Summer 7-28-2015

Ninth Grade Student Responses to Authentic Science Instruction

Michael Steven Ellison

Portland State University

Let us know how access to this document benefits you.
Follow this and additional works at: http://pdxscholar.library.pdx.edu/open_access_etds

Recommended Citation

This Dissertation is brought to you for free and open access. It has been accepted for inclusion in Dissertations and Theses by an authorized administrator of PDXScholar. For more information, please contact pdxscholar@pdx.edu.
Ninth Grade Student Responses to Authentic Science Instruction

by

Michael Steven Ellison

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Environmental Sciences and Resources

Dissertation Committee:
William G. Becker, Chair
Dalton Miller-Jones
Ronald Narode
Linda A. George
Richard C. Hugo

Portland State University
2015
ABSTRACT

This mixed methods case study documents an effort to implement authentic science and engineering instruction in one teacher’s ninth grade science classrooms in a science-focused public school. The research framework and methodology is a derivative of work developed and reported by Newmann and others (Newmann & Associates, 1996). Based on a working definition of authenticity, data were collected for eight months on the authenticity in the experienced teacher’s pedagogy and in student performance. Authenticity was defined as the degree to which a classroom lesson, an assessment task, or an example of student performance demonstrates construction of knowledge through use of the meaning-making processes of science and engineering, and has some value to students beyond demonstrating success in school (Wehlage et al., 1996). Instruments adapted for this study produced a rich description of the authenticity of the teacher’s instruction and student performance.

The pedagogical practices of the classroom teacher were measured as moderately authentic on average. However, the authenticity model revealed the teacher’s strategy of interspersing relatively low authenticity instructional units focused on building science knowledge with much higher authenticity tasks requiring students to apply these concepts and skills. The authenticity of the construction of knowledge and science meaning-making processes components of authentic pedagogy were found to be greater, than the authenticity of affordances for students to find value in classroom activities beyond demonstrating success in school. Instruction frequently included one aspect of value beyond school, connections to the world outside the classroom, but students were
infrequently afforded the opportunity to present their classwork to audiences beyond the teacher.

When the science instruction in the case was measured to afford a greater level of authentic intellectual work, a higher level of authentic student performance on science classwork was also measured. In addition, direct observation measures of student behavioral engagement showed that behavioral engagement was generally high, but not associated with the authenticity of the pedagogy. Direct observation measures of student self-regulation found evidence that when instruction focused on core science and engineering concepts and made stronger connections to the student’s world beyond the classroom, student self-regulated learning was greater, and included evidence of student ownership.

In light of the alignment between the model of authenticity used in this study and the Next Generation Science Standards (NGSS), the results suggest that further research on the value beyond school component of the model could improve understanding of student engagement and performance in response to the implementation of the NGSS. In particular, it suggests a unique role environmental education can play in affording student success in K-12 science and a tool to measure that role.
ACKNOWLEDGMENTS

I am deeply grateful to the many people that supported me and made this work possible. Chief among them are the following...

- Without my wife’s support, love, and patience I could not have done this. My daughters has been loving, encouraging and supportive.

- My committee chair, William Becker, along with Dalton Miller-Jones selflessly shared their invaluable expertise, advice, and encouragement from start to finish. I appreciate the advice and support of my other committee members, Ronald Narode, Linda George, and Richard Hugo. I am also very grateful for Patrick Edwards’ expert help.

- The teacher who welcomed me into his classroom and professional life over several years made this work possible. I learned an enormous amount about quality teaching from my time in his classroom. And I must mention the students that put up with an unusual presence in their classroom.

- The leadership and other teachers in the school within which this study was conducted were supportive and encouraging of my work and presence among them.

- Fellow science educators Carol Ramsey and Jim Martin spent many hours with me scoring student work. Their assistance and expertise was essential.

- The National Science Foundation provided financial support for this research through the Center for Teaching and Learning in the West program.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chapter 1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>1.1</td>
<td>The Status of U.S. Science, Technology, Engineering, and Mathematics (STEM) Education</td>
<td>1</td>
</tr>
<tr>
<td>1.2</td>
<td>Purpose</td>
<td>3</td>
</tr>
<tr>
<td>1.3</td>
<td>Research Questions</td>
<td>5</td>
</tr>
<tr>
<td>1.3.1</td>
<td>Correlation Study Propositions</td>
<td>6</td>
</tr>
<tr>
<td>1.4</td>
<td>Brief Description of Study</td>
<td>6</td>
</tr>
<tr>
<td>1.5</td>
<td>Overview of the Dissertation</td>
<td>8</td>
</tr>
<tr>
<td>Chapter 2</td>
<td>Review of Literature</td>
<td>10</td>
</tr>
<tr>
<td>2.1</td>
<td>Educating For Authentic Intellectual Development</td>
<td>10</td>
</tr>
<tr>
<td>2.2</td>
<td>Situated Cognition, Cultural Practices and Authenticity</td>
<td>12</td>
</tr>
<tr>
<td>2.3</td>
<td>Authenticity and Science Education Reform</td>
<td>13</td>
</tr>
<tr>
<td>2.3.1</td>
<td>Authenticity as Discourse in a Community of Practice</td>
<td>16</td>
</tr>
<tr>
<td>2.3.2</td>
<td>Authenticity as Meaningfulness, Relevance, and Interest</td>
<td>18</td>
</tr>
<tr>
<td>2.3.3</td>
<td>Authenticity as Connections to the World Beyond the Classroom</td>
<td>19</td>
</tr>
<tr>
<td>2.4</td>
<td>Engagement and Authentic Achievement</td>
<td>21</td>
</tr>
<tr>
<td>2.4.1</td>
<td>Engaging High School Youth from Groups Underrepresented in STEM</td>
<td>22</td>
</tr>
<tr>
<td>2.4.2</td>
<td>A Model of Engagement in Academic Work in Secondary School</td>
<td>23</td>
</tr>
<tr>
<td>2.5</td>
<td>Engagement and Self-Regulated Learning</td>
<td>24</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Behavioral Engagement</td>
<td>25</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Cognitive Engagement</td>
<td>25</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Self-Regulated Learning</td>
<td>25</td>
</tr>
<tr>
<td>2.6</td>
<td>Implementation of Authentic Inquiry in K-12 Science Education</td>
<td>26</td>
</tr>
<tr>
<td>2.7</td>
<td>An Authentic Pedagogy model from the School Reform Movement</td>
<td>27</td>
</tr>
<tr>
<td>2.8</td>
<td>Research Gaps</td>
<td>29</td>
</tr>
<tr>
<td>Figures</td>
<td></td>
<td>34</td>
</tr>
<tr>
<td>Chapter 3</td>
<td>Theoretical Perspective</td>
<td>36</td>
</tr>
<tr>
<td>3.1</td>
<td>Affordance</td>
<td>37</td>
</tr>
<tr>
<td>3.2</td>
<td>Authentic Science Pedagogy and Student Performance</td>
<td>37</td>
</tr>
<tr>
<td>3.3</td>
<td>Intellectual Aspects of Authentic Science Pedagogy</td>
<td>39</td>
</tr>
</tbody>
</table>
Chapter 5

5.1 Methods ................................................................. 67
  5.1.1 Measures of authentic pedagogy in the teacher’s assessment tasks 68
  5.1.2 Measures of authentic pedagogy from observations of classroom instruction 73
  5.1.3 Data Analyses .................................................. 83

5.2 Results ........................................................................ 91
  5.2.1 The teacher’s assessment tasks .................................. 92
  5.2.2 Classroom instruction across the entire study period .......... 97
  5.2.3 Classroom instruction in selected instructional units ......... 100
  5.2.4 Summary of differences in authenticity between selected instructional units ........................................ 106

5.3 Discussion ................................................................. 110
  5.3.1 Overall Authenticity Levels ....................................... 112
  5.3.2 Authenticity trends over the school year ...................... 114
  5.3.3 The teacher’s curricular strategies in the end-of-trimester project units ................................................. 116
  5.3.4 Prominent features of authenticity in the teacher’s practice .... 118
5.3.5 Do the affordances in the teacher’s assessment tasks align with those in the teacher’s classroom instruction? ..................123
5.3.6 Comparison with previous research..................................124
Tables ........................................................................128
Figures ........................................................................143
Endnotes ........................................................................149

Chapter 6
Student Responses to Authentic Pedagogy over a School Year ..........153
6.1 Methods ..................................................................154
  6.1.1 Measures of behavioral engagement and self-regulation........154
  6.1.2 Behavioral Engagement Data Analyses..........................161
  6.1.3 Self-Regulated Learning Data Analyses.........................165
  6.1.4 Measures of authenticity of student performance in science work samples ..........................................................168
  6.1.5 Authenticity of Student Performance Data Analyses ..........175
6.2 Results ..................................................................178
  6.2.1 Behavioral engagement levels over the entire study period ......178
  6.2.2 Trends in behavioral engagement during the school year .....179
  6.2.3 Differences in behavioral engagement between instructional units.................................................................181
  6.2.4 Self-regulated Learning levels during the study period ......183
  6.2.5 Self-regulated learning levels in four selected instructional units ........................................................................184
  6.2.6 Authentic Student performance levels ..........................187
6.3 Discussion ................................................................191
  6.3.1 Behavioral engagement and self-regulated learning ............191
  6.3.2 Authentic performance .............................................197
Tables ........................................................................208
Figures ........................................................................217
Endnotes ........................................................................222

Chapter 7
Relationships between authentic pedagogy and student performance, engagement, and self-regulated learning ..................225
7.1 Methods ..................................................................227
  7.1.1 Proposition 1 - Relationships between authenticity of classroom instruction and student behavioral engagement and self-regulated learning ...........................................227
  7.1.2 Proposition 2 – Relationships between authentic pedagogy and authentic student performance ...................................229
7.2 Results ..................................................................233
  7.2.1 Proposition 1 - Relationships between authenticity of classroom instruction and student behavioral engagement and self-regulated learning ...........................................233
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 4.1</td>
<td>Variables measured in the descriptive study.</td>
<td>64</td>
</tr>
<tr>
<td>Table 4.2</td>
<td>Examples of alignment between this study’s model of authentic pedagogy and student performance and the standards for pedagogy and school structure and culture identified by the Expeditionary Learning Schools network as core benchmarks.</td>
<td>65</td>
</tr>
<tr>
<td>Table 5.1</td>
<td>Variables overview for Research Question 1.</td>
<td>128</td>
</tr>
<tr>
<td>Table 5.2</td>
<td>Teacher proficiency assessment tasks scored for authenticity. The tasks are listed in chronological order. Tasks noted as alternatives are an option or an additional chance for students to show proficiency on the same Learning Target or Targets (LTs).</td>
<td>129</td>
</tr>
<tr>
<td>Table 5.3</td>
<td>Elements of authentic science pedagogy measured in the teacher’s assessment tasks.</td>
<td>132</td>
</tr>
<tr>
<td>Table 5.4</td>
<td>Instructional units observed over the study period from October to June. All the instructional units observed and scored for authenticity of classroom instruction at least one day are included. The units are listed in chronological order.</td>
<td>133</td>
</tr>
<tr>
<td>Table 5.5</td>
<td>Instructional units observed frequently enough to compare authentic pedagogy measures of the teacher’s classroom instruction with the corresponding assessment tasks. The units are listed in chronological order.</td>
<td>134</td>
</tr>
<tr>
<td>Table 5.6</td>
<td>Description of Lesson Authenticity Indicators (LAIs) and associated components of authentic science pedagogy scored during each lesson segment scored for authenticity of classroom instruction.</td>
<td>135</td>
</tr>
<tr>
<td>Table 5.7</td>
<td>Construction of the scores for the four components of authentic science pedagogy in classroom instruction from Lesson Authenticity Indicator (LAI) scores during each lesson segment.</td>
<td>137</td>
</tr>
<tr>
<td>Table 5.8</td>
<td>Task Authenticity Indicator (TAI) scores for selected teacher assessment tasks from across the entire study period. Tasks are ordered by Combined Authenticity score, then chronological order during the school year.</td>
<td>138</td>
</tr>
</tbody>
</table>
Table 5.9 Normalized component and combined task authenticity scores for selected teacher assessment tasks. Tasks are ordered by Combined Authenticity score, then chronological order during the school year. ..........139

Table 5.10 Numbers of lesson segments (LSs) and hours of class time scored for authenticity of classroom instruction from each school trimester. All lesson segments were focused on science and/or engineering instruction. .................................................................140

Table 5.11 Normalized mean authenticity of classroom instruction scores for lesson segments (LSs) observed over the entire study period. .................140

Table 5.12 Combined authenticity of classroom instruction for seven selected instructional units with sufficient observations to determine a reliable estimate of authenticity. Units are listed in chronological order. .............141

Table 5.13 Normalized means (0-1) for the eleven Lesson Authenticity Indicators used to calculate the four components of authentic classroom instruction. Instructional units are listed in chronological order. ..........142

Table 6.1 Variables overview for Research Question 2. ...............................................................208

Table 6.2 Description of the assessment tasks corresponding to the student work samples analyzed for authenticity of student performance. .........................209

Table 6.3 Elements of authentic science and engineering performance measured in student science work samples. .................................................................210

Table 6.4 Mean and median percent of observations behaviorally engaged (BE) per student for each school trimester for all students and divided by gender........................................................................................211

Table 6.5 Mean and median levels of percent of observations behaviorally engaged (BE) per student during seven instructional units from October to June in chronological order. Number of lesson segments (LSs) observed during each unit is included..............................................211

Table 6.6 Percent of all observations recording the presence of any type of self-regulated learning (SR) during each trimester and the entire study period for all students and divided by gender.........................................................212

Table 6.7 Percent of all observations recording the presence of any type of self-regulated learning (SR) during four selected units.................................................................212

Table 6.8 Percent of observations (N) recording the presence of any type self-regulated learning (SR) during subsets of observations from selected
instructional units and the entire study period. The first subset of observations is from lesson segments when the class activity has all students grouped as a whole class. The second subset is when students were working in small groups or individually. ........................................ 213

Table 6.9 Percent of self-regulated learning category variable recorded as working autonomously measured during selected instructional units. .................. 214

Table 6.10 Number of work samples (N) scored from each demographic group for each the assessment tasks from the four selected instructional units. .... 215

Table 6.11 Mean authenticity of student science performance measurements for work samples collected from four selected instructional units. For each instructional unit, means for the three Performance Authenticity Indicators (PAIs) and the Combined Authenticity are listed. ................. 216

Table 7.1 Variables in the correlation analysis for Proposition 1 ........................................ 244

Table 7.2 Variables in the correlation analysis for Proposition 2. ................................. 245

Table 7.3 Correlation coefficients (r) for classroom instruction Combined Authenticity and component authenticity variables and behavioral engagement (%BE/LS) and self-regulated learning (% presence of SR/LS). ................................................................. 246

Table 7.4 Correlation coefficients (r) for classroom instruction Lesson Authenticity Indicator variables (LAIs) and behavioral engagement (% BE/LS) and self-regulated learning (% presence of SR/LS). LAIs are categorized by the authenticity component they contribute to. Only LAIs with significant correlations are listed. ........................................ 247

Table 7.5 Correlation coefficients (r) for authenticity variables for the teacher’s assessment tasks and authenticity variables for student performance on work samples. ................................................................. 248

Table 7.6 Correlation coefficients (r) for authenticity of the teacher’s classroom instruction variables and authenticity of student performance on work samples. The classroom instruction variables are mean values for the four instructional units corresponding to the student work samples........... 249
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1 Pictorial representation of the Center for Organization and Restructuring of Schools (CORS) model of authentic pedagogy and student performance (Wehlage et al., 1996).</td>
<td>34</td>
</tr>
<tr>
<td>Figure 2.2 The Newmann model of student engagement in academic work in secondary schools (Newmann, et al., 1992). Figure reproduced with permission of the copyright holder.</td>
<td>35</td>
</tr>
<tr>
<td>Figure 3.1 Model of authentic science pedagogy and student performance developed for use in this study. All three components are necessary for the classroom environment to be highly authentic. Construction of Knowledge and Science Meaning-Making Processes are characteristics of the reasoning practices of science and engineering and ideas considered. The Value Beyond School component expresses the affordance provided when students recognize value in their classwork beyond demonstrating success in school.</td>
<td>49</td>
</tr>
<tr>
<td>Figure 3.2 Conceptual framework for this study. The variables and relationships explored between them are illustrated. Hypothesized positive relationships are shown. The other arrows indicate directional and non-directional relationships documented in previous research that are examined to the extent possible in this study. Authenticity of pedagogy includes authenticities of these components in classroom instruction and in the teacher’s assessments.</td>
<td>50</td>
</tr>
<tr>
<td>Figure 5.1 Combined Authenticity scores of 22 assessment tasks in chronological order from October to June. Authenticity levels are normalized to a 0-1 scale, where 0 equals not authentic, 0.5 is moderately authentic, and 1 is highly authentic. Each assessment task name is followed by the teacher’s Learning Target(s) it assesses in parentheses. The integers before the decimal point indicate the trimester.</td>
<td>143</td>
</tr>
<tr>
<td>Figure 5.2 Scores of three the components of authentic pedagogy for 22 teacher assessment tasks in chronological order from October to June. Each component is labeled with its normalized score. Authenticity levels for each component are normalized to a 0-1 scale, where 0 equals not authentic, 0.5 is moderately authentic, and 1 is highly authentic. The maximum possible score on the authenticity axis is 3. Each assessment task name is followed by the teacher’s Learning Target(s) it assesses in parentheses. The integers before the decimal point indicate the trimester.</td>
<td>144</td>
</tr>
</tbody>
</table>
Figure 5.3 Percent class time observed at classroom instruction authenticity levels for Combined Authenticity and each component from lesson segments observed across the entire study period. Numeric labels are percents. Authenticity levels increase clockwise from 12 o’clock position. The Science Inquiry/Engineering Design scale is a sub-component of the Science Meaning-Making Processes component. Percent of class time at different authenticity levels varies between components. ..........................145

Figure 5.4 Mean authenticity for the four components of classroom instruction and Combined Authenticity for the lesson segments scored during each of seven selected instructional units. Means are calculated from lesson segment scores weighted by the duration of the lesson segment. All scores are normalized to a scale of 0-1. All authenticity scores are normalized to a scale of 0-1, where 0 is no authenticity, 0.5 is moderately authentic, and 1 is highly authentic.................................146

Figure 5.5 Comparison of Combined Authenticity of assessment tasks with mean Combined Authenticity of classroom instruction from seven instructional units observed frequently enough to reliably estimate mean authenticity of the instruction. Instructional units are in chronological order. Classroom instruction scores are means of scores calculated from lesson segments weighted by the duration of the lesson segment. All scores are normalized to a scale of 0-1. ........................................................................147

Figure 5.6 Levels of components of authentic pedagogy in assessment tasks and classroom instruction during seven selected instructional units and corresponding assessment tasks. Classroom instruction authenticity scores are means of scores from individual lesson segments weighted by their length. Each component is labeled with its normalized score. The maximum on the Y-axis for the assessment tasks is 3.0 because 3 components are measured to construct the combined score. The maximum on the Y-axis for the classroom instruction is 4.0 because 4 components are measured.................................................................148

Figure 6.1 Boxplots of fraction of observations behaviorally engaged per student for each class period from each trimester.........................................................217

Figure 6.2 Percent of all individual observations recording the presence of any type of self-regulated learning during four instructional units selected for observation of a wide variety of the classroom activities during each unit..................................................................................218

Figure 6.3 Percent of self-regulated learning (SR) category variable observations recorded as working autonomously for each gender during two trimesters........................................................................219
Figure 6.4 Boxplots of student work sample Combined Authenticity scores for four instructional units. All of the authenticity scores are normalized to a scale of 0-1......................219

Figure 6.5 Mean combined authentic performance for four instructional units for each gender. All of the authenticity scores are normalized to a scale of 0-1. Error bars are ± 1 SD. The numbers of males (M) and females (F) are listed for each instructional unit.................................................................220

Figure 6.6 Mean Combined Authenticity for assessment tasks and student performance on work samples from four selected instructional units.........221

Figure 7.1 Relationship of mean Combined Authenticity of student performance on science work samples and Combined Authenticity of the teacher’s corresponding assessment tasks. The numbers of work samples averaged for each assessment task are given below the chart. All scores are normalized to a scale of 0-1.................................................................250

Figure 7.2 Relationship of mean Combined Authenticity of student performance on science work samples and mean Combined Authenticity of the teacher’s classroom instruction for the corresponding instructional units. Mean classroom instruction values for each instructional unit are computed from lesson segment scores weighted by the duration of the lesson segment. The numbers of work samples averaged for each instructional unit are given below the chart. All scores are normalized to a scale of 0-1.................................................................251
GLOSSARY

Affordances: environmental features that are perceived by organisms to present certain categories of activities. Affordance is used in this study for cognitive, sociocultural, and psychological environmental cues intentionally implemented in the classroom and curriculum.

Assessment task: written teacher definition of learning products students were to complete as the culmination of instructional units. Evaluation of the student product was used to decide whether student was able to demonstrate proficiency on the instructional unit’s learning target or targets. The level of proficiency demonstrated was used to assign grades.

Behavioral engagement: student attentiveness to and involvement in classroom activities. Variable defined as behaviorally engaged or not behaviorally engaged.

Class period: A regularly recurring time period during a school day when a teacher meets with the same group of students. Each school day includes a series of different class periods.

Construction of Knowledge: A component of the authentic science pedagogy and performance in this study. It is reasoning practices that promote construction rather than simple reproduction of scientific knowledge. Classroom science instruction and assignments afford Construction of Knowledge when they ask students to use higher-order thinking to synthesize, evaluate or organize complex information and to consider alternative solutions or explanations to address a fundamental science concept, a problem, or an issue.
**Instructional unit:** A period of instruction during a school trimester defined by one or more of the teacher’s learning targets. Each instructional unit was focused on the specific limited science content and skills defined by the learning target or targets.

**Learning Target:** specific learning proficiencies stated in language students understand. Each course had 5-8 learning targets for each trimester. Each learning target defined the knowledge and skills that students must demonstrate proficiency on to receive a grade and credit for an academic course.

**Lesson Authenticity Indicator (LAI):** Each of the 11 characteristics of the teacher’s classroom instruction scored for level of authentic affordances described in Table 5.6.

**Lesson segment:** Time period during a class period focused on a distinct activity with a particular interaction pattern. Each class period observed was divided into two or more lesson segments.

**Performance Authenticity Indicator (PAI):** Each the three characteristics of student performance on their science work samples scored for level of authenticity as described in Table 6.3.

**Science Meaning-Making Processes:** A component of authentic science pedagogy and performance in this study. It is the use of reasoning processes similar to those employed by science practitioners to establish new knowledge. Science Meaning-Making Processes include using core science concepts and scientific and engineering practices of inquiry and communication.

**Self-regulated learning:** Students exercise autonomy, agency, or ownership related to classroom activities as evidenced by self-initiation and self-direction.
Self-regulated learning presence variable: Records the presence or absence any type of evidence of self-regulation.

Self-regulated learning category variable: Records whether self-regulation, when observed, is (1) limited to independently following the teacher’s direction for an activity or task or (2) shows autonomy or ownership that goes beyond following the teacher’s direction.

Student work sample: The artifact produced by a student and given to the teacher for grading in response to the teacher’s expectations defined by an assessment task.

Task Authenticity Indicator (TAI): Each of the seven characteristics of the teacher’s assessment tasks scored for level of authentic affordances described in Table 5.3.

Trimester: each of the three 12-week periods that comprised the school year. Academic courses were completed each trimester. Grades were assigned to each course at the end of each trimester.

Value Beyond School: A component of authentic science pedagogy in this study. It is defined by connections conveyed by instruction and activities to problems, issues, and interests that are part of or likely to be of students’ experiences in their lives beyond the classroom. It also includes communicating ideas, producing a product, investigating a problem, or engaging in a performance for an audience beyond the teacher and the classroom.
Chapter 1  Introduction

Engaging all students in an authentic science education experience by integrating the fundamental ideas of science and engineering with the practices employed by scientists and engineers is now promoted as the primary strategy for science education to graduate science literate citizens (National Research Council, 1996, 2000, 2012; S. R. Singer, Hilton, & Schweingruber, 2005). Authentic school science inquiry is envisioned to create contexts where students can engage in meaning-making processes as much like those of science practitioners as possible (J. S. Brown, Collins, & Duguid, 1989; H.-S. Lee & Songer, 2003; McGinn & Roth, 1999).

Science inquiry and engineering design are cognitively, socially, and technologically complex activities that require sustained engagement from students. Accepting science and technological literacy for all students as the aim of K-12 science education and authentic science inquiry and engineering design as the primary strategy presents educators with the complex challenge of providing both cognitive and affective support for a high degree of cognitive engagement (H.-S. Lee & Songer, 2003; National Research Council, 2012). However, engagement and authenticity are also widely acknowledged as uncommon experiences for K-12 science students (S. R. Singer et al., 2005).

1.1  The Status of U.S. Science, Technology, Engineering, and Mathematics (STEM) Education

National policymakers consistently discuss science, technology, engineering, and mathematics (STEM) education as a national priority. Science education in the U.S. is
called upon to meet our needs for a growing STEM workforce and for citizens to
cnfidently and intelligently deal with STEM-laden dilemmas they will face. In spite of
over a century of earnest discussion on why and how to improve U.S. science education
(S. R. Singer et al., 2005), U.S. students consistently rank in the lower half in
international comparisons of science achievement and the gaps widen in high school. In
the 2009 National Assessment of Educational Progress (NAEP) results, only 21% of
twelfth grade students were judged “proficient” in science (National Center for Education

Long-term results for the NAEP science achievement test also show persistent
disparities in average scores by racial/ethnic subgroups. At grade 12, Euro American and
Asian/Pacific Islander students on average score significantly higher than their Black,
Hispanic, and American Indian counterparts (National Center for Education Statistics,
2011). However, since 1969 the achievement advantage males held over females has
narrowed significantly (S. R. Singer, et al., 2005), but has not been closed (National
Center for Education Statistics, 2011).

Rodriguez (1998) argues that because the groups that often achieve poorly are
already a significant proportion of our population (about one-third), and are the fastest
growing ethnic groups in the country, both a concern for social justice and for a strong
economic future require attention to this social inequity. The cultural boundaries these
students must cross as they enter science classrooms or STEM careers are often more
distinct and more challenging than for students from cultural groups that make up most
STEM practitioners and most science teachers (S. R. Singer et al., 2005).
In the recent past, serious and sustained national scale efforts to reform U.S. K-12 science education go back at least 25 years to the advent of the American Association for the Advancement of Science’s Project 2061 (Project 2061 (American Association for the Advancement of Science), 1989). However, the overall results of reform are still dismal. “Massive amounts of money, time, and effort associated with these reforms have effected little overall change in scientific literacy” (McGinn & Roth, 1999, p. 14).

American high schools struggle to achieve science literacy for all by engaging students in science inquiry and engineering design integrated with fundamental disciplinary concepts. According to the National Research Council’s 2005 report on the state of high school science laboratory experiences, most people in the U.S. do not understand science well enough to make informed decisions about the many science-laden public policy issues they will face, and high schools do not improve students’ science literacy (S. R. Singer et al., 2005). One important reason cited is that “typical” high school science laboratory experiences are poorly integrated into development of understanding science concepts or practices. In addition, they conclude that science laboratory experiences, which should be a prime context for authentic science inquiry, are commonly of poor quality, are ineffective in generating interest in science, and are inequitably distributed.

1.2 Purpose

This study addresses two broad questions. 1) What does a high school science curriculum that intentionally attempts to provide students with an authentic science
experience look like? 2) Does engaging students in a more authentic experience produce gains in science achievement, as well as, student engagement and self-regulation?

The purpose of this study is to document an effort to implement authentic science instruction in high school classrooms and examine the influence of experiencing authentic science pedagogy on the quality of students’ performance in science classwork, and student engagement, and self-regulation.

The theoretical model of authentic pedagogy adopted in this study includes both intellectually authentic instruction and instruction that helps students find value in their classwork beyond just showing how successful they can be in school. This investigator choose this model based on interest in how environmental education integrates with this “value beyond school” component of authenticity. As will be noted, the case selected for study was chosen with this in mind.

The update in thinking about national science education standards in A Framework for K-12 Science Education calls for a “more…authentic science education experience for all students” (National Research Council, 2012, p. 265). It calls for integrating fundamental disciplinary concepts with science inquiry and engineering design practices. Though conducted before A Framework for K-12 Science Education was published, this study documents an implementation of this goal for improving science curriculum.

The framework document calls for greater focus on engineering and technology in K-12 science education (National Research Council, 2012). This study documents teacher pedagogy and student performance in a curriculum that integrated both science and engineering. In this study science pedagogy and performance on science classwork
are used for brevity, but always measures science and/or engineering pedagogy whichever is present.

A Framework for K-12 Science Education includes a research agenda on the implementation and evolution of its recommendations. “Perhaps most important, research is needed on classroom-level contexts, materials, and discourses that engage and support a wider range of students in high-quality teaching and learning experiences” (National Research Council, 2012, p. 325). This study assessed the nature of the science and engineering pedagogy and students’ academic behaviors in high school classrooms.

A Framework for K-12 Science Education highlights the need for understanding the achievement of students when they experience authentic science instruction (National Research Council, 2012). This study was conducted in a high school that intentionally recruited students from the diverse cultural groups within the community and those that would be first generation higher education students.

Research on engagement in high schools has been conducted (Marks, 2000; National Research Council, 2003), but little has been focused on the unique characteristics of science classrooms. High school settings are important because engagement in academically demanding tasks competes with a larger array of nonacademic opportunities than exist for most middle school students, where secondary science education research is concentrated.

1.3 Research Questions

What does a high school science curriculum that intentionally attempts to provide students with an authentic science experience look like?
1. How does the presence of features of authentic science pedagogy vary over the school year?

2. How do the levels of authentic student performance, engagement, and self-regulated learning vary over the school year?

   Does engaging students in a more authentic experience produce gains in science achievement, engagement, and self-regulated learning?

3. How are levels of authentic student performance, engagement, and self-regulated learning related to the features of authentic science instruction?

1.3.1 Correlation Study Propositions

   The second broad objective of exploring student science achievement, engagement, and self-regulation in the studied classrooms is addressed by correlation analyses examining the following propositions.

   **Proposition 1**

   The level of authentic science pedagogy students experience will be positively related to student behavioral engagement and self-regulation.

   **Proposition 2**

   The level of authentic science pedagogy students experience will be positively related to the level of authentic student performance.

1.4 Brief Description of Study

   This mixed methods case study is a year-long investigation in ninth grade science classrooms in a high school with a population that includes a high proportion of students
from groups traditionally underrepresented in science, technology, engineering, and mathematics (STEM) careers. The school and the science class were intentionally designed to provide students with an authentic learning experience. The science education literature and the work of the Center on Organization and Restructuring of Schools (CORS) are used to identify pedagogical characteristics of authentic instruction. These components include provision of affordances for construction of knowledge (Donovan & Bransford, 2005; Tobin & Tippins, 1994; Wehlage, Newmann, & Secada, 1996); the use of the meaning-making processes of science (Linn & Eylon, 2006; Wehlage et al., 1996) and connecting tasks with the students’ lives and communities outside of the classroom, so that students are afforded the opportunity to find value in classroom activities beyond demonstrating success in school (Basu & Barton, 2007; Roth & Barton, 2004; Tomasek, 2006; Wehlage et al., 1996). These components align well with the recommendations of A Framework for K-12 Science Education (National Research Council, 2012).

The presence of specific components of authentic science pedagogy in instruction and assessment tasks are documented from October to June over a school year. Systematic classroom observations of classroom instruction are evaluated relative to a model of authentic science pedagogy (Research Question 1). Document analysis of teacher assessment tasks and curriculum goals and teacher interviews are used to measure levels of authenticity of pedagogy and compare them with measures from classroom observations.

Student behavioral engagement and self-regulated learning during classroom instruction was measured from classroom direct observations concurrently with measures
of the instruction’s authenticity. Analysis of student classwork measure student performance relative to the authentic of the pedagogy (Research Question 2).

Correlational analyses explore the relationships between the components of authentic science in the teacher’s instruction and students’ engagement, self-regulated learning, and their science achievement (Research Question 3).

1.5 **Overview of the Dissertation**

This introductory chapter has described the need for research on the quality of science instruction in high school classrooms and how students respond to efforts to bring instruction in line with current science education goals. The research questions addressed by the study have been described.

Chapter 2 presents a review of how the science education and school reform literature defines authentic pedagogy, the importance of engagement to successfully learn science, and gaps in the research regarding these.

Chapter 3 describes the theoretical perspective taken by this study. The specific model of authentic science pedagogy that defines the measures of classroom instruction and student performance in science classwork in this study is detailed. How the theoretical perspective is used to define the conceptual framework of the study wraps up this chapter.

Chapter 4 presents the overall case study design. This chapter introduces the case study by describing the school, the teacher and students that participated in the study. Then the limitations of the study are delineated.
Chapters 5 and 6 present the descriptive study. Chapter 5 reports on the study defined by Research Question 1, how the features of authentic science pedagogy varied in the case over the study period. The study defined by Research Question 2, how the levels of student engagement, self-regulated learning, and performance on classwork varied over the study period, is reported in Chapter 6. Each of these chapters includes the methods used and results obtained. The findings for each research question are discussed in light of the purpose of the study.

Chapter 7 addresses the correlation study in Research Question 3. The methods used to explore the relationships between the authenticity of the instruction students experience and student performance on their science classwork, their engagement, and self-regulation are described. The results and their importance in light of the purpose of the study are discussed.

Chapter 8 summarizes the important patterns found and presents conclusions regarding the study’s two broad research questions, as well as their limitations. Chapter 9 discusses the questions raised by this study and how they may be addressed in future research.
Chapter 2  Review of Literature

This literature review establishes the need for greater student engagement to meet national goals for science education reform. It describes the appeals for “authenticity” by science education reformers and explores models of authentic student performance. The relationship between authentic science pedagogy and the engagement of high school students is considered. The gaps in knowledge about this relationship addressed by this study are identified.

Research on engagement in high schools has been conducted (Marks, 2000; National Research Council, 2003), but little has been focused on the unique characteristics of science classrooms. High school settings are important because engagement in academically demanding tasks competes with a larger array of nonacademic opportunities than exist for most middle school students, where secondary science education research is concentrated.

What aspects of research on learning have influenced and been embraced by science education researchers and reformers as they address the challenge of engaging young people in the difficult cognitive work of bringing their naïve understandings into correspondence with scientific understandings of the natural world?

2.1 Educating For Authentic Intellectual Development

It is generally agreed that a central goal of school is to prepare students for successful participation in what competent adults actually do (Wehlage et al., 1996). Education researchers, and science education researchers, in particular, have addressed
characteristics of curriculum, instruction, and classroom culture that are engaging to all students and develop the complex intellectual capabilities required to function in the adult world.

Recent research in education and cognitive psychology favors a view of learners as active constructors of knowledge within social communities (Bransford, Brown, & Cocking, 1999; Tobin & Tippins, 1994). Science education research supports the value of science instruction that recognizes people are constantly creating and refining mental structures as they interact with the world (Donovan & Bransford, 2005; Linn & Eylon, 2006; J. Singer, Marx, Krajcik, & Chambers, 2000). The constructivist view proposes that people require multiple encounters with new ideas and information that challenge current conceptions to modify their current understandings (Bransford et al., 1999; Krajcik, Czerniak, & Berger, 2003). Curriculum focused on construction of knowledge addresses the development of the complex intellectual skills needed by competent adults (Bransford et al., 1999). Developing critical adult reasoning skills requires practice addressing complex, realistic, practical dilemmas (Linn, 2000; Wehlage et al., 1996).

The degree of science literacy needed to comprehend topics like the consequences of human induced environmental change, requires fostering intellectual development that includes understanding the scientific explanations of the natural world. In addition, science literacy requires understanding the reasoning processes and techniques science practitioners use to generate explanations and to judge their trustworthiness (National Research Council, 2000).

Constructivist pedagogy is now considered central to developing the intellectual skills students will need as science literate adults (Donovan & Bransford, 2005). Science
education research has acknowledged the challenge of developing pedagogical
techniques to scaffold knowledge construction that changes the naïve conceptions of
children about the natural world into models more like those accepted by scientists
(Donovan & Bransford, 2005; Hunt & Minstrell, 1994; Linn & Eylon, 2006). Taking the
time for students to develop coherent explanations through investigation and reflective
discussion results in consideration of fewer curriculum topics, but when the topics
considered are core science ideas, achievement can improve (AREA, 2007; Marx et al.,
2004).

Sustained and fruitful effort has addressed this complex pedagogical goal of
developing scientific habits of mind (AREA, 2007; Donovan & Bransford, 2005;
National Research Council, 2000, 2002; J. Singer et al., 2000). Inquiry is promoted as
the most effective strategy for the construction of new knowledge through investigations
using reasoning practices like those of science practitioners (National Research Council,
2000). The National Science Education Standards (NSES) define inquiry in classroom
practice as “the activities of students in which they develop knowledge and understanding
of scientific ideas, as well as an understanding of how scientists study the natural world”

2.2 Situated Cognition, Cultural Practices and Authenticity

Situated cognition theorists’ belief that learning cannot be separated from the
learning context has influenced science education research (H.-S. Lee & Songer, 2003;
McGinn & Roth, 1999). The situated cognition perspective stresses the importance of the

Situated cognition theorists also discuss the concept of authenticity. Situated cognition theory defines authentic activity as “ordinary practices of culture” by which “meanings and purposes are socially constructed through negotiations among present and past members” (J. S. Brown, et al., 1989, p. 34). Authentic commonly describes the true or genuine, as opposed to the misleading or artificial (Wehlage et al., 1996). Authenticity as the cultural practices of a discipline refers to the actual reasoning and discourse practices it uses to construct and share knowledge and meaning. These practices may be said to embody authentic human achievement (Wehlage et al., 1996).

Using this sense of authenticity, authentic schoolwork, to the extent appropriate and possible, situates learning in the cognitive and social practices of competent adults (J. S. Brown et al., 1989; Lave & Wenger, 1991; Resnick, 1987). Such a cognitive, social, and even physical learning environment is considered authentic. Used this way, authenticity promotes the educational goal of developing the intellectual capacities of competent adults through construction, rather than reproduction, of knowledge as discussed above.

2.3 Authenticity and Science Education Reform

The social, cultural, cognitive, and even motivational aspects and implications of this concept of authenticity have been prominent in the work of some science education researchers (Barab & Hay, 2001; Basu & Barton, 2007; H.-S. Lee & Songer, 2003; McGinn & Roth, 1999; Roth, 1995). When authenticity is used in science education
research it usually refers to correspondence with what scientists actually do (Barab & Hay, 2001). Science education researchers also define authenticity by reference to the particular reasoning processes (Chinn & Malhotra, 2002) and other problem solving and technological tools of professional scientists (National Research Council, 1996, 2000). They focus on the meaning-making processes scientists use to construct knowledge and share understanding about the natural world.

Science education researchers frequently connect authenticity and inquiry, with calls for authentic science inquiry (H.-S. Lee & Songer, 2003). Sandoval (2005) describes authentic inquiry as “using scientific methods to construct scientific knowledge” (p. 649). This connects authentic inquiry to calls for constructivist science pedagogy. Students who spent a large amount of time in inquiry activities have shown improved thinking skills and mastery of science content (Marx et al., 2004; National Research Council, 2000).

Situated cognition theorists’ belief that learning cannot be separated from the learning context has also influenced the meaning of authenticity in science education research (H.-S. Lee & Songer, 2003; McGinn & Roth, 1999). Researchers propose that active engagement with natural phenomena is a uniquely important aspect of science learning context (Krajcik et al., 2003). They claim construction of knowledge is enhanced when students ask and refine questions related to natural phenomena, make predictions and explain phenomena in the context of interactions with the actual phenomena. This study treats interactions with engineering designs as equivalent to those with natural phenomena.
This view of authenticity emphasizes the reasoning processes of scientists as social and cultural aspects of scientific practice (Roth, 1995; J. Singer et al., 2000). Some science education researchers speak of the cultural practices of scientists as the key aspect of authenticity (McGinn & Roth, 1999; Roth & Barton, 2004).

This is also the basis for another common connotation of authenticity. When the National Science Education Standards (NSES) called for “inquiry into authentic questions generated from student experiences” as “the central strategy for teaching science” (National Research Council, 1996, p. 31), they connected authenticity to respect for student perceptions that spring from their personal experiences and interests, influenced by their culture. This view, often expressed as meaningfulness or relevance, relates authenticity to the degree of connection to student experiences and interests, and to “real world” issues, problems, and contexts outside the classroom. The value of this meaning of authenticity is based on the constructivist view that effective pedagogy builds new understandings based on the current conceptions of learners (Bransford et al., 1999; Donovan & Bransford, 2005) and the belief that relevance and meaning support motivation (Newmann, Wehlage, & Lamborn, 1992).

To summarize, the concept of authenticity has been connected by science education researchers to social construction of knowledge through inquiry in a community of practice, and to connections to the “real world”, including its cultural elements, that provide relevance and meaning for students. This review will now consider these in more detail.
2.3.1 Authenticity as Discourse in a Community of Practice

McGinn and Roth (1999) use sociocultural studies of science to define authenticity. They emphasize the necessity of social construction of knowledge through inquiry and meaning-making discourse.

In classrooms modeled after science as practiced, students pursue investigations of their own interests, negotiate with their collaborators as to problem and solution frames, and debate the merits of different processes for seeking solutions. "Authentic science" requires that students pursue their activities under the constraint that they make their actions and products accountable to themselves, their peers, and their teachers; that is, classrooms are organized as knowledge-producing communities in which rhetorical dimensions similar to those in science are enacted (p. 17).

Science education researchers have explored the critical role of discourse which resembles the negotiations that occur among science practitioners highlighted by the sociocultural perspective on authenticity (A. L. Brown & Campione, 1994; Edelson, 1998; Linn & Eylon, 2006; Roth, 1995; J. Singer et al., 2000). Social negotiation in inquiry that makes students thinking visible has been shown to promote knowledge construction and retention through knowledge integration (Linn, 2000). Research demonstrates that social learning communities provide an especially rich environment for scaffolding knowledge construction (A. L. Brown & Campione, 1994; Vellom, Anderson, & Palincsar, 1993) and promoting higher-order thinking during inquiry (Wehlage et al., 1996).
Studies of inquiry instruction that provides a prominent role for discourse in meaning-making have shown positive results (Marx et al., 2004; Rivet & Krajcik, 2004; Roth, 1995). Involving students in discourse around problem solving and investigating the natural world supports construction of knowledge and understanding of the collaborative and iterative nature of scientific inquiry (Krajcik, Blumenfeld, Marx, & Soloway, 2000; Krajcik et al., 2003; White & Frederiksén, 1998). Wehlage, Newmann and Gamoran’s (1996) conception of authenticity defines extended classroom discourse during inquiry on core disciplinary content as a component of authenticity. Lee and Songer (2003) propose discourse that constructs a common understanding through negotiation, presentation, and communication characteristic of science as a way to add authenticity to tasks.

Creation of such a community of practice through scaffolding discourse that makes student thinking visible has been demonstrated (A. L. Brown & Campione, 1994; Scardamalia & Bereiter, 1991; Scardamalia, Bereiter, & Lamon, 1994). Improved science meaning-making in these communities has been demonstrated (Bell & Linn, 2000; Linn, 2000). Technology tools have also been successfully used to scaffold science reasoning practices by making student thinking visible and facilitating communication that supports a community of practice (Linn & Eylon, 2006).

Metacognitive reflection has been associated with knowledge construction and improved science performance (White & Frederiksén, 2000). Classroom collaborative discourse provides many opportunities for metacognitive reflection when students are asked to consider each other’s explanations or perspectives (Krajcik et al., 2000).
2.3.2 Authenticity as Meaningfulness, Relevance, and Interest

Authenticity is often connected to the meaningfulness and relevance of classroom activities and topics. McGinn and Roth (1999) connect authenticity with student choice and interest, a common theme in research on authenticity. Brown, Collins, and Duguid (1989) describe authentic activities as “coherent, meaningful, and purposeful” (p. 34) based largely on the social context. Welhage, Newmann, and Secada (1996) characterize “tasks that are considered meaningful, valuable, significant, and worthy of one’s effort, in contrast to those considered nonsensical, useless, contrived, trivial, and therefore unworthy of effort” (Welhage et al., 1996, p. 23), as authentic work.

The science education research literature emphasizes the advantages of students investigating their own questions (Collins, Brown, & Newman, 1989; Rahm, Miller, Hartley, & Moore, 2003; Roth, 1995; J. Singer et al., 2000). Linn and Eylon report that “design of inquiry around personally relevant questions (Linn & Hsi, 2000) and topics that evoke passion (Collins et al., 2004) has positive cognitive outcomes” (Linn & Eylon, 2006, p. 526). Reiser, Smith, Tabak et al. (2001) argue that research shows that science inquiry learning not contextualized in students’ own questions results in shallow understanding.

Some researchers use “contextualization” to describe how to add relevance and meaning to classroom activities. Rivet and Krajcik define contextualization as “the utilization of particular situations or events that occur outside of science class or are of particular interest to students to motivate and guide the presentation of science ideas and concepts” (Rivet & Krajcik, 2008, p. 80).
2.3.3 Authenticity as Connections to the World Beyond the Classroom

Connections to the ‘real world’ outside the school are a common theme in science education research and are frequently connected to authenticity (H.-S. Lee & Songer, 2003; Rahm et al., 2003; Rahm, Moore, & Martel-Reny, 2005; Roth & Barton, 2004). Newmann et al. (1992) claim that authenticity “depends largely” on connection to the “real world” (p. 26). These connections are envisioned in various ways, some of which overlap with what has already been discussed.

The most common connections are activities that mention or focus on topics or questions concerning issues or phenomena outside the classroom, thought to be recognizable to students. Authentic real world problems are complex and open-ended. Successful implementation has been shown to require sustained inquiry on the order of weeks, instead of days (O’Neill & Polman, 2004; Polman, 2000; Rahm et al., 2003; Rivet & Krajcik, 2008; Roth, 1995). Roth (1995) and Singer et al. (2000) have documented how extended inquiry within the classroom, based on student generated questions, can produce the kind of complex reasoning characteristic of science practitioners. These activities may make connections to student experience, and scaffold a community of practice in the classroom (Cognition and Technology Group at Vanderbuilt, 1992; Krajcik et al., 1998; J. Singer et al., 2000).

The value of these connections is thought to be both cognitive, supporting construction of knowledge, and motivational, creating interest and value in students’ eyes (Barlia & Beeth, 1999). Contextualization in the world beyond the classroom is thought to provide motivational support for engagement through interest and ownership. Connections to the real world allow construction of knowledge to be built on students’
prior knowledge (Donovan & Bransford, 2005; Krajcik et al., 2003). The authenticity of such investigations lies in the opportunities they provide to use the inquiry processes of science to construct knowledge around meaningful and fundamental science content and produce products with value to students beyond demonstrating success in school (Basu & Barton, 2007; Rahm et al., 2005; Wehlage et al., 1996).

Environmental education is the source of many of the premier examples of enhancing the authenticity of learning environments by connecting them to the world beyond the classroom. Curricula and programs using environmental investigations to contextualize learning exhibit combinations of all the examples of connections to the real world discussed here. Investigations in the real world of the local environment have been related to increased student enthusiasm, interest, engagement, science achievement scores, understanding of the nature of science, and transfer of knowledge to new situations (George & Becker, 2003; Lieberman & Hoody, 1998; Rivet & Krajcik, 2008; J. Singer et al., 2000; Tomasek, 2006).

Researchers at the Center on Organization and Restructuring of Schools (CORS) proposed a comprehensive model of authentic student performance illustrated in Figure 2.1 that explicitly included connections to the world beyond the classroom (Wehlage et al., 1996). Their model proposed that such connections increase student perceptions of meaningfulness and relevance. One of the ways these connections are operationalized in the CORS model is the creation of meaningful learning products for an audience beyond the school.
2.4 Engagement and Authentic Achievement

Without sustained cognitive engagement, the complexity of the reasoning of science and engineering practitioners will be inaccessible to students (National Research Council, 2003). In spite of fruitful research on designing science inquiry experiences that support conceptual change (Donovan & Bransford, 2005; Linn & Eylon, 2006; National Research Council, 2000), a large proportion of students exiting U.S. high schools demonstrate poor understanding of how to use science knowledge to address academic or real world dilemmas (Linn & Eylon, 2006). Research on typical science laboratory experiences in the U.S. showed that they only slightly increased students’ interest in science and contributed only slightly to the development of scientific reasoning (S. R. Singer et al., 2005). The National Research Council’s (2005) report on the state of high school science laboratory experiences adopts as one of the goals for these experiences “cultivating interest in science and interest in learning science” (p. 3).

In 2008 student apathy was one of the top two student problems identified on a nationwide survey of middle and high school math and science teachers, with 32% of them calling describing it as “serious” (National Science Board, 2008). Authenticity has been proposed to improve engagement (Newmann et al., 1992). Generating meaning and value through challenge and connecting learning products to the world outside of the classroom is the mechanism proposed. Marks (2000) demonstrated an association between engagement and authentic intellectual work of middle and high school students as conceived in the model of authentic pedagogy developed by the Center on Organization and Restructuring of Schools.
2.4.1 Engaging High School Youth from Groups Underrepresented in STEM

Researchers have explored the relationship between authenticity and engaging urban minority youth in science, demonstrating engagement in challenging science activities when students are supported to connect science to their values, cultural backgrounds, and everyday experiences (Basu & Barton, 2007; Rahm, 2002; Rahm et al., 2005; Seiler, Tobin, & Sokolic, 2001). These researchers call for using these students’ funds of knowledge as resources for framing scientific investigations (Basu & Barton, 2007). I interpret this as an example of connecting with the world outside the classroom. Basu and Barton (2007) found that when student learning environments are shaped by how students envision science in their own futures, by the social activity structures they value, and by putting into practice their ideas about the purpose of science, students develop a sustained interest in science, and show evidence of self-regulated learning.

These studies provide evidence that students’ perceptions that learning and using science can give them valuable resources for their lives, can motivate cognitive engagement, and promote learning. However, these findings were all in informal science settings outside of the regular classroom. Roth and Barton (2004) argue for situating school science in problems accessible and useful for citizens in their everyday lives as an equity strategy. They find that students who were labeled “learning disabled” “…actually do many things we normally associate with high ability, if only the contexts of their activities are appropriate” (p. 130).

Research has related science inquiry instruction to improved achievement among all student groups (Handelsman et al., 2004; Lieberman & Hoody, 1998), including urban students of color (Marx et al., 2004). Acknowledged, but less frequently investigated, are
students’ perceptions that science schoolwork has value in their lives and communities outside of school (Rivet & Krajcik, 2008). While some school reform research (Marks, 2000; Newmann, King, & Carmichael, 2007; Newmann, Marks, & Gamoran, 1996) and some science education research (Marx et al., 2004; Rivet & Krajcik, 2004; Rodriguez & Berryman, 2002; Seiler et al., 2001) supports the claim that pedagogy explicitly connecting learning to the world outside the classroom benefits student groups underrepresented in STEM, prominent researchers in this area describe that support as preliminary (Rivet & Krajcik, 2008).

2.4.2 A Model of Engagement in Academic Work in Secondary School

Newmann developed a model of academic engagement in secondary school depicted in Figure 2.2 (Newmann et al., 1992). This model adopts the self-system theorist’s belief in an innate human need for competence (Newmann et al., 1992; Ellen A. Skinner, Connell, Zimmer-Gembeck, Eccles, & Wellborn, 1998) and an inherent need for supportive, significant, long-lasting interpersonal relationships (Furrer & Skinner, 2003). Humans need to be effective in producing change in our environment (Ellen A. Skinner et al., 1998). This human need for competence will drive school engagement when a school environment supports a sense of membership in a school community and when its focus is on authentic work that students feel has value (Figure 2.2).

The Newmann school engagement model in Figure 2.2 identifies five factors under the control of the school contributing to the likelihood that students will commit to the school community (Newmann et al., 1992). School membership is promoted by
clarity of purpose, fairness, and personal support, providing frequent opportunities for all students to experience success, and a climate of caring (Figure 2.2).

In addition, the Newmann model identifies authentic school tasks (Figure 2.2) that students deem meaningful and worthy of effort as a contributor to school engagement. The Newmann model predicts increased engagement when students perceive that school tasks have extrinsic rewards, align with intrinsic interests, offer them a sense of ownership, are connected to their world, and include some fun (Figure 2.2). School membership and authentic work meet the need for competence by providing opportunities and support for success at work students believe has value.

2.5 Engagement and Self-Regulated Learning

Engagement, in a general sense, is “active, goal-directed, flexible, constructive, persistent, focused interactions with the social and physical environments” (Furrer & Skinner, 2003, p.149). Motivational processes in the self-system operate primarily through engagement to contribute to learning and development (Furrer & Skinner, 2003). A student engaged in learning will direct their attention, expectation, and persistent participation toward understanding or mastering the knowledge, skills, or crafts promoted by academic work (Newmann, et al., 1992).

Engagement is motivation in action marked by enthusiastic participation. Engagement in learning science is psychological investment and effort toward the goal of improved scientific understanding in specific situations. The cognitive, behavioral, and emotional aspects of engagement are all important contributors to engagement that maximizes learning and development.
2.5.1 Behavioral Engagement

Behavioral engagement is sometimes called on-task behavior and taken to indicate student acceptance of the goals and means of goal attainment specified by the teacher and school (Newmann et al., 1992). Behavioral engagement is defined as attentiveness and involvement (Klem & Connell, 2004; O. Lee & Anderson, 1993). Behavioral engagement is generally assumed to be necessary for learning and development.

2.5.2 Cognitive Engagement

Cognitive engagement in science learning has been defined as investment of effort in integrating personal knowledge with scientific knowledge and in application of scientific knowledge to explaining and predicting the natural world (O. Lee & Anderson, 1993). Observed behavioral engagement is considered an insufficient indicator of cognitive engagement. Behavioral engagement as acceptance the teacher’s goals or in pursuit of only extrinsic rewards may not indicate cognitive engagement that leads to integration of personal and scientific knowledge.

2.5.3 Self-Regulated Learning

Self-regulation describes how people choose to apply their personal resources as they interact with the world. If an innate need for competence is assumed, it follows that goal directed behavior is normative. In this view all people are self-regulating, but those with low levels of competence and control beliefs exhibit less self-regulation. Self-regulation is an expression of goal directed behavior (Paris & Winograd, 2003; Pintrich,
Students that are disaffected may exhibit self-regulation directed toward disruption or expression of their disaffection. Paris and Winograd (2003) describe self-regulated learning as a coherent construct integrating how knowledge of cognitive strategies, metacognition, and motivation interact to influence the self as active agent in learning, that is, internalized regulation of learning. So conceived, self-regulated learning is an indicator of cognitive engagement.

2.6 Implementation of Authentic Inquiry in K-12 Science Education

Research on the benefits of authentic science inquiry pedagogy has been presented. Chinn and Malhotra (2002) document that science curriculum in the U.S. promotes “inquiry” that is epistemologically different and to a considerable degree antithetical to the reasoning processes that characterize science as practiced by professionals and propose this may play a role in low science literacy. They identify 14 cognitive processes of authentic science inquiry and contrast these with the cognitive processes required by school inquiry tasks. This study uses their analysis as simplified by Tomasek (2006) in developing part of its definition of authentic science pedagogy.

Chinn and Malhotra’s analysis of 468 hands-on activities in nine middle school science textbooks revealed that on average only 0.5 of 11 elements of authentic science reasoning were required per task (Chinn & Malhotra, 2002). Nine of 11 elements of authentic science reasoning appeared in less than 5% of tasks, if they appeared at all. The exceptions were making multiple observations and the use of analog models, which occurred in 17% and 15% of the textbook tasks, respectively.
They also examined 26 inquiry tasks developed by educational and psychological researchers. These tasks frequently included most of their eleven elements of authentic reasoning. The frequencies present ranged from 12%, for generating own research questions, complex transformation of observations, and studying expert research reports to 92% for developing simple controls.

While research findings on the benefits of adding authenticity to science pedagogy has not been widely implemented, the most frequent context for authenticity has been environmental science topics and investigations. Environmental education provides not only relevant issues and topics for study; an authentic environmental context is often only steps away from a sterile classroom environment. This study considers the effects of authentic teaching of environmental topics on student engagement and self-regulated learning in the classroom studied.

