Lindenwood University
Digital Commons@Lindenwood University

Dissertations | Theses & Dissertations

Fall 11-2016

Comparison of Classroom Grades and Missouri End-of-Course Exam Scores

Rodney Delmont
Lindenwood University

Follow this and additional works at: https://digitalcommons.lindenwood.edu/dissertations

 Part of the Educational Assessment, Evaluation, and Research Commons

Recommended Citation

Delmont, Rodney, "Comparison of Classroom Grades and Missouri End-of-Course Exam Scores" (2016). *Dissertations*. 273.
https://digitalcommons.lindenwood.edu/dissertations/273

This Dissertation is brought to you for free and open access by the Theses & Dissertations at Digital Commons@Lindenwood University. It has been accepted for inclusion in Dissertations by an authorized administrator of Digital Commons@Lindenwood University. For more information, please contact phuffman@lindenwood.edu.

Comparison of Classroom Grades and Missouri End-of-Course Exam Scores

by

Rodney Delmont

November 2016

A Dissertation submitted to the Education Faculty of Lindenwood University in partial fulfillment of the requirements for the degree of

Doctor of Education

School of Education

Comparison of Classroom Grades and Missouri

End-of-Course Exam Scores

by

Rodney Delmont

This Dissertation has been approved as partial fulfillment

of the requirements for the degree of

Doctor of Education

Lindenwood University, School of Education

Dr. Shelly Fransen, Dissertation Chair — Date: 11-9-16

Dr. Sherry DeVore, Committee Member — Date: 11-9-16

Dr. Dennis Cooper, Committee Member — Date: 11-9-16

Declaration of Originality

I do hereby declare and attest to the fact that this is an original study based solely upon my own scholarly work at Lindenwood University and that I have not submitted it for any other college or university course or degree.

Full Legal Name: Rodney O. Delmont

Signature:  Date: 11-9-16

Acknowledgements

I would like to thank my dissertation committee, Dr. Shelly Fransen, Dr. Sherry DeVore, and Dr. Dennis Cooper. The support and encouragement I received from my committee was greatly appreciated. I would also like to thank the school districts and administrators who gave their time to participate in this study. I am grateful to my wife, Nancy, and to my kids for being supportive through the entire process of finishing my dissertation. Finally, to my family, friends, and to the Weaubleau R-3 Board of Education, faculty, and staff, thank you for your support and encouragement with the many hours I have spent completing my dissertation.

Abstract

Educational leaders are tasked with finding the best practices for motivating students and accurately reporting progress (Popham, 2011). Additionally, the alignment of curriculum, assessment, and pedagogy can facilitate changes in education (Klenowski & Wyatt-Smith, 2014). Therefore, as districts align curriculum to assessment standards, students' scores on assessments should increase (Squires, 2012). The purpose of this study was to determine if a correlation exists between Missouri English II end-of-course (EOC) exam scores and student grade-point averages (GPAs) earned in English II. Moreover, data were examined to determine if a correlation exists among student gender, Missouri English II EOC scores, and student GPAs earned in the course. Additionally, perceptions from high school principals concerning grades, standardized assessments, and gender bias were analyzed. Seven southwest Missouri schools from the same conference were chosen to participate. Schools submitted secondary data including gender, Missouri English II end-of-course exam scores, and GPAs earned in the English II course. Additionally, secondary principals from participating districts were asked to complete a survey. The study revealed a positive correlation between Missouri English II EOC exam scores and GPAs. Furthermore, the study showed a higher percentage of females than males earning a 3.0 or higher average GPA in English II. However, despite the correlation of EOC scores and GPAs, the majority of principals surveyed disagreed standardized tests are an accurate representation of students' subject knowledge. Likewise, a majority of principals surveyed expressed course grades are free of gender bias and the best practice for accurately reporting student progress.

Table of Contents

List of Tables

List of Figures

Chapter One: Introduction

In many districts, keeping up with directives such as the No Child Left Behind Act of 2001 has been the primary focus of educators (Schimmer, 2012; Yesbeck, 2011). The No Child Left Behind Act of 2001 was replaced by the Every Student Succeeds Act (ESSA) in December 2015 (U.S. Department of Education, 2016). The ESSA continues to require annual statewide assessments to demonstrate student progress toward meeting high standards (U.S. Department of Education, 2016). Furthermore, the ESSA ensures districts are held accountable when students are not making progress (U.S. Department of Education, 2016). Districts can be assessed penalties and federal funding can be withheld for inadequate educational performance (Kaufman & Blewett, 2012; Meier, 2012). As a result, teachers and administrators feel intense pressure to ensure test scores improve (Styron & Styron, 2012). Schools limit and change curriculum to match standardized state tests (Eastwood, D'Agostino, & Welsh, 2014). Additionally, teachers may only present to students what is covered on the standardized state tests, and classroom assessments resemble individual state's standardized tests (Eastwood et al., 2014).

Despite alignment of curriculum to state standards, there are often many components to a classroom grade (Goodwin, 2011; Tomlinson, 2014). The final grade for a course might include a percentage for homework, projects, quizzes, unit tests, participation, and attendance (Reeves, 2011). In some situations, grades might have been modified for students with disabilities, inflated with extra credit, or curved to increase the overall average scores (Erickson, 2011; O'Connor & Wormeli, 2011).

Both course grades and standardized state assessments have strengths as methods of evaluating a student's academic ability (Tomlinson, 2014). However, using either

standardized assessments or classroom grades as a definitive measure of student learning has limitations (Ikenberry, Jankowski, Kinzie, & Kuh, 2014). Despite these limitations, this research was designed to determine if a significant positive correlation exists between students' scores in a course and students' scores on standardized assessments.

In this chapter, a background of the study and overview of the research are provided. Additionally, the theoretical framework, the statement of the problem, and the purpose of the study are offered. Next, research questions used to guide this study are detailed. Furthermore, limitations to the study are explained.

Background of the Study

Educational leaders are tasked with finding the best practices for motivating students and accurately reporting progress (Popham, 2011). Additionally, the passage of the No Child Left Behind Act in 2001 created an emphasis on educational accountability in the United States (Meier, 2012). To make states more accountable, No Child Left Behind mandates tied federal funding to academic achievement (Kaufman & Blewett, 2012). In addition, districts repeatedly not meeting Adequate Yearly Progress (AYP) became unaccredited and were tasked to replace their principals and other key staff members (Kaufman & Blewett, 2012).

Further perpetuating educational accountability, Race to the Top is an educational initiative from the United States Department of Education which requires student academic performance to be used in teacher evaluations (U.S. Department of Education, 2016). Using student test scores to evaluate teacher performance is perceived as unfair by many secondary school teachers (Almy, 2011). The Every Student Succeeds Act, a successor to the No Child Left Behind Act, allows states to decide if teacher evaluations

will be tied to student performance on standardized tests (Klein, 2016). However, districts under this act are still required to test students, and states are required to identify and transform poor-performing schools (Klein, 2016).

Early high-stakes assessments came in the form of college entrance examinations. Colleges such as Princeton, Harvard, and Yale each had their own entrance examinations (Jacobsen, 2016). In 1947, the Educational Testing Service assumed responsibility for developing and administering a standardized test known as the Scholastic Aptitude Test (SAT) (Jacobsen, 2016). In 1959, the American College Testing (ACT) Program was founded by Ted McCarrel and E. F. Lindquist (Jacobsen, 2016). Lindquist suggested the ACT was an achievement test and should not be used just for admission but as an indicator of academic preparation and placement (Jacobsen, 2016). In 1965, the Elementary and Secondary Education Act required public school students to take a standardized test in order to gather information (Hout & Elliott, 2011). Ultimately, in 1988, the Elementary and Secondary Education Act was reauthorized and required districts with poor test results to develop improvement plans (Hout & Elliott, 2011).

All states' standardized assessment results are submitted to the United States Department of Education for review to determine if the assessments meet the mandated requirements; in Missouri, the assessments make up the Missouri Assessment Program, also referred to as the MAP (Missouri Department of Elementary and Secondary Education [MODESE], 2016a). This collection of grade-level and end-of-course assessments are used to provide data at state and federal levels for student achievement accountability (MODESE, 2016a). Grade-level exams are administered in grades three through eight in both English language arts and mathematics and in grades five and eight

for science (MODESE, 2016a). End-of-course exams are administered at the secondary level in Algebra I, Biology, and English II (MODESE, 2016a). The assessments consist of multiple choice, constructed response, and in some areas, performance event items (MODESE, 2016a).

Another criterion for college admission is grade point average (GPA) (Hutt & Schneider, 2014; Soares, 2012). According to Schmitt (2012), the "best predictors of college GPA are high school GPA and SAT/ACT scores" (p. 25). Johnson (2013) suggested one of the top professional responsibilities of teachers is reporting grades in a timely and consistent manner.

Grading in America represents one of the most fundamental aspects of education (Hutt & Schneider, 2014). Due to an increased mobility of students in the 19th century, school districts were pressured to standardize practices such as school organization, school calendars, and grading (Hutt & Schneider, 2014). Additionally, grades became so integral to education that a need for simple reporting short-hand in the form of letter grades and GPAs was generated (Hutt & Schneider, 2014). The creation of grade point average is credited to William Farish and was implemented in 1792 (Soh, 2011). By the 1940s, the A-F marking system had become the dominant grading system, and by 1971, 80% of schools in the United States were using letter grades (Hutt & Schneider, 2014).

Theoretical Framework

For this study, an interpretivist theoretical framework was utilized. The introduction of interpretivism in education can be referenced back to the late 1970s (Taylor & Medina, 2013). Interpretivism in education focuses on the political, social,

and economic forces influencing curriculum and instruction in districts where teachers are immersed (Taylor & Medina, 2013).

Furthermore, the key goal of interpretivism is to search for patterns of meaning (Butin, 2010). Butin (2010) suggested interpretivism assumes reality is constructed by society, and consequently, can be represented through varying viewpoints. Additionally, interpretivism "accurately and thoroughly document[s] the perspective being investigated" and "does not attempt to adjudicate between competing truth claims in order to determine the one best answer" (Butin, 2010, p. 60). Consequently, this research was focused on the correlation between classroom grades and standardized test scores as opposed to a determination of which is a more accurate representation of student mastery of subject matter.

Interpretivist researchers try to learn what is going on in particular situations while taking into account the perspectives of the people concerned in order to organize data while excluding the researcher's own prior assumptions (Wellington, 2015). By using the interpretivist research perspective, the research focus was on perspectives of standardized test scores and classroom grades as a measure of learning. Likewise, by adhering to the interpretivist goal of searching for patterns of meaning, the correlation between classroom grades and standardized test was investigated. Additional patterns of gender equity in both standardized testing scores and classroom grades were examined.

Statement of the Problem

Standardized state assessment scores and classroom grade point average based on multiple data points can provide valuable insight to student learning (Ikenberry et al., 2014). Standardized state assessment scores and classroom grades are used for

accountability in different areas (Meier, 2012). Standardized assessments provide districts with insight to drive curricular changes and are a major component for school accountability (Eastwood et al., 2014). Course grades, however, provide teachers with guidance for changes in the classroom and are often the basis for student accountability (Tomlinson, 2014). Both standardized assessments and course grades are used to demonstrate student learning and to drive curriculum changes.

Purpose of the Study

The purpose of this study was to determine if a significant positive correlation exists between Missouri English II end-of-course exam scores and the GPAs earned in the English II course. Additionally, data were examined to determine if a correlation exists among student gender, Missouri English II end-of-course exam scores, and the GPAs earned in the course. This information will be disseminated to participating schools for administrators to determine the degree of alignment between the English II end-of-course exam and the teachers' assessment of student learning.

Research questions and hypotheses. The following research questions and hypotheses guided the study:

1. What is the correlation between Missouri English II end-of-course exam scores and the corresponding end-of-year GPAs earned in English II?

$H1_0$: There is no correlation between Missouri English II end-of-course exam scores and the corresponding end-of-year GPAs earned in English II.

$H1_a$: There is a correlation between Missouri English II end-of-course exam scores and the corresponding end-of-year GPAs earned in English II.

2. What is the correlation between Missouri English II end-of-course exam scores and gender?

$H2_0$: There is no correlation between Missouri English II end-of-course exam scores and gender.

$H2_a$: There is a correlation between Missouri English II end-of-course exam scores and gender.

3. What is the correlation between gender and end-of-year GPAs earned in English II?

$H3_0$: There is no correlation between gender and end-of-year GPAs earned in English II.

$H3_a$: There is a correlation between gender and end-of-year GPAs earned in English II.

4. What are the perceptions of high school principals regarding best practices for accurately reporting student progress?

5. What are the perceptions of high school principals regarding high-stakes testing?

Significance of Study

Since insight into student learning can be acquired through standardized assessment scores and student grade point averages (Ikenberry et al., 2014), both standardized assessment results and classroom GPAs are used for accountability in different areas (Meier, 2012). However, according to Dickinson and Adelson (2015), confirmatory factor analysis models indicate student grades and standardized assessment scores reflect noticeably differing achievement constructs. Consequently, as emphasis on

content standards and standardized assessments increases, there may become an increased alignment between what is taught and assessed in the classroom and on standardized assessments (Dickinson & Adelson, 2015).

While there is current research on grades versus standards-based grading and on high-stakes testing, there is little research correlating the two. This study will contribute to the knowledge base surrounding the relationship between grade point averages and standardized test results. This is significant, because educators are tasked with finding the best practices for motivating students and accurately reporting progress (Popham, 2011).

Limitations

The following limitations were identified in this study:

Sample demographics. This study included seven rural public school districts in southwest Missouri. The schools chosen were members of the same conference and were similar in size and demographics. The relatively small sample compared to the total number of public school districts in the state is a limitation.

Timeframe. Another limitation is the number of years of data examined. Over the past several years, the Missouri English II end-of-course exam content and the course-level expectations have varied; therefore, comparing multiple years of data does not yield valid results.

Summary

The state and federal government use of standardized tests makes raising test scores the single most important gauge of school improvement (Styron & Styron, 2012). As a result, teachers and administrators feel massive pressure to align curriculum with

state standardized tests (Eastwood et al., 2014). One indicator of the level of alignment would be if a significant positive correlation exists between course grades and the students' scores on the state assessment.

In Chapter One, a background of the study and overview of the research were provided including a historical synopsis of testing and accountability. Next, the interpretivist framework was examined, stating the key goal of interpretivism is to search for patterns of meaning (Butin, 2010). In addition to the theoretical framework, the statement of the problem and the purpose of the study were introduced. Finally, research questions used to guide this study were stated and limitations to the study were explained.

