PEPPERDINE UNIVERSITY

Pepperdine University
Pepperdine Digital Commons

Theses and Dissertations

2011

Interdisciplinary authentic assessment: cognitive expectations and student performance

Jenni M. Taylor

Follow this and additional works at: https://digitalcommons.pepperdine.edu/etd

Recommended Citation

Taylor, Jenni M., "Interdisciplinary authentic assessment: cognitive expectations and student performance" (2011). *Theses and Dissertations*. 211.
https://digitalcommons.pepperdine.edu/etd/211

This Dissertation is brought to you for free and open access by Pepperdine Digital Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Pepperdine Digital Commons. For more information, please contact Katrina.Gallardo@pepperdine.edu, anna.speth@pepperdine.edu, linhgavin.do@pepperdine.edu.

Pepperdine University

Graduate School of Education and Psychology

INTERDISCIPLINARY AUTHENTIC ASSESSMENT: COGNITIVE EXPECTATIONS AND STUDENT PERFORMANCE

A dissertation submitted in partial satisfaction

of the requirements for the degree of

Doctor of Education in Educational Leadership, Administration, and Policy

by

Jenni M. Taylor

October, 2011

Christopher Lund, Ed.D, Dissertation Chairperson

This dissertation, written by

Jenni Taylor

under the guidance of a Faculty Committee and approved by its members, has been submitted to and accepted by the Graduate Faculty in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

Doctoral Committee:

Christopher Lund, Ed.D. Chairperson

Doug Leigh, Ph.D.

Jennifer Rumack, Ed.D.

TABLE OF CONTENTS

LIST OF TABLES viii

LIST OF FIGURES ix

ACKNOWLEDGEMENTS x

VITA xi

ABSTRACT xii

Chapter 1: Foundations of the Study 1

Background 1
- Theoretical Background 1
- Contextual Background 7

Statement of Problem 10
Purpose of the Study 12
Research Questions 13
Operational Definitions of Variables 13
- Cognitive Level 13
- Student Performance 15

Conceptual Definitions of Key Terms 15
- Assessment 15
- Traditional Assessment 15
- Performance Task 15
- Authentic Assessment 16
- Interdisciplinary Teaching Unit 17
- Interdisciplinary Authentic Assessment 18
- Objective 19
- Assignment Prompt 19
- Rubric 19

Nature of Intervention 19
Importance of the Study 21
- Theoretical 21
- Practical 23

Limitations 24
- Age Group 24
- Measuring Student Learning 24
- Reliability of Evaluation Practices 25
- Generality of Trends 25
- Limited Scope 26
- Generalization and Reproducibility 27

Demographics 27
Assumptions 27
Analyzing Assessments 27
Measuring Student Learning 28

Chapter 2: Literature Review 30

Overview 30
Introduction 31
Historical Review of Standards-Based Testing 31
Historical Review of Interdisciplinary Teaching 34
Theoretical Review of Interdisciplinary Teaching 35
Benefits to Teachers 36
Benefits for Students 36
Affective Development 36
Perspectives 37
Student Learning 37
Challenges 38
Limitations of Existing Literature 40
Empirical Review of Interdisciplinary Teaching 41
Outcomes of Interdisciplinary Teaching 41
Impact on Teachers 41
Student Achievement 42
Historical Review of Alternative Assessment 42
Overview of Alternative Approaches to Assessment 44
Theoretical Overview: Need for Alternative Approaches 44
Empirical Review: Perspectives on Alternative Approaches 45
Teacher and Public Perspectives 45
Student Perspectives 46
Theoretical Overview: Impact of Alternative Approach on Students 47
Empirical Review: Impact of Alternative Approach on Students 48
Theoretical Review of Alternative Assessment and Authentic Assessment 50
Alternative Assessment 50
Authentic Assessment 50
Distinguishing Authentic Assessment 54
Challenges in Reviewing Literature 55
Outcomes of Authentic Assessment 56
Theoretical Overview of Outcomes 56
Empirical Review: Teacher Perspectives and Support 57
Theoretical Review: Motivation 59
Theoretical Review: Interest 62
Empirical Review: Student-Centered Approaches, Interest, and Motivation 63
Empirical Review: Assessment and Motivation 65
Empirical Review: Motivational Impact on Learning 66
Theoretical Overview: Achievement 68
Empirical Review: Achievement, as Measured by Test Scores 69
Empirical Review: Alternative Measures of Student Learning 70

Challenges of Authentic Assessment....73
Theoretical Review: Challenges of Authentic Assessment....73
Empirical Review: Challenges of Authentic Assessment....74
Fitting within a Standards-Based Paradigm....74
Disconnect Between Need and Practices....75
Time, Resources, and Classroom Management....77
Implementation and Consistency....79
Theoretical Review of Understanding....80
Transferability and Enduring Understanding....80
Cognitive Hierarchy of Understanding....81
Bloom's Taxonomy....81
The Revised Taxonomy Table....83
The New Taxonomy....84
Hierarchical Evaluation Methods....86
Empirical Review: Assessment and Cognitive Level....87
Need for Further Study....89
Summary....90

Chapter 3: Methods....91

Research Design and Rationale....91
Overall Research Design....91
Part One: Expected Cognitive Levels....92
Part Two: Performance Outcomes....95
Sampling Methods, Sample, Participants....96
Sampling Methods....96
Sample....98
Part One: Expected Cognitive Levels....98
Part Two: Performance Outcomes....99
Participants and Handling of Non-Response....99
Part One: Expected Cognitive Levels....99
Part Two: Performance Outcomes....100
Human Subjects....103
Informed Assent and Consent....103
Part One: Expected Cognitive Levels....103
Part Two: Performance Outcomes....104
Risk Minimization and Benefit Maximization....105
Confidentiality....106
Data Collection, Setting, Procedures....107
Instrumentation....107
Part One: Expected Cognitive Levels....107
Part Two: Performance Outcomes....109
Pilot Testing....109
Threats to Internal Validity....110
Analytic Techniques....111
Part One: Expected Cognitive Levels....111
Part Two: Performance Outcomes....112

Data Storage 113

Chapter 4: Results 114

Overview 114
Research Question One 114
Frequency of Expected Cognitive Levels 115
Expected Cognitive Levels for Each Assessment 118
Panel Ranking Agreement 123
Research Question Two 123
Overall Analysis of Each Assessment 124
Overall Analysis by Cognitive Level 126
Mean Scores 126
Analysis by Assessment 127
9th Grade Assessment 127
10th Grade Assessment 128
11th Grade Assessment 129
12th Grade Assessment 131
Summary of Results 133
Part One: Expected Cognitive Levels 133
Part Two: Performance Outcomes 133

Chapter 5: Conclusions and Recommendations 134

Overview 134
Conclusions, Discussions, Literature Contributions 134
Cognitive Complexity 134
Upper Cognitive Levels 135
Lower Cognitive Levels 136
Measurement and Grading 137
Varying Interpretations 138
Concept Integration 139
Assessment Design 141
Overall Student Performance 142
Higher Cognitive Levels 143
Recommendations for Policy and Practice 144
Practice 144
Value of Interdisciplinary Approach 144
Professional Collaboration 144
Assessment Design and Consistency 145
Policy 147
Recommendations for Further Study 148
Conclusion 149

REFERENCES 150

APPENDIX A Bloom's Taxonomy: Proving Behaviors 161

APPENDIX B: Criteria for Authentic Assessment Tasks 162

APPENDIX C: Professional Development Intervention Timeline 163

APPENDIX D: Planning Framework 164

APPENDIX E: Assessment Revision Protocol 165

APPENDIX F: Overview of Assessments for Each Grade 166

APPENDIX G: Ninth Grade Assessment and Rubric 167

APPENDIX H: Tenth Grade Assessment and Rubric 172

APPENDIX I: Eleventh Grade Assessment and Rubric 175

APPENDIX J: Twelfth Grade Assessment and Rubric 180

APPENDIX K: Informed Consent For Participation in Research: Teachers 182

APPENDIX L: Informed Consent Cover Letter: Parents 184

APPENDIX M: Assent Forms for Students 186

APPENDIX N: Informed Consent For Participation in Research: Parents 190

APPENDIX O: Informed Consent For Participation in Research: Panel 194

LIST OF TABLES

Table 1. Bloom's Taxonomy: Proving Behaviors 82

Table 2. The Taxonomy Table (Anderson & Krathwohl, 2001) 84

Table 3. The New Taxonomy (Marzano & Kendall, 2007) 85

Table 4. Population. Sample Size, and Percent Participation 101

Table 5. Demographics of Participants Versus Total Population 102

Table 6. Frequency and Percent of Objectives by Cognitive Level 116

Table 7. 9th Grade Assessment: Expected Cognitive Level of Each Objective 119

Table 8. 10th Grade Assessment: Expected Cognitive Level of Each Objective 120

Table 9. 11th Grade Assessment: Expected Cognitive Level of Each Objective 121

Table 10. 12th Grade Assessment: Expected Cognitive Level of Each Objective 122

Table 11. Three-Member Panel Ranking Agreement (External Panel of Evaluators) 123

Table 12. Mean Rubric Scores for Each Assessment by Cognitive Level 125

Table 13. Mean Overall Scores by Cognitive Level 126

Table 14. 9th Grade Assessment: Mean Rubric Scores for Each Objective 127

Table 15. 10th Grade Assessment: Mean Rubric Scores for Each Objective 128

Table 16. 11th Grade Assessment: Mean Rubric Scores for Each Objective 130

Table 17. 12th Grade Assessment: Mean Rubric Scores for Each Objective 131

LIST OF FIGURES

Figure 1. Histogram of Objectives by Cognitive Level.....115

ACKNOWLEDGEMENTS

The accomplishment of creating and completing this dissertation project would not have been feasible without the cherished encouragement and support from my family, friends, and colleagues. Specifically, I wish to thank my mother, Sharon Taylor, for always supporting me and helping me believe I can accomplish anything, and my father, William Taylor, for fostering my sustained endurance for achievement. I would also like to extend appreciation to my colleagues Amy Frame, Mandy Breuer, and Hannah MacClaren, for their continued inspiration and their contagious fascination with teaching and learning; to Alison Suffet-Diaz, for the boundless opportunities to lead creatively and passionately beyond traditional confines; and to Dr. Annette Marshall, for her moral support throughout the entirety of my graduate school journey. Finally, I would like to thank my dissertation committee for their contributions. My sincere appreciation is extended to Dr. Jennifer Rumack and Dr. Doug Leigh for their critical guidance and support, and especially to Dr. Christopher Lund for his patience, intellectual inspiration, and continued leadership.

VITAE

JENNI M. TAYLOR

ACADEMIC PREPARATION

2011 Pepperdine University, California
Ed.D. Educational Leadership, Administration, and Policy

2008 Pepperdine University, California
M.S. Educational Leadership and Administration

2000 University of California, Santa Cruz
B.S. Earth Sciences
B.A. Environmental Studies

RELATED CERTIFICATIONS

2008 California Clear Administrative Services Credential

2006 National Board Certification for Teachers: Science

2005 California Clear Single Subject Teaching Credential: Science

PROFESSIONAL EXPERIENCE

Environmental Charter High School, California
Principal (2009 to present)
Assistant Principal (2006 to 2009)
Math and Science Teacher (2002 to 2007)

ABSTRACT

This two-pronged quantitative, non-experimental design study, conducted at an urban secondary school of 472 students in Los Angeles, California, was designed to gain understanding of the potential impact of interdisciplinary authentic assessment and the manner and complexity with which such tasks push students to think. Since limited research has been conducted around the results of such practices at the secondary school level, this research serves as a pilot study to examine (a) cognitive levels of Bloom's Taxonomy present within four interdisciplinary authentic assessment tasks, following an ongoing professional development intervention and (b) student performance on these assessments of varying cognitive complexity.

Panel analysis of objectives from the assessments under study revealed that 94% of objectives measured student understanding beyond knowledge and comprehension levels of Bloom's Taxonomy. Sixty two percent of these objectives measured understanding within the top three cognitive levels (analysis, synthesis, and evaluation). Middle to upper taxonomy levels were identified most frequently, particularly the application, analysis, and synthesis levels of the taxonomy at 32%, 34%, and 22%, respectively. Student performance did not increase or decrease substantially with cognitive demand; instead, students on average performed near proficiency level (3.0, on 1.0 to 4.0 scaled rubrics) on each cognitive level, indicating that students may be able to meet challenges at varying levels of cognitive demand.

From this pilot study, interdisciplinary authentic assessment appears to be an appropriate and necessary challenge for secondary school curricula, particularly with increasing pressure for accountability around standardized test performance. Such assessments should be coupled with traditional assessments to develop multiple levels of

understanding. Since issues such as lack of reliability, inconsistency in assessment design and grading, and potential for grading bias remain important challenges with authentic assessment, and since there is little existing expertise in the area of interdisciplinary curriculum development, more collaboration, accessibility, and instruction around such methods in schools should be encouraged. Although challenges with interdisciplinary authentic curricula are many, schools should rethink approaches to assessment and may need policy incentives to do so. Education policy should not limit itself to a focus on traditional testing alone.

Chapter 1: Foundations of the Study

Background

Theoretical Background. Various authors (Pink, 2006; Reese, 2002; Resnick, 1999; Tchudi & Lafer, 1996; Wiggins & McTighe, 2005) have revealed that memorizing content standards and preparing for state tests alone are not meaningful enough for our students who will soon be facing real life challenges that require creativity, critical thinking and academic rigor. Wiggins and McTighe (2005) state quite clearly that, "what few educators seem to realize… is that drilling students for state tests is a *failing* strategy" (p. 43). Similarly, there are an alarming number of students that do not graduate from high school, particularly in states that focus on high stakes testing (Walden & Kritsonis, 2008). Only around 70% of students graduate from high school nationally, but the graduation rate is closer to 50% in many cities throughout the United States (Glass & Rose, 2008). There is reason to believe that students become more focused and interested in school when they experience a rigorous and relevant school program that is "tied to the world around them" and when they see that learning "means something" (Glass & Rose, 2008, p. 11).

Traditional tests generally only call for a narrow range of cognitive skills, which are often disconnected from real life experiences that students will have outside of the classroom setting (Resnick, 1987). According to Bloom (1956), who is highly referred to for his description of a cognitive taxonomy of learning objectives, "knowledge," defined as "little more than the remembering of the idea or phenomenon in a form very close to that in which it was originally encountered" (p. 28), is frequently the primary and sometimes only type of educational objective in curriculum. We need to demand

more rigor and critical thinking from our students to more adequately prepare them for life beyond the classroom, especially at the high school level. According to Wiggins and McTighe (2005), "The goal of schooling is fluent and effective performance in the world, not mere verbal or physical response to narrow prompts," (p. 78). The expectations that we demand in our school curricula should rise to match this goal. Curriculum design is an important tool for developing high levels of cognitive thinking, for making classroom learning relevant and meaningful, and for consequently motivating students.

Students need a thinking curriculum (Resnick, 1999) that expects students to ask questions, problem-solve, think, and reason (Principals of learning, n.d.). Students need a results-based curriculum with essential questions and assignments that not only capture big ideas from the standards, but that also provoke thought, reveal meaning, and encourage students to actively use their knowledge (Wiggins and McTighe, 2005). Wiggins and McTighe (2005) also explain that true understanding of concepts is really the ability to "transfer" knowledge and skills to new settings, which "involves the capacity to take what we know and use it creatively, flexibly, fluently, in different settings or problems, on our own" (p. 40). As it is, "students in general can do low-level tasks but are universally weak in higher-order work that requires transfer" (Wiggins & McTighe, 2005, p. 45). Bransford, Brown, and Cocking (2000) similarly describe transfer as "the ability to extend what has been learned in one context to new contexts" (p. 51). Bransford et al link quality of a learning experience to the degree of transfer. Transfer can happen between problems, between courses or disciplines, and from school

to home or career. Transfer, according to Bransford et al., is the difference between true understanding and memorization.

The notion of transfer is not unlike Bloom's (1956) idea of "application" that is typically referred to in curricular design.

> If the situations… are to involve application… then they must either be situations new to the student or situations containing new elements… Ideally we are seeking a problem which will test the extent to which an individual has learned to apply the abstraction in a practical way. (p. 125)

For learning experiences to develop a deeper understanding within students, assessments should involve transfer or application of knowledge and skills to other problems, other contexts or settings, or other disciplines. Tchudi and Lafer (1996) recall that "authentic assessment refuses to accept a distinction between assessment and learning itself" (p. 191). In most authentic learning situations, student learning occurs in conjunction with and as a result of the authentic assessment experience, which may imply that the cognitive demands of an assessment may have a direct relation to what students may learn through the process of engaging in an authentic assessment experience. This concept provides the underlying reason for engaging students in authentic assessment experiences.

Authentic assessment, then, provides a framework for engaging students beyond traditional classroom learning which tends to limit their experience to memorization and test-taking. Authentic assessment, sometimes called performance based assessment (Moon, Brighton, Callahan, & Robinson, 2005), involves learning experiences that require students to demonstrate understanding by engaging in real work tasks and

scenario-based problem solving (Moon et al., 2005; Darling-Hammond, 1997). Authentic assessment is similar to Reese's (2002) idea of contextual learning, which requires situated, social and distributed learning experiences to develop higher levels of cognitive understanding. "Situated" refers to the physical and social context for the assessment, "social" refers to interaction with other people, and "distributed" refers to the individual, other people, and symbolic and physical environments (p. 41). Reese (2002) reminds us that, "To keep students learning, we must draw from their interests and personal experiences and demonstrate the connections between what they need to learn and how that learning will be used in the real world" (p. 41). Reese's description touches the core essence of authentic assessment, which requires that students perform a high level task that demonstrates new skills and understandings in a meaningful and applied context. For example, instead of (or in addition to) taking a test on water quality in a science class or analyzing data provided in a textbook, students could collect water quality samples in a local waterway and draw conclusions about impacts of environmental conditions or practices, and use their results to propose management solutions to a local governing agency. Instead of (or in addition to) solving abstract math problems on surface area and volume from a geometry text, students could design two- and three-dimensional proposals for reconstructing a school garden to demonstrate understanding of concepts and skills related to these concepts.

Use of authentic assessment in the high school classroom may provide secondary students with the opportunities they need to practice application or transfer (Bloom, 1956; Bransford, Brown, & Cocking, 2000; Wiggins & McTighe, 2005) of material and to find meaning and relevance in the work they do for their courses. Authentic learning

assessments call for a deeper explanation and use of newly acquired understandings and skills within an authentic and *real* learning context. "Authentic assessments require students to be effective performers with acquired knowledge. Traditional tests tend to reveal only whether the student can recognize, recall or 'plug in' what was learned out of context," (Wiggins, 1990, p. 2).

More contextual or authentic use of newly acquired understanding also requires students to transfer knowledge beyond one discipline. Teacher collaboration and planning around multiple disciplines enables students to draw and apply natural connections in their learning, thus deepening the learning cycle. Interdisciplinary teaching, requiring collaborative planning around big ideas and meaningful essential questions, leads to a curriculum that promotes application of knowledge, higher-order thinking, and "enduring understandings" (Wiggins & McTighe, 2005, p. 128). Such interdisciplinary teaching encourages transfer or application of concepts across disciplines and can therefore enable a more comprehensive understanding of themes and essential questions. Due to the real life nature of cross-disciplinary thinking, interdisciplinary teaching tends to result in culminating authentic assessments which can be more meaningful, integrated, and applied for students and which allow students to practice not only transfer of skills and conceptual understandings across disciplines, but also transfer within new contexts and situations.

To exemplify such interdisciplinary learning, consider this prompt from a six grade teaching unit which integrates social science, math, science, and language arts standards:

> Egypt is in the middle of a four-year drought. You are a Minister of the Waters for your region. You have been asked to conduct some tests on water from farmlands to find out about the purity of your water. Based on what you have discovered in the fields, it is your job to influence the pharaoh to make wise decisions for the kingdom. The pharaoh has asked you, "Which agricultural practices should be used, changed, or stopped in your region of the Nile?" To persuade the pharaoh to follow your proposal, you must use relevant evidence based on your expertise as a specialist Minister of the Waters (Frame, Evola, Arezina, 2010, p. 1).

This interdisciplinary unit seeks to teach students about principles of geography, ancient world history, scientific investigation, data analysis, and persuasive speech. Due to the inherent complexity of linking such widespread standards, the integration of concepts in this unit is facilitated more meaningfully in a project where students have to collect data, analyze findings, and apply findings within historical and geographical parameters using persuasion. This complex task is something one might encounter in the real world or within a job, where concepts and skills tend to be less isolated by discipline.

Authentic learning, resulting from interdisciplinary teaching and learning tasks where students must apply conceptual knowledge and skills across physical, social, and disciplinary contexts, should be the goal of curricular programs, particularly at the high school level. If this notion is to be widely accepted in the field of education, the outcomes of such teaching and learning practices must become a greater focus in research and measurement at school sites, and frameworks for developing and

implementing such practices should be more available in current literature. Additionally, more research should be conducted to highlight the high expectations and cognitive demands which are thought to result from such practices, along with analysis of how students perform on various cognitive functions related to authentic assessment tasks. This study will focus on one particular interdisciplinary teaching term, culminating in various authentic assessment tasks, with the intent of developing a deeper understanding of the implementation process, the cognitive demand of the interdisciplinary authentic assessments, and related student performance.

Contextual background. The school in this study is an independent charter school located in Los Angeles County, California. The school hosts approximately 470 9th through 12th grade students, 30 teachers and 10 other full-time staff members. The school was opened in the fall of 2001 and has undergone many programmatic changes since its inception. The mission of the school, in short, is to prepare students for success in college and to help students become active stewards of their community.

According to its charter petition, the school implements four best practices in order to realize its mission, including: (a) a small learning community, (b) a rigorous academic curriculum, (c) learning beyond the classroom walls, and (d) community partnerships. For teachers, a small learning community implies regular teacher collaboration on curriculum design, instructional practices, and student learning. Teachers implement a college preparatory curriculum and follow a school-wide grading policy which puts student growth and achievement at the forefront of instructional efforts. *Community partnerships* and *learning beyond the classroom walls* are used as strategies to make the college-preparatory curriculum more meaningful and relevant.

Community partners from local businesses and non-profit organizations often come to the school to participate in curricular experiences for students, in order to provide an opportunity for students to contextualize their learning and understand curricular content as it applies in the real world outside of the classroom. Similarly, field trips are a regular part of the curriculum, since they are also used as a tool to extend classroom learning, allowing students to experience what they are learning first-hand and to connect course content to what surrounds them outside of the classroom. The supplemental implementation of a *rigorous academic curriculum* serves to hold students to high expectations and to prepare them for college, while also allowing them to develop a keener meaning of and context for their learning through partnerships and experimental learning.

Throughout the 10 year history of this school, much time and commitment has been lent to building a professional learning community at the school which focuses on collaboration, a rigorous authentic curriculum, and the growth and achievement of the school in terms of its mission and vision. Teachers are asked to collaborate frequently, analyze student achievement data, and participate in many school-wide decisions that help shape the school. Instructionally, teachers are expected to create authentic assessments for their courses and to create interdisciplinary units with their grade level teams that result in authentic performance tasks, while still adhering to state standards and expected school-wide learning results (ESLRs). Teachers are also encouraged to use project-based learning in the classroom, to utilize the local environment as a learning context for students, and to seek service learning opportunities for their students and classrooms.

The professional development program has focused on, among other themes, allowing collaborative time for subject department planning and for interdisciplinary team planning around essential questions. Teachers meet regularly in grade level teams to create interdisciplinary instructional connections across grade levels, and teams create a month-long interdisciplinary program that results in authentic learning projects for the community at each grade level, based around subject matter standards and essential questions. Several half-day or full day professional development workshops have been dedicated for training teachers to understand the importance and technique of asking thought-provoking essential questions to encourage deeper thought and contextualization of ideas. Additionally, teachers work on summative projects for each course which serve as more authentic measurements of student learning than standardized tests, and which allow students to explore concepts, skills and attitudes, instead of only demonstrating learning through rote memorization.

Although authentic assessment and interdisciplinary teaching and learning appear to have a positive impact on the overall school culture and on the skills, attitudes, understandings, and motivation of this school's students, there is no clear analysis of the overall impact of such assessment practices on student learning. Teachers and administrators report that students complete their highest quality, most thoughtful and cognitively complex work during the month-long interdisciplinary teaching units and that this interdisciplinary assessment period, which culminates in authentic assessment tasks, is one of the most fulfilling teaching periods for teachers. However, only anecdotal data exists to support the notion that authentic projects and interdisciplinary

teaching at this school result in more complex thinking or high achievement of rigorous learning outcomes.

Furthermore, this school program appears to be relatively unique in its logistical and conceptual implementation of interdisciplinary teaching at the secondary school level. Not only does the school re-schedule students for a special term during the middle of the year so they can receive integrated instruction from the same teachers on each grade level, but these units also require collaborative planning and teaching from this same team of six teachers who teach courses of varying disciplines from math, science, social science, language arts, to varying elective courses. Furthermore, each interdisciplinary team teaching unit results in an authentic assessment task which is presented to the local community. This school does not require a prescribed rubric format for its assessment, leaving assessment and rubric creation up to the discretion of these six-member teaching teams, creating an interesting point of analysis. As a unique program, little research has been conducted around the results of such efforts; this research serves as a pilot study of the cognitive demand and performance outcomes resulting from such interdisciplinary programming.

Statement of Problem

Although interdisciplinary teaching and authentic assessment have gained more widespread recognition as educators have begun to study the value and effectiveness of teaching and learning within small professional learning communities, these practices together are not usually implemented systematically in school programs and are likewise under-researched. Since interdisciplinary learning and authentic assessment together can pose the thinking curriculum which Resnick (1999) proposes and can provide students

with opportunities to transfer new ideas (Bransford, et al., 2000; Wiggins & McTighe, 2005) in order to reach deeper stages of learning and understanding, programmatic interdisciplinary methods should be a goal for schools.

The underlying premise of this study is that if we as educators create engaging, rigorous assessment tasks for our students, they will reach the cognitive demands and, in the end, become more cognitively prepared for the world beyond testing which they will soon face. The cognitive demands of interdisciplinary thinking and authentic assessment appear to be more valuable than standardized tests; yet, there are many complexities of creating and grading these assessments which prevent us from easily measuring such value in meaningful ways. Instead, we often rely on traditional, lower-order cognitive assessments to measure student learning, rather than measuring the real cognitive complexities of authentic assessment tasks (Bloom, 1956; McNamar, 2009; Wood & Sellers, 1997). Students are often not graded with the same complexity as the tasks they perform and educators don't necessarily take the time to analyze the cognitive complexity of these assessments they implement. Furthermore, anecdotal data and perspectives in support of such programming and assessment design are often what are called on to justify the value of such learning opportunities.

Overall, the implementation and cognitive outcomes of interdisciplinary teaching and authentic assessment are underexplored and limited in scope and measurement, even in schools where such practices are valued and expected. In this day of data and measurement, teaching and learning practices should have clear measurement of cognitive demands and impact on student learning. Since literature regarding assessment and learning generally suggests that authentic assessment and interdisciplinary teaching

can motivate students and facilitate higher levels of learning (Bransford et al., 2000; Drake & Burns, 2004; Erickson, 2002; Tchudi & Lafer, 1996; Wiggins & McTighe, 2005), yet there are concerns regarding articulation, implementation, grading, and underexplored outcomes (Avery, Carmichael-Tanaka, Kunze, & Kouneski, 2000; Suurtamm, 2004; Tanner, 2001), it is important to further research on interdisciplinary teaching and authentic assessment practices in order to understand the complex cognitive demands and how students perform on such expectations. Since little empirical research exists in this specific area, this research will contribute to the field as a pilot study of the cognitive complexity of interdisciplinary authentic assessment and the ability of students to meet the demand from such practices.

Purpose of the Study

The study will examine the cognitive demands within interdisciplinary authentic assessments, following an ongoing professional development intervention, to better understand cognitive expectations and complexities of such assignments. The study will also examine how students perform on tasks of varying cognitive complexity, to provide insight into how well students are able to meet curricular demands. Specifically this quantitative study will determine:

1. The cognitive levels of Bloom's Taxonomy present in interdisciplinary authentic assessment tasks, following an ongoing professional development intervention.
2. How students perform on each cognitive level of Bloom's Taxonomy within interdisciplinary authentic assessment tasks.

Research Questions

In order to more deeply understand the cognitive expectation and outcomes of interdisciplinary learning and authentic assessments, four interdisciplinary authentic assessment prompts and related rubrics were analyzed, along with student rubric scores for these assessments, to determine the cognitive expectations for student performance from these assessments and related student performance. The following research questions guided the course of study:

1. Which cognitive levels of Bloom's Taxonomy are present in interdisciplinary authentic assessment tasks, following an ongoing professional development intervention?
2. How do students perform on each cognitive level of Bloom's Taxonomy within interdisciplinary authentic assessment tasks?

Operational Definitions of Variables

Cognitive level. Bloom's Taxonomy (1956) has been perhaps the most universally referenced literature throughout the last decade for classifying educational objectives. This taxonomy describes varying levels of cognitive performance expectations for students, including the following: knowledge, comprehension, application, analysis, synthesis, and evaluation. The taxonomy is arranged in a hierarchy, implying that each new level requires skills and abilities from the level below it. The lowest level of Bloom's Taxonomy is *knowledge*, which implies a focus on the recognition or recall of material in a form similar to the manner in which the information was learned. *Comprehension* implies an understanding of material in a way that students are able to make use of the material and communicate knowledge in a form other than its

original expression. Where comprehension implies demonstration of understanding beyond the originally presented form of the material, *application* implies correct use and function of those skills and abilities in a different context in which "no mode of solution is specified" (Bloom, 1956, p. 120). *Analysis* implies a breakdown of material, the detection of organization and relationships, or an understanding of the techniques for deriving meaning or conclusions. *Synthesis* is a relatively more creative cognitive level, which is described as the ability to recombine or integrate elements or parts of a whole to show new patterns, structure, or meaning. The highest level of the cognitive domain, as described by Bloom is *evaluation*, which implies a qualitative or quantitative judgment regarding the value of the learning material using self-determined or prescribed criteria (Bloom, 1956).

Recent revisions to the taxonomy have been proposed to better represent the deeper interrelated complexities of both cognitive and knowledge domains (Anderson & Krathwohl, 2001; Marzano & Kendall, 2007). This study, however, continues to employ the original version of Bloom's taxonomy, to ensure common understanding and simplicity of use among involved researchers. For this study, the Proving Behaviors for Bloom's Taxonomy (see Appendix A) will be used by an external panel of three independent assessment experts, with interest and expertise in authentic and/or interdisciplinary assessment, to classify objectives and assignment prompts into these hierarchical cognitive levels of Bloom's Taxonomy: knowledge, comprehension, application, analysis, synthesis, and evaluation (Proving Behaviors in Bloom's Taxonomy, n.d.).

Student performance. For this study, student performance on the selected interdisciplinary summative assessments will be measured according to teacher-created rubrics, based on specific and measureable learning objectives from various disciplines. The learning objectives used on each assessment rubric are explicitly linked to specific criteria from various disciplines with the interdisciplinary unit of study. Average student rubric scores on each objective, for each unit of study, will be used to measure student performance. Student performance on these rubrics will be graded by teams of two or more teachers, who are the students' own teachers (not independent graders).

Conceptual Definitions of Key Terms

Assessment. According to Wiggins and McTighe (2005) assessments are "techniques used to analyze student accomplishment against specific goals and criteria" (p. 337). Assessment demonstrates student learning and understanding of specific knowledge, concepts, skills, attitudes or other criteria. Assessment can include paper and pencil tests, essays, oral presentations, visual or artistic displays, or other demonstration of accomplishment measured against specific learning criteria.

Traditional assessment. Traditional assessment measures knowledge, concepts or skills using specific criteria, without necessarily requiring students to apply their learning to a performance task or applied situation. Traditional assessment often comes "in the form of pencil-and-paper multiple-choice tests" (Moon et al., 2005), prompting students to memorize and recall or recognize factual information.

Performance task. A performance task is a formal or informal, non-traditional assessment of a student's knowledge, skills, or abilities. Wiggins and McTighe (2005) define a performance task as "a task that uses one's knowledge to effectively act or bring

to fruition a complex product that reveals ones knowledge and expertise" (p. 346). Performance tasks require more than recalling or recognizing factual information; these forms of assessment require students to do something with this information. Performance tasks can be "authentic," but not all performance tasks are authentic, as they do not always require transfer of understanding to new situations, contexts, or disciplines. Similarly, a performance task can be interdisciplinary, but not all performance tasks require interdisciplinary integration.

Authentic assessment. A combination of parameters for alternative assessment and/or authentic assessment, as defined by Abadiano and Turner (2003), Moon, et al (2005), Tchudi and Lafer (1996), and Wiggins (1989), will be used as to define authentic assessment for this study. Based on their insight, authentic assessment as discussed in this study includes the following parameters (see also Appendix B for criteria):

- Has value or meaning beyond the classroom
- Focuses on big ideas or concepts of a discipline (breadth versus depth)
- Involves research or active use of conceptual knowledge
- Pushes students towards more advanced use of skills and conceptual knowledge
- Counts what students "do"
- Presents transparent performance criteria; scores according to this criteria

Similarly, the following characteristics are not necessarily essential characteristics for assessment, but do demonstrate a stronger level authenticity of an assessment. Therefore, assessments used in this study will also be held to the following criteria:

- Public; requires an audience

- Requires collaboration (with peers, professionals, or community members)
- Allows for student choice
- Allows for consideration of student learning styles, aptitudes and interest

Lastly, an important criterion for authentic assessment is cognitive complexity. However, instead of assuming cognitive complexity, this study will measure cognitive levels as an operational variable, in attempt to justify the value of authentic assessment.

Interdisciplinary teaching unit. Interdisciplinary teaching, according to Rowntree (1982), is an approach in which "two or more disciplines are brought together, preferably in such a way that the disciplines interact with one another and have some effect on one another's perspectives" (p. 135). Teacher collaboration and planning around multiple disciplines enables students to draw and apply natural connections in their learning, making the experience more meaningful and allowing students to reach new levels of thinking and discourse (Klein, 1990). This collaboration should also lead to the transferability which Wiggins and McTighe (2005) discuss.

