Validation*		Basic Stepwise Regression*		Regression Bootstrapping*	
Grade	6	8	6	8	6	8
Program Variables						
Organization for Instruction (cont.):						
* Individualized	39% (Rank 1)	100% (Rank 1)	N.S.	0.0108		
	42% (Rank 2)	100% (Rank 2)	0.0029	0.0002	N.S.	- 0.0001
	57% (NR)	100% (NR)	0.0009	0.0005	+ 0.0021	N.S.
Instructional Sites (Y/N):						
* Classrooms		64% (NR)		0.0040		
* Science Labs		99% (NR)		0.0215		
* Computer Labs	93% (NR)	65% (NR)	0.0009	N.S.	- 0.0029	
* Library/Media Center						
* School Grounds						
* Field Sites						
* Other Community Sites						
Instructional Methods (Y/N):						
* Lecture	39% (NR)	99% (NR)	0.0145	0.0001	N.S.	+ 0.0097
* Labs						
* Discussion	39% (NR)		0.0300		N.S.	
* Projects						
* Cooperative Learning						
* Inquiry						
* Hands-On						
* Service Learning						
Assessment Methods (Y/N)						
* Informal						
* Alternative	37% (NR)		0.0043			
* Traditional						
* Standardized	93% (NR)	99% (NR)	0.0009	0.0121	+ 0.0106	- 0.0025

*** Notes**: NR = no response; and N.S. = not statistically significant at $p < .05$

- Assessment Methods: school programs for which *standardized* or *alternative assessment* methods were not selected.

When these program variables were analyzed in the second step, only the ranking of *individualized* first as a way to Organize for Instruction, was not retained as a significant predictor of these student composite scores. Finally, when the remaining variables were analyzed in the third step, four variables remained as significant predictors of these scores. One of these was based on selected responses, while three were based on the absence of selected responses (i.e., NR = no response). The former included the use of *departmentalization* in school programs for the Organization of Teachers. However, this was found to be negative predictor of student composite scores. Two of the significant predictor variables that reflected the non-selection of a response was the use of *individualization* in the Organization [of Students] for Instruction and the use of *standardized* as an Assessment Method. The absence of responses for these items was found to be a positive predictor of these scores.

For the first step in the analysis of the **Phase One eighth-grade data for Research Question a**, Table F.2 indicates that variables associated with six items in the Program Information Form (PIF) were significant in 20% or more of the repeated KFCV analyses ($p < .05$). These were:

- Major Goals and Objectives: school programs for which *science*, *affective dispositions*, and *investigation skills* were not selected as major goals/objectives;
- Curriculum Organization: school programs that reported using *common themes in separate subjects*;
- Organization for Instruction: school programs for which *groups/teams* was ranked 1, 2 or not at all, and for which individualized was ranked 1, 2, or not at all;
- Instructional Sites: school programs for which *classrooms*, *science labs*, and *computer labs* were not selected;
- Instructional Methods: school programs for which *lecture* was not selected; and
- Assessment Methods: school programs for which *standardized* was not selected.

When these program variables were analyzed in the second step, four variables were not retained as significant predictors of these student composite scores: all three areas under Major Goals and Objectives, and *computer labs* as an Instructional Site. Finally, when the remaining variables were analyzed in the third step, five remained as significant predictors of these scores. Three of these were based on providing responses, while two were based on the absence of responses (i.e., NR = no response). The former included one positive predictor (*common themes in separate subjects* as the mode of Curriculum Organization), and two negative predictors (ranking the use of *teams/groups* first and i*ndividualized* second for the Organization of Instruction). Of the two significant predictor variables identified on the basis of the non-selection of responses, *lecture* as an Instructional Method was a positive predictor, while *standardized* as an Assessment Method was a negative predictor, of these student composite scores.