In Chapter Two, a literature review is included to examine relevant topics. The central topics of this study include the following: to examine accountability in education, to examine curriculum development, to analyze the impact of assessments, to examine current grading practices, and to consider gender bias in education. Next, Chapter Three consists of an explanation of the methodology, data collection, and ethical considerations employed for this study. Chapter Four includes an analysis of the quantitative data collected. Additionally, Chapter Four includes tables and figures to display the data. Finally, Chapter Five is comprised of a summary of findings, implications for practice, and conclusions regarding the correlation among Missouri English II end-of-course exam scores, GPAs earned in the English II course, and gender.

Chapter Two: Review of Literature

Schools use standardized tests as a tool to measure how students compare to other students (Hout, Elliott, & Frueh, 2012). Additionally, standardized tests are used to determine if children are ready for school; to track them into instructional tiers; to monitor for learning disabilities and other handicaps; and to decide whether to promote, retain in a grade level, or graduate many students (Chappuis, Stiggins, Chappuis, & Arter, 2012; Hout et al., 2012). As districts align curriculum to assessment standards, students' scores on assessments should increase (Squires, 2012)

Standardized state assessment scores and classroom grade point averages are effective methods for analyzing student learning (Ikenberry et al., 2014). Additionally, standardized assessments assist districts with curricular changes and are a major component for school accountability (Eastwood et al., 2014). Course grades, however, provide teachers with guidance for changes in the classroom and are often the basis for student accountability (Tomlinson, 2014).

Theoretical Framework

The interpretivist paradigm can be traced back to the sociological theories of Max Weber (Humphrey, 2013). Humphrey (2013) suggested the main principle of interpretivism and the offshoots of interpretivism is rooted in opposition of positivism. Furthermore, conflicting views between positivism and interpretivism can be traced back to the middle of the 19th century; however, the conflict appeared in the field of educational research during the second half of the 20th century (Murphy, 2016).

Butin (2010) stated interpretivism "accurately and thoroughly document[s] the perspective being investigated" and "does not attempt to adjudicate between competing

truth claims in order to determine the one best answer" (p. 60). In addition, interpretivists/social constructivists develop theories or search for meaningful patterns throughout the research process (Picciano, 2015). Furthermore, interpretivists must understand their own individual and cultural experiences shape their interpretations of situations (Creswell, 2013; Picciano, 2015).

The primary goal of an interpretivistic theoretical framework is to understand the experiences of each individual perspective (Taylor & Medina, 2013). Interpretivist researchers try to learn what is going on in particular situations while taking into account the perspectives of the people concerned in order to organize data while excluding the researcher's own prior assumptions (Wellington, 2015). For this study, an interpretivism theoretical framework was utilized.

By utilizing the interpretivism research perspective, the focus was on determining if a correlation exists between Missouri English II end-of-course exam scores and the corresponding end-of-year GPAs earned in English II. Additional patterns among gender, GPAs earned in English II, and Missouri English II end-of-course exam scores were evaluated. Reeves (2011) stated schools should be connecting the process of grading and feedback, not separating the two. Furthermore, using an average instead of other measures to calculate final grades may not reflect what students have learned or mastered (Marzano & Heflebower, 2011).

Butin (2010) emphasized the key goal in the interpretivism theoretical framework is to "search for patterns of meaning" (p. 59). Therefore, the research in this study was focused on patterns or correlations between GPAs earned in English II, Missouri English II end-of-course exam scores, and gender. Additionally, interpretivism dictates the

researcher present opposing viewpoints concerning the use of course GPAs and standardized test scores (Butin, 2010).

Assessments

Sindelar (2011) stated assessments reveal what students know before starting a lesson, the level of comprehension during the conveyance of the lesson, and details of what the students learned from the lesson. Furthermore, effective educators incorporate multiple forms of assessments (Sindelar, 2011). According to Slavin (2014), "Generation and use of data to inform instruction, as well as to measure the effectiveness of instruction, is a core part of the instructional cycle and is critical to the success of both teacher and students" (p. 1). Slavin (2014) suggested there are three categories of assessments which provide valuable data for instruction.

First, assessments fall under the Planning for Instruction category (Slavin, 2014). Sindelar (2011) stated one function of assessments is to demonstrate what the student knows before instruction begins. When initializing instruction on new content, educators can employ pre-assessments to evaluate what students already know about a specific subject (Clark, 2015; Slavin, 2014). Pre-assessment tools may include interest surveys, journaling, concept maps, or simple pre-tests (Slavin, 2014). Slavin (2014) further suggested data from pre-assessments can assist educators in selecting materials and instructional strategies to effectively engage students or in assigning groups based on knowledge and interest.

Additionally, assessments can fall under the Implementing Instruction category (Slavin, 2014). Sindelar (2011) noted this category of assessment aides the teacher in knowing whether or not students are comprehending the material being delivered.

Assessments utilized during the implementation category are referred to as formative assessments (Slavin, 2014). Formative assessments are used to monitor student learning during the learning process (University of Texas at Austin, 2015).

Formative assessments offer the most benefits for teachers and students when administered frequently and can be an important tool for educators (Chappuis et al., 2012). The use of formative assessments by educators can be formal or informal (Slavin, 2014). Additionally, formative assessments may be directed toward the whole class, small groups, or individual students (Slavin, 2014). Informal formative assessments may include discussions with students, listening to small group discussions, monitoring nonverbal cues from students, or students communicating their level of comfort with the material being presented (Slavin, 2014). Conversely, formal formative assessments may include bell work, exit slips, writing prompts, quizzes, homework assignments, or checklist (Slavin, 2014). However, Slavin (2014) cautioned against the overuse of formative assessments for grading purposes.

Formative assessments can effectively provide teachers with immediate feedback on students' level of understanding (Bartlett, 2012; Klenowski & Wyatt-Smith, 2014). However, formative assessments are only effective when employed systematically and purposefully (Slavin, 2014). Furthermore, Clark (2015) stated formative assessments encourage personal achievement. Additionally, formative assessments can guide instruction, improve student understanding and confidence, and improve summative test scores (Doubet, 2012; Edwards, 2013; Tomlinson, 2014).

An essential element of formative assessments is feeding up (Frey & Fisher, 2011), which is the process of clarifying the purpose of the lesson presented and the

assessments used (Slavin, 2014). According to Clark (2015), effective educators share criteria for success along with learning intentions. Additionally, educators should communicate the purpose with students frequently (Frey & Fisher, 2011; Slavin, 2014). Furthermore, establishing a clear purpose helps motivate the students to become actively engaged in the learning process (Frey & Fisher, 2011). Frey and Fisher (2011) contended an effective purpose "lays out the content of what will be learned, the learner's role in what will be accomplished, and the expectations for the interactions" (p. 17). Moreover, effectively communicating a clear purpose of instruction assists students with shifting information from the initial learning phase to a transference of learning (Frey & Fisher, 2011). According to Clark (2015), by providing transparency through a clear purpose, interaction and feedback are facilitated between students and instructors.

Furthermore, Slavin (2014) suggested students expect feedback when formative assessments are utilized. Additionally, effective feedback must be detailed, understandable, and returned in a timely manner (Frey & Fisher, 2011). According to Slavin (2014):

- Feedback should relay information concerning challenging goals or tasks.
- Feedback should be specific and make comparisons to what students have improved upon from past mistakes.
- Feedback is more effective if instructors focus on what students are doing right as opposed to what is wrong.
- Feedback should be encouraging as not to diminish students' confidence and self-esteem.

Effective teachers will utilize multiple forms of assessments (Sindelar, 2011). Multiple assessments help prevent judgments based on a single assessment (Slavin, 2014). Additionally, using multiple forms of assessments increases reliability and validity of assessment data (Slavin, 2014). Furthermore, using multiple assessments increases the quality and amount of feedback available to students (Slavin, 2014).

Moreover, assessments can fall under the Evaluating Instruction category (Slavin, 2014). A major component of the evaluating instruction category is the use of summative assessments (Slavin, 2014). Tomlinson and Moon (2013) stated summative assessments are more formal than pre-assessments and formative assessments. Furthermore, summative assessments often occur during transition points such as completion of chapters or units, midterm, end-of-semester, or end-of-course (Slavin, 2014). Summative assessments measure instructional outcomes and may be measured using chapter or unit tests, extensive projects or papers, midterm exams, final exams, or end-of-course assessments (Tomlinson & Moon, 2013). Furthermore, whereas pre-assessments and often formative assessments are not included in grade reporting, summative assessments are graded and usually make up a significant portion of the student's final grade (Slavin, 2014; Tomlinson & Moon, 2013).

Frequent use of summative assessments can detrimentally impact students' views of education and learning (Slavin, 2014). Furthermore, Earl (2004) stated:

> Teachers use the tests to assess the quantity and accuracy of student work, and the bulk of teacher effort is taken up in marking and grading. A strong emphasis is placed on comparing students, and feedback to students comes in the form of marks or grades, with little advice for improvement. These kinds of testing events

> indicate which students are doing well and which ones are doing poorly. Typically, they don't give much indication of mastery of particular ideas or concepts because the test content is generally too simplistic to represent the broad range of skills and knowledge that has been covered. But this lack of specificity hasn't presented a problem because the teachers' perceived purpose for the assessment is to produce a rank order of the students and assign a symbol to designate the students' position within the group, whatever group it might be. (pp. 22-23)

Summative assessments are misused when data are used more for teacher or school accountability than for educating students (Sindelar, 2011; Slavin, 2014).

Despite the finality of summative assessments, teachers and students may use summative assessment data in a formative manner (Slavin, 2014). Teachers can utilize summative assessment data to determine patterns of deficiencies and to modify curriculum (Slavin, 2014). Additionally, students may analyze data from summative assessments and apply skills and knowledge to future educational endeavors (Slavin, 2014).

Summative assessment of students' level of understanding, similar to formative assessments, provides more reliability and validity if based upon multiple assessments (Slavin, 2014). Reliability and validity are two key requirements of quality summative assessments (Tomlinson & Moon, 2013). Additionally, Tomlinson and Moon (2013) suggested quality summative assessments meet five indicators:

- The assessment should directly align with the learning goals and assessment items the student is expected to achieve.

- The content of the assessment reflects the knowledge and skills that are most significant for the student to attain competency in the subject matter.
- The format of the assessment is aligned with the depth of knowledge required to demonstrate mastery of the learning goals.
- The range of knowledge needed to respond to an assessment item is equivalent to the range of knowledge required during instruction of a specific learning goal.
- The assessment should not require students to utilize knowledge or skills outside what is required by the learning goal.

Many districts focus on two main categories of assessment: formative (low-stakes) and summative (high-stakes) (University of Texas at Austin, 2015). High-stakes assessments have existed for centuries (Slavin, 2014). Subsequently, summative assessments often have high point values (high-stakes), occur at the end of an instructional unit or course, and are developed to measure the degree to which students have achieved the desired learning outcomes (University of Texas at Austin, 2015).

Standardized assessments are often summative in nature (University of Texas at Austin, 2015). If utilized correctly, Astin (2013) suggested standardized assessments offer more meaningful information than traditional grade point averages. However, standardized assessments should not be used as one-time, end-of-course assessments (Astin, 2013). Astin (2013) noted the difficulty of any assessment is determining not only what outcomes will be assessed, but how the outcomes will be assessed. Furthermore, when assessments de-emphasize or omit skills or competencies, what is taught often lacks the same skills or competencies (Slavin, 2014).

Historically, standardized tests have been used as a tool to measure how students compare to other students (Hout et al., 2012). Standardized tests can be used to demonstrate how much a student has learned about a particular subject (Chappuis et al., 2012; Hout et al., 2012). Additionally, most standardized assessments can be used frequently to determine a student's growth (Astin, 2013).

Emanuel (2013) stated shorter, more frequent testing produces additional educational benefits. Astin (2013) suggested using assessments when students enter a program of study and again at the completion of a program demonstrates the students' growth. Additionally, using assessments multiple times has greater validity for judging teachers or districts than high-stakes, end-of-course assessments (Astin, 2013). Furthermore, the more frequently students are assessed over a given set of information, the more educational gains are made (Emanuel, 2013). Moreover, Emanuel (2013) proposed the sooner students are assessed after being exposed to new information, the greater the likelihood the information will be retained.

Inversely, Emanuel (2013) stated current use of standardized testing encourages memorization and inhibits originality. Slon (2013) cautioned the misuse of standardized assessments demoralizes teachers and alienates students. Additionally, end-of-course standardized assessments hinder creative thinking and merely create better test-takers (Emanuel, 2013; Slon, 2013). More frequent, shorter assessments not only foster creative problem solving, but improve transference of learned concepts to new situations (Emanuel, 2013).

According to Dickinson and Adelson (2015), standardized tests measuring college readiness have been used to gain admission into higher education institutions.

Additionally, due to pressure from accrediting associations and legislators, a growing number of colleges and universities are incorporating outcome-based assessments (Astin, 2013). Often, Astin (2013) suggested colleges utilize standardized assessments to demonstrate mastery of desired outcomes.

Proponents of a balanced assessment system suggest the importance of utilizing three measurement strategies (Slavin, 2014). Chappuis, Commodore, and Stiggins (2016) suggested using a balanced assessment system meets the data and decision making needs for all stakeholders. Slavin (2014) identified the three measurement strategies of balanced assessment as follows:

- Classroom assessments are often created by the teacher and are utilized to grade students or motivate students for larger assessments.
- Interim assessments are standardized assessments given multiple times throughout the year.
- Large-scale assessments are often used for accountability and are generated by assessment companies.

Educators are encouraged to balance the use of formative and summative assessments not only to grade students, but to facilitate student achievement (Chappuis et al., 2016).

However, according to Slavin (2014), only two of the three measurement strategies are supported by solid evidence. First, classroom assessments, when applied as formative assessments, are proven to be extremely effective in improving student achievement (Slavin, 2014). Additionally, large-scale assessments are deemed valuable for accountability purposes (Chappuis et al., 2016; Slavin, 2014). Inversely, Slavin (2014) suggested interim assessments lack sufficient research evidence for support.

Nonetheless, Chappuis et al. (2016) suggested interim assessments provide an assortment of useful assessment data. Interim assessments can be used to predict which students are at-risk of scoring poorly on standardized assessments (Chappuis et al., 2016; Slavin, 2014). Moreover, interim assessments can be used to evaluate the effectiveness of an educational program (Slavin, 2014). Furthermore, Slavin (2014) advocated interim assessments offer an instructional function by providing pertinent diagnostic information on students.

Grading

Dating back to the 19th century, school districts were pressured to formulate a standard grading system (Hutt & Schneider, 2014). In the 1880s, Harvard University implemented the use of letter grades as a way of evaluating student work and comparing students (Goodwin, 2011). Brookhart (2011) stated the most prevalent method of reporting student performance is letter grades. However, parents, students, and many teachers feel letter grades are subjective (Reeves, 2011).