In this study, an interdisciplinary unit requires collaborative planning around big ideas and essential questions (Wiggins & McTighe, 2005) derived from the overlap between four to six different disciplines and resulting in a culminating performance task. According to Erickson (2002) an interdisciplinary teaching unit has the following characteristics: (a) "there is a conceptual lens that forces thinking above the fact base," (b) "the topic becomes a tool for understanding conceptual ideas that transfer across time and cultures," and (c) each discipline in the web has depth and integrity as a study on its own" (p. 65). An interdisciplinary teaching unit often (but not always) culminates in a performance task which is an authentic assessment. In this study, each

interdisciplinary unit was created collaboratively by a team of six teachers and was taught through six different classes by these same six teachers, individually, during six different rotating periods throughout the course of one month. For each interdisciplinary teaching unit, all students took the same six classes from the six teachers who planned and graded the culminating performance task together.

Interdisciplinary authentic assessment. An interdisciplinary authentic assessment is an authentic assessment, as defined above, which is the culminating assessment of an interdisciplinary teaching unit (also defined above), integrating at least four disciplines. Specifically, an interdisciplinary authentic assessment:

- Has value or meaning beyond the classroom
- Focuses on big ideas or concepts of a discipline (breadth versus depth)
- Involves research or active use of conceptual knowledge
- Pushes students towards more advanced use of skills and conceptual knowledge
- Counts what students "do"
- Presents transparent performance criteria; scores according to this criteria
- Public; requires an audience
- Requires collaboration (with peers, professionals, or community members)
- Allows for student choice
- Allows for consideration of student learning styles, aptitudes and interest
- Requires integration of knowledge and/or skills from four or more disciplines

Again, it is assumed that interdisciplinary authentic assessments are cognitively complex, but this study will actually investigate the cognitive levels within the assessments being analyzed.

Objective. An educational objective is a specific goal for student performance. This study will employ the definition for objective described by Roberts and Kellough (2008): "the actual performance that students are expected to display" (p. 81). According to Roberts and Kellough, goals are more generally stated, whereas objectives indicate what students will specifically do.

Assignment prompt. An assignment prompt is a description of the assessment the students are asked to complete. The assignment prompt in this study primarily includes an overall description of the interdisciplinary authentic assessment task, but could also include supporting documents which guide the students in completing their assessment.

Rubric. A rubric is a teacher-developed tool which is used to assess the degree to which students reach expected objectives for an assessment. Roberts and Kellough (2008) describe a rubric as, "a form or outline with headings of desired performance and general degrees of satisfactory completion" (p. 124). In this study, rubrics include rows of various disciplinary objectives and columns with headers indicating varying levels of performance.

Nature of Intervention

Since a review of literature on interdisciplinary teaching suggests challenges such as time, logistics, intellectual and collaborative demands, and availability of a planning framework (Letterman & Dugan, 2004; McGehee, 2001) and since design of assessments for this study follow an ongoing professional development process, it is important for this study to address the collaborative processes which lead to the creation of the interdisciplinary authentic assessments under scrutiny.

The intervention prescribed in this study is an ongoing professional development sequence (see Appendix C) led by school leaders with the primary purpose of allowing the teachers to work collaboratively on the development of a rigorous interdisciplinary teaching curriculum and resulting authentic assessment. Following the professional development sequence which focuses on the development of an interdisciplinary teaching curriculum by a team of six teachers in each grade level, students are rescheduled for a special month-long term during the middle of year so that classes of students rotate through different periods of these same six teachers throughout the day, so that each class of students in a grade level receives the same classes and same general instruction.

Since professional development on interdisciplinary unit planning has been cyclical and ongoing at this school site for the last nine years, many teachers have participated in this process in different forms and over a multi-year period. Not only is the month-long term impacted by intervention; curriculum throughout the school year is impacted by this ongoing work, since many teachers collaboratively begin building skills and concepts for the assessments at various points throughout the term just before the interdisciplinary unit.

Within the professional development sequence over the last year, various workshop sessions sought to build common knowledge around rubric development, understanding and use of pedagogical terminology, theory on cognitive scaling of performance objectives, and approach to interdisciplinary planning of curriculum using big ideas and essential questions. Appendix C provides a simplified schematic overview of the most recent ongoing intervention and related timeline. Most teachers within the

six-person curriculum development and teaching teams also worked collaboratively beyond the allotted professional develop sessions to share curriculum, plan assessments and rubrics, and coordinate instruction.

McGehee (2001) suggests a framework for interdisciplinary planning and instruction be provided for teacher collaboration. The framework for interdisciplinary collaboration within the professional development sequence used for this study (see Appendix D) was based on a collective of ideas from Wiggins and McTighe (2005) and Erickson (2002) and developed by an external organization (Real Curriculum, 2006) and local school leaders. This guiding framework was used by each team to plan their integrated units and was referred to and utilized throughout the planning sequence. An Assessment Revision Protocol (see Appendix E) was also employed as a collaborative tool for teams to give one another feedback regarding their interdisciplinary culminating assessment plans prior to final revisions and implementation. The revision and feedback process is designed to increase rigor, clarity, alignment, and validity of each assessment.

Importance of the Study

This study will serve to expand current research regarding cognitive expectations and outcomes of interdisciplinary authentic assessments, in order to add empirical understanding to the theoretical research which suggests that such tasks push students to reach higher cognitive levels than traditional assessments (Bransford et al., 2000; Drake & Burns, 2004; Erickson, 2002; Tchudi & Lafer, 1996; Wiggins & McTighe, 2005).

Theoretical. Many theorists over the past several decades have pointed towards the need for transfer or application of knowledge for the development and demonstration of deeper understanding (Anderson & Krathwohl, 2001; Bloom, 1956; Bransford et al.,

2000; Marzano & Kendall, 2007; Tchudi & Lafer, 1996; Wiggins & McTighe, 2005). Similarly, many theorists continue to point towards contextual learning that requires meaningful and relevant application of learning in order to deepen understanding and prepare students for real world tasks (Erickson, 2002; Klein, 1990; Moon et al., 2002; Reese, 2002; Wiggins, 1989; Wiggins, 1990). Such learning theory suggests interdisciplinary teaching and authentic assessment as valuable and necessary learning experiences for students, specifically for high school students who may benefit from drawing connections and finding relevancy between what they do in school and what they are expected to do in life (Glass & Rose, 2008).

Interdisciplinary teaching is challenging, and often near impossible, in a traditional school setting. Logistics and structures within comprehensive high schools do not lend themselves to opportunities or interest for interdisciplinary planning among teaching staffs. As a greater number of small, more flexible learning environments emerge with recent small school and school-within-a-school trends of the new millennium, increasing opportunities for interdisciplinary instruction are becoming more prominent in secondary schools. Since there has been little opportunity in most public schools throughout the last century for such practices, there is little research supporting outcomes and challenges of interdisciplinary practices beyond the efforts of two teachers. This study will provide an important framework, model, and analysis of the development and outcomes of interdisciplinary practices within a highly collaborative secondary school setting.

Additionally, although interdisciplinary teaching and authentic assessment can, in theory, be effective, it is often difficult to measure and verify the effectiveness of

related authentic learning assessments. Overall, there is a lack of research measuring specific learning outcomes that occur as a result of interdisciplinary authentic assessments and, although higher-order thinking is a supposed result of such practices, there is little to no literature which analyzes expected cognitive behaviors and related student performance on such behaviors. A more in-depth analysis of such factors related to interdisciplinary authentic assessment tasks will enable educators to better understand the shortcomings and needs of designing authentic learning curriculum and monitoring student performance. This study will serve to measure and document potential higher-order expectations for students and the learning that results from public interdisciplinary authentic assessment experiences, following the implementation of an interdisciplinary teaching unit.

Practical. On a more practical level, this study will serve to document the developing impact of a localized school assessment program which prides itself on the use of interdisciplinary teaching and authentic assessment. The culture at this high school is one which values the use of authentic assessment in the classroom but has not historically defined specific parameters or grading criteria for assessing authentic learning. The nontraditional curricular program which students at this school experience, including the interdisciplinary units, appears to have strong motivational, affective, skill- and conceptual-based impacts on student learning, but only general assumptions, informal qualitative analysis, or anecdotal data have hitherto been used to gauge the specific cognitive expectations or outcomes of the different learning and assessment tools that are used throughout the program. More research needs to be done to measure and document the complexities and student performance outcomes of innovative

interdisciplinary and authentic assessment approaches that are being used in this high school program, in order to more fully understand the processes, outcomes, and challenges of using such strategies. This pilot study can provide insight and recommendations for other teachers and schools implementing authentic interdisciplinary assessments.

Limitations

Although this study seeks to exemplify and generalize results of an interdisciplinary authentic assessment period at one school, it is only a start to the amount of research which should be conducted to fully explore the outcomes of such assessment practices. This study serves as a pilot study of this type of curricular program, realizing that further study regarding cognitive outcomes is inherently necessary. There are important limitations to this particular study, which should be followed in the future by the exploration of other specific and related research questions.

Age group. This study is limited to the experience of high school students. Although similar trends may be evident for different age groups, this study will not examine those trends. It would be beneficial and informative to compare the impacts of authentic assessment on various age groups, in order to understand any trends or differences across such groups; however, such a focus is outside the scope of this specific study.

Measurement of student learning. Another limitation of this study is the subjectivity related to designing assessments and scoring guides, and assigning rubric scores for student performance on learning objectives. To enhance reliability, the assessments under study were selected partially because two or more teachers

participated in the scoring of each student's performance on selected assessment tasks. However, there is still an inherent level of subjectivity associated with designing rubrics and scoring student performance on rubrics. Additionally, inter-rater reliability of rubric scoring will not be measured in this study. This research has been developed around existing frameworks of planning, grading, methodology, and time already provided at the school site, which leaves it difficult to also measure grading inter-rater reliability. Although questions regarding inter-rater reliability of scoring are important and should continue to be explored in future research, such questions are beyond the scope of study of this project.

Reliability of evaluation practices. This study assumes a level of reliability in assessment evaluation and grading practices. Although the research methods prescribed will seek to limit skewed grading practices by the use of team grading of the assessments, there will not be specific analysis as to the inter-reliability of grading on these assessments. The variation between rubrics in this study and evaluations of those rubrics may impact the internal validity of this study. Teachers across grade levels may use different methods to design assessments and scoring guides and to evaluate student performance on rubrics. Additionally, there may be bias present in the assignment of grades by teachers for their own assessments. This potential arbitrary nature or lack of reliability of the evaluation of student performance is one of the most important limitations of this study.

Generality of trends. This study is limited to general trends. There will not be a control group and experimental group of students who are learning the same content within the same time frame, but through different forms of assessment. The study is not

designed to compare such groups, but to look for general trends within specific assessments, rubrics, and student performance scores. Comparing control and experimental groups would enable us to more accurately understand the specific impacts of learning and assessment strategies and should continue to be investigated in the future, where possible or prudent. In this study, comparison of control and experimental groups would not necessarily be ethical, since it is the belief of the school and the researcher that the value of this authentic assessment period is something that every student should experience and since it would be difficult to control other extraneous variables if comparing students at this school to a similar school with more traditional teaching and assessment practices. A form of pre- and post- survey analysis, however, may be another interesting aspect for future study.

Limited scope. The assessments represented in this study are limited to localized, teacher-created interdisciplinary assessments. Although there are other innovative forms of authentic assessment which may be used to develop rigorous student learning, this study only focuses on interdisciplinary, teacher-created assessments. Computer-based simulations and assessments, for example, often serve as an innovative form of authentic learning or authentic assessment, but such approaches are more specific in focus and are beyond the scope of this particular study. Other forms of authentic learning may also be present in student experiences that extend beyond the regular classroom experience. Internships, for example, are likely an excellent opportunity for students to develop complex cognitive understandings through direct, authentic experiences. This study, however, focuses solely on interdisciplinary authentic

classroom assessments as a supplementary or alternative form of school-wide traditional learning and assessment practices.

Generalization and reproducibility. The assessments analyzed in this study are not necessarily similar in content or rigor and the results may not necessarily be generalized for all authentic assessment tasks since all tasks will be very different in nature and form. Four different assessments were selected to ensure that an analysis of various types of assessment were explored in this study, but the conclusions drawn from analysis of these particular assessments may vary for other grade levels, schools, teachers, students, or learning contexts.

Additionally, interdisciplinary authentic assessment practices, as defined and analyzed in this study, are fairly unique to this particular school since the structure and design of traditional secondary schooling makes it logistically difficult to impossible to implement such practices in most school settings as they exist. Therefore, this study is limited in its reproducibility; the results only represent one school, one method of implementation, and specific grade levels, subjects, classes, or students.

Demographics. This study is also limited in that it will not analyze the student sample populations with regard to gender, ethnicity, age, or the level or subject of coursework being taught.

Assumptions

Analyzing assessments. It is assumed that the material covered in the different types of assessments is similar in scope and difficulty. Similarly, it is assumed that the assessment is the primary influence on student learning and that individual teachers or teaching teams or teaching style does not have a significant overall effect on student

performance. It is also assumed that the assessments were actually implemented in the manner described on the assessment prompts and rubrics provided. Although it will be difficult to ascertain that all of these assumptions do not interfere with study results, collection of data on four different assessments may allow for a more reliable study.

Bloom's Taxonomy (1956) assumes a hierarchy of learning objectives with each new level requiring prerequisite "skills and abilities which are lower in the classification order," (Bloom, 1956, p. 120). In this study, it will be difficult to know or prove, other than through suggestion from the teacher panel member, that lower cognitive levels have already been expected and tested before the culminating performance task under study. A teacher may, for example, have already expected students to memorize information and explain concepts in their own words, before being asked to apply this learning to their interdisciplinary culminating performance task. Since analysis of all related assessments prior to the culminating performance task is outside the scope of this study, it is assumed that if students are asked to perform a task at a particular cognitive level, that the prerequisite knowledge, skills, and abilities from lower cognitive levels have already been assessed or met.

Measuring student learning. It is assumed that student performance on the authentic assessment tasks measured in this study are indicative of actual student learning as a result of this unit. This assumes a few things: (a) that student learning as measured by rubric scores are actually outcomes of the prescribed unit of study, (b) that students perform to their ability on these assessments, (c) that any lack of performance on these assessments is actually due to a lack of skills, knowledge, or abilities related to the specified objectives, as opposed to any other disruption, including teaching style,

that may prevent students from performing on these tasks, and (d) that teachers have graded the assessments objectively and that the rubric scores reflect an accurate approximation of student understanding. The study focuses only on the summative assessment experience, without regard to formative instruction or assessment which occurred up to this point (traditional or non-traditional) and without regard to the manner in which instruction was delivered to the students. The study assumes that the interdisciplinary authentic assessment experience is what is impacting student learning.

It is also assumed that the teachers judge students objectively on the grading criteria using the rubrics, and that this judgment is not swayed by external judgment factors. It is assumed that all teachers understand the rubric descriptors of student performance and that there is consistency among teacher teams in the manner and understanding of terminology related to the learning objectives and performance descriptors. Since it is likely not the case that all teachers understand and grade students the same way on given rubric descriptors, assessments were chosen for analysis if they were graded by two or more teachers, to ensure there was as much consistency and inter-rater reliability as possible. Despite these efforts, however, assumptions will still be made as to the reliability of the rubrics and of teachers' scoring practices.

Chapter 2: Review of Literature

Overview

Memorizing content standards and preparing for state tests within isolated disciplines are not alone meaningful enough for our students who will soon be facing real life challenges that require creativity and critical thinking. Curriculum design is an important tool for making classroom learning relevant and meaningful, motivating students, and developing high levels of cognitive thinking.

To develop a clearer understanding of what is known about interdisciplinary teaching, authentic assessment, and student understanding, this review of literature will begin by discussing historical trends regarding standards-based testing and accountability. A historical and theoretical review of the benefits and challenges of interdisciplinary teaching will be presented, followed by an empirical review of outcomes, including affective development, enhanced perspectives, and heightened learning. A historical review of the shift towards student-centered learning and alternative assessment practices will also be discussed, followed by a theoretical and empirical review of outcomes. Theoretical justification of authentic assessment as a means for inspiring interest and motivation and for developing higher-order thinking skills for students will be discussed, along with existing challenges of such practices. Theoretical and empirical outcomes of authentic assessment will be reviewed, including student-centered approaches, increased engagement and motivation, and higher achievement.

Lastly, the idea of understanding (in its different forms) will be reviewed, including various taxonomies used to measure levels of cognition and levels. Theoretical

and empirical review of literature will highlight the need for a focus on assessment related to the development of higher-order thinking skills. In summary, the need for further study of the impact of interdisciplinary authentic assessment on various levels of cognitive understanding will be suggested.

Introduction

Education today is packaged with many challenges and multiple layers of reform efforts. Some of the most important challenges facing educators in the United States are the high number of students that do not complete high school and the poor comparative results of American students to other countries on math and science tests. Student and school performance results and expectations vary widely across schools, districts and states. As a result, national reform efforts have focused on educational standards and accountability, with an increased push over the last decade to move towards national education standards for the United States. More recent reform efforts, however, have indicated that standardized tests are not what will push our students to graduate and achieve. Recent efforts have been focused on alternative strategies, including an authentic assessment approach, to increase engagement, motivation and aptitude beyond such tests. This review of literature will explore the history, theory, challenges and outcomes of recent reform efforts, and will frame the need for more study with regards to the impact of authentic assessment and interdisciplinary teaching practices on various cognitive levels of student learning.

Historical Review of Standards-Based Testing

Evaluation of student learning began as a theoretical and practical approach to education in the 1800s (Rotham, 1995), with large-scale achievement tests dating back

to the 1840s (Koretz, 2008). Many large-scale tests were designed initially for diagnostic purposes, not to evaluate performance of classrooms and schools (Koretz). Yet standardized tests were introduced from 1880 to 1920 as a means for pushing schools to justify quantitative performance for concerned taxpayers (Dutt-Doner & Maddox, 1998). Theorist Horace Mann promoted the notion of using testing to measure and monitor the quality of instruction in classrooms and schools, allowing educators to gauge performance among teachers and schools (Rotham, 1995). This led to the use of testing as common practice throughout American schools over the last two centuries.

Standardized tests emphasizing multiple choice questions and short-answer prompts were a cheap and efficient way to objectively measure student performance in and develop public confidence in our schools (Resnick & Resnick, 1992). Grading was reliable and implementation was efficient, saving time and money. However, testing back to the 1950s and before did not have the high-stakes consequences that it does today (Koretz, 2008). Since the original development of the standardized testing movement, other legislation and reform efforts have been created to support increased pressure for accountability and standardized teaching and testing. In 1965, the Elementary and Secondary Education Act (ESEA) was established in the United States, to improve performance of students in low income schools (Koretz, 2008). Title 1 compensatory funding was established and the Title 1 Evaluation and Reporting System (TIERS), requiring reporting of norm-referenced standardized test scores, became a requirement for funding (Koretz, 2008). In the 1960s, the National Assessment of Educational Progress (NAEP) was also established as a means of measuring national progress and collecting information on the progress of different regions and subgroups;

the NAEP, however, was deliberately not intended to compare schools and districts, making it impossible to use for accountability purposes. Although neither of these initiatives bore consequences on students or teachers, they represented a change in testing purposes from diagnostic to large-scale monitoring of achievement, and eventually to standards-based accountability (Koretz, 2008).

But by the 1970s, 60% of states required mandatory testing in schools and the shift to *measurement-driven instruction* had begun to take place, signifying the change in the purpose of testing to holding students, teachers and schools accountable for test scores (Koretz, 2008). *A Nation at Risk* (United States Department of Education, 1983) was published in the 1980s, serving as a call to attention to the poor performance of American schools. This publication encouraged the *education reform movement*, noted by an increase in pressure for performance on standardized tests and increased sanctions and policies to push educators and schools to enhance student performance. The publication was followed by *Goals 2000* (United States Department of Education, 1998), an initiative designed to focus on measurement-driven instruction and reflection on achievement of objectives or standards (Herman, 1997). Following this political movement, *No Child Left Behind* (United States Department of Education, 1998) took the measurement-driven initiative to new levels of accountability and scrutiny of our public school systems. Instead of relying on high-stakes tests as a means for improving instruction or assisting struggling students, standardized tests are now primarily used to satisfy public desires to compare and scrutinize schools and teachers and to push for change in our school systems (Dutt-Doner & Maddox, 1998). The pressure from this accountability movement leaves teachers and schools scrambling to focus on pushing

students to reach standards on large-scale state mandated tests, often with the result of leaving other important curricular goals behind.

Yet Herman (1997) points out that "the match between valued goals and large scale assessments is currently imperfect," (p. 198). Most states in our nation have developed rigorous content and performance standards, yet most testing still emphasizes multiple choice testing (Bond, Braskamp, & Roeber, 1996), which does not always effectively measure complex thinking or rigorous academic skills. "In short, just because an assessment task looks like it measures thinking and is aligned with rigorous academic skills does not mean it necessarily does so," (Herman, 1997, p. 198). The historical move from evaluation from for diagnostic purposes to the measurement world full of comparison and scrutiny with regards to standards-based testing has left our nation with an imperfect assessment system. This imperfect system is what continues to drive educational reform efforts, and what has currently led many educators into the current push for interdisciplinary teaching and authentic assessment practices.

Historical Review of Interdisciplinary Teaching

From the late 1970s through the late 1980s, there was a surge in efforts to integrate educational curricula and implement interdisciplinary teaching practices (Drake & Burns, 2004; Haney, Wang, Keil, & Zoffel, 2007; Palmer, 1998). However, the standards-based reform movements of the 1990s pulled away from this momentum, as educational institutions have become more pressured to place strict focus on standards, isolated disciplinary approaches, and more traditional teaching practices. According to Haney, et al. (2007), continued budgetary struggles in American school systems also continue to halt momentum for integrated learning, since schools seek to be

"financially 'lean'" (p. 32) to meet the demands of significant budget constraints. Layoffs of teachers, limited planning time, less flexible schedules, and less opportunity for team teaching make interdisciplinary teaching all the more difficult to implement in our current education system (Haney et al., 2007). However, research over the past decade has again begun to focus on the positive outcomes of integrated curricular approaches, including interdisciplinary teaching and professional learning communities (Ackerman & Perkins, 1989; DuFour & Eaker, 1998; Field, Lee, & Field, 1994; Haney et al., 2007; Ivanitskaya, Clark, Montgomery, & Primeau, 2002; Letterman & Dugan, 2004; McGehee, 2001). Such research has contributed to a gradual theoretical shift in approach by many educators.

Theoretical Review of Interdisciplinary Teaching

Theoretical review of interdisciplinary teaching suggests that this approach can have a positive impact on secondary schooling outcomes. Potential benefits of interdisciplinary strategies include increased teacher collaboration, affective student growth, enhanced diversity of perspectives, higher levels of learning, and increased academic achievement (Ackerman & Perkins, 1989; Field, Lee, & Field, 1994; Haney et al., 2007; Ivanitskaya et al., 2002; Letterman & Dugan, 2004; McGehee, 2001). Challenges include time, logistics, intellectual demand, lack of expertise or experience, and lack of availability or use of supporting frameworks for effective planning (Letterman & Dugan, 2004; Jacobs, 1989; McGehee, 2001). Overall, empirical literature supporting interdisciplinary teaching is limited, particularly with regards to secondary schooling.

Benefits to teachers. At the secondary level, teachers have historically remained segregated by discipline and grade level. Most teachers work in isolation in their own classrooms. The idea of "professional learning communities" (DuFour & Eaker, 1998), however, has become a commonplace educational notion in the past decade, as we have realized that educators can learn so much more from each other and can have greater educational impacts as a collaborative team. One of the benefits of interdisciplinary teaching is the deviation from typical classroom isolation that teachers face (Letterman & Dugan, 2004). Interdisciplinary teaching requires deep and ongoing collaboration, resulting in opportunities for increased sharing, modeling, and learning of new ideas and strategies (Letterman & Dugan, 2004). Teachers are pushed to engage in philosophical conversations around pedagogy and disciplinary content, inspiring thoughtful discussion about curriculum, student learning, and innovation. Instead of being isolated, teachers have the opportunity to learn from each other and deepen their practice.

Benefits for students. ***Affective development.*** Theorists believe that interdisciplinary teaching engages educators in deeper philosophical dialogue around content and pedagogy and that the resulting curriculum is more interesting and more challenging (Drake & Burns, 2004; Roberts & Kellough, 2008; Kovalik, 1993; Tchudy & Lafer, 1996; Letterman & Dugan, 2004). With a more interesting and challenging curriculum, students are more likely to be engaged; literature states that this type of engagement motivates students to participate at a higher level and to push their learning to deeper levels (McGehee, 2001; Ivanitskaya et al., 2002). Interdisciplinary work also tends to engage students in peer and adult collaboration. Students interact more with each other as they contextualize learning, discuss cross-curricular ideas, and engage in

dialogue among a diversity of teachers, students, and disciplines. This increased collaboration develops students' social, communicative, and interpersonal skills (Letterman & Dugan, 2004). Additionally, student-student and student-teacher relationships are believed to be enhanced by this type of learning and thinking, along with students' self-esteem and affective beliefs about their development and abilities as learners (Ivanitskaya et al., 2002).

Perspectives. Interdisciplinary curriculum tends to bring more relevance to the curriculum for the students, helping them make personal connections with their learning experiences (Drake & Burns, 2004; Ivanitskaya et al., 2002; Kovalik ,1993; McGehee, 2001; Tchudy & Lafer, 1996). Connecting curriculum to students' lives may further increase student motivation and attitudes towards learning (Ivanstkaya et al., 2002). Additionally, the exposure students receive to more diverse intellectual perspectives, experiences, and viewpoints may help expand students' cultural sensitivity. In theory, interdisciplinary learning promotes diversity (Letterman & Dugan, 2004) and modifies "perspectives and attitudes (e.g. enhanced sensitivity to the ethical dimensions of issues)," (Field et al., 1994; Ivanitskaya et al., 2002, p. 101).

Student learning. Most importantly for this particular study, interdisciplinary teaching is thought to promote more complex student learning and higher levels of academic thought and achievement (Drake & Burns, 2004; Ivanitskaya et al., 2002; Letterman & Dugan, 2004; Kovalik, 1993; McGehee, 2001; Tchudy & Lafer, 1996). Students are given more opportunities to seek meaning within their curriculum and are engaged in deep learning, as opposed to rote learning or memorization (Ivanitskaya et al., 2002). Students are engaged in critical analysis of ideas from multiple perspectives

and are expected to apply, synthesize, generalize, and evaluate information, instead of just memorizing facts or comprehending simple concepts (Ivanitskaya et al., 2002; Letterman & Dugan, 2004). Interdisciplinary teaching promotes complex critical thinking and the development of higher-order cognitive skills, instead of just developing specific content knowledge (Ackerman & Perkins, 1989; Drake & Burns, 2004; Field et al., 1994; Ivanitskaya et al., 2002; Kovalik ,1993; McGehee, 2001; Tchudy & Lafer, 1996). As a result of this type of engagement in the learning process, students develop increased judgment and analysis skills and are better prepared to think more independently without needing as much structure and guidance (Ivanitskaya et al., 2002).

Additionally, students are expected to reflect on academic processing and their own learning as they grapple with interdisciplinary concepts. Students are asked to internalize and process ideas from different perspectives to create personal opinions and contextualized understanding. There is more self-management of cognitive processes and understanding, thus developing students' megacognitive processing skills (Ivanitskaya et al., 2002). Students theoretically increase their ability to and develop more confidence to reason, think, and make informed decisions.

Challenges. Despite the inherent theoretical benefits of interdisciplinary instruction, it is, of course, not without challenges. If such pedagogical practices were easier to implement, they would likely be implemented more often by educators throughout primary, secondary, and higher education institutions.

The most obvious challenges to interdisciplinary collaboration are time and logistics (Letterman & Dugan, 2004; McGehee, 2001). The amount of time needed for

teachers to work collaboratively on curriculum tends can be greatly underestimated. Furthermore, planning with multiple instructors inherently leads to less individual autonomy in the classroom and limited flexibility with curriculum, methods, and timing of lessons (Letterman & Dugan, 2004). Logistical coordination is particularly difficult, particularly at the high school level where students are programmed individually into several different classes with different teachers from one another. The organizational structures of most high schools simply do not lend themselves well to the scheduling structure needed for interdisciplinary teaching. Additionally, some institutions do not necessarily support this type of collaboration among educators (Letterman & Dugan, 2004).

The intellectual challenges of interdisciplinary pedagogical practices are also great. Teachers are often forced outside of their individual and collaborative comfort zones to meet the rigorous demands of interdisciplinary thinking and planning (McGehee, 2001). Teachers are not necessarily trained or experienced in this type of collaboration, which can be intellectually intimidating. For successful interdisciplinary collaboration, teachers should have more than just knowledge of their course and content. McGehee recommends teachers bring knowledge of pedagogy, subject matter, and bigger-picture curriculum scope and sequence knowledge. Teachers should have confidence in the promise of their intellectual contributions to curriculum development.

Even when teachers possess the necessary skill and experience, they are not always provided with a useful or effective planning framework which leads them to successful interdisciplinary collaboration. Many educators attempting to integrate disciplines in a meaningful context tend to run into what McGehee (2001), citing Jacobs

(1989) describes as the "Potporri Problem" and the "Polarity Problem" (p. 380).The Potporri Problem is the tendency of teachers to use a "sampling of knowledge from each discipline" (p. 380), without really focusing on the integration of big ideas from the various disciplines. This multidisciplinary approach is a common misconstruction of the essence of interdisciplinary thinking. Multidisciplinary thinking implies loose connections between topics from each discipline, whereas true interdisciplinary thinking implies a deeper overlapping integration of larger overarching concepts or big ideas which represent the core essence of a discipline. The Polarity Problem is the idea that concepts or instruction are either interdisciplinary or not; sometimes the lines are not quite that transparent. There are some concepts and disciplines have obvious interdisciplinary connections and some important concepts from specific disciplines that just won't be that easy to integrate.

Limitations of existing literature. Since interdisciplinary teaching is logistically difficult to implement at the high school level and has historically been implemented on such as small scale, there is inherently an overall lack of research regarding the results of such methods. Most empirical literature regarding interdisciplinary teaching is focused on higher education experiences and is still limited in scope. Letterman and Dugan (2004) also suggest a lack of availability of instructional information for approaching collaborative interdisciplinary planning in a meaningful and effective manner. Attempting to collaborate in a meaningful way without a well-researched supporting framework or appropriate planning materials may lead to difficulty in implementing such strategies, as described above by McGehee (2001).

Empirical Review of Interdisciplinary Teaching

Outcomes of interdisciplinary teaching. Most literature regarding the outcomes of interdisciplinary teaching is limited to anecdotal observations, to the experience of the teachers, and/or to higher education settings (Barisonzi & Thorn, 2003; Letterman & Dugan, 2004; McGehee, 2001; Nation, 2008).

Impact on teachers. A study was conducted to investigate teachers' practices and beliefs throughout a two-year professional development experience focused on the development and implementation of an interdisciplinary, problem-based curriculum in three rural and three urban public and private schools among grades six, seven, and eight in Ohio (Haney et al., 2007). This study examined context beliefs, self-efficacy, beliefs about the use of constructivism approaches, and frequency of use of traditional and reform strategies of eighteen teachers who participated in the implementation of an interdisciplinary environmental health science program called EXCITE. Eight of these teachers taught science and ten taught a wide range of other disciplines. The following tools were utilized in the study: the Context Beliefs About Teaching Science Instrument, the Science Teaching Efficacy Belief Instrument, the Constructivist Learning Environment Survey, and the Best Practices Survey. There were significant and positive differences for pre- and post- test data from teachers regarding context beliefs, self-efficacy, and constructivist beliefs following implementation of the EXCITE program, and teachers implemented traditional teaching strategies less often following program implementation. The study showed that teachers took to the interdisciplinary program opportunities and positively shifted their educational beliefs and perspectives regarding teaching strategies and instructional approach and values.

Student achievement. The study of the implementation of the interdisciplinary environmental health curriculum in Ohio middle schools (Haney et al., 2007) also examined student performance data on Ohio proficiency tests before and after the two year project and found that scores increased in all five subjects, including mathematics, science, social studies, reading, and writing. When compared to the growth in scores of their non-EXCITE peers, proficiency test scores improved more significantly for peers in the program versus other peers in writing, mathematics, and citizenship. Interdisciplinary programming, then, appears to have a significant impact on student learning and achievement.

Historical Review of Alternative Assessment

The accountability movement and emphasis on standards-based testing over the last few decades have left educators with an enhanced need to re-focus not just on a more integrated approach to learning, but on more individualized learning, multiple intelligences (Gardner, 1993), and higher-order thinking skills through a different approach to assessment (Koretz, 2008).

In the late 1980s, many large scale assessments began to enhance multiple choice tests with *performance assessments*, designed to measure higher-order thinking skills and to present students with opportunities to demonstrate the type of learning that would be required outside of school (Koretz, 2008). Although this movement grew quickly, it proved expensive, time consuming, and laden with logistical difficulty and unreliability. Along with the performance assessment movement, however, came the development of content and performance standards, of which large-scale assessments would soon be aligned to. Standards-based education has now become the norm in our public schooling

systems. As Koretz (2008) points out, "Regardless of the arguments for and against standards-based accountability, I believe that it is with us for the foreseeable future" (p. 73).

As we delve deeper into educational accountability and standards-based testing, the world workforce is moving towards one of autonomy and outsourcing (Pink, 2006). Workplace systems seek efficiency and the job market demands creativity and innovation. Within this information age (Pink, 2006), where new technology serves to bridge easy connections across developed and developing countries, our students are left competing globally, and they often are not competitive enough for what the global market demands.

As the standards-based movement progresses and global demands change, there is a surfacing call for authentic, standards-based reform (Thompson, 2001). Accountability and transparency are thought to inspire honesty, competition and results. Assessment coupled with this accountability, including incentives for good performance and sanctions for poor performance, will "motivate students to learn better, teachers to teach better, and schools to be more educationally effective" (Herman, 1997, p. 197). However, a single testing event should never result in high-stakes consequences (Thompson, 2001). Accurate knowledge of performance requires multiple avenues for demonstration of skills and understanding (Gardner, 1993). Similarly, the ultimate purpose of education, as most professional educators would likely agree, is not to raise test scores, but to enable "all students to achieve as much of their creative, intellectual, and social potential as possible" (Thompson, 2001, p. 360). Indeed, as Pink (2006) describes, this is what the global marketplace is demanding. Such intentions considered,

standards-based reform has been an important step in demanding rigorous and thoughtful education, but along with shifting expectations for the future demands on our students, we must begin to judge students on more than their ability to take standardized tests.