When these results for Phase One sixth- and eighth-grade students were compared, four program variables were found to be significant predictors of student composite scores in the second step for both sixth- and eighth-grade samples (i.e., under the Organization [of Students] for Instruction, ranking *individualized* second and not at all; under Instructional Methods, the non-selection of *lecture*; and under Assessment Methods, the non-selection of *standardized*). Of these four, only the non-selection of *standardized* as an Assessment Method was found to be a significant predictor of these scores for both grades in the third and final step. However, results indicated that this was a positve predictor of these scores for the sixth-grade sample but a negative predictor for the eighth-grade sample.

Results for Research Question b

As described in the previous sections (i.e., Methods for Research Questions a - d, and Results for Research Question a), three steps were taken in these linear multiple regression analyses. The results of all three steps for Research Question b, the effects of school variables on student composite scores for the Phase Two sixth-, seventh-, and eighth-grade samples, are summarized in Table F.3.

For the first step in the analysis of the **Phase Two sixth-grade data for Research Question b,** Table F.3 indicates that only one school variable was significant in 20% or more of the repeated KFCV analyses ($p < .05$): the percent of students in the Free Lunch Program, as an indicator of the socio-economic composition of these schools. When this school variable was analyzed in the second step, it remained a significant predictor of these students' composite scores. Finally, when this variable was analyzed in the third step, it remained as a significant predictor of student composite scores. However, results also indicated that this was a negative predictor of these scores (i.e., it had a negative regression coefficient).

For the first step in the analysis of the **Phase Two seventh-grade data**, Table F.3 indicates that two school variables were significant in 20% or more of the repeated KFCV analyses ($p < .05$): School Type III (i.e., whether or not a school was designated as a middle school); and the percent of students in the Reduced Lunch Program, also an indicator of the socio-economic composition of these schools. When these two school variables were analyzed in the second step, only the percent of students in the Reduced Lunch Program remained a significant predictor of these student composite scores. Finally, when this variable was analyzed in the third step, it remained as a significant predictor of these scores. However, results also indicated that this was a negative predictor of these scores (i.e., it had a negative regression coefficient).

For the first step in the analysis of the **Phase Two eighth-grade data**, Table F.3 indicates that none of the school variables were significant in 20% or more of the repeated KFCV analyses ($p < .05$). As a result, the research team could not proceed to the second step and third step in these analyses.

Table *F*.3

Major Results of Regression Analyses for Research Question b: School Variables by Composite Scores for the Phase 2 Sample

Step in Regression Analyses	**First Step**			**Second Step**			**Third Step**		
Analysis Method	Repeated K-Fold Cross Validation*			Basic Stepwise Regression*			Regression Bootstrapping*		
Grade	6	7	8	6	7	8	6	7	8
% Significant Models (100 runs)	100%	23%	0%						
F-Ratio				18.77	3.707	NA	18.775	6.059	NA
Significance				0.0001	0.0356	NA	0.0001	0.019	NA
R-Square Value				0.337	0.188	NA	0.337	0.155	NA
Adj. R-Square Value				0.319	0.137	NA	0.319	0.129	NA
School Variables									
School Type									
* I: Public (1) – Private (2)									
* II: Magnet/Charter (1) – Not (2)									
* III: Middle School (1) – Not (2)		24%			N.S.				
SES Composition of School									
* % Reduced Lunch		93%			0.014			- 0.019	
* % Free Lunch	100%			0.0001			- 0.0001		

*** Notes**: NR = no response; and N.S. = not statistically significant at p < .05

When these results for Phase Two sixth-, seventh-, and eighth-grade students were compared, the absence of any significant results for the eighth-grade sample indicated that no school variable(s) served as significant predictors of student composite scores across all three grades. Further, while there were significant results for the sixth- and seventh-grade samples, the closest the research team came to identifying any significant predictor(s) for these grades was the identification of socio-economic indicators of the composition of schools as significant but negative predictors of these scores (i.e., 6^{th} grade: the percent of students in a *Free Lunch* Program; and 7^{th} grade: the percent of students in a *Reduced Lunch* Program).