Hutt and Schneider (2014) suggested one of the most important aspects of education in America is grading student progress. Furthermore, Tomlinson and Moon (2013) asserted, "Grades exist to communicate to students, their parents, and others who need the information, grades are used for multiple purposes, including, but by no means limited to, communication of current student academic performance" (p. 123). Moreover, Spencer (2012) stated grades allow teachers a method of ranking students and creating hierarchies and curves. Additionally, Tomlinson and Moon (2013) proposed further purposes for grades include the following:

- Determining class rank and identifying class honors such as valedictorian or salutatorian.
- Motivating student learning through rewards or punitive measures.
- Grouping students by ability level or determining enrollment status in certain courses.
- Providing educators with information needed for planning instructional strategies.
- Evaluating and assessing teachers and programs.

According to Cox (2011), learning standards and testing at the local and national levels have improved due to national reform initiatives; however, alignment of curriculum standards and grading has made little progress. Brookhart (2011) stated, "Even though opinions about why grades are important differ, more and more educators are beginning to question traditional grading practices that were developed to sort students into learners and nonlearners, not to support learning for all" (p. 10). Subsequently, while standards-based reporting has become more prevalent at the elementary level, grade reporting at the secondary level has remained relatively unchanged (Cox, 2011).

Districts attempt to align curriculum to assessment standards in an effort to increase student scores on standardized assessments (Squires, 2012). Nevertheless, there are often many components to a classroom grade (Goodwin, 2011; Tomlinson, 2014). Reeves (2011) suggested the final grade for a course can be an average of homework, projects, quizzes, and unit tests, in addition to scores for attendance, participation, and behavior.

According to Brookhart (2011), the reporting of student grades needs to differentiate between mastery of standards and non-academic components, such as behavior and attendance. Non-academic elements in education may include behavior and attendance (Brookhart, 2011). Furthermore, using an average instead of other measures to calculate final grades may be an inaccurate representation of a student's mastery of subject matter (Marzano & Heflebower, 2011).

Marzano and Heflebower (2011) proposed many districts are changing from traditional grading systems to standards-based grading systems. Standards-based grading provides clear expectations for student learning (Marzano & Heflebower, 2011). Additionally, standards-based report cards communicate more meaningful information about content mastery to students and parents (Marzano & Heflebower, 2011). Moreover, Johnson (2013) proposed educators' main responsibility is reporting grades in a timely and consistent manner.

Spencer (2012) stated, "Standards-based grading derives from the idea that teachers ought to have clearly defined academic goals for their students, be able to determine if they've met them, and then communicate that to students and parents" (p. 5). Furthermore, correlational evidence has shown high school GPAs are a better predictor of first-year college GPAs than are standardized assessment scores (Sawyer, 2013). However, 30% of first-year students will drop out of college (Goodwin, 2011). Vatterott (2015) suggested high dropout rates in college might skew the correlation of high school GPAs and college GPAs.

Effective grading practices. According to Vatterott (2015), current changes in grading practices consist of a shift from a compliance culture defined by teachers

assigning grades as rewards for compliance to a performance-based culture defined by students demonstrating mastery of learning standards. However, Cox (2011) suggested grading reform lacks the momentum existing in other standards-based educational reforms. Moreover, effective grading practices follow a number of commonly agreed-upon attributes (Brookhart, 2011; Tomlinson & Moon, 2013).

First, grades should be based on specific learning goals that are clearly defined (Tomlinson & Moon, 2013). According to Spencer (2012), clearly defined goals not only make determining if students meet the goals easier, but clearly defined goals can be communicated more effectively to students and parents. Educators must state what students are expected to know, understand, and be able to do upon completion of the lesson (Slavin, 2014).

Next, educators should base grades on set criteria (Tomlinson & Moon, 2013). Additionally, grades should not be comparative or used to rank students (Vatterott, 2015). According to Brookhart (2011), grades should be assigned based on what the student has learned, and should not be merely based on points a student has earned.

Furthermore, Tomlinson and Moon (2013) asserted educators should avoid over-grading student work. Learning challenging content to a high level of depth of knowledge requires repetition and practice (Tomlinson & Moon, 2013). With traditional grading systems, student grades suffer when concepts are not mastered early and may continue to suffer with additional assessments (Vatterott, 2015). However, effective feedback on repetitive formative assessments is effective for student academic success; grading the assessments is not always necessary (Tomlinson & Moon, 2013).

Additionally, educators should utilize quality assessments when evaluating students (Tomlinson & Moon, 2013). Assessments must be valid, reliable, and designed to assess clearly defined learning goals (Slavin, 2014). Furthermore, Tomlinson and Moon (2013) suggested quality assessments avoid trick questions and assess the types of knowledge or skills in the same manner designated in the learning goal.

Educators should avoid anything during instruction or grading that adds confusion to the meaning of the grade (Tomlinson & Moon, 2013). Traditional grades often include extra credit or bonus points, scores from group work, attendance points, and points for behavior (Reeves, 2011). Spencer (2012) suggested separating non-academic factors from a student's grade and noting non-academic factors in a separate location on the student's report card.

Consequently, educators should avoid mathematical procedures that detract from accurately reporting student mastery of learning objectives (Tomlinson & Moon, 2013). Traditional grading commonly assigns zeros for missing work (Brookhart, 2011). Subsequently, students receiving zeros on assignments have a difficult time raising grades and can become frustrated and unmotivated (Vatterott, 2015). Additionally, Tomlinson and Moon (2013) noted by averaging all scores to obtain a final percentage, student grades are often lowered by one low score.

Educators should place greater emphasis on grading during the later stages of the grading cycle (Tomlinson & Moon, 2013). Brookhart (2011) suggested grading should encourage students and facilitate a motivation to learn. When students attain higher grades later in a grading cycle, they demonstrate effort and growth (Tomlinson & Moon, 2013).

Next, Tomlinson and Moon (2013) suggested report cards should be categorized to report three key areas:

- Student performance – Student performance includes the educator's assessment concerning what the student knows, understands, and can do.
- Student process – Educators assess the student's willingness to work in order to master content goals.
- Student progress – Educators evaluate improvements made by the student from one grading cycle to the next.

Additionally, grade reports should provide useful information given to parents and students in a timely manner (Brookhart, 2011). Furthermore, Spencer (2012) cautioned against report cards with numerous indicators or indicators written in language difficult for parents and students to understand.

Finally, educators should include students in the grading, reporting, and assessment process (Tomlinson & Moon, 2013). Slavin (2014) maintained, "For effective self-assessment to occur, students need to have the following:

- Awareness of the value of self-assessment
- Access to clear criteria on which to base the assessment
- A specific task or performance to assess
- Direct instruction in and assistance with self-assessment
- Practice
- Opportunities to revise and improve the task or performance. (p. 21)

Furthermore, self-assessment encourages students to develop the skills required to monitor their own academic achievements (Slavin, 2014).

Grade Inflation

A decade after implementing the letter grade system at Harvard, some professors were concerned higher grades were distributed too readily (Goodwin, 2011). Additionally, 27% of college presidents surveyed stated professors graded too leniently (Glenn, 2011). In the Theory of Grade Inflation, several factors have led to an increase in grades (Oleinik, 2009). Grades given by teachers should be a reliable measurement of the student's knowledge in a particular subject; however, there are critics who state teachers give high grades for mediocre work (Goodwin, 2011). Godfrey (2011) showed some schools offer higher grades and have lower levels of proficiency on Advanced Placement tests than other schools do (Godfrey, 2011).

Consequently, grade inflation decreases the validity of high school GPAs (Godfrey, 2011). Erickson (2011) suggested teachers commonly award extra points for non-academic endeavors ranging from bringing in tissue boxes and canned food donations to completing extra credit assignments. Subsequently, Godfrey (2011) stated a consequence of grade inflation is college admission offices experience difficulties when trying to compare students and make admission decisions (Godfrey, 2011). Often, high school grades are viewed as an unreliable criterion for college admission (Geiser & Santelices, 2007; Godfrey, 2011).

Over a 20-year span, the average GPA for the graduating class of 2006 increased .26 points while standardized scores on the SAT remained comparatively unchanged (Double Take, 2011). Assessment results are not the only data demonstrating grade inflation. Cohen, Germuska, and Rado (2011) analyzed data and found the average GPA for Illinois graduates from 2006 thru 2008 was 3.08; however, the average GPA for the

same students as college freshmen was 2.52. Schmidt (2007) analyzed student performance on the National Assessment of Educational Progress. While reading scores declined between 1992 and 2005, GPAs rose from 2.68 in 1990 to 2.98 in 2005 (Schmidt, 2007).

Laurie (2007) further explained a significant negative correlation exists between grade inflation and student performance on standardized tests. Consequentially, when the difference between the teacher-assigned grade and the school's average on the standardized test was large, there was a lower average score on the standardized test for the district (Laurie, 2007). Laurie (2009) suggested possible explanations for the gap between teachers' assigned grades and the schools' average grades on the standardized test.

One explanation is the standardized tests cover a whole course, while the grades assigned by a teacher are comprised of several short-term tests over smaller portions of the curriculum (University of Texas at Austin, 2015). Additionally, teacher-assigned grades are made up of other work (Laurie, 2009). Furthermore, teachers will increase scores on assignments if the class average is lower than anticipated (Erickson, 2011). Effort, while often subjective, can be a major factor in a student's grade (Carey & Carifo, 2011). Furthermore, Strauss (2014) stated high school grades are a better predictor of college success than ACT or SAT scores.

Curriculum

According to Clark (2015), "Curriculum is the totality of experiences which are planned for young learners through their education" (p. 91). Additionally, Squires (2012) stated, "Curriculum is the district's written plan incorporating aspects of time use, content

and process aligned to standards and assessments that establishes a focus for instruction, assessment, staff development and management so student achievement improves" (p. 46). Furthermore, the alignment of curriculum, assessment, and pedagogy can facilitate changes in education (Klenowski & Wyatt-Smith, 2014).

Squires (2012) noted there is no officially agreed-on method for formalizing curriculum. Likewise, Clark (2015) noted an important difference from past curriculum requirements. Attributable to the rapidly changing workplace, predicting the knowledge needed even five to 10 years from now is impossible (Clark, 2015). Therefore, a school's curriculum should be challenging, foster an enjoyment of learning, encourage higher-order thinking skills, and spark creativity and innovation (Clark, 2015). Squires (2012) suggested curriculum should be developed with the following four key areas in order to improve student achievement:

- Alignment – Curriculum should be aligned to the most current state standards and assessment criteria in order to describe what the student is to know and do (Squires, 2012).
- Structure of the Curriculum – The structure of the curriculum specifically states not only what needs taught, but specifies ways a teacher should teach the subject matter to ensure consistency (Squires, 2012).
- Implementation and Infrastructure – District administrators have systems in place to ensure the curriculum is being taught (Squires, 2012).
- Assessment – Assessment data from students are reviewed and communicated with teachers so teachers can modify instruction to better suit the needs of students (Squires, 2012).

Teachers, especially inexperienced teachers, can benefit from having a curriculum established that provides detailed guidance and materials (Eakle, 2012). Eakle (2012) further suggested having a developed curriculum allows teachers the time needed to develop classroom discipline and management strategies which enable the material to be presented more effectively. Therefore, Squires (2012) asserted curriculum can provide school districts with direction to guide improvements.

Furthermore, Eakle (2012) stated experienced teachers are seldom practiced in or trained to write curriculum. According to Squires (2012), developed curriculum can help eliminate variation in teaching outcomes. Additionally, teachers may not have the time needed to align curriculum with current research and best practices (Eakle, 2012).

Klenowski and Wyatt-Smith (2014) stated, "Alignment – or, more aptly, the extent to which it is achieved – is a marker of good education" (p. 5). Curriculum must have vertical alignment ("Hidden Curriculum," 2014). Vertical alignment ensures content taught in a grade level, course, or lesson will prepare the student for subsequent content ("Hidden Curriculum," 2014). Additionally, horizontal alignment ensures assessments are based on and aligned with what content has actually been taught ("Hidden Curriculum," 2014). Horizontal alignment warrants the content assessed and taught be aligned to the learning standards for a particular grade level or course ("Hidden Curriculum," 2014). However, curriculum alignment can be difficult due to numerous standards and assessments (Squires, 2012).

The Common Core standards were designed to become the foundation for educational accountability systems (Hess & Mcshane, 2013). The Common Core State Standards Initiative began in 2009 when nearly all state education agencies collaborated

with various associations, legislators, and business leaders (Doorey, 2012). Doorey (2012) suggested the Common Core initiative addressed key issues in education including:student mobility and the differing expectations across state lines, the ever-changing skills required for the current and future workplace, creating competitiveness in the global workplace, and alignment with standardized assessments. However, the Common Core standards were intended to identify the knowledge and skills students need, not to be a national curriculum (Hess & Mcshane, 2013).

Despite the collaborative effort by states to create the Common Core standards, many states have decided to discard implementation of the standards (Hess & Mcshane, 2013). Missouri has replaced the Common Core standards with the Missouri Learning Standards (Ballentine, 2016). The new Missouri Learning Standards were developed and reviewed by educators, legislators, parents, and academic researchers from across the state (Ballentine, 2016). Ballentine (2016) stated school districts must now develop curriculum aligned to the new standards.

The new Missouri Learning Standards not only incorporate math and English like the Common Core standards, but also address social studies and science (Ballentine, 2016). Additionally, Ballentine (2016) noted the Missouri Learning Standards reorganize math benchmarks, place an emphasis on research in language arts, and reintroduce cursive writing at the elementary level. Finally, the state will need to develop standardized state assessments aligned to the Missouri Learning Standards (Ballentine, 2016).

Conversely, Pinar (2012) countered, “Creativity, critical thinking, even the canon are all causalities of curriculum aligned with standardized examinations” (p. 65).

Furthermore, districts' focus on improving test results has taken away from other critical areas of education (Slon, 2013). Districts have reduced or eliminated funding in the areas of art, music, and sports (Slon, 2013).

Accountability

Title I of the Elementary and Secondary Education Act of 1965 first required standardized testing of students in public schools (Hout & Elliott, 2011). However, earlier high-stakes assessments came in the form of college entrance examinations. Several elite colleges such as Princeton, Harvard, and Yale each had their own entrance examination (Jacobsen, 2016). The Educational Testing Service assumed responsibility for developing and administering a standardized test known as the Scholastic Aptitude Test in 1947 (Jacobsen, 2016). Lindquist, founder of the American College Testing Program in 1959, suggested the ACT was an achievement test and should not be used just for admission, but as an indicator of academic preparation and placement (Jacobsen, 2016).

In 1965, the Elementary and Secondary Education Act required public school students to take a standardized test in order to gather information (Hout & Elliott, 2011). Additionally, college readiness tests were utilized in state accountability measures (Dickinson & Adelson, 2015). Ultimately, in 1988, the Elementary and Secondary Education Act was reauthorized and required districts with poor test results to develop improvement plans (Hout & Elliott, 2011).