2.7 An Authentic Pedagogy model from the School Reform Movement

In response to the school reform movements of the 1980’s, the nationally funded Center on Organization and Restructuring of Schools (CORS) at the University of Wisconsin-Madison developed a broad conception of authentic pedagogy and authentic student achievement designed to be applicable to all grade levels and disciplines (Newmann, Secada, & Wehlage, 1995; Wehlage et al., 1996). Their framework for authenticity was based on the nature of authentic human achievement, that is, what competent adults do. Therefore, they stress the intellectual skills exercised by competent adults as a goal of K-12 education (Newmann et al., 2007), including the view that
cognitive work in the adult world is dominated by constructing or producing knowledge rather than reproducing it (Resnick, 1987).

CORS defined authenticity in an instructional context as “the extent to which a lesson, assessment task, or sample of student performance represents construction of knowledge through the use of disciplined inquiry that has some value or meaning beyond success in school” (Newmann, et al., 1995, p. 4). They proposed construction rather than reproduction of knowledge, the use inquiry practices characteristic of the discipline, and value or meaning beyond success in school were each necessary for authenticity as illustrated in Figure 2.1.

Research has shown increased achievement on conventional and authentic measures in classrooms with high levels of authentic instruction, as measured by the use of construction of knowledge, use of disciplined inquiry, and connections made to the world beyond the classroom (Newmann et al., 2007). Evidence for increased achievement was demonstrated in middle and high school science and math classrooms (Newmann & Wehlage, 1995). Authentic instruction correlated positively with greater achievement gains on high school science assessments for students of low socioeconomic status than those of high socioeconomic status (V. E. Lee, Smith, & Croninger, 1997; Newmann & Wehlage, 1995).

The CORS authentic pedagogy and student performance model (Figure 2.1) connects authenticity to the emphasis on constructivism, inquiry, relevance, and contextualization in science education research discussed in this review. As in the CORS model, A Framework for K-12 Science Education (National Research Council, 2012) connects a constructivist view of learning and inquiry to authenticity. Both emphasize
the role of inquiry characteristic of the discipline in effective pedagogy and both view the intellectual processes of skilled disciplinary practitioners as models for organizing instruction. Both the CORS formulation and A Framework for K-12 Science Education advocate viewing science inquiry as the process by which knowledge is constructed.

The CORS model developed from effective schools research and the science education literature reviewed here propose a common view of authentic science pedagogy. Both conceptions propose contextualizing learning in the real world or everyday experiences and in the reasoning practices of science practitioners.

2.8 Research Gaps

This review has discussed how the concept of authenticity is connected to a number of important themes in science education research and to the problem of engaging students in the cognitively complex processes required to construct scientific literacy. There are few longitudinal studies in the science education literature that clarify the relationship between authentic science instruction and the engagement of high school students in the intellectual processes that build science reasoning skills. This study will further explore this relationship, including consideration of the role of connections between science classwork and the world outside the classroom, through a longitudinal study of a ninth grade science classroom with a curriculum designed to be more authentic.

Research on authenticity in science education has generally focused on two areas. First, authentic science inquiry has been defined as resemblance to the reasoning practices scientists use to construct understanding of the natural world (Chinn &
Malhotra, 2002). This use of authenticity focuses on its cognitive benefits. Chinn and Malhotra (2002) describe the “pressing need for research that develops and tests instructional approaches for fostering the development of such complex strategies” (p. 213).

Secondly, authenticity has been connected to the meaningfulness and relevance of the activities or subject matter. This meaningfulness and relevance generally comes through connections to the world outside the classroom and is considered a motivational resource encouraging cognitive engagement. The proposed cognitive benefits of these authentic contexts have also been discussed (H.-S. Lee & Songer, 2003; Lieberman & Hoody, 1998; Newmann et al., 1992).

While some common themes in the literature have been highlighted here, there is not a consensus on what constitutes authentic science activities (H.-S. Lee & Songer, 2003), especially in classroom settings. Edelson (1998) identifies authenticity as a “crucial objective for learning” (p. 318), but very hard to implement. Chinn and Malhotra’s (2002) comparison of authentic science reasoning practices with those in middle school curriculum materials highlights the challenge of fostering authentic science reasoning. Chinn and Malhotra (2002) conclude that: “At present little is known about how to foster complex reasoning” (p. 213) characteristic of the practice of science.

Scientists have experiences and resources unavailable to most adolescent students (H.-S. Lee & Songer, 2003). Studies have just begun to identify how to transform classroom tasks to include authentic science reasoning practices while making them accessible to students (H.-S. Lee & Songer, 2003; Rivet & Krajcik, 2008). Lee and Songer (2003) state that this transformation is “far from understood” (p. 928). This study
defines characteristics of authentic student performance in science classwork, as well as, behavioral engagement and self-regulated learning; tracks them over time in a classroom; and relates them to the components of authentic science pedagogy present.

The CORS model of authentic pedagogy and student performance described here provides a coherent model of authenticity that has been shown to align well with recent science education goals (National Research Council, 2012). The CORS model has previously been adapted specifically for science pedagogy (Research Institute on Secondary Education Reform for Youth with Disabilities, 2001), but published research using that adaptation is limited (Newmann et al., 2007). This study defines a broad set of characteristics of authentic science pedagogy by adapting the CORS model of authentic student performance using the synthesis of the science education research discussed here. The study then documents the implementation of these components in a high school science classroom over a school year.

This longitudinal study is designed to reveal patterns in student responses to affordances for authentic science reasoning. Longitudinal research studies that define authentic science pedagogy and achievement at the classroom level using the CORS model were not found.

The school studied has a school-wide philosophy supportive of authentic instruction. However, while some science specific curriculum support was provided, the teacher in this study developed his own curriculum. The burden of a full teaching assignment, curriculum development, and implementing school-wide reform was primarily on teachers. In the current culture of support for public schools, this represents
an above average case. Therefore, this study documents what a widespread effort to implement authentic science curriculum may look like.

Rivet and Krajcik (2008) note a paucity of studies that explore the claim that contextualizing science classroom instruction in the world outside of school improves science achievement, and that there little evidence supporting the claim: “although much literature has touted the benefits of such efforts to contextualize science instruction to improve learning, few studies have explored this relationship and little research exists to substantiate such claims” (p. 79). In a preliminary study of a small number of students, they report that student use of the contextualizing features of a curriculum rich in them was positively correlated with science learning assessments (Rivet & Krajcik, 2008). By operationally defining and documenting the connections to the world beyond the classroom in assessment tasks and instruction, this study looks for a relationship between these contextualizing aspects of authentic science pedagogy and student performance.

The profound cognitive work needed to build new scientific understanding of the natural world requires students to sustain purposive effort over time. Science education research has considered the motivational needs of students to engage in the cognitive challenge of science learning less frequently than construction of knowledge through inquiry in a social learning environment. While this motivational challenge is acknowledged, less research specifically addresses how pedagogy relates to cognitively engaging diverse students i.e., different genders and ethnicities, over time.

It is claimed that increasing authenticity by anchoring instruction in the investigation of real world problems increases students’ enthusiasm (H.-S. Lee & Songer, 2003). The model, usually implicit, in the science education literature is:
Connections in instruction to the world beyond the classroom → Student perceptions of relevance & meaning (value) → Increased motivation (interest, enthusiasm) → Increased cognitive engagement & self-regulation → Increased achievement.

Science education researchers have seldom explicitly tested this model. Lee and Anderson (1993) carefully documented 6th grade students’ cognitive engagement in a progressive science curriculum that accessed students’ prior knowledge and prioritized social construction of knowledge. They explicitly based the study on both conceptual change research and motivation research and pointed out how infrequently science education research considered both. They highlighted our lack of knowledge on how to reach students with personal agendas that leave little effort resources to engage in learning science.

This study defines authentic science pedagogy and student performance by adapting the CORS formulation in Figure 2.1. To add to research at the classroom level, this study documents the implementation of these components and student responses in high school science classrooms over a school year. This study’s longitudinal classroom level research relating authentic student performance to authentic science pedagogy hasn’t been reported before. Concurrent measurement of behavioral engagement and self-regulated learning addresses another research gap. Finally, inclusion of measures of the presence of the value beyond demonstrating success in school component of authentic pedagogy is a unique contribution.
Figures

Figure 2.1 Pictorial representation of the Center for Organization and Restructuring of Schools (CORS) model of authentic pedagogy and student performance (Wehlage et al., 1996).
Figure 2.2 The Newmann model of student engagement in academic work in secondary schools (Newmann, et al., 1992). Figure reproduced with permission of the copyright holder.
Chapter 3  Theoretical Perspective

The challenge of developing in youth the intellectual skills used by science practitioners and science literate citizens has been presented. This study takes the perspective that the natural world is complex and meaning-making practices of science are intellectually challenging. The scientific inquiry practices and scientific explanations of natural phenomenon differ markedly from the everyday problem solving skills and naïve conceptions of most students. The intellectual work of using meaning-making processes of science to construct science knowledge is difficult, taxing, and unfamiliar. It requires cognitive engagement in authentic tasks sustained through multiple episodes of what Piaget called “disequilibration” and construction of more sophisticated understandings.

High school students experience a complex educational system with many social, intellectual, and psychological components. This chapter identifies the elements of the educational system and their key interactions considered in this study. The theoretical basis for the variables measured in this study are discussed. First, the model of authentic science pedagogy used to define measures of both instruction and student performance in science classwork are presented.

Then a motivation model used to characterize the value students may perceive in their classwork is discussed. Finally, the basis for indicators of student engagement and self-regulation is detailed. First, the concept of affordance from perceptual psychology needs to be defined.
3.1 Affordance

Gibson (1977) developed the idea of affordances as environmental features that are perceived by organisms in a way that presents certain categories of activities. A tree perceived as “climbable” by a squirrel, for example, affords climbing. Different types of instructional materials or activities have been viewed as differentially engaging different cognitive processes (Cognition and Technology Group at Vanderbuilt, 1992). For example, presenting a discrepant event may afford careful observation, questioning, brainstorming explanations, and evaluating alternative explanations. By contrast, a reading about the natural phenomenon in the discrepant event is likely to afford recognizing and recording and little more.

Affordances, as used in this study, are cognitive, sociocultural, and psychological environmental cues intentionally implemented in the classroom and curriculum. The affordances of interest in this study are those that education research indicates should support student engagement in authentic school science. This study investigates whether the identified components of authentic science pedagogy are related to student engagement, self-regulation, and perceptions of value beyond demonstrating success in school.

3.2 Authentic Science Pedagogy and Student Performance

A central tenant of education is that it functions to prepare learners to function as competent adults. Therefore, this study defines authenticity in reference to the scientific explanations and meaning-making activities used by science practitioners in the community and by characteristics that afford students to find meaning and value in their
classwork. The broad perspective of this study includes such science practitioners as, professional scientists, engineers, and medical professionals, as well as those who practice citizen science (Tomasek, 2006), use science to influence public policy, and do science journalism (McGinn & Roth, 1999). The authenticity of science learning environments created by the school will be judged by a model that considers whether students are involved in using the intellectual and social tools of science practitioners to investigate, talk, read, and write about questions that students find valuable, significant, and worthy of effort (Rahm et al., 2003; Wehlage et al., 1996).

This study integrates the model of authentic pedagogy and student performance developed by the Center for Organization and Restructuring of Schools (CORS) (Newmann & Associates, 1996) and the multiple research perspectives connected to the concept of authenticity discussed in the literature review. The reasoning practices in the CORS model align with the focus of the National Research Council’s A Framework for K-12 Science Education’s (2012) emphasis on exploring foundational ideas about how the natural world and designed systems behave through application of those reasoning strategies used by competent practitioners of science and engineering.

Based on the CORS authenticity framework, this study identifies three key pedagogical affordances for authentic science that can be tracked in a high school science classroom over a school year. Figure 3.1 identifies these as Construction of Knowledge, Science Meaning-Making Processes and Value Beyond School and describes how they are defined. The first two of these are intellectual components of the reasoning practices of science and engineering practitioners. The Value Beyond School component is important because when students perceive value beyond demonstrating success in school
in their coursework, this component provides motivational support for the difficult
cognitive work required by the two intellectual components of authenticity.

The positive influences of authenticity are maximized only when all three
components of authenticity are present. Science education investigators claim that
gaining understanding of science concepts is maximized through using science inquiry
practices, and that the reverse is also required to develop proficiency in inquiry (Krajcik,
Codere, Dadsah, Bayer, & Mun, 2014). This is reflected in this model’s assertion of the
need for both the concepts and disciplinary inquiry practices emphasized in this model’s
first two components.

This chapter summarizes the model of authentic science pedagogy and student
performance used to define variables in this study.

3.3 Intellectual Aspects of Authentic Science Pedagogy

Authentic science pedagogy includes those teaching practices that afford students
opportunities to develop science literacy by practicing meaning-making methods of
science and using them to construct scientific explanations of the natural world. These
are the intellectual components of authenticity considered in this study.

3.3.1 Affordances for Reasoning Practices that Promote Construction of
Knowledge

The first component of authentic science pedagogy is student use of reasoning
practices that promote construction rather than simple reproduction of scientific
knowledge (Figure 3.1). Construction of Knowledge will be used as a brief way to refer
to this component of authentic pedagogy.
Classroom science instruction and assignments afford Construction of Knowledge when they ask students to use higher-order thinking to synthesize, evaluate or organize complex information and to consider alternative solutions or explanations to address a fundamental science concept, a problem, or an issue in ways that produce new meanings for them (Newmann et al., 1995). This study adopts the view of science education investigators that Construction of Knowledge in science is enhanced by active engagement with the natural phenomena under investigation (Krajcik et al., 2003).

Instructional activities that begin by eliciting students’ current understandings (Donovan & Bransford, 2005), and then provides opportunities to try and revise strategies and explanations (McGinn & Roth, 1999) are also defined as contributing to authenticity. Because research demonstrates that metacognitive reflection on one’s current understanding enhances Construction of Knowledge (Donovan & Bransford, 2005), activities that explicitly encourage students to reflect on their learning states affords Construction of Knowledge (White & Frederiksen, 2000).

3.3.2 **Affordances to Use the Meaning-making Reasoning Practices of Science Practitioners**

As in other fields, science has its own practices and standards for producing new meanings. The second component of authentic science pedagogy is the use of reasoning processes similar to those employed by science practitioners to establish new knowledge. For brevity this component will be referred to as *Science Meaning-Making Processes* (Figure 3.1).

Classroom activities that are authentic for this component ask students to use unifying ideas and theories central to science, as well as, the practices of inquiry,
research, model building, representation, and communication characteristic of science practitioners (Newmann et al., 1995). High authenticity in this component also requires active engagement with natural phenomena or human engineered designs (Krajcik et al., 2003).

3.3.3 **Affordances for Authentic Science Discourse – Substantive Conversation**

This study also uses the CORS conception of Substantive Conversation to measure the use of sustained classroom discourse to consider and evaluate scientific explanations or investigations. These conversations are more authentic when characterized by higher-order thinking and the meaning-making practices of science, such as, arguing from evidence. Linn (2000) argues that discourse that makes students’ thinking visible promotes construction of scientific understanding. This is used as a key component of authenticity in this study.

In such conversations, students are able to articulate their current understanding, promoting updating of their conceptions, and practice Science Meaning-Making Processes (Becker, 2008; National Research Council, 2012). The Substantive Conversation construct is, therefore, considered an aspect of the reasoning processes of science practitioners in this study’s theoretical perspective (National Research Council, 2012). The Substantive Conversation component of authenticity does not appear as a separate component in Figure 3.1 because it is measured in classroom instruction, but not in analysis of the teacher’s assessments. Substantive Conversation during classroom instruction contributes to both Construction of Knowledge and Science Meaning-Making Processes.
3.4 Affective Aspects of Authentic Science Pedagogy

Students perceive much schoolwork as trivial or contrived with no application beyond showing that they are “competent” in school (Newmann et al., 1992). Authentic intellectual work addresses solutions to meaningful problems or issues or expresses aesthetic values. The fourth characteristic of authentic science pedagogy is affordances in assessments and instruction for students to perceive value in their classwork beyond demonstrating success in school (Wehlage et al., 1996). For brevity this component will be referred to as the Value Beyond School (Figure 3.1).

Authentic science pedagogy asks students to use scientific meaning-making to communicate their knowledge, present a product or performance, or take some action with an impact on the community outside the classroom that goes beyond demonstrating their success in school (Newmann et al., 1995). Evidence of this aspect of authenticity may be found in classroom instruction and the teacher’s assessments.

Instruction and assignments that focus on questions, issues, or problems that students “have actually encountered, or are likely to encounter, in their lives beyond school” (Newmann et al., 1995, p. 84) afford them opportunities to find value in their class work beyond demonstrating success in school. Communicating ideas, producing a product, investigating a problem, or engaging in a performance for an audience beyond the teacher and the classroom may also afford this component of authenticity. For example, students might be asked to measure air pollution levels in or around their school and share their findings with a public agency.

Classroom instruction can also afford perceptions of value beyond school when it gives students opportunities to use the representational, technological, or measurement
tools of science or engineering practitioners and to work alongside such practitioners (McGinn & Roth, 1999).

A unique aspect of this study is exploration of the frequently hypothesized, but seldom explicitly investigated (Rivet & Krajcik, 2008), positive relationship between student performance and engagement, and science instruction and assessments that are meaningful and relevant to students’ lives and communities outside of school. The CORS model of authentic pedagogy proposes that classwork and instructional practices that help students find value in science classwork beyond getting a grade or a score on a standardized test increase the likelihood of cognitive engagement in using science meaning-making reasoning processes to construct scientific knowledge (Figure 2.2).

3.5 Authenticity of Student Performance

This study also adapts the CORS model for defining the authenticity of student performance on science work samples (Newmann et al., 1995). There three aspects of authenticity from the two intellectual components of the authenticity model are measured. *Construction of Knowledge* is measured by use of scientific analysis. Student use of *Science Meaning-Making Processes* is measured by use of scientific concepts and by elaborated communication about science. These are considered authentic because they reflect important aspects of science as practiced by competent adults.

Student performance of scientific analysis is indicated by evidence of organizing, synthesizing, and interpreting information with scientific content. Explicit and correct use of important scientific ideas, concepts, or theories in classwork is also an important performance indicator.
The third indicator of authentic performance, elaborated communication, is demonstrated when students successfully communicate an explanation or argument regarding a natural phenomenon and scientific support for their argument. The form and content of the communication is authentic when it resembles forms of communication used by science practitioners and is complex enough to reflect some of the nuances that adults have to deal with.

3.6 School engagement

Educators, science education investigators, and policy makers express the belief that engagement in authentic activities is critical to student success in meeting science learning goals. The authors of A Framework for K-12 Science Education (National Research Council, 2012) articulate how important student engagement in affordances for the components of this study’s authenticity model:

Students...actively engage in scientific and engineering practices and apply crosscutting concepts to deepen their understanding of the core ideas in these fields. The learning experiences provided for students should engage them with fundamental questions about the world and with how scientists have investigated and found answers to those questions (p. 8).

School engagement is a multifaceted construct involving both cognitive, behavioral, and emotional components. The Newmann model of school engagement proposes that an innate human need for competence drives engagement in academic work (Figure 2.2).
3.6.1 Cognitive Engagement

Cognitive engagement is one aspect of school engagement. Cognitive engagement in science learning has been defined as investment of effort in integrating personal knowledge with scientific knowledge and in application of scientific knowledge to explaining and predicting the natural world (O. Lee & Anderson, 1993). This definition is expressed in terms of the intentions of students engaging in science and engineering reasoning and deepening their understanding of fundamental ideas. Measuring cognitive engagement is challenging. Evidence of behavioral engagement and self-regulated learning may also indicate cognitive engagement.

3.6.2 Behavioral Engagement

Behavioral engagement is defined as attentiveness and involvement (O. Lee & Anderson, 1993) in this study. Behavioral engagement is sometimes described as on-task behavior. It signifies an acceptance of teacher-specified tasks.

3.6.3 Self-Regulated Learning

Self-regulated learning is displayed when students apply their understanding of cognitive learning strategies and their awareness of their own learning preferences and abilities to adopt particular strategies to accomplish a specific learning task. Self-regulated learning may be evidenced by deployment of “funds of knowledge” and metacognitive and motivational resources (Paris & Winograd, 2003) to valued educational goals. Investigators use “funds of knowledge” to refer to the socio-cultural
and bio-physical knowledge, understanding, and social capital all students bring with them into school settings.

Self-regulation is willful strategic action toward a self-selected goal. Students exhibit self-regulated learning when they exercise agency, autonomy, or ownership related to a class, a lesson, or learning in general. Self-regulated learning is more representative of cognitive engagement than behavioral engagement.

3.7 Conceptual Framework

Current science education goals call for students to develop the ability to apply scientific understandings and meaning-making processes as users of scientific information in daily life, participants in the discussion of important public policies, and lifelong science learners (National Research Council, 2012). They call for actively engaging students in using fundamental concepts and reasoning practices in the science classroom. The purpose of this study is to use the model of authentic science pedagogy in Figure 3.1 to document those efforts in a science classroom and investigate relationships between authentic science pedagogy, student performance, engagement, and whether students find Value Beyond School in their classwork.

Figure 3.2 presents the conceptual framework for this study. It describes the constructs considered in this study’s research questions and proposed relationships between them. These relationships are based on the theoretical perspective described in this chapter. The directions of these relationships are expressed in the propositions investigated by this study.
The model of authentic science pedagogy and performance adopted by this study is used to measure the authenticity of both pedagogy and student performance on science classroom tasks. As shown in Figure 3.2, Construction of Knowledge and Science Meaning-Making Processes are measured in both the teacher’s pedagogy and student performance on science classwork. These components of authentic pedagogy are proposed to be positively related to the authenticity of student performance on these same components. Figure 3.2 shows that affordances for finding Value Beyond School in authentic pedagogy may also be related to student performance.

Authentic pedagogy is also proposed to be associated with greater cognitive engagement, as indicated by evidence of behavioral engagement and self-regulated learning, and with student perceptions of the value they find in their classwork beyond demonstrating success in school.

Relationships between gender, ethnicity, and prior achievement and student science achievement and engagement have been previously documented in the science education and broader education literature. Where possible, this study examines the influence of these student characteristics on authentic student performance and engagement.

Gender differences in science achievement have narrowed over the last 45 years (Bacharach, Baumeister, & Furr, 2003; Maerten-Rivera, Myers, Lee, & Penfield, 2010; O’Reilly & McNamara, 2007; S. R. Singer et al., 2005), but haven’t completely closed in the latest National Assessment of Educational Progress results (National Center for Education Statistics, 2011, 2012). Persistent gaps have been documented in school science achievement, especially for African American and Hispanic students compared
with majority students (Bacharach et al., 2003; Dochy, Segers, & Buehl, 1999; O. Lee & Luykx, 2007; Maerten-Rivera et al., 2010; National Center for Education Statistics, 2011, 2012).

Gender and ethnicity have been reported to influence engagement in school (Marks, 2000). Previous research has found that mathematics and science achievement are related (Maerten-Rivera et al., 2010; Wang, 2005), and that prior science knowledge is related to achievement in science (O’Reilly & McNamara, 2007). Marks (2000) found that prior grades are related to student engagement.

Next, Chapter 4 describes the high school classrooms studied. They are the context within which these variables are measured and the relationships in the conceptual framework explored.
Figure 3.1 Model of authentic science pedagogy and student performance developed for use in this study. All three components are necessary for the classroom environment to be highly authentic. Construction of Knowledge and Science Meaning-Making Processes are characteristics of the reasoning practices of science and engineering and ideas considered. The Value Beyond School component expresses the affordance provided when students recognize value in their classwork beyond demonstrating success in school.
Figure 3.2 Conceptual framework for this study. The variables and relationships explored between them are illustrated. Hypothesized positive relationships are shown. The other arrows indicate directional and non-directional relationships documented in previous research that are examined to the extent possible in this study. Authenticity of pedagogy includes authenticities of these components in classroom instruction and in the teacher’s assessments.
Chapter 4  Case Study Design

The purpose of this study is to document the efforts of one high school science teacher to intentionally implement a curriculum with many of the characteristics of the authentic science pedagogy model described in Chapter 3. The study also examines the role that experiencing this pedagogy plays on the quality of students’ performance in science classwork, engagement and self-regulated learning. To accomplish these goals, the research was designed to answer two broad research questions.

1) What does a high school science curriculum that intentionally attempts to provide students with an authentic science experience look like?

2) Does engaging students in a more authentic experience produce gains in science achievement, engagement, and self-regulated learning?

The case investigated in this study is the 9th grade science classrooms of one teacher. To answer these questions, instruction, student engagement and student performance on the classwork are measured over the course of a school year and relationships between instruction and student performance are explored. The conceptual framework developed in Chapter 3 identifies the variables measured for the descriptive study defined by the first broad research question. Table 4.1 correlates those variables with Research Questions 1, 2, and 3.
4.1 Use of an Systemic Approach

The study was conducted over most of a school year so that both longitudinal trends and variations in the instruction students experienced could be documented and student responses could be compared with patterns over time and between different units of instruction. The study describes the teacher’s curriculum and instructional decisions across the school year as they relate to authenticity.

Teachers communicate academic expectations to students by the nature of the assessments used to assign course grades and through their day-to-day classroom instruction. These are two important sources of affordances for learning science students experience in the instructional ecosystem of a typical high school classroom. This study uses the theoretical model of authentic pedagogy from previous research to define the authentic pedagogy in the teacher’s assessments and classroom instruction. This gives a rich description of the authenticity students’ experience.

This same model of authentic pedagogy is used to operationally define the variables in Table 4.1 for student performance. The value of these inferences is affirmed by the model’s alignment with current national science and engineering education recommendations documented in the theoretical perspective and literature review (Chinn & Malhotra, 2002; Katehi, Pearson, & Feder, 2009; H.-S. Lee & Songer, 2003; McGinn & Roth, 1999; National Research Council, 2012; J. Singer et al., 2000; S. R. Singer et al., 2005).

Research supports the critical role cognitive engagement plays in helping students to bridge the distance between naïve student thinking and the meaning-making processes

This research is a study of a reform of great interest in progress, not an intervention implemented for this study. It addresses a complex research problem requiring a naturalistic setting. In situ observations of the authenticity of the teacher’s instruction and multiple aspects of student responses using a systemic approach brings depth to the description of the case. Relationships discovered between instruction and student responses will be from a range of instructional experiences over most of a school year. The use of a common model of authentic pedagogy gives greater validity to inferences regarding the relationship between the authenticity of the teacher’s instruction and student performance on class work.

4.2 Participants

4.2.1 School

The case investigated in this study is in a suburban public school district in the western U.S. facing growing poverty and diversity in its student population. The classrooms studied were in a grade 6 to 12 STEM focused options school that was identified by district information as preparing the diverse students for post-secondary education in health careers and engineering. The school intentionally recruited a larger
A large proportion of Hispanic and Latino students than the district average. A large proportion of the school staff were bilingual in English and Spanish. The school had no academic screen for admitting students. District information said it was open to all students interested in science and medicine.

This study was conducted during the third year of operation (2009-2010) of the school. The school planning staff of experienced educators believed that a “rigorous curriculum” that made learning relevant to the “students’ culture, community, and personal lives” (school website) would drive engagement and academic success for the diverse population. Therefore, the school planning staff intentionally sought inclusion in the Expeditionary Learning Schools network. The school was opened as part of the network and teachers received training from Expeditionary Learning Schools (ELS). Staff had access to on-going professional development and curriculum resources of the network and on-site visits from ELS resource people.

4.2.1.1 School-wide curriculum

The school year was divided into three 12-week periods referred to as trimesters in this study. Students received credit toward high school graduation requirements based on grades achieved in academic courses during each trimester. The school’s curriculum emphasized readiness for success in STEM college majors and careers for all of its diverse students. All students took biology, physics, and chemistry in grades 9 and 10. In grade 11 students choose medicine and health or engineering and design as a career focus. College credit courses in math, engineering, writing, and Spanish were available.
on campus as a part of the normal course of study. This was part of an intentional effort to create the expectation that all students would plan and prepare for post-secondary education.

The school’s curriculum and instruction model was informed by the ELS model (Expeditionary Learning Outward Bound, 2003). The ELS model aligns closely with this study’s model of authentic pedagogy and student performance. Table 4.2 gives examples of this alignment between this study’s model of authentic pedagogy and student performance and the ELS model. The ELS model stresses building curriculum around “compelling topics” (Expeditionary Learning Outward Bound, 2003, p. 11) that make “connection to…public problems or personal experiences” (Newmann et al., 2007, p. 44). In addition, ELS calls for learning “products [that]… meet an authentic need and have an audience… beyond… the classroom teacher” (Expeditionary Learning Outward Bound, 2003, p. 13).

This aligns with this study’s interest in the relationship between students’ self-regulated learning, and authentic student performance in science and engineering. Self-regulated learning is explicitly addressed by the ELS model when it calls for instruction to ensure that “students do the thinking and the work” (Expeditionary Learning Outward Bound, 2003, p. 19). ELS and this study’s theoretical model stress in-depth investigation of powerful disciplinary ideas through higher-order thinking and use of inquiry strategies (Table 4.2). As the teacher in this study developed and implemented curriculum, he was informed by the ELS model because his personal philosophy aligned closely with ELS
recommendations. Final curriculum decisions were those of the teacher, based on the teacher’s expertise. This was true for teachers throughout the school.

4.2.1.2 School-wide assessment system

The school used a proficiency-based assessment system for assigning student grades. The system was designed collectively by the school staff. According to school documents, it reflected beliefs that the school’s equity and student achievement goals were advanced by (1) high academic expectations for all students; (2) clear statements of specific learning proficiencies, called learning targets, in language children understand; (3) limiting learning targets to five to eight per trimester focused on higher-level thinking about key disciplinary ideas; (4) explicitly connecting daily instruction and classroom tasks to learning targets; (5) recognizing that students will learn at different rates and be able to express proficiency in different ways by providing a flexible time frame and multiple opportunities and avenues to demonstrate proficiency; and (6) excluding all consideration of the appropriateness of student behavior in the content of learning targets and assessment of proficiency. This system was designed to maximize readiness for post-secondary education in a student population with a large proportion of first-generation college students.

According to school documents, teacher assessment of student proficiency on individual learning targets was expressed to students as beginning or basic understanding (1), not fully proficient (2), and meets or exceeds the target (3). Performance at level three was to include application of knowledge, concepts, and skills using higher-order
thinking. Grades are reported to parents or guardians on official transcripts in the A through F fashion traditional in the U.S. However, they are determined solely by the proportion of learning targets scored in each proficiency category.

Learning Targets were defined by the teacher for each course and presented to students at the beginning of each of the three grading periods (trimesters) that make up the school year. Field notes from classroom observations documented that the current Learning Target was displayed to students at the beginning of nearly every class day. Students were always offered alternative tasks during each grading period to demonstrate proficiency on the learning targets. If proficiency was not demonstrated on the first assessment task, additional tasks could be attempted at a later date without a grade penalty.

4.2.2 Teacher

The one teacher that taught all of the ninth grade science courses participated in the study. The teacher is a Euro-American male who is bilingual in English and Spanish. His academic preparation is in life science. He had taught high school science for 12 years. He was experienced teaching introductory biology and physics. He was part of the planning team for this options school and was regarded by his peers as a highly skilled teacher.

The school year observed was the first time this teacher taught these two courses in this school. Implementing the school-wide proficiency based grading system and building each trimester’s curriculum around a compelling topic led to the teacher reorganizing the curriculum he had used in the past. No textbook guided the curriculum.
The first two trimesters (fall and winter) all 9th grade students were enrolled in introductory high school biology. The third trimester (spring) all 9th grade students were enrolled in introductory high school physics.

The teacher taught 4 different groups (class periods) of students that met at different times of the school day. Class periods 1, 2, and 3 met in the morning before lunch. Class period 5 met at the end of the school day, at approximately 1 pm to 2 pm. All class periods met on Mondays and Fridays for 50 minutes. During mid-week each class period met twice for 75 minutes. Observations in this study were focused in class periods 2, 3, and 5.

The school curriculum and grading policies were implemented consistently in the classroom studied. The proficiency-based assessment system shaped teacher decisions on contextualizing curriculum, when it was necessary to reteach content, and provision of multiple opportunities to demonstrate proficiency. The teacher’s flexibility in responding to student needs meant post hoc decisions on appropriate units of analysis were necessary.

4.2.2.1 Classroom curriculum context

Each trimester’s curriculum used a topic meant to be compelling to students to focus the curriculum. The fall trimester curriculum focused on human and biological factors that contribute to making a decision whether to get an influenza vaccine. Biology topics, including cell structure and replication, viral infection, viral evolution, electrophoresis technology, and the immune system were taught through this lens.
The winter trimester biology curriculum focused on the role of biology in defining human identity. The topics included how genetic variation and traits and neurology contribute to human identity. The role of measurements of human characteristics in understanding human identity was interwoven with activities on presentation and analysis of such data to identify meaningful patterns. The trimester wrapped up with a project on unique aspects of the neurology of the teenage brain and what adults should understand about it.

The spring trimester physics curriculum focused on how engineering design is used to improve human life. Mechanics and energy concepts were taught and applied to student’s own engineering designs. The culmination of the trimester was a project to design and test a wind turbine as an example of alternative energy.

Within each trimester, the curriculum was divided into instructional units defined by one or more of the teacher’s learning targets (Appendix A). Therefore, each instructional unit was focused on specific science content and skills. Each instructional unit was a series of class days that included a variety of instructional activities. Each unit culminated in an assessment completed by students. The teacher designed assessment was used to decide whether students were able to demonstrate proficiency on the instructional unit’s Learning Target or Targets and used to assign grades. For brevity these are referred to in this study as assessment tasks.

Some instructional units prepared students to demonstrate proficiency on a single learning target, while others simultaneously focused on two or more learning targets. Usually, performance on one proficiency assessment task at the end of an instructional
unit allowed the teacher to assess student proficiency on the Learning Target or learning targets that were the focus of the unit. Occasionally, there were separate tasks for each learning target, or the teacher gave students a choice of assessment formats.

Proficiency assessment tasks varied in format, complexity, and class days devoted to preparation for and completion of them. One or two page quizzes with a mix of short answer and open response questions were used to assess straightforward applications of one or two concepts or ability to describe a biological system. At the other extreme, multi-week projects at the end of each trimester generated a substantial original artifact integrating understanding of several concepts or skills covered by multiple learning targets to address a question related to the compelling topic for the trimester. Chapter 5 documents and compares the authenticity levels measured in the teacher’s assessment tasks and classroom instruction from instructional units across most of the school year.

4.2.3 Students

According to school district data, the demographic makeup of the ninth grade class at the beginning of the school year was Euro-American (42%), Latino (36%), Asian (16%), other (6%), Male (37%), Female (63%), and English language learners (16%). There were 105 students enrolled in the ninth grade science courses in the fall trimester, 106 students in winter and 102 students in the spring. The differences reflect students entering or leaving the school. A total 108 different students were enrolled in at least one trimester. Of the 108 students enrolled, 43 (40%) were male and 65 (60%) female.
The classes studied averaged 25-26 students. Class sizes varied between trimesters and class periods. The smallest was 23 and the largest class was second period during winter and spring trimesters with 32 students.

Seventy-three of 108 students and their parents or guardians gave informed consent to use their self-identified ethnicity on a survey conducted at the beginning of the year. Eleven percent identified as Asian or native Hawaiian, 19% Hispanic or Latino, 11% multiple racial categories, 40% Euro-American, and 10% other racial categories. Other racial categories included Black or African American and other self-reported descriptions. In this sample of 73 students, Euro-American students seem proportionally represented compared to school district data, but Latinos and Asians appear underrepresented. External factors that may have influenced the granting of consent contribute to questions about whether the sample is representative of the whole 9th grade class. This limits some inferences of interest in this study.

Sixty-one students consented to use school records of their prior achievement. The breakdown of mean 8th grade science and mathematics class grades was 38% A, 33% B, 21% C, 3% D, and 5% F. This distribution is unsurprising, but the 40% of students not included could change this distribution considerably. External factors that may have influenced the granting of consent also raise questions about the representativeness of this sample. This also limits some inferences in this study.
4.3 Role of this investigator

This investigator is a former high school science teacher and knew the teacher in the case for about eight years as a fellow science educator before the study. The investigator attended some meetings with the school staff to understand the design of the school for three years before starting the research. The investigator was in the teacher’s classroom almost weekly in the school year before this study’s research period to understand the teacher’s instructional style and the school context; and to refine the classroom observation protocol. At the time of the study the investigator had also known the school principal for about eight years and had collaborated with him in science professional development.

This investigator’s role was as a nonparticipant observer and occasional collaborator with the teacher. The investigator was introduced to students on the first day of school and was in the classroom the first week of school. This regular presence in the classroom from the beginning was designed to make the investigator’s presence ordinary to students. The investigator worked to be an unobtrusive observer to reduce biasing student behavior, but did not avoid interaction with students to develop comfort and a level of trust. If students asked for help, the investigator responded to that request when appropriate and possible.

These steps to develop comfort with this investigator’s presence in the classroom were among those taken to support data validity. The investigator admits to interest in the success of the teacher, the students, and the school.
4.4 Limitations of the Study

This study uses multiple data sources and convenience samples to create a rich description of instruction and student responses in a public high school science classroom within a school that has adopted goals in line with recent science education policy documents (Katehi et al., 2009; National Research Council, 2012). Multiple data sources improve internal validity and provide explanatory power within the case. The school serves a student population and community similar to many other schools and classrooms. However, statistical generalization to 9th grade science students beyond the classrooms studied is not the goal of this research. As an example of a classroom implementing a model of authentic science pedagogy in a supportive school environment, the case allows examination of whether the expected student responses are in evidence.

The manuscript of this study was shared with the participating teacher and the school principal in the case to check the accuracy and appropriateness of descriptions of the case and the results.
Table 4.1 Variables measured in the descriptive study.

<table>
<thead>
<tr>
<th>Research question</th>
<th>Variable</th>
<th>Data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How does the presence of features of authentic science pedagogy vary over the school year?</td>
<td>Authentic Teacher Pedagogy</td>
<td>Teacher’s assessment tasks, Learning Targets, and scoring guides</td>
</tr>
<tr>
<td></td>
<td>Affordances to construct knowledge</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Affordances to use science meaning-making reasoning processes</td>
<td>Classroom instruction field notes</td>
</tr>
<tr>
<td></td>
<td>Affordances to find value beyond demonstrating success in school</td>
<td></td>
</tr>
<tr>
<td>2. How do the levels of authentic student performance, engagement, and self-regulated learning vary over the school year?</td>
<td>Authentic Student Performance</td>
<td>Student class work (work samples)</td>
</tr>
<tr>
<td></td>
<td>Construction of Knowledge</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scientific Meaning-Making</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cognitive engagement</td>
<td>Classroom observations of student behavior</td>
</tr>
<tr>
<td></td>
<td>Self-regulated Learning</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Behavioral Engagement</td>
<td></td>
</tr>
</tbody>
</table>
Table 4.2 Examples of alignment between this study’s model of authentic pedagogy and student performance and the standards for pedagogy and school structure and culture identified by the Expeditionary Learning Schools network as core benchmarks.

<table>
<thead>
<tr>
<th>Authentic Pedagogy and Performance Components</th>
<th>Expeditionary Learning Core Benchmarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction of Knowledge</td>
<td></td>
</tr>
<tr>
<td>Higher-order thinking</td>
<td>“Compelling topics” invite students to “consider multiple perspectives.” (Pg. 11)</td>
</tr>
<tr>
<td>Consideration of alternatives</td>
<td></td>
</tr>
<tr>
<td>Production of knowledge</td>
<td>“Teachers activate and build on students’ prior knowledge”; for example, through beginning with an experience and inviting “students to make sense of it.” (Pg. 19)</td>
</tr>
<tr>
<td>Collaborative discourse that shares understanding</td>
<td>“Teachers structure lessons so that teachers talk less and students talk more; the students do the thinking and the work.” (Pg. 19)</td>
</tr>
<tr>
<td>Science Meaning-Making Processes</td>
<td></td>
</tr>
<tr>
<td>Sustained focus on core disciplinary knowledge</td>
<td>“In-depth investigations lead students to generalizations, concepts, and big ideas.” (Pg. 27)</td>
</tr>
<tr>
<td></td>
<td>Curriculum projects “chosen to develop deeper understanding of compelling topic.” (Pg. 13)</td>
</tr>
<tr>
<td>Practice authentic inquiry processes</td>
<td>Fieldwork is modelled…on research of professionals in the field” using appropriate inquiry strategies and “data collection tools.” (Pg. 15)</td>
</tr>
<tr>
<td>Use complex forms of disciplinary communication</td>
<td>“Students are taught to engage in disciplinary discourse that pushes their understanding.” (Pg. 27)</td>
</tr>
<tr>
<td></td>
<td>Student learning products are intentionally designed to develop literacy through reading and research, and writing in appropriate genre to explain or document the product. (Pg. 13)</td>
</tr>
<tr>
<td>Value Beyond School</td>
<td></td>
</tr>
<tr>
<td>Connections to public problems or personal experiences</td>
<td>Design curriculum units around compelling topics. (Pg. 11)</td>
</tr>
<tr>
<td>Learning products have value beyond demonstrating success in school</td>
<td>Student learning “products often meet an authentic need and have an audience and purpose beyond families or the classroom teacher.” (Pg. 13)</td>
</tr>
<tr>
<td></td>
<td>“Data collected in the field are published or presented to real, external audiences.” (Pg. 15)</td>
</tr>
<tr>
<td>Collaboration with science &amp; engineering practitioners</td>
<td>Outside experts, reflecting the community, train students in professional skills, “critique their work against professional standards. Teachers and students maintain ongoing relationships with experts.” (Pg. 15)</td>
</tr>
</tbody>
</table>

*Note. All of the page numbers for the quotes in this table refer to *Expeditionary Learning Core Benchmarks* (Expeditionary Learning Outward Bound, 2003).*
Chapter 5   Features of Authentic Pedagogy over a School Year

This chapter addresses the first broad research question by documenting the implementation of authentic science pedagogy in the classrooms studied. The purpose of this chapter is to use this study’s theoretical model of authentic pedagogy to measure certain features of the instructional ecosystem that support authentic student performance and engagement in the ninth grade science classrooms studied. This study’s theoretical model of authentic science and engineering pedagogy and performance defined three components of authentic pedagogy: affordances for the construction of knowledge; affordances for the use of science and engineering meaning-making processes; and affordances to find value beyond demonstrating success in school (Figure 3.1).

Two measures of authentic pedagogy were developed, one from document analysis of teacher designed assessment tasks and a second from field notes on classroom instruction from October through June. Table 5.1 lists the variables measured and how often they were collected.

The teacher designed assessment tasks were assignments students were given as the culmination of units of instruction. They were designed to assess whether students were able to demonstrate proficiency on a specific Learning Target or Learning Targets in keeping with the school-wide proficiency-based assessment system. They are the assessments used by the teacher to assign grades. For brevity these are generally referred to as assessment tasks.
The purpose of this study is not to measure overall teaching quality in the case. The model of authentic science pedagogy adopted by this study was designed to provide valuable information on the relationship between teaching practice and student learning. The incarnation of that model described in this chapter has not, as yet, been previously shown to improve science literacy. In addition, this model is focused on intellectual achievement goals of education, but does not measure the teacher-student relationship, which many believe is the critical dimension in student learning.

This research is descriptive, not diagnostic or evaluative. To be effective, teachers must prioritize allocation of their time between many efforts, such as, individualizing feedback on student work, differentiating learning and assessment opportunities to individual student needs, developing curriculum with as many authentic components as possible, and collaborating with colleagues to develop school cultural practices supporting excellence. This study isn’t comprehensive enough to deal with the full range of important aspects of teacher practice.

Research Question 1

How does the presence of features of authentic science pedagogy vary over the school year?

5.1 Methods

Tasks the teacher used to assess student proficiency on course Learning Targets were collected and scored for features of authentic pedagogy in Table 5.1. Most of the teacher’s assessment tasks during the year were scored to represent the range of tasks
used for assessment. Except for the assessment tasks used to establish the interrater reliability of the scoring, all assessment tasks were scored by this investigator.

Secondly, the investigator’s field notes from observations of classroom instruction from October through June were scored for a set indicators representative of this study’s theoretical model of authentic pedagogy. These indicators defined and measured the features of authentic science and engineering pedagogy present. Scores for each component of authenticity and a Combined Authenticity score shown in Table 5.1 were determined from the authenticity indicators scored. Aggregated scores were calculated for selected instructional units and over the whole study period. All of these observations were made by and scored by this investigator.

5.1.1 Measures of authentic pedagogy in the teacher’s assessment tasks

To characterize the authenticity of the teacher’s pedagogy in his assessment tasks, seven indicators representing the three components of authenticity in Table 5.1 were scored for selected assessment tasks used to measure student proficiency on specific Learning Targets. Table 5.3 describes the seven indicators measured and their relationship to the three components of authentic pedagogy.

5.1.1.1 Data Sources

The assessment task documents for 22 of these tasks, the task scoring guides provided to students by the teacher, and the Learning Targets defining learning proficiencies for the tasks were analyzed. Field notes were also consulted to clarify these documents. The documents were collected from the teacher during the study period from
October to June. The task documents were those the teacher gave students to define the tasks they were to complete. Each task was always explicitly associated with measuring student proficiency on one to three Learning Targets. Teacher statements of Learning Targets by trimester are in Appendix A. Table 5.2 describes the twenty-two assessment tasks scored for authenticity.

Table 5.2 also identifies alternative assessment tasks addressing the same Learning Targets. These alternative tasks provided students choices or multiple opportunities to demonstrate proficiency. Some alternative tasks are noted as available to students concurrently with another task the teacher presented as a first option in class. Other alternative tasks are noted as available at a later time for students that had not yet achieved proficiency on an earlier task. The 22 assessment tasks scored include an example from nearly all the Learning Targets during the study period, but they do not include all the alternative tasks made available to students.

5.1.1.2 Instrument Description

The Center for Organization and Restructuring of Schools (CORS) developed scoring guides for authentic assessment tasks in social studies and mathematics (Newmann & Associates, 1996). The CORS scoring guides measured seven characteristics of authenticity representing the three components of authentic pedagogy in the theoretical model common to this study and the CORS research. These components correspond to those in Table 5.1, affordances for Construction of Knowledge, affordances for use of Science Meaning-Making Processes, Substantive Conversation, and affordances to find Value Beyond School.
The CORS scoring guides and an adaptation of them for science assessment tasks by the Research Institute on Secondary Education Reform for Youth with Disabilities (2001) were adapted to write the scoring guides for science and engineering assessment tasks used in this study. The scoring guides used in this study are in Appendix A.

The seven task authenticity indicators (TAIs) scored for each assessment task and how they represent the three components of authentic pedagogy are summarized in Table 5.3. Scores on these indicators are a measure of the affordances for authentic intellectual work experienced by students. Each TAI was scored on a 1-3 or 1-4 scale from no to low authenticity (1) to high authenticity (3 or 4). These scores for each task were entered into SPSS, which was used to manage and analyze the data to compute the component and Combined Authenticity scores. Scores on the seven indicators are summed for scores on three authenticity components and a Combined Authenticity score (Table 5.3).

High authenticity for TAIs (coded as 3 or 4) is defined as a dominant or clear expectation. Moderate authenticity is defined as some or unclear expectations for the affordance described by the indicator. A low score (coded as 1) means that there is no or minimal affordance in the assessment task or classroom activity.

5.1.1.3 Reliability of assessment task authenticity measure

To score each TAI consistently across 22 assessment tasks with varying formats, each task was analyzed for the specific performance or performances they asked from students. These performances were analyzed for affordances to do the cognitive work measured by each of the indicators in Table 5.3.
To demonstrate the reliability of the scoring guide, the investigator and a published university science education researcher independently scored 18% of the tasks (4 tasks). Only the first six indicators were used to demonstrate interrater reliability because the seventh indicator did not require a judgment by the scorer. The seventh indicator was based on first-hand knowledge or teacher report of what occurred in class. Of the 24 individual indicator scores for these four tasks, 87.5% were in agreement.

Table 5.3 shows that the range of Indicator 5 (1-4) is greater than the other five indicators (1-3). Therefore, only the percent agreement of 75% can be reported for Indicator 5. The percent agreement for Indicators 1 – 4 and Indicator 6 is 90%. Cohen’s Kappa coefficient for judge agreement for these five indicators was determined to be 0.835, (95% CI, .942 - .728) p < .0005. This indicates very good agreement (Altman, 1999).

To summarize the level of authentic pedagogy in the 22 assessment tasks analyzed (Table 5.2), scores for each of the authentic pedagogy components and a Combined Authenticity score were calculated as presented in Table 5.3. To make scores with different ranges more easily comparable, a normalized score was calculated from: (component score – low bound of score range)/(high bound of score range – low bound of score range). This normalizes each score on a scale from 0-1, where 0 measures no authenticity, 0.5 measures moderate authenticity, and 1 measures high authenticity. A one on this scale requires that all seven indicators are scored at their highest authenticity (3 or 4).
5.1.1.4 Validity of assessment task authenticity measure

This study’s measure of authenticity of assessment tasks has validity within this study. First, construct validity is based on using the same indicators and scoring levels of each as those used in the CORS scoring guide and the previous research conducted with their instrument. In many cases this study’s scoring guide was worded the same (Appendix A). Second, content validity was maintained during the adaptation for science and engineering by careful reference of science education policy documents (Katehi et al., 2009; National Research Council, 2012), the science education literature (Chinn & Malhotra, 2002; H.-S. Lee & Songer, 2003; McGinn & Roth, 1999; J. Singer et al., 2000), and consultation with engineering education experts (C. Sneider, personal communication, February 5, 2011; G. Recktenwald, personal communication, May 16, 2011). Judgments of content validity also depend on the expertise of this investigator.

To calibrate the scoring guide and check content validity, the investigator interviewed the teacher on his application of the scoring guides (Appendix A). The teacher’s scoring of a sample of three of the assessment tasks was compared with the investigator’s scoring of the same tasks.

The sensitivity of assessment task authenticity measures to detect differences between the authenticities of assessment tasks is a function of the ranges of each of the measures in Table 5.3. The component authenticity ranges are from four to seven, so it is reasonable to categorize authenticity levels of the normalized component and combined scores as: not authentic (0-0.19), moderately low authenticity (0.2-0.39), moderate
authenticity (0.4-0.59), moderately high authenticity (0.6-0.79), and highly authentic (0.8-1).

5.1.2 Measures of authentic pedagogy from observations of classroom instruction

In addition to documenting the features of authenticity in the teacher’s assessment tasks, the authenticity of classroom instruction was measured using an instrument created for this study. Classroom instruction was observed and coded for features of authentic pedagogy from October to June. The class lessons chosen for observation and the measurement instrument are described below.

5.1.2.1 Data Sources

Direct observations of classroom instruction began in October, about half way through the fall trimester. After the first instructional unit of the school year, observations were made on at least one day during all other instructional units as shown in Table 5.4. Based on teacher interviews and the investigator’s teaching experience, lessons expected to show the influences of active inquiry and the Value Beyond School component of authenticity on student behavior were chosen for more frequent observation. Lessons devoted to preparing for and completing assessment tasks with active inquiry and Value Beyond School components were prioritized. For example, few resources were expended observing students taking quizzes or watching films. Most instructional units addressing Learning Targets focused on building basic science knowledge were not prioritized for observation.
In the fall trimester observations focused on lessons during the final two instructional units, the RNA Fingerprinting Lab, and the Immune System and Flu Vaccines unit (Table 5.4). They were observed about three days a week. Observations of earlier fall trimester units were very limited. In winter trimester, lessons during the Human Characteristics Lab Analysis unit at the beginning of the trimester and the Teen Brain unit at the end of the fall trimester were observed 3-4 days/week (Table 5.4). Lessons during other instructional units were seldom observed.

After the first two weeks of spring trimester, observations were conducted 4 days a week on average, until they were suspended the last week and a half of the trimester. Unlike the first two trimesters, the field notes during spring trimester included quite a few observations from lessons during almost all the instructional units. The Energy Concepts unit, the Introduction to Engineering Design unit and the Wind Turbine Engineering & Testing unit were chosen to have the most complete set of observations (Table 5.4).

Observations were focused on class periods 2, 3, and 5. Period 1 observations were 4.4% of the total. Because the focus class periods, 2, 3 and 5, were unevenly distributed across the mid-week days, they weren’t always observed with equal frequency, especially during a particular instructional unit.

Coding for characteristics of authentic science pedagogy was done by lesson segments, defined as time periods during a class period focused on a distinct activity with a particular interaction pattern (Stuessy, 2002). For example, a class period might include: (1) seven minutes of the teacher giving verbal directions for an activity with little student input; then (2) ten minutes of students reading texts silently; then (3) students
created concept maps in small groups for 25 minutes; and finally (4) an 8 minute teacher-led classroom discussion of how students have organized the key concepts in their maps to conclude the class period. This lesson includes four distinct lesson segments. Each of these lesson segments is scored independently for the 11 indicators of authentic pedagogy in Table 5.6.

Each identified lesson segment is coded separately because the authenticity indicators identified in Table 5.6, as well as, student behavioral engagement and self-regulated learning are likely to vary widely during these different instructional patterns. Semi-structured interviews with the teacher clarified aspects of instruction when the investigator had questions.

The resulting sampling of the teacher’s instruction is a convenience sample. Over the entire study period from October to June, a total of 500 individual lesson segments of classroom instruction were coded for authenticity of instruction using the 11 Lesson Authenticity Indicators (LAIs) in Table 5.6. These 500 lesson segments encompassed 169 hours of class time and came from about 50% of the class days comprising the entire school year. This dataset is used to characterize the authenticity of instruction for the entire study period and selected instructional units from each trimester.

5.1.2.2 Instrument Description

A Classroom Observation Protocol was developed to measure authentic pedagogy in classroom instruction, as well as, student behavioral engagement, self-regulated learning, and some other aspects of the instructional ecosystem. The instrument developed for measuring authentic pedagogy in classroom instruction included measures
for each of the three components of authenticity measured for the assessment tasks (i.e., Construction of Knowledge, Science Meaning-Making Processes, and Value Beyond School) plus another component called Substantive Conversation (Table 5.1).

Like the instrument developed to measure authenticity in the assessment tasks, the instrument for authenticity in the classroom instruction is based on the theoretical model common to this study and the CORS research. The four components of authentic classroom instruction in Table 5.1 are same ones measured by a CORS instrument (Newmann & Associates, 1996). This study renames the CORS Disciplined Inquiry component Science Meaning-Making Processes.

These four components match the three components of authentic pedagogy in the assessment task instrument because Substantive Conversation contributes to both Construction of Knowledge through articulation of ideas during discourse with peers and to Science Meaning-Making Processes because of the essential role of discourse in validating scientific explanations and engineering designs.

The scoring of the four CORS components of authentic classroom instruction are modified in three ways to fit the goals of this study and the unique aspects of science and engineering education. First, the CORS Construction of Knowledge and Disciplined Inquiry components were adapted to focus on science and engineering instruction using the science and engineering education literature on the reasoning and cultural practices of science and engineering practitioners described in the literature review and theoretical perspective. On this basis, a priori decisions were made to define the instrument.
Second, the eleven Lesson Authenticity Indicators (LAIs) in Table 5.6 were defined. These indicators were developed to be discrete enough to observe concurrently with measures of student engagement. For example: Instruction affords student reasoning practices that support higher-order thinking; Instruction affords a sustained focus on a core science concept, practice, or topic; Instruction affords connections between core science topics and either significant public problems or personal experiences in their lives outside of school. (See Table 5.6, page 48).

Table 5.7 shows how these indicators were used to construct the four components of authentic classroom instruction (i.e., Construction of Knowledge, Science Meaning-Making Processes, Substantive Conversation, and Value Beyond School). The low, moderate, or high scores for subsets of these indicators were transformed into scores for each of the four authenticity components. Some of these discrete indicators contributed to more than one of the four components of authenticity. By contrast, the CORS versions of the four components were scored directly with 1-5 scoring guides (Newmann & Associates, 1996).

Thirdly, the eleventh indicator in Table 5.6 was based on previous research (Chinn & Malhotra, 2002; Tomasek, 2006) to reflect the central role active science inquiry and engineering design play in meeting science and engineering education goals (Krajcik et al., 2003). The score on this indicator was designed to make up part of the Science and Engineering Meaning-Making component of authentic pedagogy.
5.1.2.2.1 Lesson Authenticity Indicator coding

The investigator’s field notes from each lesson segment observed were coded for the 11 LAIs in Table 5.6 using the scoring guides in Appendix B. Coding for the first ten of the indicators is high (scored 3), moderate (scored 2), or low (scored 1) authenticity for each lesson segment. A high rating indicates instruction that affords all students participation in this science reasoning practice. A high rating indicates this affordance is a key strategy strongly supportive of these reasoning practices. A moderate rating indicates that there is some affordance for the reasoning practice, but it is less explicit or given less emphasis. A low score indicates that the affordance is minimal because little or no time is specifically devoted to engaging students in that reasoning practice.