For the first step in the analysis of the **Phase Two sixth-grade data**, Table F.4 indicates that variables associated with six items in the Program Information Form (PIF) and in the Screening Survey were significant in 20% or more of the repeated KFCV analyses ($p < .05$). These were:

- Program Intensity: exposure to the school program, in hours per weeks, as reported on the Phase Two Screening Survey;
- Major Goals and Objectives: school programs for which *higher order thinking skills* was not selected as a major goal/objective;
- Curriculum Organization: school programs for which *separate subjects* and *common themes integrated across subjects* were selected;
- Organization of Teachers: school programs for which *team-teaching* was selected;
- Instructional Sites: school programs for which *classrooms*, *science labs*, *computer labs, the library/media center, school grounds, and field sites* were not selected;
- Instructional Methods: school programs for which *lecture* and *inquiry* were not selected; and
- Assessment Methods: school programs for which no Assessment Method was *ranked 4th*.

When these program variables were analyzed in the second step, the only variable that was not retained as a significant predictor of these student composite scores was *team teaching* under Organization of Teachers. Finally, when these remaining variables were analyzed in the third step, the only variable that remained as a significant predictor of these scores was *separate subjects* under Curriculum Organization. However, this variable was a negative predictor of these scores (i.e., it had a negative regression coefficient).

For the first step in the analysis of the **Phase Two seventh-grade data**, Table F.4 indicates that none of the program variables were significant in 20% or more of the repeated KFCV analyses. A more careful review of results indicated that only 11 of the 100 KFCV models were found to be valid and significant prediction models, which indicates that few, if any, of the variables within these models would be significant predictors. Due to this, no further analyses could be conducted for the second step or the third step.

Table *F*.4

Major Results of Regression Analyses for Research Question b: Program Variables by Composite Scores for the Phase 2 Sample

Steps in Regression Analysis	**First Step**			**Second Step**			**Third Step**		
Analysis Method	Repeated K-Fold Cross Validation*			Basic Stepwise Regression*			Regression Bootstrapping*		
Grade	6	7	8	6	7	8	6	7	8
% Significant Models (100 runs)	100%	11%	96%						
F-Ratio				22.058	NA	1.364	4.548	NA	NA
Significance				0.0001	NA	0.278	0.021	NA	NA
R-Square Value				0.947	NA	0.484	0.275	NA	NA
Adj. R-Square Value				0.905	NA	0.129	0.214	NA	NA
Program Variables									
Program Intensity	41%			0.0026			N.S.		
Program Duration			95%			N.S.			
Program Goals and Objectives (Y/N):									
* Science			96% (NR)			N.S.			
* Social Studies									
* Communication Skills			95% (NR)			N.S.			
* HO Thinking Skills	100% (NR)			0.0013			N.S.		
* Affective Dispositions									
* Problem/Issue Awareness									
* Investigation Skills									
* Service/Action Skills			96% (NR)			N.S.			
Curriculum Organization (Y/N):									
* Separate Subjects	100%			0.0001			- 0.0067		
* Common Themes/Sep. Subjects									
* Common Themes/Integrated	33%			0.0175			N.S.		
Organization of Teachers (Y/N):									
* Self-Contained									
* Departmentalized									
* Team Teaching	59%			N.S.					

Table F.4 (cont.)