Despite opposition from teacher unions, all states, districts, schools, teachers, and administrators are held accountable through high-stakes standardized assessments (Almy, 2011). Bushaw and Lopez (2013) stated less than 25% of Americans polled consider the

increased usage of high-stakes assessments has actually helped increase the performance of public schools. Furthermore, the majority of Americans polled are opposed to using standardized assessment scores as a part of teacher evaluations (Bushaw & Lopez, 2013).

The passage of the No Child Left Behind Act in 2001 created an emphasis on educational accountability in the United States (Meier, 2012). No Child Left Behind tied federal funding to academic achievement in an effort to make states more accountable (Kaufman & Blewett, 2012). Moreover, districts repeatedly not meeting Adequate Yearly Progress (AYP) became unaccredited and were to replace their principals and other key staff members (Kaufman & Blewett, 2012). However, since the implementation of No Child Left Behind, teachers have voiced their opinions concerning the shift in importance from teaching students how to think critically to teaching them how to pass a standardized test (Chatterji, 2013).

Further propagating educational accountability was Race to the Top. According to McNeil (2013), U.S. Secretary of Education Arne Duncan used $4 billion as Race to the Top incentives and cited incentives are the most effective method to push his education agenda. The Race to the Top educational initiative required states to use student academic performance in teacher evaluations (U.S. Department of Education, 2016). However, using student test scores to evaluate teacher performance is perceived as unfair by many secondary school teachers (Almy, 2011).

More recently was the passage of the Every Student Succeeds Act. The ESSA, a successor to the No Child Left Behind Act, allows states to decide if teacher evaluations will be tied to student performance on standardized tests (Klein, 2016). However,

districts under the ESSA are still required to test students, and states are required to identify and transform poor-performing schools (Klein, 2016).

All states' standardized assessments are submitted to the United States Department of Education for review to determine if the assessments meet the requirements of the No Child Left Behind Act of 2001 (MODESE, 2016a). Missouri Assessment Program (MAP) status is full approval (MODESE, 2016a). This collection of grade-level and end-of-course assessments are used to provide data at the state and federal levels for student achievement accountability (MODESE, 2016a). Grade-level exams are administered in grades three through eight in both English language arts and mathematics and in grades five and eight for science (MODESE, 2016a). End-of-course exams are administered at the secondary level in Algebra I, Biology, and English II (MODESE, 2016a). The assessments consist of multiple choice, constructed response, and in some areas, performance event items (MODESE, 2016a).

Gender Bias

A common finding in education research asserts females have an advantage in school and achieve higher marks (Voyer & Voyer, 2014). Additionally, Voyer and Voyer (2014) stated a significant advantage for females over their male counterparts in language courses. Conversely, males tend to receive more attention from their teachers (Stevens, 2015).

Reeves (2011) suggested males tend to be less obedient, and consequently, are more likely to fail a course despite their actual aptitude in the course. Additionally, females experience better grades because they are more compliant (Brookhart, 2011;

Reeves, 2011). Subsequently, females may not be as motivated to comprehend more challenging content due to receiving higher marks (Brookhart, 2011).

While girls tend to outperform boys on reading tests at the primary level, boys and girls tend to score the same on science and math assessments (Cornwell, Mustard, & Van Parys, 2013). However, boys who perform the same as girls on assessments are given lower grades by their teachers (Cornwell et al., 2013). Additionally, from the 1980s to the 2000s, the average high school letter grade for females shifted from "B" to "A" while the average high school letter grade for males remained at "B" (Fortin, Oreopoulos, & Phipps, 2013). In Britain, females were 7.3% more likely to earn five or more grades in the C-A range than males (Wells, 2015).

Additionally, standardized assessments given to 15-year-old students in 63 countries including the United States demonstrated a disparity in performance between male and female students (Porter, 2015). Porter (2015) stated the males who score high in math surpass the females who score high on the math assessment. However, Porter (2015) suggested males are falling behind in the other subjects.

In 2012, more females are earned college degrees than their male counterparts (Porter, 2015). Wells (2015) affirmed British females are a third more likely to attend post-secondary educational institutions than males. However, females are less likely to pursue careers in math or science (Porter, 2015).

In many countries, the number of women now surpasses men in terms of enrollment and success in higher education (Grow & Van Bavel, 2015). In contrast, males in many cultures tend to view themselves higher in terms of intelligence than their female counterparts (Kaufman, 2012; Stumm, 2014). Stumm (2014) referred to the

higher self-estimation of intelligence by males as a phenomenon called “male hubris – female humility” bias (p. 244).

Summary

High-stakes testing is the driving force behind the No Child Left Behind Act. High-stakes testing uses the promise of rewards and the threat of punishment in an attempt to make teachers accountable (Hout & Elliott, 2011). Ultimately, one effect of high-stakes testing is the narrowing of curriculum to focus more intently on what is tested (Chatterji, 2013).

Analysis of research indicates an increase in grade inflation (Goodwin, 2011). Additionally, there is a negative correlation between grade inflation and student performance on standardized tests (Laurie, 2007). Furthermore, an increase in the gender gap has occurred, and female students are receiving higher grades (Voyer & Voyer, 2014).

Chapter Three includes a detailed explanation of the methodology, data collection, and ethical considerations employed for this quantitative study. Next, a synopsis of the problem and purpose of the study are presented. Additionally, explanations of the population, sample, and instrumentation are provided. Finally, the approach for data collection and data analysis processes are examined.

Chapter Three: Methodology

The passage of the No Child Left Behind Act in 2001 created an emphasis on educational accountability in the United States (Meier, 2012). Improving test scores has, therefore, become a primary focus of educators (Schimmer, 2012; Yesbeck, 2011). Secondary-level educators often perceive the use of student test scores to evaluate teacher performance as unfair (Almy, 2011). However, using multiple factors including high school GPAs and SAT/ACT scores is best when predicting college GPAs (Schmitt, 2012).

According to Punch (2014), interpretivism is most likely to be associated with qualitative research. Adams and Lawrence (2015) stated, "Qualitative measures are non-numerical while quantitative measures are numerical" (p. 78). Furthermore, the key goal of interpretivism is to search for patterns of meaning (Butin, 2010). Creswell (2014) added that with quantitative research, patterns within the data can be analyzed. Therefore, this study was designed using the theoretical framework of interpretivism with quantitative analysis to determine if correlating patterns exist.

In Chapter Three, the problem and purpose, research design, population, and sample are discussed. Additionally, the instrument to be used in the study, data collection procedures, and data analysis are addressed. Finally, ethical considerations concerning confidentiality and anonymity are considered.

Problem and Purpose Overview

Grading in America represents one of the most fundamental aspects in education (Hutt & Schneider, 2014). Furthermore, standardized state assessment scores and classroom grade point averages based on multiple data points can provide valuable

insight to how students have learned (Ikenberry et al., 2014). However, using either standardized assessments or classroom grades as a definitive measure of student learning has limitations (Ikenberry et al., 2014). Consequentially, as emphasis on standardized assessments increases, there may become an increased alignment between what is taught and assessed in the classroom and the content standards of standardized assessments (Dickinson & Adelson, 2015). Information resulting from this study will be disseminated to participating schools in order for administrators to determine the degree of alignment between English II end-of-course exams and teacher assessments of student learning.

Research questions and hypotheses. The following research questions and hypotheses guided this study:

1. What is the correlation between Missouri English II end-of-course exam scores and the corresponding end-of-year GPAs earned in English II?

$H1_0$: There is no correlation between Missouri English II end-of-course exam scores and the corresponding end-of-year GPAs earned in English II.

$H1_a$: There is a correlation between Missouri English II end-of-course exam scores and the corresponding end-of-year GPAs earned in English II.

2. What is the correlation between Missouri English II end-of-course exam scores and gender?

$H2_0$: There is no correlation between Missouri English II end-of-course exam scores and gender.

$H2_a$: There is a correlation between Missouri English II end-of-course exam scores and gender.

3. What is the correlation between gender and end-of-year GPAs earned in English II?

$H3_0$: There is no correlation between gender and end-of-year GPAs earned in English II.

$H3_a$: There is a correlation between gender and end-of-year GPAs earned in English II.

4. What are the perceptions of high school principals regarding best practices for accurately reporting student progress?

5. What are the perceptions of high school principals regarding high-stakes testing?

Research Design

This study was conducted utilizing a quantitative research design, specifically a correlational quantitative design. According to Creswell (2014), correlational research can be defined as "procedures in quantitative research in which investigators measure the degree of association between two or more variables using the statistical procedure of correlational analysis" (p. 21). Quantitative analysis of data from the Missouri English II end-of-course exam and an average of first- and second-semester course GPAs in English II for the 2013-2014, 2014-2015, and 2015-2016 school years was conducted. Furthermore, a survey (see Appendix A) was developed to assess principals' perceptions regarding high-stakes testing and grade reporting.

Ethical Considerations

Upon approval of the Institutional Review Board (IRB) at Lindenwood University (see Appendix B), school district superintendents were sent a letter electronically (see

Appendix C) to determine each district's interest in providing data and being part of the study. School superintendents who agreed to participate in the study were asked to sign an informed consent agreement (see Appendix D). Once approval from the superintendent was obtained, a third party made contact with the proper staff responsible for student data. In most cases, this was the school principal or counselor.

The identities of the principals completing the survey remained confidential, along with the district names. Districts were assigned a letter to be used in place of names. Additionally, principals were provided an informed consent to determine whether they wished to participate in the study.

The data requested included a list of students, their grades for each semester of English II, and their English II end-of-course exam performance levels. The MOSIS numbers were omitted to maintain confidentiality and are not reported in the findings. Data were collected for the 2013-2014, 2014-2015, and 2015-2016 school years. Additionally, principals were asked to complete a short survey concerning their perceptions of high-stakes testing and grade reporting. Individual principal names or school district names were not used in the data collection; all districts were randomly assigned identification letters to maintain the anonymity of all participants.

Once the third party collected the data, these data were merged into an Excel spreadsheet (see Appendix E). The data were released to the primary researcher and were stored in a locked cabinet. Any identifiable information for the students was omitted to ensure confidentiality. After completion of the study, the statistical data and survey information will be maintained in a locked cabinet for three years as required by federal regulation.

Population and Sample

The population for this study included all public high schools in Missouri. A purposeful sample of seven public high schools was chosen for this study. Bryman (2012) stated purposive sampling units are selected because they have particular characteristics which enable detailed exploration of the questions being studied. The sample was selected to represent student scores and principal perceptions in districts with similar demographics as follows:

- only one elementary and one high school in the district,
- no charter schools,
- one teacher for the English II course,
- comparable student-to-teacher ratio, and
- affiliation with the same conference.

The population consisted of all students from the selected schools who participated in the English II end-of-course exam and the English II course during the 2013-2014, 2014-2015, or 2015-2016 school years. Additionally, principals at the participating schools were surveyed.

Instrumentation

The student data were collected by a third party who made contact with the proper staff responsible for student data. In most cases, this was the school principal or counselor who exported data from their student records program in a comma-separated values (CSV) file. Once the third party collected the data, it was merged into an Excel spreadsheet.

Principals were asked to complete a short survey concerning their perceptions of high-stakes testing and grade reporting. The survey questions were field-tested by certified personnel within the public school system to address reliability and validity. The survey was generated through SurveyMonkey utilizing a four-point Likert-type scale. A Likert-type scale is composed of a series of four or more character or personality trait items that are combined into a single composite score to provide a quantitative measure (Boone & Boone, 2012). According to Boone and Boone (2012), descriptive statistics including central tendency and mean are recommended for interval Likert-type scale items. The numerical data were charted and graphed using Excel.

Data Collection

Upon approval of the Institutional Review Board at Lindenwood University, school district superintendents were contacted to determine each district's interest in participating in the study and providing data. Once approval from the superintendent was obtained, a third party made contact with the appropriate staff responsible for student data. In most cases, this was the school principal or counselor. Consequently, students who transferred in or out of the districts during the year the course was taken were removed from the study.

The survey was generated through SurveyMonkey utilizing a Likert-type scale. A third-party data collector sent the survey link to participants. Additionally, a consent agreement was sent to the participating principals explaining the purpose of the research, voluntary participation and privacy statements, description of risk, procedures, estimated time for completion, and contact numbers for questions or concerns.

Data Analysis

A Pearson product-moment correlation coefficient (PPMC) was utilized in conducting a correlational analysis. The PPMC coefficient is useful in statistics because it indicates both the magnitude and direction of the relationship between two variables (Ary, Jacobs, Sorensen, & Walker, 2014). The overall scores were used in the PPMC calculation to determine if there is a relationship between English II end-of-course exam scores and student grade point averages in English II. Additionally, each proficiency level of the English II end-of-course exam was compared to the corresponding grade point averages to determine if a correlation exists. Finally, the data were analyzed by gender using the PPMC calculation to determine the degree of correlation, if any, between each gender's English II end-of-course exam scores and student grade point averages in English II.

Descriptive statistics were used to analyze the quantitative data from the surveys and each gender's English II end-of-course exam scores and student grade point averages in English II. Ary et al. (2014) stated, "Descriptive statistics serve to describe and summarize observations" (p. 154). Tables were created to display the descriptive data and to view a summary of the responses as a whole. Additionally, a two-tailed *t*-test was conducted to determine if there is a statistically significant difference between the frequency of male and female proficiency levels and grades (Murphy, Myors, & Wolach, 2014). In order to determine a statistical significance, the observed value of *t* was compared with the critical value of *t* (Murphy et al., 2014).

Summary

The sample population size was narrowed to include participants of similar demographics and to create a manageable sample size. Data were obtained from the participating schools through a third party. Quantitative analysis of data from the Missouri English II end-of-course exam and an average of first- and second-semester course grades in English II in a sample population for the 2013-2014, 2014-2015, and 2015-2016 school years was conducted. Additionally, principal perceptions regarding high-stakes testing and grade reporting were examined.

Chapter Four: Analysis of Data

According to Squires (2012), research in curriculum alignment exists showing the relationship among the alignment of the written curriculum, the taught curriculum, and the tested curriculum. Furthermore, Squires (2012) stated the alignment of any pair will have a positive impact on the other. The Every Student Succeeds Act continues to require annual statewide assessments which demonstrate students' progress toward meeting high standards (U.S. Department of Education, 2016). Therefore, teachers and administrators feel intense pressure to ensure test scores improve (Styron & Styron, 2012). The purpose of this study was to determine if a significant positive correlation exists between Missouri English II end-of-course exam scores and GPAs earned in the English II course. Additionally, data were examined to determine if a correlation exists among student gender, Missouri English II end-of-course exam scores, and GPAs earned in the course.