The eleventh indicator, measuring affordance for inquiry reasoning practices characteristic of science and engineering practitioners was scored only when a significant focus of the lesson was on any of the seven cognitive practices of science inquiry or engineering design in the last row of Table 5.6. The practices are present when students use them to collect, manipulate, and analyze data during active engagement with natural phenomena or engineering designs. Each practice was coded as authentic (scored 2) or simple (scored 1) or not present (scored 0). The definitions are based on Tomasek’s adaptation (2006) of Chinn and Malhotra’s (2002) contrast between simple and authentic science inquiry. When instruction in the lesson segment goes beyond simply inquiry to resemble authentic inquiry, it was scored authentic. The scoring guides are in Appendix B.
The LAIs for each lesson segment scored from the classroom observation protocol field notes were entered into SPSS, which was used to manage and analyze the data to compute the component and Combined Authenticity scores.

5.1.2.2.2 **Treatment of missing observations**

It is the nature of naturalistic observation studies that there will be instances where no observations were made resulting in missing data. One or more of the 11 Lesson Authenticity Indicators (LAIs) was not scored in 7.2% of the individual lesson segments (8.8% of hours) of the 500 lesson segments (169 hours) coded. The most frequent reason was an interruption that caused the observer to miss important details that prevented reaching a definitive score for a Lesson Authenticity Indicator. Before analysis each missing value for each LAI was replaced with the mean of the values of that LAI from other lesson segments devoted to the same instructional unit and with a similar instructional pattern.³

5.1.2.2.3 **Determining scores for each authentic pedagogy component of classroom instruction**

The authenticity level of classroom instruction was summarized by scores for each of the three components of authentic pedagogy measured in assessment tasks and a fourth component, Substantive Conversation. These authentic pedagogy component scores were calculated from the Lesson Authenticity Indicator (LAIs) as detailed in Table 5.7 for each of the 500 lesson segments observed. Each component score is normalized
to a scale with a maximum of 24 so it contributes equally to a combined score for each lesson segment.

The Construction of Knowledge score is calculated from a subset of five of the 11 LAIs as shown in Table 5.7. The higher-order thinking indicator (LAI 1) accounts for half of the score, reflecting its importance in the CORS scoring guides. The other four indicators support construction of knowledge. Revisions and tinkering (LAI 3) and multiple representations (LAI 4) are included because of their importance in scientific and engineering thinking (McGinn & Roth, 1999; Roth, 1995; Stuessy, 2002). This score corresponds to the Higher-Order Thinking standard in the CORS scoring guides (Newmann et al., 1995).

The Science Meaning-Making Processes score is determined from four LAIs. Core concepts (LAI 2) is weighted heavily to reflect the Deep Knowledge standard in the CORS scoring guide (Newmann et al., 1995). The revisions and tinkering indicator (LAI 3) and the multiple representations indicator (LAI 4) are added because they are key science and engineering inquiry practices. This authenticity component includes the subscale, Science Inquiry & Engineering Design Practices, which is calculated separately when any of the seven inquiry and design practices in LAI 11 are present in the observed lesson segment. It is normalized to be worth slightly more than one-fourth of the score.4

Substantive Conversation measures the degree to which sustained collaborative classroom discourse promotes making thinking visible, promotes shared understanding, and focuses on core science and engineering concepts. Discourse promotes construction of knowledge (Linn, 2000). It is also central to meaning-making in science and
engineering cultures (Roth & Barton, 2004). The score is the sum of two separately
determined scores based on higher-order thinking (LAI 1) and core concepts (LAI 2),
respectively, because substantive conversations are most authentic when they are focused
on both components. Each of these scores is a geometric combination of LAI scores that
resembles the levels in the CORS scoring guide.5

As was the case for the assessment tasks, the Value Beyond School component
score is determined from the sum of LAIs 8 and 9 (Table 5.7). The sum is multiplied by
the scaling factor in Table 5.7 to normalize the Value Beyond School score to the range
of the other three components.

The four component scores are totaled for a combined score for authenticity of
classroom instruction. As for assessment tasks, normalized scores for each component of
authenticity and the combined score were calculated in the same manner as assessment
tasks: (component score – low bound of score range)/(high bound of score range – low
bound of score range). This standardizes each component and combined score on a scale
from 0-1, where 0 measures no authenticity, 0.5 measures moderate authenticity, and 1
measures high authenticity.

5.1.2.3 Reliability of Lesson Authenticity Indicators scores

To maximize reliability, detailed scoring guides were developed for each
indicator before data collection. These were pilot tested for feasibility of use during
observations in the same teacher’s classrooms during the previous school year. In
addition, the investigator’s field notes from each class period were coded for the 11 LAIs
shortly after the observations were made in almost all cases. After the study period, all of
the coding was reviewed for consistency using the original field notes and notes on scoring decisions recorded at the time of coding. The scoring guides for each level of the eleven indicators are provided in the description of the classroom instruction observation protocol in Appendix B.

5.1.2.4 Validity of classroom instruction authenticity scores

This instrument has content validity based on extensive consultation of science education policy documents (National Research Council, 2012), the science education literature (Chinn & Malhotra, 2002; H.-S. Lee & Songer, 2003; McGinn & Roth, 1999; J. Singer et al., 2000), and engineering education policy documents (“Introduction to engineering design™ analysis of cognitive levels of learning and mathematics and science content,” 2008; Katehi et al., 2009; Technology for All Americans Project, 2007) during design of the instrument, and the high school science teaching experience of the investigator. Construct validity is based on use of the research foundation inherent in the CORS framework and the relationship previously found between authentic pedagogy and student science achievement (Newmann et al., 2007). The Data Analyses section explains how weighting of lesson segment scores by lesson segment duration is designed to improve construct validity for mean authenticity scores calculated at different levels of analysis.

Table 5.7 shows that there are Lesson Authenticity Indicators (LAIs) that are part of calculating more than one of the Construction of Knowledge, Substantive Conversation, and Science Meaning-Making Processes components of authentic instruction. Therefore, some correlation between these measures is built into the
instrument. However, the Value Beyond School component score is constructed with a unique set of two LAI scores (Tables 5.7).

5.1.3 Data Analyses

5.1.3.1 Assessment tasks

Research Question 1 asks whether authentic science pedagogy varied over the school year. To answer this, each of the 22 proficiency assessment tasks in Table 5.2 is analyzed to document the features of authenticity students’ experience in this aspect of the teacher’s pedagogy. Table 5.3 shows how the three component and Combined Authenticity scores are calculated. A descriptive analysis documents the patterns found in the presence of these features of authenticity across the study period and in the levels of authenticity between assessment tasks. Patterns in individual task authenticity indicators (TAIs) reveal details of the teacher’s expectations of students.

The scoring guide for authenticity of the assessment tasks is based on the presence of authentic requirements within a task rather than the task as a whole (Appendix A). Therefore, tasks with many parts that provided multiple performance opportunities to demonstrate authenticity were biased to score higher. This applies to two assessment tasks at the end of the school year. The Wind Turbine Engineering and Testing task stretched over three and a half weeks of class time, had many parts and the student artifacts were 10-20 pages.

By contrast, one week was devoted to the Energy Concepts unit, which was assessed by the Energy Concept Quiz task requiring only a 1 page artifact. Therefore, the
Wind Turbine Engineering and Testing task was divided into three parts based on natural divisions within the overall task and on the number of performance opportunities in each section. Each of these parts was scored independently using the assessment task scoring guide (Appendix A). The scores reported for the task as a whole are the means of the scores on these three parts of the overall task.

The Engineering Challenge 1 unit was shorter, but the task’s structure was similar to the Wind Turbine Engineering and Testing task. It had similar performance opportunities. Therefore, the reported scores are the means of the same three parts as the Wind Turbine task.

5.1.3.2 Determining mean authenticity levels for classroom instruction

Measures of the mean authenticity of classroom instruction for the entire study period are determined to assess this second aspect of the authentic pedagogy students experienced. The total instructional time of lesson segments with high, moderately high, moderately low, and low authenticity during the entire study period are also calculated.

Measures of the authenticity of classroom instruction were also determined for two or three selected instructional units from each trimester. All the measures of authenticity for individual lessons during these instructional units were aggregated and compared. The mean Combined Authenticity and the levels of four components of authentic pedagogy are contrasted between units and across the study period.

Lesson segment lengths varied from just a few minutes to 75 minutes. When calculating the location measures of authenticity of classroom instruction for each instructional unit and the entire study period, the combined and component scores for
each lesson segment were weighted by the length of the lesson segment in minutes using the SPSS Weight Cases procedure. This treats longer lesson segments proportionally to their frequency in the sample, giving greater construct validity to central tendency measures for these variables. Without weighting each lesson segment by its length, shorter lesson segments make up a greater proportion of the means only because they are more frequent.

Table 5.5 lists the selected seven instructional units from Table 5.4 that were observed and scored for authentic pedagogy during at least 60% of the class days associated with these units. Three hundred sixty-four (73%) of the 500 lesson segments scored are from these units. This is 123 hours of class time. The levels of authentic affordances experienced by students in the teacher’s classroom instruction and assessment tasks were determined for each of these instructional units. These affordances include each of the variables in Table 5.1.

These instructional units were chosen to be observed often enough to have a rich enough observational record for a reliable estimate of measures of central tendency for the authenticity of the classroom instruction during them. The observations for these instructional units included most of the learning activities the teacher designed to build knowledge and skills devoted to preparation for and completion of the assessment tasks the teacher used to assess student proficiency and assign grades. They also included most of the instructional patterns, such as, teacher presented task instructions, direct instruction, individual seatwork, small group work, and student-directed project work, during each of these units.
The RNA Fingerprinting Lab unit was included in the analysis because it was observed at least half of the days devoted to the unit. However, closer inspection during post hoc analysis shows that the sample of lesson segments observed isn’t as representative as the other six units. It did not include the full range of major activities in this unit. Fewer lesson segments and less class time was observed than in the other units. Only 5.5 hours of class time was observed, one-third of the mean class time observed per unit. Days students were performing the RNA Fingerprinting lab activity were observed, but the class time focused on interpreting the lab results were not observed at all.

Based on examination of the teacher’s lesson planning and class activity documents, and patterns in the teacher’s instruction in other units, it is likely the sample of lesson segments scored for the RNA Fingerprinting Lab unit biased the mean Construction of Knowledge component score to be lower than would have been measured if some of the missed lesson segments had been included. It is less likely that the mean Science Meaning-Making Processes and Value Beyond School component scores are inaccurate. Interpretation of these results for this unit take this into account.

Because of the ordinal nature of the underlying Lesson Authenticity Indicators, the convenience samples used, and the absence of interrater reliability, an appropriate approach to describing and detecting differences in mean authenticity levels for classroom instruction for each unit is the one adopted for the assessment tasks. The normalized component and combined mean scores are categorized as: not authentic (0-0.19), to moderately low authenticity (0.2-0.39), moderate authenticity (0.4-0.59), moderately high authenticity (0.6-0.79), and highly authentic (0.8-1).
5.1.3.3 Determining authentic pedagogy levels for selected instructional units

Table 5.5 shows that some of these instructional units were assessed by multiple assessment tasks. The teacher’s flexibility in designing the curriculum to meet instructional goals and respond to student needs required special treatment of the assessment task scores from three of the seven instructional units, so that mean task authenticity was a measure of student experience for the instructional unit as a whole.

First, Table 5.5 shows two assessment tasks were used by the teacher to assess student proficiency for the RNA Fingerprinting Lab Analysis instructional unit. However, scores for the lab analysis assessment task were used as the measures of authentic affordance for this unit instead of the RNA Fingerprinting Quiz because the quiz was only made available to students at the end of the trimester as an additional opportunity to demonstrate proficiency. The RNA Fingerprinting Lab Analysis task was the task the teacher was preparing students for during the classroom instruction scored for authenticity in this instructional unit.

Next, Table 5.5 shows the Immune System & Flu Vaccines instructional unit was assessed by four tasks that the teacher used to assess two Learning Targets. Even though the teacher assessed proficiency on the two Learning Targets independently, the teacher taught this content in a sequential, connected context focused on understanding the decision whether to get an influenza vaccine. Therefore, all the lesson segments addressing both these Learning Targets are considered one instructional unit. Authenticity scores for classroom instruction from all these lesson segments are aggregated to measure the authentic affordances students experienced during this unit.
The teacher focused his instruction on the public service announcement task in class, but made it clear to students that the quiz format would be an option for them. Students freely chose between these options. Students worked on either the PSAs or the quizzes during most of the lessons observed. Based on field notes, about 90% of the students chose the PSA option.7

Therefore, observations of student behavior and classroom instruction during this instructional unit were primarily influenced by the PSA tasks for Learning Targets 1.4 and 1.5. Post hoc it was decided that the most valid measure of the authenticity for these tasks comparable with the measure of the authenticity of classroom instruction for this instructional unit was the mean of the authenticity scores for only the two PSA tasks.

Lastly, Table 5.5 lists two assessment tasks and two Learning Targets for the Human Characteristics Lab Analysis unit. Data display in graphs and tables is assessed by the first of these Learning Targets. The interpretation of scientific data is assessed by the second Learning Target. This unit included both of these. In this case, the teacher judged that the first of the tasks for the unit in Table 5.5 was completed by students before sufficient instruction on all the content.

What the teacher originally intended as a combined assessment became a formative assessment. The second task was a second opportunity to show proficiency on the same Learning Targets after all of the content in both Learning Targets was taught. Therefore, the mean of the scores from the two similar tasks is used as the measure of the authenticity of the assessment tasks in this unit. This mean was compared with the mean
of authenticity scores from all the lesson segments scored for classroom instruction in the unit.

The other four instructional units were each assessed by one assessment task. The authenticity scores for that task were compared with those for classroom instruction from all the lesson segments in the unit.

5.1.3.4 Interpreting classroom instruction and assessment task authenticity levels

Scoring some of the TAIs and LAIs include common elements, as can be seen in the scoring guides in Appendix A and Appendix B, respectively. The scoring guides for LAIs include both specific observations of teacher and student behaviors and judgments on the cognitive complexity afforded by the teacher’s instruction observed during each lesson segment.\(^8\) The judgment of cognitive complexity involved in scoring LAI 1, for example, may be influenced by this investigator’s judgment of cognitive complexity when he scored TAI 1 and TAI 2 for the assessment task(s) for the same instructional unit.\(^9\) Some correlation between the two measures for the same instructional unit is an artifact the construction and implementation of the measurements.

While not completely free of overlap with the task authenticity indicators (TAIs), the Lesson Authenticity Indicators consider many factors that are not part of the TAIs. Central tendency measures of collections of lesson segment scores are also weighted by the time students experience instruction at different LAI levels, which is independent of TAI scoring judgments. The instrument for classroom instruction was designed to produce a valuable measurement of student experience of affordances for constructing knowledge, using science meaning-making process and finding value beyond
demonstrating success in school. It is believed that the methodology used produces a measurement with value beyond that in the measurement of the same affordances in the assessment tasks.

This study’s measures of two aspects of authentic pedagogy are designed to consistently implement the common theoretical model from the Center for Organization and Restructuring of Schools (CORS) underpinning both measures. Both measures construct the components of authenticity in the CORS instruments from discrete indicators, but there are differences.

This study’s scoring guide for the assessment tasks has the same indicators with the same scoring scale as the CORS scoring guide (Table 5.3). They were just adapted for science and engineering. However, the indicators and scoring protocol for the instrument for authenticity of classroom instruction is quite different. As described above, it uses the set of discrete indicators in Table 5.6 to score the same four authenticity components in the CORS instrument. Component scores are determined from algorithms described in Table 5.7 and designed to simulate the set of levels in the CORS scoring guides. This is different from choosing a level from one scoring guide for each component in the CORS instrument.

There are also differences in level of analysis. Each assessment task measurement describes that task. Table 5.5 shows when the mean of two assessment tasks are used to characterize the authentic affordances in those tasks for each of the seven selected instructional units. However, levels of authentic affordances in classroom instruction for each instructional unit are the mean authenticity scores from many lesson segments.
scored during that unit. These means also take into account class time in each lesson segment to better reflect student experience of authenticity.

There is a theoretical basis for concluding that the two instruments measure the same constructs on a scale from high authenticity to moderate to low to no authenticity. However, the differences in the instruments – and for classroom instruction, the convenience samples and time-weighting of lesson segment scores – mean that this study will not claim the scales are quantitatively calibrated. Still, large differences within and between the measures suggest real differences.

5.2 Results

To answer Research Question 1, the authenticity of the pedagogy that defines the students’ instructional ecosystem is measured in the teacher’s assessment tasks and in classroom instruction. This chapter documents the patterns in the features of authentic science pedagogy in these two variables across the study period. It presents the overall authenticity levels and how levels change over the school year and between instructional units. The authenticity components that contribute to the combined authenticities are examined. Patterns in the individual authenticity indicators used to calculate the component scores are used to illustrate the aspects of authenticity that were most prominent in the teacher’s practice.

In Chapter 7 the measurements of authenticity levels of science pedagogy discussed in this chapter are related to student performance in science classwork, as well as their behavioral engagement and self-regulated learning. This will address the second
broad research goal of this study to determine if there is evidence that more authentic pedagogy produces gains in student achievement, engagement, and self-regulated learning.

5.2.1 The teacher’s assessment tasks

The topics and formats of each of the 22 assessment tasks scored are briefly described in Table 5.2. The authentic pedagogy scores for 22 assessment task documents are in Appendix C, Table C.1. The score of 9 for the lowest scoring task, the Genes to Traits Quiz, results from all of the assessment task authenticity indicators but two scoring 1, ‘not authentic’ (Appendix C, Table C.1). The score of 19.7 for the Wind Turbine Engineering and Testing (Table 5.8), is achieved because most of the seven assessment task authenticity indicators scored at the top of their scales and none are scored ‘not authentic’ (1).

On a scale normalized to 0-1, the Combined Authenticity of tasks scored ranges from 0.13 for the Genes to Traits Quiz to 0.79 for Wind Turbine Engineering and Testing (Appendix C, Table C.2). This represents a range from low to moderately high authenticity.

Figure 5.1 shows that the most prominent pattern in the Combined Authenticity scores for the 22 assessment tasks is their variability. During the study period there wasn’t a trend of increasing or decreasing Combined Authenticity. In fact, there are several assessment tasks with highly authentic expectations in the first trimester. Figure 5.2 shows that there is a similarly high degree in variation in the authenticity component scores for these assessment tasks.
Table 5.9 shows that the means of normalized scores for the Construction of Knowledge and Science Meaning-Making Processes components of authenticity from all 22 assessment tasks are nearly identical at 0.56 and 0.57, respectively. The range in component authenticity scores for individual assessment tasks varies from 0, or nearly 0, to 1 for both of these components, indicating a wide variation among tasks.

Some of this variation can be seen in the means of Task Authenticity Indicators (TAIs) 1-6 from all tasks, which are all in a moderate to moderately high authenticity range from 0.45 to 0.66 (Table 5.9). The only TAI scoring consistently low was TAI 7, Audience beyond the school (Table 5.8).

Table 5.8 shows that the highest scoring TAI across all 22 tasks were Organization of information (TAI 1, mean = 0.66), Elaborated communication (TAI 5, mean = 0.61), and Science or engineering content (TAI 3, mean = 0.59). The moderately high TAI 1 score measures a mix of tasks with some and clear affordance for active organization of information in ways new to students over reproducing information. Elaborated scientific communication of an argument or explanation or engineering design, TAI 5, was also a frequent expectation. Only one task scored not authentic (1) for Science or engineering content, TAI 3, indicating that nearly all tasks require at least some understanding of concepts, theories, or perspectives central to science or engineering.

Table 5.8 shows that TAI 6, connection to the world beyond the classroom, is similar to the other TAI. However, the mean for TAI 7 (0.13), measuring presentation of the task artifact to an audience beyond the teacher, is responsible for the Value Beyond...
School component averaging only moderately low (0.29). Opportunities for students to present their task artifacts outside the school environment, high authenticity for TAI 7, didn’t occur. Expectations that the students would direct their product to an audience beyond the classroom, but still within the school, a TAI 7 score of 3, only occurred for the Teen Brain PSA and the Antiviral Drug Design.¹⁰

5.2.1.1 Assessment tasks with lower authenticity

There are some patterns in the Combined Authenticity scores of the individual assessment tasks. Table 5.2 identifies ten of the assessment tasks formatted as a quiz or exam. These tasks are a mixture of short answer and open response items focused on recall and application of factual knowledge or algorithms and relatively simple concepts. When completing quizzes, students were always allowed to use any notes or texts they had available.

Seven of these quizzes have low and moderately low combined authenticities. They were the assessment tasks with the lowest combined authenticities. None of these tasks scored above a 2 on any of the seven TAIs, resulting in moderate to low scores for each component of authenticity (Appendix C, Table C.2). All of these tasks include some affordance for Organization of scientific information (TAI 1), use of Science or engineering core concepts (TAI 3) and/or Elaborated communication (TAI 5). In addition, all these tasks make only limited Connections to the world beyond the classroom (TAI 6) and don’t address an Audience beyond the teacher (TAI 7) resulting in low authenticity for the Value Beyond School authenticity component (Appendix C, Table C.1).¹¹
5.2.1.2 Assessment tasks with higher authenticity

None of the assessment tasks scored high – greater than 0.8 – for Combined Authenticity (Appendix C, Table C.2). The Wind Turbine Engineering and Testing task has the highest combined score, 0.79, at the border between moderately high authenticity and highly authentic level. There are seven other assessment tasks that score moderately high Combined Authenticity, above 0.6, in Figure 5.1. Their Combined Authenticity scores are unique combinations of component scores and represent different blends of task characteristics.

Table 5.2 shows there is great variety within these tasks. For example, the Antiviral Drug Design task asks students to use knowledge regarding how a flu virus infects cells to present a marketing poster for a creative idea for a drug that could interfere with viral infection or reproduction. This task’s authenticity lies in Construction of Knowledge, communication of a model for a biological intervention (TAI 5) and Value Beyond School.

By contrast, the Memory Experiment Data Analysis task builds skills on tabular and graphic data display and interpretation, by collection and analysis of a set of class data on a particular memory effect. This task scores at the top of the authenticity scale for both Construction of Knowledge and Science Meaning-Making Processes.

However, there are some patterns in the scores. All but one of these eight assessment tasks scored moderately high or highly authentic, at least 0.70, on both the Construction of Knowledge and Science Meaning-Making authenticity components and
relatively lower on the Value Beyond School component of authentic pedagogy (Figure 5.2).

The high and moderately high scores on Construction of Knowledge resulted because many of these eight moderately high scoring assessment tasks asked for organization of experimental or engineering data, earning a high score on TAI 1 (Appendix C, Table C.1). In addition, most of these tasks also ask for choices between data presentation options or alternative explanations (TAI 2).

These assessment tasks asked students to use important science concepts (TAI 3), to write a novel argument (TAI 5) supported by a scientific model (TAI 4), or by data or design records. This earns all of these tasks, except the Teen Brain PSA, high or moderately high scores in each of the three indicators making up the Science Meaning-Making score (Table 5.8).

However, the Value Beyond School component scores for most of the assessment tasks with moderately high Combined Authenticity were only 0.4 or 0.2 (Figure 5.2). The Teen Brain PSA, the Flu Vaccines PSA and the Antiviral Drug Design task also had relatively high Value Beyond School scores, contributing to moderately high Combined Authenticity scores on these tasks.

To summarize, the analysis of 22 assessment tasks reveals a very wide range of scores for each of the indicators of authenticity of the tasks, resulting in a wide variety in component and Combined Authenticity scores between tasks. None of the tasks score highly authentic, although one is right on the edge of this level. However, students are asked to complete some moderately high Combined Authenticity tasks throughout the
school year. There are also low authenticity tasks throughout the school year. In this mix of assessments, students are afforded opportunities for authentic performance for each of the authenticity characteristics embodied in the task authenticity indicators multiples times during the year, except for audience beyond the school, TAI 7.

5.2.2 Classroom instruction across the entire study period

The 500 lesson segments scored for authenticity of classroom instruction over the entire study period from October to June came from about 50% of the class days comprising the entire school year. Table 5.10 shows the distribution of the 500 lesson segments scored by school trimester with the total for the entire study period. The amount of class time observed increased from fall to spring trimesters. The 500 lesson segments scored cover a total of 169 hours of class time observed.

The first set of scores in Table 5.11 are combined and component authenticity of classroom instruction scores on the normalized scale from 0-1 for the entire set of 500 time-weighted lesson segments over the whole study period. The mean authenticity scores for classroom instruction indicate the prevalence and the quality of affordances for authentic intellectual work in the instruction. The mean authenticity of the combined measure and all the components except Value Beyond School score at a moderate level. The Value Beyond School component scores moderately low authenticity.

It is common in high school classrooms for some class time to be used by a teacher to give students directions for activities, or to convey content without student input. During these lesson segments, many of the LAIs always or frequently score low. For example, when all of the discourse is dominated by the teacher, affordances for
students to use Higher-order thinking (LAI 1), engage in Collaborative discourse (LAI 6), or engage in inquiry reasoning practices (LAI 11) and some other LAIs are limited. For this reason, mean classroom instruction authenticity scores at or near the top of the authenticity scale are not likely during these lesson segments.

In about 30% of the class time observed, the lesson segments were dominated by teacher talk without students having much opportunity to be part of the discourse. Table 5.11 shows that when these lesson segments are removed from the dataset, the mean authenticities of the instruction for the first three authenticity components are measured moderately high authenticity and the Value Beyond School component still averages moderately low authenticity. The Combined Authenticity averages moderately high.

### 5.2.2.1 Class time that students experience authentic instruction

Student experience of authenticity can also be viewed through the amount of class time scored at high, moderate or low authentic instruction. Figure 5.3 shows that class time observed at low, moderately low, moderate, moderately high and high categories of Combined Authenticity is fairly evenly distributed. The two highest categories are lower and higher than expected for an even distribution, respectively. The peak of the distribution is about 0.6.

Figure 5.3 shows that there are differences in the distributions of class time by authenticity for different authentic instruction components. The Construction of Knowledge and Substantive Conversation components score moderately high or high almost half of class time observed. Substantive Conversation measures sustained classroom discourse that supports Construction of Knowledge and Science Meaning-
Making Processes. Table 5.7 shows the LAIs that contribute to this component. A close examination of the frequencies of the lesson segment scores shows that discourse when most students were sharing their thinking occurred, but was infrequent.13

Figure 5.3 shows that the time observed at different authenticity levels is more evenly distributed between authenticity levels for the Science Meaning-Making Processes component. Students experienced moderately high or moderate authenticity instruction about 60% of class time for this component. The Science Inquiry & Engineering Design Practices indicator is the primary reason for this difference with the Construction of Knowledge and Substantive Conversation components.

During about half of the class time observed, the Science Inquiry & Engineering Design Practices indicator scored zero because these practices were not a part of the instruction (Figure 5.3). Students only experience moderate or higher authenticity for these practices during about a quarter of the total class time observed. The mean Science Meaning-Making Processes component is still moderately authentic, like the Construction of Knowledge and Substantive Conversation components, primarily because instruction frequently afforded Sustained focus on science concepts (LAI 2). Its mean over the entire study period was 0.66, the highest of any of the LAIs.

In contrast to the other three authenticity components, Figure 5.3 shows that students seldom experienced high or moderately high authenticity for the Value Beyond School authentic instruction component. The primary reason is that very little class time is devoted to presenting work to an audience beyond the classroom (LAI 9). Less than 9% of class time observed includes students communicating or preparing to communicate
their knowledge, investigating a problem, or taking some action to influence or assist others beyond the teacher or the classroom.

5.2.3 Classroom instruction in selected instructional units

Table 5.11 shows that the mean component authenticity scores measured in classroom instruction over the entire study period are moderate at very close to 0.5 for the Construction of Knowledge, Science Meaning-Making Processes, and Substantive Conversation components, while the Value Beyond School component scores moderately low. In contrast, the component authenticity scores vary considerably between the seven instructional units with sufficient observations to estimate authenticity of instruction for each unit (Figure 5.4).

The combined authenticities for classroom instruction for the four instructional units from the fall and winter trimesters score moderately authentic (Table 5.12). By contrast, the combined authenticities for the three instructional units from the spring trimester show a greater range, from moderately low for the Energy Concepts unit to moderately high authenticity for the Wind Turbine Engineering and Testing unit. The scores show that the instrument used does detect differences between units in authenticity component contributions to overall authenticity of instruction.

The component authenticity scores vary from moderately high (0.69) to 0 (Figure 5.4). Instructional units with similar combined authenticities, such as those scored in the fall and winter trimesters, have considerable differences in the components of authenticity scores that make up each combined score (Figure 5.4).
The Energy Concepts unit had the lowest combined score for any instructional unit. It was focused on building understanding of basic science concepts and was assessed by a quiz with items that weren’t as complex as those expected in the artifacts students produced in the other units scored. Fall and winter trimesters included similarly focused units (Table 5.2), but they weren’t observed frequently enough to estimate the corresponding authenticity of classroom instruction.

Figure 5.4 shows why the Energy Concepts unit scores lower than the others. The Construction of Knowledge, Science Meaning-Making Processes, and Substantive Conversation components all score in or close to the moderately low range. However, the very low score for the Value Beyond School component contributes strongly to the lowest combined score of all the selected instructional units.

The Combined Authenticity score for the Wind Turbine Engineering and Testing unit is the highest scored. The primary reason is that the Construction of Knowledge, Science Meaning-Making Processes, and Substantive Conversation components all score moderately high (Figure 5.4). The moderate authenticity score for the Value Beyond School component doesn’t drag down the combined score like it does for the Engineering Challenge 1 unit.

**5.2.3.1 Patterns in the classroom instruction authenticity indicators**

The differences in mean Lesson Authenticity Indicator (LAI) scores give additional detail about the classroom instruction during each of the seven instructional units selected for intensive observation. There is much variation between mean LAI scores within and between these instructional units (Table 5.13).
LAIs 8 and 9 are summed to get the Value Beyond School authenticity component (Table 5.7). Table 5.13 shows that LAI 8, Connection to the world beyond the classroom, scores highly authentic in more units than any other LAI, including the three instructional units at the end of each trimester. These end-of-trimester units ask students to create an artifact that integrates science or engineering concepts with the trimester’s ‘compelling topic’. Only the Energy Concepts and Engineering Challenge 1 units have no or almost no Connection to the world beyond the classroom (LAI 8). By contrast, LAI 9, Audience beyond the classroom, scores very low for six of seven units. These scores explain why the Value Beyond School component is the lowest scoring authenticity component for four of seven units, and varies the most between units.

Table 5.13 shows that the Teen Brain unit scores highest of any unit for the Value Beyond School component because the instruction affords highly authentic Connections to the world beyond the classroom (LAI 8) and is the only unit that scores moderately authentic for audience beyond the classroom (LAI 9). This is the only unit of the seven when students were encouraged to write a product to an adult outside the classroom.

LAI 2, Core concepts, is the only indicator that scores moderate to highly authentic for each unit (Table 5.13). All seven units scored included affordances for exploring and applying core science or engineering concepts. The teacher’s instruction consistently included a focus on important science and/or engineering concepts. The first two units scored the highest because students were focused on fundamental biology concepts about how molecular mechanisms are the basis for structures and processes with observable functions and effects at the macroscopic level.
LAI 2 accounts for half of the Science Meaning-Making Processes component score. Most of the rest of that score comes from LAI 11, the Science inquiry and engineering design sub-score (Table 5.7). LAI 11 varies greatly depending on the authenticity of the science inquiry and engineering design affordances during each unit. Table 5.13 shows three units score very low. The RNA Fingerprinting Lab and the Wind Turbine Engineering and Testing units score highest for the Science Meaning-Making Processes component because they scored relatively high for both LAI 2 and LAI 11.

LAI 5 and 6 are the primary contributors to the Substantive Conversation component of authentic instruction (Table 5.7). They characterize the frequency and depth of student discourse on science and engineering. Sustained discourse, LAI 5, consistently scored moderate indicating that it occurred but didn’t involve most students (Table 5.13). LAI 6, Collaborative discourse, measures whether students are sharing their higher-order thinking about core science ideas. The mostly moderately low scores for LAI 6 indicate that Collaborative discourse was infrequent across all seven instructional units (Table 5.13).

Five LAIs contribute to the Construction of Knowledge component score (Table 5.7). LAI 1, Higher-order thinking accounts for half of the score. LAI 1 scores moderately low to moderately high authenticity across the seven units, indicating that affordances for higher-order thinking were present in instruction for all units, but did not predominate for any of the units (Table 5.13).

LAIs 3 and 4 both contribute to the Construction of Knowledge and Science Meaning-Making Processes components (Table 5.7). The Wind Turbine unit is the only
one that emphasizes Revisions and tinkering, LAI 3, as students were given considerable time to optimize their designs (Table 5.13). LAI 3 was consistently the lowest scoring indicator, except for audience beyond the classroom, LAI 9. Interactions with representations, LAI 4, scored moderate authenticity in most tasks (Table 5.13), indicating that it occurred, but much the interaction with representations is passive.

The last two LAIs that contributed to the Construction of Knowledge component are LAI 7, Metacognitive facilitation, and LAI 10, Student responsibility for learning (Table 5.7). Both LAIs ranged from low to moderately high authenticity across the seven instructional units. Metacognitive facilitation scored moderately high for each of the units at the end of each trimester. This score level indicates the teacher frequently modeled or discussed metacognitive strategies, but infrequently provided specific time or explicit support for practicing these strategies in class.

Student responsibility for managing their learning, LAI10, scores highest for the engineering units in the spring (Table 5.13). In these units students are given time to make and implement decisions on optimizing their designs. Their moderately high authenticity scores indicate that overall in these units that the teacher gives students some to most of the responsibility for important decisions about using class time and the content and format of the artifacts they produce to meet the Learning Targets.

To summarize, differences in the four components of authentic classroom instruction between the seven instructional units in Figure 5.4 are explained by the patterns in LAI scores in Table 5.13. None of the LAIs score consistently highly authentic. Only focus on Core concepts (LAI 2) and Connections to the world beyond
the classroom (LAI 8) score highly authentic in some units. Affording higher-order thinking (LAI 1), use of Multiple representations (LAI 4), and Metacognitive facilitation (LAI 7) score moderately authentic in three units each. Student responsibility for learning (LAI 10) is only moderately high authenticity during the engineering tasks in spring trimester. Moderately authentic scores for LAIs indicate those affordances were present, but didn’t predominate in the unit. LAI 11, Science inquiry and engineering design practices, is only part of some units. Revisions and tinkering (LAI 3), Sustained and Collaborative discourse (LAIs 5 and 6), and Audience beyond the school (LAI 9) are the least prominent in the teacher’s practice during these instructional units.

It is also instructive to review the Wind Turbine Engineering and Testing unit as it scores moderate authenticity (.59), the highest level of any of the seven instructional units (Table 5.12). This is because it is the only task to score moderately high authenticity for 3 of the four components of authentic instruction (Figure 5.4). Only the Value Beyond School component scores only moderately authentic. This broad base of authenticity is because seven of the 11 LAIs score moderately high authenticity and one, connections to the world beyond the classroom (LAI 8) scores highly authentic (Table 5.13). Only the discourse LAIs, 5 and 6, and Audience beyond the classroom (LAI 9) score less than moderately high authenticity. During the wind turbine unit students are observed designing, testing, and revising an engineering design, emphasizing design optimization (LAI 1, 3). Student responsibility for managing their learning (LAI 10) scores highest of any of the seven units (Table 5.13). These are the reasons this unit scores highest for Construction of Knowledge. Instruction also provides some affordance for students to
apply energy transfer and conservation concepts to their designs (LAI 2), but it is a limited part of the unit.

5.2.4 Summary of differences in authenticity between selected instructional units

5.2.4.1 Classroom instruction

Of the seven instructional units selected to have sufficient classroom observations to determine estimates of the mean authenticity of classroom instruction, the Wind Turbine Engineering and Testing unit has the greatest Combined Authenticity (Figure 5.5). The Energy Concepts unit has the lowest Combined Authenticity. As noted, the RNA Fingerprinting Lab Analysis sample makes comparisons with the Combined Authenticity measures of other instructional units unjustified. The other four instructional units have similar combined authenticities that lie between the Wind Turbine Engineering and Testing and Energy Concepts units (Figure 5.5).

The most highly afforded components of authenticity measured in the remaining six instructional units are the Construction of Knowledge component in the Immune System and Flu Vaccines unit and the Construction of Knowledge and Science Meaning-Making Processes components for the Wind Turbine Engineering and Testing unit (Figure 5.6B). All of these score moderately high authenticity. The lowest affordance measured for Construction of Knowledge is in the Energy Concepts unit. The lowest affordances for Value Beyond School are in the Energy Concepts and Engineering Challenge 1 units, both scoring low to no affordance. Removing the lesson segments that were dominated by teacher discourse doesn’t change any of these comparisons.
Most of the authenticity components for the Immune System and Flu Vaccines, Human Characteristics Data Analysis, and Teen Brain units score moderate, more like the Engineering Challenge 1 unit than the Energy Concepts or Wind Turbine units (Figure 5.6B). However, the Construction of Knowledge component for the Immune System and Flu Vaccines unit is closer to the moderately high level of the Wind Turbine unit.

The other exception is the Value Beyond School component, which is higher in the Teen Brain unit than any of the other units (Figure 5.6B). The Value Beyond School component for the Teen Brain PSA task is also the highest affordance for this component measured for any of the assessment tasks.

5.2.4.2 Assessment tasks

There is no clear trend in Combined Authenticity levels for assessment tasks for the selected instructional units (Figure 5.5). Most have moderately high Combined Authenticity scores. None of the units have highly authentic assessment tasks (≥ 0.80), but the RNA Fingerprinting Lab unit and Wind Turbine unit are close. They have the highest Combined Authenticities of all units.

The Energy Concepts unit has the lowest Combined Authenticity for its assessment task, as it does for the Combined Authenticity for its classroom instruction (Figure 5.5). The other four units have similar combined authenticities, from 0.53 to 0.66, clustered around the 0.6 level between moderate and moderately high authenticity.

Figure 5.6A shows that within the seven selected instructional units, the authenticity component scores frequently vary meaningfully, even for units with tasks with similar Combined Authenticity. There is wide variation in the Value Beyond School
component between the units. The variation in Construction of Knowledge and Science Meaning-Making Processes components is not quite as great, but varies from 0.3 to 1.0, with only one unit, Energy Concepts, scoring below 0.5.

Four of the seven units provide highly authentic affordances for Construction of Knowledge (Figure 5.6A). Similarly, three of the units provide highly authentic affordances for Science Meaning-Making Processes. As noted for the entire sample of assessment tasks, only one unit scores highly authentic for the Value Beyond School component, the Teen Brain unit. In addition, only one unit, the Immune System and Flu Vaccines unit, scores moderately authentic for this component, while the rest score 0.4 or less.

5.2.4.3 How do the measured levels of authenticity in classroom instruction and assessment tasks compare during selected instructional units?

Figure 5.5 shows that the differences in Combined Authenticity scores for classroom instruction between different instructional units tend to be smaller than those between assessment tasks for those units, but it is expected that one teacher’s practice will have common elements across units. The low variation noted in some of the Lesson Authenticity Indicators across the seven instructional units scored (Table 5.13) accounts for this.

For most of the seven selected instructional units, Figure 5.5 shows that the Combined Authenticity levels for classroom instruction are lower than those for their corresponding assessment task levels. The two exceptions are the Energy Concepts unit and the Immune System and Flu Vaccines unit. The Combined Authenticity levels
scored for both of these units are very similar. Because it can’t be determined that the authenticity scales are closely calibrated, it can only be concluded that for these two units the ratio of authenticity levels in the classroom instruction to those in the assessment tasks is relatively higher than for the other five instructional units. The instruments used in this study are able to show that the relationship between the combined levels of authentic affordances in the classroom instruction and assessment tasks is not the same in all units.

While the authenticity scales from each instrument are only theoretically calibrated, it is reasonable to conclude that differences of at least 0.3 between classroom instruction and assessment task component authenticity scores for the same instructional unit suggest meaningful differences in affordance, but the differences can’t be quantified.

During the winter trimester, Figure 5.6 shows that the Human Characteristics Data Analysis unit’s Construction of Knowledge component scores much lower for instruction (0.5) than for the assessment task (1.0). The difference between moderate authenticity for instruction and a high authenticity score for the assessment tasks for this unit do indicate a lower affordance in instruction than in the expectations of the assessment tasks for this component. The Science Meaning-Making Processes component for the Human Characteristics Data Analysis tasks also scores lower for instruction (0.5 vs. 0.8), suggesting that the affordance measured in instruction is less than in the assessment tasks.

At the beginning and end of the study period, Figure 5.6 shows that the Science Meaning-Making Processes component scores lower for instruction than for the
assessment tasks in both the RNA Fingerprinting Lab unit and the Wind Turbine Engineering and Testing unit. Figure 5.6 also shows that the Construction of Knowledge component score is also lower for instruction than for the assessment task for the Wind Turbine Engineering and Testing unit.

In addition, Figure 5.6 shows that the Value Beyond School component scores about 0.3 lower for classroom instruction (0.0) than assessment task (0.27) for the Engineering Challenge 1. Again, this suggests a lower affordance in the instruction than the task.

These six examples are the only components with great enough differences to conclude they measure different affordances. Although there are other infrequent examples of greater authenticity scored in instruction than the corresponding assessment task, none of these are large enough to have confidence in. At the instructional unit level of analysis, the instruments for authentic affordances in the teacher’s classroom instruction and assessment tasks are able to detect differences in both combined and component authenticities, but their sensitivity is limited.

5.3 Discussion

The measures of authentic pedagogy used in this study were able document patterns in the affordances for authentic intellectual work in the teacher’s assessment tasks and classroom instruction over the study period and between instructional units. Documenting the features of authentic pedagogy in the teacher’s assessment tasks and classroom instruction over the school year allows several questions to be addressed.
What authenticity levels characterized the teacher’s practice? Did these levels increase over the school year? What do authenticity levels reveal about the teacher’s curriculum strategies? What aspects of authenticity were emphasized in the teacher’s practice? Did the authentic aspects measured in the teacher’s day-to-day instruction align with authenticity measured in the teacher’s curriculum goals and embodied in his assessment tasks? How do these results compare with previous research?

Some of the units in each trimester were created and taught for the first time. Based on initial and final Learning Target documents and semi-structured interviews of the teacher, the teacher adjusted the curriculum in response to student success and difficulty meeting Learning Targets. Sometimes the teacher took time to reteach a skill or concept. In some cases this led to adjusting planned approaches to teaching another topic or removing a Learning Target from the curriculum. Designing new curriculum and being flexible to student needs and success required extra effort from the teacher.

The scoring guides for indicators of authenticity in the teacher’s assessment tasks (Appendix A) and classroom instruction (Appendix B) were designed with common descriptors of authenticity levels. Highly authentic for assessment task indicators is defined as a dominant or clear expectation. Similarly, classroom instruction defines highly authentic as the affordance for involvement in the authentic activity measured by each Lesson Authenticity Indicator is prominent and involved most students because it is a key teacher strategy during that lesson segment. Moderate authenticity is defined in assessment tasks as some or unclear expectations in regards to the indicator. In classroom instruction moderate indicates some affordance is present, but it wasn’t a
prominent teacher strategy. A low score means that there is no or minimal affordance in the assessment task or classroom activity.

5.3.1 Overall Authenticity Levels

The teacher’s assessment tasks and his day-to-day instruction are two aspects of students’ experience of authenticity in the classroom. The teacher’s assessment tasks analyzed were explicitly tied to the Learning Targets that were the basis for assigning student course grades. For the full set of 22 assessment tasks, the mean Combined Authenticity, and the mean Construction of Knowledge and Science Meaning-Making Processes authenticity components were moderately authentic (Table 5.9). The mean Value Beyond School component for these tasks was lower, only moderately low authenticity.

The mean authenticity levels for the entire sample of classroom instruction scored (Table 5.11) were similar to the mean authenticity levels for the full set of 22 assessment tasks. Both the mean combined and the mean Construction of Knowledge and Science Meaning-Making Processes components scored moderately authentic, and the Value Beyond School component scored moderately low.

These moderate means suggest that the teacher’s pedagogy, on average, afforded some overall authenticity and some affordance for Construction of Knowledge and Science Meaning-Making Processes (Figure 5.6). The lower means for the Value Beyond School component tell us that, on average, affordance for finding value beyond success in school was between some and none.
These authenticity scores suggest that students’ overall experience of affordances for authentic intellectual work measured by Combined Authenticity and corresponding components are at similar levels for both the assessment tasks and the instruction (Figures 5.5, 5.6). Consistently high authenticity is not present across the study period in the classroom instruction or the assessments tasks.

There is another similarity between the assessment tasks and classroom instruction authenticities regarding the Value Beyond School component scores. Tables 5.8 and 5.13 show that affordances for connections to the world beyond the classroom are frequently strong, but the moderately low mean for the Value Beyond School component is due to instruction very infrequently asking students to present their work to audiences beyond the classroom. Arranging the kinds of partnerships with adults outside the school that create meaningful opportunities for students to present their work is a considerable additional effort for teachers and schools.

The classroom instruction authenticity ratings allow calculation of the fraction of class time measured at different levels of Combined Authenticity. This gives greater detail on the distribution of student experience of authenticity than the means. Figure 5.3 shows that in lesson segments observed over the whole study period, class time that students experienced higher versus lower authenticity instruction wasn’t distributed the same for Combined Authenticity and each authenticity component.

Students’ experienced high or moderately high Combined Authenticity instruction for almost 40% the total class time observed (Figure 5.3). Differences between components of authentic instruction were found. Students experienced highly authentic
instruction for Construction of Knowledge 40% of the total time, but only 27% of time for Substantive Communication, 12% for Science Meaning-Making Processes, and 1% for Value Beyond School. Figure 5.3 shows that a reason less class time is scored at high authenticity for the Science Meaning-Making Processes component is that LAI 11, Science Inquiry and Engineering Design Practices, scores low or moderately low 75% of the time. Instructional units that focused on active science inquiry or engineering design weren’t common before the engineering units at the end of spring trimester when moderately authentic scores predominated.

5.3.2 Authenticity trends over the school year

There was no clear trend in the Combined Authenticity levels of all 22 assessment tasks scored from across the study period, but there was a lot of variation between tasks in their Combined Authenticity (Figure 5.1) and their formats (Table 5.2). There is also great variation in assessment tasks’ authenticity component scores (Figure 5.2). Each trimester had at least two low to moderately low authenticity quizzes and a project task at the end of the trimester with moderate to moderately high Combined Authenticity. Within the fall and winter trimesters other moderately high authenticity tasks are interspersed.14

The examination of the Learning Targets confirmed that the teacher made use of short answer format quizzes throughout the year to assess students’ understanding of basic science concepts. These quizzes frequently afforded moderately authentic Construction of Knowledge, but had lower expectations for Science Meaning-Making
Processes and Value Beyond School (Figure 5.2). Students were expected to apply many of the concepts assessed by these quizzes in the other moderate and moderately high authenticity tasks.

The project tasks at the end of each trimester are also key elements in the teacher’s curricular strategy. They integrate science or engineering skills, concepts, and practices from throughout the trimester, and connect strongly to the compelling topic for the trimester. Students were expected to create a significant artifact, such as a public service announcement or a functioning design that demonstrates their ability to communicate. Students were given substantial choice regarding the form of the final product.

There is an increase in the end-of-trimester unit combined assessment task authenticities across the year (Figure 5.5). However, the difference between the Immune System and Flu Vaccines PSAs from fall trimester and the Teen Brain PSA from winter is entirely due to the difference in the Value Beyond School component because the Teen Brain PSA is the only one to include some presentation to an audience beyond the classroom (Figure 5.6). The Wind Turbine Engineering and Testing unit in spring trimester has a higher combined task authenticity because it has highly authentic expectations for both the Construction of Knowledge and Science Meaning-Making Processes components (Figure 5.6), including a higher score for LAI 11, Science Inquiry and Engineering Design Practices (Table 5.13).

The samples of lesson segments scored for authenticity of classroom instruction from these end-of-trimester units were extensive and included the major learning
activities during the instructional units (Table 5.5). The Combined Authenticity and
Construction of Knowledge and Science Meaning-Making Processes component scores
for classroom instruction are greater for the Wind Turbine units then for the PSA units, as
they are for the assessment tasks (Figures 5.5, 5.6). However, the difference between the
Construction of Knowledge score for the Wind Turbine unit and the Immune System and
Flu Vaccines unit isn’t great enough to be confident the affordance is stronger at the end
of the year (Figure 5.6).

The Wind Turbine unit gives students an extended period for building and
optimizing an engineering design. Students were observed problem-solving more often
in the Wind Turbine unit than in the Immune System and Flu Vaccines and Teen Brain
units. In those two units from the fall and winter students spent a greater proportion of
their time assembling information to construct explanations. This difference is reflected
in the higher Construction of Knowledge and Science Meaning-Making Processes
component scores in the Wind Turbine unit. In addition, the greater Substantive
Conversation score for Wind Turbine unit indicates that on average students are engaged
in conversations focused on higher-order thinking and disciplinary processes to a greater
extent than in the Immune System and Flu Vaccines unit.

5.3.3 The teacher’s curricular strategies in the end-of-trimester project units

These authenticity scores for the end-of-trimester project assessment tasks and
units reflect differences in the teacher’s curricular strategies, between fall and winter
trimesters and spring. The scores for all five TAs for the Construction of Knowledge
and Science Meaning-Making Processes components for the Immune System and Flu Vaccines PSA tasks and the Teen Brain PSA task indicate only some affordance for authentic expectations in almost every case (Table 5.8).

The Wind Turbine Engineering and Testing project assessment task, by contrast, scores highly authentic or close on TAIs 1-5, accounting for most of the higher combined task authenticity. Different classroom instruction authenticity indicators account for the similarly higher scores for the Wind Turbine unit than the PSA units. Table 5.13 shows that for the Wind Turbine unit, Revisions and tinkering (LAI 3) and Science Inquiry & Engineering Design Practices (LAI 11) are most responsible for the higher Construction of Knowledge and Science Meaning-Making Processes scores, respectively. LAI 3 scores higher in the Wind Turbine unit because students experience cycles of creating, testing, and revising their engineering designs. Students are practicing the engineering skill of optimization. LAI 3 was also an important contributor to the moderately high authenticity of the unit’s Construction of Knowledge score for instruction.

This difference aligns with the teacher’s choice to focus on active involvement in engineering design and designing an experiment to test the design in the Wind Turbine unit. The much higher LAI 11 score for the Wind Turbine unit than the PSA units is the primary result of choosing this strategy. The PSA units score only moderate or moderately low authenticity for the Science Meaning-Making Processes component of authenticity (Figure 5.4).

The teacher’s different strategies during the three project units at the end of each trimester were based partly on the designation of the first two trimesters as a biology
course preparing students for a focus on health careers and the spring trimester as a physics course preparing students to focus on engineering. The compelling topics chosen for the fall and winter trimesters, deciding whether to get a flu shot, and biological systems as a foundation for human uniqueness, respectively, were not seen as amenable to designing units focused on active inquiry on natural phenomena or a design. The time constraints a full-time teacher experiences played a role. The physics course offered a more easily created opportunity for an extended experience of designing and optimizing a mechanical system. The teacher also reported “it is hard to motivate them this time of year.”

The teacher also implemented a sequence of instructional units based on Learning Targets to develop skills over time for success in a more open inquiry task. The teacher included a unit in the winter trimester focused on preparing students for organizing and interpreting data using scientific standards, the Human Characteristics Data Analysis unit. In the Wind Turbine unit students were expected to be successful combining data display and analysis skills with a controlled experimental test of their own design and interpretation of their findings. Experimental design skills were first practiced in the Engineering Challenge 1 unit.

5.3.4 Prominent features of authenticity in the teacher’s practice

The individual indicators in the assessment task and classroom instruction instruments give additional detail about specific aspects of authenticity the teacher emphasized in his expectations for student performance and his day-to-day classroom
instruction. Tables 5.8 and 5.13 present the indicator scores by task and instructional unit, respectively.\textsuperscript{16}

None of the individual assessment task or classroom instruction indicators averaged highly authentic. The highest scoring task and instruction indicators are contributors to the Science Meaning-Making Processes component. Both TAI 3, affordance to show understanding of Core science concepts, and LAI 2, sustained focus on Core science concepts, have among the highest means and occasionally score highly authentic. Some to strong affordance for use of fundamental science and engineering concepts was a consistent focus of the teacher’s pedagogy.

TAI 5, Elaborated scientific communication, another contributor to the Science Meaning-Making Processes component, was a frequent focus of the assessment tasks. Thirteen of the twenty-two assessment tasks scored moderately high or high authenticity for elaborated communication, meaning they required a scientific or engineering argument and some appropriate support for it.

In spite of a moderately high mean score for LAI 2, the mean Science Meaning-Making Processes component score for classroom instruction is moderately authentic (Figure 5.4). An important reason is that LAI 11, Science Inquiry & Engineering Design Practices, averages moderately low. However, in the four units the teacher chose to focus on active science inquiry or engineering design practices (Table 5.13), the mean is 0.51, moderate authenticity.

These moderate scores for LAI 11 indicates that in the units with a focus on inquiry and design practices, there was a mix of simple and authentic inquiry practices
dominated by simple inquiry. Science inquiry and engineering design activities were only observed in about a quarter of the total time observed. The authentic practices observed outnumbered the simple practices observed in less than 10% of this class time. This suggests teacher’s instruction favored simple over authentic inquiry and design practices.

Both TAI 1 and LAI 1 measure affordance to use higher-order thinking in a science or engineering problem solving context. They are key contributors to Construction of Knowledge component of authenticity and each have among the highest means of all indicators. TAI 1 is scored highly authentic for more than one third of the tasks. LAI 1 scores moderate to moderately high authenticity for classroom instruction in 5 of 7 instructional units. Most of the teacher’s pedagogy included at least some affordance for thinking consistent with constructing rather than reproducing knowledge.

LAI 4, interactions with Multiple representations, is another of the four highest averaging Lesson Authenticity Indicators. Representations of data, models, and designs are important tools used by scientists and engineers (McGinn & Roth, 1999). LAI 4, scored moderate authenticity in most tasks (Table 5.13). These scores indicate much of the interaction with representations, while common, is passive. LAI 4 scores moderately high authenticity in three tasks, indicating students are interpreting or creating representations some of the time.

TAI 6 and LAI 8, both measuring affordances for students to make connections to the world beyond the classroom are common across tasks and in classroom instruction. However, the rarity of affordances for students to present their learning products to
audiences beyond the teacher theoretically reduces learning opportunities and making the motivational potential of these affordances lower than those the authenticity model proposes.

To summarize, the studied teacher’s authentic pedagogy profile includes emphases on higher-order thinking applied to constructing scientific knowledge and building understanding through focusing instruction on a limited number of fundamental science concepts and skills. The teacher frequently integrates and has students create scientific representations. The teacher also frequently asks students to write scientific explanations and arguments, including support for their conclusions. In addition, with the exception of some assessment tasks and instructional units focused on basic science background, the teacher’s pedagogy includes frequent clear connections to the world beyond the classroom.

At least two of these emphases are unusual for high school science teachers. Most high school science teachers don’t frequently ask students to create written explanations and arguments, eschewing the extra grading time (Banilower et al., 2013). In addition, in this case students were encouraged to revise their writing if it didn’t meet the proficiency standard for the grade they aspired to.

The frequent clear connections to the world beyond the classroom also distinguishes pedagogy in this classroom. In a typical high school science curriculum, such connections are less integrated into the focus of study, being relegated to infrequent asides or perhaps a special unit, if they are considered at all. These are the two largest
differences between the pedagogy in the case studied and typical high school science classrooms.

Also notable in the teacher’s practice is the variety the teacher builds into the task contexts and the products students create, as shown in Table 5.2. The quizzes mentioned have somewhat similar formats, with short answer and open response items. The other tasks show great variety from long-form persuasive essays, posters using diagrams to explain biological models, public service announcements in electronic or pamphlet form, graphical displays of data with written interpretation, to engineering notebooks.

In many cases students were given a choice of artifact formats for a particular task. For example, the Immune System and Flu Vaccines PSA task instructions suggested “Project Options” including a “cartoon sequence”, a “brochure”, a “video”, or “your own idea”. This approach gives students choices, within which they still organize information in scientific ways (TAI 1) and express understanding of important concepts (TAI 3) by presenting arguments with appropriate scientific support (TAI 5).

Variety may provide a benefit in its own right. However, the teacher was also providing students with choices to fulfill their need for autonomy (E. A. Skinner, 1995) and a diversity of ways to demonstrate proficiency on Learning Targets. This is coherent with current thinking that “exemplary assessment practice recognizes that there are multiple ways in which students might express their developing understanding” (National Research Council, 2012, p. 287).
5.3.5 Do the affordances in the teacher’s assessment tasks align with those in the teacher’s classroom instruction?