Steps in Regression Analysis	**First Step**			**Second Step**			**Third Step**		
Analysis Method	Repeated K-Fold Cross Validation*			Basic Stepwise Regression*			Regression Bootstrapping*		
Grade	6	7	8	6	7	8	6	7	8
Program Variables									
Organization for Instruction (Rank):									
* Whole Class			95% (Rank 3)			N.S.			
* Groups/Teams			48% (Rank 3)			N.S.			
* Individualized									
Instructional Sites (Y/N):									
* Classrooms	100% (NR)			0.0093			N.S.		
* Science Labs	100% (NR)			0.0007			N.S.		
* Computer Labs	92% (NR)		96% (NR)	0.0104		N.S.	N.S.		
* Library/Media Center	59% (NR)			0.0113			N.S.		
* School Grounds	100% (NR)		96% (NR)	0.0373		N.S.	N.S.		
* Field Sites	92% (NR)		96% (NR)	0.0332		N.S.	N.S.		
* Other Community Sites			100% (NR)			N.S.			
Instructional Methods (Y/N):									
* Lecture	100% (NR)			0.0001			N.S.		
* Labs									
* Discussion									
* Projects									
* Cooperative Learning									
* Inquiry	100% (NR)			0.0001			N.S.		
* Hands-On									
* Service Learning									
Assessment Methods (Rank)	41% (Rank 4 - NR)			0.044			N.S.		
* Informal									
* Alternative			96% (Rank 2)			N.S.			
* Traditional									
* Standardized									

*** Notes**: NR = no response; and N.S. = not statistically significant at $p < .05$

For the first step in the analysis of the **Phase Two eighth-grade data**, Table F.12 indicates that variables associated with five items in the Program Information Form (PIF) and in the Screening Survey were significant in 20% or more of the repeated KFCV analyses ($p < .05$). These were:

- Program Duration: exposure to the school program, in weeks per school year, as reported on the Phase Two Screening Survey;
- Major Goals and Objectives: school programs for which *science, communication skills, and service/action skills* were not selected as a major goal/objective;
- Organization [of Students] for Instruction: school programs for which *whole class* and *groups/teams* were each ranked 3rd;
- Instructional Sites: school programs for which *computer labs, school grounds, field sites*, and *other community sites* were not selected; and
- Assessment Methods: school programs for which *alternative assessment* was ranked 2nd.

When these program variables were analyzed in the second step, none were retained as significant predictors of these student composite scores. Due to this, no further analyses could be conducted for the third step.

When these results for Phase Two sixth-, seventh-, and eighth-grade students were compared, what stood out was that only one program variable was found to be significant in the analysis of sixth grade data (i.e., under Curriculum Organization, the use of *separate subjects*, which was found to be a negative predictor), and none were found to be significant in the analysis of either seventh or eighth grade data.

Results for Exploratory Research Question c

Research Question c pertains to the effects of teacher and student variables on composite scores for the Phase One sixth- and eighth-grade samples. In Phase One, each school was asked to select one sixth- and one eighth-grade class. For this reason, the number of participating schools and teachers was nearly equal in each grade (i.e., 6^{th}: 48 schools and 50 teachers; 8^{th}: 48 schools and 50 teachers). Thus, in analyses for Research Question c, the school served as the unit of analysis. For Research Question c, data for three predictor variables were collected on the MSELS: age, gender, and ethnicity. However, MSELS responses were available for each student, so component and composite scores for each student were calculated. Thus, in analyses for Research Question c, the individual student served as the unit of analysis.

As described in previous sections (i.e., Methods for Research Questions a. – d., and Results for Research Question a), three steps were taken in these linear multiple regression analyses. The results of all three steps for Research Question c., the effects of teacher variables on student composite scores for the Phase One sixth- and eighth-grade samples, are summarized in Table F.5.

Table *F*.5
Major Results of Regression Analyses for Research Question c: Teacher Variables by Composite Scores for the Phase 1 Sample