A 10-question survey concerning perceptions of high school principals was utilized for this study. The survey topics included the following:

- course grades to demonstrate mastery of content
- gender bias in course grades
- standardized assessments to demonstrate mastery of content
- gender bias in standardized assessments
- use of standardized assessments for teacher accountability
- use of standardized assessments for principal accountability
- alignment of curriculum
- correlation of course grades and standardized test results

The survey employed a four-point Likert-type scale to demonstrate the participant's level of agreement to a particular question. A weighted average and a bar graph are utilized to illustrate participants' perceptions in response to each question.

English II course grades, English II End-of-Course assessment proficiency levels, and gender for each of the three years of data and for the combined three years of data were analyzed. Analysis included correlation of students as follows:

- A in English II and proficiency level
- B in English II and proficiency level
- C in English II and proficiency level
- D in English II and proficiency level
- advanced and grade in English II
- proficient and grade in English II
- basic and grade in English II
- below basic and grade in English II
- overall gender and proficiency level
- overall gender and grade

The Pearson product-moment correlation coefficient was performed to measure the magnitude and direction of a linear correlation between the two variables (Bluman, 2012). Figures and tables were used to visually represent the numerical data in a concise and easy-to-comprehend format.

High School Demographics

A sample of Missouri high schools from the same conference was selected, and demographic data were collected from the MODESE (2016b). The sample consisted of

seven schools from the southwest region in the state of Missouri which agreed to participate in the study. However, only six of the seven schools participated in either the principal survey or submission of data concerning gender, Missouri English II end-of-course exam proficiency level, and course grades in English II for the 2013-2014, 2014-2015, and 2015-2016 school years.

Table 1

Number of Students to Classroom Teacher

School	2013	2014	2015
School A	12	13	13
School B	14	14	16
School C	19	19	18
School D	15	17	15
School E	11	11	12
School F	14	14	13
School G	16	14	13

Note. Data from the MODESE (2016b).

As shown in Table 1, the number of students per classroom teacher is comparable for the majority of schools. Some demographics were significantly different between the Missouri state average for percentage of students participating in the free and reduced lunch program and the data for high schools that participated in the study, as shown in Figure 2. School D had the highest average percentage of free and reduced lunch

participation at 88.5%. Dissimilarly, School E, the only school below the state average, had the lowest average percentage of free and reduced lunch participation at 49.0%.

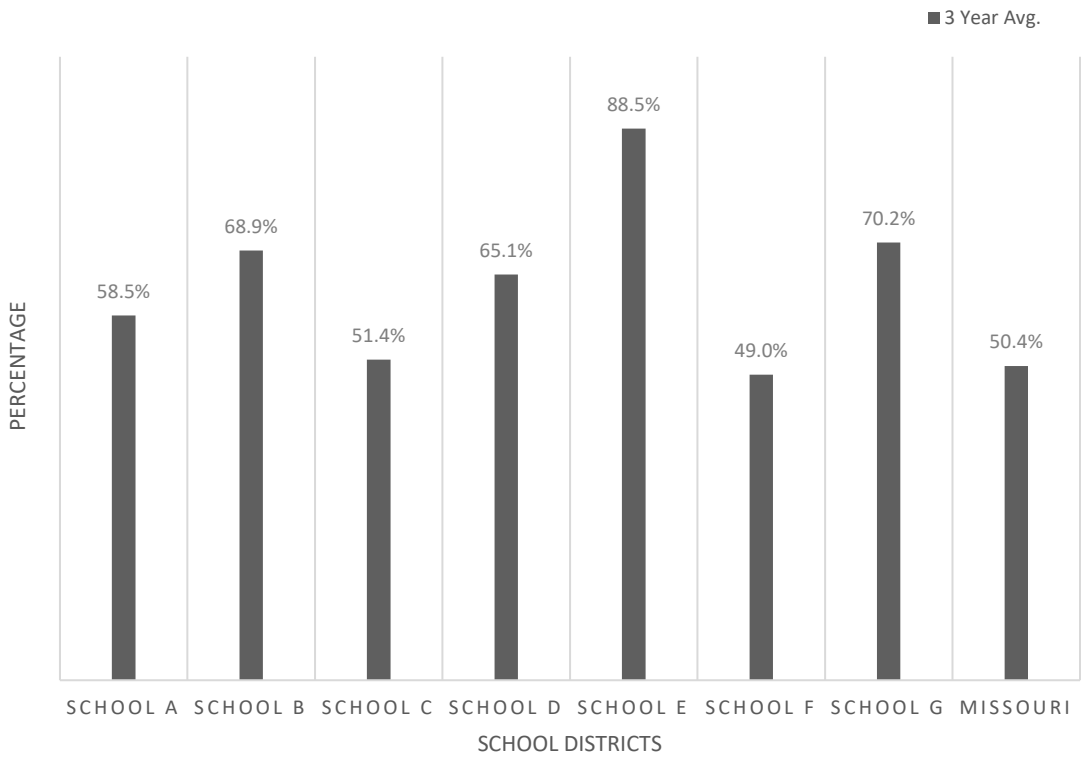


Figure 1. 2013-2015 free and reduced price meals percentages (MODESE, 2016b).

Analysis of Quantitative Data–PPMC

A PPMC was performed using proficiency levels on the English II end-of-course assessment and the course grade point averages for two semesters of English II. Three years of data and the combined data of all three years were analyzed to determine the magnitude and direction of the correlation. The Pearson correlation coefficient, *r*, has a value ranging from +1 to -1, where a value of 0 indicates no correlation between the two

variables (Ary et al., 2014). Additionally, the coefficient of determination, r^2, was calculated to show the amount of variation shared by both variables (Lund & Lund, 2013). According to Lund and Lund (2013), as seen in Table 2, the following guidelines have been proposed:

Table 2

Proposed Guidelines for Interpreting Pearson's Correlation Coefficient

Strength of Association	Coefficient, *r* Positive	Coefficient, *r* Negative
Small	.1 to .3	-0.1 to -0.3
Medium	.5 to .5	-0.3 to -0.5
Large	.5 to 1.0	-0.5 to -1.0

Furthermore, for an *r* value to be considered statistically significant, the size of the sample, *n*, must be evaluated (Illowsky, 2016). In order to find the minimum *r* value to be significant, the degree of freedom must be calculated by subtracting two from *n* (Illowsky, 2016).

EOC Proficiency Level vs. Course Grade

The first PPMC was performed comparing proficiency levels on the English II end-of-course assessment and the average course grade point average for English II for the 2014 school year. The value of *r* was 0.4681 ($n = 95$), and the value of r^2, the coefficient of determination, was 0.2191. Although technically a positive correlation, the relationship was of medium strength, as seen in Figure 2.

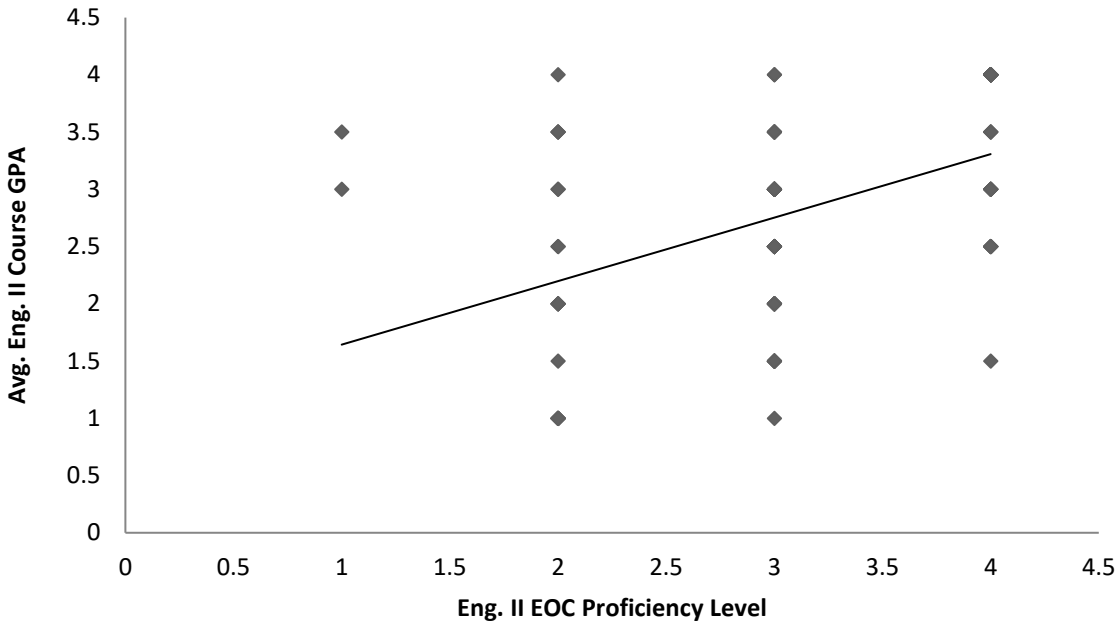


Figure 2. Correlation between 2014 EOC proficiency level and grade.

For the 2015 school year, the PPMC was performed comparing proficiency levels on the English II end-of-course assessment and the average course grade point average for English II. The value of r was 0.461 (n = 92), and the value of r^2, the coefficient of determination, was 0.2125. Although showing a positive correlation, the relationship was of medium strength as seen in Figure 3.

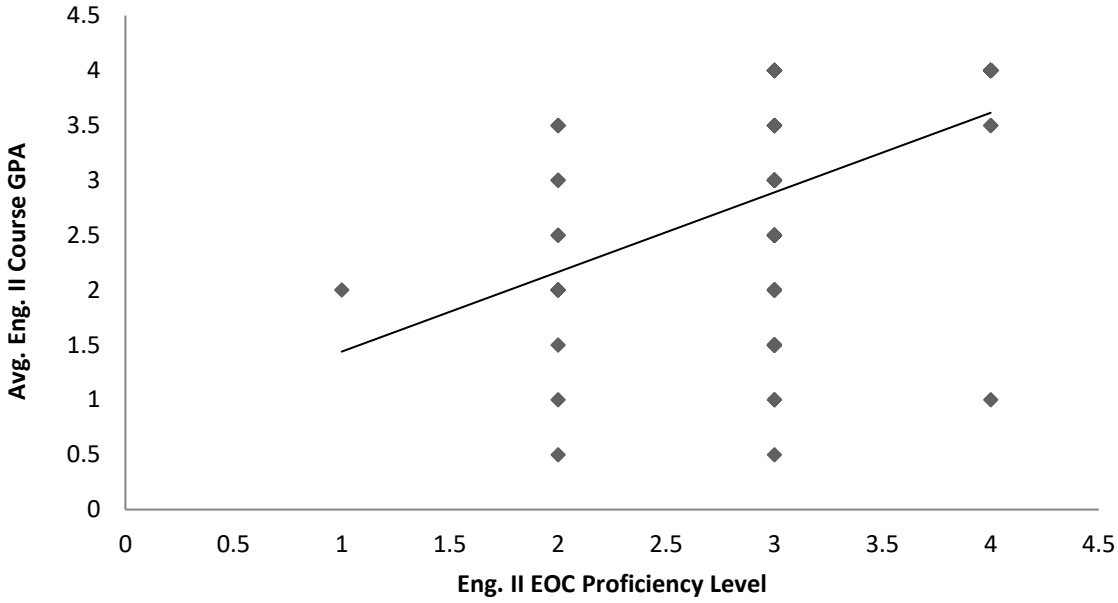


Figure 3. Correlation between 2015 EOC proficiency level and grade.

Next, the PPMC was performed comparing proficiency levels on the English II end-of-course assessment and the average course grade point average for English II for the 2016 school year. The value of r was 0.4182 (n = 114), and the value of r^2, the coefficient of determination, was 0.1749. Although still a positive correlation, the relationship was of medium strength as seen in Figure 4.

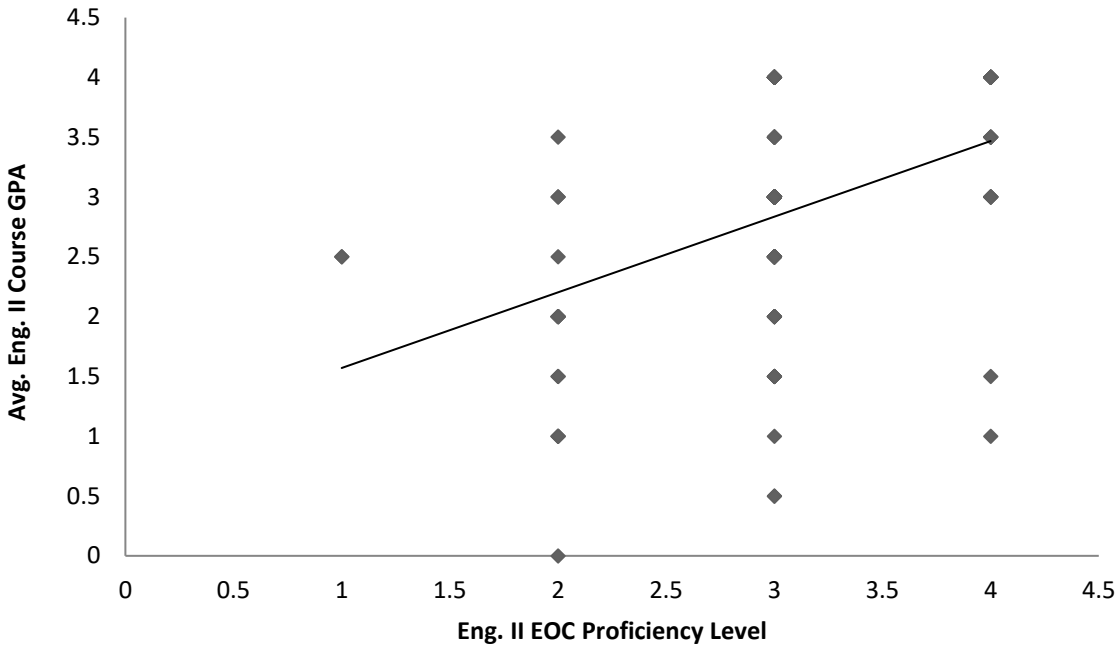


Figure 4. Correlation between 2016 EOC proficiency level and grade.

Finally, the PPMC was performed comparing proficiency levels on the English II end-of-course assessment and the average course grade point average for English II for all three school years combined. The value of r was 0.446 (n = 301), and the value of r^2, the coefficient of determination, was 0.1989. Although theoretically a positive correlation, the relationship was of medium strength as seen in Figure 5.