Because this study’s classroom instruction authenticity scores are determined differently than the assessment task scores, confidence the two measures’ scales are calibrated the same is limited. Therefore, a conservative approach concludes that the differences between authentic affordances in instruction and assessment tasks for the same unit are only detected for 30% of the components measured in the seven instructional units reliably scored for both authentic pedagogy measures. All of these differences suggest there are lower affordances for authentic intellectual work measured in the classroom instruction than in the assessment tasks. These differences are measured for:

- Engineering Challenge 1: Value Beyond School component.

Analysis of the differences between the combined measures of authentic affordances in the teacher’s classroom instruction and assessment tasks shows smaller differences than seen in the components of authenticity. It is concluded that for the Immune System and Flu Vaccines and Energy Concepts units the ratio of Combined
Authenticity levels in the classroom instruction to those in the assessment tasks is relatively higher than for the other five instructional units.

Using this conservative approach, it is reasonable to conclude the instruments are capable of detecting differences at the instructional unit level for both combined and component authenticities. Whether there is evidence those differences influence student performance will be discussed in later chapters.

This study’s classroom observation protocol also recorded student behavioral engagement and self-regulated learning in the lesson segments observed during the seven selected instructional units. The measurements of both aspects of authentic pedagogy from this chapter are used to answer correlation study Research Question 3 in Chapter 7. In addition, analysis of the authenticity levels in student work samples from four of these seven instructional units are used to address Research Questions 2 and 3 in Chapters 6 and 7.

5.3.6 Comparison with previous research

The model of authentic pedagogy adapted for science education in this research was developed by the Center on Organization and Restructuring of Schools (CORS). The CORS School Restructuring Study investigated authentic intellectual work in a range of public schools engaged in substantial school restructuring that emphasized at least some elements in the Newmann model of student engagement in academic work in secondary schools (Newmann & Associates, 1996; Newmann et al., 1992). The School Restructuring Study used the CORS model of authentic pedagogy to measure the
The SRS also found great variation among teachers, including within schools. The SRS high school teachers with the highest authenticity ratings received a normalized score of about 0.7 for pooled Combined Authenticity of assessment tasks and classroom instruction. The SRS characterizes their results as confirming earlier research on the “enormous difficulties” (Marks et al., 1996, p. 49) in implementing their vision of authentic pedagogy.

The mean Combined Authenticity of all classroom instruction observations in this study is $0.46 \pm 0.25$ ($SD$) (Table 5.11). The mean Combined Authenticity of all assessment tasks is 0.48 (Table 5.9). These scores are from a larger number of instructional units per classroom than in the School Restructuring Study. The SRS scores are based on observations of instruction from four or five classroom visits of each teacher and from two assessment tasks chosen by each teacher from the eight high schools. This study’s measure of authenticity in assessment tasks is closely based on the CORS instrument used in the SRS. It is reasonable to believe the measures for the assessment tasks in both studies have similarly calibrated authenticity scales. However, it can’t be
concluded the classroom instruction authenticity measure’s scale is calibrated the same as the scale of the measure used in the SRS.

The assessment task and classroom instruction means for the teacher in this study are at the midpoint of the Combined Authenticity scales in this study. Each of these are greater than the result for the mean pooled task and instruction measures in the SRS, which is below the midpoint of their scale. In this study it was observed that nearly all the lesson segments dominated by teacher discourse were the teacher explaining task directions. Comparison of the current study’s scoring guides with those used in the SRS indicates that these lesson segments likely negatively influenced the mean classroom instruction scores reported by this study to a greater extent than they would have if the SRS scoring guides (Newmann et al., 1995) had been used. Therefore, it is reasonable to consider how the classroom instruction measure without the teacher dominated lesson segments compares with the SRS results.

Table 5.11 shows that when the 30% of class time dominated by teacher discourse in this study is removed, the mean Combined Authenticity for all classroom instruction observations is 0.58. This is closer to the high end (0.7) of teachers in the SRS. Considering these differences in the samples and instruments used, it is reasonable to conclude that the authenticity observed in the pedagogy of the teacher in this study compares favorably with some of the higher scoring teachers in the CORS study.

One study that used the CORS framework reports on Value Beyond School separately from the other components of authenticity. King, Schroeder, and Chawszczewski (2001) report that assessment tasks that include connections to the world
outside the classroom are uncommon. Only one of the 16 teachers in their study included moderately authentic connection to an issue or experience students could recognize in their lives. Again, the teacher in this case compares favorably with teachers in their study.
### Tables

#### Table 5.1 Variables overview for Research Question 1.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Data Sources</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affordances for authentic intellectual work in teacher-designed assessment tasks</td>
<td>Twenty-two selected teacher assessment tasks from 16 instructional units Teacher Learning Targets and rubrics provided to students for 16 units Teacher semi-structured interviews</td>
<td>Teacher Assessment Task Authenticity Indicators scoring guides in Appendix A</td>
</tr>
<tr>
<td>Components of authenticity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for Construction of Knowledge score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for Use of Science Meaning-Making Processes score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances to Find Value Beyond School score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined Authenticity score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for authentic intellectual work in the teacher’s classroom instruction</td>
<td>Classroom instruction field notes from observations from October to June Lesson Authenticity Indicator scoring guides in Appendix B</td>
<td></td>
</tr>
<tr>
<td>Components of authenticity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for Construction of Knowledge score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for Use of Science Meaning-Making Processes score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances to Find Value Beyond School score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for Substantive Conversation score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined Authenticity score</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.2 Teacher proficiency assessment tasks scored for authenticity. The tasks are listed in chronological order. Tasks noted as alternatives are an option or an additional chance for students to show proficiency on the same Learning Target or Targets (LTs).

<table>
<thead>
<tr>
<th>LTs</th>
<th>Assessment Task</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall Trimester - Biology 1. Compelling Topic – Should I get a flu shot?</td>
<td>1.1 &amp; 1.7 Antiviral Drug Design Propose a drug that interrupts viral infection.</td>
<td>Small group poster &amp; presentation to peers with elaborated theoretical support.</td>
</tr>
<tr>
<td></td>
<td>1.1 Viral Structure &amp; Replication Quiz Use diagrams to describe viral structure &amp; replication within a cell.</td>
<td>Mixture of short answer and open response items.</td>
</tr>
<tr>
<td></td>
<td>1.2 Replication, Transcription &amp; Translation Quiz Use diagrams to describe &amp; identify DNA replication, translation &amp; transcription within cells.</td>
<td>Mixture of short answer and open response items.</td>
</tr>
<tr>
<td></td>
<td>1.3 &amp; 1.7 Viral Evolution Essays Use mechanisms of genetic change to account for flu vaccine design &amp; emergence of new human flu strains.</td>
<td>Long-form persuasive essays integrating theoretical support.</td>
</tr>
<tr>
<td></td>
<td>1.3 Viral Evolution Quiz Use words &amp; diagrams to describe the mechanisms of genetic change that lead to viral evolution.</td>
<td>Mixture of short answer and open response items.</td>
</tr>
<tr>
<td></td>
<td>1.6 RNA Fingerprinting Lab Analysis Describe how electrophoresis produces a RNA fingerprint &amp; use molecular mechanisms to interpret lab results.</td>
<td>Open response items including written presentation of data &amp; interpretation requiring elaborated theoretical support.</td>
</tr>
<tr>
<td></td>
<td>1.6 RNA Fingerprinting Lab Quiz (alternative task) Describe how electrophoresis produces a RNA fingerprint &amp; use molecular mechanisms to interpret a contrived fingerprint.</td>
<td>Mixture of short answer and open response items including data interpretation and theoretical explanation.</td>
</tr>
<tr>
<td></td>
<td>1.4 Immune System Public Service Announcement Describe how the chemical &amp; cellular components of the immune system work together to fight the influenza virus.</td>
<td>Choice of creative PSA formats, such as, poster or cartoon sequence. Mostly description with limited theoretical explanation.</td>
</tr>
<tr>
<td></td>
<td>1.5 Flu Vaccines Public Service Announcement Describe how the influenza vaccine works &amp; evaluate the pros &amp; cons of getting an influenza vaccine.</td>
<td>Mixture of short answer items and open response items.</td>
</tr>
<tr>
<td></td>
<td>1.4 Immune System Quiz (alternative task) Describe how the chemical &amp; cellular components of the immune system work together to fight the influenza virus.</td>
<td>Mixture of short answer items and open response items.</td>
</tr>
<tr>
<td></td>
<td>1.5 Flu Vaccines Quiz (alternative task) Describe how the influenza vaccine works &amp; evaluate the pros &amp; cons of getting an influenza vaccine.</td>
<td>Mixture of short answer items and open response items.</td>
</tr>
<tr>
<td>LTs</td>
<td>Assessment Task</td>
<td>Format</td>
</tr>
<tr>
<td>-----</td>
<td>-----------------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td><strong>Winter Trimester - Biology 2. Who am I? How does biology contribute to understanding human identity?</strong></td>
<td></td>
</tr>
<tr>
<td>2.6 &amp; 2.7</td>
<td>Human Characteristics Data Analysis tasks</td>
<td>Written presentation of data and descriptive and graphical analysis of trends.</td>
</tr>
<tr>
<td>2.7</td>
<td>Represent in tables and graphs, and interpret collected and provided data on human physical &amp; behavioral characteristics.</td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>Genes to Traits Quiz</td>
<td>Open response items.</td>
</tr>
<tr>
<td></td>
<td>Describe how human genes are related to human proteins and how proteins are related to human phenotypes.</td>
<td></td>
</tr>
<tr>
<td>2.2 &amp; 2.7</td>
<td>Genetic Disease Project (alternative task for LT 1.7)</td>
<td>Report in form of a poster; and oral presentation to older students in medical careers track.</td>
</tr>
<tr>
<td>1.7</td>
<td>Describe how a genetic mutation results in a human disease including changes in proteins and their relationship to symptoms. Describe treatment and prevention strategies.</td>
<td></td>
</tr>
<tr>
<td>2.3</td>
<td>Neuron Function Quiz</td>
<td>Mixture of multiple choice and open response items including limited application of a biological model.</td>
</tr>
<tr>
<td></td>
<td>Describe the structure &amp; function of neurons, the process of neurotransmission and neural pathways.</td>
<td></td>
</tr>
<tr>
<td>2.6 &amp; 2.7</td>
<td>Memory Experiment Data Analysis (alternative task)</td>
<td>Written presentation of data &amp; descriptive &amp; graphical analysis of trends with limited application to theory.</td>
</tr>
<tr>
<td>2.7</td>
<td>Represent and interpret data on the recency/primacy effect collected on the class members.</td>
<td></td>
</tr>
<tr>
<td>2.4 &amp; 2.8</td>
<td>Teen Brain Public Service Announcement</td>
<td>Choice of creative PSA formats. Includes personal narrative of teen brain uniqueness.</td>
</tr>
<tr>
<td>2.8</td>
<td>Describe unique aspects of the neurology of the teenage brain.</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Spring Trimester - Physics 1. How can engineering supply our future energy needs?</strong></td>
<td></td>
</tr>
<tr>
<td>3.1 &amp; 3.7</td>
<td>Simple Machines Lab Analysis</td>
<td>Specified tabular and graphical display of data and short answer items.</td>
</tr>
<tr>
<td>3.7</td>
<td>Make prescribed measurements of levers &amp; pulley systems to illustrate factors determining mechanical advantage.</td>
<td></td>
</tr>
<tr>
<td>3.1</td>
<td>Mechanics Concepts Quiz (alternative task)</td>
<td>Mixture of short answer and open response items requiring some written conceptual support.</td>
</tr>
<tr>
<td></td>
<td>Use Newton’s laws of motion &amp; concepts of friction, work &amp; power to do simple calculations and explain motion and simple machines.</td>
<td></td>
</tr>
<tr>
<td>3.2</td>
<td>Energy Concepts Quiz</td>
<td>Mixture of short answer and open response items requiring some written conceptual support.</td>
</tr>
<tr>
<td></td>
<td>Identify energy transformations in familiar situations and use these to illustrate conservation of energy and entropy.</td>
<td></td>
</tr>
<tr>
<td>3.5, 3.6 &amp; 3.8</td>
<td>Engineering Challenge 1 (alternative task)</td>
<td>Engineering design notes and open response items including testing data record &amp; explanation of design performance. Oral presentation.</td>
</tr>
<tr>
<td>3.6</td>
<td>Use an engineering design process to develop &amp; test a car made from commercial interlocking toy bricks to illustrate a mechanics or energy concept.</td>
<td></td>
</tr>
<tr>
<td>LTs</td>
<td>Assessment Task</td>
<td>Format</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------</td>
</tr>
<tr>
<td>3.5, 3.6, 3.7 &amp; 3.8</td>
<td>Wind Turbine Engineering and Testing Use an engineering design process to brainstorm, prototype, optimize &amp; test a simple wind turbine from available materials.</td>
<td>Written engineering design notes; and open response items including data presentation and evaluation of design performance.</td>
</tr>
</tbody>
</table>

- Short answer items: Responses requested can be completed with one or two sentences, or sentence fragments, or by providing labels on a diagram.
- Open response items: longer form answers requested, such as, 3-5 sentence explanations or creating representations of laboratory data or designs.
- This alternative was available at end-of-trimester for students that didn’t complete the original task, or were unsuccessful demonstrating proficiency.
- Students chose between the PSA and Quiz versions of the tasks and completed them during the same class days.
- These two Learning Targets were complementary and were assessed together; one task for each Learning Target.
Table 5.3 Elements of authentic science pedagogy measured in the teacher’s assessment tasks.

<table>
<thead>
<tr>
<th>Component of Authentic Science Pedagogy (range)</th>
<th>Task Authenticity Indicator (TAI)</th>
<th>Description of Authentic Pedagogy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reasoning practices promote Construction of Knowledge (2-6)</td>
<td>Organization of scientific information (TAI 1)</td>
<td>The task asks students to organize, synthesize, interpret, explain, or evaluate complex information in addressing a concept, problem, or issue, dealing with describing or explaining a natural phenomenon or engineering design.</td>
</tr>
<tr>
<td></td>
<td>Consideration of alternatives (TAI 2)</td>
<td>The task asks students to consider alternative solutions, strategies, explanations, or points of view as they address a concept, problem, or issue.</td>
</tr>
<tr>
<td>Use of Science Meaning-Making Processes (3-10)</td>
<td>Science or engineering content (TAI 3)</td>
<td>The task asks students to show understanding and/or use of ideas, theories, or perspectives considered central to science or engineering.</td>
</tr>
<tr>
<td></td>
<td>Science or engineering process (TAI 4)</td>
<td>The task asks students to use methods of inquiry, research, or communication characteristic of science or engineering.</td>
</tr>
<tr>
<td></td>
<td>Elaborated communication (TAI 5)</td>
<td>The task asks students to elaborate on their understanding, explanations, or conclusions through extended writing, an oral presentation, a technology presentation, or a combination of these.</td>
</tr>
<tr>
<td>Affordances to find Value Beyond School (2-7)</td>
<td>Connection to the world beyond the classroom (TAI 6)</td>
<td>The task asks students to address a concept, problem, or issue that is similar to one that they have encountered, or are likely to encounter, in life beyond the classroom.</td>
</tr>
<tr>
<td></td>
<td>Audience beyond the school (TAI 7)</td>
<td>The task asks students to communicate their knowledge, present a product or performance, or take some action for an audience beyond the teacher, classroom, or school building.</td>
</tr>
</tbody>
</table>

**Note.** The range for Task authenticity indicators 1-4 and 6 is 1-3, corresponding to low, moderate and high authenticity. The range of Task authenticity indicators 5 and 7 is 1-4, corresponding to low, moderately low, moderately high, and high authenticity.
Table 5.4 Instructional units observed over the study period from October to June. All the instructional units observed and scored for authenticity of classroom instruction at least one day are included. The units are listed in chronological order.

<table>
<thead>
<tr>
<th>Learning Target(s)</th>
<th>Instructional Unit</th>
<th>Observation frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fall Trimester - Biology 1. Should I get flu shot?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.1</td>
<td>Viral Structure &amp; Replication</td>
<td>One class period.</td>
</tr>
<tr>
<td>1.2</td>
<td>Replication, Transcription &amp; Translation</td>
<td>Two class days (of five).</td>
</tr>
<tr>
<td>1.3</td>
<td>Viral Evolution</td>
<td>Three class days (of eight).</td>
</tr>
<tr>
<td>1.6</td>
<td>RNA Fingerprinting Lab</td>
<td>Three class days (of five).</td>
</tr>
<tr>
<td>1.4 &amp; 1.5</td>
<td>The Immune System &amp; Flu Vaccines</td>
<td>Eight class days (of ten).</td>
</tr>
<tr>
<td><strong>Winter Trimester - Biology 2. Who am I? How does biology contribute to understanding human identity?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.1</td>
<td>Genes to Traits</td>
<td>One class day (of six).</td>
</tr>
<tr>
<td>2.2</td>
<td>Human Genetic Diseases</td>
<td>One class day (of five).</td>
</tr>
<tr>
<td>2.3</td>
<td>Neuron Function</td>
<td>One class day (of seven).</td>
</tr>
<tr>
<td>2.6 &amp; 2.7</td>
<td>Memory Experiment Data Analysis</td>
<td>One &amp; one-half days (of two).</td>
</tr>
<tr>
<td>2.4 &amp; 2.8</td>
<td>Teen Brain</td>
<td>Ten class days (of 14).</td>
</tr>
<tr>
<td><strong>Spring Trimester - Physics 1. How can engineering supply our future energy needs?</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.1 &amp; 3.7</td>
<td>Mechanics Concepts and Simple Machines</td>
<td>Six class days (of 13).</td>
</tr>
<tr>
<td>3.2</td>
<td>Energy Concepts</td>
<td>Five class days (of six).</td>
</tr>
<tr>
<td>3.5, 3.6 &amp; 3.8</td>
<td>Engineering Challenge 1</td>
<td>Four class days (of five).</td>
</tr>
<tr>
<td>3.4</td>
<td>Wind Power</td>
<td>Two class days (of five).</td>
</tr>
<tr>
<td>3.5, 3.6, 3.7, &amp; 3.8</td>
<td>Wind Turbine Engineering and Testing</td>
<td>Thirteen class days (of 17).</td>
</tr>
</tbody>
</table>
Table 5.5 Instructional units observed frequently enough to compare authentic pedagogy measures of the teacher’s classroom instruction with the corresponding assessment tasks. The units are listed in chronological order.

<table>
<thead>
<tr>
<th>Learning Target(s)</th>
<th>Instructional Unit and Assessment Task(s)</th>
<th>Observation frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Fall Trimester - Biology 1</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| 1.6 | RNA Fingerprinting Lab  
RNA Fingerprinting Lab Analysis  
RNA Fingerprinting Quiz | Three class days (of five) |
| 1.4 & 1.5 | The Immune System & Flu Vaccines  
Immune System Public Service Announcement  
Flu Vaccines Public Service Announcement | Eight class days (of ten) |
| **Winter Trimester - Biology 2** | | |
| 2.6 & 2.7 | Human Characteristics Data Analysis  
Personal Data Graphing and Analysis  
Craniometry Data Analysis | Nine class days (of eleven) |
| 2.4 & 2.8 | Teen Brain | Ten class days (of 14) |
| **Spring Trimester - Physics 1** | | |
| 3.2 | Energy Concepts  
Energy Concepts Quiz | Five class days (of six) |
| 3.5, 3.6 | Engineering Challenge 1  
Engineering Challenge 1 | Four class days (of five) |
| 3.5, 3.6, 3.7, & 3.8 | Wind Turbine Engineering and Testing | Thirteen class days |
| 3.7, & 3.8 | Wind Turbine Engineering and Testing | (of 17) |
Table 5.6 Description of Lesson Authenticity Indicators (LAIs) and associated components of authentic science pedagogy scored during each lesson segment scored for authenticity of classroom instruction.

<table>
<thead>
<tr>
<th>LAI</th>
<th>Component</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Instruction affords student reasoning practices that support higher-order thinking.</td>
<td>Construction of Knowledge</td>
</tr>
<tr>
<td></td>
<td>Substantive Conversation</td>
</tr>
<tr>
<td>2. Instruction affords a sustained focus on a core science concept,</td>
<td>Science Meaning-Making</td>
</tr>
<tr>
<td>practice, or topic.</td>
<td>Processes</td>
</tr>
<tr>
<td></td>
<td>Substantive Conversation</td>
</tr>
<tr>
<td>3. Instruction affords students opportunities to try out, reflect on,</td>
<td>Construction of Knowledge</td>
</tr>
<tr>
<td>and then revise their science inquiry questions, engineering designs,</td>
<td>Science Meaning-Making</td>
</tr>
<tr>
<td>or explanations, or their problem solutions or elaborated</td>
<td>Processes</td>
</tr>
<tr>
<td>communication.</td>
<td>Substantive Conversation</td>
</tr>
<tr>
<td>4. Instruction affords interactions with a variety of representations</td>
<td>Construction of Knowledge</td>
</tr>
<tr>
<td>to illustrate or explain phenomena, data, theories, conclusions, or</td>
<td>Science Meaning-Making</td>
</tr>
<tr>
<td>communicate science to others.</td>
<td>Processes</td>
</tr>
<tr>
<td>5. There was sustained discourse among students or between</td>
<td>Construction of Knowledge</td>
</tr>
<tr>
<td>students and teacher or other adults.</td>
<td>Science Meaning-Making</td>
</tr>
<tr>
<td></td>
<td>Processes</td>
</tr>
<tr>
<td></td>
<td>Substantive Conversation</td>
</tr>
<tr>
<td>6. Instruction affords shared understanding through collaborative</td>
<td>Construction of Knowledge</td>
</tr>
<tr>
<td>discourse among peers, teacher, and/or community partners that</td>
<td>Science Meaning-Making</td>
</tr>
<tr>
<td>makes each participant’s thinking visible.</td>
<td>Processes</td>
</tr>
<tr>
<td></td>
<td>Substantive Conversation</td>
</tr>
<tr>
<td>7. Instruction models or otherwise affords “the use of metacognitive</td>
<td>Construction of Knowledge</td>
</tr>
<tr>
<td>strategies to identify, monitor, and regulate learning” (Becker,</td>
<td></td>
</tr>
<tr>
<td>2008).</td>
<td></td>
</tr>
<tr>
<td>8. Instruction affords connections between core science topics and</td>
<td>Value Beyond School</td>
</tr>
<tr>
<td>either significant public problems or personal experiences in their</td>
<td></td>
</tr>
<tr>
<td>lives outside of school.</td>
<td></td>
</tr>
<tr>
<td>9. Students are or are preparing to communicate their knowledge,</td>
<td>Value Beyond School</td>
</tr>
<tr>
<td>investigate a problem, or take some action to influence or assist</td>
<td></td>
</tr>
<tr>
<td>others beyond the teacher and classroom.</td>
<td></td>
</tr>
<tr>
<td>10. By limiting teacher intervention, instruction affords students</td>
<td>Construction of Knowledge</td>
</tr>
<tr>
<td>primary responsibility for managing their learning or how they</td>
<td>Science Meaning-Making</td>
</tr>
<tr>
<td>show proficiency.</td>
<td>Processes</td>
</tr>
<tr>
<td></td>
<td>Substantive Conversation</td>
</tr>
<tr>
<td>LAI</td>
<td>Component</td>
</tr>
<tr>
<td>-----</td>
<td>-----------</td>
</tr>
<tr>
<td>a. Generating Research Questions</td>
<td></td>
</tr>
<tr>
<td>Identifying a Design Opportunity</td>
<td></td>
</tr>
<tr>
<td>b. Designing a Study</td>
<td></td>
</tr>
<tr>
<td>Generating Design Solutions</td>
<td></td>
</tr>
<tr>
<td>c. Making Observations</td>
<td></td>
</tr>
<tr>
<td>Building Models and Prototypes</td>
<td></td>
</tr>
<tr>
<td>d. Explaining Results</td>
<td></td>
</tr>
<tr>
<td>Testing &amp; Evaluating Design Solutions</td>
<td></td>
</tr>
<tr>
<td>e. Developing Theories</td>
<td></td>
</tr>
<tr>
<td>Revising Design Solutions</td>
<td></td>
</tr>
<tr>
<td>f. Studying Others’ Research/Designs</td>
<td></td>
</tr>
<tr>
<td>g. Discourse During Inquiry</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* The range for indicators 1-10 is 1 to 3, representing low (1), moderate (2), and high (3) authenticity. The range for indicator 11 is 0-6. Each of the seven practices in indicator 11 is scored not afforded (0), simple inquiry afforded (1), or authentic inquiry afforded (2).
Table 5.7 Construction of the scores for the four components of authentic science pedagogy in classroom instruction from Lesson Authenticity Indicator (LAI) scores during each lesson segment.

<table>
<thead>
<tr>
<th>Component</th>
<th>Contributing LAIs (#)</th>
<th>Calculation</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction of Knowledge</td>
<td>Higher-order thinking (LAI 1)</td>
<td>LAI 1($\times 4$) + LAI 3 + LAI 4 + LAI 7 + LAI 10</td>
<td>8-24</td>
</tr>
<tr>
<td></td>
<td>Revisions and tinkering (LAI 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multiple representations (LAI 4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Metacognitive facilitation (LAI 7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student responsibility (LAI 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Science Meaning-Making</td>
<td>Core concepts (LAI 2)</td>
<td>LAI 2($\times 4$) + LAI 3 + LAI 4 + LAI 11</td>
<td>6-24</td>
</tr>
<tr>
<td>Processes</td>
<td>Revisions and tinkering (LAI 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Multiple representations (LAI 4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Science inquiry &amp; engineering design practices (LAI 11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Substantive Conversation</td>
<td>Higher-order thinking (LAI 1)</td>
<td>First, (Minimum of LAI 5 &amp; LAI 6) * (Mean of LAI 5, LAI 6, LAI 10) * LAI 1 is determined; then transformed to a range of 4-12. This is repeated with LAI 2 &amp; the sum is used.</td>
<td>8-24</td>
</tr>
<tr>
<td></td>
<td>Core concepts (LAI 2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sustained discourse (LAI 5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Collaborative discourse (LAI 6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Student responsibility (LAI 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Value Beyond School</td>
<td>Connections to the world beyond classroom (LAI 8)</td>
<td>(LAI 8 + LAI 9) * 4</td>
<td>8-24</td>
</tr>
<tr>
<td></td>
<td>Audience beyond the classroom (LAI 9)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 5.8 Task Authenticity Indicator (TAI) scores for selected teacher assessment tasks from across the entire study period. Tasks are ordered by Combined Authenticity score, then chronological order during the school year.

<table>
<thead>
<tr>
<th>Learning Target(s)</th>
<th>Assessment Task</th>
<th>Construction of Knowledge</th>
<th>Science Meaning-Making Processes</th>
<th>Value Beyond School</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5, 6, 7, 8</td>
<td>Wind Turbine Engineering &amp; Testing</td>
<td>TAI 1: 3.0</td>
<td>TAI 2: 3.0</td>
<td>TAI 3: 2.7</td>
</tr>
<tr>
<td>1.6</td>
<td>RNA Fingerprinting Lab Analysis</td>
<td>TAI 1: 3.0</td>
<td>TAI 2: 2.0</td>
<td>TAI 3: 3.0</td>
</tr>
<tr>
<td>2.6, 7, 8</td>
<td>Human Characteristics Data Analysis</td>
<td>TAI 1: 3.0</td>
<td>TAI 2: 3.0</td>
<td>TAI 3: 2.5</td>
</tr>
<tr>
<td>1.5</td>
<td>Flu Vaccines Public Service Announcement</td>
<td>TAI 1: 2.0</td>
<td>TAI 2: 3.0</td>
<td>TAI 3: 2.0</td>
</tr>
<tr>
<td>2.4, 2.8</td>
<td>Teen Brain Public Service Announcement</td>
<td>TAI 1: 3.0</td>
<td>TAI 2: 1.0</td>
<td>TAI 3: 2.0</td>
</tr>
<tr>
<td>3.5, 6, 8</td>
<td>Engineering Challenge 1</td>
<td>TAI 1: 3.0</td>
<td>TAI 2: 2.0</td>
<td>TAI 3: 2.0</td>
</tr>
<tr>
<td>1.6</td>
<td>RNA Fingerprinting Quiz</td>
<td>TAI 1: 2.0</td>
<td>TAI 2: 3.0</td>
<td>TAI 3: 2.0</td>
</tr>
<tr>
<td>1.4</td>
<td>Immune System Public Service Announcement</td>
<td>TAI 1: 2.0</td>
<td>TAI 2: 1.0</td>
<td>TAI 3: 2.0</td>
</tr>
<tr>
<td>1.5</td>
<td>Flu Vaccines Quiz</td>
<td>TAI 1: 2.0</td>
<td>TAI 2: 2.0</td>
<td>TAI 3: 2.0</td>
</tr>
<tr>
<td>1.4</td>
<td>Immune System Quiz</td>
<td>TAI 1: 2.0</td>
<td>TAI 2: 1.0</td>
<td>TAI 3: 2.0</td>
</tr>
<tr>
<td>3.2</td>
<td>Energy Concepts Quiz</td>
<td>TAI 1: 2.0</td>
<td>TAI 2: 2.0</td>
<td>TAI 3: 2.0</td>
</tr>
<tr>
<td>Mean of all 22 Tasks</td>
<td>Mean normalized to a 0-1 scale</td>
<td>TAI 1: 2.3</td>
<td>TAI 2: 1.9</td>
<td>TAI 3: 2.2</td>
</tr>
</tbody>
</table>

Note. The range for ATIs 1, 2, 3, 4 and 6 is 1-3. The range for ATIs 5 and 7 is 1-4. The full set of 22 assessment task TAI scores are in Appendix C, Table C.1.
Table 5.9 Normalized component and combined task authenticity scores for selected teacher assessment tasks. Tasks are ordered by Combined Authenticity score, then chronological order during the school year.

<table>
<thead>
<tr>
<th>Learning Target(s)</th>
<th>Assessment Task</th>
<th>Construction of Knowledge</th>
<th>Science Making Processes</th>
<th>Value Beyond School</th>
<th>Combined Authenticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5, 6, 7, 8</td>
<td>Wind Turbine Engineering &amp; Testing</td>
<td>1.00</td>
<td>0.95</td>
<td>0.40</td>
<td>0.79</td>
</tr>
<tr>
<td>1.6</td>
<td>RNA Fingerprinting Lab Analysis</td>
<td>0.75</td>
<td>1.00</td>
<td>0.40</td>
<td>0.75</td>
</tr>
<tr>
<td>2.6, 7, 8</td>
<td>Human Characteristics Data Analysis tasks</td>
<td>1.00</td>
<td>0.79</td>
<td>0.20</td>
<td>0.66</td>
</tr>
<tr>
<td>1.5</td>
<td>Flu Vaccines Public Service Announcement</td>
<td>0.75</td>
<td>0.57</td>
<td>0.60</td>
<td>0.63</td>
</tr>
<tr>
<td>2.4, 2.8</td>
<td>Teen Brain Public Service Announcement</td>
<td>0.50</td>
<td>0.57</td>
<td>0.80</td>
<td>0.63</td>
</tr>
<tr>
<td>3.5, 6, 8</td>
<td>Engineering Challenge 1</td>
<td>0.75</td>
<td>0.67</td>
<td>0.27</td>
<td>0.56</td>
</tr>
<tr>
<td>1.6</td>
<td>RNA Fingerprinting Quiz</td>
<td>0.50</td>
<td>0.71</td>
<td>0.20</td>
<td>0.50</td>
</tr>
<tr>
<td>1.4</td>
<td>Immune System Public Service Announcement</td>
<td>0.25</td>
<td>0.57</td>
<td>0.40</td>
<td>0.44</td>
</tr>
<tr>
<td>1.5</td>
<td>Flu Vaccines Quiz</td>
<td>0.50</td>
<td>0.43</td>
<td>0.40</td>
<td>0.44</td>
</tr>
<tr>
<td>1.4</td>
<td>Immune System Quiz</td>
<td>0.25</td>
<td>0.43</td>
<td>0.20</td>
<td>0.31</td>
</tr>
<tr>
<td>3.2</td>
<td>Energy Concepts Quiz</td>
<td>0.50</td>
<td>0.29</td>
<td>0.00</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Mean of all Tasks

<table>
<thead>
<tr>
<th></th>
<th>Mean of all Tasks</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.56</td>
<td>0.00-1.00</td>
</tr>
<tr>
<td></td>
<td>0.57</td>
<td>0.14-1.00</td>
</tr>
<tr>
<td></td>
<td>0.29</td>
<td>0.00-0.80</td>
</tr>
<tr>
<td></td>
<td>0.48</td>
<td>0.13-0.79</td>
</tr>
</tbody>
</table>

Note. The range of each of the normalized component scores is 0-1 with 0 indicating no authenticity; 0.5 is moderate authenticity; and 1 indicates high authenticity. The Combined Authenticity scores are not the sum of the component scores because of the methodology described in Table 5.3. The full set of 22 assessment task scores are in Appendix C, Table C.2.
Table 5.10 Numbers of lesson segments (LSs) and hours of class time scored for authenticity of classroom instruction from each school trimester. All lesson segments were focused on science and/or engineering instruction.

<table>
<thead>
<tr>
<th>School Trimester</th>
<th>No. of LSs observed</th>
<th>Total hours observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall</td>
<td>106</td>
<td>38.3</td>
</tr>
<tr>
<td>Winter</td>
<td>178</td>
<td>56.2</td>
</tr>
<tr>
<td>Spring</td>
<td>216</td>
<td>74.5</td>
</tr>
<tr>
<td>Total</td>
<td>500</td>
<td>168.9</td>
</tr>
</tbody>
</table>

Table 5.11 Normalized mean authenticity of classroom instruction scores for lesson segments (LSs) observed over the entire study period.

<table>
<thead>
<tr>
<th></th>
<th>Construction of Knowledge</th>
<th>Science Meaning-Making Processes</th>
<th>Substantive Conversation</th>
<th>Value Beyond School</th>
<th>Combined Authenticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>All LSs</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>.52</td>
<td>.50</td>
<td>.49</td>
<td>.34</td>
<td>.46</td>
</tr>
<tr>
<td>Median</td>
<td>.50</td>
<td>.56</td>
<td>.56</td>
<td>.50</td>
<td>.53</td>
</tr>
<tr>
<td>No. of LSs</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
<td>500</td>
</tr>
<tr>
<td>SD</td>
<td>.34</td>
<td>.26</td>
<td>.36</td>
<td>.24</td>
<td>.25</td>
</tr>
</tbody>
</table>

|                          |                           |                                 |                          |                     |                       |
| Only LSs when teacher doesn’t dominate discourse |                           |                                 |                          |                     |                       |
| Mean                     | .68                       | .60                             | .66                      | .38                 | .58                   |
| Median                   | .81                       | .61                             | .75                      | .50                 | .60                   |
| No. of LSs               | 264                       | 264                             | 264                      | 264                 | 264                   |
| SD                       | .27                       | .19                             | .27                      | .23                 | .18                   |

Note. Means are calculated from lesson segment scores weighted by the duration of the lesson segment. All scores are normalized to a scale of 0-1. Lesson segments when the teacher doesn’t dominate the classroom discourse are those when teacher discourse is less than 80% of discourse.
Table 5.12 Combined authenticity of classroom instruction for seven selected instructional units with sufficient observations to determine a reliable estimate of authenticity. Units are listed in chronological order.

<table>
<thead>
<tr>
<th>Instructional Unit</th>
<th>Combined Authenticity</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>No. of LSs</th>
<th>Class time (hr)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall Trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA Fingerprinting Lab</td>
<td></td>
<td>.46</td>
<td>.44</td>
<td>.15</td>
<td>16</td>
<td>5.5</td>
</tr>
<tr>
<td>Immune System &amp; Flu Vaccines unit</td>
<td></td>
<td>.48</td>
<td>.58</td>
<td>.19</td>
<td>56</td>
<td>22.0</td>
</tr>
<tr>
<td>Winter Trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human Characteristics Data Analysis</td>
<td></td>
<td>.47</td>
<td>.55</td>
<td>.20</td>
<td>68</td>
<td>19.9</td>
</tr>
<tr>
<td>Teen Brain</td>
<td></td>
<td>.49</td>
<td>.52</td>
<td>.26</td>
<td>96</td>
<td>22.2</td>
</tr>
<tr>
<td>Spring Trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy Concepts</td>
<td></td>
<td>.29</td>
<td>.31</td>
<td>.17</td>
<td>45</td>
<td>9.6</td>
</tr>
<tr>
<td>Engineering Challenge 1</td>
<td></td>
<td>.41</td>
<td>.55</td>
<td>.24</td>
<td>25</td>
<td>7.1</td>
</tr>
<tr>
<td>Wind Turbine Engineering &amp; Testing</td>
<td></td>
<td>.59</td>
<td>.70</td>
<td>.24</td>
<td>85</td>
<td>36.3</td>
</tr>
<tr>
<td>Mean from total time observed</td>
<td></td>
<td>.49</td>
<td>.55</td>
<td>.25</td>
<td>364</td>
<td>122.5</td>
</tr>
<tr>
<td>Mean of all unit scores</td>
<td></td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Means are calculated from lesson segment scores weighted by the duration of the lesson segment. All scores are normalized to a scale of 0-1.
Table 5.13 Normalized means (0-1) for the eleven Lesson Authenticity Indicators used to calculate the four components of authentic classroom instruction. Instructional units are listed in chronological order.

<table>
<thead>
<tr>
<th>Instructional Unit</th>
<th>Lesson Authenticity Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LAI 1</td>
</tr>
<tr>
<td><strong>Fall Trimester</strong></td>
<td></td>
</tr>
<tr>
<td>RNA Fingerprinting Lab</td>
<td>.36</td>
</tr>
<tr>
<td>Immune System &amp; Flu Vaccines unit</td>
<td>.73</td>
</tr>
<tr>
<td><strong>Winter Trimester</strong></td>
<td></td>
</tr>
<tr>
<td>Human Characteristics Data Analysis</td>
<td>.53</td>
</tr>
<tr>
<td>Teen Brain</td>
<td>.57</td>
</tr>
<tr>
<td><strong>Spring Trimester</strong></td>
<td></td>
</tr>
<tr>
<td>Energy Concepts</td>
<td>.38</td>
</tr>
<tr>
<td>Engineering Challenge 1</td>
<td>.70</td>
</tr>
<tr>
<td>Wind Turbine Engineering &amp; Testing</td>
<td>.72</td>
</tr>
<tr>
<td><strong>Mean of unit scores</strong></td>
<td>.57</td>
</tr>
<tr>
<td><strong>Minimum unit score</strong></td>
<td>.36</td>
</tr>
<tr>
<td><strong>Maximum unit score</strong></td>
<td>.73</td>
</tr>
</tbody>
</table>

*Note.* Means are calculated from lesson segment scores weighted by the duration of the lesson segment. All scores are normalized to a scale of 0-1. The means of the unit scores give equal weight to each unit score.
Figure 5.1 Combined Authenticity scores of 22 assessment tasks in chronological order from October to June. Authenticity levels are normalized to a 0-1 scale, where 0 equals not authentic, 0.5 is moderately authentic, and 1 is highly authentic. Each assessment task name is followed by the teacher’s Learning Target(s) it assesses in parentheses. The integers before the decimal point indicate the trimester.
Figure 5.2: Scores of three components of authentic pedagogy for 22 teacher assessment tasks in chronological order from October to June. Each component is labeled with its normalized score. Authenticity levels for each component are normalized to a 0-1 scale, where 0 equals not authentic, 0.5 is moderately authentic, and 1 is highly authentic. The maximum possible score on the authenticity axis is 3. Each assessment task name is followed by the teacher's Learning Target(s) it assesses in parentheses. The integers before the decimal point indicate the trimester.
Figure 5.3 Percent class time observed at classroom instruction authenticity levels for Combined Authenticity and each component from lesson segments observed across the entire study period. Numeric labels are percents. Authenticity levels increase clockwise from 12 o’clock position. The Science Inquiry/Engineering Design scale is a sub-component of the Science Meaning-Making Processes component. Percent of class time at different authenticity levels varies between components.
Figure 5.4 Mean authenticity for the four components of classroom instruction and Combined Authenticity for the lesson segments scored during each of seven selected instructional units. Means are calculated from lesson segment scores weighted by the duration of the lesson segment. All scores are normalized to a scale of 0-1. All authenticity scores are normalized to a scale of 0-1, where 0 is no authenticity, 0.5 is moderately authentic, and 1 is highly authentic.
Figure 5.5 Comparison of Combined Authenticity of assessment tasks with mean Combined Authenticity of classroom instruction from seven instructional units observed frequently enough to reliably estimate mean authenticity of the instruction. Instructional units are in chronological order. Classroom instruction scores are means of scores calculated from lesson segments weighted by the duration of the lesson segment. All scores are normalized to a scale of 0-1.
Figure 5.6 Levels of components of authentic pedagogy in assessment tasks and classroom instruction during seven selected instructional units and corresponding assessment tasks. Classroom instruction authenticity scores are means of scores from individual lesson segments weighted by their length. Each component is labeled with its normalized score. The maximum on the Y-axis for the assessment tasks is 3.0 because 3 components are measured to construct the combined score. The maximum on the Y-axis for the classroom instruction is 4.0 because 4 components are measured.
Endnotes

1 The curriculum included both science and engineering content and both are considered in the specification of assessment task and classroom instruction scoring guides. For brevity in the text Science Meaning-Making Processes will be used, but where engineering content and skills are relevant, those are as central in the scoring as science is when it is the central focus.

2 Field notes from a total of 559 lesson segments were scored for authentic pedagogy from October to June. Fifty-nine of these lesson segments did not include instruction or activities related to science or engineering instruction. Therefore, instructional authenticity scores from these lesson segments were not included in the sample used to characterize the science and engineering instruction experienced by students. These lesson segments were 4.6% of the total class time observed.

3 For the purpose of calculating replacement values, lesson segments with similar instructional patterns were defined as those in the same category of domination by teacher talk. Lesson segments with at least 80% teacher talk were considered dominated by teacher talk. Lesson segments with at least 65% student talk were classified as dominated by student talk. The other lesson segments were categorized as mixed teacher and student talk. In most cases the correlation between the missing LAI and the domination of discourse by teacher talk category was greater than 0.5 (Kendall tau-b). All correlations were checked by inspecting the data distribution. If the correlation of discourse domination category and LAI was not strong (<0.5) or there was not a clear trend in the distribution of scores, the mean of all the LAI scores for all lesson segments devoted to the same assessment task replaced the missing value.

This definition of similar instructional patterns is reasonable because most of the missing LAI observations were the discourse indicators LAI 5 and LAI 6. Seventy-five percent of the replacements were LAI 5, sustained discourse, LAI 6, collaborative discourse, or LAI 7, metacognitive facilitation. Also, it was observed over the school year that most of the teacher’s instructional patterns were in one of the discourse domination categories. The teacher typically spent some class time giving instructions or advice, or notes for students to record presented in a lecture format. These correlated strongly with teacher domination of discourse. A second common instructional pattern was a class discussion, which was usually mixed teacher and student talk not dominated by either. Thirdly, the teacher frequently asked students to work in groups or individually, which corresponded to student domination of discourse. A more robust missing values analysis was beyond the scope of this dissertation.

4 LAI 11 measures the additional authenticity afforded when lessons involve students in using the practices listed in Table 5.6 to collect, manipulate, and analyze data about natural phenomena or designs they are actively engaged with. Each of the practices present in the lesson segment is scored as authentic (2) or simple (1) according to the scoring guides in Tables B.1 and B.2 in Appendix B. The mean of the scores on the practices present is then transformed to the 1.1-6 scale designed for this sub-score.
Three theoretically derived points were used to define the transformation needed. Because of the presence of some aspect of inquiry, the low point of one on this sub-score was defined as 1.1, slightly more than the low point on the 1-3 scale for LAIs 1-10. A mean of 2 corresponds to the high point of 3 on that scale. When there are equal numbers of simple and authentic practices scored, the presence of the practices scored authentic is defined as having greater weight than the practices scored simple and the mean of 1.5 is mapped to 2.5 on the 1-3 authenticity scale. Therefore, a nonlinear transformation was desired. The transformation, \(-1.2*(\text{mean of LAI 11 scores})^2 + 6.3*(\text{mean of LAI 11 scores}) - 1.8\), based on these three points, converts all the possible means of the practices present to a scale with the needed maximum of 6. When no practices are present the value of the Science Inquiry & Engineering Design Practices sub-score is 0.

5 The Substantive Conversation component measures conversations that develop greater understanding and shared meaning. The CORS study measures Substantive Conversation on a scale of one to five. Sustained discourse (LAI 5), Collaborative discourse (LAI 6), Student responsibility (LAI 10), Higher-order thinking (LAI 1) and sustained focus on Core science or engineering concepts (LAI 2) measure the characteristics in the CORS rubric. Separate sub-scores for Construction of Knowledge and focus on core science or engineering concepts are calculated from the LAIs and the sum makes up the final score.

The transformation, \(x = (\text{Minimum of LAI 5 and LAI 6})*(\text{Mean of LAI5, LAI 6, LAI 10})*\) (LAI 1 or LAI 2), distributes the raw scores along a continuum that maps the possible LAI combinations onto the CORS scores. The range produced is 1 (13) to 27 (33).

The minimum of LAI 5 and LAI 6 is used because both are needed for the highest level of authenticity representing sustained Collaborative conversation. The second factor includes the influence of student responsibility for managing their learning during the conversation. The final factor is the extent to which the conversation is in the context of Higher-order thinking (LAI 1) or sustained focus on a Core science or engineering concept (LAI 2). A separate sub-score is calculated for LAI 1 and LAI 2.

The desired range for each of the two sub-scores was 4-12, so the sum is 8-24 matching the scores for the other three authentic pedagogy components. This transformation best converted the scores to this range: \(-0.0184x^2 + 0.8206x + 3.1514\) rounded to the nearest 0.1. The sum of the separate sub-scores for LAI 1 and then for LAI 2 is the final component score.

6 The lesson segments observed focused on performing the lab procedure, rather than the full range of activities in this instructional unit. About two-thirds of the time observed was teacher direct instruction and the lab procedure required careful adherence to directions, rather than designing an investigation. Therefore, LAI 1, Higher-order thinking, scored relatively low, contributing to a low score for Construction of Knowledge measured for the unit. The unobserved lesson segments focused on interpretation of the laboratory probably would have scored higher for LAI 1. In contrast, because the class time observed was focused on preparing for and completing the lab activity, and included the conceptual content in the electrophoresis technique, the lesson segments observed scored very high for LAI 2, Core concepts, and relatively high for LAI 11, Science inquiry practices. These contribute to a moderately high score for Science Meaning-Making Processes. According to the teacher’s lesson planning documents, the unobserved lesson segments focused on interpretation of the laboratory
results, which mean they likely continued to focus on LAI 2. This means that the Science
Meaning-Making Processes component score measured is likely to be accurate for the whole unit.
This analysis applies to the instructional unit level of analysis. It does not apply at the lesson
segment level of analysis.

7 Work samples turned in by 52 students, about half of the total number of students were
analyzed.

8 For example, to score high on LAI 1, affordances for Higher-order thinking, three
characteristics need to be observed. These are (a) an estimate that at least 75% of the discourse is
by students; (b) student thinking is characterized by activities as complex as, explaining,
predicting, interpreting, or generalizing; and (c) the object of student thinking is a natural
phenomenon or designed system.

9 Especially during class periods when almost all of the time is spent continuing work on
completing an assessment task presented several days before, the researcher’s judgments of the
authenticity of the assessment task have some influence on assignment of some LAI scores.

10 The Teen Brain PSA assignment instructions include writing a “personal statement”
asking students to apply what they have learned about the uniqueness of the teenage brain to how
they see themselves. The Teen Brain PSA was scored 3 on TAI 7, presentation to audience
beyond the classroom but within the school, instead of 4, presentation to an audience beyond the
school for two reasons. First, based on classroom field notes, the teacher implemented this as a
statement written to an adult expressing what that adult should know about teenagers. However,
the assignment instructions and the Learning Target were explicit about writing a personal
statement, but not that it must be written to or presented to an adult outside of school. Secondly,
the Teen Brain PSA addressed two Learning Targets. This “personal statement” addressed LT 8,
“craft a written product that describes the identity of 9th grade students at school name.” Because
Learning Targets were assessed for grading purposes and the presence of multiple opportunities
to show proficiency on any Learning Target, completion of this part of the Teen Brain PSA was
optional. The score of 3 was judged to best fit the scoring guide scale.

11 There were three other assessment tasks in a quiz format that scored moderate Combined
Authenticity. The RNA Fingerprinting Quiz includes greater expectations for a longer form
written argument (TAI 5). The Viral Evolution quiz and the Flu Vaccines quiz scored moderately
authentic based on authentic connections to the world beyond the classroom (TAI 6). The four
other assessment tasks scoring moderately authentic have diverse formats and combinations of
authenticity component scores.

12 The classroom observation protocol included an estimate of the degree of teacher vs.
student domination of talk during each LS. In about 50% of the lesson segments in the 500 lesson
segments from across the entire study period it was estimated that at least 80% of all the
discourse was teacher talk. This corresponds to 30% of the total class time in the dataset. Almost
all of these lesson segments were characterized by teacher task directions and some instances of
direct instruction, for example, delivering a set of notes for students to copy.
Figure 5.3 also shows that these components score low authenticity about one-third of class time observed. This corresponds with the 30% of class time observed that the teacher dominates classroom discourse. The mean authenticities are less than 0.1 for both of these components of authenticity when the teacher dominates classroom discourse.

Figure 5.2 shows that the spring trimester has two low authenticity quizzes focused on basic physics concepts, but no moderately high authenticity task. Based on field notes and teacher interviews the teacher’s strategy here was focused on getting students to the final engineering task with plenty of class time to use to optimize and test their design to maximize understanding of engineering design skills. The end-of-trimester project task used the time that was devoted to an additional moderately high authenticity task in fall and winter trimesters.

The teacher conveyed to the investigator that students struggled to demonstrate proficiency on the tasks for this unit. The Memory Experiment Data Analysis task was another opportunity for students to develop proficiency on data display and analysis skills. The Engineering Challenge 1 task was designed as a first activity to apply these data analysis skills as part of a task that included designing a controlled test of a design.

Classroom instruction means for each Lesson Authenticity Indicator (LAI) for the entire set of lesson segments observed without grouping by instructional unit were compared with the means of the unit scores in Table 5.13. These means for each LAI rounded to the nearest 0.1 were the same and their ranks were very similar. There was only one rank difference of more than one rank. Therefore, the values in Table 5.13 are used here to describe trends in the data.
Chapter 6  Student Responses to Authentic Pedagogy over a School Year

This chapter addresses the first broad research question by documenting student responses to the teacher’s pedagogy. Three student responses to the instructional ecosystem in the classrooms studied were measured: behavioral engagement, self-regulated learning, and authenticity of performance on samples of science class work. The behavioral engagement and self-regulated learning variables measure student behavior within the classroom during instruction. These observations coincide closely with the class days measured for authentic pedagogy discussed in Chapter 5.

The authenticity of student performance on science class work variable measures academic performance using this study’s model of authentic pedagogy and performance. The characteristics of student performance measured are based on two components in the model of authenticity used to describe authentic pedagogy in Chapter 5. The student work samples chosen for analysis were from the fall and spring trimesters to see how authenticity of student work changed from October to June.

Chapter 7 will explore the relations between teacher pedagogy variables measured in Chapter 5 with the student response variables in this chapter.

Research Question 2

How do the levels of authentic student performance, engagement, and self-regulated learning vary over the school year?
6.1 Methods

To document behavioral engagement and self-regulated learning, classroom field notes from October through June were analyzed. Levels of behavioral engagement were determined for the entire study period, each trimester, and selected instructional units (Table 6.1). Levels of self-regulated learning were determined for the winter and spring trimesters and four selected instructional units (Table 6.1).

To document the authenticity of student performance in science classwork over the study period and variation in student performance between instructional units, student work samples were collected from four instructional units, two in the fall trimester and two in the spring trimester (Table 6.1). These instructional units were also analyzed for levels of authentic pedagogy and behavioral engagement.

6.1.1 Measures of behavioral engagement and self-regulation

Naturalistic observation studies set in school classrooms must be flexible and responsive to the uncontrolled local conditions in the case studied. Infrequently, interruptions required this investigator to make choices between variables field notes should capture or interfered with data collection. For example, fire drills, respecting a student’s request for assistance, or taking a phone message for the teacher might limit the number of behavioral engagement and self-regulated learning observations collected. Like the sample of lesson segments scored for authenticity of the teacher’s classroom instruction, the samples of behavioral engagement and self-regulation are convenience samples.
6.1.1.1 Data sources

Behavioral engagement and self-regulated learning observations were collected from a wide variety of classroom activities across the whole study period. Observations of behavioral engagement were made on class days during all 16 instructional units in Table 5.4. Most of the observations were during the seven instructional units chosen for intensive study and examined in Chapter 5 to answer Research Question 1. As noted in Chapter 5, classroom activities and instructional units that gave students opportunities to for self-regulated learning were prioritized for observation. Except for the observations used to establish the interrater reliability of these measurements, all observations were made by this investigator.

Behavioral engagement was scored during 399 lesson segments during the study period from October to June. A total of 18,508 discrete observations were recorded of 108 students enrolled in the classrooms studied for at least one trimester. Students that were not present during a behavioral engagement scan because they left the classroom temporarily or when it wasn’t possible to make definitive judgement were treated as missing data. The remaining 17,716 observations are used in the analysis. The field notes from 386 of these lesson segments were also scored for authenticity of the classroom instruction as documented in Chapter 5. One hundred fourteen lesson segments were scored for authenticity of the classroom instruction, but not behavioral engagement.

Self-regulated learning was scored during 184 lesson segments from October to June. This is 46% of the lesson segments scored for behavioral engagement. Behavioral engagement was also measured during all of these lesson segments. A total of 6,982
discrete observations were recorded for self-regulation. Slightly less than 3% of these observations were recorded as ambiguous or that the student was temporarily out of the classroom. These are treated as missing data leaving 6787 observations used in the analysis. The field notes from 179 of these lesson segments were also scored for authenticity of the classroom instruction.

6.1.1.2 Instrument description

Behaviorally engaged is defined as appearing attentive or involved (O. Lee & Anderson, 1993). Attentiveness was recognized as active listening, including watching the teacher, or responding to teacher questions or those of peers. Involvement was recognized by working on classroom activities or volunteering questions or comments about the lesson content. Off-task was defined as appearing to attend to something other than teacher or the classroom activity and recorded as not behaviorally engaged. The behavioral engagement variable is a binary variable that recorded its presence (behaviorally engaged) or absence (not engaged). These two values for the behavioral engagement variable are mutually exclusive and exhaustive.

Behavioral engagement was coded for each student by the investigator at 8-10 minute intervals during lesson segment observations by evaluating each student in turn for 5-15 sec to determine if behavioral engagement predominated. Full detail on observation and coding of the lesson segments selected for study of behavioral engagement and self-regulation is in Appendix D.

Self-regulated learning is defined as exercising autonomy, agency, or ownership related to classroom activities (McCombs, 2001; Zimmerman, 2001). In this study these
are operationalized as self-initiation and self-direction (Zimmerman, 2001). Self-regulated learning observations were coded into two variables. The first self-regulated learning variable is a binary variable recorded as present (evidence of self-regulated learning) or absent (no evidence of self-regulated learning) during all three trimesters. The codes for behaviorally engaged and self-regulating in these presence variables are not mutually exclusive.

To produce a richer description of self-regulation behaviors observed, a series of codes for student behaviors that operationalized a classification of self-regulated learning were developed during classroom observations of the same classrooms during the fall trimester. The codes are listed in Appendix D, Table C.1. These codes for self-regulation behavior types collected during winter and spring trimesters were recoded into two categories of self-regulated learning behaviors. The first category was students observed working independently without continuing teacher direction, but was limited to following the teacher’s directions for the classroom activity.

The second category of self-regulated learning behaviors includes observations of behaviors that go beyond independently following directions and indicate stronger evidence of autonomy and ownership. These include initiating or maintaining lesson-related activity, making lesson-related choices, and spontaneously articulating interest in or the value of the lesson content – all with no or minimal teacher prompting or direction.

The second self-regulated learning category variable recorded these two categories of self-regulated learning described above: (1) working independently on classroom activity as defined by the teacher and (2) working autonomously by taking initiative beyond teacher directions. If any of the category (2) behaviors were observed,
even if category (1) behavior was also observed, the observation was coded as category (2). These categories for the self-regulated learning category variable are not mutually exclusive or exhaustive.