Steps in Regression Analysis	First Step		Second Step		Third Step	
Analysis Method	Repeated K-Fold Cross Validation*		Basic Stepwise Regression*		Regression Bootstrapping*	
Grade	6	8	6	8	6	8
% Significant Models (100 runs)	100%	100%				
F-Ratio			7.996	2.800	6.354	6.413
Significance			0.0012	0.0176	0.015	0.0007
R-Square Value			0.285	0.474	0.121	0.453
Adj. R-Square Value			0.249	0.304	0.102	0.382
Teacher Variables						
Teacher Certification:						
* Yes						
* No		70%		N.S.		
* Working On		63%		0.0440		N.S.
Highest Degree:						
* Bachelor's		86%		0.0075		- 0.003
* Master's	45%	99%	0.0105	0.02	+ 0.0152	N.S.
* Master's + 30						
* Specialist						
Highest Degree, Level of Schooling:						
* Early Childhood	93%		0.004		N.S.	
* Elementary		100%		N.S.		
* Middle						
* Secondary		84%		N.S.		
* Multiple						
* Other						
# of Teacher Inservices:						
* Total						
* < 1 Day						
* 1-2 Days		79%		0.0075		+ 0.002
* 3-7 Days		79%		0.02		- 0.003
* > 1 week						

Table F.5 (cont.)

Steps in Regression Analysis	**First Step**		**Second Step**		**Third Step**	
Analysis Method	Repeated K-Fold Cross Validation*		Basic Stepwise Regression*		Regression Bootstrapping*	
Grade	6	8	6	8	6	8
Teacher Variables						
Teacher Age:						
* < 21						
* 21-30						
* 31-40						
* 41-50		86%		0.002		- 0.0001
* 51-60						
* > 60						

* **Notes**: NR = no response; and N.S. = not statistically significant at p < .05

For the first step in the analysis of the **Phase One sixth-grade data for Research Question c,** Table F.5 indicates that variables associated with two items in the Teacher Information Form (TIF) were significant in 20% or more of the repeated KFCV analyses ($p < .05$). These were:

- Highest Degree Earned: teacher completion of a *Master's degree*; and
- Highest Degree Earned, Level of Schooling: completion of a degree in the area of *early childhood* education.

When these teacher variables were analyzed in the second step, both of these variables were retained as significant predictors of these student composite scores. Finally, when these two variables were analyzed in the third step, the only variable that remained as a significant predictor of these scores was *Master's degree* under Highest Degree Earned. This variable was a positive predictor of these scores (i.e., it had a positive regression coefficient)

For the first step in the analysis of the **Phase One eighth-grade data**, Table F.5 indicates that variables associated with five items in the Teacher Information Form (TIF) were significant in 20% or more of the repeated KFCV analyses ($p < .05$). These were:

- Teacher Certification: whether teachers were *working on* a teaching certificate or were *not certified*;
- Highest Degree Earned: teacher completion of a *Bachelor's* and a *Master's degree*;
- Highest Degree Earned, Level of Schooling: completion of a degree in the area of *elementary education* and *secondary education*.
- Number of Teacher Inservices: the number of EE-related inservices that lasted *1-2 days* and that lasted *3-7 days*;
- Teacher Age: whether a teacher's age ranges from *41-50* years old.

When these teacher variables were analyzed in the second step, six of these variables were retained as significant predictors of these student composite scores (i.e., the three non-significant variables under Teacher Certification: *not certified*; and under Highest Degree Earned - Level of School: both *elementary* and *secondary education*). Finally, when these six variables were analyzed in the third step, four variables remained as significant predictors of these scores. The only positive predictor variable was the number of teacher inservices that lasted *1-2 days* under Number of Teacher Inservices. The three variables found to be negative predictors of these scores were: completion of a *Bachelor's degree* under Highest Degree Earned; the number of EE-related inservices that lasted *3-7 days* under Number of Teacher Inservices; and teacher age in the *41-50* age range under Teacher Age (i.e., each had a negative regression coefficient).

When these results for Phase One sixth- and eighth-grade students are compared, what stands out is that only one teacher variable was found to be significant in the second step of the analysis for both grades (i.e., whether teachers had earned a *Master's* degree), although this did not remain a significant predictor in the third step for the eighth grade.

For the <u>first step</u> in the analysis of the **Phase One sixth-grade data for Research Question c**, Table F.6 indicates that variables associated with two items in the MSELS were significant in 20% or more of the repeated KFCV analyses (p < .05). These were:

- Student Age: whether these sixth-graders were *11, 12 or 13* years old at the time of this assessment; and
- Student Ethnicity: whether these sixth graders were African-American (*Black*).