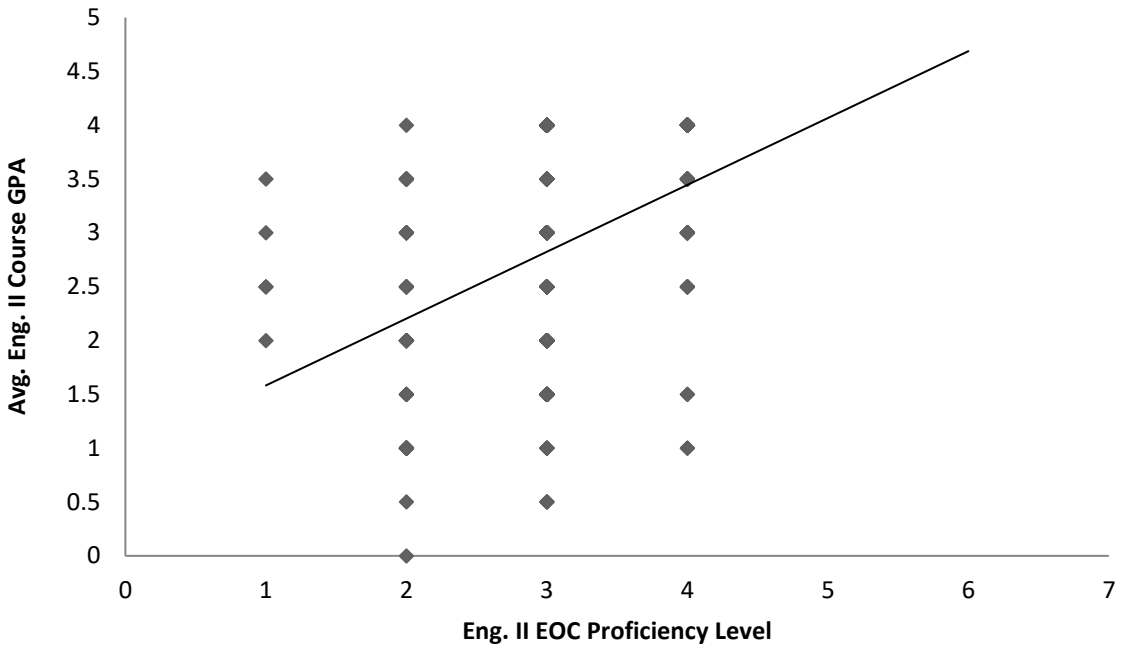


Figure 5. Correlation between three-year EOC proficiency level and grade.

Advanced and Proficient vs. Course Grade

The next PPMC was performed comparing the EOC advanced and proficient levels on the English II end-of-course assessment and the average course grade point average for English II for the 2014 school year. The value of r was 0.5446 (n = 74), and the value of r^2, the coefficient of determination, was 0.2966. Although technically a positive correlation, the relationship was on the lower range of strong as seen in Figure 6.

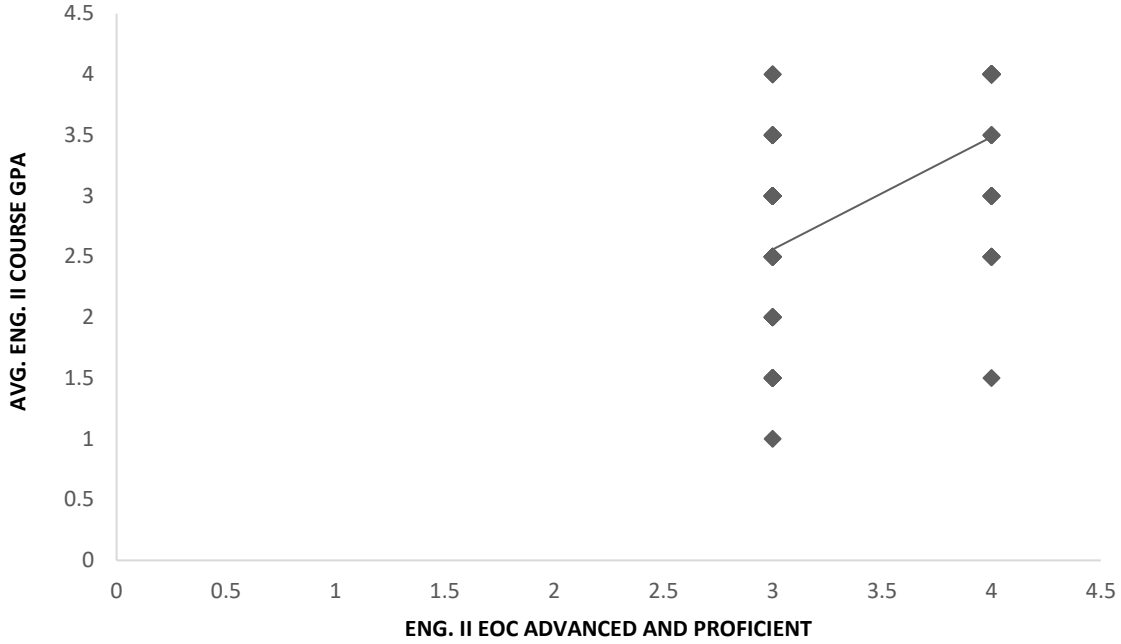


Figure 6. Advanced and proficient vs. grades for 2014.

Additionally, the PPMC was performed comparing the EOC advanced and proficient levels on the English II end-of-course assessment and the average course grade point average for English II for the 2015 school year. The value of r was 0.4445 (n = 77), and the value of r^2, the coefficient of determination, was 0.1976. Although technically a positive correlation, the relationship was of medium strength as seen in Figure 7.

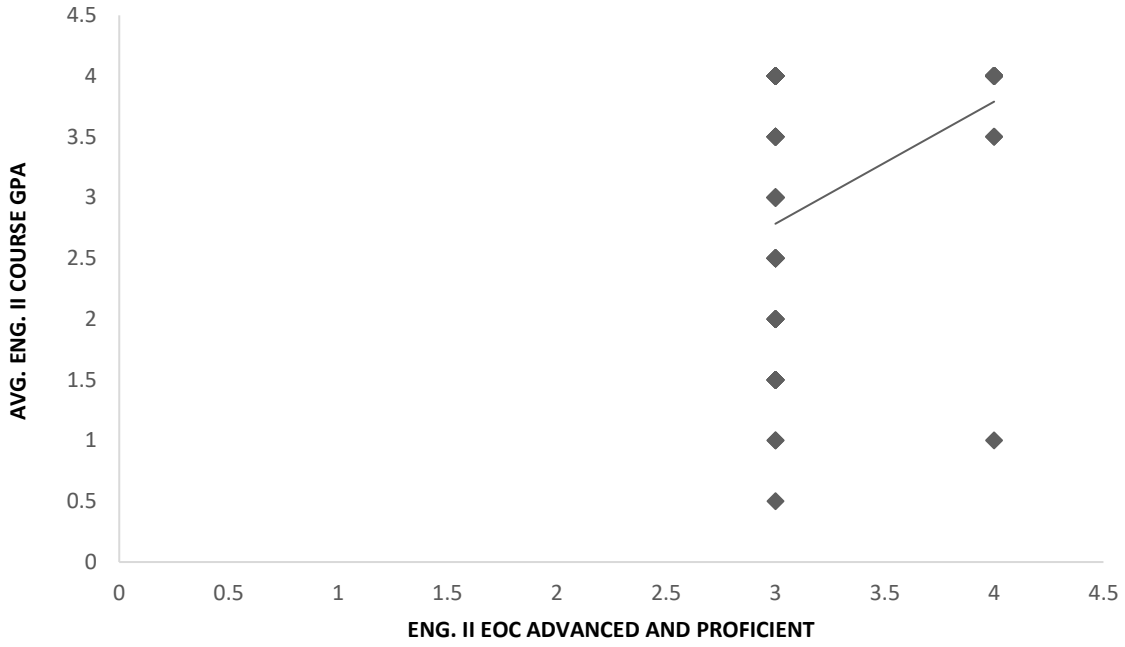


Figure 7. Advanced and proficient vs. grades for 2015.

For the 2016 school year, the PPMC was performed comparing the EOC advanced and proficient levels on the English II end-of-course assessment and the average course grade point average for English II. The value of r was 0.2442 (n = 97), and the value of r^2, the coefficient of determination, was 0.0596. Although technically a positive correlation, the relationship was weak as seen in Figure 8.

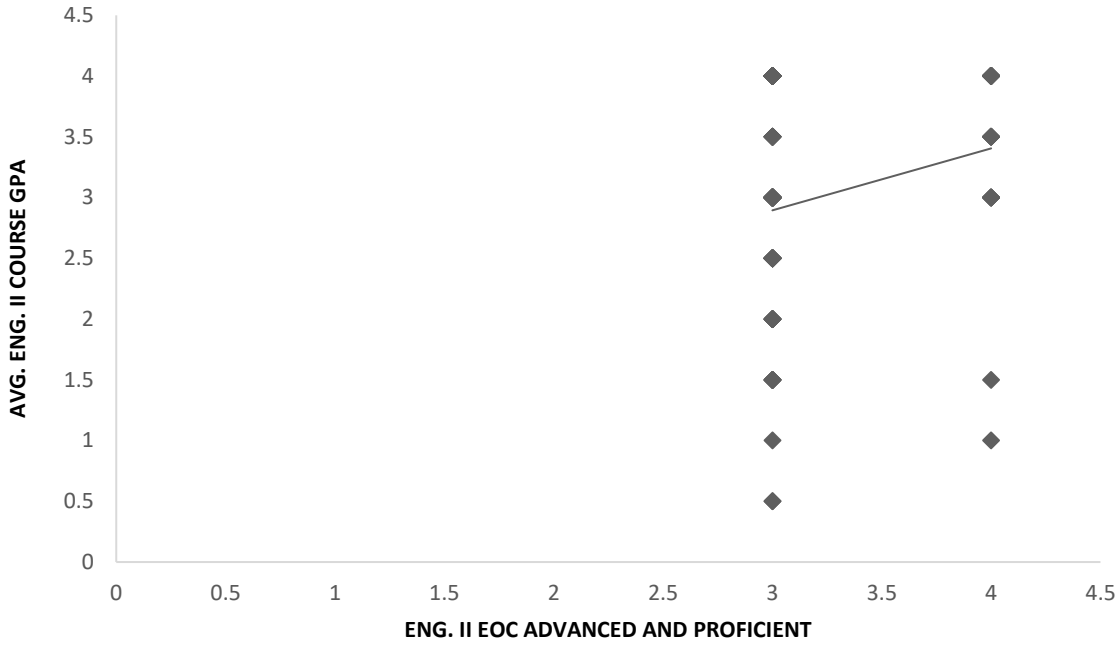


Figure 8. Advanced and proficient vs. grades for 2016.

Finally, the PPMC was performed comparing the EOC advanced and proficient levels on the English II end-of-course assessment and the average course grade point average for English II for all three school years combined. The value of r was 0.3882 (n = 248), and the value of r^2, the coefficient of determination, was 0.1507. Although technically a positive correlation, the relationship was of medium strength as seen in Figure 9.

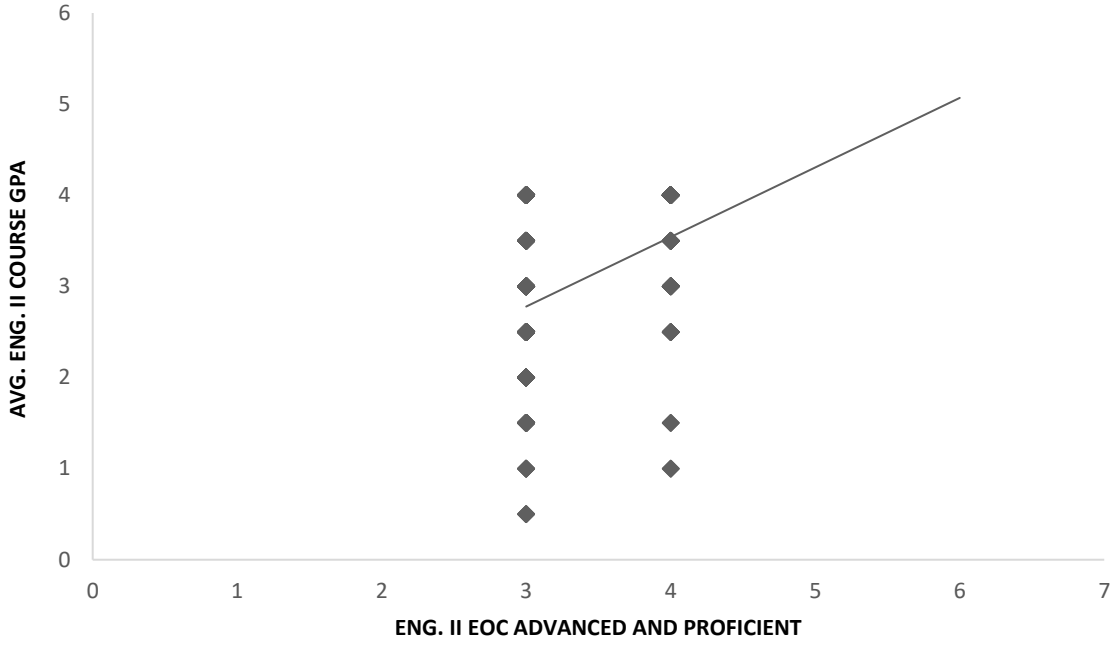


Figure 9. Correlation between three-year advanced and proficient and grades.

Basic and Below Basic vs. Course Grade

Subsequently, the PPMC was performed comparing the EOC basic and below basic levels on the English II end-of-course assessment and the average course grade point average for English II for the 2014 school year. The value of r was -0.3182 (n = 21), and the value of r^2, the coefficient of determination, was 0.1013. Although technically a negative correlation, the relationship was of medium strength as seen in Figure 10.