As our nation begins to recognize that tests scores do not always mean success, critical thinking and the ability to problem-solve are becoming more and more important. To address some of these growing concerns, education theorists have begun to take a look at the way we engage students in our classrooms and have also explored the need to shift our instructional and assessment practices in order to more adequately prepare students for the real world. Concurrently, the standards-based movement and high-stakes testing have continued to grow, leaving many unanswered questions about new educational theories regarding learning and assessment, the practical application in classroom settings, and how to measure success. This review of theoretical and empirical research will further explore existing research regarding alternative assessment, particularly authentic assessment, and will supplement existing research on the measurement of outcomes of such assessment practices.

Overview of Alternative Approaches to Assessment

Theoretical overview: Need for alternative approach. Numerous studies have described student perspectives of traditional schooling as boring (Darling-Hammond, 1997; Johnson, 2008; Lee, Bryk, & Smith, 1993). Traditional schooling, as defined by Johnson, is schooling where all decisions are made by faculty with no student input, grades are used without much additional feedback, learning goals are all prescribed by teachers, and lecture is the primary means of instruction. Wiggins and McTighe (2005)

and others (Moon, Brighton, Callahan, & Robinson, 2005; Glass & Rose, 2008) remind us that direct instruction and drilling for tests are not the answer for inspiring our students to perform.

Empirical review: Perspectives on alternative approaches. ***Teacher and public perspectives.*** Educators face an enormous challenge in working within the existing paradigm of standards-based testing and the measurement world. While teachers generally tend to support alternative approaches to assessment, parents and other public stakeholders tend to be more supportive of standardized testing, an indication of the surmounting public pressure and perspective which accountability and standards-based reform have pushed. A study by Trepanier-Street, McNair, and Donegan (2001), surveyed 298 teachers in a metropolitan area in the United States, and compared perspectives and practices of 172 lower (first and second) grade and 126 upper (third, fourth, and fifth) grade elementary school teachers in five counties regarding assessment. The participants were mostly Caucasian-American women, who taught in their own classrooms and who had a wide variety of teaching experience. Teachers were surveyed through faculty and university outreach, using a survey specifically designed to rate (a) factors used in decision-making about assessment, (b) amount of trust lent to various sources of information regarding student progress, and (c) attitudes regarding parent involvement (Trepanier-Street et al., 2001).

In their study, Trepanier-Street et al. (2001) found that most teachers both found value in and implemented a variety of assessment and alternative assessment strategies. This study found that assessment strategies were chosen by teachers because of the objectives these assessments met, for their usefulness for individual student planning

and for ease of implementation (Trepanier-Street et al., 2001). However, the study found that parents and public stakeholders held inconsistent beliefs about alternative assessment. The values of these other public constituents were evidently more aligned to a standardized testing system, while teachers reported inherent value in alternative assessment practices. This study indicates that approaches to education are changing and that there is still much room to develop more public support for alternative assessment practices (Trepanier-Street et al.).

Student perspectives. Similarly, Paris, Roth, and Turner (2000) reported findings from three studies of students' perceptions of standardized achievement tests which indicated an overall disillusionment with standardized tests, especially as students grow older. In Study 1, Paris et al. examined attitudes of 974 students in grades two to eleven to standardized achievement tests in 46 classrooms in Michigan, California, Arizona, and Florida. The tests referred to for these participants were of the following: California Achievement Test (CAT); Stanford Achievement Test (SAT); Iowa Test of Basic Skills (ITBS); or the Comprehensive Test of Basic Skills (CTBS), all which assess reading and mathematics using similar methods. Study 2 compared attitudes of 240 high and low-achieving students in grades four, seven and ten towards state-mandated tests. Study 3 compared 61 fifth grade and 65 eighth grade students' views and reactions to both standardized tests and to routine classroom tests. Overall, after examining results from all three studies, Paris et al. (2000) found not only that students grew more negative and perceived the test as less important as they grew older, but that lower-performing students were particularly more disillusioned by the test and often showed feelings of higher anxiety and concern about embarrassment as a result of their test scores. These

studies indicated that standardized tests can have a negative impact on students' self-confidence and self-worth. A more student-centered approach to assessment would seek to assess students on other measures aside from these standardized tests to develop students' applied skills and individual interests within the curriculum.

Theoretical overview: Impact of alternative approach on students. In addition to teacher perspectives regarding the importance of alternative assessment measures, some theorists (Farr & Trumball, 1997) believe that educational practices can have lasting or potentially permanent influences on students. These authors also conclude that student feelings of self-worth and self-efficacy can be damaged by achievement tests and test scores, especially for lower performing students.

Along with the importance of addressing affective growth and developing self-worth and positive attitudes towards learning, it is important to consider strategies for enhancing cognitive development and the development of lasting understanding. Theorists such as Moon et al. (2005) suggest we help students "gain understanding through the construction of their own knowledge and making interconnections among facts and concepts within and across disciplines" (p. 129). Regarded theorists Wiggins and McTighe (2005) describe this interconnected construction of knowledge *enduring understandings*, a concept which suggests the need for teaching and learning to extend beyond traditional instructional strategies and assessment techniques. Despite the national push for standards and test-preparation, students still must be prepared to be critical thinkers who can thrive in the work place as creative innovators and skilled academics.

In consideration of the development of affective growth, multiple learning styles, student interest and development of enduring understanding, education theory has been gradually shifting from the preference of direct instruction and traditional testing methods to focus around strategies that more interactively involve students with their learning. Consequently, many schools have shifted from traditional structures and traditional curricular approaches to more student-centered learning environments. Some of these schools have shown considerable results in favor of non-traditional learning environments.

Empirical review: Impact of alternative approach on students. A study of student motivation at a traditional vs. non-traditional school reveals that students may become more engaged in non-traditional settings (Johnson, 2008). Johnson's study focused on a comparison of two schools: one with a traditional, adult-directed environment with little student input and a traditional teaching and grading system, and another with a non-traditional structure allowing much student input, evaluations instead of grades, and a collaborative learning environment where teachers learn alongside the students. The traditional school was located in the Midwestern United States and the non-traditional school was located in the Northwestern United States. Both schools held similar demographics, in terms of median income, school admissions, gender, grade level, state test scores and graduation and college rates. The traditional school had a lower socioeconomic factor (10% free and reduced lunch versus 30% free and reduced lunch) and was more ethnically diverse, but both schools served a general student population and did not specifically focused on an *at-risk* student population. Forty tenth grade and 40 twelfth grade students from each school were matched for analysis. The

study measured student engagement at the two schools using the *experience sampling method*, which alerted student participants at each school to log their level of interest, enjoyment and concentration, along with what activity they were participating in, at regular intervals eight times per day.

Johnson (2008) found that students displayed significantly higher levels of engagement at the non-traditional school compared to the traditional school, and that student engagement generally increased when students were in school versus at home or in another location. At both schools in Johnson's study, students displayed high levels of concentration, but in the traditional school students displayed significantly lower levels of interest and enjoyment. Furthermore, these lower interest and enjoyment levels were most prominent during lectures and independent work periods, suggesting that traditional schooling methods do not engage students as well as alternative instructional and assessment methods might.

Such results indicate that low levels of student engagement may not be a "student-problem" but a "teacher-problem" (Johnson, 2008, p. 81) or a larger issue with the way our school systems are designed, indicating a need for shifting the way we practice instruction and engage students in school. The gradual shift away from traditional direct-instruction methods of teaching would move teachers towards using more student-centered instructional strategies that involve students in real-world problem solving where they collaborate, are faced with important decisions, and drive their own learning process with more increased levels of independence.

In summary, student-centered learning environments lend focus to allowing students to employ multiple learning styles and allow for multiple pathways of

demonstrating knowledge and skills. Such approaches to schooling may enhance student engagement and thus increase care for and interaction with learning. One important feature of a student-centered learning environment is the use of *alternative assessment* approaches, to be defined in the following section.

Theoretical review of alternative assessment and authentic assessment.

Alternative assessment. The simplest way to define alternative assessment is as a curricular approach that moves away from traditional instructional and assessment methods such as direct instruction (lecture) and standardized multiple choice testing (Davies & Wavering, 1999). Alternative assessments are derived from the philosophy that there are multiple intelligences (Gardner, 1993) and individual styles of learning (Dunn & Dunn, 1978), and that students should be able to demonstrate learning through multiple pathways. According to Davies and Wavering, alternative assessment serves four purposes: (a) promotion of learning, (b) promotion of higher-order thinking, (c) consideration of diverse learning styles, and (d) communication of evaluation criteria. Abadiano and Turner's (2003) criteria for alternative assessment include: link to standards, transparent scoring criteria, cognitive complexity, skills integration, multiple pathways of learning, multiple intelligences, meaningful learning, self-assessment, and culturally responsiveness. Alternative assessments include performances, exhibitions, portfolios, simulations, cooperative learning projects, journals, and other non-traditional measures. Alternative assessments are designed to "focus on process as well as product" (Davies & Wavering, 1999, p. 40).

Authentic assessment. Authentic assessment is one form of alternative assessment, which serves to make classroom learning more meaningful and relevant.

Since many researchers and theorists propose different meanings of *authentic assessment*, and often use this term interchangeably with *alternative assessment*, the sections that follow will explore literature definitions and distinguish between these two types of assessment. In addition, the problematic nature of having so many existing definitions and practices of authentic assessment will be discussed, since these differences make it challenging to compare studies and outcomes of such assessment practices.

In this study, authentic assessment will be defined in one particular manner, based on literature review and a specific application of alternative assessment. It is important to note that *authentic assessment* is described from many different theoretical approaches and through many different lenses, definitions and criteria (Palm, 2008). There are so many vague, specific and conflicting descriptions of what authentic assessment is. In attempt to decipher and generalize the range of explanations of authentic assessment, Palm conducted a study of abstracts through the online ERIC database and the math education database MATHDI, to further explore the definitions of the terms *authentic assessment*, *authentic*, *authenticity* and *performance assessment* in order to delineate the similarities, differences and miscommunications regarding these often interchangeable terms and to generalize overall classifications regarding each term. Palm's study points out the Cambridge Dictionary definition of the word *authentic* as something that is *real* or *true*. From this definition, the researchers describe their categorized descriptors of authentic assessment in terms of (a) "what it is that is supposed to be real or true" and (b) "what it is that it is supposed to be true to" (Palm, 2008, p. 6). In describing the second focus, Palm points out that the assessment is

assumed to be true to three main perspectives: (a) life beyond school, (b) curriculum and classroom practice, and (c) learning and instruction (p. 6). Three main foci were described for "what it is that is supposed to be real or true," including: (a) processes and products, (b) conditions and (c) figurative context (Palm, 2008, p. 6).

In other words, authentic assessment must help students learn and meet standards, fit within existing curricular goals, and relate to life beyond the classroom. The *authentic* context, or how the assessment relates to life, can be represented with the materials used, the setting, or the nature of the task at hand. Some aspect of the assessment experience must relate to or replicate a situation that would occur in the "real world" beyond the classroom.

Similarly, Tchudi and Lafer (1996) highlight three criteria for authentic assessment, as described by Archblad and Newmann (1988), which include: "disciplined inquiry," "integration of knowledge," and "value beyond evaluation" (p. 184-185). Wiggins (1990) also provides a detailed list of criteria for defining the characteristics of authentic assessments. Of these criteria, the indicators that will be explored in this study include "public" (require an audience), require collaboration, explore "essential" concepts (breadth not depth), are "enabling" (advance students' skills and knowledge), are contextualized, explore research or content knowledge, involve clear indicators of success, involve self-assessment, and allow room for different learning styles, aptitudes, and interests (Tchudi & Lafer, p. 186).

Moon et al. (2005) also present some overlapping parameters for defining authentic assessment, including: focus on big ideas or concepts, in depth, realistic for the classroom, focus on performance rather than right answer, promote development of

strengths, have criteria that are understood ahead of time, allowance for multiple paths to meet criteria, and require scoring that focuses on the true purpose of the task. These descriptions fit within Palm's (2008) generalization of authentic assessment as being true to (a) life beyond school, (b) curriculum and classroom practice, and (c) learning and instruction.

Essentially, authentic assessment is one specific, form of alternative assessment, which requires students to demonstrate understanding through engagement in real work tasks and scenario-based problem solving (Moon et al., 2005; Darling-Hammond, 1997). Authentic assessment requires that students perform a high level task that demonstrates new skills and understandings in a meaningful and applied context and provides students with opportunities they need to use their learning in an applied context or *authentic* setting representing what scholars and professionals in that discipline might do. Authentic assessment challenges students to conduct work that is meaningful and relevant beyond the scope of the classroom setting. The concept behind authentic assessment is that "instead of trying to judge students through tests and grades, teachers should look at their performance in authentic settings" or "what students do" (Tchudi & Lafer, 1996, p. 183). As Wiggins (1990) describes, authentic assessments are ""illstructured' challenges and roles that help students rehearse for the complex ambiguities of the 'game' of adult and professional life" (p. 2). Authentic assessments are opportunities to use new concepts and processes to grapple with complex ideas and abilities and perform meaningful and relevant tasks. Authentic assessment provides an opportunity to judge students on what they "do" in these new settings, as opposed to what they can remember on an abstract test or other derived representation of learning.

Additionally, as previously mentioned, interdisciplinary teaching often leads to an authentic assessment experience where students reach new levels of thinking and discourse (Klein, 1990). Since the real world beyond the classroom does not tend to operate in a context which separates concepts or skills into separate or isolated disciplines, the development of integrated thinking skills beyond a single-discipline can provide the context for an authentic learning experience. Although the isolation of skills and concepts into disciplines can help build depth of single subject knowledge, it can also stifle higher-order thinking, holistic understanding, and the creation of "meaningful connections between and among disciplines" (Ivanitskaya et al., 2002, p. 97). Therefore, interdisciplinary teaching, requiring the integration of knowledge across disciplines, enhances the authenticity of an assessment experience.

Overall, Palm's (2008) generalization of what authentic assessment encompasses provides the simplest definition of authentic assessment. To be clear, authentic assessment is an alternative assessment task which sets out to achieve specific learning and curricular objectives, but which specifically engages students with products, processes, conditions, or other context which they would face in the real world outside of the classroom setting.

Distinguishing authentic assessment. Although authentic assessment is specifically defined for the sake of this study, many researchers and theorists use the term *authentic assessment* as a misnomer. Alternative assessment, including performance-based tasks, portfolio displays, project-based learning, presentations, and just about anything else that is not just direct instruction, lecture, reading and multiple choice testing, is often inaccurately (for the sake of this study) referred to as *authentic*

assessment. Although these performance tasks are likely much more authentic than a multiple choice test, not all of these activities would necessarily be considered authentic assessment tasks, unless they were to meet the criteria described above. The specific discrepancy in terms, and perhaps one of the more difficult qualities to simulate in developing authentic assessment tasks, appears to be the connection to life beyond school. A well-developed alternative assessment will be true to curriculum and classroom practice, and to learning and instruction (Palms, 2008), but will not necessarily connect students explicitly with something they would encounter outside of the classroom setting.

It is important to clarify that an alternative assessment could be, but is not always, an authentic assessment. An alternative assessment, conversely, is a form of authentic assessment. Authentic assessment captures the core essence of alternative assessment, but applies alternative assessment practices on a more meaningful and applied level. Authentic assessment is essentially an alternative assessment which replicates a real-life task through the processes, products, conditions or context which are simulated during the activity. Authentic assessment is more than an experience or a project, it is an assessment which requires students to demonstrate their understanding or skill of one of the core essences of a discipline, in a way they would in a real-life setting outside of the classroom. This distinguishing feature is an important one which many researchers overlook when describing authentic assessment.

Challenge in reviewing literature. The inconsistent use of the term *authentic assessment* presents a challenge in the attempt to decipher and articulate learning outcomes and challenges of such tasks. If each researcher poses a slightly different

definition of authentic assessment, how can one ensure the outcomes measured are a result of the same conditions? Although this study has developed specific criteria with which to determine authenticity of an assignment, not all literature reviewed matches the description of such assessment. With so many different descriptions and understandings of what authentic assessment is in existing empirical and theoretical literature, it remains difficult to compare outcomes and challenges of authentic assessment when educators cannot agree upon criteria for such assessment. Given this difficulty, outcomes explored in this review of literature will compare both authentic assessment approaches, as described in this study, as well as general alternative assessment approaches, which appear to possess a similarly nature to the definition of authentic assessment implied by this study.

Outcomes of Alternative Assessment

Theoretical overview of outcomes. The ultimate theoretical purpose of any alternative assessment would be the development of *transferability* (Bransford et al., 2000; Wiggins & McTighe, 2005) of concepts and skills and the development of higher-order conceptual and skill-based understandings. With this development, students should become more effective, capable and innovative contributors to the global workforce. Within this larger purpose, authentic assessment also sets out to inspire students to become more engaged in school, and thus to develop a more intrinsic interest and motivation to performance. The hope is that this intrinsic interest or motivation will further enhance effort, academic performance, and sustained competency for higher levels of thinking. With performance tasks calling for the demonstration of student understanding beyond black and white *true-false* or multiple choice answers, the door

also opens to allow teachers to better understand student strengths and weaknesses, in order to change instruction to better meet student needs. A review of theoretical and empirical literature of all of these potential outcomes of authentic assessment follows.

Empirical review: Teacher perspectives and support. One general outcome of alternative assessment practices is the change in teachers' approach and their consequential ability to understand student learning on a deeper level. Research (Abadiano & Turner, 2003; Avery et al., 2000; Bauer & Garcia, 2002) shows that teachers are able to understand more about their students from authentic assessment tasks, that these tasks encourage teachers to conduct more student-centered classrooms, and that students consequently became self-directed and interested in their learning.

One such outcome of alternative assessment can be gleaned from a study by Bauer and Garcia (2002), as described by Abadiano and Turner (2003). In this study, Bauer and Garcia found that using alternative assessment measures can change the way educators understand and work with students in the classroom. Specifically, 50 classroom observations were conducted in a second grade classroom to determine (a) the link between the use of alternative assessment and student-centered instruction in reading and (b) the factors that helped support the implementation of an alternative literacy task and the move towards student-centered classroom practices. Observations focused on which students participated, how they participated, and the actions of the teachers. Collected data were qualitatively coded to compare assessment data with instructional data.

The results of the study indicated that prior to the introduction of alternative assessment, the classroom focused on teacher-centered instruction, teacher-selected texts

and teacher-directed discourse, and that after the introduction of an authentic assessment, classroom approaches became more student-centered; students met with their teacher individually and chose their own texts, and the teacher led student-oriented discussions and focused more on individual student needs (Bauer & Garcia, 2002). Bauer and Garcia generalized that alternative assessments can encourage student self-direction and evaluation and enhance important teacher knowledge of students, which enhances instruction capability (Abadiano & Turner, 2003).

Avery et al. (2000) also measured teacher perspectives and perceived outcomes from one type authentic assessment experience in a United States History course. Five teachers at a high school in Minneapolis implemented an authentic assessment task during a month-long immigration unit, using data from students' families to develop an understanding of the common experience of United States immigrants and their perspectives. The five teachers had different backgrounds and teaching styles, and for some this task required "significant changes in their teaching style—from being teacher-centered to more student-centered" (Avery et al., p. 374) while for others the task was more similar to previous teaching methods. The teachers were interviewed about their implementation of the authentic assessment task and participating students were surveyed about their perceptions of the task.

Of the five teachers who implemented the immigration task, the mean rating of their likelihood to use the task in their classroom again was an 8.8 on a scale of 1 to 10, where 1 is *definitely not* and 10 is *definitely*. Teachers were positive in response to the assignment and noted an increased level of engagement amongst their students, higher-order thinking than more traditional assignments, and an enhanced sense of community

within their classrooms. Specifically, the task appeared to help students connect their course material to the outside world (Avery et al., 2000). Student surveys similarly indicated a favorable response to the authentic assessment task. Students identified that the assessment task "was more interesting, made them think more, helped them understand information better, and caused them to consider a variety of perspectives" (Avery et al., p. 375).

Authentic assessment approaches, then, appear to have an impact on teacher perspectives and on teacher abilities to better understand their students' academic strengths and challenges. Additionally, students appear to be more self-directed and engaged as a result of the tasks at hand.

Theoretical review: Motivation. In addition to teacher perspectives and abilities, the shift towards student-centered classroom approaches is likely to motivate more students to participate and learn. This section will examine the theoretical forms and causes of motivation and the link between student-centered approaches, engagement and motivation.

A lingering question that has been explored for decades and continues to arouse much discussion and further study as we try to inspire our students to graduate from high school and learn the most they can in school is: what motivates students to learn? According to Linnenbrink and Pintrink (2002), motivation is scaled, meaning that a student is not just either motivated or unmotivated to engage in learning tasks, but that they may fall somewhere in-between. Motivation is multi-faceted and there are many different types and indicators of motivation.

Fortunately, Linnenbrink and Pintrich (2002) point out that "motivation is *not* a stable trait of an individual, but is more situated, contextual, and domain-specific" (p. 314). Similarly, Price (2008) explains that "motivation can be cultivated and inculcated" (p. 43) and that "lack of motivation is reversible" (p. 27). This notion implies that teachers' instructional strategies can in fact change students' interests and motivations to complete tasks and achieve academically. According to Linnenbrink and Pintrich (2002), "instruction and design can make a difference" (p. 314), which means that the role of the teacher and the course materials, instructional design, and assessment tools may all play a role in motivating students to learn. The idea that levels of student motivation can be altered through our own actions as educators makes our understanding of student motivation very important for our curricular approaches to education reform.

Glass and Rose (2008) also discuss the importance of studies being connected to the outside world that students live in. "They see [learning] really means something" (Glass & Rose, 2008, p. 11). Johnson (2008) explains that motivation also comes from belonging and competence. Overall, the combined results from these theoretical and empirical studies imply that instructional technique, meaning and relevance, collaboration, responsibility, and peer dependence are important qualifiers for motivating students to learn. When students feel connected to their learning experience and see meaning, when others collaborate with and depend on them, and when they feel like they have the skills they need to be successful, they are more motivated to be involved in their learning.

According to Price (2008), the following factors, among others, influence student motivation:

- Ability to see connection between academic achievement and opportunity for success in life (p. 27)
- Feelings of being valued and appreciated (p. 30)
- Feelings of belonging to a socially acceptable group (p. 36)
- Feelings of acceptance from and connection to peers (p. 26)
- Feelings of being important and depended on (presence is noticed) (p. 37)
- Feelings of being given "opportunity to share the status of responsible and competent adults" (p. 32)
- Perception of being challenged (p. 31)

Such factors are naturally assumed to be more likely to be present within an authentic assessment experience than a traditional assessment framework, implying that motivation may be developed more prominently through the use of alternative assessment practices.

The two most common types of motivation are extrinsic and intrinsic motivation. As defined by Sergiovanni (2007), extrinsic motivation is "based upon the value a person receives from the external context of the work" (p. 128), such as receiving grades, money, or other rewards. Extrinsic motivation is generally done to "please others" (p. 129), receive rewards or avoid punishments. Intrinsic motivation, on the other hand, is "based upon the value received from the work itself" (Sergiovanni, 2007, p. 127). Although Sergiovanni discusses motivation as applied to educators and leaders, he warns us of the more generally applicable dangers of relying on extrinsic motivators. He states that "extrinsic rewards can actually diminish one's intrinsic interest in the

work" (Sergiovanni, 2007, p. 129) since people may become discouraged or not care once external motivators are no longer present.

None of the above literature regarding motivation refer to grades or external rewards, further supporting Sergiovanni's (2007) suggestion that extrinsic rewards may not be as meaningful or effective as intrinsic rewards. Likewise, results from a study done by Pederson and Williams (2004) comparing different ways of assessing students in an instructionally stimulating seventh grade classroom implied that grading (as an extrinsic reward) is not very valuable in learning environments that are already engaging for students. Pederson and Williams' study was designed to study the effects of assessment grading practices on the learning and motivation of 77 seventh-grade science students during a student-centered computerized assessment in three classes in a Southwestern city in the United States. The assessment method was the independent variable, and included: (a) in-class assignments, (b) a collection of artifacts and (c) peer-and-self-assessment. The dependent variables were measured using: (a) a factual science knowledge test, (b) the *Scale of Intrinsic versus Extrinsic Orientation in the Classroom* and (c) student interviews. Although these particular assessment methods within this student-centered instructional context did not reveal any differences in student performance or motivation, students did reveal that the use of grades for this particular instructional sequence did not motivate them to work harder on the assignments. Such results again indicate that grades and other external motivators are not what drive students to want to learn (Pederson & Williams, 2004).

Theoretical review: Interest. Linnenbrink and Pintrich (2002) build off Mitchell's (1993) work to explain that interest in situational interest in classroom

activities can be thought about in two categories called *catch* and *hold*. Catch factors are those that stimulate student interest initially, which generally consist of attention-grabbing instructional techniques that facilitate interest quickly. Hold factors are those that empower students to sustain interest and engagement over time, such as making academic content meaningful and relevant or facilitating deep student involvement. Linnenbrink and Pintrich suggest that activities that facilitate this type of sustained involvement in the learning process tend to include collaborative learning, engaging discussion, and decision-making. The hold factor is thought to develop a longer-term intrinsic interest in learning, whereas the catch factor would be likely to provide immediate short-term interest, but not necessarily a deep and sustained interest over time. Although catch and hold terminology is not commonly used in regular educational jargon and there is not yet a high school measurement tool for catch and hold interest indicators (Linnenbrink & Pintrich, 2002), the idea of inspiring students beyond the immediate threshold of engagement is commonplace, as most educators and school missions seek to inspire students to be lifelong learners. Sustaining a lasting interest in learning is perceptibly more important than engaging students at the on-set of a lesson.

Empirical review: Student-centered approaches, interest and motivation. It is assumed that developing intrinsic motivation will assist in developing the lasting interest that Linnenbrink and Pintrich (2002) describe. Pedersen and Williams (2004) further explain that the intrinsic motivation that we desire of our adolescent students may be developed through student-centered learning approaches. With traditional instructional and assessment approaches, teachers are the focus of classroom instruction. Teachers generally set the learning objectives, determine activities and related problem-

solving strategies for students, and assess students' progress (Pedersen & Williams, 2004). With student-centered approaches to classroom instruction, students are given complex tasks and identify their own needs for obtaining skills, information and resources and teachers act more as facilitators that help push students to collaborate and problem solve (Pedersen & Williams, 2004). Such tasks capture the nature of authentic assessment. Pederson and Williams explain that empowering student decision-making around learning and progress can lead to increased levels of intrinsic motivation among adolescent students.

The impact of a student-centered approach was evident in a study by Wood and Sellers (1997) which examined motivation of elementary students across seven schools in first, second, third and fourth grade after implementation of a problem-based and traditional instructional setting. Participants included rural to urban students from twelve elementary, middle and high schools in a district with mostly Caucasian-American students from a variety of socioeconomic backgrounds. Motivation and achievement among the following groups of students was compared: (a) students receiving one year of problem-based mathematics instruction, (b) students receiving two years of problem-based mathematics instruction, and (c) students receiving only textbook instruction. A *Personal Goals and Beliefs Questionnaire* was used for analysis of students' beliefs and motivation for mathematics. (The results of academic performance will be described in the *Achievement as Measured by Test Scores* section of this literature review.) The motivational subscales of the survey include: working hard, making sense and collaborating, being superior to peers and not having to work. Motivational results of this study indicate that students outperforming their peers did not motivate the students,

but the desire to find their own problem-solving methods, indicating that student-centered learning may lead to enhanced efforts to truly understand, rather than to increased peer competition.

Such studies on the motivation of students imply that a more student-centered approach to teaching and learning may lead to increased engagement and motivation to participate in school. This motivation does not appear to be linked to extrinsic rewards such as grades.

Empirical review: Assessment and motivation. Given the link between student-centered practices and student engagement and motivation, it seems that the type of assessment used in the classroom would naturally have an impact on the type and level of motivation of participating students. In this regard, Kember, Ho, and Hong (2008) conducted a study on students in higher education to determine what motivated or de-motivated college-aged students. Thirty-six third-year undergraduate students in nine programs, across three universities in Hong Kong, were interviewed to determine the aspects of teaching and learning that affected their level of motivation.

Kember et al. (2008) found eight factors were most important for motivating students to learn and achieve. These factors were: "establishing relevance, establishing interest, allowing choice of courses so that interest can be followed, learning activities, teaching for understanding, assessment of learning activities, close teacher–student relationships and sense of belonging between classmates" (Kember et al., 2008, p. 249). The most commonly cited factor, and thus likely the most important factor, among the surveyed students was *establishing relevance*. Students reported that not seeing relevance in their learning could easily de-motivate their learning (Kember et al., 2008).

Although this study was conducted at the college level, this finding has important implications for the way we design curriculum and assessment in secondary schooling as well. It can be assumed, then, that where there is more relevance in the curriculum, students are more motivated to participate and do well, implying an innate importance for authentic assessment strategies.

Empirical review: Motivational impact on learning. Additionally, student engagement and motivation are considered good predictors of learning because the more engaged students are, the more they tend to interact with and learn the material (Carini, Kuh, and Klein, 2006, Johnson, 2008; Linnenbrink & Pintrich, 2002; Price, 2008; Steinmayr & Spinath, 2009). It seems straightforward enough that students would learn more if they were motivated or inspired to learn. Linnenbrink and Pintrich (2002) describe that motivation and cognition are highly integrated and together ultimately influence student learning. Johnson (2008) explains one theory on motivation suggesting that active learning nurtures academic engagement, which will result in increased opportunity for learning. Price (2008) also suggests that "motivation and mind-set towards academic achievement can affect how [students] will do in school" (p. 27).

Similarly, Carini, Kuh, and Klein (2006) sought to explore the relationship between student engagement and learning outcomes, among other questions. For this study, RAND researchers studied 1,058 college students within 14 different colleges and universities and used the Graduate Record Examination (GRE), cognitive tests developed by RAND researchers, and grade point average (GPA) for cognitive and academic measures, and the National Survey of Student Engagement (NSSE) instrument to measure student engagement and students' dedication of time and energy to

purposeful activities. Types of student engagement measured on the NSSE self-report include, among other factors, *level of challenge*, *active and collaborative learning*, *student-faculty interaction*, *enriching educational experiences*, *quality of relationships*, and *higher-order thinking* (Carini et al., 2006). The types and levels of self-reported student engagement were then compared to RAND cognitive tests, designed to measure critical thinking, to GRE test essays prompts, and to grade point average (GPA).

The study conducted by Carini et al. (2006) found statistically significant correlations between several student engagement factors and scores on the RAND and GRE tests, indicating that there is a positive relationship between how engaged students are in their learning and how they perform academically. None of the statistical correlations between student engagement factors and RAND results were negative, and the greatest correlations with RAND results were with: quality of relationships (0.14), reading and writing (0.12), level of academic challenge (0.11), integration of diversity into coursework (0.10), enriching educational experiences (0.09), higher-order thinking (0.08) and supportive campus climate (0.08). Most statistical correlations between student engagement factors and GRE results were positive and the greatest correlation with GRE results was with: reading and writing (0.16), enriching educational experiences (0.13), integration of diversity into coursework (0.12), level of academic challenge (0.10), and higher-order thinking (0.08). Student engagement through reading and writing, enriching educational experiences, integration of diversity, level of academic challenge, and higher-order thinking were significant in both comparisons, implying that the way schools engage students in enriching experiences and higher order thinking challenges may have a positive impact on cognitive achievement. Although this

study was conducted at the post-secondary level, it is assumed that similar trends exist at the high school level as well.

Another important study linked motivation to school achievement. Steinmayr and Spinath (2009) examined how different motivational attributes and intelligence predicted school achievement in language and mathematics in a 342 eleventh and twelfth grade students in a high school in Germany. An *Intelligence Structure Test, Achievement Motives Scale* and other Likert scale survey measures of students' goal orientations, ability self-perception and values were measured and compared to grades (school performance) in German and mathematics (Steinmayr & Spinath, 2009). Among other findings, Steinmayr and Spinath found most of the motivational variables were positively correlated with school achievement, but that intelligence was not necessarily a significant predictor of achievement. This indicates that regardless of initial academic aptitude or intelligence, students can achieve at high levels, and thus that instructional and assessment factors can positively or negatively influence student achievement.

Overall, motivation is contextually dependent (Linnenbrink & Pintrich, 2002) and depends on factors such as relevance of curriculum (Kember et al., 2008). If motivation can outweigh the impact of intelligence on academic achievement (Steinmayr & Spinath, 2009), then the way we design curriculum and assessment practices for students may have a significant effect on student learning.

Theoretical overview: Achievement. The purpose of authentic assessment is to enhance student interest, motivation and achievement through engagement in a real and meaningful learning context derived to foster student learning. Understanding is most typically measured through student performance on standardized tests, but there are

other important and effective ways to measure deeper understanding. This section will explore the relationship between alternative assessments and performance achievement tests, but will also justify the need for alternative measurements of student understanding at various levels.

Empirical review: Achievement as measured by test scores. Since authentic assessment allows students to make connections between academic achievement and their lives, encourages collaboration, and presents students with challenges that allow them to feel like competent and responsible adults, it should inspire the engagement, intrinsic motivation, and lasting academic interest that are so important for the academic development of adolescents. Furthermore, authentic assessment aims to allow students to grapple with challenging situations and apply or transfer specific concepts and processes to new situations and concepts. Students must use more abstract abilities to make sense of abstract situations, make connections between big ideas, and evaluate their thinking and learning from a more complex perspective, theoretically encouraging the development of the type of understanding that we hope for our students to achieve. Many studies have measured outcomes of student understanding through the use of standardized tests. Although standardized tests tend to measure lower cognitive levels of learning, they are still an important measurement of student understanding as these results help us understand at a basic level whether students are learning the suggested material.

Wood and Sellers' (1997) study of students in a problem-based learning environment suggested that this type of alternative assessment approach can result in higher achievement and increased relational interests than traditional textbook learning.

This study examined math performance of elementary students on standardized achievement tests using the Indiana Sequential Test of Educational Progress (ISTEP), with a focus on *Computation* and *Concepts and Applications* and a locally-developed Arithmetic Test, measuring students' computation skills along with students' understanding of arithmetic. This study indicated similar achievement of students taking problem-centered and textbook-taught classes after one year. However, after two years of textbook-based learning versus problem-centered learning, students in problem-centered courses outperformed their peers on both achievement tests (Wood & Sellers. 1997), indicating that student-centered curricular approaches may have an impact on student learning over time.