This study uses observations of self-regulated learning as indirect evidence that students are cognitively engaged in science learning activities (O. Lee & Anderson, 1993). Using the definitions in this study, observations of self-regulated learning provide stronger evidence of cognitive engagement than observations of behavioral engagement alone. The working autonomously category of self-regulated learning is the strongest evidence of cognitive engagement.

Two data collection conditions for self-regulation observations were dictated by classroom instructional patterns during particular lesson segments. In the first condition, students were grouped as an entire class in a common activity led by the teacher. These lesson segments were often characterized by the teacher delivering task directions or a class discussion. Occasionally, students’ roles were to copy lecture notes or watch a video.

During such lesson segments with students grouped as whole class the only observable convincing evidences of student self-regulated learning were self-initiated oral commenting, or questioning, or answering with conceptual content; or articulating personal interest in or value of the lesson content. Gathering additional observations was too intrusive. These behaviors were coded as self-regulating into the binary self-regulated learning variable; and into the self-regulated learning category variable as category (2) as they occurred. Students not exhibiting these behaviors were scored not self-regulating in the binary self-regulation variable. This results in a conservative
estimate of self-regulated behavior for the binary variable during these lesson segments. Eighteen percent of the total binary variable observations were collected under this condition.

The second data collection condition was characterized by students working individually or in small groups to complete any of a wide variety of activities not dominated by attending to the teacher. During such lesson segments selected for observation of self-regulated learning, the presence or absence of self-regulation was coded into the presence variable and the category of self-regulation was also coded when behaviors in those two categories were observed. Appendix D details the protocol used during lesson segments with each of the two observation conditions.

Category (2) behaviors could only be observed when they were overt. Student dispositions to express themselves overtly in class influence the measurement. The protocol used likely produces a conservative estimate of these self-regulated learning behaviors.

The behavioral engagement and self-regulated learning observations for individual students for each lesson segment coded from the classroom observation protocol field notes were entered into SPSS, which was used to manage and manipulate the data to compute the percent engagement and self-regulation levels for the entire study period, trimesters, and selected instructional units. Observations of each student were encoded into data files using numeric student codes to preserve confidentiality.
6.1.1.3 Reliability and Validity

To demonstrate reliability of the scoring for behavioral engagement, the investigator and a published university science education researcher independently coded 2.5% of the behavioral engagement observations, after training in the protocol. The lesson segments were chosen based on the availability of the second coders. In addition, a science education graduate student with experience in K-12 science education and the investigator independently scored 0.2% of the behavioral engagement observations, after training. The total coded by two observers is 470. Of all of these behavioral engagement scores, 90.7% were coded the same by both scorers. A Cohen's κ was calculated indicating good agreement (Altman, 1999), κ= .678, 95% CI [.590 to .766], p < .0005.

Similarly, a sample of observations of the self-regulated learning presence variable equal to 3.9% of the total observations analyzed were independently coded by the investigator and the same published university science education researcher. An additional 0.5% were coded by the same science education graduate student and the investigator. The total coded by two observers is 302. Of all of these self-regulated learning scores, 84% were coded the same by both scorers. A Cohen's κ was calculated indicating good agreement, κ= .665, 95% CI [.579 to .751], p < .0005.

Content validity for the behavioral engagement measurement is based on the use of a definition from a classroom observation study in the science education literature (O. Lee & Anderson, 1993). The content validity for the self-regulated learning measurement is based on theoretical models (McCombs, 2001; Paris & Winograd, 2003; Zimmerman, 2001). Judgments of content validity also depend on the expertise of this investigator who has long experience teaching high school science. The observation
protocol in Appendix D, including the schedule of observations, and evidence of behavioral engagement, off-task behaviors, and self-regulated learning categories, was used consistently to ensure internal validity. The fact that all the observations were in the same teacher’s classrooms and largely the same group of students also contributed to internal validity.

6.1.2 Behavioral Engagement Data Analyses

The purpose of this chapter is to document patterns in student responses within the instructional ecosystem of the classrooms studied. To answer Research Question 2 regarding student behavioral engagement during the school year, the mean of students’ percent of observations recorded as behaviorally engaged is used to describe behavioral engagement over the entire study period, during each trimester, and during the seven instructional units selected for intensive study.³

In naturalistic observational studies, such as this one, the investigator doesn’t control which students are present during observations, or the lesson or instructional unit that would be observed. Planned data analyses were sometimes modified based on the convenience samples obtained. An exploratory approach was used to make the most fruitful data analysis decisions.

Using the percent of observations behaviorally engaged for each student at each level of analysis gives each student an equal influence on the mean behavioral engagement in the sample. It also removes some of the noise in the raw data. The same measure is used to explore the influence of student groups including class periods, gender, ethnicity and prior achievement in school to the extent possible. The criteria used
to screen students for inclusion in the analytic sample at each level of analysis are described here.

Behavioral engagement observations were collected from four class periods. Observations were focused on class periods 2, 3, and 5. Less than 4% of behavioral engagement observations were made in period 1.

6.1.2.1 The entire study period and each trimester

The criteria for inclusion of a student in measures of behavioral engagement for the entire study period is observation during a variety of lessons during at least two trimesters. Based on the mean observations per student per class period observed, it was determined that at least 32 observations makes it likely the observations for that student came from an average of 8 class days. This is judged enough days to makes it reasonable to assume that the student was observed in a variety of lessons. The same number of observations per student per trimester is used as the criteria for inclusion in analysis by individual trimester.

To investigate the role of individual student disposition to be very highly engaged independent of variations in classroom activities, students with consistently high levels of behavioral engagement were identified. First, the lesson segments during which almost all students were engaged (≥ 90%) were removed. Then students with at least 90% of observations behaviorally engaged in the remaining lesson segments during both winter and spring trimesters were identified. Only four students met this criterion for being consistently very highly engaged.
Preliminary analysis confirmed that in most cases the influence of consistently highly engaged students is not detectible in measures of mean behavioral engagement for the whole sample or divisions by time period or student group. Therefore, these students are included in the analysis.

Eighty-three students are included in the analysis at the entire study period and trimester levels of analysis. The gender breakdown of these students is 61% (N=51) females and 39% (N=32) males, very close to the gender proportions in the entire ninth grade science student population. The trimester measures and gender samples are not independent. A nonparametric test was chosen because of the presence of outliers that there was no justification to remove from the data and the variable is not normally distributed for any of the trimesters.

Fifty-three students that self-reported ethnicity and gave consent for its use are included in the analysis of the influence of ethnicity. This leaves low numbers of students (<13) in all but the Euro-American ethnic category. Forty-seven students that granted consent for use of their school records are included in the analysis of the influence of prior achievement in math and science classes. The statistical power requirements for an ANOVA analysis at all levels of analysis are not met because of the low numbers of students in most groups and the lack of independence between the groups. In addition, concerns have been noted about whether each category is represented proportionally to the larger sample. Therefore, only a description of behavioral engagement levels measured by ethnic category and prior achievement are presented.
6.1.2.2 Selected instructional units

The seven instructional units selected for sufficient observations for a reliable estimate of authentic pedagogy levels discussed in Chapter 5 were also chosen for frequent observation of their behavioral engagement levels. A total of 301 lesson segments were scored for behavioral engagement during these seven units. Between eighty and ninety percent of the lesson segments scored for authentic pedagogy during each instructional unit were also scored for behavioral engagement. Some lesson segments not scored for authentic instruction during these units are included in the analysis in this chapter.

The criteria for inclusion of a student in measures of behavioral engagement for the seven selected instructional units is at least seven observations in at least four lesson segments during at least two class days. This makes it likely students were observed in a variety of instructional contexts representative of the unit. Mean and median percent behavioral engagement per student are determined for each instructional unit and differences explored for patterns between units.

In Chapter 5, a reliable estimate of authentic pedagogy levels was predicated on observing lesson segments from all the major activities or activity types during the unit. It was noted in Chapter 5 that the sample of lesson segments scored for authenticity of classroom instruction during the RNA Fingerprinting Lab unit is much smaller than the other six selected units and doesn’t include an instructionally distinct section of the unit. Behavioral engagement was scored for almost all of these same lesson segments. The observations from the RNA Fingerprinting Lab unit only include 49 students. Therefore, the behavioral engagement sample doesn’t represent the unit or student population in
these classrooms, but it does represent the same lessons scored for authentic pedagogy.

In this chapter, the behavioral engagement level measured for the RNA Fingerprinting Lab unit is analyzed qualitatively. The other six instructional units scored for authentic pedagogy in Chapter 5 meet the requirement for a reliable estimate of behavioral engagement in the unit.

A test for significant differences in behavioral engagement levels was conducted for units that met an additional criterion, based on the convenience samples obtained. The samples of students between instructional units were not independent. Therefore, a repeated-measures analysis is used to identify significant differences in mean percent behavioral engagement per student between units. This limits the analysis to only students included in each instructional unit. The five units were included that allowed at least 80% of the 85 students in analysis of the entire study period to be in this analysis. A nonparametric test is used when the assumptions of a repeated-measures ANOVA are not met.

6.1.3 Self-Regulated Learning Data Analyses

To answer Research Question 2 regarding levels of student self-regulated learning during the school year, patterns in self-regulated learning levels were documented over the study period. Self-regulated learning observations were collected from 4 class periods across all three school trimesters. Observations were focused on class periods 2, 3, and 5. Less than 3% of self-regulated learning observations were made in period 1.

The goal of the analysis is to explore the prevalence of self-regulated learning over the entire study period by comparing the frequency of self-regulated learning
between trimesters and instructional units selected for intensive study. The presence of two types of self-regulated learning indicating different degrees of ownership are described during selected trimesters and instructional units. This allows examination of the nature of self-regulation in these classrooms. Consistency in the observation protocol allows for qualitative comparisons of these categories between selected trimesters and instructional units. In addition, the gender differences are described at the trimester and instructional unit levels.

Trimesters and instructional units are included in this analysis when the observations of the self-regulated learning variables include observations from the variety of classroom activities observed for behavioral engagement and authentic pedagogy. At each level of analysis the percent of observations indicating self-regulation from all students without aggregation was used as the measurement of the self-regulated learning presence variable. The percent category (2) behaviors was determined from the total set of individual observations of the self-regulated learning category variable at each level of analysis and used as the measure of this variable.

In order to describe how self-regulated learning varied over the school year, this study was designed to collect observations of the self-regulated learning presence variable from across the study period to describe the frequency of student self-regulation in a broad variety of classroom activities, primarily from instructional units selected for more intensive study. In addition, examples of classroom activities expected to allow for self-regulation were also chosen for categorization of the type of self-regulated learning behaviors.
One-hundred eighty-four lesson segments were selected for coding the self-regulated learning presence variable. The self-regulated learning category variable was coded during winter and spring trimesters. One-hundred forty-four lesson segments were selected for coding this category variable. The number of self-regulated learning presence variable observations collected was 38% of the behavioral engagement observations.

Across the entire study period, observations of the self-regulated learning presence variable were made in 15 of 16 instructional units observed for behavioral engagement and authenticity of classroom instruction (Table 5.4). This sample of observations from the entire study period is not equally representative of each trimester, but the gender proportions in this sample were the same as in the student population in these classrooms, to the nearest percent. Therefore, the sample is used only to compare the relative rates of self-regulation by gender over the entire study period.

Self-regulated learning presence variable observations from the winter (N = 1809) and spring (N = 4434) trimesters include a wide variety of activities, but only a small number of specific activities were observed during the fall. The percent self-regulating in the self-regulated learning presence variable during the winter and spring trimesters are compared. Because the gender proportions in the samples from each of these trimesters were the same as in the student population in these classrooms to the nearest percent, rates of self-regulated learning by gender are compared. In addition, sufficient self-regulated learning category observations were collected during winter and spring trimesters to describe the incidence of these categories from the variety of lessons from each of these trimesters.
The lesson segments scored for authentic pedagogy and behavioral engagement include observations of both of those variables that are broadly representative of the major activities during all of the seven instructional units selected for intensive study except the RNA Fingerprinting Lab unit. The self-regulated learning presence variable observations were spread throughout the different activities during the Teen Brain (N = 650), Energy Concepts (N = 287), Engineering Challenge 1 (N = 341), and Wind Turbine Engineering and Testing (N = 2504) instructional units. The rates of self-regulating vs. not self-regulating in all of the individual observations are compared for these units. The incidence of two categories of self-regulated learning behaviors are also compared between these units.

6.1.4 Measures of authenticity of student performance in science work samples

To address Research Question 2, work samples were collected from two instructional units in the first trimester and two units from the third trimester to determine if performance levels changed between October and November in the fall trimester to May and June in the spring trimester (Table 6.2).

6.1.4.1 Data Sources

The work samples collected were the artifacts students produced in response to the proficiency assessment tasks the teacher assigned them to complete as the culmination of those four instructional units. The teacher assigned proficiency grades to the work samples for each of the Learning Targets associated with the corresponding assessment tasks. These grades were used to determine students’ course grades.
A goal of scoring 25 work samples from each of the four selected instructional units was set. Several criteria determined inclusion of student work samples in the study. The first was granting of informed consent from students and their parents or guardians. The second criterion was that students were enrolled in class periods 2, 3, and 5, which were chosen for more intensive study. The third criterion was the availability of the work sample to this investigator for photocopying. Fourth, work samples from the Immune System and Flu Vaccines and Wind Turbine Engineering and Testing units were included only if there was student work in each subsection of the task.

The fifth criterion was to include a student’s work when samples from 3 or 4 of the four units was available. Finally, when sufficient work samples were available, work samples from a mixture of gender, ethnicity, and prior achievement combinations were chosen. The samples of student work collected from each of the four instructional units was a convenience sample.

A total of 96 student work samples were collected and scored for authenticity of student performance. Different numbers of work samples were collected from each of the four instructional units (Table 6.2). The lowest number of samples were scored for the RNA Fingerprinting Lab unit, 14, and they were equally divided between each to the two assessment task alternatives. The greatest number of work samples were scored for the Energy Concepts unit, 34. A very small number of work samples that were responses to the alternative tasks for the Immune System and Flu Vaccines and Energy Concepts units (Table 6.2) were scored because a small minority of students chose to complete those tasks.
6.1.4.2 Instrument Description

Student work samples are scored for the Construction of Knowledge and Science Meaning-Making Processes\textsuperscript{10} components of authentic intellectual work in the model of authentic science pedagogy and performance adopted by this study (Figure 3.1). Each work sample is scored for the three Performance Authenticity Indicators (PAIs) described in Table 6.3. The Construction of Knowledge component is scored by the Scientific analysis indicator (PAI 1) and the Science Meaning-Making Processes component is scored by two indicators, Scientific or engineering analysis (PAI 2) and Elaborated communication (PAI 3). Appendix F details the scoring guides used in this study for each of the three PAIs.

These three PAIs reflect the scoring approach used in the research of the Center for Organization and Restructuring of Schools (CORS). CORS developed scoring guides for student work samples in social studies and mathematics (Newmann et al., 1995). The CORS scoring guides and an adaptation of them for science student work samples by the Research Institute on Secondary Education Reform for Youth with Disabilities (2001) were adapted to write the scoring guides for student work samples used in this study.

Table 6.3 shows that each PAI was scored on a 1-4 scale from no to low authenticity (1) to high authenticity (4). Use of Scientific or engineering analysis (PAI 1) is scored on a scale of 1 for unsuccessful or no use of analysis to 4 for substantial evidence of analysis in most of the work. This indicator measures the degree to which disciplinary higher-order thinking skills are used to construct knowledge.

Use of Core science concepts (PAI 1) is scored on a scale of 1 for no or unsuccessful application of a fundamental science or engineering concept to 4 for...
exemplary understanding evidenced by correct application of a fundamental science or engineering concept with enough detail to provide confidence there was understanding.

The use of Elaborated communication (PAI 3) is scored from 1 for no or unsuccessful communication to 4 for easily understood accurate scientific explanations, arguments, representations, or engineering designs with convincing support in a scientific or engineering format. The scores for these three indicators are totaled for a combined authentic performance score with a range of 3-12. The Science Meaning-Making Processes component has twice the weight of the Construction of Knowledge component in the Combined Authenticity score.

The scoring guides are meant to be inclusive of the range of scientific or engineering intellectual processes of practitioners in the community, from professional scientists or engineers, to educators, to those that use science to influence public policy. The combined scores are designed to measure the degree of successful application of developmentally appropriate intellectual processes resembling those of competent adult science or engineering practitioners.

The three scores for each work sample were entered into SPSS, which was used to manage and analyze the data. Scores for each student were encoded into data files using numeric student codes to preserve confidentiality.

6.1.4.3 Scoring decisions regarding particular work samples

The proficiency assessment tasks from the four instructional units selected for collection of student work samples varied in format and other characteristics. The authenticity of these assessment tasks were analyzed in Chapter 5.
The scoring guides for the three PAIs in Appendix F are based on the presence of authentic performance within the work sample rather than the work sample as a whole. This is similar to the scoring guides for the assessment task indicators in Appendix A. Both assessment tasks and student work samples that expected artifacts with multiple parts that provided students many authentic performance opportunities are biased to score higher. Therefore, the same guidelines used in Chapter 5 for deciding whether an assessment task should be subdivided before scoring were applied to the work samples collected.

Table 6.2 describes each task. The RNA Fingerprinting Lab unit in fall trimester and the Energy Concepts unit in the spring trimester were completed in five or six class days. The assessments tasks from these units assessed a single Learning Target. Students spent one or two class days completing their work samples for these assessment tasks. The work samples students produced were one to 3 pages in length. These artifacts were student paper and pencil responses to teacher specified open response and short answer items. This was true of both task alternatives in Table 6.2 for each of these units. Therefore, the RNA Fingerprinting Lab and Energy Concepts units’ work samples did not meet the criteria for division into subsections for scoring.

The instructional units at the end of the fall and spring trimesters were much longer, from 10 to 17 days. The work samples collected from these units met the criteria for scoring in subsections as did their corresponding assessment tasks in Chapter 5.

Table 6.2 shows that students chose between a public service announcement (PSA) format task and a pencil and paper quiz format for the Immune System and Flu Vaccines unit at the end of fall trimester. Both of these assessment tasks addressed two
Learning Targets. The student work samples from this unit were divided into one subsection presenting basic information on the human immune system and a second subsection presenting information about flu vaccines and reasons for or against getting a flu vaccine. These subsections were completed concurrently in class. The subsections were scored separately producing the three indicator scores for each subsection.

Student work samples were also collected from the Wind Turbine Engineering and Testing unit at the end of the spring trimester. These work samples consisted of engineering design diagrams, notes, testing results, and teacher specified open response items. The work samples were typically 10-20 pages long. Work samples from this unit were divided into the same three subsections as this unit’s assessment task and scored separately. The scores reported for the task as a whole are the means of the scores on these three parts of the overall task.

6.1.4.4 Reliability and Validity

To demonstrate that the scoring guides for student work samples could be used reliably, this investigator, a published university science education researcher, and two retired science teachers with experience teaching science to ninth graders and training in using other student work scoring guides independently coded a total of 19-2/3 work samples chosen randomly from work samples from the Energy Concepts and each of the three subsections of the Wind Turbine Engineering and Testing units. This number of work samples is 31% of the work samples from these two units and 20% of all the work samples included in the study. This investigator paired with one of the three other scorers
for the Energy Concepts unit, another for the first two parts of the Wind Turbine Engineering and Testing unit, and the third for part three of Wind Turbine unit.

For all the work samples coded independently the percent agreement was 86%. A Cohen's $\kappa$ was calculated to determine consistency of the raters. It indicated very good agreement (Altman, 1999), $\kappa = .817, 95\% \text{ CI } [.699, .935], p < .0005$.

The limited number of disagreements between the independent scorers on the work samples used for determining inter-rater agreement were decided by the judgement of this investigator.

This study’s measure of authenticity of work samples has validity within this study. First, construct validity is based on using the same indicators and scoring levels of each as those used in the CORS scoring guide and the previous research conducted with their instrument. In many cases this study’s scoring guide was worded the same (Appendix A). Second, content validity was maintained during the adaptation for science and engineering by careful reference of science education policy documents (Katehi et al., 2009; National Research Council, 2012), the science education literature (Chinn & Malhotra, 2002; H.-S. Lee & Songer, 2003; McGinn & Roth, 1999; J. Singer et al., 2000), and consultation with engineering education experts (C. Sneider, personal communication, February 5, 2011; G. Recktenwald, personal communication, May 16, 2011). Judgments of content validity also depend on the expertise of this investigator who has long experience teaching high school science.

Content validity was guarded by analyzing each of the assessment tasks corresponding to the work samples scored for the scientific and engineering analysis skills, fundamental science and engineering concepts, and science and engineering
communications, if any, required by the task. Standards documents (American Association for the Advancement of Science, 1993; Katehi et al., 2009; National Research Council, 2012) were used to define these opportunities for student performances in each work sample. These were listed, which clarifies the scoring guides for each task. This also improves internal validity.

6.1.5 Authenticity of Student Performance Data Analyses

Research Question 2 includes documenting how the authenticity of student performance on science work samples varied between the four instructional units studied to investigate changes during the study period. In addition, differences between gender, ethnicity, and prior achievement levels within instructional units are described.

Combined authenticity of performance was calculated for each work sample by adding the scores for the three performance authentic indicators (Table 6.3). Each PAI and Combined Authenticity score is normalized to a range of 0-1 employing the same calculation used for assessment task and classroom instruction authenticity scores. This standardizes each component and combined score on a scale from 0-1, where 0 measures no authenticity, 0.5 measures moderate authenticity, and 1 measures high authenticity.

Beyond the work samples scored for inter-rater agreement, some were scored by consensus of two of the scorers and others were scored by this investigator or one of the other scorers. For the Energy Concepts unit, 13 (38%) of the work samples were scored by consensus of the two scorers. The other 16 (47%) work samples were scored by one or the other of the two scorers. For the three parts of the Wind Turbine Engineering and Testing unit, nine of these work samples were scored by consensus of the two scorers and
the rest were scored by this investigator. All of the work samples from the two instructional units selected from the fall trimester, the RNA Fingerprinting Lab and the Immune System and Flu Vaccines units, were scored by this investigator.

For the analyses in this chapter, all the student work samples from alternative assessment tasks for the same instructional unit were included in the summary statistics for that unit (Table 6.2). All work samples from alternative assessment tasks for the same instructional unit were scored in light of the same listing of core science concepts and practices addressed by their common learning targets. So, work samples from alternative assessments are treated the same in these analyses. The Immune System and Flu Vaccines and Wind Turbine Engineering and Testing units were scored in subsections. The averages of those two or three subsections are used as the combined and PAI authenticity scores for each of those units in these analyses.

6.1.5.1 Differences between instructional units

The science work samples scored for each of the four instructional units come from different groups of students and the number of students scored for each unit varies widely (Table 6.10). The work samples collected from each of the four instructional units include members from each of the groups from all three of the demographic variables considered in this study, gender, ethnicity and mean 8th grade math and science course grades, with one exception (Table 6.10). The exception is students with D and F mean 8th grade math and science course grades.13

However, females are over-represented in the work samples scored for each instructional unit compared to their proportion in the entire ninth grade, but the degree of
over-representation varies between instructional units (Table 6.10). Between instructional units, the proportions of ethnic and prior achievement groups also varies (Table 6.10). The differences between central tendency measures for these units can’t be tested statistically. The performance levels between the four instructional units are compared qualitatively.

The differences in performance by gender in each instructional unit are examined compared to variation in the data. However, the low numbers of male students in all but the Energy Concepts unit (Table 6.10) limit statistically based conclusions. Conclusions regarding differences between ethnic categories and prior achievement groups are not possible because of concerns discussed for the behavioral engagement and self-regulated learning analyses. Group sizes are small and the students that gave consent are not likely representative of the whole ninth grade. The mean performance levels for each group are documented, but comparisons can’t be made.

6.1.5.2 Differences between Performance Authenticity Indicators

The three mean Performance Authenticity Indicators (PAIs) scores for each instructional unit are computed from the same set of work samples from the same students. Because scores for each student are related, a repeated-measures analysis is used to compare differences between the three PAIs for each instructional unit. Where the requirements of an ANOVA are not met, a nonparametric alternative is used.

Conclusions based on central tendency measures of the entire dataset of 96 work samples are limited because of the widely unequal proportions of work samples from each unit in the total sample and the unrepresentative samples from each of the four
instructional units (Table 6.10). Therefore, the means for each PAI and Combined Authenticity from each instructional unit are used to calculate the overall mean authenticity levels as a best estimate. These means of means are compared qualitatively.

### 6.2 Results

Research Question 2 asks how three student response variables, behavioral engagement, self-regulated learning, and authenticity of student science performance, vary over the school year. To answer this question, each of the three student response variables are considered in turn. The means of percent behaviorally engaged per student for the entire study period from October to June, each trimester, and selected instructional units were determined. Patterns at each level of analysis are described.

#### 6.2.1 Behavioral engagement levels over the entire study period

Mean percent behaviorally engaged per student is reported here as $M\% \pm SD, N$. Using percent of observations behaviorally engaged for each student, the mean behavioral engagement for all students in the analytic sample over the entire study period is $74\% \pm 14, N = 83$. The mean percent behaviorally engaged for males, $75\% \pm 12, N = 32$, is slightly higher than females, $73\% \pm 15, N = 51$ (Appendix E, Table E.1). The gender differences are not statistically significant.

#### 6.2.1.1 The influence of demographic factors

Differences in behavioral engagement levels for the entire study period between ethnic groups can’t be statistically tested. The means of percent behaviorally engaged per student for ethnic groups fall between $84\%$ and $72\%$ (Appendix E, Table E.2). The
behavioral engagement levels for all groups are measured as similar and relatively high. The highest scoring categories, Asian and Native Hawaiian (84%) and Other Race (81%), have such small N’s they can’t be evaluated statistically. The largest categories, Hispanic or Latino and Euro-American, were measured as 73% ± 13, N=12 and 77% ± 10, N=24 behaviorally engaged, respectively. However, lack of independence and small group sizes indicate the statistical power needed for tests of group differences isn’t present.

The means of percent of observations behaviorally engaged per student increased with increasing mean 8th grade science and math grades. The mean percent behaviorally engaged range from 81% ± 8, N=19 for students that averaged A to 54% ± 5, N=3 for those with an F average (Appendix E, Table E.3). The differences in behavioral engagement levels between grades A, B, and C are small relative to their standard deviations, while lower engagement was measured for students with mean grades of F. This result fits what would generally be predicted. As with the analysis of ethnic groups, statistical significance of group differences isn’t presented.

### 6.2.2 Trends in behavioral engagement during the school year

The mean percent behavioral engagement per student for each trimester varied little from the 74% level measured for the entire study period. The behavioral engagement level in the fall trimester was slightly higher than both the winter and spring trimesters, which were very similar (Table 6.4). A Friedman test (α=0.05) was run to determine if there were differences in behavioral engagement levels for each trimester for the 70 students present in all three trimesters. The percent behavioral engagement per student decreased between fall (Mdn = 79) and winter (Mdn = 75), and then remained the
same in spring \((Mdn = 75)\), but the differences were not statistically significant, \((\chi^2(2) = 4.394, p = 0.111)\).

The mean percent behavioral engagement per student measures for each trimester were determined for each gender. Slightly higher rates of mean percent behaviorally engaged per student per trimester were measured for female students during fall trimester and male students measured higher for winter and spring, but there were no significant differences between genders (Table 6.4).\(^{15}\)

The only ethnicity categories that varied more than 4% in mean percent behavioral engagement per student between trimesters were the Asian and Native Hawaiian, Multiple Categories Selected and Other Race groups, with only 5 or 6 members. There were differences of 5-10% in behavioral engagement levels between trimesters (Appendix E, Table E.4), but no conclusions can be reached. Only small differences in the Hispanic or Latino and Euro-American categories were measured between trimesters.

6.2.2.1 The influence of class periods

Means of percent behavioral engagement per student for class periods 2, 3 and 5 changed between each school trimester (Appendix E Table E.4), but the differences were within the variation in the data as shown by the boxplots in Figure 6.1. The mean percent behaviorally engaged per student decreases slightly over the year for periods 2 and 5, but period 3 is measured lowest during the winter trimester. During each trimester, period 5 has the lowest mean percent behaviorally engaged, but the only significant difference is between class periods 3 and 5 during the spring trimester.\(^{16}\)
6.2.3 Differences in behavioral engagement between instructional units

The means of percent behavioral engagement per student vary between seven instructional units from across the school year chosen for intensive study, but there is no trend across the school year (Table 6.5). The behavioral engagement levels range from a high of 89% ± 11, \( N=49 \) for the RNA Fingerprinting Lab unit in the middle of the first trimester to 66% ± 22, \( N=81 \) for the Teen Brain unit at the end of the winter trimester. The relatively large standard deviations demonstrate that there was great variation in the percent engagement of individual students during each assessment task.

The Immune System and Flu Vaccine, the Teen Brain, and the Wind Turbine Engineering and Testing units are all end-of-trimester project tasks focused on asking students to produce a significant artifact integrating science concepts and the compelling topic of the trimester. The mean behavioral engagement per student levels for the units at the end of the fall and winter trimesters are similar, 68% ± 18, \( N=84 \) and 66% ± 22, \( N=81 \), respectively (Table 6.5). By contrast, an 8% greater mean behavioral engagement per student was measured during the Wind Turbine project at the end of the school year, 75% ± 16, \( N=80 \). This level is similar to the average over the entire study period, suggesting that the students maintained the same average level of engagement up until near the end of the school year.

To compare behavioral engagement levels between instructional units, a repeated-measures analysis was used. Inclusion of five of seven of the instructional units selected for measurement of behavioral engagement levels resulted in 69 students with a measured percent behaviorally engaged for each unit. The RNA Fingerprinting Lab unit and the Engineering Design Challenge 1 unit in Table 6.5 are not included because fewer
students were measured during those instructional units. A Friedman test was used to determine if the distributions of percent behaviorally engaged for these 69 students were statistically significantly different. Pairwise comparisons were performed (SPSS, 2013) with a Bonferroni correction\(^1\) for multiple comparisons. Differences between instructional units were found \(\chi^2(4) = 25.384, p < .001\). This is a small effect, \(W = 0.092\) (Kinnear & Gray, 2010).

Post hoc analysis found statistically significant differences between the distributions of percent behaviorally engaged per student in four pairwise comparisons. The distribution of percent behaviorally engaged per student for the Immune System & Flu Vaccines unit \((Mdn = 71.0)\) is lower than that for Human Characteristics Data Analysis \((Mdn = 79.2)\) \((p = .04)\). The distribution of percent behaviorally engaged per student for the Immune System & Flu Vaccines unit \((Mdn = 71.0)\) is also lower than that for Energy Concepts Quiz \((Mdn = 83.3)\) \((p = .002)\). In addition, the distribution of percent behaviorally engaged per student for Human Characteristics Data Analysis \((Mdn = 79.2)\) was greater than that for Teen Brain \((Mdn = 72.4)\) \((p = .015)\). Finally, the distribution for the Energy Concepts unit \((Mdn = 83.3)\) is greater than that for the Human Characteristics Data Analysis \((Mdn = 79.2)\) \((p > .001)\). The median for the Wind Turbine Engineering & Testing task is 78.4, but no pairwise significant differences were found.

The Friedman Test results indicate that for the 69 students in the analysis, the greatest behavioral engagement is in the Energy Concepts unit and then the Human Characteristics Data Analysis unit. The lowest behavioral engagement levels were measured for the Immune System and Flu Vaccines and the Teen Brain units. The repeated-measures analysis doesn’t tell where the Wind Turbine Engineering and Testing
unit is ranked. The small effect sizes in differences detected between medians of percent behaviorally engaged per student for some instructional units reflects that the differences are small relative the variation in the data (Table 6.5).

6.2.4 Self-regulated Learning levels during the study period

The self-regulated learning presence variable records the presence or absence of any type of evidence of student self-regulation. The self-regulated learning category variable records whether self-regulation, when observed, is (1) limited to independently following the teacher’s direction for an activity or task or (2) shows autonomy or ownership that goes beyond following the teacher’s direction. For both variables, the observations are related, not independent. Throughout this section results for these variables are presented as percent of observations indicating self-regulation (for the presence variable) or working autonomously (for the categorical variable) out of the total observations made (N).

Over the entire school year, the self-regulated learning presence variable measured evidence of self-regulation for 54%, N=6787, of the individual observations (Table 6.6). The ratio of observations of females and males in the data collected over the entire study period and for winter and spring trimesters is the same, to the nearest 2%, as their proportions in the student population in these classrooms, 60% female and 40% male.18

The rate of self-regulating learning observed during winter and spring trimesters for all students was 49%, N=1809 and 52%, N=4434, respectively (Table 6.6). These small differences suggest the levels of self-regulated learning behavior changed little
between these trimesters. Females were measured to have a slightly greater rate of self-regulated learning during fall trimester, which is reversed for winter and spring (Table 6.6). These results do not show evidence for a gender effect between winter and spring trimesters.

Working independently was recorded more often than working autonomously for the self-regulated learning category variable during winter and spring trimesters. During winter trimester 29%, \( N=829 \), of observations were recorded as working autonomously, while the rate in spring was about half, 13%, \( N=2280 \). A slightly higher rate of working autonomously is measured for males in both the winter and spring trimesters, but the results do not provide evidence to confirm a gender difference (Figure 6.3).

### 6.2.5 Self-regulated learning levels in four selected instructional units

The percent of observations exhibiting self-regulated learning measured by the presence variable varied between the samples of lesson segments observed for the four instructional units with self-regulated learning observations from a variety of classroom activities representative of the unit (Figure 6.2). These four units were selected from the seven instructional units selected for intensive study of the authenticity of the pedagogy. The lowest percent self-regulating is measured for the Energy Concepts unit, 36%, \( N=287 \) and the highest for the Wind Turbine Engineering and Testing unit, 60%, \( N=2504 \). Both the Teen Brain (42%, \( N=650 \)) and Engineering Challenge 1 (43%, \( N=341 \)) units scored slightly higher than the Energy Concepts unit.

The percent of observations with evidence of self-regulated learning of any type during the last four of the seven selected instructional units showed differences by gender
Males were measured as 5-6% more self-regulating than females during the three units in the spring trimester, but 4% less self-regulating in the Teen Brain unit at the end of the winter trimester. These results are consistent with presence and categorical self-regulated learning results for gender differences by trimester, but although they consistently measure greater male self-regulation during the spring trimester, this study is not able to conclude this is representative of the entire 9th grade population in these classrooms.

Exploration of the entire set of self-regulated learning presence variable observations found that when class activities grouped the students as a whole classroom, self-regulation was observed 22% of the time, while it was observed 60% of the time when students were working in small groups or individually (Table 6.8). For three of the four instructional units included in the analysis of self-regulation, the fraction of observations of students working in small groups or individually versus grouped as a whole is positively associated with the percent of observations students were observed self-regulating. This suggests that comparison of self-regulated learning presence variable observations of students working in small groups and individually and those collected when the classrooms were grouped a whole classroom should be compared separately.

Differences in the grouping of students during class activities influence the levels of self-regulated learning observed (Table 6.8). In all four selected instructional units percent of observations self-regulating was greater during observations with students working in small groups or individually than when the classroom was grouped as a whole class. The percent self-regulating for students working in small groups or individually
for the Teen Brain, Engineering Challenge 1, and Wind Turbine Engineering and Testing units was slightly greater than for all observations, while the percent self-regulating for the Energy Concepts unit increased substantially, changing its percent self-regulating ranking from lowest for all observations to second greatest for only when students working in small groups or individually (Table 6.8).

There are two complimentary patterns in the observations in Table 6.8 that may influence the contrasting Energy Concepts unit results. First, nearly half of the Energy Concepts unit observations were during activities that grouped the classroom as a whole, a larger fraction than in the other three units, and during those lesson segments the percent self-regulating was only 18%. The difference in percent of observations self-regulating between the two categories of student grouping for all but the Teen Brain unit was as great as for the Energy Concepts unit, but the proportion of total observations with the classrooms grouped as a whole was lower. It is also found that the differential between the two categories of student grouping patterns is lower for the Teen Brain unit. The role of student grouping during classroom activities should be considered when interpreting these self-regulated learning levels.

The number of observations when students were found to be working autonomously was a pretty small fraction of those observations that recorded students only working independently for all four selected instructional units (Table 6.9). The Teen Brain, Engineering Challenge 1, and Wind Turbine Engineering and Testing units were measured to have similar levels of percent working autonomously. However, the Energy Concepts unit was measured at twice their level. Table 6.9 shows that evidence of students working independently and autonomously was found in each of the four
selected instructional units. The variation in percent working autonomously between units suggests that student responses varied between units.

6.2.6 Authentic Student performance levels

6.2.6.1 Differences between four instructional units

The mean combined authenticities of student performance on science work samples from the four different units (Table 6.11) are measured as different. The mean student performance levels varied from moderately high authenticity for the RNA Fingerprinting Lab and the Wind Turbine Engineering and Testing units, to moderately authentic for the Immune System and Flu Vaccines unit, and moderately low authenticity for the Energy Concepts unit (Table 6.11).

The large standard deviations in each instructional unit’s mean combined authenticity scores suggest that boxplots (Figure 6.4) are a useful descriptive tool for work sample performance differences between the units. The median Combined Authenticity for student work samples from each of the four instructional units studied varied from a highly authentic level of .83 for the RNA Fingerprinting Lab unit to similar moderately high values of .61 and .63 for the Immune System and Flu Vaccines and Wind Turbine Engineering and Testing units, respectively, and finally a moderately low authenticity value of .33 for the Energy Concepts unit (Figure 6.4).

The boxplots illustrate great variation in Combined Authenticity of student performance for each unit (Figure 6.4). The greatest variation is in the Energy Concepts unit, showing that 5 students were scored at the bottom of the authentic performance scale and one student at the top of the scale (Figure 6.4). Most of the students scored
moderately low or low authenticity. Very few of the students scored for the Energy Concepts unit performed above moderate authenticity.

Most of the student work samples for the two end-of-trimester instructional units, the Immune System and Flu Vaccines unit in the fall trimester and Wind Turbine Engineering and Testing unit in the spring trimester, scored greater than moderate authenticity (Figure 6.4). Only a few work samples from the Immune System and Flu Vaccines unit, and somewhat more from the Wind Turbine unit, score in the highly authentic range. However, the distribution of student performance scores was higher for the Wind Turbine unit, accounting for its higher mean score. In contrast to the Wind Turbine unit, a much larger proportion of students performed in the moderately low authenticity range of the scale for the Immune System and Flu Vaccines unit, and some scored in the low authenticity range.

The RNA Fingerprinting Lab and Wind Turbine units had similar ranges of scores, but most of the students scored for the RNA Fingerprinting Lab unit performed in the highly authentic range, while only one-third of students were scored highly authentic for the Wind Turbine unit. The other difference is that nearly half of the students scored for the Wind Turbine unit performed in the moderately authentic range.

6.2.6.2 Differences between the three PAIs for each instructional unit

Each work sample scored for each instructional unit was scored for all three PAIs. A repeated-measures analysis tested whether mean authenticity of student performance levels measured for the set of work samples scored for each PAI for each unit are the same (Table 6.11). The scores for each of three PAIs are not normally distributed for any
of the four instructional units. Therefore, a nonparametric Friedman test was run for each instructional unit to determine if the distributions of mean performance scores for each PAI are different.

For the RNA Fingerprinting Lab unit, students scored greater on Scientific Analysis, PAI 1, and Elaborated Communication, PAI 3, (both \(Mdn = .83\)) were greater than Science Concepts, PAI 2, (\(Mdn = .67\)) but the differences are not statistically significant, \(\chi^2(2) = 17.50, p = .417\). For the Energy Concepts unit, the medians for all three PAIs were .33 and no statistically significant differences were found, \(\chi^2(2) = 2.419, p = .298\).

For the Immune System and Flu Vaccines unit, the null hypothesis that the distributions of all three PAIs are same is rejected, \(\chi^2(2) = 8.842, p = .012\). Students scored lower on Scientific Analysis, PAI 1, (\(Mdn = .50\)) than Science Concepts, PAI 2, and Elaborated Communication, PAI 3, (both \(Mdn = .67\)). However, pairwise comparisons were performed (SPSS, 2013) with a Bonferroni correction for multiple comparisons and found no significant differences.

For the Wind Turbine Engineering and Testing unit, pairwise comparisons were performed (SPSS, 2013) with a Bonferroni correction for multiple comparisons. Statistically significantly different student performance scores were found between the three PAIs, \(\chi^2(2) = 36.447, p < .0005\). Post hoc analysis revealed statistically significant differences in Scientific Analysis, PAI 1, (\(Mdn = .89\)) and Science Concepts, PAI 2, (\(Mdn = .56\)), and between Scientific Analysis, PAI 1, and Elaborated Communication, PAI 3, (\(Mdn = .44\)).
6.2.6.3 Differences by demographic variables

In the samples of student work samples scored for each of the four selected instructional units the mean Combined Authenticity for the male students was greater than the female students (Figure 6.5). For the RNA Fingerprinting Lab unit the difference is .26, but there are only 3 males in the sample. For the other three units the differences are close to .10 (Figure 6.5). This consistent pattern suggests that authenticity of performance for males in the sample is slightly greater, but the large standard deviations (Figure 6.5) show that the differences are not significant for any of the instructional units.

The ranks of mean Combined Authenticity of performance for the five ethnic groups in this study for the RNA Fingerprinting Lab are different than the other units. The numbers of students per category are too small to make even qualitative comparisons with the other units (Appendix E, Table E.5). For all of the other three units, the Asian or native Hawaiian ranked highest mean and median authenticity and the Euro-American category ranked second highest. This is the only consistent pattern in the data. For the last three units, the other categories varied in their rank orders. The differences by ethnic group only describe the work samples scored for each instructional unit because group numbers are too small for appropriate statistical tests (Appendix E, Table E.5).

For all four instructional units, students with average grades of A in math and science courses in 8th grade ranked highest mean and median Combined Authenticity of student performance (Appendix E, Table E.6). While the statistical significance of the differences can’t be tested, this is what would generally be predicted. There are no
consistent patterns in the ranking of authentic performance levels for students with B, C, or D averages.

6.3 Discussion

6.3.1 Behavioral engagement and self-regulated learning

6.3.1.1 Patterns over the school year

Over the school year and during selected instructional units, observations of behavioral engagement and self-regulated learning focused on the students in the three classrooms that met during class periods 2, 3, and 5. Table 6.4 shows that the mean percent behaviorally engaged per student per trimester remained stable through the year at about 75%. Similarly, Table 6.6 shows that during the winter and spring trimesters when the sample of observations of the self-regulated learning presence variable included representation from a broad variety of the classroom activities from across each trimester, the measured rate of self-regulation finds the teacher’s pedagogy allowed students to work independently, and strongly suggests they did so at least half the time.

Gender differences in behavioral engagement (Table 6.4) and self-regulated learning (Table 6.6) rates are measured at the trimester level and over the entire study period, but these are small. The gender differences for behavioral engagement are not statistically significant. The self-regulating learning data also does not confirm gender differences. Both male and female students in these classrooms were behaviorally engaged and self-regulating at similar rates, and these rates don’t change substantially over the study period.
Over the entire study period, the behavioral engagement rates for the largest ethnic groups, Euro-American and Hispanic or Latino, did not deviate more than four percent from the measured means for all students studied (Appendix E, Table E.2). Similarly, the largest groups with prior year math and science grades available, students with A, B, or C averages, are measured with greater behavioral engagement for the students with A and B averages, than those with C averages, as expected (Appendix E, Table E.3). It is a limitation of this study that the differences measured in these behavioral engagement levels can’t be shown to represent the larger population of 9th graders in this study because of concerns about how obtaining informed consent influenced the representativeness of these samples, the small numbers of group members, and the relatedness in the samples.

6.3.1.2 Patterns during selected instructional units

Both the measurements of mean percent behaviorally engaged per student during seven selected instructional units (Table 6.5) and of percents of individual observations of any type of self-regulated learning for four selected instructional units (Figure 6.2) are measured to vary between instructional units. Although both of these measures vary quite differently between units, they both suggest that the rates of behavioral engagement and self-regulated learning during the Wind Turbine Engineering and Testing unit, which extends into the last eight days of the school year, are at or above the rate for the year. Students remained engaged and self-regulating through very near the end of the school year.
The variation in behavioral engagement rates between the seven instructional units in Table 6.5 and the consistent behavioral engagement rates by trimester in Table 6.4 suggest that differences in classroom activities and interactions during different instructional units may have greater influence on behavioral engagement rate than time in the school year. This suggests that the accumulation of time the students are exposed to this teacher’s pedagogical practices as the school year progresses may have less influence on student attention and involvement than differences between instructional units. The same appears to be true for student self-regulated learning, as shown by the consistency of self-regulation rates measured during winter and spring trimesters (Table 6.6) and variation between instructional units in (Figure 6.2). Chapter 7 will investigate whether differences in behavioral engagement and self-regulated learning are associated with changes in the authenticity of the pedagogy between these units.

6.3.1.3 The Influence of some instructional factors on self-regulated learning

The self-regulated learning data provides a richer description of student responses to instruction than the rate of attention and involvement measured by the behavioral engagement variable. This chapter investigates the presence of evidence of cognitive engagement, as measured indirectly by self-regulation, during selected instructional units.

Even though the repeated-measures analysis did not detect a significant difference between the behavioral engagement levels measured for the Teen Brain and the Wind Turbine Engineering and Testing units, the relationship between measured levels is similar for both the behavioral engagement and self-regulated learning presence variables for these units. Although the pattern can’t be statistically confirmed, the Teen Brain unit
is measured lower than the average over the entire study period, while the Wind Turbine Engineering and Testing unit is measured at or above the averages for both variables.

There was evidence of self-regulated learning in both of these units, but the data suggests the prevalence was lower during the Teen Brain unit. However, for both of these units, only a small fraction, around 10%, of the observations of self-regulated learning discerned student action beyond independently following directions (Table 6.9). This suggests students exhibited the type of self-regulated learning most indicative of cognitive engagement in both of these units. However, the protocol used in this study isn’t robust enough to be confident how prevalent this behavior was or whether the rates were similar.

The Teen Brain unit is the final unit in the winter trimester and the Wind Turbine unit the final unit in the spring. As noted in Chapter 5, the instructional units at the end of each trimester addressed similar instructional goals. Each of them asked students to apply and integrate several concepts taught during the trimester to create a substantial artifact addressing the compelling topic of the trimester. In addition, in each of these units considerable class time was devoted to students creating that artifact as an assessment product. Comparing the Teen Brain and Wind Turbine units illustrates the contrasts measured in student responses during units with some common characteristics.

The Teen Brain unit was focused more on evaluating and assembling information, than on assembling and evaluating a mechanical design, which characterized the Wind Turbine unit. A greater proportion of class time during the Teen Brain unit was spent on pencil and paper activities, as opposed to the large fraction of class time spent on designing, building, and testing cycles in the Wind Turbine unit. Classroom field notes
show that active engagement with their designs was promoted by the instruction during about 90% of the class time observed in the Wind Turbine unit, but that less than 10% of the class time observed during the Teen Brain unit gave students the opportunity to be directly engaged with a natural phenomenon.

Table 6.8 shows that during the class time students were working in small groups or individually, the pattern of greater self-regulated learning during the Wind Turbine unit than the Teen Brain unit is repeated. The difference in time engaged in creating and tinkering with a physical design is likely an important factor in the increased student response. Chapter 5 measured greater authenticity in the Wind Turbine assessment task and classroom instruction than in the Teen Brain unit. The time spent actively designing is measured by some of the authentic instruction indicators. Chapter 7 will further investigate how instructional characteristics are associated with student engagement.

When students work in small groups or individually they have more responsibility for managing their own learning. Therefore, the results in Table 6.8 suggest that during the class time students in these classrooms have more of this responsibility, they consistently show evidence of self-regulation that indicates cognitive engagement without a need for constant teacher supervision. Table 6.8 shows that the Wind Turbine Engineering and Testing unit appears to be particularly effective in this regard.

For all four instructional units studied, Table 6.8 also suggests that percent of individual observations indicating any type of self-regulated learning were greater when students worked in small groups or individually than when doing an activity with the classroom functioning as a whole. It is a limitation of this study that the conservative nature of the observation protocol in Appendix D during whole class observations and
student disposition may limit observation of student expression of self-regulation during these lesson segments.

Therefore, it clarifies the description of differences in the levels of self-regulated between instructional units in Table 6.8 to also compare levels for the same student grouping during classroom activities. This is helpful in assessing how the level of self-regulated learning in the Energy Concepts unit compares with the levels during the Teen Brain and Wind Turbine Engineering and Testing units.

Chapter 5 also noted that during each trimester, the teacher used shorter instructional units to focus students on developing an understanding of a small number of important science concepts. The Energy Concepts unit is an example of an instructional unit exhibiting this pattern, which contrasts strongly in several ways with the end-of-trimester project units, such as the Teen Brain and the Wind Turbine units. Much of the class time in the Energy Concepts unit was spent on lectures and worksheets. The assessment task was a one page quiz consisting of short answer and open response conceptual physics items, instead of an individually designed PSA with selected content or an engineering design notebook constructed over a week or more of class time.

One of the highest levels of behavioral engagement, 79%, was measured during the Energy Concepts unit (Table 6.5). By contrast, the self-regulated learning presence level was 36%, the lowest of any of the four selected instructional units (Table 6.8). Table 6.8 also shows that about half of the self-regulated learning observations during the Energy Concepts unit were made during activities when the students were grouped as an entire classroom. This is a greater proportion than for the Teen Brain and the Wind Turbine units.
When only observations of students working in small groups or individually are considered, the measured self-regulated learning presence level for the Energy Concepts unit is between the corresponding levels for the Teen Brain and the Wind Turbine units (Table 6.8). This shows another example of a higher presence of self-regulated learning being measured when students had more responsibility for their learning.

Table 6.9 adds additional detail regarding how students respond to the Energy Concepts unit. A larger fraction, 26%, of the observations of the self-regulated learning category variable during the Energy Concepts unit were recorded as working autonomously than for either of the end-of-trimester project units. There is evidence here that the teacher had at least moderate success keeping students cognitively engaged during a unit focused on developing basic science conceptual knowledge.

The self-regulated learning convenience samples in this study don’t allow statistical analysis of the patterns found between trimesters and instructional units. However, self-regulated learning was observed in the four instructional units observed enough to estimate its occurrence, including up until the last two weeks of the school year. There is evidence that self-regulated learning was measured at a consistent level across the winter and spring trimesters. In addition, self-regulated learning was observed during the lesson segments that gave students greater responsibility for managing their own learning.

**6.3.2 Authentic performance**

Documentation of the authenticity of student performance on work samples allows this study to describe this aspect of student responses to performance expectations
communicated by this teacher’s assessment tasks and pedagogy. This study assesses the overall level of student performance and whether there is evidence of trends across the study period from October to June.

There are several threats to measured performance authenticity levels for each of the four instructional units. The most important threat to inferences regarding performance differences between instructional units and trends over the study period is the differing numbers and students in the sets of work samples collected for each instructional unit. The dependence on the availability of work samples to this investigator prevented collection of matched sets of students from each instructional unit. In addition, for units with work samples scored in subsections, work samples without work completed on each subsection were not included. Some students may have been excluded because of difficulty completing work on time.

The set of work samples scored for each of the instructional units includes representation from students from each of the demographic groups considered in this study, but these are still convenience samples. This study doesn’t claim that the samples for each instructional unit are generalizable to the entire ninth grade population.

The measured performance levels for each instructional unit are also influenced by the teacher’s implementation of the school’s proficiency-based grading system. Because students were allowed multiple opportunities to show proficiency on Learning Targets, the work samples collected can’t be known to be the highest performance level of each student. For example, a work sample scored for this study may later have been improved by the student after feedback from the teacher, who allowed the student to rewrite a section of the work, or explain a concept orally to the teacher.
Alternatively, a student may have decided to put a preliminary effort into a particular work sample, with the intention to perform better on a different alternative task. This decision may have been strategic. For example, occasionally students expressed a desire to put more effort into a different class at a particular time. Other reasons for making a limited effort may have included unease about preparation for the task, procrastination, or peer influence.

Nearly all of the work samples in the study were collected very near the end of the instructional unit they represent.21 This investigator’s observations over the school year showed that a large majority of students made a strong effort to complete a proficient performance by the end of the instructional unit. The measures of student performance in this study are measures of student performance primarily at the end of each instructional unit, but don’t characterize the highest level of achievement reached for each of the students.

Student maturation over the period data collection in this study is expected. Because all the students are at the same grade level, differences between student participants aren’t expected to be systematic (Creswell, 2008). In addition, students in this study were not surveyed regarding any experiences during the school year that improved their study skills or motivation apart from the experience within the classrooms studied (Creswell, 2008). This investigator did not record any program with these goals within the school with participation of some of the students within this study. However, observations were not complete enough to rule it out. It is unlikely this threat would change the conclusions of this study.
In addition, students that created the work samples scored in this study were known to this investigator while he was scoring. One of the three other work sample scorers had visited the classroom several times during the school year to observe students’ behavioral engagement and self-regulation. However, student names were removed from the work samples prior to scoring. The other two scorers had no experience with the school or any participants in this study. This threat was minimized by analysis of each of the assessment tasks to clearly identify the science and engineering concepts and skills to be assessed by the scoring guide for each task.

In spite of these limitations, this study reaches valuable qualitative conclusions regarding differences between instructional units and trends between them.

6.3.2.1 Trends within the school year

The authenticity of student performance differences measured between the sets of work samples from each of the four instructional units studied can’t be assessed statistically. Qualitative analysis reveals some patterns in the results that suggest relationships between student performance and the performance expectations in the teacher’s assessment tasks for these units.

The ranks of mean Combined Authenticity of student performance levels measured for each of the four instructional units are generally positively associated with the ranks of combined assessment scores (Figure 6.6). This suggests that on average students’ performance represents a response to the teachers’ expectations, that is, when the assessment tasks for these units call for greater authenticity, these students respond with higher levels of performance. Examples of cases that fit this pattern and some that
don’t are discussed here. Together these results suggest that the authenticity of the assessment tasks plays a larger role in students’ performance than progress in the school year. Means here are presented as $Mean \pm SD$.

Another feature of these results is the large degree of variation in the individual work sample scores from each instructional unit (Table 6.11). There are examples of individual students performing at higher levels than expectations of the assessment task. For example, one student scored at the very top of the authentic performance scale for the Energy Concepts quiz, even though the Combined Authenticity of the assessment task is only 0.25. It appears likely that any signal showing improvement due to longer exposure to the instructional environment in these classrooms is smaller than the variation in the data.

The teacher’s Wind Turbine Engineering and Testing assessment task scored at the top of the authenticity scale for Construction of Knowledge (Figure 6.6). The Wind Turbine unit focused on engineering design processes and included opportunities to apply science concepts to designs. It included multiple opportunities to organize information to inform designs; organize design records in an engineering format; brainstorm designs; build, test, and evaluate designs; and display testing data in a scientific format. These opportunities resulted in the high score for Construction of Knowledge for the assessment task. The student work samples scored for this unit showed that many students were successful constructing knowledge in these ways. The mean for PAI 1 was $0.83 \pm 0.17$ (Table 6.11).

Although a smaller set, the work samples scored for the RNA Fingerprinting Lab also showed that this is another example when at least some students responded
successfully to high expectations for Construction of Knowledge. The mean Construction of Knowledge score for the RNA Fingerprinting Lab assessment tasks was 0.63, and the corresponding mean student performance score for Construction of Knowledge was 0.79 ± 0.25 (Table 6.11).

The set of students scored for the Wind Turbine Engineering and Testing unit were, on average, somewhat less successful using Science Concepts (PAI 2 = 0.61 ± 0.22) and Elaborated Communication (PAI 3 = 0.57 ± 0.22) even though the assessment task for this unit also scored at the top of the authenticity scale for Science Meaning-Making Processes (Figure 6.6). The assessment task asks students to use systematic engineering strategies to pursue a design goal. Students are expected to use build-test-modify design cycles to iteratively improve a design; predict successful design improvements by applying energy conservation concepts; acknowledge design tradeoffs; and collect useful evidence to evaluate design function. The task also asks students to communicate testing results, reason from those results, and evaluate their design’s success relative to their goal.

The set of work samples scored for the RNA Fingerprinting Lab is small, but performance on Science Meaning-Making Processes (0.73) appears to contrast with the relatively less successful responses to the teacher’s expectations for the Wind Turbine Engineering and Testing unit (Table 6.11). All three mean Performance Authenticity Indicators (PAIs) and the Combined Authenticity for the RNA Fingerprinting Lab work samples scored moderately high authenticity (Table 6.11). The focus of each of the two versions of the assessment task for this instructional unit is interpretation of an electrophoresis gel in terms of the molecular interactions of RNA, restriction enzymes,
and the factors controlling the movement of molecular fragments through an electrophoresis gel to produce a visible pattern. This focus accounts for the assessment task’s highly authentic scores for the Science Meaning-Making Processes component (Figure 6.6). A majority of the students scored for the RNA Fingerprinting Lab performed at moderately high authenticity or above on use of Science Concepts and Elaborated Communication in their work samples, giving an example of students that responded successfully to the teacher’s expectations.