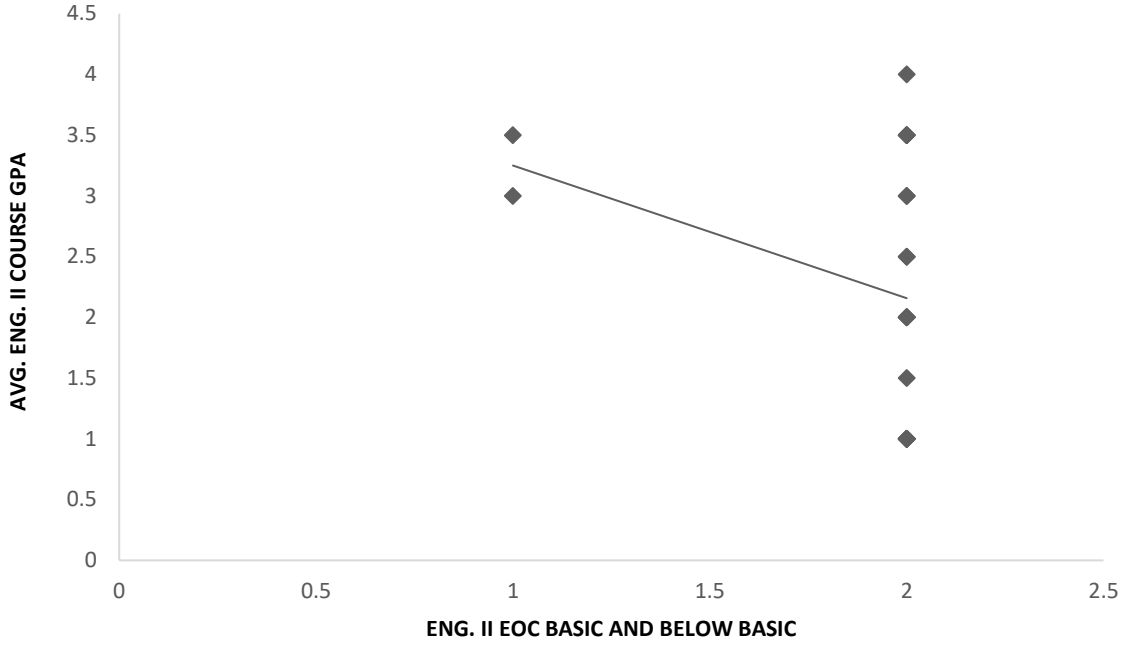


Figure 10. Correlation 2014 basic/below basic and grades.

Additionally, the PPMC was performed comparing the EOC basic and below basic levels on the English II end-of-course assessment and the average course grade point average for English II for the 2015 school year. The value of r was 0.092 (n = 15), and the value of r^2, the coefficient of determination, was 0.0085. Although technically a positive correlation, the relationship was weak as seen in Figure 11.

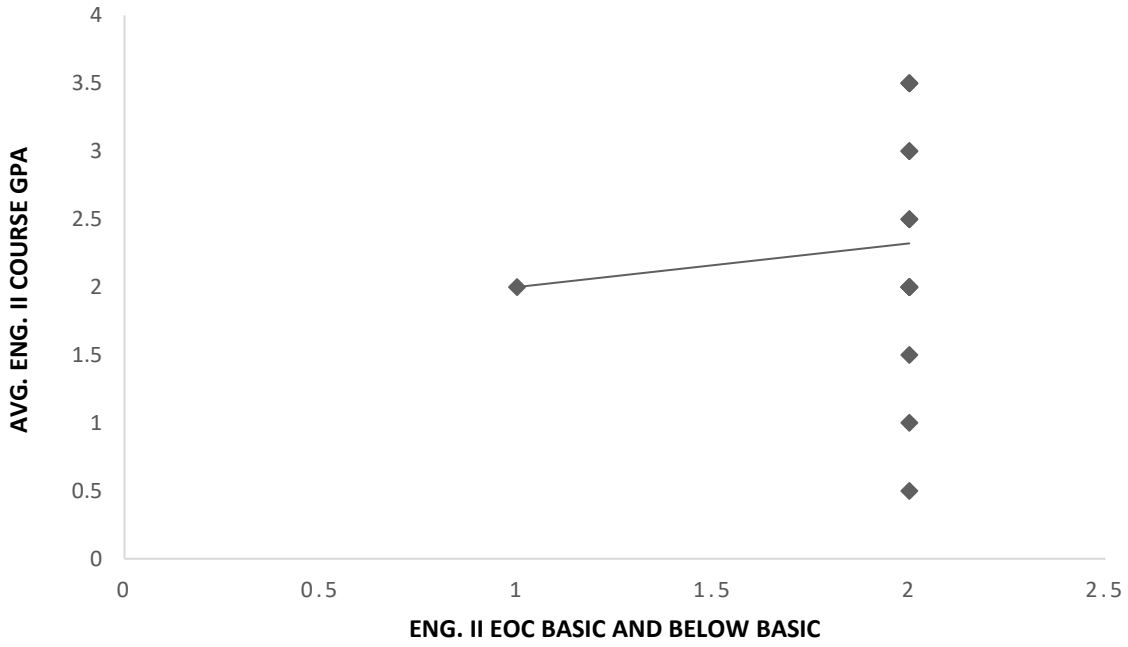


Figure 11. Correlation 2015 basic/below basic and grades.

For the 2016 school year, the PPMC was performed comparing the EOC basic and below basic levels on the English II end-of-course assessment and the average course grade point average for English II. The value of r was -0.2385 (n = 17), and the value of r^2, the coefficient of determination, was 0.0569. Although technically a negative correlation, the relationship was weak as seen in Figure 12.

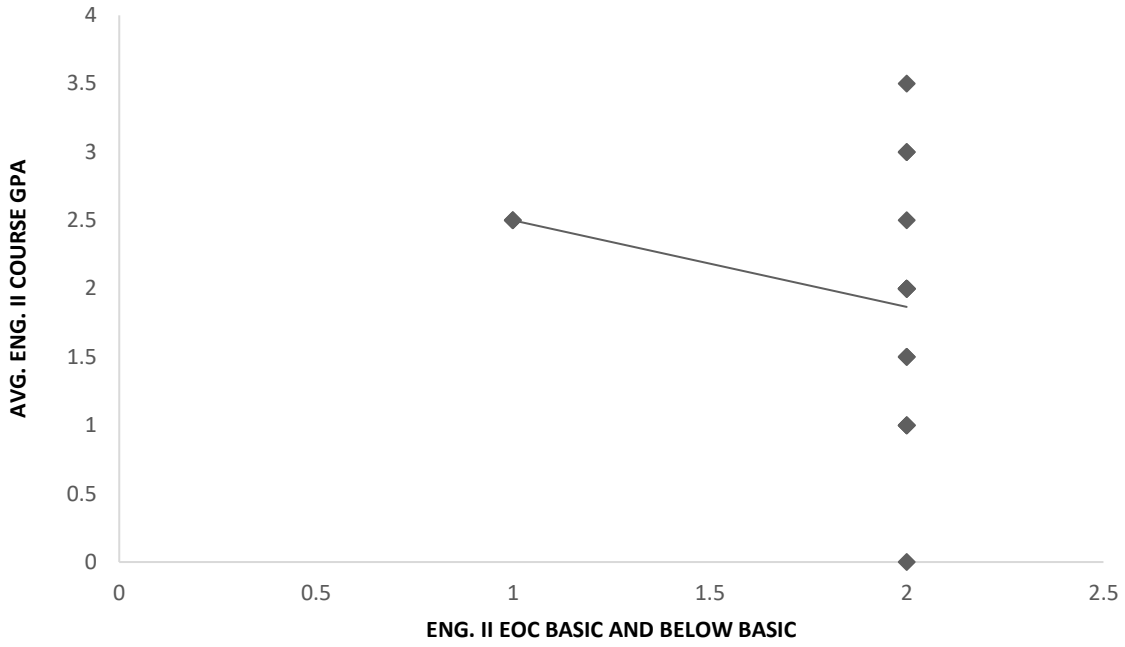


Figure 12. Correlation 2016 basic/below basic and grades.

Finally, the PPMC was performed comparing the EOC basic and below basic levels on the English II end-of-course assessment and the average course grade point average for English II for all three school years combined. The value of r was -0.1827 (n = 53), and the value of r^2, the coefficient of determination, was 0.0334. Although technically a negative correlation, the strength of the relationship was weak as seen in Figure 13.

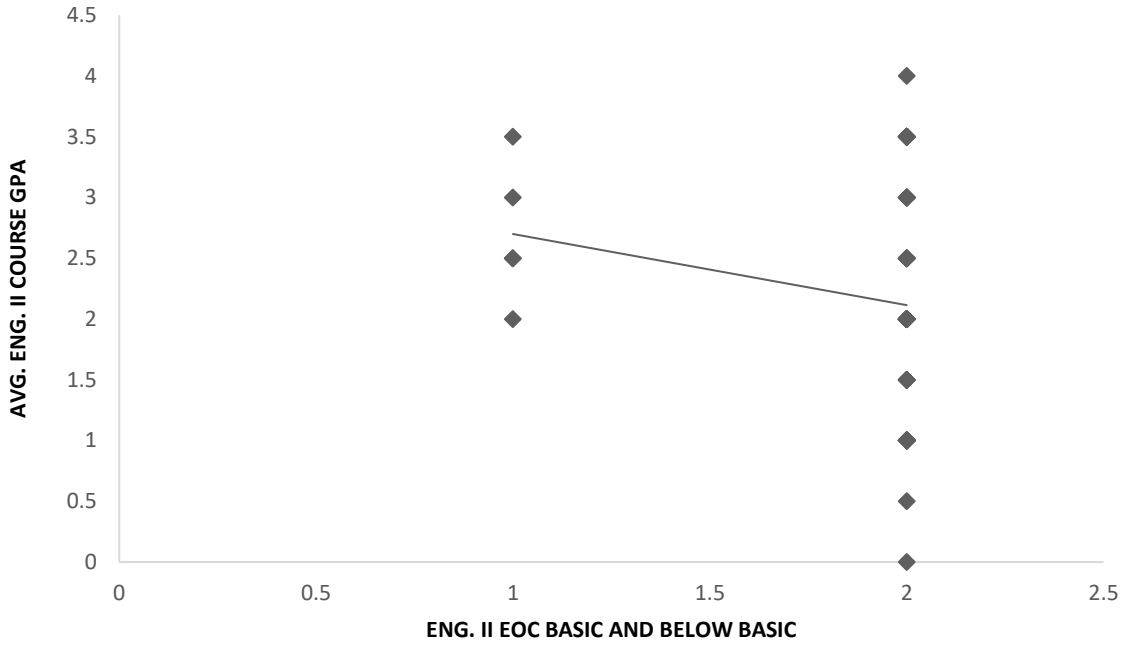


Figure 13. Correlation between three-year basic/below basic and grades.

A and B Grades vs. Proficiency Level

Inversely, the PPMC was performed comparing students with an A or B course grade point average for English II with EOC proficiency levels on the English II end-of-course assessment for the 2014 school year. The value of r was 0.4111 (n = 52), and the value of r^2, the coefficient of determination, was 0.169. Although technically a positive correlation, the relationship was of medium strength as seen in Figure 14.

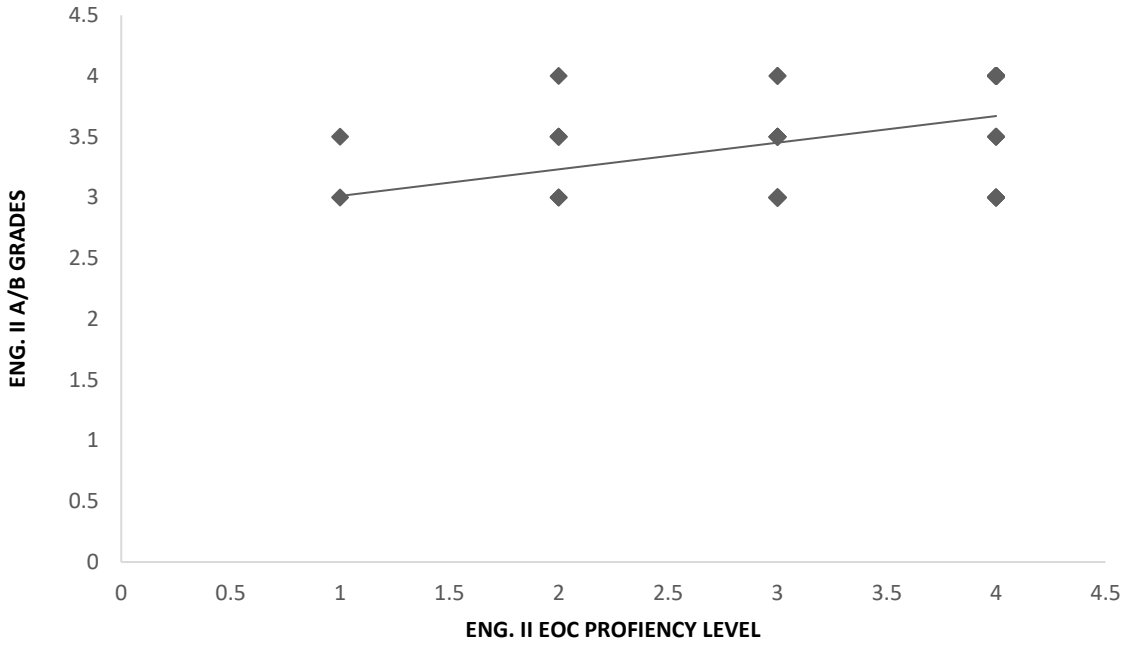


Figure 14. Correlation between 2014 A/B grades and EOC proficiency level.

For the 2015 school year, the PPMC was performed comparing students with an A or B course grade point average for English II with EOC proficiency levels on the English II end-of-course assessment. The value of r was 0.5319 (n = 54), and the value of r^2, the coefficient of determination, was 0.2829. Although showing a positive correlation, the relationship was at the lower range of strong as seen in Figure 15.

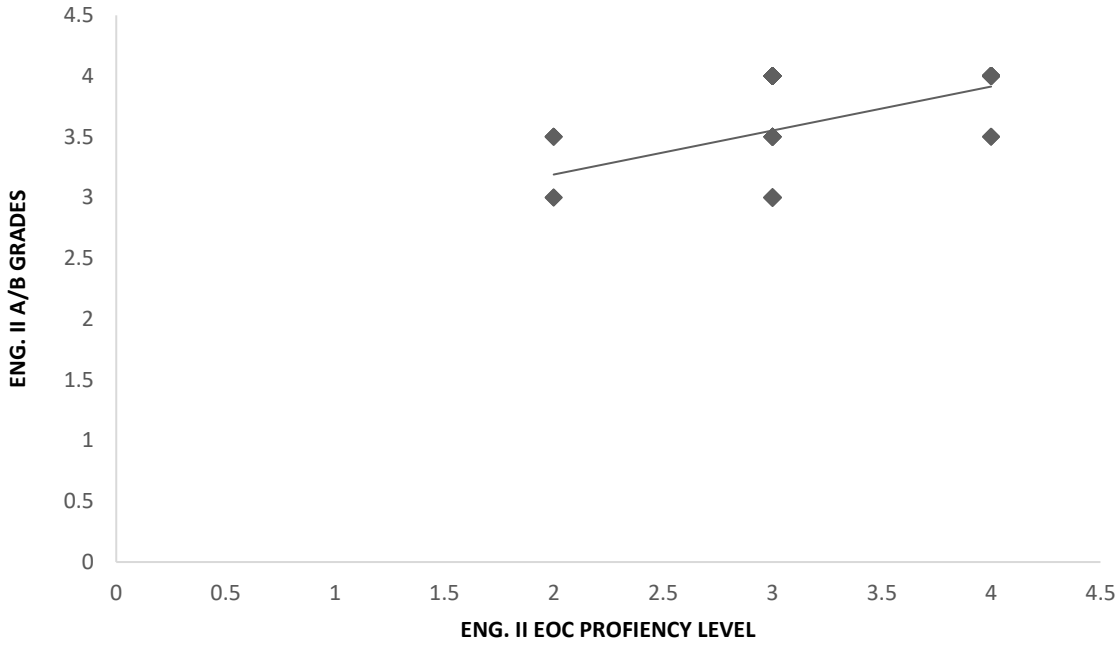


Figure 15. Correlation between 2015 A/B grades and EOC proficiency level.