Such studies show us that student-centered assessment strategies, such as authentic assessment, may not only provide a broader and more useful indication of student learning and individual challenges, in order to instruct and assist students more effectively, but that such approaches can also have an impact on students' achievement when implemented consistently over time. Authentic assessment approaches, however, as a specific form of alternative assessment, present a deeper promise for impacting students' affective engagement and cognitive understanding, since such approaches can link students' classroom learning to real situations and contexts which are useful and relevant beyond the classroom walls, and thus may likely engage students even further.

Empirical review: Alternative measures of student learning. A study conducted by Gulikers, Kester, Kirschner, and Bastiaens (2008) examined the relationship between student perception of assessments authenticity and perception of skill development. Although this study was conducted at the college level, the outcomes

should still be considered relevant to this literature review. Two groups of students, including 81 freshmen and 118 seniors from a Vocational Education and Training college, participated in two authentic assessment tasks and data was collected regarding their experience and perceived learning. These students' perception of authenticity of the tasks was examined using a 24-item questionnaire, which asked the students how similar the activity, physical and social context, format, and criteria resembled "real world" professional practice. Perceived learning was measured using a *Perceived Generic Skill Development* subscale of a *Course Experience Questionnaire*, which measured how students believed the learning activities contributed to development of transferable skills (Gulikers et al., 2008). The outcomes of Gulikers et al.'s study that relate to this review of literature are that both groups (freshmen and seniors) who perceived the activity and the physical setting as more authentic signified a deeper level of studying and a higher level of perceived skill development than the students who characterized the assessments as being less authentic. Perception of enhanced skill development is an important measure of potential student learning beyond traditional testing, since this factor may indicate improvement in understanding not measured by standardized tests alone. However, student perception of skill development is merely a perception. Although there is likely some impetus to such responses, such measurements do not necessarily describe or represent actual cognitive learning.

Another study by McNamar (2009) studied the achievement results of challenged seventh-grade students in four math classrooms who participated in an authentic learning experience in a math classroom. The school in this study served a student population of 98%, with 90% of students receiving free or reduced lunch. The

performance of the participating students had been poor, with the majority of students not turning in class work and not passing their math classes. The classes had been studying surface area in their math classes for several months before the project was initiated. The project for this study involved the use of industrial flooring samples and quotes from a tile company, with the task of having the students prepare bids for flooring the school. Pre-post tests were used to examine student understanding of mathematics, particularly concepts related to surface area and profit, along with the use of a 0-3 scale for analysis of student understanding.

According to pre-post tests, the floor bidding activity resulted in a 31% increase of students who could give an "accurate or nearly accurate description" of surface area and a 28% increase of students who could give an "accurate or nearly accurate description" of profit (McNamar, 2009, p. 144). Similarly, 24% less students answered "I don't know" or left the answers about surface area blank, and 27% less responded similarly regarding profit (McNamar, 2009). It is important to note that while only 41 students turned in the pre-test, 84 students turned in the post-test; this occurrence may have skewed the results. More importantly, the students appeared more attentive, engaged, and eager to share their learning and discuss the project than before (McNamar, 2009). This approach to measuring student achievement, through analysis of open-ended responses from students regarding their learning, allows us to think more deeply about what it is that students are learning and with what accuracy level they are able to articulate that learning. Such measures are an important step beyond the limitation of traditional standardized tests, but would be enhanced with a better description of the levels of learning inherent in student responses, beyond *I don't know*,

some knowledge, *nearly accurate,* and *accurate*. Such descriptions still limit student thinking and analysis of understanding to a duality of responses (White, 2007).

Since authentic assessment appears to increase motivation and interest, and since motivation and interest may enhance student learning, it is assumed that authentic assessment practices may enhance student performance. Furthermore, since alternative assessment appears to enhance test scores, perceived skill development, and accuracy responses to open-ended prompts, it is highly likely that such performance indicators may be representative of the potential impact of authentic assessment practices on higher levels of student understanding. However, this direct link is underexplored and more research should be done to investigate the levels of understanding developed through such means of assessment, particularly at the secondary school level, where drop-out rates inform us of the need for more student-engagement in school. Furthermore, more research should be done on the outcomes of higher-order thinking skills, as discussed in a following section of this literature review.

Challenges of Authentic Assessment

Theoretical review: Challenges of authentic assessment. Although perception of skill development, higher achievement, increased teacher knowledge, and enhanced motivation and interest all appear to be possible outcomes of the implementation of authentic assessment in classrooms, there are still challenges that come along with such practices. More inquiry on the direct relationship between authentic assessment tactics and higher levels of learning should be explored, along with the remaining underlying challenges that educators face with the implementation of such alternative assessment strategies. Public perception, use of time, availability of resources and support, and

implementation and consistency, all continue to pull focus away from the potential and empirically-proven positive outcomes of authentic assessment on motivation student learning. These challenges must be recognized and honored.

Empirical review: Challenges of authentic assessment. ***Fitting within a standards-based paradigm.*** At first glance, authentic assessment may appear to stray from the intentions of the standards-based educational movement, which is supposed to ensure that all kids are working towards demonstration of common certain proficiency indicators throughout the state or nation. Common assumptions regarding standards-based teaching and testing are that concrete skills and facts should be directly taught to students and assessed through traditional means in order to prepare students adequately for meeting standards and the expectations of high-stakes tests. Trepanier-Street et al.'s (2001) study points out from a stakeholder survey about assessment that while teachers deem much value in alternative assessments and less in standardized testing, parents and the general public disagree and generally look to these tests to validate the quality of education students receive. Despite public pressure and increased accountability measures, theorists such as Resnick (1987) and McTighe and Wiggins (2004) argue that traditional standardized tests are not generally appropriate measures of genuine student achievement since they often fail to assess deep conceptual understanding, abilities or dispositions that are actually used within the disciplines.

Given the current state of education and standards-based accountability, teachers and schools must find ways to fit authentic assessment practices within traditional reporting methods (Suurtamm, 2004). Suurtamm's study interviewed five teacher participants at four different secondary schools in Ontario, Canada, throughout one year

to determine the attitudes, practices and concerns of these teachers as they implemented authentic assessment in their mathematics curriculum within traditional school settings. Suurtamm's qualitative study found that these teachers often felt isolated as they attempted to implement authentic math assessments within their traditional settings, and that they would have benefited from more administrative support and enhanced professional development in support of their curricular efforts. This study seeks to bridge the gap between public expectations and teacher educational beliefs by bringing more credibility to authentic assessment practices, demonstrating that there may be value beyond standardized testing outcomes. This implies a need to surface more validity and empirical evidence.

Disconnect between need and practice. Similarly, there is a further need to connect theoretical needs to classroom practices. Although theorists and researchers (Abadiano & Turner, 2003; Avery et al., 2000; Trepanier-Street et al., 2001) indicate that educators support alternative assessment practices which focus on student-centered learning, the value of student-centered practices and authentic learning is not necessarily what drives the assessment and instructional choices teachers make in their classrooms. Implementation of such student-centered practices is not yet the norm, especially within socio-economically disadvantaged classrooms. The lack of alignment between theory and practice may continue to widen the achievement gap between economically challenged students and their more privileged counterparts.

In a study examining teachers' assessments of their students' skills, Stokking, Van der Schaaf, Jaspers, and Erkens (2004) looked closely at the ways teachers implement alternative assessments in various subjects. Specifically, Stokking et al.

surveyed 214 teachers in the Netherlands and judged 50 sets of assessment materials across five disciplines to determine teachers' perspectives on research skills, activities, instruction, and assessment. From the survey results on teachers' assessment practices, Stokking et al. (2004) analyzed the specific goals targeted by teachers. From the assignments reviewed by a panel of professional educators, Stokking et al. determined that most teachers develop their own assessments and/or use textbook examples and that the most important criteria used in selecting assignments (used by 90 to 95% of teachers) were that the assignments were challenging, provided choice, offered a multitude of ways to solve problems, and matched students' cognitive levels. Of concern in this literature review is that the *least* important criteria for teachers in this study (used by 55 to 65% of teachers) was "to provide a realistic context" and "to cover the subject matter" (Stokking et al., p. 104). Alternative assessment, then, may be more easily supported and implemented in classrooms than authentic assessment, as the distinguishing factor between these two terms is the realistic context.

These findings indicate a possible need to help teachers take the value they see in alternative assessment to the next level of impact for our students: bringing meaning and relevance to these assessments. Abadiano and Turner (2003) suggest that most teachers support the theory and use of authentic assessments, yet Stokking et al. (2004) describe that not all teachers connect their assessments to a realistic context, making learning relevant for students' lives and for their success in contexts beyond the classroom. Theory supports the need to develop such practices in the classroom, yet most teachers are not at implementation stage in their classrooms. Based on existing research, there is a need to begin to narrow this gap between theoretical need and practical application.

Of particular concern regarding the disconnect between the theoretical importance and existing practices of the implementation of authentic assessment measures as a means of enhancing students' self-efficacy, interest, cognitive development and critical-thinking skills, is the inequity which may be assumed in the current existence of such practices (Herman, 1997). Lower economic groups and culturally diverse students continue to face with inequities with such curricular opportunities (Darling-Hammond, 1995; Gordon, 1996; Herman & Klein, 1996 as cited by Herman, p. 199). Students in socioeconomically challenged classrooms are most susceptible to the "overemphasis on standardize multiple choice tests… and the 'drill and kill' curriculum that such emphasis often entails, leaving precious little if any opportunity to engage in disciplinary thinking and problem solving," (Herman, 1997, p. 199). Where authentic practices do exist, it seems they do not generally tend to reach the classrooms that may benefit the most. The inequitable distribution of resources and vest practices may therefore further separate students across the socio-economical divide.

Time, resources and classroom management. Teachers interviewed in the studies by Avery et al. (2000) and Suurtamm (2004) suggested other challenges with the implementation of authentic assessment. Student-centered assessment tasks often take more time to implement, and can possibly lead to omission of important content in order to focus on more deep exploration of fewer objectives and the cognitive task at hand (Avery et al., 2000). Suurtamm also found that teachers needed more time to be able to implement authentic assessment in their math courses more effectively. This time, ideally, would include professional development time to establish stakeholders'

understanding of authentic assessment measures in order to develop more support for these methods (Suurtamm, 2004).

This notion of time for development of higher order tasks and teacher collaboration is shifting as educator beliefs in alternative assessment practices also continue to shift. DuFour and Eaker's (1998) suggestions for the development of Professional Learning Communities, where teachers meet regularly in small teams to set learning goals, plan, analyze performance and refine practices, are beginning to become commonly accepted best practices in schools. Such practices are in line with needs for educators to spend more time developing, implement and reflecting on authentic assessment tasks in professional learning teams.

Another challenge according to teachers in the study by Avery et al. (2000) and in Suurtamm's (2004) study is the availability of resources. Student-centered classrooms often lend themselves to the need for more available resources to provide ample exploration of the task. Direct instruction can be much more straightforward and does not always require the same type of creativity, resources, and individualized planning as a student-centered task might. Similarly, in student-centered classrooms where students are all focusing on differing tasks or objectives, this becomes difficult to manage, particularly for new teachers or in schools where teachers face more difficult behavioral challenges. Authentic assessment, then, often requires a highly structured classroom to allow for students to conduct different tasks and may also require a plethora of available resources to further individual student learning. Organization and management of classroom tasks and resources are important challenges to consider.

Implementation and consistency. In addition, there is skepticism regarding the implementation and consistency of authentic assessment tasks, which is not difficult to imagine given the nature and implications of authenticity. Real world challenges are not tidy experiences that can necessarily fit into a black and white multiple choice format and evaluating performance of such challenges is even more complicated. Evaluation of open-ended performance, just as in any job, is always subject to arbitrary decisions and a level of subjectivity. Tanner (2001) and other theorists indicate that grading scales used to measure performance on authentic assessment tasks are often subjective or arbitrary and that they lack reliability and validity. Authentic assessment design, as with the design of any problem-based assessment, can often be weak and unreliably connect objectives and/or grading criteria to explicit standards. Tanner also points out that little study has been conducted with regards to judgment reliability for authentic assessment measures.

From this same notion of skepticism, a five-year research project was conducted at the University of Virginia to develop authentic assessment tasks for a middle school setting in various content areas and to measure the scoring reliability of associated objectives (Moon et al., 2005). The assessments were reviewed by 46 experts to determine the validity of the content and inter-rater reliability of grading. The implementation of the authentic assessment tasks was favorable to students and teachers overall and grading reliability appeared to be consistent. Moon et al.'s (2005) study suggests that authentic assessments can actually provide consistent and quantifiable information about student learning and that such assessment can measure students' attainment of academic standards. The study also implies that grading reliability and

validity can exist with authentic assessment if done right, but that expert collaboration for authentic assessment task development may be an important factor in its success. Additionally, grading criteria should be linked explicitly to the development of higher-order thinking skills and varying levels of understanding, as described below.

Theoretical Review of Understanding

Transferability and enduring understanding. Since it is assumed that the lasting motivation or "hold" factor, which we would hope for our students to obtain, is not developed through extrinsic means it is important to look beyond grades and test scores to identify learning outcomes. Furthermore, modern educational theories remind us that grades and scores do not necessarily inform us whether a student can perform outside of academic settings or not. As Resnick (1987) points out, traditional teacher tests generally only call for a narrow range of cognitive skills, which are often disconnected from experiences that students will encounter in life. To prepare our adolescent students for the world they will face beyond high school, we need to teach them to be critical thinkers that are able to apply learning to new contexts, disciplines, and settings (Bransford et al., 2000; Wiggins & McTighe, 2005). We need to seek student performance beyond "response to narrow prompts" (Wiggins & McTighe, p. 78) and towards "fluent and effective performance in the world" (p. 78). As such, it is important to distinguish learning, as demonstrated by achievement on standardized tests, from what Wiggins and McTighe call *enduring understanding*. Enduring understanding of concepts is the ability to transfer knowledge and skills to new settings, which "involves the capacity to take what we know and use it creatively, flexibly, fluently, in different settings or problems, on our own" (Wiggins & McTighe, 2005, p. 40).

Wiggins and McTighe (2005) also describe varying levels of understanding, including facts and skills (most basic level), *big ideas* (slightly more complex) and principles and generalizations (desired level of understanding). The difference between each level of understanding is the idea of *transferability*, which implies that students are able to transfer knowledge, skills and big ideas to different contexts, situations and disciplines (Wiggins & McTighe, 2005). Facts are simple and declarative, requiring memorization, and skills are procedural, requiring the discrete ability to do something in isolation. Facts and skills do not really require *transfer* to different contexts, situations, and disciplines. Big ideas require understanding of concepts and processes, which are declarative and procedural in nature, but require abstract constructs or a combination of skills to reach results. Big ideas require some transfer to topics, contexts, or disciplines, indicating a higher level of understanding. The highest level of enduring understanding is the ability to transfer knowledge about principles and generalizations, which requires abstractions, the linking two or more concepts, and transferability.

Cognitive hierarchy of understanding. *Bloom's taxonomy.* From a similar approach to describing a hierarchy of understanding and abilities, Bloom (1956) described six cognitive domain levels for classifying educational objectives, including: *knowledge*, *understanding*, *application*, *analysis*, *synthesis,* and *evaluation* (see descriptions in Table 1). The first two categories of Bloom's scale (knowledge and comprehension) do not require critical-thinking, whereas the last four classifications (application, analysis, synthesis, and evaluation) require critical thinking or higher-order thinking (Bissell & Lemons, 2006). Bloom's (1956) cognitive levels of thinking are commonly referred to in education, and describe the same basic concept which Wiggins

and McTighe (2005) explore, that students should be prompted to think beyond knowledge and comprehension. Bloom described that 95% of test questions require students to think at only the lowest cognitive level, despite the value of teaching students how to think at higher ability levels. This implies an inherent need to rethink assessment design and the development of effective educational objectives.

Table 1.

Bloom's Taxonomy: Proving Behaviors

Cognitive Level	Description
Knowledge	Recalling facts and information
Comprehension	Showing understanding of acquired knowledge
Application	Adapting/applying known information
Analysis	Breaking material down into component parts
Synthesis	Putting information together in a new way
Evaluation	Judging the outcome

Bloom's description of *application* is similar to Wiggins and McTighe's (2005) concept of *transferability*, yet Bloom would recommend that students take their learning a few steps beyond application in order to critically analyze and evaluate their learning and to synthesize or create new ideas based on the acquired concepts and processes. This is also the type of learning that the principles of *Understanding By Design* (Wiggins & McTighe, 2005) encourage teachers to instruct and assess. In teaching our students not only to be able to transfer or apply learning to other problems, contexts, settings, or disciplines, but to also judge and self-assessment the work they do and justify and

critique their performance, we are teaching them to think from a more complex cognitive thought process and with a deeper understanding of the material they interact with. Such thought processes are what teach students to think critically, problem-solve, and evaluate learning from multiple perspectives, preparing graduates to be more fluent and effective contributors to the workforce, who bring creativity and innovation.

Although Bloom (1956) has been probably the most referred to author on the cognitive domain for several decades, Bloom's original taxonomy is often criticized for its simplicity. There have been many revisions of Bloom's taxonomy, most notably and recently by Anderson and Krathwohl (2001) and Marzano and Kendall (2007). Many of revisions of this hierarchical taxonomy are more complex and involve multiple dimensions.

The revised taxonomy table. Anderson and Krathwohl's (2001) revision of Bloom's taxonomy (1956) is two dimensional and includes a knowledge domain and a cognitive process domain. Anderson and Krathwohl propose that to properly classify educational objectives, both dimensions must be analyzed. The knowledge domain includes *factual*, *conceptual*, *procedural*, and *meta-cognitive* knowledge. This domain includes increasing recognition from cognitive theorists that with development, students become more intimately knowledgeable of and accountable for their own thinking and that, upon acting on this reflective knowledge, they tend to learn better (Pintrich, 2002). The cognitive process domain in Anderson and Krathwohl's (2001) scale moves along a continuum similar to Bloom's taxonomy: *remember*, *understand*, *apply*, *analyze*, *evaluate*, and *create*.

Table 2.

The Taxonomy Table (Anderson & Krathwohl, 2001)

The Knowledge Dimension	The Cognitive Process Dimension					
	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual Knowledge						
Conceptual Knowledge						
Procedural Knowledge						
Meta-Cognitive Knowledge						

The new taxonomy. Marzano and Kendall's *New Taxonomy of Educational Objectives* (2007) similarly proposes a two-dimensional analysis for analyzing educational objectives (see Table 3). Marzano and Kendall (2007) discuss problems with the original taxonomy, mainly the vast oversimplification proposed to represent thinking of learning relationships. Marzano and Kendall's *New Taxonomy* describes six hierarchical *levels of processing,* including (from lowest to highest): *retrieval, comprehension, analysis, knowledge, utilization, meta-cognitive system,* and *self-system.* The first four levels are part of a cognitive system. The other dimension for analysis in this revised taxonomy, similar to the arrangement presented by Anderson and Krathwohl (2001), is the knowledge domain. Marzano and Kendall's knowledge domain includes: *information, mental,* and *psychomotor*. Literature from both taxonomy revisions bring

light to an important notion, regarding the complex nature of classifying educational objectives.

Table 3.

The New Taxonomy (Marzano & Kendall, 2007)

	Knowledge Domains		
Levels of Processing	Information	Mental Procedures	Psychomotor Procedures
Level 6: Self-system			
Level 5: Metacognitive System			
Level 4: Knowledge Utilization			
Level 3: Analysis			
Level 2: Comprehension			
Level 1: Retrieval			

Aside from classification, it is important to recognize the complex and imperfect hierarchy of educational objectives and understanding. Marzano and Kendall (2007), in reference to Rohwer and Sloane (1994) point out that taxonomical classification may resemble a hierarchy, more than actually being a perfect hierarchy. In different cases, it may be difficult to distinguish a hierarchical level of difficulty between *evaluation* and *synthesis*, for example. Despite the challenges of imperfection with any classification system for understanding learning and understanding, Bloom's taxonomy (1956) brought forth an expanded notion of non-unidimensional levels of thinking and

evaluating. This concept brought forth an important notion for the design of curriculum, objectives, and methods of evaluation.

Hierarchical evaluation methods. Similar to Bloom's taxonomy (1956) for educational objectives, White (2007) cites Perry's (1968) description of learning along a continuum, with more advanced levels of understanding demonstrating that students understand the complexity of given concepts, and that there may be more than one truth to a particular assertion. A simpler understanding of material would mean that students believe that problems have a *duality* of answers, or just a right and wrong response (White, 2007). Most importantly, as White points out, Bloom (1956) explains that to provide instruction at different maturation levels of understanding teachers must use varied instructional techniques. White ascribes that to evaluate knowledge or understanding at these different cognitive stages, educators must use different methods of evaluation. This concept implies the need for assessment and evaluation beyond knowledge and comprehension, which is an underlying principle for the design of this study.

The described continuum of learning levels, as cited by White (2007), moves from recitation, to comprehension, to application, to synthesis, to evaluation. Lecture and assigned readings, followed by repetition of material, are a way to assess student *recitation*, but do not necessarily inform us of actual *comprehension*. Comprehension may be revealed through a discussion of new material in students' own language, whereas this does not inform us if students are able to *apply* or *transfer* (Wiggins & McTighe, 2005) this new material to other contexts, disciplines, or settings. *Application* signifies ability to use concepts in meaningful manner to determine whether something

works or not, for example. Students at a more advanced level would be able to *synthesize* their learning into new models or theories. White believes that the synthesis level is not really expected from students until they are engaged in graduate level university work.

Unfortunately, as one educational theorist describes, "The more complex and sophisticated the cognitive function, the more difficult it is to measure," ("The Intelligence We Never Measure," 2007, p. 83). Complex assessment is not always practical. The challenges of examining student performance on authentic assessment tasks will be described in the following section of this review. Most empirical studies of the influences of alternative assessment, then, measure outcomes such as (a) teacher response or perspectives, (b) student interest or motivation, or (c) student learning, as indicated by grades or test scores. These variables are important factors to measure, and should continue to be investigated, especially as related to the description of authentic assessment defined for this study. However, there is little research which attempts to measure student learning in terms of outcomes beyond standardized test scores, including transferability, the development of enduring understandings, or analysis of multiple levels of cognitive thinking. Some empirical research has been conducted to determine student perceptions of authentic assessment tasks and their influence on learning, but little research exists to exemplify the various levels of learning and complex cognitive achievement.

Empirical Review: Assessment and Cognitive Level

A few studies have attempted to use cognition taxonomies, particularly Bloom's taxonomy (1956) and Anderson and Krathwohl's (2001) Revised Taxonomy to measure

student approach, performance, and scaled cognitive difficulty.

Bittel and Hernandez (2006) conducted an interesting study examining student selection of differentiated assessment tasks after linking grades to various levels of Bloom's taxonomy on a final astronomy project. Eighth grade students in science classes in Tucson, Arizona, were told they would get an A for completing specific assignments fitting into the *analysis, synthesis, or evaluation* levels of the taxonomy, a B for completing assignments in the *application* level, a C for the *comprehension* level, and a D for the *knowledge* level. When students had selection choice regarding projects tied to letter grades, 98% selected projects requiring higher-thinking skills (beyond knowledge and comprehension). Furthermore, even though only 80% of students earned an A or a B, 90% of students demonstrated the ability to perform at the higher thinking levels (the remaining 10% failed to complete all of their work). This study is a creative way to measure student performance as a consideration of the cognitive hierarchy of learning and understanding.

Another study in a college-level physics program in Canada was designed to determine correlations between (a) the approach to learning by college students using a "Study Process Questionnaire," (b) the intellectual demand of physics exams on a scale derived from Bloom's taxonomy, and (c) the pre-post performance of students on related a concept-based test called the "Force Concept Inventory" at the beginning and end of the academic year (Dickie, 2003). The study found a clear correlation between intellectual demands of exams, student performance, and the approach to learning (measured by the "Study Process Questionnaire as *Surface*, *Deep*, or *Achieving*). A Deep Approach to learning involves application of principles to real life situations, instead of

learning to memorize and apply formulas (the Surface Approach). Most notably, the lower cognitive demand exams in this study resulted in increased usage of Surface Approach methods among the college students, indicating that students' approach to learning is directly related to the intellectual demand of assessment emphasized.

The proven notion that evaluation choices influence student approaches to learning and can create or stifle higher-order thinking, makes it increasingly more important to design assessment experiences in a manner that intentionally engage higher-order thinking skills. The relationship between interdisciplinary authentic assessment and the impact on levels of understanding has important implications which should continue to be explored.

Need for Further Study

Overall, there is little research measuring specific cognitive expectations and outcomes that occur as a result of authentic assessment at the high school level, particularly following an interdisciplinary teaching experience. The research that does exist is primarily limited to outcomes such as teacher perspectives, motivational outcomes, and standardized test scores, without much exploration of higher-order cognitive outcomes. Furthermore, where there is empirical research, the results appear to be limited in scope and focus very narrowly on specific assessments or programs, without capturing the core essence of what it means to explore a meaningful and relevant authentic task within an integrated or interdisciplinary context. A broader scope of study of the impacts of authentic assessment at the secondary level will help move from the theoretical realm into the practical and to develop a better understanding of the potential outcomes and shortcomings of authentic assessment practices. The analysis

may also serve to assist other schools, teachers, and education leaders at the secondary school level seeking practical guidance regarding the implementation of authentic interdisciplinary assessments.

Summary

More research needs to be done to clarify the relationship between interdisciplinary authentic assessment, cognitive learning levels, and student performance at the secondary school level. Instructional practices are in the midst of changing from direct instruction, lecture and textbooks to student-centered classrooms with activities such as authentic assessment that serve empower students to problem solve and interact more with the materials and their own skills. All the while, standards-based learning and high-stakes testing have become increasingly prevalent and these trends will likely not change any time soon. In today's world of data and measurement, the impacts of authentic assessment on student engagement, and ultimately on student learning and higher-order thinking, are important to investigate and articulate as we shift into a new era of consciousness and understanding about what motivates students to be intrinsically interested in school and what truly has a lasting and deep impact on cognitive understanding. Recent educational theories point in the direction of authentic assessment, but little research answers the pressing questions about how high school students respond to and learn from cognitively complex assessment tasks. Measuring cognitive difficulty of such tasks and impacts on levels of performance of various interdisciplinary authentic assessment tasks will help lead us closer to the practical and measureable understanding that we need to make learning more meaningful and lasting for our youth.

Chapter 3: Methods

Research Design and Rationale

Overall research design. Creswell (2009) provides a framework to align research design and methods to the beliefs or worldviews which drive the purpose of a study. Creswell describes a "postpositivist worldview" which recognizes that "we cannot be 'positive' about our claims of knowledge when studying the behavior and actions of humans" (Creswell, 2009, p. 7). Within this worldview, causes are believed to determine outcomes, though absolute certainty is not considered possible. Therefore, postpositivists seek to scientifically describe the relationships between causes and outcomes, to reduce ideas into variables, and to test or verify theories. Postpositivists realize that "absolute truth can never be found" and that "evidence established in research is always imperfect and fallible," but that "data, evidence and rational considerations shape knowledge" (Creswell, 2009, p. 7). The research methodology in this study was developed to encompass the philosophies of this worldview.

This study sought to explore the intended cognitive behaviors and student outcomes resulting from four interdisciplinary authentic assessments at the high school level. The table in Appendix F displays an overview of each of the four interdisciplinary authentic assessments analyzed in this study and Appendices G, H, I, and J reveal more detailed descriptions of and rubrics for each task. From the postpositivist approach, a quantitative research design strategy was most appropriate for this empirical study (Creswell, 2009). It would have been ethically and procedurally difficult to use control and experiment group comparisons to test the outcomes of authentic assessment practices in this particular study. It would also have proved challenging to adequately measure

complex cognitive understanding using pre- and post- group analysis. Cognitive performance was measured through the execution of specific complex assessments, requiring input from multiple disciplines, public performance, and, in some cases, several weeks or months of preparation. The complex and time-consuming demand of these authentic assessments made comparing this complex performance both before and after the interdisciplinary unit a uniquely difficult and unrealistic measurement. Therefore, non-experimental design strategies were most suitable for this research project.

This study, then, is a two-pronged quantitative, non-experimental design analysis of (a) the cognitive levels of Bloom's Taxonomy which exist within interdisciplinary authentic assessments, following an ongoing professional development intervention, and (b) how students perform on each cognitive level of Bloom's Taxonomy within these tasks. Both components of this study serve as critical stages of analysis for generalizing the cognitive expectations and related learning outcomes which may result from interdisciplinary authentic assessments.

Part one: expected cognitive levels. The first part of this study examines the cognitive levels within interdisciplinary authentic assessments, following an ongoing professional development intervention. Analyzing cognitive levels expected of our students, as indicated through assignment prompts and rubrics, helps identify the cognitive range of abilities and understandings which teachers expect students to demonstrate during an interdisciplinary authentic assessment period. Theoretically, authentic assessment tends to engage students at higher cognitive levels than traditional assessment forms (Bransford et al., 2000; Drake & Burns, 2004; Erickson, 2002; Tchudi & Lafer, 1996; Wiggins & McTighe, 2005). Analysis of cognitive expectations will help

surface theoretical assumptions regarding the rigor and complexity of such assessments.

The interdisciplinary authentic assessments used for analysis (see Appendices G, H, I, J) were developed by grade-level teaching teams, following the intervention of an ongoing professional development experience, and were executed during a month-long authentic assessment term. The professional development sequence included ongoing training about authentic assessment and scaling of learning objectives, based on interdisciplinary, concept-based, and cognitive learning principles described by Erickson (2002), Marzano (2006), Tchudi and Lafer (1996), and Wiggins and McTighe (2005), along with structured collaborative planning time for grade level teams of teachers. Specifically, teachers participated in professional development in-services on the use and challenges of authentic assessment along with collaborative workshops designed to practice scaling objectives in terms of learning levels (Marzano, 2006). This professional development work was part of an ongoing sequence of development at this school site over the last several years. Throughout this intervention, teachers worked in professional learning community teams (Dufour & Eaker, 1998) to create interdisciplinary units, culminating authentic assessments, interdisciplinary unit exams, and associated rubrics, based on the overlap of conceptual themes and big ideas from each discipline (Erickson, 2002; Wiggins & McTighe, 2005). The culminating interdisciplinary authentic assessments were designed for presentation within a community forum setting, where community members, parents, educators, school partners, and students across grade levels engaged in critical dialogue with students about their work.

Theoretically, this type of public format for student presentation of interdisciplinary work engages students in authentic and cognitively complex learning,

but cognitive complexity is merely an assumed outcome of these assessments and has not necessarily been measured or analyzed. This deeper analysis serves to help determine whether interdisciplinary collaboration contributes to the development of cognitively complex authentic assessments.

The data analyzed in Part One is cross-sectional, providing a snapshot analysis of several authentic assessments used at one point during the year, when each student was to participate in an interdisciplinary authentic assessment for a collaborative project at each grade level (9th, 10th, 11th, and 12th). Panel analysis of assessment prompts and rubrics by a team of experts was used to analyze the level of cognitive expectation in terms of Bloom's Taxonomy (knowledge, comprehension, application, analysis, synthesis, and evaluation) for Part One of the study. Assessment descriptions and rubrics were the major entities of analysis for Part One of this study, since they were the primary sources used to draw conclusions regarding cognitive expectations for students on the interdisciplinary authentic assessments.

To increase reliability, teams of three external panelists (described below) ranked the objectives from each assessment into categories of Bloom's Taxonomy. Although there are more recent revisions to the taxonomy (Anderson & Krathwohl, 2001; Marzano & Kendall, 2007), the original taxonomy still appears to be the most commonly referred to and understood cognitive taxonomy throughout education communities. The original version was chosen for this study with the intention that this tool would be more straightforward to use, as it is one-dimensional and tends to be more commonly referred to. Ease of use and common understanding was hoped to increase inter-rater reliability of the ranking of objectives by independent panel members.

Individual panelists first categorized each objective on their own and then compared rankings. When panel agreement did not occur initially, panel members discussed reasoning, and referred to supporting documents for more information, to come to a consensus on placing each objective into the varying cognitive categories. Although methodology allowed the option for panelists to categorize objectives into multiple levels within the taxonomy, should they not agree on placement, the three panel members were able to reach consensus on each objective following discussion.

A three-member panel was chosen for this study to establish more reliability and consistency with ranking. Two-member rater agreement, with a third panel member available for arbitration when there is disagreement, is a practice commonly used to establish more consistency and reliability in scoring practices (Oregon Department of Education, 2011). Although this study does not examine inter-rater reliability of the ranking process, the researcher noted the frequency of panel agreement on placement prior discussion, to highlight any apparent issues or trends with regard to reliability.

Part two: performance outcomes. Part Two of this study serves to analyze student performance on varying cognitive expectations. This component of the research attempts to better understand how students respond when asked to think and perform using varying levels of cognition. Where Part One of this study seeks to note cognitive complexity of assessment expectations, Part Two seeks to articulate how well students meet objectives of varying cognitively complexity. The average results of student performance on the various cognitive levels, as measured by teacher-created rubrics, was measured and reported. In Part Two, cross-sectional data on student performance was

collected from one interdisciplinary authentic assessment term across four grade levels (9th, 10th, 11th, and 12th).

Human perception of student performance on learning objectives was the major entity of study for Part Two of this study. Although rubrics exist to measure student performance, there is still subjectivity in the creation of the rubrics and in making judgment of student performance on specific learning objectives and grading on the rubrics. To increase reliability, graded assessments and student scores were only selected for use in this study if they were team graded by two or more teachers. The assessments analyzed in this study were all graded by teams of two or more teachers who taught the students directly during this assessment period.

Sampling Methods, Sample, Participants

Sampling methods. Since this school boasts authentic assessment and interdisciplinary teaching as best practices, and actually implements these practices at least once each year, this school appeared on the surface to be an appropriate school setting at which to study the effects of authentic, interdisciplinary learning experiences. The mid-year interdisciplinary units at this school, designed by six teachers, including all core disciplines, and resulting in authentic performance tasks, appeared to be relatively unique to this specific learning environment. The variables surrounding the design of these units fit within the operational definitions of authentic assessment and interdisciplinary teaching described within this study. Additionally, interdisciplinary teaching units at schools examined for potential study sites were not necessarily as diversely represented by the various disciplines and/or tended to be taught around thematic concepts, but not necessarily integrated throughout from planning stages, to

instruction, to performance assessment, and to grading practices.