One of the distinguishing features of the teacher’s pedagogy noted in Chapter 5 was a consistent emphasis on writing about science in assessments. None of the teacher’s assessment tasks analyzed were dominated by short answer questions. Even quizzes such as, the Energy Concepts quizzes, were dominated by open response items. Except for the Energy Concepts quizzes, the assessment tasks corresponding to the other three instructional units analyzed expected extended writing in the form of an explanation or argument. Students were measured at higher mean performance levels on Elaborated Communication (PAI 3) for the RNA Fingerprinting Lab, Immune System and Flu Vaccines, and Wind Turbine units, than the Energy Concepts quizzes (Table 6.11), in line with the teacher’s higher expectations.

The Immune System and Flu Vaccines is the unit at the end of fall trimester and the Wind Turbine Engineering and Testing unit is its counterpart at the end of the spring trimester. As noted, the assessment tasks for these units asked students to use several important concepts and skills from each trimester to create a more complex artifact that addressed the compelling question for the trimester. The assessment task for the Wind Turbine unit was measured to have much higher expectations than the task for the
Immune System and Flu Vaccines unit for Construction of Knowledge (1.00 vs. 0.50) and Science Meaning-Making Processes (0.95 vs. 0.57). However, greater performance is measured in the student work samples for Construction of Knowledge (0.83 vs. 0.53), but the differences between all the Science Meaning-Making Processes PAIs are less than 0.10 (Table 5.9). While this study can’t confirm these contrasts between performance on the two components aren’t due to sampling differences, the results suggest that when students were asked to perform at a higher level in the spring, they responded successfully for Construction of Knowledge, but less so for Science Meaning-Making Processes. Students were successful meeting some, but not all, of the teacher’s increased expectations in the spring trimester culminating project relative to the culminating project in the fall trimester.

The Energy Concepts unit is an example of another recurring teacher instructional strategy. The teacher focused on building knowledge of energy conservation concepts, which students applied in the Wind Turbine unit at the end of the trimester. The Combined Authenticity of the assessment task for the Energy Concepts unit was the lowest of the four instructional units scored for authenticity of student performance (Figure 6.6). Students scored lower mean levels of performance than the other three selected units for both Construction of Knowledge and Science Meaning-Making Processes, reflecting the lower Combined Authenticity of the assessment task (Figure 6.6).

Any influence of the Value Beyond School component of assessment task authenticity on student performance levels is difficult to discern. The Energy Concepts unit assessment tasks score 0 for the Value Beyond School component, while the other
three instructional units have similar scores from 0.40 to 0.50, in the moderate authenticity range. The mean performance levels for the set of student work samples from the Energy Concepts unit are measured lower than other units, but at this level it is not possible to separate the influence of the Value Beyond School component from the Science Meaning-Making Processes component which is also lower for this unit.

6.3.2.2 Influences of demographic variables

There is a small persistent gender difference in the mean authenticity of performance for each instructional unit. Males score slightly higher than females, but the difference is not significant. Students with prior year math and science grades averaging A ranked highest performance level for all four instructional units, but this can’t be confirmed statistically. In all but the RNA Fingerprinting Lab unit, students with Asian or Native Hawaiian, then Euro-American ethnicities had the highest mean combined performance scores, but statistical confirmation of the differences wasn’t possible. For the RNA Fingerprinting Lab, two Hispanic or Latino females scored the highest Combined Authenticity possible. Within the entire set of work samples scored, there were examples of relatively high levels of authentic performance, at least 0.78, for every ethnic category.

6.3.2.3 The performance of the authentic student performance instrument

Although few of them can be confirmed statistically, the authentic performance scoring guides are able to detect differences within student performance for each of the PAIs. They measure performance level differences between PAIs within instructional
units. In addition, differences are measured for each PAI between instructional units. Many of the differences are similar to those in the assessment task authenticity levels.

Achieving inter-rater reliability was challenging, but a very good level of agreement was shown to be possible. In this study it was found that clear articulation or expert understanding of the relevant concepts, practices and skills, and appropriate levels of performance were needed to use the scoring guides consistently.

This student performance instrument could be useful to assess student performance changes relative to authentic performance expectations. The framework of authenticity in the instrument is adaptable to measuring any cognitive science and engineering learning standards. There is a potential for using the instrument to track student success across grade levels and subject areas within a common framework.

6.3.2.4 Comparison with previous research

This study shares a common model of authentic student performance with the Center on Organization and Restructuring of Schools (CORS). This study’s scoring guides were adapted from the CORS study guides for authentic student performance in high school mathematics and social studies (Newmann et al., 1995). The CORS School Restructuring Study (SRS) measured authentic student performance in 44 mathematics and social studies classrooms from eight high schools (Marks et al., 1996). The mean combined performance level measured for 642 student work samples was below the mid-point of the scoring scale, at 0.39 ± .22 (SD), when the mean is normalized as in this study.
Similar to this study’s results, the range of work sample scores in the SRS report was from very near the top to the bottom of the authenticity scale. In addition, they report examples of student work samples that were responses to a one-page worksheet of short answer items to 10-page reports that culminate many weeks of study and that called on students to integrate several concepts and skills as they answered a compelling question in a creative way (Marks et al., 1996).

The mean of the Combined Authenticity scores for all 96 of the work samples collected from four instructional units in this study was somewhat higher, $0.55 \pm 0.27$ ($SD$), than the mean in the SRS study. In order to consistently calibrate scoring, this study consulted the science and engineering standards literature to identify the authentic performance opportunities to attend to when scoring work samples from each instructional unit. Consultation with a researcher in the CORS project confirmed that this approach to adapting the scoring guides to each assessment task was the same approach the SRS used to achieve reliability (John Balwit, April 24, 2013, personal communication).

It is reasonable to believe this study’s use of its adaptation of the CORS scoring guides is calibrated similarly to scoring in the SRS, but that can’t be known. Both studies measured work samples from a range of assessment task formats in schools pursuing similar reform goals. The mean student performance levels measured in this study were likely to be as high as or higher than those measured in the SRS.
<table>
<thead>
<tr>
<th>Variable</th>
<th>Data Sources</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student behavioral engagement</td>
<td>Classroom observations of student behavior from October to June (17716 individual observations)</td>
<td>Behavioral Engagement Observation Protocol in Appendix D</td>
</tr>
<tr>
<td>Mean of percent behaviorally engaged per student for all students for entire study period; three school trimesters; seven instructional units.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student self-regulated learning</td>
<td>Classroom observations of student behavior from October to June (6787 individual observations)</td>
<td>Self-Regulated Learning Observation Protocol in Appendix D</td>
</tr>
<tr>
<td>Self-regulated learning presence (self-regulating/not self-regulating)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent individual observations self-regulating for entire study period; two school trimesters; four selected instructional units.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-regulated learning category</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(independent/autonomous)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent of individual observations with evidence of autonomy for two trimesters; four selected instructional units.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Authenticity of student performance in science</td>
<td>Student work samples from four instructional units (14-34 work samples each unit)</td>
<td>Student work sample scoring guides in Appendix F</td>
</tr>
<tr>
<td>Scientific analysis (Construction of Knowledge)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Core science concepts (Science Meaning-Making Processes)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elaborated communication (Science Meaning-Making Processes)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Five of these units included in repeated-measures analysis.

* Component of authenticity in this study’s model of Authentic Pedagogy and Performance measured by this indicator.
Table 6.2 Description of the assessment tasks corresponding to the student work samples analyzed for authenticity of student performance.

<table>
<thead>
<tr>
<th>LT(s)</th>
<th>Assessment task(s)</th>
<th>Description of assessment task</th>
<th>Format of assessment task</th>
<th>No. of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.6</td>
<td>RNA Fingerprinting Lab Analysis (2 days)</td>
<td>Describe how electrophoresis produces a RNA fingerprint &amp; use molecular mechanisms to interpret lab results.</td>
<td>Open response items including written presentation of data &amp; interpretation integrated requiring elaborated theoretical support.</td>
<td>7</td>
</tr>
<tr>
<td>1.6</td>
<td>RNA Fingerprinting Quiz (1 day)</td>
<td>Describe how electrophoresis produces a RNA fingerprint &amp; use molecular mechanisms interpret a contrived fingerprint.</td>
<td>Mixture of short answer and open response items including data interpretation and theoretical explanation.</td>
<td>7</td>
</tr>
<tr>
<td>1.4</td>
<td>Immune System PSA &amp; Flu Vaccines PSA (4 days for both parts)</td>
<td>Describe how the chemical &amp; cellular components of the immune system work together to fight the influenza virus. Describe how the influenza vaccine works &amp; evaluate the pros &amp; cons of getting an influenza vaccine.</td>
<td>Choice of creative PSA formats, such as, poster or cartoon sequence. Mostly description with limited theoretical explanation.</td>
<td>20</td>
</tr>
<tr>
<td>1.4</td>
<td>Immune System Quiz &amp; Flu Vaccines Quiz (2 days for both parts)</td>
<td>Describe how the chemical &amp; cellular components of the immune system work together to fight the influenza virus. Describe how the influenza vaccine works &amp; evaluate the pros &amp; cons of getting an influenza vaccine.</td>
<td>Mostly short answer items and an open response items.</td>
<td>1</td>
</tr>
<tr>
<td>3.2</td>
<td>Energy Concepts Quiz 1 &amp; Energy Concepts Quiz 2* (1 day)</td>
<td>Identify energy transformations in familiar situations and use these to illustrate conservation of energy and entropy.</td>
<td>Mixture of short answer and open response items requiring some written conceptual support.</td>
<td>32 and 2</td>
</tr>
<tr>
<td>3.5, 3.6, 3.7, 3.8</td>
<td>Wind Turbine Engineering &amp; Testing (most of the 17 days)</td>
<td>Use an engineering design process to brainstorm, prototype, optimize &amp; test a simple wind turbine from available materials.</td>
<td>Written engineering design notes; and open response items including presentation of testing data and evaluation of design performance.</td>
<td>27</td>
</tr>
</tbody>
</table>

Note. The instructional units are listed in chronological order. The first two units were during the fall trimester and the last two were in the spring. The second of two tasks addressing the same Learning Targets (LT) is the alternative task provided as an option to students. The teacher’s Learning Targets are found in Appendix A.
Table 6.3 Elements of authentic science and engineering performance measured in student science work samples.

<table>
<thead>
<tr>
<th>Component of Intellectual Authenticity</th>
<th>Performance Authenticity Indicator (PAI)</th>
<th>Evidence of Authentic Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reasoning includes Construction of Knowledge</td>
<td>Scientific or engineering analysis (PAI 1)</td>
<td>Student performance demonstrates organizing and interpreting information with scientific content during activities such as, predicting, experimenting, representing data, and constructing scientific arguments or engineering designs.</td>
</tr>
<tr>
<td>Use of Science Meaning-Making Processes</td>
<td>Core science or engineering concepts (PAI 2)</td>
<td>Student performance demonstrates an understanding of fundamental scientific ideas, concepts, or theories used by competent adults by explicitly using them in tasks such as, explaining phenomena, creating a design, or addressing a societal problem.</td>
</tr>
<tr>
<td></td>
<td>Elaborated communication (PAI 3)</td>
<td>Student performance demonstrates successful communication of an explanation, design, or argument and elaborated support for it. The form and content of the communication resembles forms of communication used by science practitioners.</td>
</tr>
<tr>
<td>Combined Authentic Performance score</td>
<td>Sum of all indicators (3-12)</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* All three of the PAIs are scored on a 1-4 scale from no to high authenticity.
Table 6.4 Mean and median percent of observations behaviorally engaged (BE) per student for each school trimester for all students and divided by gender.

<table>
<thead>
<tr>
<th>School Trimester</th>
<th>Mean % BE/student</th>
<th>Median % BE/student</th>
<th>No. of students</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>M</td>
<td>F</td>
<td>All</td>
</tr>
<tr>
<td>Fall</td>
<td>76</td>
<td>74</td>
<td>78</td>
<td>79</td>
</tr>
<tr>
<td>Winter</td>
<td>73</td>
<td>74</td>
<td>73</td>
<td>75</td>
</tr>
<tr>
<td>Spring</td>
<td>74</td>
<td>76</td>
<td>72</td>
<td>75</td>
</tr>
</tbody>
</table>

*Note.* Only students that were observed at least 32 times in a trimester are included. Eighty-five students are included in at least one trimester. Seventy students are included in all three trimesters.

Table 6.5 Mean and median levels of percent of observations behaviorally engaged (BE) per student during seven instructional units from October to June in chronological order. Number of lesson segments (LSs) observed during each unit is included.

<table>
<thead>
<tr>
<th>Instructional Unit</th>
<th>Mean % BE/student</th>
<th>SD</th>
<th>Median % BE/student</th>
<th>No. of students</th>
<th>No. of LSs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fall Trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA Fingerprinting Lab</td>
<td>89</td>
<td>11</td>
<td>93</td>
<td>49</td>
<td>15</td>
</tr>
<tr>
<td>Immune System &amp; Flu Vaccines</td>
<td>68</td>
<td>18</td>
<td>70</td>
<td>105</td>
<td>48</td>
</tr>
<tr>
<td>Winter Trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human Characteristics Data Analysis</td>
<td>75</td>
<td>16</td>
<td>76</td>
<td>84</td>
<td>54</td>
</tr>
<tr>
<td>Teen Brain</td>
<td>66</td>
<td>22</td>
<td>69</td>
<td>81</td>
<td>57</td>
</tr>
<tr>
<td>Spring Trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy Concepts</td>
<td>79</td>
<td>17</td>
<td>84</td>
<td>73</td>
<td>33</td>
</tr>
<tr>
<td>Engineering Challenge 1</td>
<td>72</td>
<td>25</td>
<td>75</td>
<td>65</td>
<td>18</td>
</tr>
<tr>
<td>Wind Turbine Engineering &amp; Testing</td>
<td>75</td>
<td>16</td>
<td>79</td>
<td>80</td>
<td>69</td>
</tr>
</tbody>
</table>

*Note.* Students with at least seven observations from two days and four lesson segments are included.
Table 6.6 Percent of all observations recording the presence of any type of self-regulated learning (SR) during each trimester and the entire study period for all students and divided by gender.

<table>
<thead>
<tr>
<th>School Trimester</th>
<th>% SR</th>
<th>No. of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>M</td>
</tr>
<tr>
<td>Fall</td>
<td>72</td>
<td>68</td>
</tr>
<tr>
<td>Winter</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>Spring</td>
<td>53</td>
<td>56</td>
</tr>
<tr>
<td>Entire Study Period</td>
<td>54</td>
<td>55</td>
</tr>
</tbody>
</table>

Table 6.7 Percent of all observations recording the presence of any type of self-regulated learning (SR) for all observations and each gender during four selected units.

<table>
<thead>
<tr>
<th>Instructional Unit</th>
<th>% SR</th>
<th>No. of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>M</td>
</tr>
<tr>
<td>Winter Trimester</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teen Brain</td>
<td>42</td>
<td>39</td>
</tr>
<tr>
<td>Spring Trimester</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy Concepts</td>
<td>36</td>
<td>39</td>
</tr>
<tr>
<td>Engineering Challenge 1</td>
<td>43</td>
<td>47</td>
</tr>
<tr>
<td>Wind Turbine Engineering &amp; Testing</td>
<td>60</td>
<td>63</td>
</tr>
</tbody>
</table>

Note. Gender proportions in each instructional unit are within 4% of those in the entire student sample. Instructional units are listed in chronological order.
Table 6.8 Percent of observations (N) recording the presence of any type self-regulated learning (SR) during subsets of observations from selected instructional units and the entire study period. The first subset of observations is from lesson segments when the class activity has all students grouped as a whole class. The second subset is when students were working in small groups or individually.

<table>
<thead>
<tr>
<th>Instructional Unit</th>
<th>% SR</th>
<th>No. of Observations</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>All</td>
<td>Grouped as whole class</td>
<td>Working in small groups or individually</td>
</tr>
<tr>
<td></td>
<td></td>
<td>All</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winter Trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teen Brain</td>
<td>42</td>
<td>37</td>
<td>44</td>
<td>650</td>
<td>189</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spring Trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy Concepts</td>
<td>36</td>
<td>18</td>
<td>54</td>
<td>287</td>
<td>142</td>
</tr>
<tr>
<td>Engineering Challenge 1</td>
<td>43</td>
<td>4</td>
<td>46</td>
<td>341</td>
<td>24</td>
</tr>
<tr>
<td>Wind Turbine Engineering &amp; Testing</td>
<td>60</td>
<td>19</td>
<td>63</td>
<td>2504</td>
<td>165</td>
</tr>
<tr>
<td>Entire Study Period</td>
<td>54</td>
<td>22</td>
<td>60</td>
<td>6787</td>
<td>1167</td>
</tr>
</tbody>
</table>

Note. There are a few lesson segments without social grouping classification. Therefore, the subsets of observations at the entire study period level have lower N’s than without classification by social grouping. Instructional units are listed in chronological order.
Table 6.9 Percent of self-regulated learning category variable recorded as working autonomously measured during selected instructional units.

<table>
<thead>
<tr>
<th>Instructional Unit</th>
<th>% SR working autonomously</th>
<th>No. of observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Winter Trimester</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teen Brain</td>
<td>8</td>
<td>264</td>
</tr>
<tr>
<td>Spring Trimester</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy Concepts</td>
<td>26</td>
<td>101</td>
</tr>
<tr>
<td>Engineering Challenge 1</td>
<td>9</td>
<td>140</td>
</tr>
<tr>
<td>Wind Turbine Engineering &amp; Testing</td>
<td>12</td>
<td>1455</td>
</tr>
</tbody>
</table>

*Note.* Instructional units are listed in chronological order. The total number of observations is 1960.
Table 6.10 Number of work samples (N) scored from each demographic group for each the assessment tasks from the four selected instructional units.

<table>
<thead>
<tr>
<th>Assessment task</th>
<th>Gender</th>
<th>Ethnicity</th>
<th>Mean 8th grade math and science grades</th>
<th>Total N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>F</td>
<td>AH</td>
<td>HL</td>
</tr>
<tr>
<td>Fall Trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA Fingerprinting Lab Quiz</td>
<td>2</td>
<td>5</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>RNA Fingerprinting Lab Analysis</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Immune System &amp; Flu Vaccine PSAs</td>
<td>5</td>
<td>15</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Immune System &amp; Flu Vaccine Quizzes</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Spring Trimester</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy Concepts Quizzes</td>
<td>14</td>
<td>20</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Wind Turbine Engineering &amp; Testing</td>
<td>8</td>
<td>19</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Total from All Work Samples</td>
<td>30</td>
<td>66</td>
<td>12</td>
<td>17</td>
</tr>
<tr>
<td>% of All Work Samples</td>
<td>31</td>
<td>69</td>
<td>13</td>
<td>19</td>
</tr>
<tr>
<td>% of Demographic group in whole 9th grade</td>
<td>40</td>
<td>60</td>
<td>11</td>
<td>19</td>
</tr>
</tbody>
</table>

Note. The total column includes work samples scored, but for which ethnicity group or prior achievement level was not available. The ethnic category columns labels are AH - Asian or native Hawaiian; HL - Hispanic or Latino; MC - Multiple racial categories selected; OR - Other race; and EA - Euro-American. Instructional units are listed in chronological order.
Table 6.11 Mean authenticity of student science performance measurements for work samples collected from four selected instructional units. For each instructional unit, means for the three Performance Authenticity Indicators (PAIs) and the Combined Authenticity are listed.

<table>
<thead>
<tr>
<th>Instructional Unit</th>
<th>Mean Authenticity</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PAI 1</td>
<td>PAI 2</td>
</tr>
<tr>
<td><strong>Fall Trimester</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA Fingerprinting Lab</td>
<td>.79</td>
<td>.76</td>
</tr>
<tr>
<td>Immune System &amp; Flu Vaccines</td>
<td>.53</td>
<td>.52</td>
</tr>
<tr>
<td><strong>Spring Trimester</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Energy Concepts</td>
<td>.34</td>
<td>.39</td>
</tr>
<tr>
<td>Wind Turbine Engineering &amp; Testing</td>
<td>.83</td>
<td>.61</td>
</tr>
<tr>
<td>Means of the means for each instructional unit</td>
<td>.62</td>
<td>.57</td>
</tr>
</tbody>
</table>

*Note.* Instructional units are in chronological order.
Figures

Figure 6.1 Boxplots of fraction of observations behaviorally engaged per student for each class period from each trimester.
Figure 6.2 Percent of all individual observations recording the presence of any type of self-regulated learning during four instructional units selected for observation of a wide variety of the classroom activities during each unit.
Figure 6.3 Percent of self-regulated learning (SR) category variable observations recorded as working autonomously for each gender during two trimesters.

Figure 6.4 Boxplots of student work sample Combined Authenticity scores for four instructional units. All of the authenticity scores are normalized to a scale of 0-1.
Figure 6.5 Mean combined authentic performance for four instructional units for each gender. All of the authenticity scores are normalized to a scale of 0-1. Error bars are ± 1 SD. The numbers of males (M) and females (F) are listed for each instructional unit.
Figure 6.6 Mean Combined Authenticity for assessment tasks and student performance on work samples from four selected instructional units.
Endnotes

1 Observations recorded as ambiguous were 1.6% of the total 18,508 observations. Students were recorded as having left the classroom were an additional 1.6% of total observations. A further 1.1% of observations were coded as missing because the student couldn’t be observed without interrupting instruction or influencing student behavior.

2 Ninety percent of behavioral engagement observations were from scans that recorded an observation of every student present. Infrequently, scans of the entire class were interrupted by the end of a lesson segment or a student question because the researcher made himself available as a resource at student request to develop comfort and trust in his presence. Also, this behavioral engagement observation protocol couldn’t be used during video viewing if the room was darkened.

3 The number of behavioral engagement and self-regulated learning observations per lesson segment varied because the length of lesson segments varied from a few minutes to 75 minutes. The number of observations per student varied by lesson segment, by trimester, and over the entire study period. The variation in the length of lesson segments, numbers of observations from each class period, student absences, and students transferring into or out of the school during the study period contributed to this variation in the numbers of observations per student.

4 Fall trimester percent behaviorally engaged per student was not considered because the number of lesson segments observed was less than half the numbers observed in winter and spring. In addition, some students were only enrolled during two trimesters. Only one student was measured behaviorally engaged at least 90% of observations in all three trimesters.

5 The number of students in the samples for the different instructional units varies substantially because the length of the units varies by a factor of three, from 5 class days to 17 class days. In short instructional units students were not always observed enough to meet this minimum determined to represent a variety of lessons during that unit. The RNA Fingerprinting Lab Analysis, Energy Concepts, and Engineering Challenge 1 units were only a week long. The four other instructional units were from two to three weeks long. The data screen used affects sample size during the three short instructional units. The screen only removed students with a large number of absences from samples for the four longer units. The RNA Fingerprinting Lab Analysis sample is much smaller than those of other units. It includes only 49 students, because only three class days were observed and only class periods 2 and 3 were observed more than once. The Immune System & Flu Vaccines unit sample of students is the largest, 105 students, because it includes observations of period 1 students, in addition to the observations of periods 2, 3, and 5 observed during the other six units.
The criteria for inclusion did not include a minimum number of observations per student for each level of analysis used for behavioral engagement. Because of the smaller numbers of self-regulated learning observations, the criterion used for behavioral engagement at the trimester level would remove all the students in fall and winter trimesters and more than a quarter of the students in the third trimester. At the instructional unit level, the behavioral engagement criterion would similarly reduce the number of students in even the longer units by two-thirds to three-quarters, except for the Wind Turbine Engineering & Testing unit.

Sixty-five percent of self-regulation presence variable observations are from the spring, 27% from the winter and 8% from the fall.

Availability was the greatest factor the limiting number of work samples scored in the fall trimester. In most cases this investigator needed to be present when work samples were turned in to the teacher to have an opportunity to make photocopies of work samples before the teacher needed them for grading. Not all work samples were turned in on the same days, so this limited the ability to collect samples in some cases.

Availability of work samples limited the number of students in the sample with a work sample from all four instructional units to five. Samples from at least three of the four units were scored for 17 different students. The entire sample includes work samples from 40 different students.

The curriculum included both science and engineering content and both are considered in the specification of assessment task and classroom instruction scoring guides. For brevity in the text Science Meaning-Making Processes is used, but where engineering content and skills are relevant, those are as central in the scoring as science is when it is the central focus.

Due to an illness, one of the independent scorers was not able to complete a score for one indicator on one work sample.

The normalized score was calculated from: (component score – low bound of score range)/(high bound of score range – low range)/(high bound of score range – low bound of score range). A one on this scale requires that all three PAIs are scored at their highest authenticity (4).

There were only five students that consented to use of these records in the study and work samples were collected from only one of them, but only for two instructional units.

The nonparametric test is used because there were outliers that there was no reason to exclude and none of the trimesters had a normal distribution of behavioral engagement measures.

Mann-Whitney U tests for each trimester found no significant differences at the 0.05 significance level.
One-way ANOVA tests were conducted to test the null hypothesis that the mean behaviorally engaged levels for periods 2, 3 and 5 were equal during each trimester independently. One outlier was removed. In fall and winter trimesters the null hypotheses were not rejected. In spring trimester the null hypothesis is rejected, $F(2, 76) = 4.297$, $p < .017$. Tukey post hoc analysis revealed that the mean behaviorally engaged level in period 3 was statistically significantly greater than the mean behaviorally engaged level in period 5 ($0.12, 95\% \text{ CI} [0.02, 0.22], p = .012$). This is the only statistically significant difference found.

Between the fall and winter trimester six students moved into the afternoon period (5) and seven students moved out of period 5 into morning periods. Most of the students that moved into period 5 decreased behavioral engagement (percent behaviorally engaged/trimester) and most of the students that moved out of period 5 increased their behavioral engagement. Fall to winter differences on average are about 10% for both groups of students.

A procedure for controlling experimentwise Type I error rate by applying a stricter standard for significance.

The smaller sample in fall trimester is 65% female.

The 72%, $N=544$, self-regulated learning measure in fall trimester seen in Table 6.6 doesn’t meet the criterion of sampling a wide enough variety of lessons to be considered representative of the fall trimester.

Analysis of the influence of student grouping during classroom activities on behavioral engagement rates shows that the influence is small.

In the RNA Fingerprinting Lab unit, half of the work samples in the set scored were from the quiz version of the assessment task. Those work samples were from about three weeks later than the end of the unit when the lab analysis work samples were collected. The quiz was offered to the students only at that later time.
Chapter 7  Relationships between authentic pedagogy and student performance, engagement, and self-regulated learning

This study finds elements of authenticity in the pedagogy of the teacher and in student performance in this case as documented in Chapters 5 and 6. In addition, student behavioral engagement levels are shown to be relatively high throughout the year and to vary somewhat between instructional units. Evidence of student self-regulated learning is also shown to be present and there is evidence of variation in self-regulation between instructional units.

The second broad objective of this study is to explore whether there is evidence for gains in science achievement, engagement, and self-regulated learning when students experience authentic pedagogy, as defined by the model of authentic science pedagogy and performance adopted for this study. That model holds potential for pointing out a unique contribution of environmental education to current concerns for improving STEM education.

In this Chapter the relationships between authentic pedagogy and student responses to the aspects of authenticity found to be present in their classroom environment are explored further through correlation analyses. The relationships between two aspects of authentic science pedagogy, authenticity in the teacher’s assessment tasks and in classroom instruction, and student behavioral engagement and self-regulated learning are examined. Then the relationships between those two aspects of authentic science pedagogy and the authenticity of student performance on science work samples from the classrooms are explored.

225
The common model of authenticity used for all the authenticity variables measured allows comparisons of components of authenticity between the teacher’s pedagogy and student performance. Whether there are relationships found that indicate the model could contribute to measuring the value of environmental education is considered.

Research Question 3 and two propositions guide the research reported in this chapter.

Research Question 3

How are levels of authentic student performance, engagement, and self-regulated learning related to the features of authentic science instruction?

Proposition 1

The level of authentic science pedagogy students experience will be positively related to student behavioral engagement and self-regulation.

Proposition 2

The level of authentic science pedagogy students experience will be positively related to the level of authentic student performance.

Clarity in the methodology will be improved by considering the two aspects of authentic pedagogy separately.

2.1: The level of authenticity in the assessment tasks students respond to will be positively related to the level of authentic student performance on science work samples.
2.2: The level of authenticity in classroom instruction students experience will be positively related to the level of authentic student performance on science work samples.

7.1 Methods

7.1.1 Proposition 1 - Relationships between authenticity of classroom instruction and student behavioral engagement and self-regulated learning

7.1.1.1 Data Sources

The assertion in Proposition 1 that student behavioral engagement and self-regulation levels would be positively related to the authenticity of the teacher’s classroom instruction is investigated by variables measured at the lesson segment level (Table 7.1). The correlation analyses for the behavioral engagement and the self-regulated learning presence variables were conducted independently.

Lesson segments that were scored for authenticity of classroom instruction, required at least 19 observations of behavioral engagement and category self-regulation to be included in this analysis.\(^1\) The additional criterion for inclusion was that students were working in small groups or independently during the lesson segment.\(^2\)

The sample for behavioral engagement includes 170 lesson segments comprising 90 hours of class time from 14 of the 16 instructional units observed during the study period (Table G.1). The sample for self-regulated learning includes 94 lesson segments comprising 58 hours of class time from 13 instructional units (Table G.1). All of the
lesson segments in the analytic sample for self-regulated learning are also in the analytic sample for behavioral engagement.³

7.1.1.2 Measures included in the analyses

The percent of observations recorded as behaviorally engaged per lesson segment and percent of observations recording the presence of any type of self-regulated learning per lesson segment are the measures used for these student response variables (Table 7.1). Classroom instruction authenticity measures include the 11 Lesson Authenticity Indicators (LAIs) scored during each of the lesson segments, as described in Chapter 5 (Table 5.6). In addition, the Combined Authenticity and the four authenticity component scores that were calculated from the LAI scores coded for each of the lesson segments (Table 5.7) are also included.⁴

7.1.1.3 Data Analyses

To examine whether the authenticity of classroom instruction is positively correlated with student behavioral engagement and self-regulated learning, the Pearson correlation coefficients were first determined for percent behaviorally engaged and the classroom instruction authenticity measures from each of the 170 lesson segments in the analytic sample. The Pearson correlation coefficients were also determined for percent self-regulating and the classroom instruction authenticity measures from each of the 94 lesson segments in this analytic sample.
7.1.2 Proposition 2 – Relationships between authentic pedagogy and authentic student performance

7.1.2.1 Data Sources

The set of 96 student work samples analyzed in Chapter 6 are included in this analysis (Table 7.2). They come from four instructional units, the RNA Fingerprinting Lab and the Immune System and Flu Vaccines units from the fall trimester; and Energy Concepts and Wind Turbine Engineering and Testing units in the spring trimester. The numbers of work samples analyzed vary for each instructional unit (Table 6.10).

7.1.2.1.1 Proposition 2.1 – The teacher’s assessment tasks and student performance

The assessment task documents and learning targets³ from these four instructional units are from those analyzed for authenticity and reported in Chapter 5. The two assessment tasks from the RNA Fingerprinting Lab unit, the two tasks from the Immune System and Flu Vaccines unit, the two tasks from the Energy Concepts unit, and the single task from the Wind Turbine Engineering and Testing are included in this analysis (Table 6.2).

Only one student work sample was analyzed that was a response to the quiz version of Immune System and Flu Vaccines unit assessment task. By contrast, 20 work samples were scored for Public Service Announcement (PSA) version of assessment task from that unit. Therefore, a weighted average of each of the four assessment task authenticity scores from Table 5.9 is used in the analysis here. In addition, the two assessment tasks from the Energy Concepts unit are very similar and have the same authenticity score, so they are treated as the same in the analysis here, as they were in
Chapter 6. Therefore, the analysis is conducted with five assessment task scores, two from the RNA Fingerprinting Lab unit, and one each from the other three units.

7.1.2.1.2 Proposition 2.2 – The teacher’s classroom instruction and student performance

The 96 student work samples in this analysis were student responses to the same four instructional units. These instructional units were also analyzed in Chapter 5 for authenticity of classroom instruction (Table 5.12). Classroom instruction authenticity measures from the 202 lesson segments scored from these four units were used in this analysis.6

7.1.2.2 Measures included in the analyses

The Combined Authenticity of student performance and each of the Performance Authenticity Indicators (PAIs) from each of the individual student work samples described in Chapter 6 are used (Table 6.3). PAI 1, Scientific analysis, corresponds to the Construction of Knowledge component of authenticity. PAI 2, Core science concepts, and PAI 3, Elaborated communication, together represent the Science Meaning-Making Processes component (Table 7.2). Therefore, the average of PAI 2 and PAI 3 is used to measure the Science Meaning-Making Processes component of authenticity of each work sample. These five authenticity variables from each of the 96 student work samples are analyzed to answer Propositions 2.1 and 2.2 (Table 7.2).
7.1.2.2.1 Proposition 2.1 – The teacher’s assessment tasks and student performance

The means of the Combined Authenticity and the three authenticity component scores for each of the five assessment tasks are used for these analyses. These authenticity components of the assessment tasks are Construction of Knowledge, Science Meaning-Making Processes, and Value Beyond School (Table 7.2).

In addition, the mean Combined Authenticity of student performance for the sets of work samples corresponding to each of the five assessment tasks are used in the analyses.7

7.1.2.2.2 Proposition 2.2 – The teacher’s classroom instruction and student performance

The means for the Construction of Knowledge, Science Meaning-Making Processes, Substantive Conversation, and Value Beyond School components of the authenticity of classroom instruction, and the mean Combined Authenticity from each of the four instructional units are used in this analysis (Table 7.2).8 The classroom instruction variable means are those documented in Chapter 5 (Table 5.12, Table 5.13, and Figure 5.4).

In addition, the mean Combined Authenticity of student performance for the sets of work samples corresponding to each of four instructional units are also used in a separate correlation (Table 7.2).9
7.1.2.3 Data Analyses

7.1.2.3.1 Proposition 2.1 – The teacher’s assessment tasks and student performance

The Pearson correlation coefficients were determined for the associations between the three PAIs, the Science Meaning-Making component of authenticity, and the Combined Authenticity from each student work sample (Table 7.2) and the three components of authenticity and the Combined Authenticity for each of the six corresponding assessment tasks (Table 7.2). The correlation matrix was reviewed for correlations that were statistically significant and for their relative strengths.

In addition, the Pearson correlation coefficient was determined for the associations between the mean Combined Authenticity of performance for the work samples from each of the five corresponding assessment tasks and the Combined Authenticity of the assessment tasks. The paired values were plotted to describe the relationship.

7.1.2.3.2 Proposition 2.2 – The teacher’s classroom instruction and student performance

The Pearson correlation coefficients were determined for the associations between the three PAIs, the Science Meaning-Making component of authenticity, and the Combined Authenticity from each student work sample (Table 7.2) and the mean Combined Authenticity and four mean component authenticities for each of the four corresponding instructional units. The correlation matrix was reviewed for correlations that were statistically significant and for their relative strengths.
In addition, the Pearson correlation coefficient was determined for the associations between the mean Combined Authenticity of performance for the work samples from each of the four instructional units with the six corresponding mean combined classroom instruction authenticities for the four instructional units and paired values were plotted to describe the relationship.

### 7.2 Results

#### 7.2.1 Proposition 1 - Relationships between authenticity of classroom instruction and student behavioral engagement and self-regulated learning

The Pearson correlation coefficients were determined for measures of the authenticity of classroom instruction and behavioral engagement and self-regulated learning measures to examine whether the level of authenticity of the pedagogy of the teacher in this case was associated with these student behavioral responses in the classroom.

The correlations (r) between the Combined Authenticity of classroom instruction and of each of its authenticity components and the percent behaviorally engaged per lesson segment are positive, but negligible or small, except for the Construction of Knowledge, which is negative (Table 7.3). The only authenticity component that is statistically significantly correlated is Science Meaning Making Processes (Table 7.3). The correlation (r = 0.175) indicates a small effect size.

Scores for each of the components of authenticity of the teacher’s classroom instruction were calculated from 11 Lesson Authenticity Indicators (LAIs) scored during each lesson segment. None of the 11 LAIs accounted for more than a small amount of
the variation in behavioral engagement levels and most of the correlation coefficients suggest a negligible effect size within this sample (Table 7.4).

The results show a small positive association \((r = 0.162, p = .034)\) of collaborative discourse, LAI 6, with percent behaviorally engaged per lesson segment (Table 7.4). However, LAI 5, sustained discourse, was not correlated with behavioral engagement (Table 7.4). Both LAI 5 and 6 scoring guides (Appendix B) include a requirement that most students are participating in lesson-related discourse to score high. This also requires students to be behaviorally engaged, accounting for some of the correlation observed in LAI 6. However, the difference between the correlations for LAI 5 and LAI 6, indicates it is possible that discourse that encouraged students to share their thinking on science topics kept students on task to a greater degree than discourse that was sustained but not as focused on understanding.

The correlation between Science Inquiry and Engineering Design Practices, LAI 11, and behavioral engagement is small \((r = 0.165, p = .031)\), but significant (Table 7.4), as it is for the Science Meaning Making Processes component (Table 7.3). LAI 11 is a subcomponent of Science Meaning Making Processes. It is possible that students find that physical interactions with natural phenomena during the lesson segments that scored higher for LAI 11 play a role in the correlations measured.

The correlations between aspects of the authenticity of the teacher’s classroom instruction and the percent self-regulated learning per lesson segment are positive and stronger than for behavioral engagement (Table 7.3). The Combined Authenticity of classroom instruction has a small significant correlation with percent of observations recording any type of self-regulated learning \((r = 0.240, p = .02)\). The authenticity level
of the Science Meaning Making Processes component has a medium-sized significant
correlation ($r = 0.303, p \leq .01$) with this self-regulated learning variable. The
Construction of Knowledge component has only a negligible correlation (Table 7.3). The
authenticity level of the Value Beyond School component has a small positive correlation
($0.204, p = .049$) with percent self-regulated learning.

Only two of the LAIs show significant correlations with percent of self-regulated
learning observed. The level of Focus on core scientific concepts or practices, LAI 2, had
a medium-sized positive association ($0.436, p \leq .01$) with self-regulation level (Table
7.4). LAI 2 accounts for half of the Science Meaning Making Processes component
score.

Connections to the world beyond the classroom, LAI 8, also shows a small
significant positive correlation ($0.233; p = .024$) with percent student self-regulation
observed (Table 7.4). When the influence of Focus on core scientific concepts or
practices is controlled for a significant association remains ($r=0.218, p = .036$).

Presentation of classwork to an audience beyond the classroom, LAI 9, is not
correlated with self-regulation. As described in Chapter 5, LAI 9 was the lowest scoring
of any of the LAIs (Table G.3) and had the lowest degree of variation of any of the LAIs.
Therefore, the positive association of LAI 8 was the reason the Value Beyond School
component of authenticity was positively correlated with student self-regulated learning.
7.2.2 Proposition 2 - Relationships between mean authentic pedagogy variables and student performance on work samples

The Pearson correlation coefficients were determined for measures of two aspects of the authenticity of the teacher’s pedagogy (i.e., classroom instruction and assessment tasks students completed), and the authenticity of student performance in their work samples that were responses to the corresponding assessment tasks. All of the correlations were determined to be statistically significant at the 0.01 level.

The authenticity of student performance on work samples correlates positively with both the authenticity of teacher’s assessment tasks (Table 7.5) and classroom instruction (Table 7.6). The degrees of association calculated are similar for the Combined Authenticity of the assessment tasks and mean Combined Authenticity of classroom instruction. The Construction of Knowledge and Science Meaning-Making Process components of authenticity are also positively correlated with their matching component scores for student performance for both the teacher’s assessment tasks (Table 7.5) and classroom instruction (Table 7.6).

The Value Beyond School component of the teacher’s assessment tasks and classroom instruction isn’t measured as a component of the authenticity of student performance. However, the Value Beyond School components of both pedagogy variables are found to be positively correlated with the Combined Authenticity of student performance (Table 7.5 and 7.6). Similarly, the authenticity of the Substantive Conversation component of classroom instruction is positively correlated with the Combined Authenticity of student performance (Table 7.6).
There are relatively small differences between the measured strengths of the associations of the two authenticity components of student performance and the corresponding components of either the teacher’s assessment tasks (Table 7.5) or the classroom instruction (Table 7.6). All of the correlations discussed in this section measure meaningful effects, but the meaning is limited to showing that about a quarter or less of the variance in authenticity of student performance is explained by the similarly defined authenticities of the classroom instruction and assessment tasks.

The mean Combined Authenticity of student performance for work samples completed for each assessment task is also found to be positively related to the Combined Authenticity of the assessment tasks, but it is not statistically significant (Figure 7.1). A similar analysis for mean Combined Authenticity of student performance on work samples from each instructional unit and the corresponding mean Combined Authenticity of classroom instruction for the four units found a positive relationship, but it is not statistically significant (Figure 7.2).

### 7.2.3 Summary

Recognizing the limits of the current study’s convenience samples and analyses, some meaningful associations were found that apply to this case and point to fruitful further investigation. This study doesn’t find any meaningful relationships between classroom instruction authenticity variables and student behavioral engagement in the analyses done at the lesson segment level. The analysis conducted indicates only a small amount of the variation in percent behavioral engagement between lesson segments is accounted for by the classroom instruction authenticity variables measured during each
lesson segment. These results show that the influence of the specific aspects of authentic pedagogy measured could not be separated from influence of other not included factors by this analysis.

Two of the components of authentic classroom instruction, Science Meaning-Making Processes and Value Beyond School were positively associated with the percent of observations recorded as self-regulating. Connections to the world beyond the classroom, LAI 8, and Focus on core science concepts, LAI 2, are the most important contributors to these relationships and the correlation between Combined Authenticity of student performance and self-regulated learning.

Although these analyses can’t specify the magnitude of the associations, consistent positive correlations between all authentic characteristics in this study’s model of authentic pedagogy for the teacher’s assessment tasks and classroom instruction and the authenticity of student performance on both authenticity components are found. This suggests these relationships are meaningful. When the teacher’s pedagogy afforded and expected authentic performance, students were more likely produce authentic science work.

The correlation coefficients and statistical significances calculated in this study are limited by some characteristics of the analysis and the data. The correlation results don’t account for the lack of independence in the data. For example, two aspects of the student work sample scores are not accounted for. First, these students are in the same classrooms. Some of them work together on classroom activities, such as, building and testing a wind turbine during the final instructional unit of the school year. In addition,
some work samples in these analyses were collected from the same students. Because of both of these conditions, the work sample scores are not independent measures.

These analyses also don’t take into account the complex dimensions of doing this kind of naturalistic data collection. For example, there are potential contributions to variations in the data arising from differences between the composition of the three class periods, influences from the sequences of science units as the year progresses, and other factors arising from the multilevel nature of the data (Chan, 2006). While all students were taught by the same teacher, they were grouped into three class periods.

Type I error is inflated in the analyses using all 96 work samples correlated with the authenticity variables measured for five assessment tasks (Table 7.5) and four instructional units (Table 7.6). The analyses of the relationships between student behavioral engagement and self-regulated learning and classroom instruction authenticity measures were less affected by multilevel influences because both were measured for each lesson segment, but may have been influenced by class period effects.

Therefore, these correlation analyses can only be considered qualitative conclusions. And they only describe the samples of students and lesson segments in the analyses. They can’t be generalized to the entire population of four ninth grade science classrooms taught by the teacher in the study.
7.3 Discussion

This chapter uses the instruments defined by the model of authentic science pedagogy and performance adopted by this study, to ask if students showed greater engagement with learning and did more authentic intellectual work when the pedagogy was more authentic. The findings apply to the samples of students and class time from one teacher’s classrooms included in these analyses.

7.3.1 Authentic pedagogy and student behavioral engagement

The results show that student behavioral engagement did not consistently increase when the teacher’s classroom instruction was more authentic using the authenticity measures in this study (Table 7.3). The correlations were weak (Table 7.3) and difficult to interpret. The small or negligible effect sizes indicate that other factors not captured by the authenticity measurements may have greater influence on whether students are behaviorally engaged during a lesson segment than the authenticity of the teacher’s instruction. The teacher’s positive relationships with students may be one of those influences (Klem & Connell, 2004).

Inspection of the percent behavioral engagement per lesson segment values shows a large degree of variation. Percent behavioral engagement per lesson segment varies from a minimum of 24% per lesson segment to 100% behaviorally engaged per lesson segment. Further inspection of the bottom decile of lesson segments for this variable shows a wide range of possible reasons for lower percent engagement. For example, some lesson segments were at the end of an instructional unit, when this investigator’s field notes recorded that many students had completed the project earlier in the class.
period or even the previous day. This condition isn’t directly taken into account by the observation protocol in this study.

In Chapter 6 it was shown that, by trimester, mean percent behaviorally engaged per student remained about 75% with only small variations. At the instructional unit level mean percent behaviorally engaged per student varied from 66% to 79%, but the unit with the highest behavioral engagement had the lowest mean authenticity levels of any of the instructional units analyzed in Chapter 5 (Figure 5.5). The largest influence on student behavioral engagement is likely to be student selection of this school and the rapport a veteran teacher has with the students in his classes.

7.3.2 Authentic pedagogy and student self-regulated learning

Unlike behavioral engagement, student self-regulated learning at the lesson segment level was positively correlated with Focus on core science concepts and Connections to the world beyond the classroom (Table 7.3) in the teacher’s classroom instruction. These aspects of authentic classroom instruction were two of the most the common authentic features of the teacher’s instruction. The association of Connections to the world beyond the classroom, part of the Value Beyond School component of authenticity, with self-regulated learning is weak, but is a unique finding. The Value Beyond School component is the most distinctive aspect of this study’s model of authenticity (Rivet & Krajcik, 2008). The influence of this characteristic of instruction hasn’t be studied over time in a science classroom.

Although there are fewer lesson segments in the sample, the percent self-regulated learning observed per lesson segment in these analyses are a more representative measure
of self-regulated learning of the population than the analyses in Chapter 6. In Chapter 6 percent self-regulated learning was determined from all of the observations from an instructional unit or trimester. Percent self-regulated learning was measured as 63% during the Wind Turbine Engineering and Testing unit at the end of the school year, clearly indicating that students were frequently working independently or autonomously during an instructional unit at the end of the school year that gave students much time to be responsible for their own learning. Most of the lesson segments in the correlation analyses were also from the Wind Turbine unit. The associations found apply to class time when there was an appreciable amount of self-regulated learning observed.

The positive associations of Focus on core science concepts and Connections to the world beyond the classroom with self-regulated learning suggest that science instruction that includes these authentic characteristics may encourage high school science students to be more engaged with learning and take more initiative in the classroom. The unique role of the characteristics of the Value Beyond School component of authenticity in science pedagogy is worthy of further study, as is proposed in Chapter 9. In particular, its potential contribution to environmental education research will be discussed.

### 7.3.3 Authentic pedagogy and student authentic performance

While limited to the convenience samples analyzed in the current study, the generally moderate positive correlations of student authentic performance scores with the authenticity of the teacher’s assessment tasks and classroom instruction confirm an expected result. This teacher designed a considerable degree of authenticity into his
assessment tasks and classroom activities. Students responded by producing work with levels of authentic performance that increased as the authenticity level of the teacher’s expectations increased.

In this case, with a veteran teacher committed to a vision of science instruction and school culture aligned with the model of authentic science pedagogy and performance of this study, the level of student performance isn’t surprising. The results confer a degree of construct validity to the instrumentation used. This is limited by the fact that one investigator did nearly all of the scoring for the authenticity variables.

Further research is needed on the relative contributions of specific characteristics of authentic pedagogy to levels of student performance using this dataset. Investigation of the influence of class period is also important.

Further research is warranted into the responses of individual students as defined by their authentic performance, especially students that were less successful. Investigation of patterns in behaviors and dispositions of individual students that are associated with student success meeting the expectations of the teacher is needed.
### Tables

#### Table 7.1 Variables in the correlation analysis for Proposition 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Data Sources</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Student behavioral engagement</td>
<td>Classroom observations of student behavior from October to June (170 lesson segments)</td>
<td>Classroom Observation Protocol in Appendix D</td>
</tr>
<tr>
<td>Mean of percent of all observations recorded as behaviorally engaged during a lesson segment.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student self-regulated learning presence (self-regulating/not self-regulating)</td>
<td>Classroom observations of student behavior from October to June (94 lesson segments)</td>
<td>Classroom Observation Protocol in Appendix D</td>
</tr>
<tr>
<td>Mean of percent of all observations recorded as self-regulating during a lesson segment.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for authentic intellectual work in the teacher’s classroom instruction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eleven Lesson Authenticity Indicator scores</td>
<td>Classroom instruction field notes from October to June (170 lesson segments)</td>
<td>Lesson Authenticity Indicator scoring guides in Appendix B</td>
</tr>
<tr>
<td><em>Components of authenticity</em></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for Construction of Knowledge score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for use of Science Meaning-Making Processes score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances to find Value Beyond School score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for Substantive Conversation score</td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Combined Authenticity score</em></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7.2 Variables in the correlation analysis for Proposition 2.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Data Sources</th>
<th>Instrument</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authenticity of Student Performance in Science</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Components of authenticity of performance</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Construction of Knowledge score</td>
<td>Student work samples from four instructional units (14 to 34 work samples each unit, for a total of 96)</td>
<td>Student work sample scoring guides in Appendix F</td>
</tr>
<tr>
<td>Science Meaning-Making Processes score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Mean of Science Concepts, PAI 2, and Elaborated Communication, PAI 3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined Authenticity of performance score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for authentic intellectual work in Assessment Tasks (Proposition 2.1)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Components of authenticity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for Construction of Knowledge score</td>
<td>Five teacher-designed assessment tasks from four instructional units; Teacher Learning Targets for each unit; Task scoring guides provided to students by the teacher</td>
<td>Teacher Assessment Task Authenticity Indicators scoring guides in Appendix A</td>
</tr>
<tr>
<td>Affordances for use of Science Meaning-Making Processes score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances to find Value Beyond School score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined Authenticity score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for authentic intellectual work in the teacher’s Classroom Instruction (Proposition 2.2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Components of authenticity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for Construction of Knowledge score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for Use of Science Meaning-Making Processes score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances to Find Value Beyond School score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Affordances for Substantive Conversation score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined Authenticity score</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 7.3 Correlation coefficients (r) for classroom instruction Combined Authenticity and component authenticity variables and behavioral engagement (%BE/LS) and self-regulated learning (% presence of SR/LS).

<table>
<thead>
<tr>
<th>Classroom instruction variable</th>
<th>% BE/LS</th>
<th>% presence of SR/LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined Authenticity</td>
<td>.055</td>
<td>.240*</td>
</tr>
<tr>
<td>Construction of Knowledge component</td>
<td>-.126</td>
<td>.094</td>
</tr>
<tr>
<td>Science Meaning Making Processes component</td>
<td>.175*</td>
<td>.303**</td>
</tr>
<tr>
<td>Substantive Conversation component</td>
<td>.086</td>
<td>.156</td>
</tr>
<tr>
<td>Value Beyond School component</td>
<td>.043</td>
<td>.204*</td>
</tr>
</tbody>
</table>

*Note. The 170 lesson segments included in the analysis for behavioral engagement have a total duration of 90 hours, and the 94 lesson segments included for self-regulated learning have a total duration of 58 hours. Mean values for the correlation variables are in Appendix G, Table G.2.

*p < 0.05. **p < 0.01.
Table 7.4 Correlation coefficients (r) for classroom instruction Lesson Authenticity Indicator variables (LAI) and behavioral engagement (% BE/LS) and self-regulated learning (% presence of SR/LS). LAI are categorized by the authenticity component they contribute to. Only LAI with significant correlations are listed.

<table>
<thead>
<tr>
<th>Classroom instruction variables</th>
<th>% BE/LS</th>
<th>% presence of SR/LS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Construction of Knowledge component</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacognitive facilitation (LAI 7)</td>
<td>-.168*</td>
<td></td>
</tr>
<tr>
<td>Science Meaning Making Processes component</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Focus on core concepts (LAI 2)</td>
<td>.436**</td>
<td></td>
</tr>
<tr>
<td>Science Inquiry &amp; Engineering Design Practices (LAI 11)</td>
<td>.165*</td>
<td></td>
</tr>
<tr>
<td>Substantive Conversation component</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Collaborative discourse (LAI 6)</td>
<td>.162*</td>
<td></td>
</tr>
<tr>
<td>Value Beyond School component</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Connections to world beyond classroom (LAI 8)</td>
<td>.233*</td>
<td></td>
</tr>
</tbody>
</table>

Note. The 170 lesson segments included in the analysis for behavioral engagement have a total duration of 90 hours, and the 94 lesson segments included for self-regulated learning have a total duration of 58 hours. Complete correlation analysis results are in Appendix G, Table G.4. Mean values for the LAI variables are in Appendix G, Table G.3. *p < 0.05. **p < 0.01.
Table 7.5 Correlation coefficients (r) for authenticity variables for the teacher’s assessment tasks and authenticity variables for student performance on work samples.

<table>
<thead>
<tr>
<th>Authentication of <em>Assessment Task</em> variable</th>
<th>Authenticity of Student Performance on Science Work Samples</th>
<th>Correlations for matching Student Performance and <em>Assessment Task</em> components</th>
<th>Correlations for Combined Student Performance score with <em>Assessment Task</em> components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined Authenticity</td>
<td></td>
<td>.488**</td>
<td>.488**</td>
</tr>
<tr>
<td>Construction of Knowledge component</td>
<td></td>
<td>.550**</td>
<td>.344*</td>
</tr>
<tr>
<td>Science Meaning-Making Processes component</td>
<td></td>
<td>.382**</td>
<td>.512**</td>
</tr>
<tr>
<td>Value Beyond School component</td>
<td></td>
<td></td>
<td>.404**</td>
</tr>
</tbody>
</table>

*Note.* The number of work samples in the correlations is 96 in all cases (N = 96). Complete correlation analysis results are in Appendix G, Table G.6. Mean values for the LAI variables are in Appendix G, Table G.5.  

**p < 0.01.**
Table 7.6 Correlation coefficients (r) for authenticity of the teacher’s classroom instruction variables and authenticity of student performance on work samples. The classroom instruction variables are mean values for the four instructional units corresponding to the student work samples.

<table>
<thead>
<tr>
<th>Authenticity of Classroom Instruction variable</th>
<th>Authenticity of Student Performance on Science Work Samples</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correlations for matching Student Performance and Instruction components</td>
</tr>
<tr>
<td>Mean Combined Authenticity</td>
<td>.466**</td>
</tr>
<tr>
<td>Mean Construction of Knowledge component</td>
<td>.402**</td>
</tr>
<tr>
<td>Mean Science Meaning-Making Processes component</td>
<td>.414**</td>
</tr>
<tr>
<td>Mean Value Beyond School component</td>
<td></td>
</tr>
<tr>
<td>Mean Substantive Conversation component</td>
<td></td>
</tr>
</tbody>
</table>

Note. Mean classroom instruction values for each instructional unit are computed from lesson segment scores weighted by the duration of the lesson segment. The number of work samples in the correlations is 96 in all cases (N = 96). Complete correlation analysis results are in Appendix G, Table G.8. Mean values for the LAI variables are in Appendix G, Table G.7. **p < 0.01.
Figures

Figure 7.1 Relationship of mean Combined Authenticity of student performance on science work samples and Combined Authenticity of the teacher’s corresponding assessment tasks. The numbers of work samples averaged for each assessment task are given below the chart. All scores are normalized to a scale of 0-1.

Assessment Task
1 - RNA Fingerprinting Lab analysis (N = 7)
2 - RNA Fingerprinting Lab quiz (N = 7)
3 - Immune System & Flu Vaccine PSA's (N = 21)
4 - Energy Concepts quiz (N = 34)
5 - Wind Turbine Engineering & Testing (N = 27)
Figure 7.2 Relationship of mean Combined Authenticity of student performance on science work samples and mean Combined Authenticity of the teacher’s classroom instruction for the corresponding instructional units. Mean classroom instruction values for each instructional unit are computed from lesson segment scores weighted by the duration of the lesson segment. The numbers of work samples averaged for each instructional unit are given below the chart. All scores are normalized to a scale of 0-1.