The number of student predictor variables is very small in comparison to the number of School, Program, and Teacher variables. Further, individual students served as the unit of analysis here, rather than schools as the unit of analysis (i.e., for School, Program and Teacher variables). Therefore, due to the small number of predictor variables and large sample size (n = 964) there was no need to determine the stability of results using the <u>second step</u>. Thus, when these four student variables were analyzed in the <u>third step</u>, all four were found to be significant predictors of these individual student scores. Two of these student variables were found to be positive predictors (i.e., under Student Age: students who were *11 and 12*), and two were found to be negative predictors (i.e., under Student Age, students who were *13*, and under Student Ethnicity, student who were *Black,* both had negative regression coefficient).

For the <u>first step</u> in the analysis of the **Phase One eighth-grade data**, Table F.6 indicates that variables associated with two items in the MSELS were significant in 20% or more of the repeated KFCV analyses (p < .05). These were:

- Student Gender: whether these sixth-graders were *female*; and
- Student Ethnicity: whether these sixth graders were African-American (*Black*).

As noted above, there was no need to determine the stability of results for the eighth grade (n = 917) using the <u>second step</u>. Thus, when these two student variables were analyzed in the <u>third step</u>, both were found to be significant predictors of these individual student scores. One of these student variables was found to be a positive predictor (i.e., under Student Gender: *female*), and one was found to be a negative predictor (i.e., under Student Ethnicity, student who were *Black*). The former has a positive and the latter a negative regression coefficient.

When these results for Phase One sixth- and eighth-grade students were compared, what stood out was that only one student variable was found to be significant in the <u>second step</u> of the analysis for both grades (i.e., whether students were *Black*; for both grades, it was found to be a negative predictor.

Table *F*.6

Major Results of Regression Analyses for Research Question c: Student Variables by Composite Scores for the Phase 1 Sample

Steps in Regression Analysis	**First Step**		**Second Step**		**Third Step**	
Analysis Method	Repeated K-Fold Cross Validation*		Basic Stepwise Regression*		Regression Bootstrapping*	
Grade	6	8	6	8	6	8
% Significant Models (100 runs)	100%	100%				
F-Ratio					20.969	40.4272
Significance					0.0001	0.0001
R-Square Value					0.0826	0.082
Adj. R-Square Value					0.0786	0.080
Student Variables						
Age:						
* 11	100%				+ 0.0042	
* 12	100%				+ 0.0172	
* 13	100%				- 0.011	
* 14						
* 15						
Gender:						
* Male						
* Female		100%				+ 0.0224
Ethnicity:						
* Native Am./Alaskan Native						
* Asian/Pacific Islander						
* Hispanic						
* Black	100%	100%			- 0.0001	- 0.0001
* Caucasian						

* **Notes**: NR = no response; and N.S. = not statistically significant at p < .05

Results for Exploratory Research Question d

Research Question d pertain to the effects of teacher and student variables on composite scores for the Phase Two sixth-, seventh-, and eighth-grade samples. The analysis for Research Question d was comparable to the analysis for Research Question c due the small number of student demographic items on the MSELS.

However, the analysis for Research Question d was not as simple as for Research Question c. In Phase One (Research Question c), there was almost a 1-to-1 correspondence between schools and teachers. However, in Phase Two (Research Question d), the number of participating teachers from each school varied from 1 to as many as 5. Thus, the steps taken to prepare the teacher data files for Research Question d. were more involved. The first step involved matching teacher names at the top of Teacher Information Forms (TIFs) to teacher names at the top of student Scantron forms. To ensure that teachers who completed a TIF could be matched to their students who had completed the MSELS, only those teachers whose names appeared on both TIF and student Scantron forms were included in this data file and these analyses (Grade 6: 57 teachers from 35 schools; Grade 7: 42 teachers from 30 schools; and Grade 8: 32 teachers from 23 schools). The team's inability to make these matches resulted in the loss of both TIF and MSELS data. In turn, the number of schools in the data set was reduced to a size that would no longer allow schools to serve as a viable unit of analysis in multiple linear regression analyses. For this reason, individual teachers were used as the unit of analysis for Research Question d.