After obtaining teacher consent, the four authentic assessments developed for 9th, 10th, 11th, and 12th grade classes at this school (see Appendices G, H, I, and J), and administered during a four week interdisciplinary term, were approved for this study by the external three-member panel described below as meeting the criteria for authentic assessment developed for this study (see Appendix B). To maximize trustworthiness of assessment design, the researcher ensured that each grade level team of teachers creating the assessments used for study included at least three teachers with more than five years of teaching experience and at least three teachers with two previous years of participation in interdisciplinary unit planning at this school site. Each grade-level teaching team consisted of six teachers, who each taught a different class on that grade level as part of the interdisciplinary unit. All teachers participated in an ongoing professional development intervention on interdisciplinary planning and authentic assessment.

For Part One of the study, the expert panel classified learning objectives for each assessment into expected cognitive levels according to Bloom's Taxonomy. The criteria for expert panel members included: (a) practitioner teacher-leaders, administrators, or other educational leadership contributors with specific interest and experience in assessment design, interdisciplinary teaching, and/or authentic assessment and (b) current practice in a coaching and/or professional development capacity, with at least five years of experience in this field. To classify the objectives into cognitive domain levels, each objective listed within the four summative authentic assessment prompts and/or rubrics was first classified by each panelist individually using the Proving Behaviors tool (Appendix A). Following individual placement, panel members checked for agreement

on ranking. Where there was disagreement in placement on Bloom's Taxonomy, panelists discussed their perspectives, referred to supporting documents, and came to consensus on placement. A three-member panel was used to increase reliability should disagreement occur. Two-member rater agreement, with a third panel member available for arbitration, is commonly used in ranking and scoring practices to establish more consistency and reliability (Oregon Department of Education, 2011). Since it is possible that a learning objective could require various levels of cognitive thinking, the panel had the option of placing objectives in more than one level on the taxonomy. However, the panel was able to reach consensus through discussion to place each learning objective into only one category of cognition.

For Part Two of the study, the researcher provided teachers with the lists of participants for each grade level and collected rubric scores, graded by the teachers, on each objective within the assessments for these students. Rubric scores were placed into spreadsheets for summative analysis of scores according to each objective and according to expected cognitive level, as determined by the panel of external evaluators.

Sample. ***Part one: expected cognitive levels.*** Since all teachers from each grade level were involved in the design of the interdisciplinary authentic assessments (see Appendices G, H, I, and J), each teacher from this school was asked to participate indirectly in this study. At the time consent forms were distributed, the teams had already developed and implemented the assessments for their interdisciplinary units. One hundred percent of the teacher population provided consent for this study.

At this school, teachers work on mixed-experienced grade level teams, according to subject and grade level. The teams who created and graded the interdisciplinary

teaching units (i.e. the assessment prompts and rubrics) each consisted of six teachers from different disciplines, with a varying range of teaching experience. Each team consisted of: (a) an English teacher, (b) a science teacher, (c) one or two math teachers, (d) a history teacher or elective teacher, and other potential elective teachers. The population of teachers at this school during the interdisciplinary term was diverse both in experience and other demographics. The teaching staff consisted of 13% African-American, 29% Hispanic or Latino, and 58% Caucasian teachers. 46% of the teaching population was female and 54% was male. 42% of these teachers had been teaching for less than five years and 58% had been teaching for five or more years. These extraneous teacher demographics are informational and will not be considered for analysis in this study.

Part two: performance outcomes. The student population considered for analysis of performance results in Part Two of the study consisted of the entire study body enrolled at this school during the interdisciplinary term. Census sampling of the 472 students was used, so as to capture the wide range of cognitive performance results that might have otherwise be overlooked with a smaller sample size. Student performance in this study is analyzed by grade level, since each grade level of students received a different instructional unit and a separate authentic assessment task.

Participants and handling of non-response. ***Part one: expected cognitive levels.*** Since there are only four assessments being selected for examination in this study, it was ideal to have full participation from each grade level teaching team. Participation in this study was voluntary, so teachers were given the right not to participate. Since team collaboration for interdisciplinary planning, sharing and critique of work, and analysis of

student performance data is a regular occurrence during professional development meetings at this school, it is not surprising to have received consent for full participation among the teaching staff.

Part two: performance outcomes. The student participants in Part Two of the study completed the interdisciplinary authentic assessments for their regular school coursework. This study did not require students to perform outside of their already-occurring curriculum and did not present students with any additional responsibilities. Therefore, most students completed the assessments. There was occasional non-performance due to absences, failure to complete necessary work, or other potential interruptions. Since participation in this study was voluntary, there were also students or parents who did not respond to requests for assent or consent, or who responded negatively to such requests for participation in this study. Absent students had been encouraged by teaching staff to schedule make-up performances of their missed work. Only some of these occurrences were successful. The researcher also sent home notices to parents with students and/or called non-responder families to discuss the research project and answer questions. The attempts to maximize the number of student participants were important to ensuring the census sample was as representative as possible of the overall population of students.

The overall student population at the time of the assessments consisted of 131 9th grader students, 128 10th grader students, 112 11th grader students, and 101 12th grader students. Of this population, a percentage of students were unable to complete or make up the interdisciplinary assessments, had checked out of the school by the time consent and assent forms were distributed, or did not receive parental consent or provide assent to

participate in the study. With these factors considered, the total population, sample size, and representative percent of the population are described in Table 4 below, indicating 76% participation for 9th graders, 87% participation for 10th graders, 86% participation for 11th graders, 90% participation for 12th graders, and an 85% participation rate overall. Speculation as to the differences in participation throughout the grade levels was made, but is only anecdotal. Possible reasons for a lower participation rate among 9th graders, compared to other grade levels, include lower stability in enrollment for 9th graders between assessment period and study period, less developed relationships between school and families and potential impact on trust for the research project, and more fear among students regarding public presentation of their work. Potential reasons for a higher response rate among 12th graders could include the number of students who were age 18 or over and did not need parental consent for participation. Again, these assumptions are only speculative.

Table 4.

Population, Sample Size, and Percent Participation

Grade Level	Total Population	Participants	Percent Participation
9th Grade	131	99	76%
10th Grade	128	111	87%
11th Grade	112	96	86%
12th Grade	101	91	90%
Total	472	397	85%

The total population of students during the term of study consisted of 45% male and 55% female students. The overall ethnicity of the student body at that time was 63% Latino or Hispanic, 33% African-American, 2% Asian, 4% other ethnicities. During this term, 33 of the population were classified as English Language Learners and 28 students were classified as special education students. Demographics of the student body only provide extraneous details of the subjects for this study. Since student performance data was analyzed anonymously, subgroup performance was not a focus of analysis in this particular study. However, to ensure the census group was representative of the overall population of students, the demographics of the two groups were compared. Table 5 below displays very similar demographics between the participant group and the total population in each category, implying overall adequacy of the participant group as a representative sample of the larger population:

Table 5.

Demographics of Participants Versus Total Population

Subgroup Populations	Participants	Total Population
Hispanic or Latino	74%	75%
African-American	13%	12%
Asian, Filipino, Pacific Islander	7%	6%
White (not of Hispanic origin)	5%	6%
Female	55%	54%
Male	45%	46%
Special Education	6%	4%

(continued)

Demographics of Participants Versus Total Population (continued)

Subgroup Populations	Participants	Total Population
English Learner	7%	8%
Socioeconomically Disadvantaged*	77%	76%

*percent of students eligible for free or reduced federal lunch program

Human Subjects

Informed assent and consent. ***Part one: Expected cognitive levels.*** Although teachers did not participate in this study beyond their normally assigned duties, including the design and implementation of the interdisciplinary assessments, full, informed consent from teachers was obtained for the assessment analysis in Part One of the study. Informed consent forms (see Appendix K) were given to teachers during an all-staff meeting before any research by the external three-member panel or the primary investigator took place. Notice was given to all twenty-four teachers as to the nature of the research and how each participant was to be included in the study (see Appendix K). Notice included the purpose of the research, the expectations for participation, the time period when the research would occur, and the methods which would be used to conduct the research, along with information about the inherent benefits and minimal risks of the study. A sample of the Proving Behaviors tool (see Appendix A) which was used to identify levels of cognitive assignments within this study was also shown to the teachers, along with a description of how student performance would be analyzed in Part Two of the study. Teachers were not asked to for consent of the study until after their assessments had been created, implemented, and graded, to ensure the assessments were not designed or graded differently due to knowledge of the research.

Following teacher consent for the study, three expert panel members (who were not teachers at this school, yet possessed professional expertise in the design and execution of interdisciplinary curriculum and/or authentic assessment) were invited via email, phone call, or personal conversation to join the study. After a meeting was set up with the group of three external panelists, full, informed panel member consent was obtained, after the researcher reviewed the study, protocol, timeline, and potential benefits and risks. After research for Part One of the study was completed, the panel was provided with gift cards of fifty dollars each for their participation.

Part two: Performance outcomes. For Part Two of the study, informed consent forms for parents (see Appendix L) were distributed to all 447 students enrolled at the time the research began to take place. The notice given to parents included information about the nature and purpose of the study, how each participant was to be included, the time period when research would occur, and the methods to be used to conduct the research, along with information about the benefits and minimal risks of the study, as described in the following section. It was hoped that risk would be minimized by notifying students and parents of the study after the assessments were completed, since students performed as they would regularly be expected to for school purposes. The consent forms indicated that the students would be participating in a voluntary study, if the students and parents choose to accept, and that public results from the study would be analyzed for the entire grade level of students, as opposed to individually. Consent forms were returned into the school office and/or to teachers for collection, who submitted them to the researcher. The researcher kept track of responses and followed up with non-response families to ensure maximum participation. Where there was non-response with

consent forms, follow up calls and/or meetings with parents were used to determine the reason for the non-response and to obtain consent directly if the students and parents chose to accept. Assent forms (see Appendix M) were distributed to students directly following parent consent, with the same explanation of benefits, potential risks, confidentiality, and the voluntary nature of the study, and the same collection procedures.

Risk minimization and benefit maximization. All participants were informed about the nature of the study, the reason the researcher is interested in interdisciplinary authentic assessment, and the impact the research may have on the field of education, and consequently on societal understanding of assessment and cognitive performance in the long run. Participants were informed of potential risks, to be described below.

The researcher held a meeting with teachers participating in the assessment implementation to distribute and complete consent forms directly and personally, and to explain the minimal risk and maximum benefits of the study. Minimal risks to teachers include their perceptions of being judged by the researcher, the school or others by their assessment implementation and student performance. To help minimize risks, assessments analyzed in this study were created by a team of teachers, not one individually, which may help prevent obvious connections between results and one particular teacher. Teachers were informed that the risk of participation is not of a degree greater than that which would normally be expected during professional development workshops at their school, through the focus on student performance analysis, other instructional data analysis, and assessment revision protocols, which occur regularly in collaborative staff meetings.

The consent forms distributed to and collected from parents also included explanation regarding potential benefits the study and of the minimal risks of participation for students. Minimal risks associated with this study could include possible discomfort among students from knowing they are part of a research project. The forms explained that the risks of participation are not of a degree greater than those which would be normally be experienced in a classroom setting, since students participated in the assessment experience anyway, as a component of their regular coursework, especially since the students were not asked to be a part of a study until after their performance on the assessment and since school and student names would not be used in the study. Following parental consent, students were given assent forms that outline the same details, including maximization of benefits of the study and minimization of risks. The researcher's contact information was included, for follow-up with any participating members or parents who had additional questions about the purpose, benefits, and/or risks of the study. Several student, parent, and teacher participants contacted the researcher to ask additional questions about the study.

Confidentiality. Teachers' names and the school name are intentionally not referenced in this study. Although it is not possible for teacher participation to remain completely anonymous, team participation in the creation of assessments will likely help prevent judgment regarding assessment design from being linked to any teacher individually. Names are not used in the study and teachers will not be judged or evaluated on their work or on the performance of students.

Full confidentiality was employed in the handling of student responses. There was no reason for the researcher to collect or indicate names of students directly in this study.

Although anonymity is not possible, since teachers assigned grades for grade book purposes, the researcher analyzed rubrics and scores without student names.

Assent and consent forms were explained to and distributed to students, parents, and teachers through direct communication by the researcher (teachers), through direct communication from a teacher participant in this study (students), or through a written letter sent home with students (parents). A contact telephone number and email address for the researcher was provided for any questions related to the study. Student responses were tracked by the researcher and individual follow up calls were made by the researcher to remind parents and students about the forms and further explain the nature and intent of the research, and the risks and benefits associated with the study, if the forms were not received promptly.

Data Collection, Setting and Procedures

Written approval of this research was obtained from the Executive Director of the school via email (see Appendix P) and approval from Pepperdine University's Institutional Review Board (IRB) was confirmed before all participants were accessed for this study. Assessment information and census sampling of student performance data for this study were collected from teachers at the school site in Los Angeles County, California, after the researcher confirmed the list of participating students. The expert panel meeting for this study took place at a separate location from the school, outside of the school day.

Instrumentation

Part one: Expected cognitive levels. The four authentic assessments were reviewed by an external three-member expert panel using Bloom's Taxonomy: Proving

Behaviors. Each member of the three-person panel reviewed the assessment objectives, prompts, rubric and any other supporting documents, and classified the objectives using the Proving Behaviors tool (see Appendix A). Bloom's Taxonomy (1956) was selected as a guiding framework for analysis in Part One of this study since, although revised models (Anderson & Krathwohl, 2001; Marzano & Kendall, 2007) are well-justified through literature and have strong and important rationale to support them, the original taxonomy appears to be the most commonly referred to and understood cognitive taxonomy in education. The simplicity and common use of Bloom's Taxonomy was expected to enhance ease of use of the related tool for panel members and to therefore increase inner-rater reliability. The Proving Behaviors tool (see Appendix A) was specifically selected for use in this study, since it provides a guiding list of behaviors which help make cognitive classification more straightforward than a more theoretical description of the taxonomy. Each of the four assessments in this study were analyzed and recorded using a separate table which lists the objectives of the assessment in the first column, followed by columns to categorize the expected cognitive level of each objective (see Tables 7, 8, 9, and 10).

Prior to use of this tool for the study, the panelists were re-familiarized with the Proving Behaviors tool and related literature on Bloom's Taxonomy. The panel practiced use of the Proving Behaviors tool by applying its use first for a set of other interdisciplinary authentic assessments, prior to analysis of the assessments for this study. This processed helped the panel normalize their use of the tool, to develop common understanding assessment design and cognitive performance levels, and to test the validity and usability of the tool. This work also helped test the tool for appropriateness

and reliability of use prior to implementation of the study. When using the tool, panelists first reviewed and placed objectives and activities on the classification table on their own, according to the six cognitive levels of Bloom's Taxonomy; then they compared answers. In most cases two panel members agreed after initial placement; in some cases, the three panel members discussed reasoning and used supporting documents to reach consensus on panel placement. The three-member panel was used to help establish reliability and consistency in ranking (Oregon Department of Education, 2011). Supporting documents provided by the teachers included daily lesson plans, formative assessments, student work, and other curricular materials; these documents helped clarify the type of work the students were asked to perform.

Part two: Performance outcomes. The collaborative teacher-created rubrics were used as instruments to measure student performance in Part Two of the study. Professional development meetings had been used, prior to implementation of the assessments, to help teachers identify and clarify objectives for the assessments and align components of the rubrics more closely to these objectives.

Pilot testing. Since Blooms Taxonomy (1956) has been used extensively throughout the last decade and is now commonplace in educational theory, research, and practice, the taxonomy itself has established reputable validity and reliability. The original taxonomy was selected for use over recent revisions to the taxonomy (Anderson & Krathwohl, 2001; Marzano & Kendall, 2007) since it is more commonly referred to throughout education communities and since it is one-dimensional and therefore less complex and likely easier to establish inter-rater reliability. The Proving Behaviors tool (Appendix A) is also used extensively in various settings, but to increase reliability, the

assessment review panelists practiced use of the tool on other assessments prior to implementation for this study. Practicing use of the tool helped develop consistency among the panelists and added to the trustworthiness of the study.

Threats to internal validity. Social interaction threat could play a role among teacher participants in this study. Since this school values interdisciplinary teaching and authentic assessment as best practices, and expects these strategies to be used in the classroom, it is possible that student scores could have been excluded from or modified for this study by teachers. The researcher discussed confidentiality of research results during initial presentation of the study and expressed the desire for a more genuine study, as opposed to attempt to prove the worthiness of authentic assessment or their teaching methods. The collaborative nature of the assessments should also likely assist in diminishing this threat.

Most importantly, internal validity may be threatened by the arbitrary nature of assessment and rubric development. Teachers underwent the same intervention, including a series of professional development workshops, leading them into similar understandings of rubric design and development, but the nature of each project and the different style, experience and disciplines that exist on each teaching team, led to different approaches to rubric development, different levels of expectations, and a level of arbitrariness in the analysis of student performance. The different approaches and styles are partly what inform the analysis and results of this study, but may also threaten internal validity due to inconsistent approaches to designing quality assessments and rubrics, and analyzing student performance on different assessments. Teacher bias in analyzing student performance on their own assessments also threatens the internal

validity of this study. Reliability of alternative assessments or open-ended assessment items is often challenged (Burry-Stock, Shaw, Laurie, & Chissom, 1996; Tanner, 2001); this inherent potential for lack of consistency is the biggest threat to the validity of this study.

Non-consent of students or parents, or non-performance of students due to absence or lack of work completion, could also interfere with the validity of the data. The researcher attempted to minimize this threat to internal validity, by tracking responses and following through with families to include as many student participants as possible. However, it is possible that non-performance could have resulted from the failure of lower-performing students to complete their assessments, therefore contributing to a bias in the exclusion of the non-responding student population.

Analytic Techniques

Part One and Part Two of this study were analyzed in separate sections of this study, and overall conclusions draw from results of both sections in order to generalize outcomes of interdisciplinary authentic assessment. The following sections describe analytic and statistical methods which were employed for this study.

Part one: Expected cognitive levels. To answer the research question regarding the cognitive levels of prompts and objectives for a series of authentic and interdisciplinary assessments, a three-member expert panel of independent raters used the Proving Behaviors tool (Appendix A) to determine which cognitive levels were expected within each of the four interdisciplinary assessments selected for study. Four tables, one per culminating assessment, display the list of objectives for each task and the corresponding level or levels of cognitive thinking identified by the external panel of

evaluators (see Tables 7, 8, 9, and 10). A summative table and a histogram displaying the frequency of each expected cognitive level, for all of the assessments combined, is also displayed (see Table 6 and Figure 1).

A table analyzing panel agreement on placement of each objective, prior to discussion is also provided, to lend insight into the reliability of the process of ranking objectives into Bloom's Taxonomy (see Table 11).

Part two: performance outcomes. To answer the research question regarding student performance on each cognitive level of Bloom's Taxonomy, scores for each student, describing performance on each learning objective, were measured by grade-level teaching teams on rubrics created by each team of six teachers. For each assessment, participant rubric scores for each objective were averaged by the researcher and tables display mean student scores on each objective associated with that assessment, along with the expected cognitive level or levels determined by the external panel in Part One (see Tables 14, 15, 16, and 17). Additionally, mean rubric scores for each assessment by cognitive level and overall mean rubric scores on each cognitive level are displayed (see Table 12 and 13).

Rubric scores in this study are all based on a 1.0 to 4.0 scale, where 1.0 is the lowest score and 4.0 is the highest score possible. Although teams used different content and styles in their rubrics, all teaching teams used a four-point grading scale, where a score of 3.0 means the student is "proficient," or has adequately met the objective, and a score of 4.0 means the student has exceeded expectations.

Data Storage

Paper records related to the data collection and compilation of data for all parts of this study are maintained by the researcher in a secure, locking file cabinet. Original notes from the expert panel meeting, including worksheets revealing individual and group placement of objectives into Bloom's Taxonomy are included, along with the researcher's notes from the panel discussion. Rubrics from the assessments are stored without any identifying student information, along data analysis records. Since the assessments were also part of students' course grades, teachers have also stored students' summative grades on the assessments in a secure password-protected student information system.

Chapter 4: Results

Overview

The study was designed to help generalize cognitive expectations and related learning outcomes which may result from interdisciplinary authentic assessments at the secondary school level and beyond. The overarching purpose of such analysis is to continue to explore the potential impact of such assessment methods on cognitive development by surfacing theoretical assumptions regarding the rigor and complexity of such assessments, and to generalize potential trends in student performance on expectations of varying cognitive levels. Since the interdisciplinary term at this school is unique in nature, this study serves as pilot research for the development and implementation of interdisciplinary authentic assessment at the secondary level.

Two research questions guided the inquiries in this two-pronged quantitative, non-experimental design study. The first question examined the cognitive levels of Bloom's Taxonomy which exist within interdisciplinary authentic assessments, following an ongoing professional development intervention. The second question examined student performance on varying cognitive levels within these interdisciplinary authentic assessments. This chapter is organized into two sections. The first section examines Research Question One and the second section examines Research Question Two.

Research Question One

This part of the study examines the following question: which cognitive levels of Bloom's Taxonomy are present in interdisciplinary authentic assessment tasks, following an ongoing professional development intervention?

Frequency of expected cognitive levels. Ninety four percent of objectives in this study were determined to assess student understanding beyond the knowledge and comprehension levels of Bloom's Taxonomy. Sixty four percent of objectives were categorized into the highest three cognitive levels of the taxonomy (analysis, synthesis, and evaluation). Overall, the interdisciplinary authentic assessments analyzed in this study reveal varied cognitive complexity. Figure 1 and Table 6 show the frequency and percent of cognitive levels expected of students in each of the four assessments and overall:

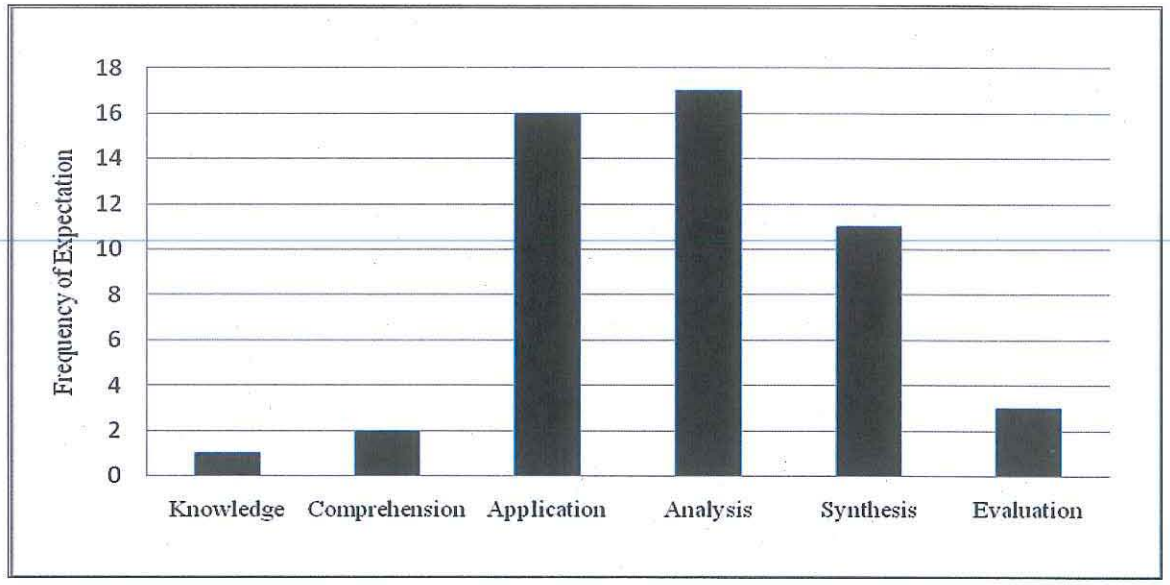


Figure 1. Histogram of objectives by cognitive level.

Table 6.

Frequency and Percent of Objectives by Cognitive Level

	Frequency of Objectives by Cognitive Level					
Cognitive Level	9th Grade Interdisciplinary Assessment (n=13 objectives)	10th Grade Interdisciplinary Assessment (n=12 objectives)	11th Grade Interdisciplinary Assessment (n=13 objectives)	12th Grade Interdisciplinary Assessment (n=12 objectives)	Total	Overall Percent of Objectives by Cognitive Level
Knowledge	1	0	0	0	1	2%
Comprehension	0	0	2	0	2	4%
Application	5	3	4	4	16	32%
Analysis	2	8	3	4	17	34%
Synthesis	3	0	4	4	11	22%
Evaluation	2	1	0	0	3	6%

According to analysis by the external three-member expert panel, and as displayed in Table 6, 94% of objectives from the authentic interdisciplinary assessments fell into levels of Bloom's Taxonomy beyond knowledge and comprehension (see Table 6). Sixty two percent of these objectives fell into the higher levels of the taxonomy (analysis, synthesis, and evaluation). Cognitive levels in the middle to upper end of Bloom's Taxonomy (application, analysis, and synthesis) appear to be assessed most frequently on these tasks, while lower and upper ends of the taxonomy (knowledge, comprehension, and evaluation) appear to be assessed less frequently (see Figure 1). More specifically, according to external panel analysis, 88% of objectives appeared to assess analysis, application, and synthesis levels; 34% were categorized as analysis, 32% as application, and 22% as synthesis. Knowledge, comprehension, and evaluation cognitive levels were each expected 6% or less of the time; only one objective was classified in the knowledge level of Bloom's, two in the comprehension level, and three in the evaluation level.

As displayed in Table 6, the 9th grade interdisciplinary assessment appears to assess the widest range of cognitive levels, with five different levels of cognitive complexity (knowledge, application, analysis, synthesis, and evaluation) identified. The 11th grade assessment appears to assess the next widest range of cognitive levels, with four middle-upper levels of Bloom's taxonomy (comprehension, application, analysis, and synthesis) identified by the panel. The 10th and 12th grade assessments appear to assess three cognitive levels of Bloom's taxonomy; the panel determined that the 10th grade assessment measures the application, analysis, and evaluation cognitive levels and

that the 12th grade assessment measures the application, analysis, and synthesis cognitive levels of the taxonomy.

Expected Cognitive Levels for Each Assessment

Tables 7, 8, 9, and 10 display the expected cognitive levels of each objective for the four assessments examined in this study, as determined by the three-member panel of external evaluators. Again, according to panel analysis, the 9th grade assessment appears to assess five different cognitive levels (all but comprehension). In this assessment, application was the most frequently assessed cognitive level. The 10th grade assessment appears to assess three cognitive levels (application, analysis, and evaluation). Analysis was the most frequently assessed cognitive level. The 11th grade assessment appears to assess four middle-upper cognitive levels (all but knowledge and evaluation). Application and synthesis were the most frequently assessed cognitive levels. The 12th grade assessment appears to assess three cognitive levels; application, analysis, and synthesis were each assessed with the same frequency.

Table 7.

9th Grade Assessment: Expected Cognitive Level of Each Objective

Objective	Cognitive Level					
	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
Grammar / Word Choice	X					
Data to Support Design			X			
Delivery / Engagement			X			
Professionalism			X			
Resource Conservation			X			
Scaling / Explanations			X			
Design / Aesthetics				X		
Real World Problems				X		
Composition / Structure					X	
Participation / Creativity					X	
Schematic / Plan					X	
Collaboration / Decisions						X
Sustainable Design						X

Table 8.

10th Grade Assessment: Expected Cognitive Level of Each Objective

Objective	Cognitive Level					
	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
Presentation			X			
Speaking			X			
Carrying Capacity				X		
Dominant Language				X		
Genetic Engineering				X		
Imperialism				X		
Industrial Revolution				X		
Persuasion				X		
Port of Los Angeles				X		
Sweatshop Production				X		
Debate						X

Table 9.

11th Grade Assessment: Expected Cognitive Level of Each Objective

Objective	Cognitive Level					
	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
American Dream		X				
History: Commentary		X				
Math: Concrete Details			X			
Public Presentation			X			
Science: Concrete Details			X			
Website			X			
Math: Commentary				X		
Science: Commentary				X		
Writing / Language				X		
Collaboration					X	
History: Concrete Detail					X	
Mission Statement					X	
Problem Solving					X	

Table 10.

12th Grade Assessment: Expected Cognitive Level of Each Objective

Objective	Cognitive Level					
	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
Enunciation/Language			X			
Oral / Video Connection			X			
Presentation / Clarity			X			
Video Production			X			
Survey Analysis				X		
Graph of Correlation				X		
Introduction / Conclusion				X		
Investigative Reporting				X		
Creativity / Service					X	
Hooks					X	
Primary Research					X	
Transitions					X	

Panel ranking agreement. Table 11 depicts initial agreement among two out of three of the panelists on cognitive level placement, following individual analysis:

Table 11

Three-Member Panel Ranking Agreement (External Panel of Evaluators)

	Occurrences	Occurrence Percent
Initial Agreement Among Two or More Members	43	86%
No Initial Panel Agreement	7	14%

Prior to group discussion, there was two-member panel agreement on placement 86% of the time. Only 14% of the time did the three-member panel members not place objectives in the same categories prior to discussion and consensus. Since the panel did not always agree initially on placement, discussion and supporting documents were used to work towards consensus of placement. Binders containing supporting curricular and instructional materials helped the external panel identify areas where language in objectives was unclear or where the task appeared to possibly be executed differently than described in the objective. Through discussion, the three-member panel was able to reach consensus for placement 100% of the time.

Research Question Two

The second part of the study examines the following question: how do students perform on each cognitive level of Bloom's Taxonomy within interdisciplinary authentic assessment tasks? For this research question, rubric scores for participants (assigned by six-member teaching teams) were analyzed. Overall student scores were tabulated and organized by the researcher, by each objective, assessment, and cognitive level.

Overall analysis of each assessment. Table 12 displays mean rubric scores for each interdisciplinary assessment by cognitive level along with overall mean rubric scores for each assessment (see also Table 13). The mean score in this table represents the average student rubric score on the overall interdisciplinary assessment for each grade level. Rubric scores are based on a 1.0 to 4.0 scale, where 1.0 is the lowest score possible and 4.0 is the highest score possible. Although there was flexibility across grade level teams in the content or style, all teams used a four-point grading scale, where a score of 3.0 represents proficiency, or adequate performance on the objective, and a score of 4.0 means student performance has exceeded expectations.

Mean scores on each cognitive level do not appear to increase or decrease with hierarchical progression of Bloom's Taxonomy. Overall mean rubric score on each cognitive level ranged from 2.7 to 3.1. Mean scores on each cognitive level across assessments ranged from 1.9 to 3.1. The largest variation in mean scores was for the analysis level, ranging from 1.9 to 3.2, and for the evaluation level, ranging from 2.5 to 3.2. Of particular note are the relatively high rubric scores on the synthesis objectives, the second highest level of the taxonomy. Scores on synthesis objectives all ranked above the proficiency level of 3.0, indicating that students, on average, are able to meet expectations at higher levels of cognitive demand.

Table 12.

Mean Rubric Scores for Each Assessment by Cognitive Level

	Assessment				
Mean Rubric Score	9th Grade Interdisciplinary Assessment (n = 99 students)	10th Grade Interdisciplinary Assessment (n = 111 students)	11th Grade Interdisciplinary Assessment (n = 96 students)	12th Grade Interdisciplinary Assessment (n = 91 students)	School-Wide (n = 397 students)
Knowledge	3.1	-	-	-	3.1
Comprehension	-	-	2.7	-	2.7
Application	3.0	2.7	2.7	3.0	2.9
Analysis	1.9	3.2	2.4	3.1	2.9
Synthesis	3.2	-	3.1	3.0	3.1
Evaluation	2.5	3.2	-	-	2.8
Overall Mean Rubric Score	2.8	3.1	2.7	3.0	2.9

Overall analysis by cognitive level. Table 13 displays mean overall rubric scores according to each cognitive level. Mean scores on each objective from all four assessments were sorted by cognitive level and averaged with all mean scores for that cognitive level.

Table 13.

Mean Overall Scores by Cognitive Level

Cognitive Level	Total Number of Objectives Analyzed	Mean Score by Cognitive Level (on 4.0 Scale Rubric)
Knowledge	1	3.1
Comprehension	2	2.7
Application	16	2.9
Analysis	17	2.9
Synthesis	11	3.1
Evaluation	3	2.8

($N = 397$)

Mean scores. As shown in Table 13, there does not appear to be any negative or positive trend in student performance across the hierarchy of Bloom's Taxonomy. Since a 3.0 rubric score represents student "proficiency" and most mean scores were close to a 3.0, Table 13 reveals that most students appear to be proficient, or to meet the learning objectives, at each cognitive level. This is a substantial finding, indicating that students are generally able to meet assigned objectives, regardless of cognitive complexity.

Furthermore, knowledge and synthesis cognitive levels had the highest mean performance scores (3.1 for each). Higher relative mean rubric scores at the synthesis level is also a substantial finding, since synthesis is the second highest level of understanding in the taxonomy. Comprehension and evaluation cognitive levels had the lowest mean performance (2.7 and 2.8, respectively); these mean scores, however, are only based on student performance on two and three objectives.

Analysis by assessment. Tables 14, 15, 16, and 17 break down the mean rubric scores and standard deviations for each objective within the four interdisciplinary assessments.

9th grade assessment. Table 14 displays the mean rubric scores and standard deviations for each objective on the 9th grade assessment.

Table 14.

9th Grade Assessment: Mean Rubric Scores for Each Objective

Objective	Level	Mean	Standard Deviation
Grammar / Word Choice	knowledge	3.1	0.98
Data to Support Design	application	2.6	0.54
Delivery / Engagement	application	3.3	0.39
Professionalism	application	4.0	0.00
Resource Conservation	application	2.6	0.53
Scaling / Explanations	application	2.8	0.99
Design / Aesthetics	analysis	1.3	0.90
Real World Problems	analysis	2.5	1.09

(continued)

9th Grade Assessment: Mean Rubric Scores for Each Objective (continued)

Objective	Level	Mean	Standard Deviation
Composition / Structure	synthesis	3.0	1.08
Participation / Creativity	synthesis	3.5	0.45
Schematic / Plan	synthesis	3.0	0.00
Collaboration / Decisions	evaluation	2.5	1.02
Sustainable Design	evaluation	2.5	0.52

($N = 99$)

There was much variation in student performance on the 9th Grade Assessment; mean student scores ranged from 1.3 to 4.0, with standard deviations ranging from 0.0 to 1.09. Analysis and evaluation objectives display the lowest mean scores (no higher than 2.5). Knowledge and synthesis display the highest mean scores (all higher than 3.0). Again, higher relative scores on synthesis objectives indicate that students can perform as well on objectives of higher cognitive complexity.