Assessment Task
1 - RNA Fingerprinting Lab (N = 14)
2 - Immune System & Flu Vaccines (N = 21)
3 - Energy Concepts (N = 34)
4 - Wind Turbine Engineering & Testing (N = 27)
Endnotes

1 Lesson segments with very low numbers of observations resulted when the investigator was interrupted or cut off by the end of the lesson segment. Observations from these lesson segments are less likely to be representative of the whole class. Nineteen observations is about 75% of the average class size of 26.

2 As discussed in Chapter 6, the observation protocol for self-regulated learning was different when students were grouped as a whole class. Observations during lesson segments when students were grouped as a whole class were not as representative of self-regulated learning frequency of the whole class as those with students working independently or in small groups. Also, only lesson segments when students were working in small groups or independently were chosen for inclusion in the behavioral engagement correlation analysis because students were frequently given more responsibility for their own learning during these lesson segments. This sample addresses the research question more meaningfully than lesson segments dominated by the teacher delivering task directions or lecture notes.

3 These lesson segments are not evenly distributed between trimesters or instructional units. Half of the lesson segments in the sample for behavioral engagement are in the spring trimester. More than 60% of these lesson segments analyzed for self-regulation are in the spring trimester and 60% of those are during the Wind Turbine Engineering and Testing unit (Table G.1).

4 As in Chapter 5, the classroom instruction authenticity variables are normalized to a scale from 0-1, where 0 measures no authenticity, 0.5 measures moderate authenticity, and 1 measures high authenticity. The normalized score was calculated from: (component score – low bound of score range)/(high bound of score range – low bound of score range).

5 The teacher’s Learning Targets are in Appendix A.

6 The 202 lesson segments, totaling 73 hours of class time, from the four instructional units in this analysis were scored from field notes on classroom instruction for 11 Lesson Authenticity Indicators (LAIs), as described in Chapter 5 (Table 5.6). Then, the Combined Authenticity and each of the four authenticity component scores for each of the lesson segments were calculated from the LAI scores (Table 5.7).

7 For the work samples from the unit, the mean of all 21 work samples corresponding to both of the assessment tasks for the unit is used. If means had been used from each of the two assessments for this unit had been used. The score from the one work sample from the quiz version of the assessment task for the unit would have had undue influence on correlation results. Therefore, the more conservative approach was taken.

8 When the means are calculated the scores from each lesson segment are weighted by the duration of that lesson segment, as in the analysis in Chapter 5.
As in Chapters 5 and 6, the authenticity variables are normalized to a scale from 0-1, where 0 measures no authenticity, 0.5 measures moderate authenticity, and 1 measures high authenticity. The same method is used for authentic student performance variables, and assessment task and classroom instruction variables. The normalized score was calculated from:

\[
\frac{\text{component score} - \text{low bound of score range}}{\text{high bound of score range} - \text{low bound of score range}}
\]
8.1 Research Overview

This mixed methods exploratory case study was designed to document an effort to implement authentic science instruction in a high school and examine student responses to their experiences of authentic science pedagogy. This study conducted a systematic analysis of authenticity in the ninth grade science classrooms of a single teacher. The analysis was guided by a theoretical framework based on a model of authenticity developed for research on effective school restructuring (Wehlage et al., 1996). This model focused on a primary goal of schooling, to develop in students the capacity for authentic intellectual work. Authentic intellectual work was conceived as the capabilities of competent adults in a particular field (Newmann et al., 2007). This study explored student responses to authentic pedagogy by examining the evidence for authentic intellectual work in their science classwork, and behavioral engagement and self-regulation.

The model of authentic pedagogy and student performance adopted by this study (Table 3.1) was chosen to explore its relevance to the various emphases found in the frequent calls for authentic science instruction (Chinn & Malhotra, 2002; McGinn & Roth, 1999; National Research Council, 2012). The authenticity model adopted claims that instruction with characteristics of intellectual authenticity and specific elements believed to support students’ perception of value in their school work beyond being successful in school (Wehlage et al., 1996) are needed to promote students’ capability to
do authentic intellectual work. Previous research has demonstrated a greater level of authentic performance in science, and other subjects, when high school students experience higher levels of authentic pedagogy (Newmann et al., 2007).

The case studied was chosen because it was in a diverse public options school with a vision closely aligned with this study’s model of authentic pedagogy and student performance. The school was science-focused and committed to addressing current concerns among science education policy makers for diversity in the STEM pipeline. By intention, the school population was more diverse than the district as whole. Based on the philosophy of its founders and its association with the Expeditionary Learning Schools network, the school’s vision included curriculum focused on problems in the community beyond the school. This vision aligned closely with the authenticity model used in this study.

The teacher, whose classrooms defined the case studied, was part of the planning committee that developed the school’s vision. He was an experienced teacher, with good classroom management skills. It has been reported that without strong classroom management skills, authentic pedagogy, as defined in this study, wasn’t observed (Avery, 2000). The teacher had personal experience conducting scientific research. He had also supervised original student science research within and outside of the regular classroom curriculum.

The model of authentic pedagogy and student performance has not previously been used in a longitudinal study of trends in authentic pedagogy during a school year and its application to science classrooms has been limited. This study documented how an experienced science teacher with a curriculum vision closely aligned with the model
of authenticity adopted in this study, and with colleagues sharing this vision, implemented elements of authentic instruction across most of a school year and how students responded.

The study was conducted in a naturalistic setting and was not an intervention. Therefore, it used convenience samples to document authenticity in the teacher’s pedagogy and student performance and explore behavioral responses of students. The convenience samples collected were used produce a rich description of the case and allowed examination of the propositions in the study but not generalization beyond the samples collected (Yin, 2014).

The importance of this study is its detailed examination of the aspects of this model implemented in a high school science classroom with a highly qualified teacher in a diverse school designed to be supportive of authentic pedagogy. A unique aspect of this study is to demonstrate use of the same model of authenticity to characterize both science and engineering pedagogy and student performance as they were integrated in the curriculum. The usefulness of the instruments adapted and developed to document characteristics of authenticity in the teacher’s pedagogy and student classwork are examined.

### 8.2 Providing Students an Authentic Science and Engineering Experience

This study used two instruments derived from the model of authentic science pedagogy it adopted and the science and engineering education literature to measure authenticity in the teacher’s assessment tasks and day-to-day classroom instruction. The instruments were not designed to be a comprehensive measure of teacher effectiveness.
They focus on pedagogical characteristics shown to be associated with authentic intellectual work by high school science students and students in other subject areas (Newmann et al., 2007).

These instruments were able to produce a richly descriptive profile of the characteristics of authenticity in the teacher’s science pedagogy. These measurements revealed patterns that point out the teacher’s instructional priorities and strategies.

For example, the teacher organized some authentic aspects of the curriculum similarly during each trimester. Each school trimester ended with an instructional unit assessed by a task that asked students to create a substantial artifact that integrated several science and engineering concepts and skills to reach a conclusion about a personal response to a biological system or to engineer and test a design connected to meeting a societal need. In addition, during each trimester the teacher used shorter units focused on developing basic knowledge of science concepts. Students were expected to apply many of these concepts in other instructional units.

The instruments documented lower authenticities of both the assessment tasks and classroom instruction matching the teacher’s goals in shorter units more limited to conceptual knowledge. The measurements revealed the relatively higher authenticity of the assessments for instructional units at the end of the trimester and how characteristics of authenticity were implemented in classroom instruction during these and some other units designed to prepare students for tasks with higher expectations.

The authenticity scores on the Task Authenticity Indicators and the Lesson Authenticity indicators measured which aspects of authenticity were commonly present in the teacher’s assessment tasks and instruction. In this case, the classroom instruction
instrument found that, by instructional unit, the teacher most commonly afforded students to use fundamental science concepts at a moderately high to high level of authenticity. The other common aspects of authenticity were affordances to make connections to the world outside the classroom, to use higher-order thinking, and to interact with multiple representations of science concepts and data. Three of these four most common authentic affordances in the teacher’s instruction were contributors to the Construction of Knowledge and Science Meaning-Making Processes components of authentic pedagogy.

The study was able to show that the Construction of Knowledge and Science Meaning-Making Processes components of authentic pedagogy were measured at higher levels than affording value beyond success in school in the teacher’s pedagogy. Two indicators define the Value Beyond School component of authenticity, connections to the world beyond the classroom, and presenting classwork to audiences beyond the classroom. The lower affordance for the Value Beyond School component was largely due to infrequent affordances for presentations to audiences beyond the teacher. By contrast, both assessment tasks and classroom instruction made frequent connections to personal and societal concerns outside the classroom. This is not a surprising finding, but it is important. It will be discussed more below and in Chapter 9.

The teacher’s average authentic pedagogy level was measured as similar to higher scoring teachers in the School Restructuring Study (Marks et al., 1996) that used the authenticity model adapted for this study. It is important to document that presentation to audiences beyond the classroom is the least common aspect of authenticity in the teacher’s pedagogy. This, of course, begs the question of what influence greater authenticity in this regard would have had.
8.3 Authentic Pedagogy and Student Performance, Engagement, and Self-Regulated Learning

It was shown that this study’s instrument for measuring the authenticity levels of student performance on their science classwork could be implemented reliably. The instrument was useful to determine differences in the performance levels between students. Performance levels varied from the top to the bottom of the authenticity scale.

The student performance instrument was based on the same model as measurements of authenticity levels in the teacher’s pedagogy. This allowed comparison of the common authenticity components within the student work and the teacher’s classroom instruction and assessment tasks. The preliminary quantitative analysis conducted in this study showed correlations between the levels of authenticity in student performance and the classroom instruction and the assessment tasks the student work was a response to.

In work samples from the fall and spring trimesters, student performance was measured as responsive to the teacher’s expectations. This study did not find a qualitative trend in student performance from fall to spring. The average level of authenticity of student performance in the convenience samples collected was above the midpoint on the authenticity scale. This compares favorably with previous research using the same authenticity model in a study of high school math and social studies classrooms from schools engaged in significant school restructuring, similar to the school in this case (Marks et al., 1996).
The level of student behavioral engagement, defined as attentiveness and involvement in classroom activities, was generally high across the school year. The correlation analysis showed negligible associations between authentic aspects of authentic instruction and this student response variable. While there were differences in behavioral engagement between instructional units in the samples collected, they weren’t clearly related to the authenticity, as proposed by Proposition 1.

It is notable that the behavioral engagement levels remained around 75% during the Wind Turbine Engineering and Testing unit when students were given most of three weeks of class time to manage their own work on an engineering task.

In over half of the observations of student self-regulated learning, students were recorded as working independently or autonomously. In the convenience samples collected, the correlation analysis found evidence that instruction focused on core science and engineering concepts and connections to the real world might have encouraged student self-regulation. This suggests greater student ownership in the learning task, as projected in Proposition 1.

All the student response variables were measured as relatively high during the final unit at the end of the school year focused in engineering a wind turbine. This was the only unit during the school year with a clear environmental theme. Engineering design was the focus of the activities, but the unit was contextualized by the need to meet our society’s need for environmentally sustainable energy.

Gender differences in behavioral engagement levels were so small as to strongly suggest no gender effect. Self-regulation measures also suggest small or no differences between males and females. The authentic performance on work samples of males is
measured slightly greater than females at the instructional unit level, but the high degree of variation in the scores doesn’t allow a statistically based conclusion. In a population that was 60% female in a science focused options school, no clear gender differences were evident in the classroom of a male teacher.

8.4 Limitations of This Study

The levels of authentic pedagogy, authentic student performance, as well as, behavioral engagement and self-regulation were likely influenced strongly by the expertise of the teacher of the classrooms studied. His strong rapport with students and classroom management skills played a role. His experience and commitment to the vision of the school influenced his implementation of pedagogy with authentic characteristics. Even though school personnel, including this teacher, worked hard to recruit a diverse population that included first generation college students from groups underrepresented in STEM careers, students or their parents or guardians chose the school. That influence must also be acknowledged when considering implications for other contexts.

The advantage of choosing this case for the study was the likelihood of seeing some influence of authenticity pedagogy in the classrooms of a teacher able to implement it. That goal was achieved by this study, but the disadvantage was that reaching conclusions regarding the study propositions was restricted by limited differences in the authenticity of classroom instruction.

The convenience samples collected in a naturalistic context allowed for description, but limited the generalizations that can be made. The teacher demonstrated
best practices by adapting the curriculum on an ongoing basis in response to frequent formative assessments of student progress. Another best practice was providing multiple and customized assessments of student proficiency. The result was that collecting sets of student work samples that could be generalized to the entire ninth grade population in the case wasn’t achieved within this study.

The positive relationship between the authenticity of student performance and the teacher’s assessment tasks and classroom instruction hypothesized in Proposition 2 was supported by the correlation analysis for Combined Authenticity, but isn’t generalizable beyond the convenience samples collected. The correlation analysis for Proposition 2 was preliminary and didn’t explore the relative degrees of association for different components of authenticity in the teacher’s pedagogy and student performance. Nor did it consider the influence of possible confounding variables, such as, class period or gender. The samples did include representation from different gender, ethnicity, and prior achievement groups, but the samples analyzed weren’t random samples representative of all groups within these classrooms.

Chapter 6 detailed another limitation regarding interpretation of the authentic performance levels. The teacher’s implementation of the school-wide proficiency-based grading system meant that it is not known whether work samples collected represent the highest level of achievement reached for each student on the learning proficiencies being assessed.
8.4.1 A limitation of particular concern of this investigator

This investigator adopted the model of authenticity used by this study and conducted the research in the case described based on a strong interest in the exploring the unique contribution that affording students value beyond demonstrating success in school in their classwork would play in student learning. The investigator’s strong concern for promoting environmental literacy is behind this interest. When this investigator began meeting with the planning staff for the school, the plan was for frequent contextualization of learning in environmental science topics and concepts.

During the school planning process that focus was changed by the school district to health science and engineering concentrations within the high school curriculum. However, the vision to implement connections within the curriculum to issues students recognized outside the classroom and connecting with community partners remained. This strongly limited the direct application of this study’s findings to environmental education.

The teacher in this case during the school year studied was teaching the biology and physics courses for the first time in this particular school. The teacher used some curriculum from his long experience in the classroom, but the curriculum implemented was either developed new for that year or repackaged around the compelling topics chosen to make connections to the world beyond the classroom and encourage student engagement.

Among the teacher’s challenges to develop a new, conceptually rich curriculum with a coherent learning progression, teach effectively for a diverse student population, including frequent individual feedback to student’s on their level of understanding and
skill, and collaborate with colleagues to develop a new school culture, the intention to have students create learning products that met community needs through collaborations with community partners wasn’t implemented within the classrooms. Connections to the world beyond the classroom were common in the teacher’s pedagogy, but not presentations to audiences beyond the classroom. This limited this study’s intention to explore both aspects included in the Value Beyond School component of authenticity.

In addition, based on the concept of authenticity in the science education literature (Roth & Barton, 2004), this investigator’s instrument for measuring the Value Beyond School component of authentic classroom instruction originally included a third indicator: Activities occur in environments or contexts that are physically, socially, and/or technologically similar to those in which competent adults practice science (H.-S. Lee & Songer, 2003; McGinn & Roth, 1999; Rahm et al., 2003). Only two lesson segments scored moderately authentic for this indicator, so it was not included in the study. It wasn’t possible to explore this hypothesized element of authenticity’s influence on students.

Examination of the full of effect of the Value Beyond School component of the model of authentic pedagogy wasn’t possible because there weren’t examples of higher authenticity on the measure of presentations to audiences beyond the classroom. However, the finding that other aspects of authenticity were implemented without this one in a case where there was an explicit acknowledgement of the value of both aspects of the Value Beyond School component in the original authenticity model is important.

In an unstructured interview in September after the study period, the teacher commented on his thinking regarding his curriculum-related priorities in the upcoming
school year. He conveyed that school leadership was calling for curriculum that included learning activities and learning products that met community needs through collaborations with community partners to be increased. He said that he believes that fieldwork is valuable and more authentic, but that he needs to take care of himself and not be “tired and bummed out”. He said that given finite time resources, it is probably more valuable for the students for him to do the best job he can in the context of his own classroom without pursuing opportunities for interdisciplinary curriculum and fieldwork in the community.

This teacher’s reaction points to important challenges that need to be considered in future research and the concerns of communities for the quality and humanity of their schools.

8.5 Summary

The instruments used to assess authentic pedagogy and student performance used in this study were able to produce rich descriptions of the implementation of characteristics of authentic pedagogy and student performance. Connections between authentic pedagogy and student achievement and self-regulated learning were found within the case.

The lower level authenticity found for the component of authenticity designed to provide a context that challenges and motivates students to put the intellectual skills of competent adults into practice, highlights the need for research on how accomplish this goal in our schools.
There are two primary ways this study extends current research on science instruction and student achievement. First, it adapts for application to high school science classrooms earlier research on intellectually challenging and effective instruction, and research on best practices that support student learning and motivation. Second, it uses a framework based on that research to document how it can be used to describe instruction and student achievement in high school science classrooms. This provides an example of the effectiveness and challenges of implementing this comprehensive view of authenticity, that hasn’t been available to science education researchers before.
Chapter 9  Further Research

This study compared the authenticity of student performance with the authenticity of two critical elements of the teacher’s pedagogy, assessment tasks and classroom instruction. Questions of interest that were not answered fully within the scope of the analysis in this study include:

- Are any of the associations found between authentic student performance and authentic pedagogy stronger for some student groups than others?
- Did the Value Beyond School component have a measurably distinct role in predicting student performance?
- Does authentic instruction that includes presentations to audiences beyond the school contribute to student performance distinct from other aspects of authentic instruction?
- What supports are needed for teachers to enable them to implement all of the elements of authenticity theory says are needed to maximize student learning?
- In addition, are measures of the values students perceive in their classwork associated with increased Value Beyond School as predicted by this study’s authenticity model?

Current K-12 science education in the U.S. is concerned with implementing the Next Generation Science Standards (NGSS), as well as, how they integrate with the Common Core State Standards (CCSS) for English language arts and literacy and for mathematics (NGSS Lead States, 2013). There is a common focus on higher-level
thinking, depth of knowledge, and disciplinary meaning-making practices in these standards. The instruments used in this study to measure affordances for authentic intellectual work in teacher pedagogy and authentic student performance in science can be adapted to document the impact of NGSS adoption. The NGSS identifies performance expectations that are designed to encourage teaching strategies that integrate the three dimensions of science education envisioned in *A Framework for K-12 Science Education* (National Research Council, 2012). Students are to use science and engineering practices to apply crosscutting concepts, the big ideas of science, and disciplinary core ideas to solving problems and build understanding of the interconnected knowledge that defines science literacy.

The NGSS approach is similar to the model of authentic pedagogy and student performance used in this study. Both stress the claim that integration of all instructional elements in their models are essential to student learning (Krajcik et al., 2014). Major features of the theoretical model used in this study could contribute to research on the implementation of the NGSS and other current standards.

The development and field testing of the instruments described in this study also allows research questions to be investigated that could help address assessment of progress implementing the NGSS. Adaptation of the authenticity indicators from this study would be necessary to fit the NGSS framework. However, the clarity of contributions from core ideas, crosscutting concepts, and practices in the NGSS Performance Expectations would make reliability and calibration of research instrumentation easier to achieve than it was in this study.
The research agenda called for in *A Framework for K-12 Science Education*, the theoretical foundation for the NGSS, highlights the primacy of research on “classroom-level contexts, materials, and discourses that engage and support a wider range of students in high-quality teaching and learning experiences with the concepts, ideas, and practices” (National Research Council, 2012, p. 325). An adaptation of the investigative framework used in this study to examine classroom pedagogy and corresponding student performance could help explore the following NGSS implementation questions:

- What is the baseline of NGSS-compliant assessments, classroom instruction, and student performance within a school district?
- Did a particular teacher professional develop program measurably impact NGSS implementation and connected student achievement?
- Can aspects of this study’s model of authenticity adapted for the NGSS help track the development of beginning teachers?

This study’s model of authentic pedagogy claims that student performance is maximized when Value Beyond School is implemented in school pedagogy. This element of this study’s model is not fully embraced by *A Framework for K-12 Science Education* and the NGSS (NGSS Lead States, 2013). However, the connections to the world beyond the classroom indicator is emphasized in some NGSS Performance Expectations, but this is uneven. Some science educators have spoken of their belief that NGSS Performance Expectations promote students finding value in the work they require. Juliana Texley, President of the National Science Teachers Association, has spoken recent about the NGSS promoting student “empowerment” (Texley, 2014).
The theoretical model used in this study is more explicit in its inclusion and definition of student value in classwork beyond demonstrating success in school. Research on whether this component of the model of authenticity contributes to student achievement of the NGSS Performance Expectations is needed. *A Framework for K-12 Science Education* calls for research on “how best to develop and sustain students’ interest in science” (National Research Council, 2012, p. 314). The effect of instruction that includes experiences designed to explicitly promote student’s perceptions that classwork has value beyond success in school is important to investigate further.

The North American Association for Environmental Education has published recommendations for linking environmental education standards to the NGSS (North American Association for Environmental Education, 2014). They include the promotion of environmental education as an instructional context for NGSS in K-12 schools. I believe environmental education and the experience of the environmental education community can play a unique role in defining a research agenda for determining the role of value beyond demonstrating success in school in sustaining interest and engagement in science learning.

There is expertise in the environmental education community regarding the role learning experiences with value to students beyond demonstrating success in school play in promoting intellectual work, active citizenship, and environmental protection (Oregon Environmental Literacy Task Force, 2013). Environmental education standards include “direct experience with the environment, environmental issues, and society” (North American Association for Environmental Education, 2010, p. 3) as essential for developing environmental literacy. Environmental educators make connections between
the classroom and audiences beyond the school more frequently than most other science teachers.

I’ve suggested further research on the role of value beyond demonstrating success in school in sustaining motivation and increasing achievement in science classrooms. That research ought to include Value Beyond School as envisioned in environmental literacy standards. Research can also be designed to determine how environmental education that includes value beyond demonstrating success in school contributes to achieving the learning outcomes that are shared by the NGSS and environmental literacy standards. Strategies for partnerships with environmental educators outside of K-12 classrooms that enable teachers to include and sustain the implementation of Value Beyond School in their pedagogy should be investigated.

Environmental educators and researchers could help adapt and improve assessment tools, such as those described in this study, to validate the centrality of student values for place and sustaining our life-support system in a complete model of effective learning that empowers citizenship.

Possible applications to teacher education

Planning curriculum aligned with the Next Generation Science Standards

Some of what was learned in the current study is informative for teacher education programs. Loughran (2014) highlights the value of using standards to guide student teachers’ reflection on their practice. The authentic pedagogy model modified for this study may be adaptable for this purpose, considering its general alignment with the NGSS. The Task Authenticity Indicators (TAIs) are adaptable to more specifically
address the NGSS. Science or engineering content (TAI 3) can be mapped to Disciplinary Core Ideas (DCIs) and Crosscutting Concepts (CC) in the NGSS, and TAI 4 can be mapped to the NGSS Science and Engineering Practices (SEP).

As student teachers consider NGSS recommendations, they may find the authenticity model embodied in these adapted TAIs a helpful tool for reflection on their curriculum planning and summative assessments. Supervising teachers could use the TAIs to provide feedback. The TAIs might play a similar helpful role for a professional learning community.

The original Center for Organization and Restructuring of Schools (CORS) classroom instruction scoring guides consist of a single scoring guide for each of the four components of authenticity (Newmann et al., 1995). I believe the original form would be more useful in teacher education than the adaptation in this study. Like the TAIs, the CORS classroom instruction scoring guides can also be revised to more specifically align with the NGSS. They may serve as a helpful tool to analyze video or direct observation of teacher practice in preservice teacher education. They could also be used by a professional learning community, perhaps in a lesson study context (Lewis, Perry, & Hurd, 2004).

Use of a consistent model for evaluating instruction, curriculum, and assessment could be beneficial for teachers. The broad framework in the authentic pedagogy model also could be adaptable to introducing science teachers to commonalities in the NGSS and the Common Core State Standards (CCSS) for English language arts and literacy and for mathematics (Common Core State Standards Initiative, 2010a, 2010b).

Implementing the Value Beyond School component of authentic pedagogy
The most unique aspect of the authenticity model used in this study is the operational definition of the Value Beyond School component. The concept of classwork that has value to students beyond demonstrating success in school and a tool to assess the inclusion of this concept in curriculum planning and assessment design, could provide a context for preservice teachers to consider some aspects of motivation and the value of student-centered instruction (Lambert & McCombs, 1998).

Connecting instructional activities to students’ experiences outside school requires teachers to consider developmental and cultural perspectives on what students will recognize and value. These perspectives are also important to implementing the presentation to audiences beyond the school. The theory is that a significant aspect of the authenticity of these presentations will be the extent that students see themselves as bringing something of value to those audiences (Basu & Barton, 2007).

*The curriculum strategy for implementing Value Beyond School from this study*

The teacher in this study used compelling questions that contextualized the curriculum over each trimester to provide connections to the world beyond school. In the science education literature, the concept of a “driving question” from project-based science is similar (Krajcik, McNeill, & Reiser, 2008; J. Singer et al., 2000). In this study a compelling question created a need to know that guided curriculum decisions regarding the sequence of concepts needed to address assessment tasks with authentic affordances for constructing knowledge and applying science and engineering meaning-making processes. In the context of the NGSS, this teacher’s degree of success points to a useful example of a curriculum-planning model that could inform the training of teachers and continued teacher learning.
Performance Expectations in the NGSS define the DCIs, CCs, and SEPs instruction needs to address (NGSS). A compelling question is proposed to provide a motivating context for classroom activities. A compelling question can also bring focus to selecting the ideas and topics a curriculum planner needs to weave into a motivating instructional sequence that provides practice solving realistic problems.

The teacher described in this study was an experienced educator. That certainly played a role in his success using compelling questions to improve the authenticity of his practice. Teacher educators will need to consider where implementation of the Value Beyond School component of authenticity from this study fits in the developmental sequence of training beginning teachers.

Implications for a research agenda

To advance teacher education, further research is needed to confirm the unique contributions that both of the elements of the Value Beyond School component of authenticity bring to student engagement and achievement. This research needs to include developmental and cultural variables. It needs to assess the values students actually perceive at different developmental levels and extend research on the importance of connecting to students’ cultural funds of knowledge (Barton, Tan, & O’Neill, 2014; Basu & Barton, 2007; Gonsalves, Rahm, & Carvalho, 2013).

Research is needed into the learning experiences that help preservice teachers find connections to the world outside the classroom their students can recognize. This is especially important for training teachers with different cultural backgrounds than the students they will teach. We need to test ideas for school districts to support and equip teachers to develop successful relationships with community partners that provide
opportunities for students to share meaningful learning products with audiences outside of the school.
References


Appendix A  Scoring Science and Engineering Assessment Tasks for Authenticity

This instrument is a modified version of those reported in Newmann, F. M., Secada, W. G., & Wehlage, G. (1995). *A guide to authentic instruction and assessment: Vision, standards and scoring*. Madison, WI: Wisconsin Center for Education Research and Research Institute on Secondary Education Reform (RISER) for Youth with Disabilities (2001). *Standards and Scoring Criteria for Assessment Tasks and Student Performance*. Madison, WI: Wisconsin Center for Education Research. The original text of the Newmann, et al. document is in regular text. Quotes of the RISER document are in Calibri font. *Modifications made for this study are in italics.*

*These scoring guides were used to score selected proficiency assessment tasks for their affordances for each component of authentic performance in the model of authentic science pedagogy developed for this study. Each task was scored for all seven indicators.*

General Guidelines

*The source material for scoring assessments for a particular Learning Target are first, the assessments themselves, second, criteria for evaluation indicated by the teacher interview or teacher scoring guide or key, and thirdly, the statements of the learning targets.*

- If a task has different parts with different levels of expectations, the score may be based on the inclusion of the indicator in the assessment rather than the entire assessment. The best guideline is that a high score should be given when the indicator is required to be scored highly proficient by teacher. The teacher's scoring rubric and learning targets associated with the task should be used to clarify the teacher's expectations. Therefore, if a major part of the assessment meets the high level for a particular indicator, the whole assessment can be scored high for the indicator.

- *Scores should reflect reasonable ninth grade level expectations.*

- When it is difficult to decide between two scores, give the higher score only when a persuasive case can be made that the task meets minimal criteria for the higher score.

- If the specific wording of the criteria is not helpful in making judgments, base the score on the general intent or spirit of the standard described in the introductory paragraphs of the standard.
CONSTRUCTION OF KNOWLEDGE

Indicator 1: Organization of Information

The task asks students to organize, synthesize, interpret, explain, or evaluate complex information in addressing a concept, problem, design, or issue, dealing with describing, representing, explaining, predicting, or manipulating a natural phenomenon or designed system.

Consider the extent to which the task asks the student to organize information, rather than to retrieve or to reproduce isolated fragments of knowledge or to...apply previously learned algorithms and procedures. To score high, the task should call for interpretation of nuances of the information or a topic that goes deeper than surface exposure or familiarity by considering the meaning of the information. It should also require describing, explaining, predicting, or manipulating a natural phenomenon or designed system.

When students are asked to gather information for reports, this indicates some selectivity and organizing beyond mechanical copying, but if they are not asked to consider the meaning of the information by interpreting, evaluating, or synthesizing, give a score of 2.

Scoring guide
3 = The task’s dominant expectation is for students to organize and interpret information related to a natural phenomenon or a designed system.
2 = There is some expectation for students to organize and interpret information.
1 = There is very little or no expectation for students to organize...and interpret information. Its dominant expectation is for students to retrieve or reproduce fragments of knowledge or to...apply previously learned information and procedures.

Indicator 2: Consideration of Alternatives

The task asks students to consider alternative solutions, strategies, explanations, or points of view as they address a concept, problem, theory, design, or issue.

To what extent does success in the task require consideration of such alternatives? To score high, the task should clearly involve students in considering alternatives, either through explicit presentation of the alternatives or through an activity that cannot be successfully completed without examination of alternatives implicit in the work. It is not necessary that the students’ final conclusions include listing or weighing of alternatives, but this could be an impressive indicator that it was an expectation of the task.

Scoring guide
3 = The task’s dominant expectation is for students to analyze and evaluate alternatives.
2 = There is some expectation for students to analyze and evaluate alternatives.
1 = There is very little or no expectation for students to analyze and evaluate alternatives. Its dominant expectation is for students to retrieve or reproduce fragments of knowledge or to apply previously learned information and procedures.
**SCIENCE MEANING-MAKING REASONING PROCESSES**

**Indicator 3: Science and Engineering Content**

The task asks students to show understanding of concepts, theories, or perspectives considered central to science or engineering.

*To what extent does the task promote students’ understanding of and thinking about concepts, theories or perspectives considered seminal or critical to science or engineering?* Reference to isolated factual claims, definitions, or algorithms—though necessary to inquiry within science or engineering—will not be considered indicators of significant science content unless this is in a context requiring students to apply fundamental powerful science or engineering ideas that organize and interpret the information.

**Scoring guide**

3 = Success in the task clearly requires understanding of concepts, theories, or perspectives central to science or engineering AND application of these concepts to a natural phenomenon or engineering design.

2 = Success in the task seems to require understanding of concepts, theories, or perspectives central to science or engineering, but the task does not make these very explicit.

1 = Success in the task can be achieved with a very superficial (or…without any) understanding of concepts, theories, or perspectives central to science or engineering.
Indicator 4: Science and Engineering Processes

The task asks students to use methods of inquiry, research, or communication characteristic of science or engineering.
To what extent does the task lead students to use methods of inquiry, research, communication, or discourse characteristic of science or engineering? Some powerful processes of inquiry may not be linked uniquely to any specific discipline (e.g., interpreting graphs), but they are valued here if the task calls for their use in ways similar to important uses in science or engineering.

Scoring guide

3 = Success in the task clearly requires use of methods of inquiry or communication important to the conduct of science or engineering AND application of these processes to a natural phenomenon or engineering design.

2 = Success in the task seems to require use of scientific or engineering methods of inquiry or communication, but these are not very explicit, are not central to the conduct of science or engineering, or are incidental to the main task.

1 = Success in the task can be achieved without use of any specific methods of inquiry or communication.
Indicator 5: Elaborated Written Communication

The task asks students to elaborate on their understanding, explanations, or conclusions through extended writing.

*To what extent does the task require students to elaborate on their understanding, explanation, or conclusion regarding natural phenomena, the role of science in a public issue, or engineering design through an extended piece of scientific or engineering communication? Examples include tasks requiring students to use observations to argue the adequacy of an explanation of a natural phenomenon; construct an argument or reach a conclusion from evidence; critique or justify a research question, an experimental design, or an engineering design; create or revise a model to explain data; or use science knowledge in a defense of a position on an issue.*

*Graphs, tables, equations, diagrams, sketches, computer models, mapping (GIS), spreadsheets, or imaging, as well as prose, may be part of the communication. The communication may be in the form of an oral presentation, a technology presentation, a poster, or a combination of these, but must include some explanatory prose.*

**Scoring guide**

4 = Analysis/Persuasion/Theory

*The task explicitly calls for students to present a generalization, conclusion, or solution AND provide elaborated support for their thinking. Elaborated support is justification of their thinking or solution path with evidence such as data, theories, illustrations, or examples. To clarify, the task must include BOTH a “what” component (generalization or solution) that is connected to a “why” or “how” component (support).*

3 = Report/Summary

*The task calls for generalization OR supporting information. The task asks students, using narrative or expository writing, either to present a generalization, conclusion, argument, or solution OR to offer examples, summaries, illustrations, details, or reasons, but not both. The task focuses on “what”, “how”, or “why” but doesn’t use one to justify or explain the other.*

2 = Short-answer exercises

*The task or its parts can be answered with only one or two sentences, clauses, or phrasal fragments that complete a thought. Students may be asked to show some work or give some examples or reasons, “but this is not emphasized and very limited detail is requested.*

1 = Multiple choice or fill-in-the-blank exercises

*Answered with less than a sentence. The task requires ... only giving solutions or definitions.*
VALUE BEYOND SCHOOL

Indicator 6: Problem Connected to the World Beyond the Classroom

The task asks student to address a concept, problem, or issue that is similar to one that they have encountered, or are likely to encounter, in life beyond the classroom.

To what extent does the task present students with a question, issue, or problem that they have actually encountered or are likely to encounter, in their lives beyond school?

Task demands for culturally valued, “basic” knowledge will not be counted here unless the task requires applying such knowledge to a specific problem likely to be encountered beyond the classroom.

When students are allowed to choose topics of interest to them, this might also indicate likely application of knowledge beyond the instructional setting. But tasks that allow student choice do not necessarily connect to issues beyond the classroom. To score high on this standard, it must be clear that the question, issue, or problem which students confront resembles one that students have encountered, or are likely to encounter, in life beyond school.

Scoring guide

3 = The question, issue, or problem clearly resembles one that students have encountered, or are likely to encounter, in life beyond school. The resemblance is so clear that teacher explanation is not necessary for most students to grasp it.

2 = The question, issue, or problem bears some resemblance to real world experiences of the students, but the connections are not immediately apparent. The connections would be reasonably clear if explained by the teacher, but the task need not include such explanations to be rated 2. The task offers the opportunity for students to connect the topic to experiences, observations, feelings, or situations significant in their lives, but does not explicitly call for them to do so.

1 = The problem has virtually no resemblance to questions, issues, or problems that students have encountered, or are likely to encounter, beyond school. Even if the teacher tried to show the connections, it would be difficult to make a persuasive argument.
**Indicator 7: Audience Beyond the School**

The task asks students to communicate their knowledge, present a product or performance, or take some action for an audience beyond the teacher, classroom, and school building. Authenticity increases when students complete the task with the intention of communicating their knowledge to an audience beyond the teacher and when they actually communicate with that audience. Such communication can include informing others, trying to persuade others, performing, and taking other actions beyond the classroom. This refers not to the process of working on the task, but to the nature of the student’s final product.

**Scoring guide**

4  =  Final product is presented to an audience beyond the school.
3  =  Final product is presented to an audience beyond the classroom, but within the school.
2  =  Final product is presented to peers within the classroom.
1  =  Final product is presented only to the teacher.
Teacher’s Learning Targets

9th Grade Biology Learning Targets – Trimester 1 – Influenza

LT 1: I can explain the connection between form & function in viral structure.
- I can describe the structure of a flu virus in detail.
- I can explain what neuraminidase, hemagglutinin, RNA, the lipid bilayer, matrix proteins, ion channels, & NEPs do for a virus.
- I can explain how a flu virus enters a cell and takes over the cell to produce more new virus.

LT 2: I can explain how a genes become proteins.
- I can describe the function of the nucleus & the ribosomes in cell.
- I can compare and contrast the structures and functions of RNA and DNA.
- I can explain the processes of replication, transcription & translation.

LT 3: I can explain the mechanisms of genetic change that lead to viral evolution.
- I can explain how a change in the genetic code of a virus can change the way that virus affects human health.
- I can explain the concept of antigenetic drift.
- I can explain the concept of antigenetic shift.

LT 4: I can explain how the immune system defends the body from viral infection.
- I can describe the systems in the body that are part of the immune system.
- I can describe the cellular & chemical components of the immune system.
- I can explain how the immune system defends the body against influenza.

LT 5: I can explain how vaccines can be used to defend the body from viral infection.
- I can describe the main types of flu vaccines & what each is made up of.
- I can explain how a vaccine stimulates the immune system.
- I can explain the pros and cons of getting the flu vaccine.

LT 6: I can use the technique of electrophoresis to determine the identity of an unknown virus.
- I can name and describe the use for each piece of equipment used in electrophoresis.
- I can demonstrate proper techniques during an electrophoresis lab.
- I can explain the connection between restriction enzymes and electrophoresis.
- I can analyze electrophoresis results to compare unknown viruses to known viruses.

LT 7: I can craft a product that communicates a scientific concept to a variety of community members with varying levels of scientific expertise.
- I can write a clear explanation of a science concept.
- I can demonstrate control of standard writing conventions (correct spelling, punctuation, capitalization, paragraph breaks, & grammar).
- I can use images to effectively enhance communication about a science concept.
9th Grade Biology Learning Targets – Trimester 2 – Who Am I?

LT 1: I can describe how the information contained in human DNA is the source of human genetic variation.
- I can explain what a gene is, where they are located, & how they are used to create proteins.
- I can describe what proteins do including structural support, enzymes, cell function & immune system function.
- I can describe how single or multiple genes can lead to specific traits.
- I can explain how genes and environmental factors interact to create certain traits.

LT 2: I can explain how an understanding of the human genome impacts human health.
- I can explain how a genetic mutation can cause disease, including its effect on protein structure & protein function.
- I can explain how proteins can impact human health, including effects on structure, organ function, & behavior.
- I can explain how genetics applies to the health care strategies of preventive medicine, pharmacogenomics, gene therapy, and drug therapy.

LT 3: I can explain how a nerve cell works.
- I can describe the structure of a nerve cell including the cell body, axon, dendrites, myelin sheath, synapse and explain what each of these structures does.
- I can explain how nerve cells communicate with each other chemically and electrically.
- I can explain how nerve cells communicate with other cell types in the body.

LT 4: I can explain how the brain works.
- I can identify the main regions of the brain and describe in general terms what each region does.
- I can describe in some detail how the processes of memory & emotions are controlled by the brain.
- I can explain is some detail at least one specific aspect of the brain that is relevant to teenagers.

LT 6: I can carefully collect and clearly & accurately display scientific data.
- I can carefully follow a set of directions for accurately collecting a specific type of data.
- I can create a well organized data table that clearly identifies the independent variable, the dependent variable, any derived variables, repeated trials, correct units of measurement, and a clear, descriptive title.
- I can identify the correct type of graph (pie, line, bar, or scatter plot) for displaying a specific type of data.
- I can create a graph that positions the independent & dependent variables correctly, includes accurately scaled & labeled axes, clear, correct units of measurement, and a clear, descriptive title.
LT 7: I can clearly & accurately analyze data.
- I can clearly & accurately describe patterns shown by a set of data, including highs/lows, averages, & any correlation or lack of correlation.
- I can propose a clear, reasonable, scientific, explanation of the patterns identified in a set of data.
- I can pose a reasonable, scientific hypothesis that could be tested in a future experiment to prove or disprove some idea related to this set of data.
- I can clearly & accurately identify potential sources of error in a set of data. These might include problems of accuracy, precision, or calculation.

LT 8: I can craft a written product that describes the identity of 9th grade students at HS2.
- I can accurately describe a scientific characteristic related to identity;
- I can explain in a narrative how and why that characteristic is meaningful to me;
- I can create a data table & graph that displays the statistics for the characteristic among 9th grade HS2 students.

LT 1: I can explain & apply basic mechanics concepts.
- I can explain & calculate the concepts of speed, velocity, acceleration, mass, force, friction, gravity, work, & power
- I can explain & apply Newton's first 3 laws to specific, practical situations.
- I can identify & describe how the six simple machines work & can form parts of larger, more complex machines.

LT 2: I can explain & apply energy concepts.
- I can describe & identify types of energy involved in a specific, practical situation.
- I can explain the concept of conservation of energy and apply it to specific, practical situations.
- I can identify energy transformations that occur in a variety of mechanical situations.

LT 4: I can identify & explain the role of wind energy in a sustainable future on this planet.
- I can describe how a basic wind turbine converts the energy in wind into electricity.
- I can explain the pros & cons of wind energy as a sustainable source of energy for the future.
- I can design, build, & test a wind turbines & explain how it can work as a sustainable energy source.

LT 5: I can describe & use the design process to solve simple mechanical problems.
- I can identify the steps of the design process.
- I can explain how the design process is used to produce solutions to specific design problems.
- I can use the design process to develop solutions to specific design problems.

LT 6: I can use a design journal to consistently keep a record of my work.
- I can accurately record dates, page numbers, learning targets, & project title for each journal entry.
- I can accurately use diagrams to describe physics concepts & design ideas.
- I can accurately use dimensions, labels, and keys to clearly & specifically describe diagrams.
- I can record detailed notes & explanations related to each physics concept & design challenge.

LT 7 I can accurately use the process of scientific inquiry to test products designed to accomplish as specific task.
- I can outline a specific set of steps that describe how to collect a specific.
- I can use a data table accurately collect & record data related to a specific aspect of a design.
- I can clearly & accurately graph data so that it can be analyzed to better understand a design.
- I can identify patterns in data in order to draw conclusions about a specific design.
LT 8: I can consistently do work that is of high quality and craftsmanship.
- I can complete written assignments with minimal errors of grammar & spelling & neat handwriting or typed text.
- I can create diagrams & sketches that are neat, detailed, well organized, & that help to clarify & explain ideas.
- I can complete construction/design assignments that are meet design requirements, are sturdy, functional, & safe.
Appendix B  Classroom Instruction Observation Protocol

Protocol instructions

At the beginning of the class period, record the position of each student on the seating map. Behavioral engagement and self-regulated learning are also scored during most of the lesson segments scored for classroom instruction authenticity.

Field Notes

Use the Description of Events table to record field notes of lesson activity. When possible, begin each new lesson segment on a new page. The “Look for” column contains reminders of key observations. Record the time of discrete events in the left-hand column.

At the end of the segment. Record social grouping as dominated by whole class working together (lecture, class discussion), small group work (lab activities, research, worksheets), or individual (reading, seatwork, independent research). Estimate the proportion of teacher to student discourse during the segment. If lesson is focused on science inquiry or engineering design, note this in space provided. Note presence of community partners. Note time of end of segment.

At the end of the class period. Complete the Lesson Authenticity Indicators worksheet. Rate each segment individually using the scoring guides for each indicator.
Classroom Observation Protocol

Observer: ________________________  # Descrp Events pgs _______  # BE/SR Tables _____

Date: ___________  Time Start: ______  Time End: ______  Per: ___  Whole period?  Y / N

School: __________  Tchr ___________  Course:  __________________  Grade Level(s): ______

Students:  Total _______  Number of Males ________  Number of Females ________

Complete any part of this table that isn’t covered by the handouts collected from the teacher.

<table>
<thead>
<tr>
<th>Major activity or activities planned</th>
<th>Purpose (Topics, Concepts, Skills or Learning Target(s))</th>
<th>Assessment Strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Homework:

Copies of instructional materials obtained?
Record observations to help complete ratings (use new page for new lesson segment)

<table>
<thead>
<tr>
<th>Time</th>
<th>Description of Events</th>
<th>Look for:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>• BE: involved, attentive (participating)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• SR: agency, autonomy, ownership (initiates, independent, innovates, interest)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Teacher questions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Student questions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Cognitive level</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Core science content</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Science meaning-making processes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Sustained discourse</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Collaborative discourse</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Revisions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Representations used</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Metacognitive facilitation</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Form Assmt/Adjustment</td>
</tr>
</tbody>
</table>

End of Segment:  
Community Partners:  
☐ Present  ☐ Collaborating  ☐ Not Present  
Science inquiry focused?  
☐ Yes  ☐ No  
Tchr talk =  
Stdnt talk

Science Inquiry Processes (A = authentic/S = simple)  
RQs _____ Design _____ Make _____ Explain _____ Develop _____ Study _____  
Study _____ Obs _____ Results _____ Theories _____ Others _____  

Wh Class ☐  
Sm. Grp ☐  
Individual ☐
Lesson Authenticity Indicators Worksheet

Observer __________  Date _______ Per ___  Class _____ Total LSs ___  LSs skipped ___

The general meaning of a high rating is that the lesson segment is designed to afford all students participation in the processes of learning in the characteristic. This affordance is prominent, that is, it is a key strategy in the segment. A moderate rating indicates that there is some affordance for this characteristic. A low score indicates that there is no or minimal affordance.

### Lesson Authenticity Indicators

<table>
<thead>
<tr>
<th>Rating</th>
<th>Low</th>
<th>High</th>
<th>Don’t know</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>10.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>11.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

Ratings 1-10: Record for the number of each lesson segment in blank beside its score.

Rating 11: Write ‘a#’ or ‘s#’ where a=authentic inquiry, s=simple inquiry, and # = lesson segment in the blank for the reasoning processes present in the lesson segment. Leave blank those practices not observed. Use the definitions of simple and authentic inquiry in the tables of Reasoning Practices of Science Inquiry (Table B.2) and Reasoning Practices of Engineering Design (Table B.3).
Lesson Authenticity Indicators Scoring Guides

The general meaning of a high rating is that the lesson segment is *designed* to afford all students participation in the reasoning processes in the characteristic. This affordance is prominent, that is, it is a key strategy in the segment. The rationale is that the affordance is strongly supportive of these reasoning processes. A moderate rating indicates that there is some affordance for this characteristic, but it is less explicit or given less emphasis (e.g., time). A low score indicates that the affordance is minimal or not present.

#1. Instruction affords student reasoning practices that support higher order thinking.
Instruction involves students in reasoning practices similar to those used by people practicing science and complex enough to **produce new meanings or understandings for students**. The focus of this indicator is on quality of student thinking during the lesson and whether it involves students in gaining insights to new relationships (constructing knowledge) rather than just reproducing knowledge.

<table>
<thead>
<tr>
<th>Low Rating (lower order thinking - LOT)</th>
<th>High Rating (higher order thinking - HOT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students receive or recite factual information or employ rules and algorithms through repetitive routines. Students receive or recall <strong>teacher-specified knowledge</strong>. Instruction affords transmission of teacher-specified knowledge or practice of procedural routines. Activities can be somewhat complex, but students only follow pre-specified steps and routines or employ algorithms in a rote fashion.</td>
<td>Instruction affords students to explore connecting facts, observations, concepts, equipment, visual representations, physical environments, or models in ways that <strong>produce new meanings or understandings</strong> for them. Student gain insights to new relationships through manipulating ideas and information by explaining, predicting, solving a problem, evaluating, arriving at a conclusion or an interpretation, synthesizing, or generalizing, all in the context of natural phenomena or an engineering design.</td>
</tr>
<tr>
<td><strong>Low (1)</strong></td>
<td><strong>Moderate (2)</strong></td>
</tr>
<tr>
<td>Instruction affords cognitive activity no more complex than lower order thinking described above.</td>
<td>Instruction is dominated by affordances for LOT, but includes some affordance for higher order thinking or affordances for HOT are unclear/confusing or a lack of resources or time limit opportunity for HOT.</td>
</tr>
</tbody>
</table>
#2. Instruction affords a sustained focus on a core science or engineering concept, practice, or topic.
Students are considering topics that science or engineering practitioners would recognize as scientific and important, that is, the concepts, practices and topics they work with.

<table>
<thead>
<tr>
<th>Low Rating</th>
<th>Moderate (2)</th>
<th>High (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction or activities deal with topics that are not science, or don’t deal with understanding or manipulating core science or engineering topics.</td>
<td>Sustained focus = about 15 minutes or more on understanding a core science topic. Includes consideration of ideas with enough thoroughness to explore connections and relationships affording relatively (in-depth) complex understandings. <strong>High includes consideration of natural phenomena or an engineering design.</strong></td>
<td>There is a sustained focus on core science ideas or practices. Few enough ideas are considered for their use to express relatively complex (in-depth) understanding.</td>
</tr>
<tr>
<td>Low (1)</td>
<td>Students do not deal with understanding or manipulating any core science concept, skill, or topic during the segment OR the segment is dominated (&gt;=75%) by teacher directions.</td>
<td>Students do focus on core science ideas, but for too short a time to construct knowledge about them (~about 15 minutes) OR too many ideas are considered to sustain focus on any one OR the connection to core science ideas is superficial.</td>
</tr>
</tbody>
</table>

#3. Instruction affords students opportunities to try out, reflect on, and then revise their science inquiry questions, designs, or explanations, or their problem solutions or elaborated communication, in ways resembling the kinds of trial and error or revisions of science or engineering practitioners.

<table>
<thead>
<tr>
<th>Lowest Rating (1)</th>
<th>Moderate (2)</th>
<th>High (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The instruction does not ask students to revise their ideas or retry a skill, OR there is only a reflection without revision, OR it is clear that any “revision” will not be implemented, that is, the revision is just an exercise with little meaning beyond completing the exercise.</td>
<td>Revisions are a minor part of the segment, or written revisions occur without affordance for feedback from peers or an adult. Segment may be dominated by discussion of revision if there is still clear intent to implement the revision. During student directed segments, score 2 if revisions are an option and less than about two-thirds of engaged students do them.</td>
<td>Revisions are a prominent feature of the lesson segment. That is, they take up most of the time, or are a key instructional strategy, and almost all of engaged students participate in some revision. The revisions are implemented during this segment.</td>
</tr>
</tbody>
</table>
#4. Instruction affords interaction with, use of, or creation of a variety of representations to illustrate or explain phenomena, data, theories, or conclusions, or communicate science to others.

**Explanatory comments**

Interaction means there is an element of student input and response to a subsequent output, such as, controlling a computer simulation, building a computer-based model, or creation and critique of a representation. Representations include physical or computer models, drawings, images, graphs, physical materials, symbolic representations, mathematical representations, simulations, video, animations, etc.

<table>
<thead>
<tr>
<th>Lowest Rating (1)</th>
<th>Moderate (2)</th>
<th>High (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The lesson does not afford the use of visual or physical representations of scientific knowledge, data, or of a design; OR no students interact with representations.</td>
<td>Students passively receive one or more representations of scientific knowledge, data, or of a design, but they do not manipulate or create them as part of meaning making.</td>
<td>Students are afforded interaction with, that is, use of, interpretation of, or creation of at least one type of representation beyond text. Alternatively, the teacher affords student understanding by using more than one representation beyond text to support meaning making.</td>
</tr>
</tbody>
</table>

#5. There was sustained discourse among students or between students and teacher or other adults.

**Explanatory comments**

This measures whether instruction relies on discourse as the primary means of construction of knowledge and/or meaning making during the observed segment. Sustained is taken to mean that multiple, connected 2-way interchanges are common. Subject of discourse can be anything related to lesson content.

<table>
<thead>
<tr>
<th>Low (1)</th>
<th>Moderate (2)</th>
<th>High (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>There is no sustained discourse.</td>
<td>There is some sustained discourse, but it occurs infrequently - less than half of the students participate or occurs during less than half of the lesson segment.</td>
<td>There is sustained discourse involving most of the students during most of the lesson segment.</td>
</tr>
</tbody>
</table>

#6. Instruction affords collaborative discourse among peers, teachers, and/or community partners that makes each participant’s thinking “visible”.

**Explanatory comments**

Discourse that makes students thinking visible means that students are sharing how they think about the science content in the lesson and responding to each other, as opposed to trying to give the “right” or desired answer. The focus is on sharing their thoughts during meaning-making processes. Such discourse may occur in pairs, small groups, or in a presentation or discussion involving the whole class.

<table>
<thead>
<tr>
<th>Low (1)</th>
<th>Moderate (2)</th>
<th>High (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students may be conversing with each other, but they are not sharing their thinking with others.</td>
<td>The role of discourse in meaning making is limited because sharing is one-way without negotiation or mutual comparison, or it is not the primary strategy for meaning making in the lesson (less than half the students are sharing or it occurs during less than half the segment).</td>
<td>Discourse is a primary strategy for meaning making during the lesson. Most students share their thinking and compare it with others during most of the segment. Sharing involves some of the characteristics above.</td>
</tr>
</tbody>
</table>
#7. Instruction models or otherwise “facilitates the use of metacognitive strategies to identify, monitor, and regulate learning” (W. Becker, 2008).

<table>
<thead>
<tr>
<th>Explanatory comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support for reflection (justifying a high rating) could include the use of metacognitive scaffolds that through repetition have become routine rather than explicit.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low (1)</th>
<th>Moderate (2)</th>
<th>High (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instruction does not model, or give them any time, or explicitly encourage students to reflect on their own learning.</td>
<td>Instruction discusses or models metacognitive strategies but does not explicitly scaffold the use of them. Or instruction calls for reflection, but does not give students the time to do it.</td>
<td>Explicit encouragement and support for student reflection on their learning is a key part of the lesson segment, that is, reflection is scaffolded and students are given time for it. Reflection is not presented as optional. Instructor modeling could be part of this.</td>
</tr>
</tbody>
</table>

#8. Instruction affords connections between science-related content or issues and either significant public problems or personal experiences in their lives outside of school.

<table>
<thead>
<tr>
<th>Explanatory comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>If the only connection is a general sense of usefulness in future studies, e.g., college, score no more than a 2. If the issue is presented in a way that students use their own knowledge to influence initial exploration of the topic or to make choices that shape the final product, this contributes to a High score.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low (1)</th>
<th>Moderate (2)</th>
<th>High (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>The lesson topic and activities have either no clear connection to anything beyond themselves or no science content. The teacher may try to connect to students’ experiences or to a contemporary public situation; i.e., the teacher may inform the students that there is unspecified potential value in the knowledge being studied because it relates to the world beyond the classroom; however, the connection is unspecified and there is no evidence that the students make the connection.</td>
<td>Students study a topic, problem, or issue with at least some science content that the teacher connects to students’ actual experiences or to a contemporary public situation. These connections are not made explicit to students. They remain vague, abstract, hypothetical, or optional.</td>
<td>Students study or work on a science-related topic, problem, or issue that the teacher or learning materials explicitly connects to actual contemporary public situations or student personal experiences. The connections between classroom activities or learning products and situations outside the classroom are explicit and are not optional.</td>
</tr>
</tbody>
</table>

#9. Students are communicating or are preparing to communicate their knowledge, investigate a problem, or take some action to influence or assist others beyond the teacher and classroom.

**Explanatory comments**

Score Moderate (2) if a lesson segment includes preparation of learning products where presentation with the intention of influencing an audience beyond the classroom is an option, and fewer than about three-fourths of students choose that option. If the instruction explicitly specifies a product that is a written to an external audience, but not actually presented, may be scored as 2.

<table>
<thead>
<tr>
<th>Low (1)</th>
<th>Moderate (2)</th>
<th>High (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Students are not involved in nor preparing to communicate their knowledge, investigate a problem, or take some action to influence or assist others beyond the teacher and classroom.</td>
<td>Students are presenting or creating learning products that will be presented to an audience beyond their classroom, with the intention of illustrating their learning progress. An example would be preparing for student-led conferences with parents.</td>
<td>Students are presenting or creating learning products with the intention to influence a larger audience beyond their classroom in one of the following ways: by communicating knowledge to others (including within the school), advocating solutions to social problems, providing assistance to people, or creating performances of products with utilitarian or aesthetic value.</td>
</tr>
</tbody>
</table>


#10. Instruction affords students primary responsibility for managing how they spend classroom time, the learning strategies and resources they use, or how they show proficiency.

**Explanatory comments**

The high rating requires very limited teacher intervention and that students are explicitly put in charge of most of the decisions regarding how class time is used and some decisions related to product content. Includes a role in how inquiry question is investigated. A high rating limits teacher intervention to limited "prompting", while a 2 may include more prompting and some "guiding". Teacher “prompting” to revise below refers to a reminder or provision of scaffolding on factors to consider (including, for example, a model graph caption). “Guided” refers to specific teacher suggested revisions or solutions. A High rating (3) should not be given unless the segment is at least 10 minutes long.