For the <u>first step</u> in the analysis of the **Phase Two sixth-grade data for Research Question d**, Table F.7 indicates that variables associated with two items in the MSELS were significant in 20% or more of the repeated KFCV analyses ($p < .05$). These were:

- Teacher Certification Level: whether or not teachers were certified to teach at the *secondary* level; and
- Number of Teacher Inservices: the number of EE-related inservices that lasted *1-2 days.*

When these teacher variables were analyzed in the <u>second step</u>, only teacher certification at the *secondary* level was retained as a significant predictor of these student composite scores. Finally, when this variable was analyzed in the <u>third step</u>, it remained as a significant predictor of these scores, although this was found to be a negative predictor (i.e., it had a negative regression coefficient).

For the <u>first step</u> in the analysis of the **Two seventh-grade** and of the **eighth-grade data**, Table F.7 indicates that none of the variables associated with items in the Teacher Information Form (TIF) were significant in 20% or more of the repeated KFCV analyses ($p < .05$). As a result, the research team could not proceed to the second step and third step in these analyses for either the seventh- or eighth-grade sample. Further, in light of this, a comparison of results across all three grades indicates that no teacher variables were found to serve as predictors of these student composite scores.

Table *F*.7
Major Results of Regression Analyses for Research Question d: Teacher Variables by Composite Scores for the Phase 2 Sample

Steps in Regression Analysis	**First Step**			**Second Step**			**Third Step**		
Analysis Method	Repeated K-Fold Cross Validation*			Basic Stepwise Regression*			Regression Bootstrapping*		
Grade	6	7	8	6	7	8	6	7	8
% Significant Models (100 runs)	97%	0%	0%						
F-Ratio				6.643	NA	NA	6.746	7.808	NA
Significance				0.0028	NA	NA	0.0121	0.008	NA
R-Square Value				0.213	NA	NA	0.113	0.167	NA
Adj. R-Square Value				0.181	NA	NA	0.096	0.145	NA
Teacher Variables									
Teacher Certification, Level:									
* Elementary									
* Middle									
* Secondary	98%			0.0026			- 0.012		
* Multiple									
* Other									
Highest Degree, Subject Area:									
* Science									
* Social Studies/History									
* English/Language Arts									
* Math									
* Multiple									
* Other									
* None									
# of Teacher Inservices:									
* Total									
* < 1 Day	93%			N.S.					
* 1-2 Days									
* 3-7 Days									
* > 1 week									

Table F.7 (cont.)

Steps in Regression Analysis	**First Step**			**Second Step**			**Third Step**		
Analysis Method	Repeated K-Fold Cross Validation*			Basic Stepwise Regression*			Regression Bootstrapping*		
	6	7	8	6	7	8	6	7	8
Teacher Variables									
Age:									
* < 21									
* 21-30									
* 31-40									
* 41-50									
* 51-60					- 0.0139			- 0.008	
* > 60									

* **Notes**: NR = no response; and N.S. = not statistically significant at p < .05

For the first step in the analysis of the **Phase Two sixth-grade data for Research Question d**, Table F.8 indicates that variables associated with three items in the MSELS were significant in 20% or more of the repeated KFCV analyses ($p < .05$). These were:

- Student Age: whether these sixth-graders were *14* years old at the time of this assessment;
- Student Gender: whether these sixth graders were *female*; and
- Student Ethnicity: whether these sixth graders were *Asian/Pacific Islander*, *Black*, and *Caucasian*.