10th grade assessment. Table 15 displays mean rubric scores and standard deviations for each objective on the 10th grade assessment.

Table 15.

10th Grade Assessment: Mean Rubric Scores for Each Objective

Objective	Level	Mean	Standard Deviation
Presentation	Application	3.5	0.63
Speaking	Application	4.0	0.54

(continued)

10th Grade Assessment: Mean Rubric Scores for Each Objective (continued)

Objective	Level	Mean	Standard Deviation
Carrying Capacity	Analysis	3.0	0.69
Dominant Language	Analysis	2.9	0.95
Genetic Engineering	Analysis	3.1	0.59
Imperialism	Analysis	3.4	0.69
Industrial Revolution	Analysis	3.4	0.69
Persuasion	Analysis	3.3	0.89
Port of Los Angeles	Analysis	3.2	0.64
Sweatshop Production	Analysis	3.2	0.64
Debate	Evaluation	3.2	0.62

($N = 111$)

Mean scores on the 10th grade assessment ranged from 2.9 to 4.0, with standard deviations ranging from 0.54 to 0.95. Generally, students performed slightly better on the lower cognitive objectives; mean scores on application objectives range are both above 3.5, mean scores on the analysis objectives ranged from 2.9 to 3.4, and the mean score on the evaluation objective was 3.2. Since a 3.0 score indicates that students have met learning objectives, it is important to note from this chart that the mean score in each cognitive level assessed was near or above proficiency, even for the highest cognitive level (evaluation).

11th grade assessment. Table 16 displays the mean rubric scores and standard deviations for each objective on the 11th grade assessment.

Table 16.

11th Grade Assessment: Mean Rubric Scores for Each Objective

Objective	Level	Mean	Standard Deviation
American Dream	comprehension	1.3	0.69
History: Commentary	comprehension	4.0	0.00
Science: Concrete Details	application	3.2	1.24
Math: Concrete Details	application	2.3	1.15
Website	application	2.5	0.77
Presentation*	application	*	*
Science: Commentary	analysis	2.7	1.01
Math: Commentary	analysis	2.3	1.15
Writing / Language	analysis	2.2	0.94
Mission Statement	synthesis	2.2	0.94
History: Concrete Detail	synthesis	4.0	0.00
Collaboration*	synthesis	*	*
Problem Solving*	synthesis	*	*

($N = 96$)

* This particular objective was not graded by the teaching team. There was no information provided as to why several objectives listed in the project descriptions from the 11th grade teaching team were not assessed on their rubric.

Again, there was a wide range of student performance on the 11th grade assessment. Mean student scores ranged from 1.3 to 4.0, with standard deviations ranging from 0.00 to 1.24. Students received both the lowest (1.3) and highest (4.0) mean scores

on the two comprehension objectives. Students received similarly low (2.2) and high (4.0) mean scores on the two synthesis objectives. Application scores ranged from 2.3 to 3.2 and analysis scores ranged from 2.2 to 2.7. Although there is wide range of scores at each cognitive level, it is important to note that there are some higher scores in the second highest levels of the taxonomy (synthesis).

It is interesting to note that the presentation, collaboration, and problem-solving objectives were not scored by the 11th grade teaching team. The external panel of evaluators speculated as to why these objectives did not appear on the rubric. Two of the objectives fall in the synthesis level of the taxonomy, leaving the panel to speculate as to whether objectives in the upper levels of Bloom's taxonomy might present more difficulty in the design of assessments and/or scoring guides. Collaboration and problem-solving are not very concrete demonstrations of learning and perhaps the teaching team felt as though the evaluation and measurement of these skills would not be easy to delineate with objectivity. It is uncertain, however, why the teaching team did not decide to grade students on the presentation of their projects.

12th grade assessment. Table 17 displays the mean rubric scores and standard deviations for each objective on the 12th grade assessment.

Table 17.

12th Grade Assessment: Mean Rubric Scores for Each Objective

Objective	Level	Mean	Standard Deviation
Enunciation/Language	application	3.1	0.61
Oral / Video Connection	application	3.2	0.61

(continued)

12th Grade Assessment: Mean Rubric Scores for Each Objective (continued)

Objective	Level	Mean	Standard Deviation
Presentation / Clarity	application	3.0	0.32
Video Production	application	2.9	0.52
Graph of Correlation	analysis	3.0	0.49
Introduction / Conclusion	analysis	3.2	0.49
Investigative Reporting	analysis	3.1	0.55
Survey Analysis	analysis	3.0	0.68
Creativity / Service	synthesis	3.0	0.61
Hooks	synthesis	3.0	0.72
Primary Research	synthesis	3.1	0.56
Transitions	synthesis	3.0	0.73

($N = 91$)

The 12th grade assessment displayed the smallest range of student performance. Mean scores ranged from 2.9 to 3.2, with standard deviation ranging from 0.32 to 0.73. Overall, there was no substantial variation in student performance on each objective, nor was there much variation according to cognitive level. Application scores ranged from 2.9 to 3.0; analysis scores ranged from 3.0 to 3.2; and synthesis scores ranged from 3.0 to 3.1. On average, most mean scores within this assessment were at proficient level (3.0) or above for each cognitive level, including the second highest cognitive level (synthesis). The lack of substantial variation in student scores across objectives indicates that students performed just as well on higher cognitive levels as on lower cognitive levels.

Summary of Results

Part one: Expected cognitive levels. Overall, the interdisciplinary authentic assessments analyzed for this study revealed some cognitive complexity, with the upper-middle levels of the taxonomy most frequently assessed. Most of the objectives analyzed fell within the application, analysis, and synthesis levels of the taxonomy. Very few objectives were classified in the knowledge, comprehension, or evaluation taxonomy levels. The overall cognitive complexity of each assessment fell within the application or analysis level of Bloom's Taxonomy (closer to analysis). In most cases, two of the three members agreed initially without discussion; in the remaining 14% of the time, panel members discussed reasoning and used supporting documents to reach consensus on panel placement.

Part two: Performance outcomes. Although there was no consistent increasing or decreasing trend in mean rubric scores across the hierarchy of Bloom's Taxonomy, students on average scored just below, at, or above the 3.0 proficiency level. Rubric scores indicate that students were able to meet proficiency on cognitively demanding objectives, just as they were on less challenging objectives. The lowest mean student scores overall were seen on comprehension and evaluation level objectives, but some mean scores on those objectives were also high. The highest mean student scores were seen on knowledge, application, and synthesis objectives. Three objectives on one assessment were not actually scored; two of these objectives were placed in the synthesis level of Bloom's taxonomy and one in the application level.

Chapter 5: Conclusions and Recommendations

Overview

This two-pronged quantitative, non-experimental design study was guided by two research questions, which examined a) the cognitive levels of Bloom's Taxonomy present within four interdisciplinary authentic assessments, following an ongoing professional development intervention and b) student performance on these assessments. This conclusion will summarize the findings of each research question, draw conclusions from these findings and relate these conclusions to relevant literature, and make recommendations for policy, practice, and further study.

Conclusions, Discussions, Literature Contributions

Cognitive complexity. Since it is commonly believed and discussed in literature that traditional testing engages students primarily in lower-order thinking, or knowledge and comprehension levels of Bloom's Taxonomy (Bloom, 1956; Dana & Trippins, 1993; Herman, 1997; McTighe & Wiggins, 2004; Resnick, 1987;), this study helps us understand that authentic interdisciplinary assessments may be an important and necessary component of school curricula, since such tasks appear to diversify the levels of cognition expected of students, particularly beyond knowledge and comprehension levels. As noted in Table 6, 94% of objectives were designed to measure student understanding beyond knowledge and comprehension levels of the taxonomy; 62% of the objectives were in the top three cognitive levels of the taxonomy (analysis, synthesis, and evaluation). This study supports the theoretical work of many education theorists who believe that authentic assessment tends to engage students at higher cognitive levels than traditional assessments (Bransford et al, 2000; Drake & Burns, 2004; Erickson,

2002; Tchudi & Lafer, 1996; Wiggins & McTighe, 2005), and the work of theorists who believe that interdisciplinary curricular design promotes more complex student learning and higher levels of cognition (Drake & Burns, 2004; Ivanitskaya et al, 2002; Letterman & Dugan, 2004; Kovalik, 1993; McGehee, 2001; Tchudy & Lafer, 1996).

Upper cognitive levels. Only six percent of objectives were determined to assess the evaluation level of Bloom's Taxonomy. However, 22% of objectives were determined to assess the synthesis level of understanding, and 34% the analysis level. Additionally, students performed just as adequately on these higher cognitively complex objectives as they were on the lower cognitive objectives. Such data imply that students are capable of performing at higher cognitive levels when challenged to do so.

White (2007) explained that the synthesis level of cognition is not generally expected from students until they are engaged in graduate level university work. That over twenty percent of objectives on the interdisciplinary assessments were determined to measure the synthesis cognitive level implies that the assessments analyzed are cognitively demanding assessments which may push students to think beyond traditional expectations for most high school students. White (2007) further explains that "dissertation review and journal editing are examples of this <evaluation> level of knowledge," (p. 161). Although six percent of objectives in this study were classified in the evaluation level, this is likely still a push beyond traditional and standardized secondary school assessments, which tend to test lower cognitive levels of student understanding (Bloom, 1956; Dana & Trippins, 1993; Herman, 1997; Resnick, 1987; Wiggins & McTighe, 2004). More importantly, students in this study demonstrated the capacity to meet the challenges at the varying cognitive levels assigned.

Lower cognitive levels. Some education theorists suggest the importance of assessing students at multiple levels of cognition (Heflebower, 2010; Marzano, 2010; White, 2007), so that all students have the opportunity to demonstrate understanding at their level of cognition. As White (2007) explains, "Unfortunately, if questions are posed <only> at very high levels of learning, students with less ability may find themselves unable to receive any credit at all thereby underrepresenting their grade," (p. 162). If a diversity of cognitive assessment levels is important, there is room within or in conjunction with such interdisciplinary performance tasks to also include assessment of lower cognitive levels, since very few objectives within only two of the assessments analyzed assessed lower levels of cognition (knowledge and comprehension).

Additionally, although the taxonomy specifically assumes a hierarchy of learning objectives, with each new level requiring prerequisite "skills and abilities which are lower in the classification order," (Bloom, 1956, p. 120), and since this study assumed that lower level knowledge was required for the performance on upper cognitive levels, it was difficult for the external panel members to tell anecdotally from materials analyzed whether knowledge and comprehension were actually required for the performance of many of these objectives. For example, it occasionally appeared possible for students to apply, analyze, synthesize, or evaluate information without really having clear knowledge or comprehension of particular concepts. This is not necessarily a judgment of the assessments, rather an anecdotal observation regarding expected behaviors. Development of higher-order thinking skills without the ability to recall rote learning may not necessarily be a negative attribute of these assessments. However, if it is in fact important to assess varying levels of understanding, then coupling of traditional

and interdisciplinary authentic assessment, or more intentional inclusion of lower level understandings within these alternative assessments, may be a valuable consideration.

Measurement and grading. On another note, since there is much subjectivity with regards to assessment and scoring (Avery, Carmichael-Tanaka, Kunze & Kouneski, 2000; Suurtamm, 2004; Tanner, 2001), it may be that the assessments themselves and the expected levels of complexity may be more valuable than the way we actually evaluate and grade students; this value, however, is difficult to measure.

According to "The Intelligence We Never Measure," there is implied difficulty in measuring cognitively complex understanding. "The more complex and sophisticated the cognitive function, the more difficult it is to measure," ("The Intelligence We Never Measure," 2007, p. 83). In the assessments in this study, not all objectives in the assessments were actually graded or measured by teachers. Although reasons for the lacking measurement on these three objectives are only speculative, this could be related to the difficulty of measurement alluded to in the aforementioned article. However, students were still asked to perform in a complex manner for these assessments, regardless of the measurement of performance. Although grades for alternative assessment measures can be difficult to assign and are often subjective, this should not prevent educators from continuing to assign students cognitively complex thinking tasks.

From this study, lingering questions regarding the potential impact of unmeasured learning objectives remain. Some investigation in this area already exists (Bittel & Hernandez, 2006; Johnson, 2008, Pederson & Williams, 2004), yet further

research around the impact of assigning grades for student performance of complex tasks would also be useful.

Varying interpretations. In observing the panel classification process, the researcher made note of comments during the panel discussion, to help inform our understanding of the results and any future needs for additional investigation. This became particularly important to the researcher during the study while observing the variation in individual panel categorization of the objectives and the clarifying discussion.

As the panel discussed each objective, they often referred to supporting documents to help clarify actual expectations. One panelist noted that objectives were "densely packed with several implied tasks and behaviors," making it more difficult to place objectives into one level and for panel members to reach initial agreement on classification. Another member added that the ranked cognitive level of many of the objectives would "depend on how exactly the assessment was implemented by the teachers." The panel agreed many components of the assessments, as written, could be executed very differently by different teachers or in different school settings. The potential impact of instruction, then, is an important limitation of this study which should be further explored.

Tanner (2001) describes the arbitrary nature of authentic assessment, which also implies that tasks are interpretable and could be executed differently, depending on the teacher or context. Actual classroom instruction and feedback may, then, play a larger role in shaping how students are cognitively assessed, regardless of how each objective or rubric is written. The written task may not be as telling as the execution of that task

might be, implying a need for further investigation regarding implementation or teaching of the interdisciplinary authentic assessments. In this study, since there was only one teacher per discipline per grade, instruction was consistent for each student; students on each grade level all had the same six teachers delivering instruction. However, instruction by a different set of teachers may have resulted in varying different delivery and/or varying learning outcomes for students.

There were also stages of panel analysis where one word within each objective appeared to be interpreted differently by the different external panel members. Descriptive words used in the Proving Behaviors tool occasionally seemed to be used by teachers in a different manner or context in some of the objectives analyzed. One panel member noted as they determined placement that, "There may be a significant difference between what teachers actually expect students to do and the language used to describe these behaviors." The language in the objectives did not always stand well alone and, as written, could be taught or executed in many different manners, depending on teacher execution. There is apparent room, then, to develop more common language and consistency in use of the language among teachers and other education professionals. Additionally, in-depth discussion of objectives and analysis of the teaching tasks may add important insight and clarity to understanding among teaching teams.

Concept integration. According to Appendix B, one criterion for an authentic assessment is that the assessment "requires integration of knowledge from overlapping concepts or disciplines." In referring to some of the prompts and rubrics for the assessments in this study, one external panel member asked the others whether all of the assessments actually integrated the separate disciplines. For the 10^{th} grade assessment,

for example, students were graded on whether they understood material regarding specific concepts, which were designed from an overarching theme and essential question, either through debate and/or through a written assignment, but it was unclear whether the debates or the writing assessment clearly required the students to integrate concepts from multiple disciplines. Similarly, although the 9th Grade Assessment required students to integrate several disciplines in order to design, construct, and justify their structures, each disciplinary concept within the writing task was divided into separate writing components. There appeared to be a tendency within each assessment to isolate the disciplines for scoring or grading purposes, even though these assessments resulted in only one overall transcript grade for each student for the six interdisciplinary courses. The lack of integration within the assessments analyzed does not necessarily imply lower or higher cognitive expectation, but rather is an anecdotal observation important to our core understanding of interdisciplinary authentic assessment and the design and implementation of such practices within one school.

There are several potential reasons for the underdeveloped integration of concepts within these interdisciplinary assessments, also referred to as the Potpourri Problem (Jacobs, as cited by McGehee, 2001). Even with a professional development intervention, the tendency to use a "sampling of knowledge from each discipline" (McGehee, 2001, p. 380), without really focusing on integration, is still present.

Authentic assessment takes more time to implement and pushes students into deeper exploration of fewer objectives, leaving less room for holding students accountable for learning a high number of standards (Avery et al., 2000), as teachers might in preparation for standardized tests. It is possible that increased pressure and

scrutiny related to standards-based teaching and student achievement (Dutt-Doner & Maddox, 1998; Herman, 1997; Koretz, 2008; Resnick & Resnick, 1992) may interfere with teachers' full commitment to integrating disciplines.

It could also be that teachers are accustomed to having more individual control over the teaching and learning in their classrooms. Planning with other instructors, as Letterman and Dugan (2004) describe, leads to less autonomy in the classroom and limited flexibility with curriculum, methods, and timing. It is possible that isolating student performance by discipline allows teachers to feel more potential for holding students accountable to specific lessons delivered in their own classrooms.

The different teaching teams' success in integrating disciplinary understandings is not entirely unexpected, as several authors point to a lack of expertise, training, and availability of planning frameworks concerning the creation of interdisciplinary teaching units (Letterman & Dugan, 2004; Jacobs, 1989; McGehee, 2001). Furthermore, teachers are often pushed outside of comfort zones and experience levels when challenged with the opportunity for interdisciplinary planning (McGehee, 2001). The varying or questionable integration of concepts could result from either the need for more professional instruction or guidance on how to create an interdisciplinary assessment. Letterman and Dugan and McGehee, and the results of this study, suggest a potential need for increased availability of, and support with, instructional information and supporting frameworks for interdisciplinary planning.

Assessment design. Similarly, the design of each assessment was different in approach across the grade level teams. Most assessments included a performance task and a writing assessment, although some assessments focused more heavily on one or

the other. Each assessment took a slightly different approach to integration of concepts and each had a slightly different approach to rubric design and/or grading, even with the professional development intervention. This variation in assessment creation brings up the question of the arbitrary nature of assessment design, and the challenges regarding articulation, implementation, and grading (Avery et al., 2000; Suurtamm, 2004; Tanner, 2001). More research should be done to explore such varying approaches to assessment design and the impact on student performance.

Similarly, although anecdotal, the ambiguity discovered by the panel with regards to the language and writing style of the objectives, led to skepticism on behalf of the researcher with regard to scoring reliability of the rubrics. This is one of the most important criticisms of authentic assessment and also a reason why standardized testing has become such a commonplace measurement of school success (Dutt-Doner & Maddox, 1998; Herman, 1997; Koretz, 2008; Resnick & Resnick, 1992). It would be ideal if, at least throughout one school, there was more consistency with regard to language usage, creation of rubrics, and scaling of objectives or proficiency levels.

It is important, however, to remember that authentic assessment has been and can be consistent and reliable (Moon et al., 2005). Expert collaboration and much dedicated time for this collaboration may be necessary, but this collaborative work can lead to more consistency and objectivity in assessment design and implementation.

Overall student performance. Overall student performance did not appear, for the most part, to depend on the level of cognition expected. Instead, a wide range of student scores were mostly displayed throughout the hierarchy of cognitive levels. Similarly, students demonstrated success on interdisciplinary assessments at each grade

level and each level of objectives, indicating an ability of students to meet higher level objectives when challenged. Wiggins and McTighe (2005) stated that, "Students in general can do low-level tasks but are universally weak in higher-order work that requires transfer," (p. 45). The overall lack of disparity throughout the study between student performance on low and higher level objectives may mean the universal trend Wiggins and McTighe identified is likely not due to student capabilities, but perhaps due to the levels with which with we challenge students to achieve in traditional school curricula.

This outcome of the research supports the study of Bittel and Hernandez (2006) which found that 98% of students, when given the choice, chose to engage in more cognitively complex work for higher grades, and that 90% of these students were able to perform at the higher thinking levels once engaged in those cognitively complex tasks. The work of Bittel and Hernandez, and the outcomes of this study, indicate that perhaps student achievement does not generally depend on difficulty or cognitive demand, but that students will still achieve and perform just as consistently when challenged.

Higher cognitive levels. Still, there is some speculation based on student rubric scores, as to whether there was a pattern with regard to student performance on the evaluation objectives. Since there were only three evaluation objectives, the results from this study are speculative and not significant enough to be entirely conclusive. Students did perform less proficiently on evaluation-ranked objectives for one of the assessments, implying that this highest cognitive demand may be difficult for high school students to achieve. Since White (2007) describes the evaluation level of cognition as a typical level of thinking for dissertation writing or peer review of journals, which are advanced

college-level thinking skills, it would make sense that these objectives would be more difficult for secondary students than the lower-level objectives. Still, students were able to perform well enough on these objectives to indicate that they should in fact be pushed to think beyond lower levels of the taxonomy. Further study specifically regarding student performance on evaluation objectives is recommended.

Recommendations for Policy and Practice

Practice. ***Value of interdisciplinary approach.*** Overall, interdisciplinary authentic assessments appear to appropriately challenge students to reach a variety of cognitive thinking levels, particularly application, analysis, and synthesis levels of Bloom's Taxonomy. Student performance does not seem dependent, for the most part, on the level of cognitive challenge. Therefore, educators and schools should work to develop more such learning experiences for students. Darling-Hammond (1995) mentions, specifically, that economically disadvantaged and culturally diverse groups face a particular inequity with regard to such experiences. Since this study did not reveal any lack of ability of such groups to perform on such cognitively complex assessments, teachers and schools serving these groups of students should specifically focus on developing interdisciplinary authentic assessments, as they may challenge these students into new levels of thinking and cognitive performance. The need for the development of such curricula is particularly important in the wave of increased accountability of schools to perform on standardized tests (Dutt-Doner & Maddox, 1998; Haney, et al., 2007; Herman, 1997; Koretz, 2008; Rotham, 1995).

Professional collaboration. For this shift in implementation of a varied curriculum to take place, there needs to be much more time within the school day, or

within existing collaboration structures, spent on collaborative curricular development and on the writing and analysis of learning objectives among teaching teams. More accessibility and instruction around interdisciplinary curriculum design and teaching methods should be encouraged in schools (Letterman & Dugan, 2004; Jacobs, 1989; McGehee, 2001).

Expertise in this type of collaborative work does not happen naturally or quickly. Teachers should visit other schools implementing similar curricula to learn from others' collaboration sessions, assessment designs, instruction, and grading practices. Teachers should participate in the analysis of existing assessments to evaluate together strengths and weaknesses of various assessments, to strength common team understanding of criteria for effective interdisciplinary assessment design. Similarly, participation of teachers in analysis of their own and others' objectives and rubrics, similar to the analysis of the expert panel in this study, would add a valuable inquiry component to teachers' reflective practices surrounding curriculum development, writing of objectives, and normalization of language usage. Teachers generally need more collaborative professional development time within the school day to be able to engage in such inquiry processes together, which may imply a needed shift in the way our school systems operate.

Assessment design and consistency. The varied nature of assessments, objectives language, rubric design, concept integration, and scoring in this study also imply a need for more consistency within schools. There tends to be much isolation and autonomy across classrooms, and schools should strive to lessen the arbitrary nature of

the design of authentic assessments, where they exist, to lessen scrutiny or concern with regards to design and scoring practices.

Schools should specifically adopt similar or shared structures for assessment and rubric design and, if implementing interdisciplinary assessment practices, should have specific formats and criteria for how concepts are integrated and how writing tasks are prompted and how scoring is approached. Teachers could collaboratively decide on their own criteria after analyzing strengths and challenges of other existing programs and assessments together. Most importantly, teachers should work to develop consistency and reliability in their grading practices and use of language.

Since this study implies value in assessing all cognitive levels of Bloom's Taxonomy, formats and criteria should assess students at varying levels of cognition, to ensure consistency and a variety of critical thinking opportunities. Teachers could use a checklist of criteria, with specific reference to the varying cognitive taxonomy levels, to ensure through collaborative review and analysis that the assessments evaluate students at multiple levels of learning. Similarly, interdisciplinary authentic assessments could intentionally be coupled with more traditional assessments to provide opportunities for students to demonstrate understanding at various levels.

As literature suggests, teachers need more instruction and specific framework designs and samples for developing expertise with regards to interdisciplinary teaching (Letterman & Dugan, 2004; Jacobs, 1989; McGehee, 2001). Schools can develop their own formats and guidelines or could use existing frameworks. One example of a format which encourages assessment of the multiple levels of understanding is the "Diploma

Plus Competency Rubric" (Diploma Plus, 2011), which provides a template for designing rubrics specifically around taxonomy levels.

Policy. As federal and state policies continue to push for standardization of learning and assessment, and as schools face increased scrutiny and accountability for pushing student achievement on standardized multiple choice tests, our nation misses a critical opportunity to develop students who are ready for the creative and cognitively complex demands of the workforce and higher education. If education policy continues to limit itself to pushing multiple-choice tests and teaching around these tests, teachers and schools will continue to focus on developing low levels of cognitive thought and other possibilities among our youth.

Although the challenges of developing and implementing interdisciplinary authentic curricula in schools are many (Avery et al., 2000; Darling-Hammond, 1995; Letterman & Dugan, 2004; McGehee, 2001; Suurtamm, 2004; Stokking et al., 2004; Trepanier-Street et al., 2001; Tanner, 2001), many schools will never be willing to rethink their approach to alternative education if the standardized-assessment movement continues to take precedence over the development of integrated curriculum design and cognitively complex assessment. Issues of consistency and reliability with regards to authentic assessment may be real concerns, but this does not mean that we should not encourage our students to participate in a performance-based education. Education policy should broaden to encourage and support schools and teacher education programs in the development and analysis of such curricula. This does not necessarily imply that standardized testing should be curbed, but it cannot be the only policy focus for our schools.

Recommendations for Further Study

- A study specifically examining student performance on evaluation objectives.
- Further study which explores the impact of grades or measurement of student performance on complex tasks.
- A pre-post study measuring the development of knowledge and comprehension objectives, along with measurement of performance on more complex tasks.
- A pre-post study of perceived skill development related to such assessments.
- A pre-post study of student motivation, attitudes, or interest related to such assessments.
- A study measuring the quality of rubric design related to such assessments.
- A study of varying assessment design approaches and the related impact on student performance.
- A qualitative study of teacher collaboration around the of interdisciplinary authentic assessment.
- A qualitative study of the professional development series, which notes observations and measures teacher response and implementation.
- A qualitative study of the implementation or teaching of the interdisciplinary authentic assessments.
- A follow-up study of the long term impact of such assessment experiences on what students remember and their perceived skill development as related to the interdisciplinary authentic assessment experiences.
- A study which examines and/or compares the performance of various demographic groups (i.e. age, socioeconomic, and ethnic groups).

Conclusion

The study served to generalize cognitive expectations and related learning outcomes which may result from the design and implementation of interdisciplinary authentic assessment, in order to contribute to existing literature on our understanding of the manner and complexity with which such assessments push secondary school students to think. Assumptions regarding the cognitive complexity of such assessments were examined and used to generalize potential trends in student performance, or lack thereof, on expectations of varying cognitive levels.

Overall, the interdisciplinary authentic assessments analyzed in this study revealed cognitive complexity, with the upper-middle levels of Bloom's Taxonomy (application, analysis, and synthesis) expected most frequently. The cognitive level of each objective did not appear to correlate directly or inversely to student performance; rather, students were able to meet expectations at the various levels of the taxonomy. Such trends imply that student performance may not, in fact, depend on cognitive demand of the task, and also merit the need for further study of student performance on assessments of varying cognitive difficulty.

As a pilot study for a unique curricular approach at one secondary school, this research points out the potential benefits of such programming for the development of complex cognitive thought among students. The study not only highlights important considerations for high schools implementing similar programs, but recommends broad policy direction for the global education community.

REFERENCES

Ackerman, D.B. & Perkins, D.N. (1989). Integrating thinking and learning skills across the curriculum. In H.H. Jacobs (Ed.), *Interdisciplinary Curriculum: Design and Implementation.* (pp 77-96). Alexandria, VA: Association for Supervision and Curriculum Development.

Abadiano, H. R. & Turner, J. (2003). Thinking it through: Re-examining our beliefs about assessment for diverse students. *The New England Reading Association Journal, 39*(1), 58-63.

Anderson, L.W. & Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing.* New York, NY: Addison Wesley Longman, Inc.

Avery, P.G., Carmichael-Tanaka, D., Kunze, J., & Kouneski, N.P. (2000): Writing about immigration: Authentic assessment for U.S. History students. *Social Education, 64*(6), 372-375. Retrieved from http://www.socialstudies.org/socialeducation

Barisonzi, J. & Thorn, M. (2003). Teaching revolution: Issues in interdisciplinary education. *College Teaching, 51*(1), 5-8. doi: 10.1080/87567550309596402

Bauer, E.B. & Garcia, G. E. (2002). Lessons from a classroom teacher's use of alternative literacy assessment. *Research in the Teaching of English, 36*, 462-494. Retrieved from http://www.ncte.org/journals/rte

Bissell, A. N. & Lemons, P. P. (2006). A new method for assessing critical thinking in the classroom. *BioScience, 56*(1), 66-72. Retrieved from http://www.ias.ac.in/jbiosci

Bittel, K. & Hernandez, D. (2006). Differentiated assessment. *Science Scope, 30*(4), 49-51.

Bloom, B. S. (Ed). (1956). *Taxonomy of educational objectives: Classification of educational goals. Handbook 1: Cognitive domain.* New York, NY: Longman, Green & Co.

Bond, L.A., Braskamp, D., & Roeber, E. (1996). *The status report of the assessment programs in the United States.* Oakbrook, IL: North Central Regional Educational Laboratory (NCREL)/Council of Chief State School Officers.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How People Learn: Brain, Mind, Experience, and School.* Washington, D. C.: National Academy Press.

Burry-Stock, Shaw, Laurie, & Chissom (1996). Rater agreement indexes for performance assessment. *Educational and Psychological Measurement, 56*, 251-262. doi: 10.1177/0013164496056002006

Carini, R. M., Kuh, G. D., & Klein, S. P. (2006). Student engagement and student learning: testing the linkages. *Research in Higher Education, 47*(1), 1-32. doi: 10.1007/s11162-005-8150-9

Cresswell. J. W. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches.* Thousand Oaks, CA: Sage Publications.

Dana, T. M. & Tippins, D. J. (1993). Considering alternative assessment for middle level learners. *Middle School Journal, 25*(2), 3–5. Retrieved from http://www.amle.org/Publications/MiddleSchoolJournal/tabid/435/Default.aspx

Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work.* San Francisco, CA: Jossey-Bass.

Davies, M.A. & Wavering, M. (1999). Alternative assessment: New directions in teaching and learning. *Contemporary Education, 71*(1): 39-45.

Dickie, O. (2003). Approach to learning, the cognitive demands of assessment, and achievement in physics. *The Canadian Journal of Higher Education, 33*(1): 87-112. Retrieved from http://ojs.library.ubc.ca/index.php/cjhe

Diploma Plus (2011). DP Competencies and Rubrics. *DiplomaPlus.net.* Retrieved June 25, 2011, from https://network.diplomaplus.net

Drake, S. M. & Burns, R. C. (2004). *Meeting Standards Through Integrated Curricula.* Alexandria, VA: Association for Supervision and Curriculum Development.

DuFour, R., & Eaker, R. (1998). *Professional Learning Communities at Work: Best Practices for Enhancing Student Achievement.* Alexandria, VA: Association for Supervision and Curriculum Development.

Dunn, R. & Dunn, K. (1978). Teaching students through their individual learning styles: A practical approach. Reston, VA: Reston Publishing Company- a Prentice-Hall Division.

Dutt-Doner, K. M. & Maddox, R. (1998). Implementing authentic assessment. *Kappa Delta Pi, 35*(4), 135-137. Retrieved from http://www.kdp.org/publications/

Erickson, H. L. (2002). *Concept-Based Curriculum and Instruction: Teaching Beyond the Facts.* Thousand Oaks, CA: Corwin Press.

Farr, B., & Trumbull, E. (1997). *Assessment alternatives for diverse classrooms.* Norwood, MA: Christopher-Gordon.

Field, M., Lee, R., Field, & M. L. (1994). Assessing interdisciplinary learning. *New Directions for Teaching and Learning, 58*, 69-84. Retrieved from http://www.josseybass.com

Frame, A., Evola, S., Arezina, N. (2010, November). *Interdisciplinary benchmark project, unit 2: Influencing river systems.* Assessment description for Environmental Charter Middle School, Inglewood, California.

Gardner, H. (1993). *Multiple Intelligences: Theory Into Practice.* New York, NY: Basic Books.

Glass, R., & Rose, M. (2008). Tune out, turn off, drop out. *American Teacher, 93*(3), 8-10, 21. Retrieved from http://www.aft.org/newspubs

Gulikers, J. T. M., Kester, L., Kirschner, P. A. & Bastiaens, T. J. (2008). The effect of practical experience on perceptions of assessment authenticity, study approach, and learning outcomes. *Learning and Instruction, 18,*172-186.

Haney, J. J., Wang, J., Keil, C., & Zoffel, J. (2007). Enhancing teachers' beliefs and practices through problem-based learning focused on pertinent issues of environmental health sciences. *Journal of Environmental Education, 38*(4), 25-33. doi: 10.3200/JOEE.38.4.25-33

Heflebower, T. (2010, April). *Formative assessment and standards-based grading.* Conference presentation for Solution Tree, Albuquerque, NM.

Herman, J. (1997). Assessing new assessments. *Theory into Practice, 36*, 196-204. Retrieved from http://ehe.osu.edu/tip/order

The intelligence we never measure. (2007, January). *Roeper Review, 29*(2), 83-84. doi: 10.1080/02783190709554390

Ivanitskaya, L.,Clark, D., Montgomery, G., & Primeau, R. (2002). Interdisciplinary learning: Process and outcomes. *Innovative Higher Education, 27*(2), 95-111. Retrieved from http://www.uga.edu/ihe/ihe.html

Jacobs, H. H. (1989). The growing need for interdisciplinary curriculum content. In H. H. Jacobs (Ed.), *Interdisciplinary Curriculum: Design and Implementation.* Alexandria, VA: Association for Supervision and Curriculum Development.

Johnson, L. (2008). Relationship of instructional methods to student engagement in two public high schools. *American Secondary Education. 36*(2), 69-87. Retrieved from http://www.ashland.edu/ase

Kember, D., Ho, A., & Hong, C. (2008). The importance of establishing relevance in motivating student learning. *Active learning in higher education, 9*(3), 249-263. doi: 10.1177/1469787408095849

Klein. J. T. (1990). *Interdisciplinarity: History, theory and practice.* Detroit: Wayne State University Press.

Kovalik, S. (1993). *ITI: The Model: Integrated Thematic Instruction.* Village of Oak Creek, AZ: Books for Educators.