<table>
<thead>
<tr>
<th>Low (1)</th>
<th>Moderate (2)</th>
<th>High (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use of class time, knowledge considered, learning strategies used, how questions are investigated, and social interactions are teacher specified. The only real choice the students are given is whether they will do what the teacher is telling them to do. Characterized by lecture, reading, independent practice, or teacher-led question and answer. Score an assessment lesson segment low if the students have no input on how to demonstrate proficiency.</td>
<td>Use of class time, knowledge and questions considered and social interactions are teacher specified, but activities give students some responsibility for learning strategies used, how inquiry questions are investigated. The teacher’s role includes “guiding”. May include teacher-supervised discussion with students responding to each other, cooperative learning, guided inquiry activities, verification laboratory, or jigsaw. Score an assessment lesson segment moderate if the students have partial responsibility for how to demonstrate proficiency.</td>
<td>Students are given primary responsibility for decisions about use of class time, learning strategies they use, or researching a topic. Score 3 if any one of these or a combination dominates the segment and the teacher’s role is limited to “prompting” with students loosely supervised. Characterized by open-ended inquiry, project work, independent research, student led discussion, or student directed preparation for an assessment. Score an assessment lesson segment high only if the students have full responsibility for how to demonstrate proficiency.</td>
</tr>
</tbody>
</table>
Table B.1 Reasoning Practices of Science Inquiry

When instruction in the lesson segment goes beyond simply inquiry to resemble authentic inquiry at a 9th grade level, score authentic. The instrument described in this table is a modified version of one reported in a doctoral dissertation entitled: Student Cognition and Motivation During the Classroom BirdWatch Citizen Science Project, (Tomasek, 2006), which was based on of Chinn and Malhotra’s (2002) comparison of the cognitive processes used by scientists with those most frequently found in school curriculum materials. The regular text below is from Tomasek. Modifications and additions are in italics.

<table>
<thead>
<tr>
<th>Reasoning Practices</th>
<th>Simple Inquiry examples</th>
<th>Authentic Scientific Inquiry examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Generating Research Questions</td>
<td>Questions or hypotheses provided by teacher or curriculum</td>
<td>Develop own questions based on observations, the work of others’, community concerns, or personal curiosity or interests. Questions or hypotheses are testable.6</td>
</tr>
<tr>
<td>b. Designing a Study</td>
<td>Single pre-defined independent and dependent variable Observing prescribed features No controlled conditions Single controlled condition Single prescribed outcome measure Prescribed steps</td>
<td>Variables selected by researcher Steps designed or invented by students Experimentation may be done on an analog model, when advantageous.1 Controls selected based on causal models Multiple controlled factors or conditions External controls verifying procedures or equipment Intervening and multiple outcomes Critical reflection on methods Work on ill-defined problems4</td>
</tr>
<tr>
<td>c. Making Observations</td>
<td>Making specified measurements Data quality evaluation limited to simple measurement 'errors' (e.g., correct reading of instrument)</td>
<td>Evaluate data quality (e.g., employ technique to guard against perceptual bias) Employ professional measurement tools6 Legitimate peripheral participation/apprenticeship with professional(s); “at the elbows of scientists”5</td>
</tr>
<tr>
<td>d. Explaining Results</td>
<td>Data transformation, drawing Data transformation, graphing No data transformation Simple error analysis Direct reasoning Indirect reasoning defined by text Manipulated variables same as theoretical variables Simple contrastive or causal reasoning Simple inductive reasoning Simple deductive reasoning</td>
<td>Data transformation, statistics Data transformation, other Complex error analysis (methodological flaws) Indirect reasoning Complex chain of inferences to defend knowledge claims6 Manipulated variables different from theoretical variables Complex reasoning, responses to anomalous data: consider alternative mechanism to explain results ignore, reject or express degree of uncertainty about data Complex reasoning, consider ways to verify validity of methods Generalizations to new situations Use professional tools for analysis of data.3 Employ appropriate technology.6</td>
</tr>
<tr>
<td>e. Developing Theories</td>
<td>Direct observations not connected to theory Direct observations illustrating stated theory</td>
<td>Construct or revise3 theoretical explanations based on evidence Study at level of observable regularity Study at level of mechanism (may be unobservable) Coordinate results from multiple studies Strategies to resolve inconsistencies in multiple studies</td>
</tr>
<tr>
<td>f. Studying Others’ Research</td>
<td>Reading about topic in science tradebooks</td>
<td>Reading research of others Building on the work of others Peer review of findings</td>
</tr>
<tr>
<td>Reasoning Practices</td>
<td>Simple Inquiry examples</td>
<td>Authentic Scientific Inquiry examples</td>
</tr>
<tr>
<td>---------------------</td>
<td>-------------------------</td>
<td>--------------------------------------</td>
</tr>
<tr>
<td>g. Discourse during inquiry</td>
<td>Small groups discuss completing activity by following instructions given. Ask teacher how to complete activity by following instructions given.</td>
<td>Develop and share inquiry ideas and resources with peers. Collaboration with other students, community members, and/or scientists. May include negotiation with and/or challenging these colleagues. Membership in a research community through shared knowledge, practices, resources, and discourses. Communicate and defend results. Use feedback from scientists, teachers or peers to improve written communication.</td>
</tr>
</tbody>
</table>

1 From Chinn and Malhotra’s analysis of cognitive processes used by scientists and the epistemology of science (Chinn & Malhotra, 2002).  
2 Based on Expeditionary Learning Core Practice Benchmarks (Expeditionary Learning Outward Bound, 2003).  
3 Based on Rahm et al. (Rahm et al., 2003)  
4 Based on Roth (Roth, 1995).  
5 Based on White and Fredricksen’s work on metacognitive facilitation (White & Fredricksen, 2000).  
6 Based on Evaluation Criteria for a Student-Centered Science Inquiry Activity (Becker, 2008).
Table B.2 Reasoning Practices of Engineering Design

When instruction in the lesson segment goes beyond simply inquiry to resemble authentic inquiry at a 9th grade level, score authentic. The reasoning practice levels are defined based on several primary sources.

<table>
<thead>
<tr>
<th>Reasoning Practices</th>
<th>Simple Design examples</th>
<th>Authentic Design examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Identifying a Design Opportunity</td>
<td>Design criteria and constraints provided by teacher or curriculum, and limited to physical constraints. Design need is focused on completion of the task.</td>
<td>Develop a design opportunity based on personal curiosity or interests; or needs of users. Begin with ill-defined problems. Development of design specifications is based on others’ work Design opportunity sets multiple classes of constraints, such as, scientific, physical, environmental, economic, social, ethical, and professional.</td>
</tr>
<tr>
<td>b. Generating Design Solutions</td>
<td>Limited design alternatives are provided or suggested. Only one or two design alternatives generated. Creativity and innovation are severely limited by time, materials, or equipment. Evaluation of design alternatives based on hunches.</td>
<td>Many design alternatives are brainstormed. Includes decomposing the system into interacting components. Evaluation of design alternatives based on knowledge of previous designs. Evaluation of design alternatives uses conceptual models. Evaluation of design alternatives is explicitly based on design specifications and systematic evaluation of tradeoffs.</td>
</tr>
<tr>
<td>d. Testing and Evaluating Design Solutions</td>
<td>Making specified measurements. Testing plan provided Testing is trial and error Testing results are qualitative Only the overall function of design is tested. Evaluation of design alternatives based on results for overall function Evaluation is based on one or two design specification or constraints Direct reasoning Indirect reasoning defined by text</td>
<td>Testing plan, what and how tested, designed by the engineer. Testing plan based on design analysis and predictions. Mathematics used to test or evaluate design. Both components and systems tested for design performance. Testing of design includes reliability and robustness. Multiple design alternatives tested. Multiple controlled factors or conditions. Complex chain of inferences to evaluate testing results. Evaluation of design explicitly based on tradeoffs, multiple design specifications and constraints, such as, feasibility, safety, and aesthetics. Employ professional measurement and data analysis tools. Employ appropriate technology.</td>
</tr>
<tr>
<td>Reasoning Practices</td>
<td>Simple Design examples</td>
<td>Authentic Design examples</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>e. Revising Design Solutions</td>
<td>Redesign does not consider results of testing components or subsystems. Redesign based on trial &amp; error. Return to previous practice (a-d) as instructed. Iterate based on instructions, not design function.</td>
<td>Revise design solution model from test results. Revise design based on analysis from improved model of design solution. Revise only components needing revision. Decide which previous practice (a-d) to return to. Continue to iterate until design meets specifications.</td>
</tr>
<tr>
<td>f. Studying Others’ Research and Designs</td>
<td>Reading about topic in science or engineering tradebooks (address topics generally). Design constraints, alternatives, or technical background supplied by teacher. Learn from the experiences of classmates.</td>
<td>Reviewing design solutions to similar problems and the research of others. Building on the research of others. Get feedback on design from people in many roles (engineering colleagues, scientists, customers, end users, craftspeople, regulators, and people who market and sell products). Give constructive feedback on the written or oral communication of peers. Identify pre-existing solutions or components available for use.</td>
</tr>
<tr>
<td>g. Discourse During Design</td>
<td>Small groups discuss completing activity by following instructions given. Ask teacher how to complete activity by following instructions given. Written reporting limited to what was done and what happened.</td>
<td>Innovation and creativity facilitated by consultation in a team. Develop and share design ideas and resources with colleagues. Collaboration with other students, community members, clients, and/or engineers. May include negotiation with and/or challenging these colleagues. Membership in an engineering community through shared knowledge, practices, resources, and discourses. Communicate and defend design process, the chosen design solution, and ethics. Use feedback from customers, engineers, scientists, technicians, teachers or peers to improve written communication.</td>
</tr>
</tbody>
</table>

1 From Chinn and Malhotra’s analysis of cognitive processes used by scientists and the epistemology of science (Chinn & Malhotra, 2002).
2 Engineering in K-12 Education (Katehi, 2009)
3 Based on Rahm et al. (Rahm et al, 2003)
4 Based on Roth (Roth, 1995).
5 Based on Evaluation Criteria for a Student-Centered Science Inquiry Activity (Becker, 2008)
6 Standards for Technological Literacy (Technology for All Americans Project, 2007)
Appendix C  Supplementary Assessment Tasks Tables

This appendix contains tables C.1 and C.2 which include the full data on authenticity measures of all 22 assessment tasks scored over the study period.
Table C.1. Task Authenticity Indicator (TAI) scores for 22 teacher proficiency assessment tasks from across the entire study period. Tasks are ordered by combined authenticity score, then chronological order during the school year.

<table>
<thead>
<tr>
<th>Learning Target(s)</th>
<th>Assessment Task</th>
<th>Construction of Knowledge</th>
<th>Meaning-Making Processes</th>
<th>Value Beyond School</th>
<th>Combined Authenticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5, 6, 7, 8</td>
<td>Wind Turbine Engineering &amp; Testing</td>
<td>3.0 3.0 2.7 3.0 4.0 2.7 1.3</td>
<td></td>
<td></td>
<td>19.7</td>
</tr>
<tr>
<td>1.1, 1.7</td>
<td>Antiviral Drug Design</td>
<td>3.0 3.0 2.0 3.0 4.0 3.0 2.0</td>
<td></td>
<td></td>
<td>19.0</td>
</tr>
<tr>
<td>1.6</td>
<td>RNA Fingerprinting Lab Analysis</td>
<td>3.0 2.0 3.0 3.0 4.0 3.0 1.0</td>
<td></td>
<td></td>
<td>19.0</td>
</tr>
<tr>
<td>2.6, 7, 8</td>
<td>Memory Experiment Data Analysis</td>
<td>3.0 3.0 3.0 3.0 4.0 2.0 1.0</td>
<td></td>
<td></td>
<td>19.0</td>
</tr>
<tr>
<td>1.3, 1.7</td>
<td>Viral Evolution Essays</td>
<td>3.0 2.0 3.0 2.0 4.0 3.0 1.0</td>
<td></td>
<td></td>
<td>18.0</td>
</tr>
<tr>
<td>2.6, 7, 8</td>
<td>Human Characteristics Data Analysis tasks</td>
<td>3.0 3.0 2.5 3.0 3.0 2.0 1.0</td>
<td></td>
<td></td>
<td>17.5</td>
</tr>
<tr>
<td>1.5</td>
<td>Flu Vaccines Public Service Announcement</td>
<td>2.0 3.0 2.0 3.0 3.0 2.0 1.0</td>
<td></td>
<td></td>
<td>17.0</td>
</tr>
<tr>
<td>2.4, 2.8</td>
<td>Teen Brain Public Service Announcement</td>
<td>3.0 1.0 2.0 2.0 3.0 3.0 3.0</td>
<td></td>
<td></td>
<td>17.0</td>
</tr>
<tr>
<td>3.5, 6, 8, 10</td>
<td>Engineering Challenge 1</td>
<td>3.0 2.0 2.0 2.7 3.0 1.3 2.0</td>
<td></td>
<td></td>
<td>16.0</td>
</tr>
<tr>
<td>1.6</td>
<td>RNA Fingerprinting Quiz</td>
<td>2.0 2.0 3.0 2.0 3.0 2.0 1.0</td>
<td></td>
<td></td>
<td>15.0</td>
</tr>
<tr>
<td>2.2 &amp; 1.7</td>
<td>Genetic Disease Project</td>
<td>2.0 1.0 2.0 2.0 3.0 2.0 3.0</td>
<td></td>
<td></td>
<td>15.0</td>
</tr>
<tr>
<td>1.3</td>
<td>Viral Evolution Quiz</td>
<td>2.0 2.0 2.0 2.0 2.0 3.0 1.0</td>
<td></td>
<td></td>
<td>14.0</td>
</tr>
<tr>
<td>1.4</td>
<td>Immune System PSA</td>
<td>2.0 1.0 2.0 2.0 3.0 2.0 2.0</td>
<td></td>
<td></td>
<td>14.0</td>
</tr>
<tr>
<td>1.5</td>
<td>Flu Vaccines Quiz</td>
<td>2.0 2.0 2.0 2.0 2.0 3.0 1.0</td>
<td></td>
<td></td>
<td>14.0</td>
</tr>
<tr>
<td>3.1</td>
<td>Simple Machines Lab Analysis</td>
<td>2.0 2.0 2.0 3.0 3.0 1.0 1.0</td>
<td></td>
<td></td>
<td>14.0</td>
</tr>
<tr>
<td>1.1</td>
<td>Viral Structure &amp; Replication Quiz</td>
<td>2.0 2.0 2.0 1.0 2.0 2.0 1.0</td>
<td></td>
<td></td>
<td>12.0</td>
</tr>
<tr>
<td>1.4</td>
<td>Immune System Quiz</td>
<td>2.0 1.0 2.0 2.0 2.0 2.0 1.0</td>
<td></td>
<td></td>
<td>12.0</td>
</tr>
<tr>
<td>1.2</td>
<td>Replication, Transcription &amp; Translation Quiz</td>
<td>2.0 2.0 2.0 1.0 2.0 1.0 1.0</td>
<td></td>
<td></td>
<td>11.0</td>
</tr>
<tr>
<td>3.2</td>
<td>Energy Concepts Quiz</td>
<td>2.0 2.0 2.0 1.0 2.0 1.0 1.0</td>
<td></td>
<td></td>
<td>11.0</td>
</tr>
<tr>
<td>2.3</td>
<td>Neuron Function Quiz</td>
<td>2.0 1.0 1.0 1.0 1.0 2.0 1.0</td>
<td></td>
<td></td>
<td>10.0</td>
</tr>
<tr>
<td>3.1</td>
<td>Mechanics Concepts Quiz</td>
<td>2.0 1.0 2.0 1.0 2.0 1.0 1.0</td>
<td></td>
<td></td>
<td>10.0</td>
</tr>
<tr>
<td>2.1</td>
<td>Genes to Traits Quiz</td>
<td>1.0 1.0 2.0 1.0 2.0 1.0 1.0</td>
<td></td>
<td></td>
<td>9.0</td>
</tr>
<tr>
<td>Mean of all Tasks</td>
<td></td>
<td>2.3 1.9 2.2 2.0 2.8 2.1 1.4</td>
<td></td>
<td></td>
<td>14.7</td>
</tr>
<tr>
<td>Mean normalized to a 0-1 scale</td>
<td></td>
<td>0.66 0.45 0.59 0.49 0.61 0.55 0.13</td>
<td></td>
<td></td>
<td>0.48</td>
</tr>
</tbody>
</table>

*Note.* The range for ATIs 1, 2, 3, 4 and 6 is 1-3. The range for ATIs 5 and 7 is 1-4.
Table C.2. Normalized (0-1) component and combined task authenticity scores for 22 teacher proficiency assessment tasks. Tasks are ordered by combined authenticity score, then chronological order during the school year.

<table>
<thead>
<tr>
<th>Learning Target(s)</th>
<th>Assessment Task</th>
<th>Construction of Knowledge</th>
<th>Science Making Processes</th>
<th>Value Beyond School</th>
<th>Combined Authenticity</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.5, 6, 7, 8</td>
<td>Wind Turbine Engineering &amp; Testing</td>
<td>1.00</td>
<td>0.95</td>
<td>0.40</td>
<td>0.79</td>
</tr>
<tr>
<td>1.1, 1.7</td>
<td>Antiviral Drug Design</td>
<td>1.00</td>
<td>0.71</td>
<td>0.60</td>
<td>0.75</td>
</tr>
<tr>
<td>1.6</td>
<td>RNA Fingerprinting Lab Analysis</td>
<td>0.75</td>
<td>1.00</td>
<td>0.40</td>
<td>0.75</td>
</tr>
<tr>
<td>2.6, 7, 8</td>
<td>Memory Experiment Data Analysis</td>
<td>1.00</td>
<td>1.00</td>
<td>0.20</td>
<td>0.75</td>
</tr>
<tr>
<td>1.3, 1.7</td>
<td>Viral Evolution Essays</td>
<td>0.75</td>
<td>0.86</td>
<td>0.40</td>
<td>0.69</td>
</tr>
<tr>
<td>2.6, 7, 8</td>
<td>Human Characteristics Data Analysis</td>
<td>1.00</td>
<td>0.79</td>
<td>0.20</td>
<td>0.66</td>
</tr>
<tr>
<td>1.5</td>
<td>Flu Vaccines Public Service Announcement</td>
<td>0.75</td>
<td>0.57</td>
<td>0.60</td>
<td>0.63</td>
</tr>
<tr>
<td>2.4, 2.8</td>
<td>Teen Brain Public Service Announcement</td>
<td>0.50</td>
<td>0.57</td>
<td>0.80</td>
<td>0.63</td>
</tr>
<tr>
<td>3.5, 6, 8</td>
<td>Engineering Challenge 1</td>
<td>0.75</td>
<td>0.67</td>
<td>0.27</td>
<td>0.56</td>
</tr>
<tr>
<td>1.6</td>
<td>RNA Fingerprinting Quiz</td>
<td>0.50</td>
<td>0.71</td>
<td>0.20</td>
<td>0.50</td>
</tr>
<tr>
<td>2.2 &amp; 1.7</td>
<td>Genetic Disease Project</td>
<td>0.25</td>
<td>0.57</td>
<td>0.60</td>
<td>0.50</td>
</tr>
<tr>
<td>1.3</td>
<td>Viral Evolution Quiz</td>
<td>0.50</td>
<td>0.43</td>
<td>0.40</td>
<td>0.44</td>
</tr>
<tr>
<td>1.4</td>
<td>Immune System Public Service Announcement</td>
<td>0.25</td>
<td>0.57</td>
<td>0.40</td>
<td>0.44</td>
</tr>
<tr>
<td>1.5</td>
<td>Flu Vaccines Quiz</td>
<td>0.50</td>
<td>0.43</td>
<td>0.40</td>
<td>0.44</td>
</tr>
<tr>
<td>3.1</td>
<td>Simple Machines Lab Analysis</td>
<td>0.50</td>
<td>0.71</td>
<td>0.00</td>
<td>0.44</td>
</tr>
<tr>
<td>1.1</td>
<td>Viral Structure &amp; Replication Quiz</td>
<td>0.50</td>
<td>0.29</td>
<td>0.20</td>
<td>0.31</td>
</tr>
<tr>
<td>1.4</td>
<td>Immune System Quiz</td>
<td>0.25</td>
<td>0.43</td>
<td>0.20</td>
<td>0.31</td>
</tr>
<tr>
<td>1.2</td>
<td>Replication, Transcription &amp; Translation Quiz</td>
<td>0.50</td>
<td>0.29</td>
<td>0.00</td>
<td>0.25</td>
</tr>
<tr>
<td>3.2</td>
<td>Energy Concepts Quiz</td>
<td>0.50</td>
<td>0.29</td>
<td>0.00</td>
<td>0.25</td>
</tr>
<tr>
<td>2.3</td>
<td>Neuron Function Quiz</td>
<td>0.25</td>
<td>0.14</td>
<td>0.20</td>
<td>0.19</td>
</tr>
<tr>
<td>3.1</td>
<td>Mechanics Concepts Quiz</td>
<td>0.25</td>
<td>0.29</td>
<td>0.00</td>
<td>0.19</td>
</tr>
<tr>
<td>2.1</td>
<td>Genes to Traits Quiz</td>
<td>0.00</td>
<td>0.29</td>
<td>0.00</td>
<td>0.13</td>
</tr>
</tbody>
</table>

Mean of all Tasks: 0.56, 0.57, 0.29, 0.48

Range: 0.00-1.00, 0.14-1.00, 0.00-0.80, 0.13-0.79

Note. The range of each of the normalized component scores is 0-1 with 0 indicating no authenticity; 0.5 is moderate authenticity; and 1 indicates high authenticity. The combined authenticity scores are not the sum of the component scores because of the methodology described in Table 5.3.
Appendix D  Behavioral Engagement and Self-Regulated Learning Observation

Protocol

Observation Protocol

At the beginning of the class period, record the position of each student on the class seating map. Behavioral engagement and self-regulated learning are scored during most of the same lesson segments as authenticity of classroom instruction. Appendix B describes the classroom instruction scoring protocol.

Recording behavioral engagement (BE) and self-regulated learning (SR)

During lesson segments when students are grouped as a whole class and are teacher-directed

At 8 to 10 minute intervals, scan the students present for behavioral engagement and record codes in the table beside the class map. Record the start and end times for the scan. Whenever self-regulation is observed throughout the lesson segment, record the code from Table C.1 in the table in the SR column for preceding time point.

During lesson segments where students work individually or in groups

At 8 to 10 minute intervals, scan the participating students for behavioral engagement and self-regulated learning. Record the codes for behavioral engagement and the self-regulated learning presence and category variables in the table beside the class map. There were some of these lesson segments when only behavioral engagement was observed.

In lesson segments chosen for intensive self-regulated learning observations, also make approximately 3-minute observations of 2-6 students seated together at a table between the whole class scans. Record codes for all three variables in the table beside the class map. Systematically switch these observations between tables or groups for each subsequent scan. Table groups of 2-6 students was the normal seating arrangement during these lesson segments.
**Student Behaviors Coding**

**Behavioral engagement (BE) coding**

Recognizing behavioral engagement

Students appear *attentive* or *involved*. *Attentiveness* can be recognized by active listening or responding to teacher questions or those of peers. *Involvement* can be recognized by working on assigned tasks and by asking or answering questions or making comments related to the lesson content.

Coding behavioral engagement

Behaviorally engaged (2) – Any of the student actions above indicating attentiveness or involvement are observed. It is common to see students switch between on-task and off-task behavior during the period of the scan. Spend up to one minute per table to determine if on- or off-task behavior predominates for each student.

Not behaviorally engaged (off-task) (1) – Off-task behaviors are observed.

Ambiguous (0) – Can’t tell whether student is on- or off-task, for example, because student can’t be clearly observed without disrupting the classroom.

**Self-regulated learning (SR) coding**

Recognizing self-regulated learning

Students are self-regulating when they exercise agency, autonomy, or ownership related to the class activities, or learning science in general. There are five types of behavioral evidence: working independently; initiating lesson-related activity or continuing it autonomously; expressing personal interest or value; innovating, that is, going beyond the lesson expectations; and articulating answers, questions, or comments with conceptual content regarding or extending the lesson. Codes used for these behaviors are in Table C.1.

Coding self-regulated learning

SR presence variable

Self-regulating (2) – Student actions indicating autonomy, agency, or ownership above are observed. (Only recorded during teacher-directed lesson segments as Yes when one of the category variable behaviors are observed.)

Not self-regulating (1) – None of the behaviors indicating agency, autonomy, or ownership are observed. (Not recorded during teacher-directed lesson segments.)

Ambiguous (0) – some evidence of self-regulation, but can’t be certain. (Not recorded during teacher-directed lesson segments.)

**SR category variable**

\[
\begin{align*}
SR &= 1 \quad \text{Working independently} \\
SR &= 2 \quad \text{Working autonomously (includes initiates, innovates, interest, ownership, and conceptual questions, answers or comments)}
\end{align*}
\]

Additional specific student behavior codes are listed in Table C.1. Whether they indicated behavioral engagement of self-regulated learning and coding is included.
Table C.1 Student behavior codes recorded.

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>BE Code</th>
<th>pSR Code</th>
<th>cSR Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q</td>
<td>Asks a “question”.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>A</td>
<td>“Answers” a question from teacher or another student.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C</td>
<td>Makes an unsolicited “comment”.</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>sQ/A/C</td>
<td>Prefix for “simple” used with Q, A, or C</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>A question, answer or comment that pertains to the content of the lesson,</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>but is only dealing with lower order thinking, such as, definitions or</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>simple relationships.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pQ/A/C</td>
<td>Prefix for “procedural” used with Q, A, or C</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>A question, answer or comment about the logistics of the lesson or the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>class.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cQ/A/C</td>
<td>Prefix for “conceptual” used with Q, A, or C</td>
<td>yes</td>
<td>yes</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>A question, answer or comment that exhibits interaction with or</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>extension of the conceptual content of the lesson.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>“initiates” or works autonomously</td>
<td>yes</td>
<td>yes</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Student starts or maintains lesson-related activity or makes lesson-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>related choices with minimal teacher prompting or direction in a way</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>that goes beyond just following directions. Indicates ownership and</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>agency.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>works “independently”</td>
<td>yes</td>
<td>yes</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Student works on lesson in response to teacher direction without</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>noticeable student choice – just following directions, but without</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>continuing teacher direction.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>v</td>
<td>“innovates”</td>
<td>yes</td>
<td>yes</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Unsolicited student discourse or action beyond specific lesson</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>directions or content, that extends the activity. Indicates ownership and</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>agency.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t or w</td>
<td>“interest” or “ownership”</td>
<td>yes</td>
<td>yes</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Expresses personal interest in or the personal value of the activity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>spontaneously without teacher prompting.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>r</td>
<td>“reading” non-relevant mats. Evidence of being off-task.</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>oc</td>
<td>Working on classwork from “another class”. Evidence of being off-task.</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>ph</td>
<td>Attending to “phone” instead of attending to class activity. Evidence of</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td></td>
<td>being off-task.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note.


cSR – category of self-regulated learning variable.

s - For example: “Is this a bar graph?” or “Does attachment come before endocytosis?” The discourse context of the question matters.

p - For example: “will we have to turn this in?” or “Where do you want us to record notes on the reading?” Or “Is it time to start the lab?” or “When will we have another opportunity to pass this learning target?”

c - For example: "why is a bar graph better for this data?” or “A new vaccine is developed each year because viruses mutate quickly.”
Record of Behavioral Engagement and Self-Regulation at regular time points

Codes BE as on-task (2), off-task (1), ambiguous (0). Codes SR as self-regulating (2), not self-regulating (1), ambiguous (0) = SR codes.

<table>
<thead>
<tr>
<th>Time or Start Time</th>
<th>Score Start Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1</td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td></td>
</tr>
<tr>
<td>T3</td>
<td></td>
</tr>
<tr>
<td>T4</td>
<td></td>
</tr>
<tr>
<td>T5</td>
<td></td>
</tr>
<tr>
<td>T6</td>
<td></td>
</tr>
<tr>
<td>T7</td>
<td></td>
</tr>
<tr>
<td>T8</td>
<td></td>
</tr>
</tbody>
</table>

Scan End Time
Appendix E  Supplementary Behavioral Engagement, Self-Regulated Learning, and Student Performance Tables

Table E.1. Mean and median percent observations behaviorally engaged (BE) by gender over the entire study period.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Mean % BE/Student</th>
<th>Median % BE/Student</th>
<th>SD</th>
<th>Number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>75</td>
<td>73</td>
<td>11.8</td>
<td>32</td>
</tr>
<tr>
<td>Female</td>
<td>73</td>
<td>76</td>
<td>14.8</td>
<td>51</td>
</tr>
<tr>
<td>Total</td>
<td>74</td>
<td>74</td>
<td>13.8</td>
<td>83</td>
</tr>
</tbody>
</table>

Table E.2. Means and medians of percent of observations per student from the entire study period behaviorally engaged (BE) from each ethnic category.

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Mean % BE/student</th>
<th>Median % BE/student</th>
<th>SD</th>
<th>Number of students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian or native Hawaiian</td>
<td>84</td>
<td>79</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Hispanic or Latino</td>
<td>73</td>
<td>75</td>
<td>13</td>
<td>12</td>
</tr>
<tr>
<td>Multiple racial categories selected</td>
<td>72</td>
<td>71</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>Other race</td>
<td>81</td>
<td>79</td>
<td>7</td>
<td>6</td>
</tr>
<tr>
<td>Euro-American</td>
<td>77</td>
<td>76</td>
<td>10</td>
<td>24</td>
</tr>
<tr>
<td>All students</td>
<td>77</td>
<td>77</td>
<td>11</td>
<td>53</td>
</tr>
</tbody>
</table>

*Note.* Students with less than 32 observations per trimester from at least two trimesters and students without consent to use their ethnicity are not included.
Table E.3. Mean of percent of observations per student from the entire study period behaviorally engaged (BE) from each mean 8th grade science and math grade category.

<table>
<thead>
<tr>
<th>Mean of 8th grade science &amp; math grades</th>
<th>Mean % BE/student</th>
<th>Median % BE/student</th>
<th>Number of students</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>81</td>
<td>79</td>
<td>19</td>
<td>8.4</td>
</tr>
<tr>
<td>B</td>
<td>79</td>
<td>77</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>C</td>
<td>72</td>
<td>71</td>
<td>8</td>
<td>10</td>
</tr>
<tr>
<td>D</td>
<td>67</td>
<td>67</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>54</td>
<td>56</td>
<td>3</td>
<td>5.4</td>
</tr>
<tr>
<td>All Students</td>
<td>77</td>
<td>77</td>
<td>45</td>
<td>11</td>
</tr>
</tbody>
</table>

Table E.4. Mean percent behavioral engagement per student per trimester for class periods in each trimester.

<table>
<thead>
<tr>
<th>School Trimester</th>
<th>Mean % BE/student</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Period 2</td>
<td>Period 3</td>
<td>Period 5</td>
</tr>
<tr>
<td>Fall</td>
<td>79</td>
<td>80</td>
<td>72</td>
</tr>
<tr>
<td>Winter</td>
<td>77</td>
<td>73</td>
<td>69</td>
</tr>
<tr>
<td>Spring</td>
<td>74</td>
<td>80</td>
<td>68</td>
</tr>
</tbody>
</table>

Note. Only students with at least 32 observations per trimester are included for that trimester.
Table E.5. Combined authenticity of student performance on science work samples for each ethnicity category from four selected instructional units.

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RNA Fingerprinting Lab unit (fall)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian or native Hawaiian (AH)</td>
<td>.94</td>
<td>.94</td>
<td>.08</td>
<td>2</td>
</tr>
<tr>
<td>Hispanic or Latino (HL)</td>
<td>1.00</td>
<td>1.00</td>
<td>.00</td>
<td>2</td>
</tr>
<tr>
<td>Multiple racial categories selected (MC)</td>
<td>.67</td>
<td>.56</td>
<td>.29</td>
<td>3</td>
</tr>
<tr>
<td>Other race (OR)</td>
<td>.61</td>
<td>.61</td>
<td>.08</td>
<td>2</td>
</tr>
<tr>
<td>Euro-American (EA)</td>
<td>.69</td>
<td>.78</td>
<td>.24</td>
<td>5</td>
</tr>
<tr>
<td><strong>Immune System &amp; Flu Vaccines unit (fall)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian or native Hawaiian (AH)</td>
<td>.78</td>
<td>.78</td>
<td>.00</td>
<td>2</td>
</tr>
<tr>
<td>Hispanic or Latino (HL)</td>
<td>.39</td>
<td>.22</td>
<td>.27</td>
<td>6</td>
</tr>
<tr>
<td>Multiple racial categories selected (MC)</td>
<td>.44</td>
<td>.44</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Other race (OR)</td>
<td>.54</td>
<td>.61</td>
<td>.29</td>
<td>3</td>
</tr>
<tr>
<td>Euro-American (EA)</td>
<td>.62</td>
<td>.72</td>
<td>.21</td>
<td>9</td>
</tr>
<tr>
<td><strong>Energy Concepts unit (spring)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian or native Hawaiian (AH)</td>
<td>.44</td>
<td>.44</td>
<td>.20</td>
<td>4</td>
</tr>
<tr>
<td>Hispanic or Latino (HL)</td>
<td>.33</td>
<td>.33</td>
<td>.25</td>
<td>5</td>
</tr>
<tr>
<td>Multiple racial categories selected (MC)</td>
<td>.30</td>
<td>.22</td>
<td>.13</td>
<td>3</td>
</tr>
<tr>
<td>Other race (OR)</td>
<td>.31</td>
<td>.33</td>
<td>.23</td>
<td>4</td>
</tr>
<tr>
<td>Euro-American (EA)</td>
<td>.41</td>
<td>.33</td>
<td>.30</td>
<td>15</td>
</tr>
<tr>
<td><strong>Wind Turbine Engineering &amp; Testing unit (spring)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian or native Hawaiian (AH)</td>
<td>.81</td>
<td>.87</td>
<td>.15</td>
<td>4</td>
</tr>
<tr>
<td>Hispanic or Latino (HL)</td>
<td>.54</td>
<td>.43</td>
<td>.24</td>
<td>4</td>
</tr>
<tr>
<td>Multiple racial categories selected (MC)</td>
<td>.37</td>
<td>.37</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Other race (OR)</td>
<td>.52</td>
<td>.52</td>
<td>.07</td>
<td>3</td>
</tr>
<tr>
<td>Euro-American (EA)</td>
<td>.73</td>
<td>.74</td>
<td>.15</td>
<td>13</td>
</tr>
</tbody>
</table>

*Note.* Students that did not give consent for use of their ethnicity or who didn’t report it are not included.
Table E.6. Combined authenticity of student performance on science work samples for each grade category of mean 8th grade math and science grades from four selected instructional units.

<table>
<thead>
<tr>
<th>Mean 8th grade math and science grades</th>
<th>Combined Authenticity of Student Performance</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA Fingerprinting Lab unit (fall)</td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>A</td>
<td>.83</td>
<td>.89</td>
</tr>
<tr>
<td>B</td>
<td>.44</td>
<td>.44</td>
</tr>
<tr>
<td>C</td>
<td>.70</td>
<td>.67</td>
</tr>
<tr>
<td>Immune System &amp; Flu Vaccines unit (fall)</td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>A</td>
<td>.75</td>
<td>.78</td>
</tr>
<tr>
<td>B</td>
<td>.43</td>
<td>.44</td>
</tr>
<tr>
<td>C</td>
<td>.22</td>
<td>.22</td>
</tr>
<tr>
<td>D</td>
<td>.72</td>
<td>.72</td>
</tr>
<tr>
<td>Energy Concepts unit (spring)</td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>A</td>
<td>.44</td>
<td>.33</td>
</tr>
<tr>
<td>B</td>
<td>.31</td>
<td>.28</td>
</tr>
<tr>
<td>C</td>
<td>.30</td>
<td>.28</td>
</tr>
<tr>
<td>D</td>
<td>.33</td>
<td>.33</td>
</tr>
<tr>
<td>Wind Turbine Engineering &amp; Testing unit (spring)</td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>A</td>
<td>.74</td>
<td>.74</td>
</tr>
<tr>
<td>B</td>
<td>.65</td>
<td>.63</td>
</tr>
<tr>
<td>C</td>
<td>.46</td>
<td>.41</td>
</tr>
</tbody>
</table>

Note. Students that did not give consent for use of their school records or for whom records were not available are not included.
Appendix F  Scoring Authentic Student Performance in Science and Engineering

This instrument is a modified version of those reported in Newmann, F. M., Secada, W. G., & Wehlage, G. (1995). *A guide to authentic instruction and assessment: Vision, standards and scoring*. Madison, WI: Wisconsin Center for Education Research and Research Institute on Secondary Education Reform (RISER) for Youth with Disabilities (2001). *Standards and Scoring Criteria for Assessment Tasks and Student Performance*. Madison, WI: Wisconsin Center for Education Research. The original text of the Newmann, et al. document is in regular text. Quotes of the RISER document are in Calibri font. *Modifications made for this study are in italics.*

These scoring guides will be used to score student work samples from the science class. Each work sample is scored for all 3 indicators.

Guidelines

A. Scores should be based only on evidence in the student’s performance relevant to the criteria. Matters such as whether the student followed directions, neatness, correct spelling, etc. should not be considered unless they are relevant to the criteria.

B. Scores may be limited by tasks which fail to demand analysis or the other criteria, but the scores must be based only upon the student’s work in the sample, unless classroom observation notes provide information on the originality of the student’s work.

C. Scores should take into account what students can reasonably be expected to do at the 9th grade level. However, scores should still be assigned only according to “absolute” criteria in the standards, not relative to other samples …previously scored.

D. When it is difficult to decide between two scores, give the higher score only when a persuasive case can be made that the work meets minimal criteria for the higher score.

E. If the specific wording of the criteria are not helpful in making this judgment, base the score on the general intent or spirit of the standard described in the introductory paragraphs of the indicator.

F. Completion of the task is not necessary to score high.
Construction of Knowledge

Indicator 1: Scientific or Engineering Analysis

Student work demonstrates thinking with scientific or engineering content by organizing, synthesizing, interpreting, hypothesizing, using a model to predict, describing patterns, generalizing from data, constructing or critiquing scientific arguments or models, creating simulations of phenomena or representations of data, or designing engineering solutions or experiments.

Guiding question

This indicator is intended to measure the extent to which students demonstrate cognitive activity that goes beyond mechanically recording or reporting information, or applying rules, definitions, or algorithms. The essential question is whether students demonstrate construction of knowledge through organizing and interpreting information, versus reproduction of knowledge by restating what has been previously given to them or mechanical application of an algorithm.

To score a 3 or 4, the work must appear reasonably original. The scorer should be reasonably confident that no significant portion of the response has been virtually copied from some other source.

Scoring guide

First, identify statements where the student uses an analysis characteristic of science or engineering. Then use all the guidelines below to determine the prevalence of analysis in the work sample.

4 =

- At least three statements indicate that the student has successfully organized, generalized, interpreted, or synthesized specific information exhibiting evidence of scientific or engineering analysis.
- Substantial evidence of analysis.
- Most of the student’s work includes analysis.

3 =

- At least two statements indicate that the student has successfully organized, generalized, interpreted, or synthesized specific information exhibiting evidence of scientific or engineering analysis.
- Moderate evidence of scientific or engineering analysis.
- A central portion of the student’s work includes analysis.
2 =
- At least one statement shows that the student has successfully organized, generalized, interpreted, or synthesized specific information exhibiting evidence of scientific or engineering analysis.
- Some evidence of analysis.
- A small, but not central, portion of the student’s work includes analysis.

1 =
- Almost all statements consist of recording, or reporting specific information, without evidence of the students’ organizing it or reflecting upon it; OR virtually all analysis offered is unsuccessful or in error.
- No evidence of scientific or engineering analysis.

Science or Engineering Meaning-Making Reasoning Processes

Indicator 2: Understanding of Disciplinary Concepts

Student performance demonstrates an understanding of important scientific ideas, concepts, or theories by explicitly using them to explain phenomena, explain the use or meaning of a representation, test a hypothesis, create a design, or address a societal problem. Use of important ideas implies the work includes more than correct use of vocabulary or application of an algorithm, but demonstrates conceptual understanding. Evidence of student understanding includes elaborating on definitions, representing concepts in alternate ways or contexts, making connections to other scientific or engineering concepts, to other disciplines, or to real-world situations.

Guiding question

Does the student show understanding through correct application of at least one of the fundamental ideas relevant to the science or engineering needed in the task?

Prior to scoring, the rater should identify what science or engineering concepts are required to succeed in the task or are relevant to expressing understanding of the task. If the task does not require the use of fundamental science or engineering concepts, and students don’t add related fundamental concepts on their own, the task should be scored 1.

The main issue is the extent to which the student has demonstrated understanding of fundamental science or engineering concepts by correctly applying them to organize, explain, interpret, or represent data; or solve a problem. The score should be based on
the quality of the application of science or engineering concepts, not on the proportion of the work sample that reflects application of the science or engineering concepts. The student does not have to complete the task to show this understanding. A correct answer without work shown can be scored a 3 or 4 if the question or problem clearly requires conceptual understanding for success.

**Scoring guide**

4 = The student demonstrates an exemplary understanding of at least one scientific or engineering concept that is central to the task by correctly using it to organize, explain, interpret, represent, model, or solve a problem. The application of the concept is appropriate and enough detail is included to be certain of understanding.

3 = There is substantial evidence that the student understands at least one scientific or engineering concept that is central to the task. The student applies the concept to the task appropriately; however, there may be minor flaws in the application, or details may be missing.

2 = There is some evidence that the student understands at least one scientific or engineering concept that is central to the task. Where the student used appropriate scientific or engineering concepts, the application of those concepts is clearly flawed or significantly incomplete.

1 = There is little or no evidence that the student understands any of the scientific or engineering concepts that are central to the task, or the scientific concepts that are used are totally inappropriate to the task, or they are applied in inappropriate ways.

**Indicator 3: Elaborated Written Communication**

Student performance demonstrates an elaboration of his or her understanding or explanations through extended writing.

Elaboration consists of two major parts: a conclusion, generalization, or argument regarding natural phenomena, the role of science in a public issue, or engineering design AND support for it, in the form of example, summary, illustration, detail, or reason. Elaboration is coherent when the examples, summaries, illustrations, details, or reasons do indeed provide appropriate, consistent support for the conclusions, generalizations, or arguments.

Consider the extent to which the student presents a clear and convincing explanation or argument. Student-generated commonly used scientific or engineering representations such as graphing, mathematics, drawings, drafting of designs, computer models, mapping (GIS), spreadsheets, or imaging may be a substantial part of the communication.
if they are original. To score high on this standard the student must communicate in writing an accurate and convincing explanation or argument.

Guiding question

Does the scientific explanation or engineering design contain a generalization or argument AND clear, accurate, coherent, and convincing support for it?

Scoring guide

The score should not be based on the proportion of student work that contains generalization or argument but on the quality of the scientific or engineering communication. One generalization, explanation, conclusion, or argument with substantial elaborated support that meets the criteria for a score below is enough.

4 = Includes a scientific explanation or engineering design that contains a generalization or argument AND accurate, coherent, convincing support for it with no significant errors. The communication is well articulated - unambiguous and easily understood.

3 = Includes a scientific explanation or engineering design that contains a generalization or argument AND reasonably coherent and accurate, but less convincing support for it. Either the generalization or argument, or the support is incomplete, ambiguous, or incorrect in minor ways.

2 = Includes a scientific explanation or engineering design with less coherent elaboration because the generalization or argument OR its support is either missing or significantly incomplete. They are incomplete because they have not been finished, they omit a major part of an argument, or they contain several scientific, mathematical, or logical errors. Generally complete, appropriate, and correct representations (e.g., a graph or diagram) should be scored a 2 if they are not integrated into the basis for a generalization or argument OR the support for a generalization or argument.

1 = Scientific explanations, arguments, or representations or engineering designs are absent or, if present, are seriously incomplete, inappropriate, or incorrect. This may be because the task did not ask for argument or explanation, e.g., fill-in-the-blank or multiple-choice questions, or reproducing a simple definition in words or pictures.
## Appendix G Supplementary Correlation Analyses Tables

Table G.1 Distribution of the behavioral engagement (BE) and self-regulated learning (SR) lesson segments (LS) included in the correlation analysis for each instructional unit.

<table>
<thead>
<tr>
<th>Instructional Unit</th>
<th>No. LSs observed</th>
<th>BE</th>
<th>Presence of SR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Viral Evolution</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>RNA Fingerprinting Lab</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Immune System &amp; Flu Vaccines</td>
<td>21</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>Human Characteristics Data Analysis</td>
<td>27</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Genes to Traits</td>
<td>3</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Human Genetic Diseases</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Neuron Function</td>
<td>3</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Memory Experiment Data Analysis</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teen Brain</td>
<td>19</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>Mechanics Concepts and Simple Machines</td>
<td>17</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>Energy Concepts</td>
<td>11</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Engineering Challenge 1</td>
<td>9</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Wind Power</td>
<td>7</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Wind Turbine Engineering &amp; Testing</td>
<td>38</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>170</strong></td>
<td><strong>94</strong></td>
<td></td>
</tr>
</tbody>
</table>
Table G.2. Mean values for the authentic instruction variables correlated with students’ behavioral engagement and self-regulation. Correlation analysis results are in Table 7.3.

<table>
<thead>
<tr>
<th>Correlation variable</th>
<th>Behavioral engagement</th>
<th></th>
<th></th>
<th>Self-regulated learning</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
<td>Mean</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Authenticity of classroom instruction</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined Authenticity/LS</td>
<td>.57</td>
<td>.18</td>
<td>170</td>
<td>.61</td>
<td>.18</td>
<td>94</td>
</tr>
<tr>
<td>Construction of Knowledge component/LS</td>
<td>.69</td>
<td>.24</td>
<td>170</td>
<td>.74</td>
<td>.23</td>
<td>94</td>
</tr>
<tr>
<td>Science Meaning-Making Processes component/LS</td>
<td>.58</td>
<td>.19</td>
<td>170</td>
<td>.61</td>
<td>.20</td>
<td>94</td>
</tr>
<tr>
<td>Substantive Conversation component/LS</td>
<td>.64</td>
<td>.28</td>
<td>170</td>
<td>.68</td>
<td>.25</td>
<td>94</td>
</tr>
<tr>
<td>Value Beyond School component/LS</td>
<td>.37</td>
<td>.23</td>
<td>170</td>
<td>.40</td>
<td>.23</td>
<td>94</td>
</tr>
<tr>
<td>Student response</td>
<td>71%</td>
<td>/LS</td>
<td>170</td>
<td>60%</td>
<td>/LS</td>
<td>21</td>
</tr>
</tbody>
</table>

Note. The mean number of observations of behavioral engagement per lesson segment is 59 ± 31 (SD). The mean lesson segment duration in minutes for the behavioral engagement lesson segments is 32 ± 18 (SD). The mean number of observations of self-regulated learning per lesson segment is 55 ± 29 (SD). The mean lesson segment duration in minutes for the self-regulated learning lesson segments is 37 ± 18 (SD).
Table G.3. Mean values for the Lesson Authenticity Indicator variables correlated with students’ behavioral engagement and self-regulation. Correlation analysis results are in Table G.4.

<table>
<thead>
<tr>
<th>Lesson Authenticity Indicator (LAI)</th>
<th>For behavioral engagement lesson segments</th>
<th>For self-regulated learning lesson segments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>SD</td>
</tr>
<tr>
<td>Higher-order thinking (LAI 1)</td>
<td>.81</td>
<td>.29</td>
</tr>
<tr>
<td>Focus on core concepts (LAI 2)</td>
<td>.76</td>
<td>.28</td>
</tr>
<tr>
<td>Revisions and tinkering (LAI 3)</td>
<td>.33</td>
<td>.36</td>
</tr>
<tr>
<td>Use multiple representations (LAI 4)</td>
<td>.73</td>
<td>.36</td>
</tr>
<tr>
<td>Sustained discourse (LAI 5)</td>
<td>.59</td>
<td>.28</td>
</tr>
<tr>
<td>Collaborative discourse (LAI 6)</td>
<td>.42</td>
<td>.31</td>
</tr>
<tr>
<td>Metacognitive facilitation (LAI 7)</td>
<td>.51</td>
<td>.42</td>
</tr>
<tr>
<td>Connections to world beyond classroom (LAI 8)</td>
<td>.69</td>
<td>.40</td>
</tr>
<tr>
<td>Audience beyond the classroom (LAI 9)</td>
<td>.05</td>
<td>.17</td>
</tr>
<tr>
<td>Student responsibility for learning (LAI 10)</td>
<td>.69</td>
<td>.36</td>
</tr>
<tr>
<td>Science Inquiry &amp; Engineering Design Practices (LAI 11)</td>
<td>.29</td>
<td>.31</td>
</tr>
</tbody>
</table>

*Note.* The number of behavioral engagement and self-regulated learning lesson segments in the analysis were 170 and 94, respectively.
Table G.4. Correlation coefficients (r) for classroom instruction Lesson Authenticity Indicators with behavioral engagement (BE) and self-regulated learning (SR). The student response variables are measured as percent observations per lesson segment (LS) recorded as behavioral engaged or self-regulating for lesson segments from the entire study period with at least 19 BE or SR observations.

<table>
<thead>
<tr>
<th>Lesson Authenticity Indicator (LAI)</th>
<th>% BE/LS</th>
<th>Sig.</th>
<th>% SR/LS</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher-order thinking (LAI 1)</td>
<td>-.054</td>
<td>.485</td>
<td>.108</td>
<td>.302</td>
</tr>
<tr>
<td>Focus on core concepts (LAI 2)</td>
<td>.142</td>
<td>.065</td>
<td>.436**</td>
<td>.000</td>
</tr>
<tr>
<td>Revisions and tinkering (LAI 3)</td>
<td>-.097</td>
<td>.211</td>
<td>-.021</td>
<td>.838</td>
</tr>
<tr>
<td>Use multiple representations (LAI 4)</td>
<td>-.064</td>
<td>.405</td>
<td>.171</td>
<td>.098</td>
</tr>
<tr>
<td>Sustained discourse (LAI 5)</td>
<td>-.003</td>
<td>.971</td>
<td>.037</td>
<td>.721</td>
</tr>
<tr>
<td>Collaborative discourse (LAI 6)</td>
<td>.162*</td>
<td>.034</td>
<td>-.002</td>
<td>.985</td>
</tr>
<tr>
<td>Metacognitive facilitation (LAI 7)</td>
<td>-.168*</td>
<td>.028</td>
<td>-.106</td>
<td>.310</td>
</tr>
<tr>
<td>Connections to world beyond classroom (LAI 8)</td>
<td>.075</td>
<td>.329</td>
<td>.233*</td>
<td>.024</td>
</tr>
<tr>
<td>Audience beyond the classroom (LAI 9)</td>
<td>-.061</td>
<td>.433</td>
<td>.011</td>
<td>.919</td>
</tr>
<tr>
<td>Student responsibility for learning (LAI 10)</td>
<td>-.144</td>
<td>.061</td>
<td>.162</td>
<td>.118</td>
</tr>
<tr>
<td>Science Inquiry &amp; Engineering Design Practices (LAI 11)</td>
<td>.165*</td>
<td>.031</td>
<td>.022</td>
<td>.831</td>
</tr>
</tbody>
</table>

*Note.* There are 170 lesson segments that are included in the BE analysis and 94 lesson segments in the SR analysis.

*p < 0.05. **p < 0.01.
Table G.5. Mean values for the authenticity of student performance variables correlated with the teacher’s assessment task variables in the Proposition 2.1 analysis. Correlation analysis results are in Table G.6.

<table>
<thead>
<tr>
<th>Correlation variable</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Authenticity of Student Performance on Science Work Samples</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Combined Authenticity</td>
<td>.55</td>
<td>.27</td>
<td>96</td>
</tr>
<tr>
<td>PAI 1 - Scientific Analysis (Construction of Knowledge (^a))</td>
<td>.59</td>
<td>.32</td>
<td>96</td>
</tr>
<tr>
<td>PAI 2 - Science Concepts</td>
<td>.54</td>
<td>.27</td>
<td>96</td>
</tr>
<tr>
<td>PAI 3 - Elaborated Communication</td>
<td>.53</td>
<td>.28</td>
<td>96</td>
</tr>
<tr>
<td>Science Meaning Making Processes component</td>
<td>.53</td>
<td>.27</td>
<td>96</td>
</tr>
<tr>
<td><strong>Authenticity of Assessment Tasks</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Combined Authenticity</td>
<td>.52</td>
<td>.23</td>
<td>5</td>
</tr>
<tr>
<td>Construction of Knowledge component</td>
<td>.66</td>
<td>.22</td>
<td>5</td>
</tr>
<tr>
<td>Science Meaning Making Processes component</td>
<td>.62</td>
<td>.29</td>
<td>5</td>
</tr>
<tr>
<td>Value Beyond School component</td>
<td>.26</td>
<td>.21</td>
<td>5</td>
</tr>
</tbody>
</table>

\(^a\) Component of authenticity in this study’s model of Authentic Pedagogy and Performance measured by this indicator.
Table G.6. Correlation coefficients (r) for authenticity variables for the teacher’s assessment tasks and authenticity variables for student performance on work samples. The number of work samples in the correlations is 96 in all cells (N = 96).

<table>
<thead>
<tr>
<th>Mean authenticity of student performance</th>
<th>Authenticity of Assessment Tasks variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Combined Authenticity</td>
</tr>
<tr>
<td></td>
<td>r</td>
</tr>
<tr>
<td>Combined Authenticity</td>
<td>.488**</td>
</tr>
<tr>
<td>PAI 1</td>
<td>.645**</td>
</tr>
<tr>
<td>PAI 2</td>
<td>.382**</td>
</tr>
<tr>
<td>PAI 3</td>
<td>.304**</td>
</tr>
<tr>
<td>Science Meaning Making Processes</td>
<td>.357**</td>
</tr>
</tbody>
</table>

*p < 0.05.  **p < 0.01.
Table G.7. Mean values for the authenticity of student performance variables correlated with the teacher’s classroom instruction variables in the Proposition 2.2 analysis. Correlation analysis results are in Table G.8.

<table>
<thead>
<tr>
<th>Correlation variable</th>
<th>Mean</th>
<th>SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Authenticity of Student Performance on Science Work Samples</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Combined Authenticity</td>
<td>.55</td>
<td>.27</td>
<td>96</td>
</tr>
<tr>
<td>PAI 1 - Scientific Analysis (Construction of Knowledge a)</td>
<td>.59</td>
<td>.32</td>
<td>96</td>
</tr>
<tr>
<td>PAI 2 - Science Concepts</td>
<td>.54</td>
<td>.27</td>
<td>96</td>
</tr>
<tr>
<td>PAI 3 - Elaborated Communication</td>
<td>.53</td>
<td>.28</td>
<td>96</td>
</tr>
<tr>
<td>Science Meaning Making Processes component</td>
<td>.53</td>
<td>.27</td>
<td>96</td>
</tr>
<tr>
<td><strong>Authenticity of Classroom Instruction</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean Combined Authenticity</td>
<td>.44</td>
<td>.12</td>
<td>4</td>
</tr>
<tr>
<td>Mean Construction of Knowledge component</td>
<td>.47</td>
<td>.19</td>
<td>4</td>
</tr>
<tr>
<td>Mean Science Meaning Making Processes component</td>
<td>.52</td>
<td>.11</td>
<td>4</td>
</tr>
<tr>
<td>Mean Value Beyond School component</td>
<td>.28</td>
<td>.18</td>
<td>4</td>
</tr>
<tr>
<td>Mean Substantive Conversation component</td>
<td>.47</td>
<td>.09</td>
<td>4</td>
</tr>
</tbody>
</table>
Table G.8 Correlation correlations (r) for authenticity of the teacher’s classroom instruction and authenticity of student performance on work samples. The classroom instruction variables are mean values for the instructional units corresponding to the student work samples. The number of work samples in the correlations is 96 in all cases (N = 96).

<table>
<thead>
<tr>
<th>Mean authenticity of student performance</th>
<th>Authenticity of Classroom Instruction variables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Combined Authenticity</td>
</tr>
<tr>
<td>Combined Authenticity</td>
<td>.466** .000</td>
</tr>
<tr>
<td>PAI 1</td>
<td>.601** .000</td>
</tr>
<tr>
<td>PAI 2</td>
<td>.351** .000</td>
</tr>
<tr>
<td>PAI 3</td>
<td>.320** .002</td>
</tr>
<tr>
<td>Science Meaning Making Processes</td>
<td>.349** .000</td>
</tr>
</tbody>
</table>

Note. Mean classroom instruction values for each instructional unit are computed from lesson segment scores weighted by the duration of the lesson segment.

*p < 0.05. **p < 0.01.