As mentioned for Research Question c, there was no need to determine the stability of results for the sixth grade (n = 3,128) using the second step. Thus, when these student variables were analyzed in the third step, all five of these variables remained as significant predictors of these scores. Of these, *female, Asian/Pacific Islander,* and *Caucasian* were positive predictors (i.e., they had positive regression coefficients), while the other two variables were negative predictors of these scores.

For the first step in the analysis of the **Phase Two seventh-grade data**, Table F.8 indicates that student variables associated with three items in the MSELS were significant in 20% or more of the repeated KFCV analyses ($p < .05$). These variables were:

- Student Age: whether these seventh graders were *12* and *13* years old at the time of this assessment;
- Student Gender: whether these seventh graders were *female*; and
- Student Ethnicity: whether these seventh graders were *Asian/Pacific Islander*, African-American (*Black*), and *Caucasian*.

As noted above, there was no need to determine the stability of results for the seventh grade (n = 2,678) using the second step. Thus, when these student variables were analyzed in the third step, all six of these variables remained as significant predictors of these scores. Of these, five variables were found to be positive predictors of these scores (i.e., to have positive regression coefficients); only *Black* was found to be a negative predictor.

For the first step in the analysis of the **Phase Two eighth-grade data**, Table F.8 indicates that variables associated with three items in the MSELS were significant in 20% or more of the repeated KFCV analyses ($p < .05$). These were:

- Student Age: whether these eighth graders were *13* and *14* years old at the time of this assessment;
- Student Gender: whether these eighth graders were *female*; and
- Student Ethnicity: whether these eighth graders were *Hispanic*, *Black*, and *Caucasian*.

As noted above, there was no need to determine the stability of results for the eighth grade (n = 1,853) using the second step. Thus, when these student variables were analyzed in the

Table *F*.8
Major Results of Regression Analyses for Research Question d: Student Variables by Composite Scores for the Phase 2 Sample

Steps in Regression Analysis	First Step			Second Step			Third Step		
Analysis Method	Repeated K-Fold Cross Validation*			Basic Stepwise Regression*			Regression Bootstrapping*		
Grade	6	7	8	6	7	8	6	7	8
% Significant Models (100 runs)	100%	100%	100%						
F-Ratio							33.87	45.5026	53.278
Significance							0.0001	0.0001	0.0001
R-Square Value							0.058	0.1008	0.164
Adj. R-Square Value							0.056	0.098	0.161
Student Variables									
Age:									
* 11									
* 12		100%						+ 0.0001	
* 13		100%	100%					+ 0.0001	+ 0.0001
* 14	100%		100%				- 0.0011		+ 0.0001
* 15									
Gender:									
* Male									
* Female	100%	100%	100%				+ 0.0001	+ 0.0001	+ 0.0001
Ethnicity:									
* Native Am./Alaskan Native									
* Asian/Pacific Islander	100%	100%					+ 0.0001	+ 0.0194	
* Hispanic			100%						- 0.008
* Black	100%	100%	100%				- 0.0001	- 0.0001	- 0.0001
* Caucasian	100%	100%	100%				+ 0.0001	+ 0.0001	+ 0.0001

* **Notes**: NR = no response; and N.S. = not statistically significant at p < .05

third step, all six of these variables remained as significant predictors of student composite scores. Of these, four variables were found to be positive predictors of these scores (i.e., to have positive regression coefficients). The negative predictors of these scores were the two variables under Ethnicity, *Hispanic* and *Black* (i.e., these had negative regression coefficients).

When these results for Phase Two sixth-, seventh-, and eighth-grade students were compared, what stood out was that three variables were found to be significant predictors of student composite scores in all three grades. Of these, under Gender, *female*, and under Ethnicity, *Caucasian*, were found to be positive predictors of these scores, while under Ethnicity, *Black* was found to be a negative predictor of these scores. Both student Age and student affiliation with other Ethnicities served as predictors of these scores, but not on a consistent basis across all three grades.