Koretz, D. (2008). *Measuring Up: What Educational Testing Really Tells Us.* Cambridge, Massachusetts: Harvard University Press.

Lee, V. E., Bryk, A. S., & Smith, J. B. (1993). The organization of effective secondary schools. *Review of Research in Education, 19*, 171-267. Retrieved from http://rre.sagepub.com

Letterman, M. R. & Dugan, K. B., (2004). Team teaching a cross-disciplinary honors course: Preparation and development. *College Teaching, 52*(2):76-79.

Linnenbrink, E. A., & Pintrich, P. R. (2002). Motivation as an enabler for academic success. *School Psychology Review, 31*(3), 313-327. Retrieved from http://www.nasponline.org/publications/spr/index-list.aspx

Marzano, R. J. (2010). *Formative assessment and standards-based grading.* Bloomington, IN: Marzano Research Laboratory.

Marzano, R. J. (2006). *Classroom assessment and grading that work.* Alexandria, VA: Association for Supervision and Curriculum Development.

Marzano, R. J. & Kendall, J. S. (2007). *The new taxonomy of educational objectives.* Thousand Oaks, California: Corwin Press.

McGehee, J. J. (2001). Developing interdisciplinary units: A strategy based on problem solving. *School Science and Mathematics, 101*(7), 380-389. doi: 10.1111/j.1949-8594.2001.tb17972.x

McNamar, M. M. (2009). Using a real life contract bid for students to learn mathematics. *Journal of Instructional Psychology, 36*(2), 142-147. Retrieved from http://www.projectinnovation.biz/jip_2006.html

McTighe, J. & Wiggins, G. (2004). *Understanding by Design: Professional Development Workbook.* Alexandria, VA: Association for Supervision and Curriculum Development.

Meaghan, D.E. & Casas, F. R. (1995). On the testing of standards and standardized achievement testing: Paneca, placebo, or Pandora's box? *Interchange, 26*(1), 33-58.

Mitchell, M. (1993). Situational interest: Its multifaceted structure in the secondary school mathematics classroom. *Journal of Educational Psychology, 85*,424-436. doi: 10.1037/0022-0663.85.3.424

Moon, T. R., Brighton, C.M., Callahan, C.M., & Robinson, A. (2005). Authentic assessments for middle school. *The Journal of Secondary Gifted Education, XVI*(2/3), 119-133.

Moon, T. R., Callahan, C. M., Brighton, C. M., & Tomlinson, C. A. (2002). Development of differentiated performance assessment tasks for middle school classrooms (Report No. RM02160). Washington D.C.: National Research Center on the Gifted and Talented. Retrieved from http://www.gifted.uconn.edu

Nation, M. L. (2008). Project-based learning for sustainable development. *Journal of Geography, 107*(3): 102-111. doi: 10.1080/00221340802470685

Oregon Department of Education (2011). *The Scoring Process*. Retrieved from http://www.ode.state.or.us/teachlearn/subjects/elarts/writing/assessment/scoringprocess.pdf

Palm, T. (2008). Performance assessment and authentic assessment: A conceptual analysis of the literature. *Practical Assessment, Research and Evaluation, 13*(4), 1-11. Retrieved from http://pareonline.net

Palmer, J. (1998). *Environmental Education in the 21st Century: Theory, Practice, Progress and Promise.* London, England: Routledge.

Paris, S. G., Roth, J. L., & Turner, J.C. (2000). Developing disillusionment: Students' perceptions of academic achievement tests. *Issues in Education, 6*(1/2), 17-46.

Pederson, S. (2003). Motivational orientation in a problem-based learning environment. *Journal of Interactive Learning Research, 14*(1), 51-77. Retrieved from http://www.aace.org/pubs/jilr

Pedersen, S., & Williams, D. (2004). A comparison of assessment practices and their effects on learning and motivation in a student-centered learning environment. *Journal of Educational Multimedia and Hypermedia, 13*(3), 283-306. Retrieved from http://www.aace.org/pubs/jemh

Pink, D. H. (2006). *A Whole new mind: Why right-brainers will rule the future.* Toronto, Ontario: Penguin Group.

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory Into Practice, 41*(4), 219-225. doi: 10.1207/s15430421tip4104_3

Price, H. (2008). *Mobilizing the community to help students succeed.* Alexandria, Virginia: Association for Supervision and Curriculum Development.

Principals of learning. (n.d.). Retrieved September 26, 2011, from University of Pittsburgh, Institute for Learning website, http://ifl.lrdc.pitt.edu/ifl/index.php

Proving behaviors in Bloom's taxonomy. (n.d.). Retrieved September 26, 2011, from Paramount Unified School District website, http://www.paramount.k12.ca.us/schools/hollydale/resources.html

Real Curriculum (2006). *The Process of Unit and Instructional Design.* Unpublished diagram.

Reese, S. (2002). Contextual teaching and learning. *Techniques. 77*(1), 40-41.

Reeves, D. B. (2002). *The leader's guide to standards: A blueprint for educational equity and excellence.* San Francisco, CA: Jossey-Bass.

Resnick, L. B. (1987). Education and learning to think. Washington, D.C.: National Academy Press.

Resnick, L. B. (1999). Making America smarter. *Education Week Century Series, 18*(40), 38-40.

Resnick, L. B. & Resnick D.P. (1992). Standards, curriculum, and performance: A historical and comparative perspective. *Education Researcher, 20*(8), 15-21. doi: 10.3102/0013189X014004005

Roberts, P. L. & Kellough, R. D. (2008). *A guide for developing interdisciplinary thematic units.* Upper Saddle River, NJ: Pearson Prentice Hall.

Rohwer, W. D. & Sloan, K. (1994). Psychological perspectives. In L.W. Anderson & L.A. Sosniak (Eds.), *Bloom's Taxonomy: A forty-year retrospective: Ninety-third yearbook of the National Society for the Study of Education* (pp. 41-63). Chicago, IL: University of Chicago Press.

Rotham, R. (1995). *Measuring up: Standards, assessment, and school reform.* San Francisco: Jossey-Bass.

Rowntree, D. (1982). *A dictionary of education.* Totowa, NJ: Barnes & Noble Books.

Sergiovanni, T. J. (2007). *Rethinking leadership: A collection of articles*. Thousand Oaks, CA: Corwin Press.

Stokking, K., Van der Schaaf, M., Jaspers, J. & Erkens, G. (2004). Teachers' assessment of students' research skills. *British Educational Research Journal, 30*(1), 93-116. Retrieved from http://www.tandf.co.uk/journals

Steinmayr, R. & Spinath, B. (2009). The importance of motivation as a predictor of school achievement. *Learning and Individual Differences, 19*, 80-90. doi: 10.1016/j.lindif.2008.05.004

Suurtamm, C.A. (2004). Developing authentic assessment: Case studies of secondary school mathematics teachers' experiences. *Canadian Journal of Science, Mathematics and Technology Education, 4*(4), 497-513. Retrieved from http://connection.ebscohost.com

Tanner, D. E. (2001). Authentic assessment: A solution or part of the problem? *High School Journal, 85*(1), 24-30. Retrieved from http://muse.jhu.edu/journals/high_school_journal

Tchudi, S., & Lafer, S. (1996). *The interdisciplinary teachers' handbook: Integrated teaching across the curriculum.* Portsmouth, NH: Boynton/Cook Publishers.

Thompson, S. (2001). The authentic standards movement and its evil twin. *Phi Delta Kappan, 82*(5), 358-362. Retrieved from http://www.kappanmagazine.org

Trepanier-Street, M., McNair, S., & Donegan, M. (2001). The views of teachers on assessment: A comparison of lower and upper elementary teachers. *Journal of Research in Childhood Education, 15*(2), 234-241. doi: 10.1080/02568540109594963

United States Department of Education. (1983, April). A nation at risk: The imperative for educational reform. Retrieved May 26, 2011, from http://www2.ed.gov/pubs/NatAtRisk/index.html

United States Department of Education. (1998, April). Goals 2000: Reforming education to improve student achievement. Retrieved from http://www2.ed.gov/pubs/G2KReforming/index.html

United States Department of Education. (2001). Public law print of PL 107-110, the No child left behind act of 2001. Retrieved May 26, 2011, from http://www2.ed.gov/policy/elsec/leg/esea02/107-110.pdf

Walden, L. M., & Kritsonis, W. A. (2008). The impact of the correlation between the No Child Left Behind Act's high stakes testing and the high drop-out rates of minority students. *National Journal for Publishing and Mentoring Doctoral Student Research, 5*(1), 1-6. Retrieved from http://www.nationalforum.com

White, C. S. (2007). Levels of understanding—A guide to the teaching and assessment of knowledge. *Journal of Education for Business, 82*(3), 156-162. doi: 10.3200/JOEB.82.3.159-163

Wiggins, G. (1989). Teaching to the authentic test. *Educational Leadership, 48*(4), 41-47. Retrieved from http://www.ascd.org/Default.aspx

Wiggins, G. (1990). The case for authentic assessment. *Practical Assessment, Research & Evaluation*, 2(2). Retrieved from http://pareonline.net

Wiggins, G., & McTighe, J. (2005). *Understanding by design.* Alexandria, VA: Association for Supervision and Curriculum Development.

Wood, T. & Sellers, P. (1997). Deepening the analysis: Longitudinal assessment of a problem-centered mathematics program. *Journal of Research in Mathematics Education, 28*,163-18. doi: 10.2307/749760

Appendix A: Proving behaviors

Bloom's Taxonomy: Proving Behaviors

Cognitive Level	Description	Behavior
Knowledge	Recalling facts and information	Labeling, listing, memorizing, matching, underlining, orally stating, filling in the blank, pointing, repeating
Comprehension	Showing understanding of acquired knowledge	Explaining, describing, paraphrasing, summarizing, rewriting in your own words, retelling, reporting, translating, discussing
Application	Adapting/applying known information	Computing, using, making, solving, demonstrating, operating, illustrating, constructing, drawing, calculating
Analysis	Breaking material down into component parts	Analyzing, categorizing, classifying, comparing, contrasting, developing, deducing, diagnosing, examining, specifying
Synthesis	Putting information together in a new way	Changing, composing, constructing, creating, designing, formulating, generating, inventing, producing, revising
Evaluation	Judging the outcome	Appraising, comparing, deciding, defending, evaluating, judging, prioritizing, supporting

Appendix B: Criteria for authentic assessment

Criteria for Authentic Assessment

Authentic Assessment Criteria
Has value or meaning beyond the classroom
Focuses on big ideas or concepts of a discipline(s): depth vs. breadth
Involves research or active use of conceptual knowledge
Pushes students towards more advanced use of skills and conceptual knowledge
Counts what students "do"
Presents transparent performance criteria, scores according to this criteria
Requires integration of knowledge from overlapping concepts or disciplines
Public; requires an audience
Requires collaboration (with peers, professionals, or community members)
Allows for student choice
Allows for consideration of student learning styles, aptitudes and interest

Appendix C: Professional development intervention timeline

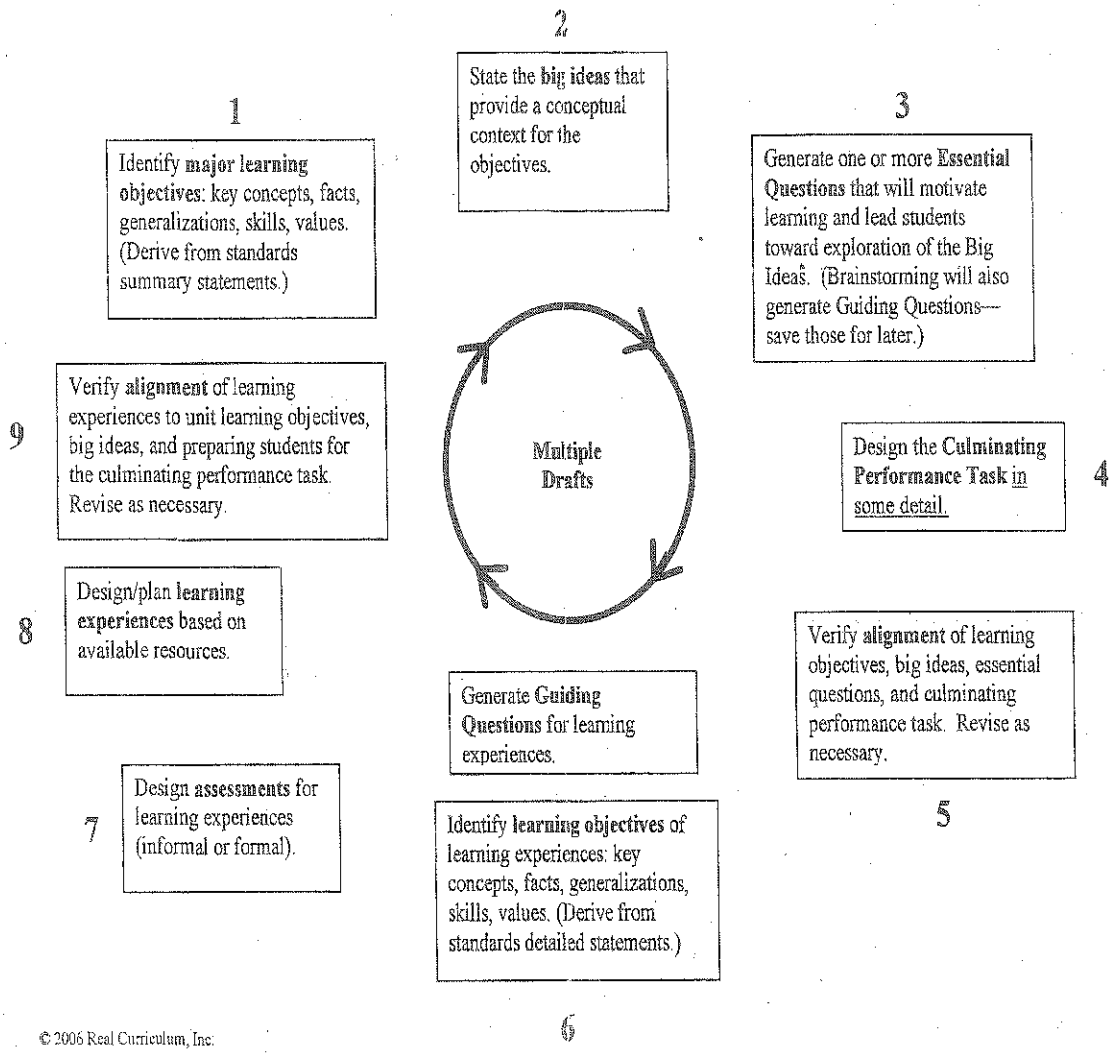
Professional Development Intervention Timeline

Timeline	Topic/Explanation
Ongoing: year to two years prior	Development and revision of "power standards," from state standards and expected school-wide learning results (Reeves, 2002)
Year prior: half day	Individual and team debrief of previous interdisciplinary units
Summer prior: half day	Workshop: defining authentic assessment and "scaling activity" (Marzano, 2010; Wiggins & McTighe, 2005; Solution Tree, 2010)
Summer prior: half day	Use of power standards to choose big ideas for interdisciplinary units (Erickson, 2002; Wiggins & McTighe, 2005)
Summer prior: one day	Identification and development of overlapping concepts, big ideas, and essential questions (Wiggins & McTighe, 2005; see Appendix D)
Semester prior: ongoing, twice a month	Collaborative development, reflection, planning of interdisciplinary authentic assessments and rubrics
Month or two prior: one day	"Assessment Revision Protocol" (see Appendix E) for collaborative reflection and feedback on assessments/rubrics from one grade level to another; team revision of assessments
Month prior: one day, plus individual	Planning of individual instructional sequences to build student knowledge and skills around interdisciplinary concepts and planned assessment

Appendix D: Planning framework

Real Curriculum

The Process of Unit Instructional Design



© 2006 Real Curriculum, Inc.

Appendix E: Assessment revision protocol

ASSESSMENT REVISION PROTOCOL

- Provide copies of the following to a colleague or to your subject or grade level team: Learning Targets /Objectives, Assessment Task, Rubric or Scoring Guide, Unit Plan
- Verbally provide a 2-minute overview of your objectives and how the assessment will be conducted without making any evaluative comments. Answer any clarifying questions.
- Allow time for the panel to review the above documents, individually complete the survey, and gather additional thoughts.
- Have the panel share thoughts with the group, to give feedback that will inform teaching and learning. Have everybody take notes using the questions below. Return summarized notes to presenter(s).

Question	Response			Comments
a. Does the benchmark capture the main purpose of the Learning Targets being assessed (including standards and/or ESLRs)?	Yes	No	Sort Of	
b. Does the benchmark clearly provide an opportunity for students to demonstrate proficiency on specific Learning Targets?	Yes	No	Sort Of	
c. Does the task have sufficient complexity and rigor? Are all elements of the task grade-level appropriate?	Yes	No	Sort Of	
d. Does the assessment provide enough opportunity for students at ALL levels to demonstrate varying levels of skills/knowledge?	Yes	No	Sort Of	
e. Does the assessment allow students to demonstrate higher-order understandings and abilities? (think about Bloom's taxonomy and conceptual levels of tasks and/or understandings)	Yes	No	Sort Of	
f. Does the benchmark incorporate environmental principles?	Yes	No	Sort Of	
g. Does the benchmark clearly and objectively express the grading criteria in a well-defined rubric?	Yes	No	Sort Of	
h. Does the rubric clearly measure individual performance of defined Learning Targets?	Yes	No	Sort Of	
i. Are all of your Learning Targets represented and measured by your rubric?	Yes	No	Sort Of	
j. Are all components of your rubric aligned to your Learning Targets?	Yes	No	Sort Of	
k. Does the benchmark avoid bias or places where the assessment task could unfairly penalize certain students?	Yes	No	Sort Of	
l. If available, does the Unit Plan appear to provide instruction and formative feedback to help students work towards proficiency of these Learning Targets?	Yes	No	Sort Of	

Overview of Interdisciplinary Authentic Assessments for Each Grade

Grade/Project	Description	Related Courses
Ninth: Sustainable College Campuses	Students will design, construct, present, and justify two- and three-dimensional college campus models using math, aesthetics, and sustainability.	English, Math, Art, Environmental Science, College-Prep
Tenth: Debates on Progress	Students will debate the measurement of "progress" within imperialism, the industrial revolution, sweatshops, the Port of Los Angeles, multilingualism, genetic engineering, genetically-modified foods, and carrying capacity.	English, World History, Math, Biology, "Green Ambassadors"
Eleventh: American Dream Websites	Students will create and present websites about how to enhance the sustainability of the American Dream in terms of a) environment, b) economics, and c) equality for all.	English, United States History, Math, Science, College Prep
Twelfth: Social Justice Newscasts	Students will research a societal issue, conduct surveys, collect and analyze correlation data, defend conclusions, and recommend actions regarding their a social injustice in the community.	English, Government, Math, Science, Senior Seminar

Appendix G: 9th grade assessment and rubric

9th Grade Sustainable Design Project
Interdisciplinary Unit Exam

TASK:

Create a written proposal explain the pros and cons of your building ideas. Within your proposal you will explain how your building was made, the advantages to your building, why it should be built and you will display the knowledge you have learned in each of your classes. Use posted examples of professional proposals for ideas on layout, tone, content, and use of graphics when writing your proposal.

In writing the proposal your goal is to convince a panel of prospective clients to buy your idea and build your campus building.

PROCESS:

1. Collect information about your building design, the design's advantages and information from each of your 5 classes about how green buildings affect my quality of life and the community.

2. Write two, two chunk paragraph organizers about each of the following topics.

    a. **Campus Building Challenge**. What is it? What are you being asked to create during Intersession?

    b. **Details of your building**. What does it look like? Where on the campus is it located? If the building was to be build to size, what materials would be used?

    c. **Resources and Sustainability**. What effect does my selection of materials and natural resources have on local and global communities? How can research and planning improve sustainability? What role does conservation and sustainability play in the cost of a building and operating a campus structure?

    d. **Building Scale**. What scale was used to create your building model? How do I use math to describe the world around me? How do rational numbers operations connected to problem-solving in real life? How does creating and using models made to scale affect the sustainability of my community?

    e. **Building Artistic Contributions**. How does your building make an artistic contribution to the community? What risks were taken in creating

an aesthetically pleasing building? How can taking (appropriate) risks enhance your quality of life? How does effort, energy and work ethic relate to quality of life?

f. **Group and Personal decision making to building design.** How did your group come together to make decisions on the building design? How can you connect decisions made toward building design to your every day decision making? How can I use data to support or refute my decisions? Do my decisions have positive or negative outcomes? How do I make the right decisions? How do my decisions affect my community?

g. **Building funding**. If necessary how could your group create funding to create your building? How could investment in stocks provide for enough funding to create your building? Are you able to prove numerically and rationally why your choice of stocks will lead to positive gains?

3. Write two, two chunk paragraphs summarizing the intersession and what it taught you about quality of life.

4. Now that the rough draft is complete type a final draft of your essay. Make sure to watch out for grammar mistakes and typos.

5. Create a cover page with company logo and names of members; include section headings and a table of contents.

9th Grade Sustainable Design Project Rubric

GROUP SCORE		MASTERY - 4	PROFICIENT - 3	PROGRESSING - 2	BASIC - 1
	3D Model	▪ Each member has 100% participation in 3D model. ▪ Model has at least 3 examples of Bio-Mimicry ▪ Model has an expert level of craftsmanship, creativity, and risk.	▪ Each member has 100% participation in 3D model. Model has at least 2 examples of Bio-Mimicry ▪ Model has a high level of craftsmanship, creativity, and risk.	▪ Members have 50% participation in 3D model. ▪ Model has at least 1 examples of Bio-Mimicry ▪ Model demonstrates some level of craftsmanship, creativity, and risk.	▪ Members have 50% or less participation in 3D model. ▪ Model does not have any examples of Bio-mimicry. ▪ Model does not show enough craftsmanship, creativity, and/or risk.
	Schematic/Plan	▪ Each group has a highly original, creative and aesthetically pleasing Building schematic/plan. ▪ Plan has 3 elements overhead site, elevation, and 3D concept.	▪ Each group has an original Building schematic /plan. Plan has 3 elements overhead site, elevation, and 3D concept.	▪ Each group has a Building schematic /plan. Plan has only 2 elements: overhead site, or elevation, or 3D concept.	▪ Each group has a Building schematic /plan. Plan only has 1 element: overhead site, or elevation, or 3D concept.
	Delivery & Engagement	▪ Effective tone, voice projection, enunciation and eye contact most of the time. ▪ Effective pace most of the time. ▪ Displays energy and enthusiasm	▪ Acceptable tone, voice projection, enunciation and eye contact. ▪ Pace varies. ▪ Displays some energy and enthusiasm.	▪ Needs to improve tone, voice projection, enunciation, eye contact or pacing. ▪ Limited energy and enthusiasm.	▪ Tone, voice projection, enunciation, eye contact are limited or not effective. ▪ Not paced well. ▪ Lacks energy and enthusiasm.
	Professional	▪ Proper body language, posture and/or hand gestures. ▪ Dress pants or knee-length dress/ skirt, collared shirt (not tight, baggy or revealing), dress shoes	▪ Proper body language, posture and/or hand gestures. ▪ Dress is appropriate, but a little more casual.	▪ Casual body language, lacks posture and/or hand gestures. ▪ Made some effort to dress professionally, but is missing required articles.	▪ Body language, posture, and/or hand gestures not professional. ▪ Wearing uniform or violating uniform and must be sent to the Dean.

9th Grade Sustainable Design Writing Rubric

		MASTERY - 4	PROFICIENT - 3	PROGRESSING - 2	BASIC - 1
Environmental Design: Art and Science	**Green Architectural space. College Building**	Student can clearly analyze, discuss and rationalize green design and the aesthetics of bio mimicry as it relates to their project. Essay must include at least 6 art vocabulary words.	Student can analyze and discuss green design and the aesthetics of bio mimicry as it relates to their project. Essay must include at least 4 art vocabulary words.	Student can discuss green design and bio mimicry as it relates to their project. Essay must include at least 2 art vocabulary words.	Student cannot discuss green design or bio mimicry as it relates to their project. Art vocabulary not present
	Data is used to support design.	☐ Multiple Authoritative data sources ☐ Careful attribution/citation ☐ Highly relevant and current ☐ Presented in data tables and graphs ☐ Method of analysis is clearly explained ☐ Graphs clearly demonstrate analysis ☐ Analysis and calculation are complete and accurate. ☐ Analysis is applied to design	☐ Presented in data tables and graphs ☐ Multiple data sources ☐ Careful attribution/citation ☐ Graphs illustrate data ☐ Clear analysis	☐ Graphs and tables are relevant careful attribution/citation ☐ More that one data source ☐ Data is discussed in the paper ☐ Some analysis is evident	☐ Graphs and tables are present ☐ Data is discussed in the paper
	Sustainability central to the design:	☐ Integrated into Most elements of the design ☐ Demonstrated with data and analysis. ☐ Compromises are discussed supported with data ☐ Compared to non- sustainable options with data ☐ Evaluates the inaction between quality of life and sustainability	☐ Integrated into Many elements of the design ☐ Demonstrated with data ☐ Compromises are discussed ☐ Compared to non- sustainable options	☐ Integrated into more than 1 element of the design ☐ Reasoning explained clearly (CD)	☐ Integrated into 1 element of the design
	Natural Resource conservation is central to the design, construction, and operation the project.	☐ Explains use of passive design elements. ☐ Explains how local sourcing reduces carbon footprint and cost. ☐ Creative design reduces operation cost with data ☐ Evaluates impact of demolition ☐ Accurately applies principles of conservation ☐ Analyze how the design saves resources (CD and data).	☐ Demonstrates how the design saves resources (CD and data). ☐ Design reduces operation cost ☐ Accurately applies principles of conservation	☐ Explain how the design saves resources (CD). ☐ Design uses stainable materials ☐ Applies principles of conservation	☐ References principles of conservation

Math	**Rates, ratios, and proportions to construct models made to scale.**	Calculations and explanations of proportional reasoning are clear, concise, and thorough.	Calculations and explanations of proportional reasoning are clear and concise.	Calculations and explanations of proportional reasoning are included but are difficult to understand.	Calculations and explanations of proportional reasoning are not included and/or lack significant detail.
	Use appropriate expressions (signs) to solve real world problems (word problems) and justify individual process using correct terminology.	Calculations and Explanations of investment choices are completely justifiable and proven by thorough analysis of annual reports and internet research.	Calculations and Explanations of investment choices are justifiable and proven by an adequate analysis of annual reports and internet research.	Calculations and Explanation of investment choices have been researched but are not proven justifiable.	Calculations and Explanation of investment choices lack any meaningful research.
Decision Making	**Collaborate, challenges and quality decision making.**	☐ Design decisions are thoroughly justified by the use of the 4 steps of the decision making model. ☐ Describes how decisions were reached based upon values and how challenges were overcome. ☐ Considers realistic positive and negative effects of each decision.	☐ Design decisions are justified by the use of the 4 steps of the decision making model. ☐ Describes how decisions were reached and challenges were discussed. ☐ Considers some positive and negative effects of each decision.	☐ Design decisions are justified by the using parts of the decision making model. ☐ Describes decisions and challenges. ☐ Considers effects of each decision.	☐ Design decisions attempt to be justified by the using parts of the decision making model. ☐ Attempts to describe decisions and challenges. ☐ Considers effects of decisions.
Writing / Structure	**Original Composition and Writing Structure**	The paper will contain a Thesis that is clearly written and addresses the prompt; including 3 Major points an opinion and a subject. The paper will also contain Body Paragraphs made up of Three or more Concrete Details introduced in context and Six or more Commentary sentences are used to explain the Concrete Details. Commentary sentences are used to analyze and synthesize Concrete Details as they relate back to the topic sentence.	The paper will contain a Thesis that is clearly written and addresses the prompt; including 3 Major points, an opinion and a subject. The paper will also contain Body Paragraphs made up of Two Concrete Details introduced in context and Four Commentary sentences are used to explain the Concrete Details. Commentary sentences are used to analyze Concrete Details as they relate back to the topic sentence.	The paper will contain a Thesis that is clearly written and addresses the prompt; including 3 Major points, an opinion and a subject. The paper will also contain Body Paragraphs made up of Two Concrete Details introduced in context and Four Commentary sentences are used to explain the Concrete Details.	The paper will contain a Thesis that is clearly written and addresses the prompt; including 3 Major points, an opinion and a subject. The paper will also contain Body Paragraphs made up of less then Two Concrete Details introduced in context and less then Four Commentary sentences are used to explain the Concrete Details.
	Grammar/ Word Choice		Grammatical errors do not interfere with understanding and sentences are logically constructed with minimal errors in use of Standard American English.		Grammatical errors interfere with understanding and sentences are not logically constructed.

Appendix H: 10th grade assessment and rubric

10th Grade Interdisciplinary Unit Exam
What is progress and how is it measured?

During Intersession you have studied a variety of issues related to the idea of progress and how it is measured.

Question: Do the changes we have studied in intersession represent more of a positive or negative change? In other words have the changes we studied helped or hindered progress?

Please answer the question using evidence that addresses 4 of the following groups.

You may omit the group that you covered in your debate topic. (For example if you covered the Industrial Revolution in your debate you may omit Group A).

For Each Group explain

Group A: History
1. The Industrial Revolution: Positive and Negative
2. Imperialism: Positive and Negative

Group B: Green Ambassadors
1. Are Sweatshops Necessary for Progress?
2. Is the Port of LA Helping or hurting progress in Los Angeles?

Group C: Biology
1. Genetic engineering for kids
2. Genetically modified foods

Group D: Spanish
1. What are the positives and negatives of requiring citizens to speak the dominant language of their country?

Group E: Math
Carrying Capacity
1. Have we reached our carrying capacity?

10th Grade Debate Project

What is progress and how is it measured?

	Below Basic - 1	Basic - 2	Proficient – 3	Advanced - 4
History: Industrial Revolution	Student uses provided research to explain either the positive or negative effects of the Industrial Revolution	Student uses provided research to explain the positive and negative effects of the Industrial Revolution.	Student uses provided research to explain and evaluate the positive and negative effects of the Industrial Revolution.	Student adds their own research to explain and evaluate the positive and negative effects of the Industrial Revolution
History: Imperialism	Student uses provided research to explain either the positive or negative effects of the Imperialism	Student uses provided research to explain the positive and negative effects of the Imperialism	Student uses provided research to explain and evaluate the positive and negative effects of the Imperialism	Student adds their own research to explain and evaluate the positive and negative effects of the Imperialism
Green Ambassadors Port of LA	Student uses provided research to explain either the positive or negative effects of the Port of LA on economic and environmental sustainability for Los Angeles.	Student uses provided research to explain the positive and negative effects of the Port of LA on economic and environmental sustainability for Los Angeles.	Student uses provided research to explain, evaluate and advocate for or against the positive and negative effects of the Port of LA on economic and environmental sustainability for Los Angeles.	Student adds their own research to explain and evaluate the positive and negative effects of the Port of LA on economic and environmental sustainability for Los Angeles. Student demonstrates an understanding of the interconnections between social, environmental and economic cost/benefits.
Green Ambassadors: Sweatshops	Student uses provided research to explain either the positive or negative effects of sweatshop manufacturing to long term economic and environmental sustainability.	Student uses provided research to explain the positive and negative effects of sweatshop manufacturing to long term economic and environmental sustainability.	Student uses provided research to explain, evaluate and advocate for or against the positive and negative effects of sweatshop manufacturing to long term economic and environmental sustainability.	Student adds their own research to explain, evaluate and advocate for or against the positive and negative effects of sweatshop manufacturing to long term economic and environmental sustainability. Student demonstrates an understanding of the interconnections between social, environmental and economic cost/benefits.

	Below Basic - 1	Basic - 2	Proficient - 3	Advanced - 4
Spanish: Dominant National Language	Student uses provided research to explain either the positive or negative effects of requiring citizens to speak the dominant language of their country of residence.	Student uses provided research to explain the positive and negative effects of requiring citizens to speak the dominant language of their country of residence.	Student uses provided research to explain and evaluate the positive and negative effects of requiring citizens to speak the dominant language of their country of residence.	Student adds their own research to explain and evaluate the positive and negative effects of requiring citizens to speak the dominant language of their country of residence and uses additional research.
Math: Area, S.A Perimeter, & Volume	Students cannot make connections between carrying capacity and the development of sustainable technology.	-Students make an attempt to explain how the development of sustainable technology is related to carrying capacity.	-Students can explain with detailed examples how the development of sustainable technology is related to carrying capacity.	- Students can explain with detailed examples and their own research how the development of sustainable technology is related to carrying capacity.
Biology: Genetic Engineering Pros and cons of genetic engineering	Student does not demonstrate understanding of the topic and does not use provided research to accurately support their position	Student uses provided research to explain the positive and negative effects of genetic engineering.	Student uses provided research to create convincing arguments with strong evidence to support their position	Student goes beyond provided research and demonstrates a deep understanding of the topic. Arguments and evidence demonstrate original thinking.
English: Oral Presentation and Persuasion	Student does not demonstrate understanding of topic nor are arguments or evidence convincing. Students do not demonstrate effective presentation and debate skills.	Student may understand topic but do not create convincing arguments or gather convincing evidence. Students will demonstrate effective presentation and debate skills.	Student uses provided research to create convincing arguments with strong evidence that shows solid understanding of topic. Students will create effective and convincing arguments, evidence and rebuttals. Students will demonstrate effective presentation and debate skills.	Student goes beyond provided research and demonstrates a deep understanding of the topic. Arguments and evidence demonstrate original thinking by creating effective and convincing arguments, evidence and rebuttals. Students demonstrate effective presentation and debate skills.

Appendix I: 11th grade assessment and rubric

11th Grade Final Project
How Can the American Dream Be Made Sustainable?

Task: Your task is to form an Advocacy Group centered on answering the above essential question. An Advocacy Group is a group of activists who address political and social issues by identifying problems and developing and implementing solutions. In order to answer the essential question, you will need to answer the following:

1) How can the American Dream respect the **ENVIRONMENT**?
2) How can the American Dream be **ECONOMICALLY** accessible to those who want it?
3) How can the American Dream guarantee **EQUALITY FOR ALL**?

THE ADVOCACY GROUP MUST INCLUDE THE FOLLOWING:

1) An Advocacy Group name and slogan
2) A Mission Statement
    a. Your Mission Statement is an essay that answers the essential question
    b. It is the focus of your Advocacy Group and gives people reason to support you.
3) Statement of the American Dream
    a. In a 2 paragraph essay you should define the American Dream and its changing definitions over time. You should be sure to include the events and people that have influenced the American Dream.
4) Discussion of the issues and solutions
    a. Examine three specific issues (one for each of the Three E's) and create solutions
5) Website
    a. Your group website will incorporate all of the above
    b. Website will be composed & published in xxxxxxxx's class. Go to the following website to see an example: http://www.xxxxxxxxxxxxxx.com

Through this project, you will be demonstrating proficiency in the following Grade 11 Learning Targets:

English	- **Organized & Original Writing:** Students will be able to structure ideas and arguments in a sustained, persuasive, and sophisticated way and support them with precise and relevant examples; students will be able to write in a variety of styles using language in natural, fresh, and vivid ways to establish a specific tone and engage readers. Students' writing will demonstrate command of the language (i.e., grammar, syntax, mechanics, and conventions), college-level typed format, and MLA documentation of resources. - **Public Presentations:** Deliver effective and engaging multimedia and/or oral presentations.
U.S. History	- **Industrialization & Immigration**: Students analyze the relationship among the rise of industrialization, large-scale rural to urban migration, and massive immigration from Southern and Eastern Europe, including working conditions, urbanization, Americanization, immigration, industrial leaders, Social Darwinism, Social Gospel
Chemistry	- **Investigation and Experimentation Standards** 1.d)Select and use appropriate tools and technology to performs tests, collect data, analyze relations and display data. 1.m) Investigate a science-based societal issue by researching the literature, analyzing data, and communicating the findings
Math	- **Power Standard 1:** Linear Equations- Solve problems using tables, graphs, equations, and words. - **Power Standard 2:** Quadratics- Solve problems using tables, graphs, equations, and words.
College Prep	- **College Prep Power Standard 7:** Students will collaborate in groups and demonstrate mastery of essential Critical Skills dispositions when completing challenges: 11th grade: Character, problem solving, management and critical thinking

Section Details

1. Mission Statement:

Each group must have a clear and focused statement that lays out the purpose of your group. All of your solutions to the issues of sustainability should relate back to this Mission Statement.

2. State of the Dream

In this section, each group will discuss current realities of the American Dream and factors that led to these realities.

In this section, students will define the American Dream and discuss its changing definitions over time, including the people and events that have influenced the American Dream. Each group will discuss current realities of the American Dream and factors that led to these realities.

3. Sustainability: Issues and Solutions (3 Sections)

In these sections, each group will present issues related to the sustainability of the America Dream and then propose a solution to these issues. Each of these issues and solutions must be supported by one or more of the types of evidence/details mentioned on the following page. These issues/solutions should be related to the central mission of your advocacy group. For example, you might choose a single topic about which your group is passionate, and then try to connect your three issues/solutions to that.

4. Conclusion: A Vision of Hope:
In this final section, your group will present a brief vision of the American Dream after your solutions have been implemented and how America will be better.

Types of Evidence/Details:
Each of your issues and solutions must be supported using details reflecting the learning from one of your classes. You may use any class as evidence to support any of the issues or solutions, but one argument must be supported with math, one must be supported with science, and at least two must be supported by different social sciences (College Prep, History, English).

Science Evidence: Use facts relating to data or technology as the concrete detail. The commentary expands upon the costs, benefits, and implementation of the concrete detail.

- Evidence/Details: Should be pulled from a technology discussed in xxxxxxxx's class. For example, fire allowed humans to unleash the stored energy in wood for light & heat.
- Commentary: Fire allowed us to cook food, see after sunset, and improved our general health. On the other hand, this forced people to be tied to sources of wood, and began the process of deforestation and pollution.

Math Evidence: Use graphs and numbers as the concrete detail. The commentary discusses the implications of notable trends and predictions.

- Evidence/Details: A graph relating to the topic.
- Commentary: A detailed discussion of the visible trends shown in the graph, identifying specific points, interpreting the changes or differences between different points on the graph, making the connection on what these interpretations have to do with the argument.

Social Sciences Evidence: Use details of historical and cultural events and quotes as the concrete details. Commentary will explain the significance of these details.

- Evidence/Details: Recounts an event or a quote from a relevant figure from Industrial Revolution/Immigration or the riots.
- Commentary: Incorporates the significance and context of the above to support your argument

11th Grade Rubric: How Can the American Dream Be Made Sustainable?

	Advanced - 4	Proficient - 3	Basic - 2	Below - 1
Mission Statement	* Centered around the American Dream, the Mission Statement *thoroughly and effectively* describes the *specific purpose* of the group, its *specific* goals, and how it plans to accomplish goals * 4-6 sentences	* Centered around the American Dream, Mission Statement describes the purpose of group, goals, and how it plans to accomplish goals * 4-6 sentences	* Mission statement describes the purpose and goals in general sense. May not be centered on American Dream. * More than 6 or less than 4 sentences	* Mission statement is bland, gives no specific purpose, is not centered on American Dream. * More than 6 or less than 4 sentences
Definition and state of American Dream	Thorough definition of American Dream, accurate & thorough discussion of both influences and changes	Definition of American Dream, basic discussion of changes/influences	Basic definition of American Dream; simplistic discussion of changes or influences	Inaccurate or insufficient definition of AD and/or influences & changes; too much is missing
Science: Concrete Details	Details mention new or influential technology, its effects, costs, and scientific principles behind it	Details mention a new or influential technology, its effects and information on its costs	Details mention a new or influential technology and some of its effects	Details mention a new or influential technology
Science: Commentary	Benefits, and costs of the technology are effective detailed and a cost-benefit analysis is provided.	Benefits of the technology and effects are directly linked to the solution or issue	Technology is related vaguely to the issue	Technology is not related to the issue whatsoever
Math: Concrete details	An extremely relevant graph is included and significant features and trends are referenced in text.	A graph is included and referenced in detail, including trends and data points	A simple graph is referenced, though features of the graphs may be unmentioned	No graph or unrelated graph
Math: Commentary	Commentary insightfully uses trends and other features of the graph to strongly support the argument.	Commentary relates meaningful features of the graph (slope, trends) to the argument	Commentary relates the graph to the argument, but may miss certain points	Commentary mentions the graph but does not contribute to the argument

Social Science: Concrete Details	Quote or event is extremely relevant and clearly strengthens argument	Quote or event chosen clearly relates to the argument at hand and adequately strengthens argument	Student uses a well documented quote or event, but may not fully relate to the argument and/or only somewhat strengthens the argument	Quote or event does not meet basic standards
Social Science: Commentary	Commentary explains the context and background needed to understand the concrete detail's significance	Commentary explains the context and background, but does not discuss significance	Commentary merely summarizes the evidence	Commentary is irrelevant
Writing & Language	Writing is eloquent and fluid, organized in logical "chunk" style, using original academic vocabulary	Writing is organized in logical "chunk" style, using academic vocabulary	Writing has basic organization; points may not be clear; academic vocabulary not used; mechanics, grammar, spelling, punctuation may be poor.	Writing is poor, unorganized, does not use "chunk" style; points unclear; mechanics, grammar, spelling, punctuation is poor
Website	- Layout is eye-catching & unique, maintains professionalism while enhancing content - Images are powerful & unique, enhance content paragraphs - Site design allows users to easily locate any information; pages cross-referenced for legibility	- Layout is neat & orderly with appealing colors; complements your platform - Images make a strong statement; complement work but may not add value - Site design allows users to easily search for information, but no cross-referencing	- Layout is organized but not easy to use; color scheme may be unprofessional or distracting - Images are fairly bland and banal; use of images is forced rather than flowing from the material - Site design requires scrolling or searching to locate specific information; difficult to search	- No use of color or particular layouts; color, if any, seems to detract rather than add - Images are nonsensical or absent; images clash with meaning or displayed inappropriately - Site design is unorganized; specific information can only be located by reading the whole site

Appendix J: 12th grade assessment and rubric

12th Grade Project: How Am I Powerful?

The Investigative Reporting Challenge

The head executives of Global News Network have approached your news magazine to assemble a news package addressing current issues in the Los Angeles metro area for a highly-touted prime time slot during the 2012 Super Bowl. This broadcast time is worth over $25 million.

Specifically, Global News Network is looking for a 20-minute investigative report with a collection of vignettes on these current issues. Each beat reporter must submit a news story that runs for at least 2 minutes on the issue you have been researching for your senior thesis. The story must include at least three soundbites that show the following: an expert opinion, public opinion & an analysis of your survey results (stand up). You will film your stand up at the end of this month in xxxxxxxxxx's class with the results of your survey displayed on a graph behind you. You will have 30 seconds to clearly explain the graph and the correlation or lack thereof. Additionally, the story must include related images to your topic (B-roll footage) and voice overs to guide the audience through your news piece.

Once your 2-minute vignette is completed, you must work with other reporters in your news station to assemble the pre-recorded 20-minute newscast. The stories must be edited together, thoughtfully organized (stacked), and flow with an objective and a purpose. You will need to designate roles amongst your news crew, i.e., lead anchor(s), cameraperson(s), editor(s), assignment director, wardrobe, producer(s). The newscast will also need to include your call numbers, throws to anchors/reporters, graphics & a creative background.

Between each 2 minute vignette anchors should conclude the story and introduce the next segment, which can include weather, sports, entertainment, a commercial break, and/or the next vignette. Be creative! Your commercials can consist of advertisements or public service announcements for your service learning organizations or upcoming events on xxxxxx's campus.

The final newscast will be shown to the panel of Global News Network executives on _____ at 8 am. You are competing for the network's prime time spot amongst fierce competition. The best news magazine, as determined by the panel, will receive the prime time slot and bonus points!

12th Grade News Cast Rubric ~ How Am I Powerful?

	Excelling - 4	Proficient - 3	Progressing - 2
Content and skills	▪ Information presented is clear, concise and enhances viewer insight into the topic that they may not have known if they had not seen the video piece ▪ Stand up clearly communicates analysis of the survey results and mentions major assumptions, limitations, etc. ▪ Contains 2 sound bites from an expert, and public opinion that ▪ Are transitioned seamlessly into the news story ▪ Graph is accurate and properly labeled, easy to understand visually without any external prompting ▪ Demonstrates at least 3 qualities of a strong investigative reporter ▪ Proper enunciation and academic language is used effectively and sustained throughout	▪ Information presented is appropriate, and gives a clear view of the thesis topic ▪ Stand up communicates the correlation, if any, clearly and analyzes the survey results ▪ Contains 2 sound bites from an expert, and public opinion that ▪ contributes to the news story ▪ Graph is appropriate, accurate, and properly labeled, easy to understand visually ▪ Attempts at least 3 qualities of a strong investigative reporter ▪ Proper enunciation and academic language is used effectively	▪ Information presented is not relevant, and does not give a clear view of the thesis topic ▪ Stand up does not communicate your analysis of the survey results ▪ Contains less than 2 sound bites and/or they are unrelated or interrupt flow of the story ▪ Graph is not appropriate, accurate, or properly labeled; has a major flaw that inhibits understanding ▪ Does not attempt to demonstrate qualities of a strong investigative reporter ▪ Does not use proper enunciation and academic language effectively
Organization of video	▪ Each segment of the news cast flows smoothly from one story to the next with creative throws from the anchors and reporters ▪ Anchor lead ins are engaging and preview the upcoming story in an original manner ▪ The news magazine contains a clear and creative beginning and ending. ▪ Includes commercials that are creative and advertise service learning organizations and use propaganda techniques	▪ Each segment of the news cast flows from one story to the next with throws from the anchors and reporters ▪ Anchor lead ins are engaging and preview the upcoming story ▪ The news casts contains a clear beginning and ending. ▪ Includes commercials that are creative and advertise service learning organizations	▪ Segment transitions are choppy ▪ Anchor lead ins are not engaging and/or do not preview the upcoming story ▪ The news casts does not contain a clear beginning and ending. ▪ Commercials are not creative and uninspiring.
Video Quality	▪ The technical elements (text, images, sound) of the video enhances the student's ability to demonstrate understanding ▪ Voiceover enhances the meaning of images on screen (b-roll)	▪ The technical elements (text, images, sound) of the video does not get in the way of student's ability to demonstrate understanding ▪ Voiceover relates to images on screen (b-roll)	▪ The technical elements (text, images, sound) of the video gets in the way of the student's ability to demonstrate understanding ▪ Voiceover does not relate to images on screen (b-roll)

Appendix K: Informed consent for participation in research: Teachers

Participant/ Teacher: _____

Principal Investigator: Jenni Taylor, Doctoral Student in Pepperdine University's Graduate School of Education and Psychology.

Faculty Advisor: Dr. Christopher Lund, Faculty Member at Pepperdine University's Graduate School of Education and Psychology.

Title of Project: *Interdisciplinary Authentic Assessment: Cognitive Expectations and Student Performance*

1. I, _____ (Teacher Name), agree to participate in the research study being conducted by Jenni Taylor under the direction of her advisor Dr. Christopher Lund.

2. The overall purpose of this study is to expand current limited research regarding the cognitive expectations and outcomes of interdisciplinary authentic assessment. The study will examine the cognitive complexity of learning objectives within authentic assessment tasks and will also examine levels of student performance resulting from these tasks. The study will also explore potential relationships between the cognitive complexity of expectations and related student performance. A panel of assessment experts will be categorizing the learning objectives of your assessments into Bloom's Taxonomy. The researcher will also summarize students' scores and compare results according to cognitive complexity.

3. Participants at this school have been selected for this study, due to the unique approach to interdisciplinary authentic assessment at this site. As a participant, I have been involved in the creation and/or implementation of an interdisciplinary authentic assessment task with my grade level team. I have also participated with a panel of teachers in the scoring of this assessment using a team-created rubric. My participation in the study, which will take place over the next month, will involve sharing our grade level interdisciplinary assessments, allowing the researcher to use the assessments to analyze cognitive demand, and sharing scores of student participants from the associated rubrics, without any identifying student information.

4. My participation in the study will not require me to do anything other than share the assessments created by my teaching team and student performance results on rubrics, without any identifying student information.

5. I understand that the possible benefits to myself or society from this research are: a deeper societal understanding of assessment design and related cognitive performance, and the potential this understanding may have for the field of education and curricular reform in schools and for future students.

6. I understand that there are certain risks and discomforts that might be associated with this research. These risks or discomforts include: the perception of potential judgment of the school or researcher regarding assessment implementation or student performance.

7. I understand that I may choose not to participate in this research.

8. I understand that my participation is voluntary and that I may refuse to participate and/or withdraw my consent and discontinue participation in the project or activity at any time without penalty or loss of benefits to which I am otherwise entitled.

9. I understand that the investigators will take all reasonable measures to protect the confidentiality of my records and my identity will not be revealed in any publication that may result from this project. The confidentiality of my records will be maintained in accordance with applicable state and federal laws. Under California law, there are exceptions to confidentiality, including suspicion that a child, elder, or dependent adult is being abused, or if an individual discloses an intent to harm him/herself or others.

10. I understand that the investigator is willing to answer any inquiries I may have concerning the research herein described. I understand that I may contact Jenni Taylor (xxxxxxxx@xxxxxxxx.xxx) or her supervisor, Dr. Christopher Lund (xxxxxxxx@xxxxxxxx.xxx) if I have other questions or concerns about this research. If I have questions about my rights as a research participant, I understand that I can contact the Chairperson of the Graduate and Professional Schools Institutional Review Board (GPS IRB), Dr. Yuying Tsong, at xxx-xxx-xxxx or at xxxxxxxx@xxxxxxxx.xxx.

11. I understand to my satisfaction the information regarding participation in the research project. All my questions have been answered to my satisfaction. I have received a copy of this informed consent form which I have read and understand. I hereby consent to participate in the research described above.

_____ _____

Participant's Signature Date

I have explained and defined in detail the research procedure in which the subject has consented to participate. Having explained this and answered any questions, I am cosigning this form and accepting this person's consent.

_____ _____

Principal Investigator Date

Appendix L: Informed consent cover letter: Parents

Greetings Parents,

As many of you may know, I am currently in the process of completing my doctoral studies at Pepperdine University. For my dissertation research, I am hoping to learn more about the complexity of the work students and teachers at xxxxxxxx do every year. Specifically, I am studying the cognitive demands of interdisciplinary authentic assessments and the related student performance.

I would like your permission to include your student's scores on their Intersession assessments in my study. Students will not have to do anything more than they already do for their classes and their participation in this research project is completely voluntary.

Attached to this letter you will find a consent form which will allow your student's scores to be used for this research project. No student scores will be identified by name in the study, nor will the school name be used.

If you any questions, please do not hesitate to call or email me at xxxxxxxx@xxxxxxxx.xxx or (xxx) xxx-xxxx. I will also host a question and answer session on _____ at _____ in the school library.

Thank you,

Jenni Taylor
Doctoral Student at Pepperdine University

Estimados Padres,

Como muchos de ustedes pueden saber, estoy actualmente en el proceso de completar mis estudios doctorales en la Universidad de Pepperdine. Para mi investigación de disertación, yo espero aprender más acerca de la complexidad del trabajo que hacen los estudiantes y maestros de xxxxxxxx todos los años. Específicamente, estudio las demandas cognoscitivas de evaluaciones auténticas interdisciplinarias y el desempeño que muestran los estudiantes

Querría que su permiso en incluir los resultados de su estudiante en sus evaluaciones de Intercesión en mi estudio. Los estudiantes no tendrán que hacer nada más que ellos ya hacen para sus clases y su participación en este proyecto de investigación es completamente voluntario.

Incluida en esta carta usted encontrará una forma de consentimiento que permitirá que los resultados de su estudiante sean utilizados para este proyecto de investigación. Ningunos resultados de ningún estudiante serán identificadas por nombre, ni por el estudio, ni el nombre de la escuela será utilizada.

Si usted tiene alguna pregunta, por favor no dude en llamar me o mandarme un correo electrónico a xxxxxxxx@xxxxxxxx.xxx o llámeme al (xxx) xxx-xxxx. Yo también acogeré una sesión de pregunta y respuesta él _____ a las _____ en la biblioteca de la escuela.

Gracias,

Jenni Taylor
Estudiante de Doctorado en la Universidad de Pepperdine

Appendix M: Assent forms for use with minors

Dear Student,

My name is Jenni Taylor, a student in the Educational Leadership Doctoral Program at Pepperdine University. Your parents have given me their permission to allow you to participate in a study, similar to your upcoming senior thesis assignment! I would like to invite you to participate in this study if you are interested.

This study is a research study, which I am conducting for the purposes of completing a doctoral degree in Educational Leadership. Before I explain more about the study, I want you to know that the choice to participate is completely voluntary. No one is going to force you to do something you are not interested in doing. The refusal to participate or discontinuing participation at any time will not involve any penalty or loss of benefits to which you are entitled, and will certainly not affect any of your grades or your status as a student.

To participate in this study, you don't have to do anything if you have completed the assessments for your Intersession Term (your Interdisciplinary Unit Exams and/or your Community Forum Presentation). Your rubric scores for these assessments will be used to measure how well the overall student body does on each part of their assessments.

Your scores will be kept confidential and the researchers will not have direct access to your name when accessing rubric scores. When the results of this study are published or presented to professional audiences, the scores will totaled and averaged; your score will not be reported anywhere with your name next to it. The names of the people who participated in the study will not be revealed. The school name will also not be identified.

The purpose of the study is to learn more about interdisciplinary learning and authentic assessment, so that teachers can better understand assessment design and student performance. Your participation in this study may not provide information that will be directly helpful to you, but I hope it can be helpful future students who are undergoing a similar experience.

If you have any questions, you may contact me at (xxx) xxx-xxxx or xxxxxxxx@xxxxxxxx.xxx. You may also contact my faculty advisor, Dr. Christopher Lund, at xxxxxxxx@xxxxxxxx.xxx. You may keep a copy of this form if you wish.

If I have questions about my rights as a research participant, I may contact Pepperdine University Graduate and Professional Schools Institutional Review Board (GPS IRB) at (xxx) xxx-xxxx or at xxxxxxxx@xxxxxxxx.xxx.

Student's signature _____ Date _____

I have explained and defined in detail the research procedure in which the subject has consented to participate. Having explained this and answered any questions, I am cosigning this form and accepting this person's consent.

Signature of Jenni Taylor, Principal Investigator, Doctoral Student at Pepperdine University _____ Date _____

Estimado Estudiante,

Me llamo Jenni Taylor, soy una estudiante en el Programa del Liderazgo Educativo Doctoral de la Universidad de Pepperdine. ¡Sus padres me han dado su permiso para que ustedes puedan tomar parte en un estudio, similar a él tesis que pronto entregaran!

Este estudio es un estudio de investigación que estoy realizando a los efectos de completar un doctorado en Liderazgo Educativo. Querría invitarlos a tomar parte en este estudio si están interesados. Antes que les explique más acerca del estudio, yo quisiera saber que ustedes han elegido participar por su propia voluntad completamente voluntaria. Nadie los forzará a hacer algo que usted no están interesados en hacer. La negativa a participar o dejar de participar en cualquier momento, no implica ninguna sanción o pérdida de beneficios a que tiene derecho, y no van a afectar a cualquiera de sus grados o su condición de estudiante.

Para tomar parte en este estudio, ustedes no tienen que hacer nada si ha completado las evaluaciones de Intercesión (sus Exámenes Interdisciplinarios de la Unidad y/o su Presentación del Foro de la Comunidad). Sus rubrics para estas evaluaciones serán utilizados para medir que bien los alumnos en general hacen en cada parte de sus evaluaciones.

Sus resultados serán mantenidos confidenciales y los investigadores no tendrán acceso directo a el nombre del estudiante cuando están revisando sus resultados. Cuando los resultados de este estudio sean publicados o son presentados a audiencias profesionales, las resultados en su totalización y los promedios; de sus resultados **no serán** reportados con su nombre. Los nombres de las personas que tomaron parte en el estudio no serán revelados. El nombre de la escuela también no será identificada.

El propósito del estudio es de aprender más acerca del aprendizaje interdisciplinario y evaluación auténtica, para que maestros puedan comprender mejor el diseño de evaluación y el desempeño de estudiantes. Su participación en este estudio tal vez no puede proporcionar información que será directamente útil a ustedes, pero espero que pueda ser útil para futuros estudiantes que experimentan una experiencia semejante.

Si tiene cualquier pregunta, usted me puede llamar al (xxx) xxx-xxxx o mandarme un correo electrónico xxxxxxxx@xxxxxxxx.xxx. Usted también puede contactar a mi consejero de la facultad, el Dr. Christopher Lund, en xxxxxxxx@xxxxxxxx.xxx. Usted puede mantener una copia de este formulario si desea.

Si tengo preguntas sobre mis derechos como participante de la investigación, puedo comunicarse con la Universidad de Pepperdine Postgrados y Colegios Profesionales Junta de Revisión Institucional (IRB GPS) al (xxx) xxx-xxxx o en xxxxxxxx@xxxxxxxx.xxx.

La firma de estudiante	Fecha

Le he explicado y definido en detalle el procedimiento de investigación en la que el sujeto haya dado su consentimiento para participar. Después de explique y dio respuestas a cualquier preguntas, estoy firmando para aceptar este formulario de consentimiento y de esta personaje.

La firma de Jenni Taylor, investigador, Estudiante de Doctorado en la Universidad de Pepperdine	Fecha asentimiento obtenido

Appendix N: Informed consent for participation in research: Parents

Participant/ Student: _____

Principal Investigator: Jenni Taylor, Doctoral Student in Pepperdine University's Graduate School of Education and Psychology.

Faculty Advisor: Dr. Christopher Lund, Faculty Member at Pepperdine University's Graduate School of Education and Psychology.

Title of Project: *Interdisciplinary Authentic Assessment: Cognitive Expectations and Student Performance*

1. I, parent of _____ (Student Name), agree to allow my student to participate in the research study being conducted by Jenni Taylor under the direction of her advisor Dr. Christopher Lund.

2. The overall purpose of this study is to expand current research regarding the expectations and outcomes of interdisciplinary authentic assessment. The study will examine the cognitive complexity of learning objectives within authentic assessment tasks and will also examine levels of student understanding resulting from these tasks. The study will also explore potential relationships between the cognitive complexity of expectations and related student performance.

3. My student is already participating in the interdisciplinary assessment (Interdisciplinary Unit Exam and/or Community Forum presentation). My consent will allow allowing the researcher to use my student's scores as part of the study.

4. Participation in the study will not require my student(s) to do anything other than take the regular assessments for his or her assigned classes.

5. I understand that the possible benefits to myself or society from this research are: a deeper societal understanding of assessment design and related cognitive performance, and the potential this understanding may have for the field of education and curricular reform in schools and for future students.

6. I understand that there are certain risks and discomforts that might be associated with this research. These risks include: possible discomfort among students from knowing they are part of a research project.

7. I understand that I may choose not to allow my students to participate in this research.

8. I understand that my student's participation is voluntary and that I may refuse to have him/her participate and/or withdraw my consent and discontinue participation in the project or activity at any time without penalty or loss of benefits to which I am otherwise entitled.

9. I understand that the investigators will take all reasonable measures to protect the confidentiality of my student's records and my student's identity will not be revealed in any publication that may result from this project. The confidentiality of records will be maintained in accordance with applicable state and federal laws. Under California law, there are exceptions to confidentiality, including suspicion that a child, elder, or dependent adult is being abused, or if an individual discloses an intent to harm him/herself or others.

10. I understand that the investigator is willing to answer any inquiries I may have concerning the research herein described. I understand that I may contact Jenni Taylor (xxx-xxx-xxxx or xxxxxxxx@xxxxxxxx.xxx) or her supervisor, Dr. Christopher Lund (xxxxxxxx@xxxxxxxx.xxx) if I have other questions or concerns about this research. If I have questions about my rights as a research participant, I understand that I can contact the Chairperson of the Graduate and Professional Schools Institutional Review Board (GPS IRB), Dr. Yuying Tsong, at xxx-xxx-xxxx or at xxxxxxxx@xxxxxxxx.xxx.

11. I understand to my satisfaction the information regarding participation in the research project. All my questions have been answered to my satisfaction. I have received a copy of this informed consent form which I have read and understand. I hereby consent to allow my student to participate in the research described above.

Parent or legal guardian's signature on participant's behalf if participant is less than 18 years of age or not legally competent.	Date

I have explained and defined in detail the research procedure in which the subject has consented to participate. Having explained this and answered any questions, I am cosigning this form and accepting this person's consent.

Principal Investigator	Date

El participante/Estudiante: _____

Principal Investigador: Jenni Taylor, Estudiante Doctoral en los cursos de posgraduado de la Universidad de Pepperdine de Educación y Psicología.

Asesor de la Facultad: el Dr. Christopher Lund, facultad de la Escuela de Postgrado de la Universidad Pepperdine de Educación y Psicología

El título del Proyecto: La Evaluación Auténtica de interdisciplinaria: Expectativas cognoscitivas y El Desempeño del Estudiante

1. Yo, el padre (tutor) de _____ (Nombre de Estudiante), concuerdo en permitir a mi estudiante para tomar parte en la investigación o estudio que será realizado por Jenni Taylor bajo la dirección de su consejero el Doctor. Christopher Lund.

2. El propósito general de este estudio es de expandir investigaciones actuales con respecto a las expectaciones y resultados de evaluación auténtica interdisciplinaria. El estudio examinará la complejidad cognoscitiva de objetivos de aprendizaje dentro de tareas auténticas de evaluación y también examinará los niveles de la comprensión de estudiante que resulta de estas tareas. El estudio también explorará las relaciones potenciales entre la complejidad cognoscitiva de expectaciones y el desempeño relacionado del estudiante

3. Mi estudiante ya toma parte en la evaluación interdisciplinaria (Examen Interdisciplinario de Unidad y/o presentación del Foro de Comunidad). Mi consentimiento permitirá permitiendo el investigador para utilizar los resultados de mi estudiante como parte del estudio

4. La participación en el estudio no requerirá a mis estudiantes a hacer nada de otra manera que toma las evaluaciones regulares para sus clases asignadas

5. La participación en el estudio no requerirá a mis estudiantes a hacer nada de otra manera que toma las evaluaciones regulares para sus clases asignadas.

6. Yo Comprendo que hay ciertos riesgos y molestias que quizás sean asociados con esta investigación. Estos riesgos incluyen: molestias entre estudiantes de saber que ellos forman parte de un proyecto de investigación.

7. Comprendo que puedo escoger no permitir a mi(s) estudiantes tomar parte en esta investigación

8. Comprendo que la participación de mi estudiante es voluntaria y que puedo negarme a que él/ella participen y/o retira mi consentimiento y discontinúa participación en el proyecto o la actividad en tiempo sin pena o pérdida de beneficios a que yo de otro modo soy permitido.

9. Comprendo que los investigadores tomarán todas las medidas razonables para proteger la confidencialidad de los documentos de mi estudiante y la identidad de mi estudiante no será revelada en ninguna publicación que puede resultar de este proyecto. La confidencialidad de documentos será mantenida de acuerdo con el estado aplicable y las leyes federales. Bajo la ley de California, hay excepciones a la confidencialidad, si hay sospecha que un niño, anciano, o adulto son abusados, o si un individuo revela que hay una intención para dañarse a él/ella misma o a otros.

10. Comprendo que el investigador está dispuesto a contestar cualquier indagación que puedo tener con respecto a la investigación que se ha mencionado. Comprendo que puedo llamar a Jenni Taylor (xxxxxxxx@xxxxxxxx.xxx) o su supervisor, el Doctor. Christopher Lund (xxxxxxxx@xxxxxxxx.xxx) si tengo otras preguntas o preocupaciones acerca de esta investigación. Si tengo preguntas acerca de mis derechos como un participante de investigación, yo comprendo que puedo contactar el Presidente de la Posgrado y Consejo Colegios Profesionales de Revisión Institucional (IRB GPS), el Dr. Yuying Tsong, en xxx-xxx-xxxx o por correo electrónico xxxxxxxx@xxxxxxxx.xxx.

11. Comprendo a mi satisfacción la información con respecto a participación en el proyecto de investigación. Todas mis preguntas han sido contestadas a mi satisfacción. He recibido una copia de este formulario de consentimiento que he leído y he comprendido. Yo presente consiento en permitir a mi estudiante tomar parte en la investigación descrita arriba

_____ _____

Firma de padre o tutor legal del participante si el participante es menor de 18 años de edad o no legalmente competente.

Fecha

He explicado y he definido en detalles el procedimiento de investigación en el que el sujeto ha consentido en participar. Habiendo explicado este y contestado cualquier pregunta(s), yo consigno esta forma y aceptar el consentimiento de esta persona.

_____ _____

La firma de Jenni Taylor, investigador principal, Estudiante de Doctorado en la Universidad de Pepperdine

Fecha

Appendix O: Informed consent for participation in research: Panel member

Participant/ Teacher: _____

Principal Investigator: Jenni Taylor, Doctoral Student in Pepperdine University's Graduate School of Education and Psychology.

Faculty Advisor: Dr. Christopher Lund, Faculty Member at Pepperdine University's Graduate School of Education and Psychology.

Title of Project: *Interdisciplinary Authentic Assessment: Cognitive Expectations and Student Performance*

1. I, _____ (Panel Member Name), agree to participate in the research study being conducted by Jenni Taylor under the direction of her advisor Dr. Christopher Lund.

2. The overall purpose of this study is to expand current limited research regarding the cognitive expectations and outcomes of interdisciplinary authentic assessment. The study will examine the cognitive complexity of learning objectives within authentic assessment tasks and will also examine levels of student performance resulting from these tasks. The study will also explore potential relationships between the cognitive complexity of expectations and related student performance. The panel of assessment experts will be categorizing the learning objectives of assessments into Bloom's Taxonomy. The researcher will also summarize students' scores and compare results according to cognitive complexity.

3. This school has been selected for this study, due to the unique approach to interdisciplinary authentic assessment at this site. I have been specifically selected as a panel member for this research, due to my expertise and interest in the area of assessment, interdisciplinary teaching and learning, or cognitive theory. As a participant, I will work with a team of other panel members to categorize learning objectives from four different assessment tasks into levels of Bloom's Taxonomy using a Proving Behaviors tool.

4. My participation in the study will require me to pilot the use of the Bloom's Taxonomy Proving Behaviors tool with two other panel members and to normalize the categorization of assessment objectives into levels of the taxonomy. After we are able to categorize objectives within a range of one cognitive level, we will be asked to place the objectives for the interdisciplinary assessments for this research, marking an asterisk next to placement which could not be agreed upon between the three panel members.

5. I understand that the possible benefits to myself or society from this research are: a deeper societal understanding of assessment design and related cognitive performance, and the potential this understanding may have for the field of education and curricular reform in schools and for future students.

6. I understand that there are certain risks and discomforts that might be associated with this research. These risks or discomforts include judgment or frustration of the school our teachers regarding our placement of objectives into categories, and possible scrutiny from a wider educational community regarding our expertise and decisions.

7. I understand that I may choose not to participate in this research.

8. I understand that my participation is voluntary and that I may refuse to participate and/or withdraw my consent and discontinue participation in the project or activity at any time without penalty or loss of benefits to which I am otherwise entitled.

9. I understand that the investigators will take all reasonable measures to protect the confidentiality of my records and my identity will not be revealed in any publication that may result from this project. The confidentiality of my records will be maintained in accordance with applicable state and federal laws. Under California law, there are exceptions to confidentiality, including suspicion that a child, elder, or dependent adult is being abused, or if an individual discloses an intent to harm him/herself or others.

10. I understand that the investigator is willing to answer any inquiries I may have concerning the research herein described. I understand that I may contact Jenni Taylor (xxxxxxxx@xxxxxxxx.xxx) or her supervisor, Dr. Christopher Lund (xxxxxxxx@xxxxxxxx.xxx) if I have other questions or concerns about this research. If I have questions about my rights as a research participant, I understand that I can contact the Chairperson of the Graduate and Professional Schools Institutional Review Board (GPS IRB), Dr. Yuying Tsong, at xxx-xxx-xxxx or at xxxxxxxx@xxxxxxxx.xxx.

11. I understand to my satisfaction the information regarding participation in the research project. All my questions have been answered to my satisfaction. I have received a copy of this informed consent form which I have read and understand. I hereby consent to participate in the research described above.

_____ _____

Participant's Signature Date

I have explained and defined in detail the research procedure in which the subject has consented to participate. Having explained this and answered any questions, I am cosigning this form and accepting this person's consent.

_____ _____

Principal Investigator Date