Next, the PPMC was performed comparing students with an A or B course grade point average for English II with EOC proficiency levels on the English II end-of-course assessment for the 2016 school year. The value of r was 0.267 (n = 71), and the value of r^2, the coefficient of determination, was 0.0713. Although still a positive correlation, the relationship was weak as seen in Figure 16.

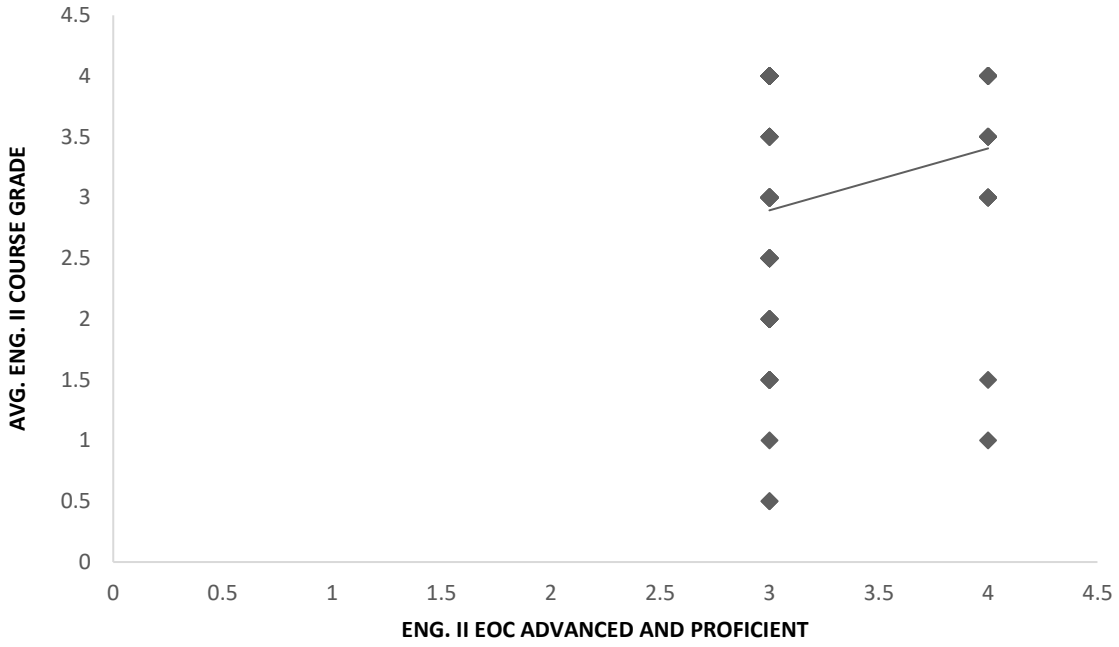


Figure 16. Correlation between 2016 A/B grades and EOC proficiency level.

Lastly, the PPMC was performed comparing students with an A or B course grade point average for English II with EOC proficiency levels on the English II end-of-course assessment for all three school years combined. The value of r was 0.3807 (n = 177), and the value of r^2, the coefficient of determination, was 0.1449. Although technically a positive correlation, the strength of the relationship was medium as seen in Figure 17.

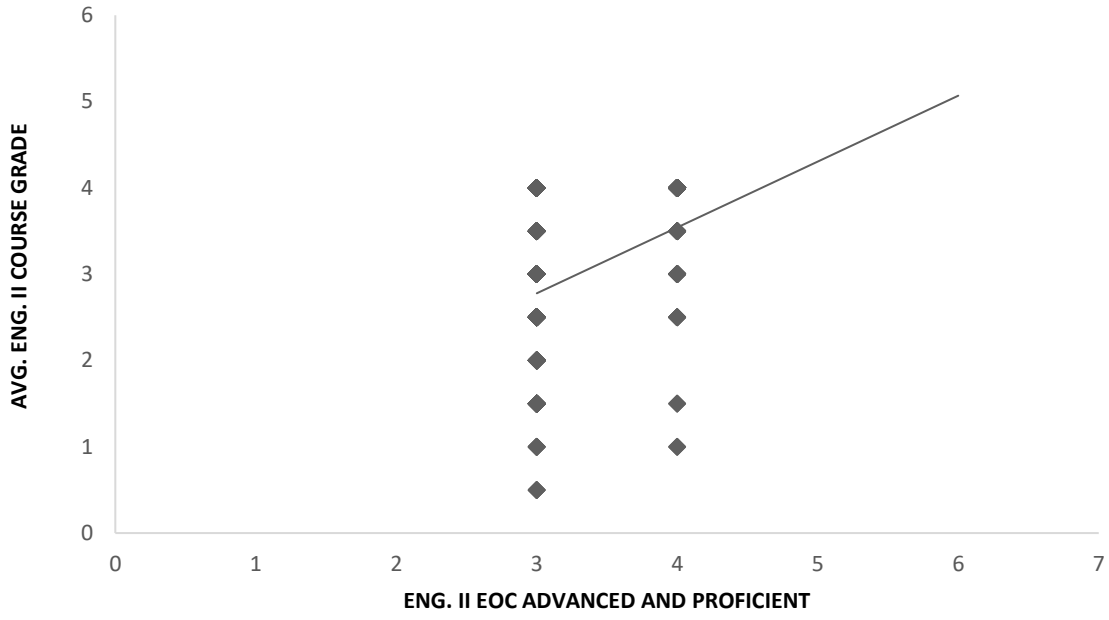


Figure 17. Correlation between three-year A/B grades and EOC proficiency level.

C and Below Grades vs. Proficiency Level

Next, the PPMC was performed comparing students with a C or below course grade point average for English II with EOC proficiency levels on the English II end-of-course assessment for the 2014 school year. The value of r was 0.4636 (n = 43), and the value of r^2, the coefficient of determination, was 0.2149. Although technically a positive correlation, the relationship was of medium strength as seen in Figure 18.

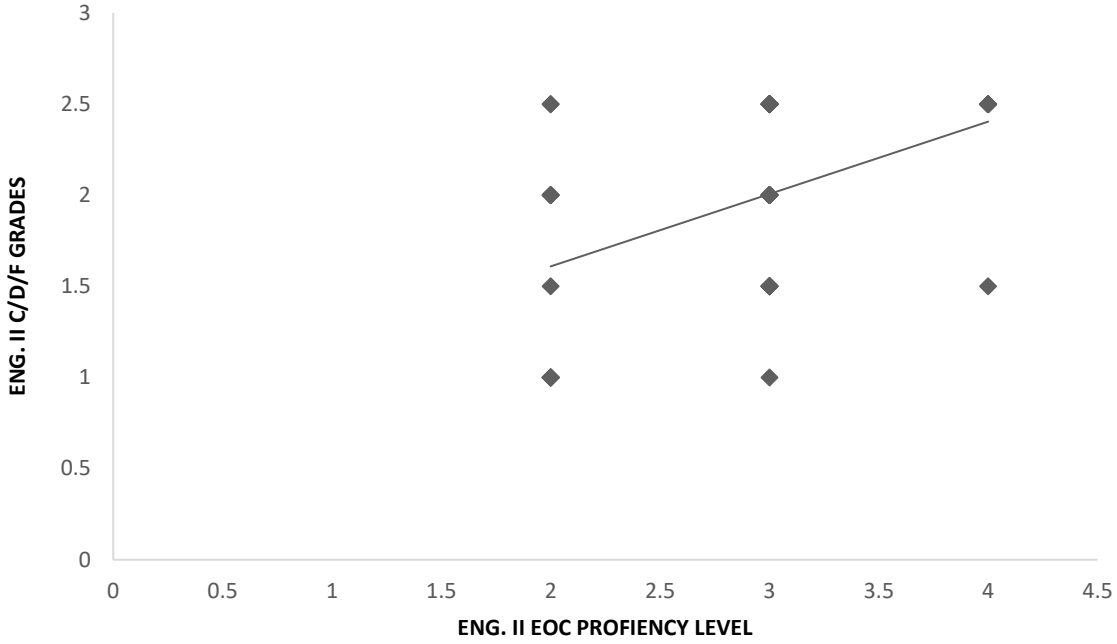


Figure 18. Correlation between 2014 C/D/F grades and EOC proficiency level.

Additionally, the PPMC was performed comparing students with a C or below course grade point average for English II with EOC proficiency levels on the English II end-of-course assessment for the 2015 school year. The value of r was -0.015 (n = 38), and the value of r^2, the coefficient of determination, was 0.0002. Although technically a negative correlation, the relationship was of weak strength as seen in Figure 19.

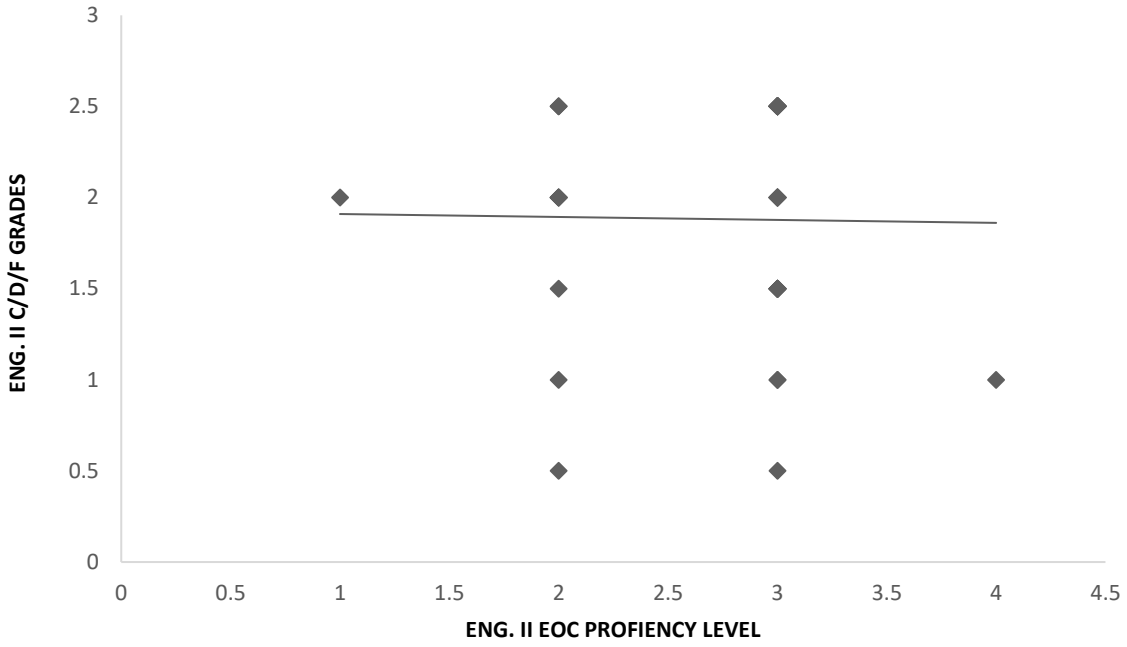


Figure 19. Correlation between 2015 C/D/F grades and EOC proficiency level.

For the 2016 school year, the PPMC was performed comparing students with a C or below course grade point average for English II with EOC proficiency levels on the English II end-of-course assessment. The value of r was 0.0026 (n = 43), and the value of r^2, the coefficient of determination, was 0. Although technically a positive correlation, the relationship was weak as seen in Figure 20.

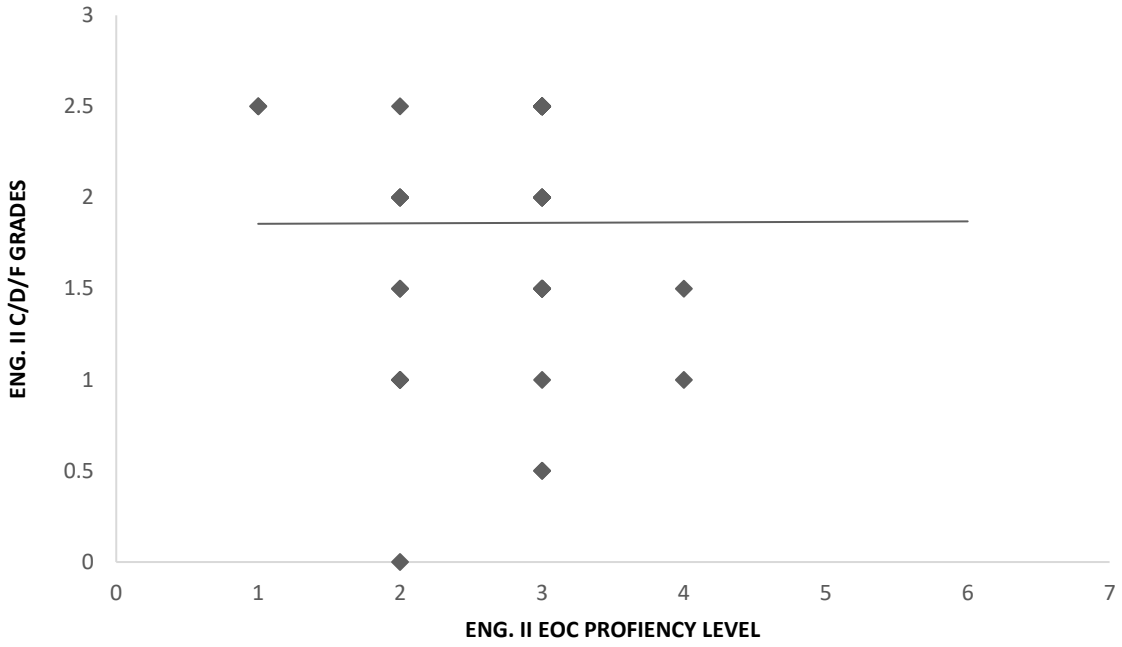


Figure 20. Correlation between 2016 C/D/F grades and EOC proficiency level.

Finally, the PPMC was performed comparing students with a C or below course grade point average for English II with EOC proficiency levels on the English II end-of-course assessment for all three school years combined. The value of r was 0.157 (n = 124), and the value of r^2, the coefficient of determination, was 0.0246. Although technically a positive correlation, the strength of the relationship was weak as seen in Figure 21.

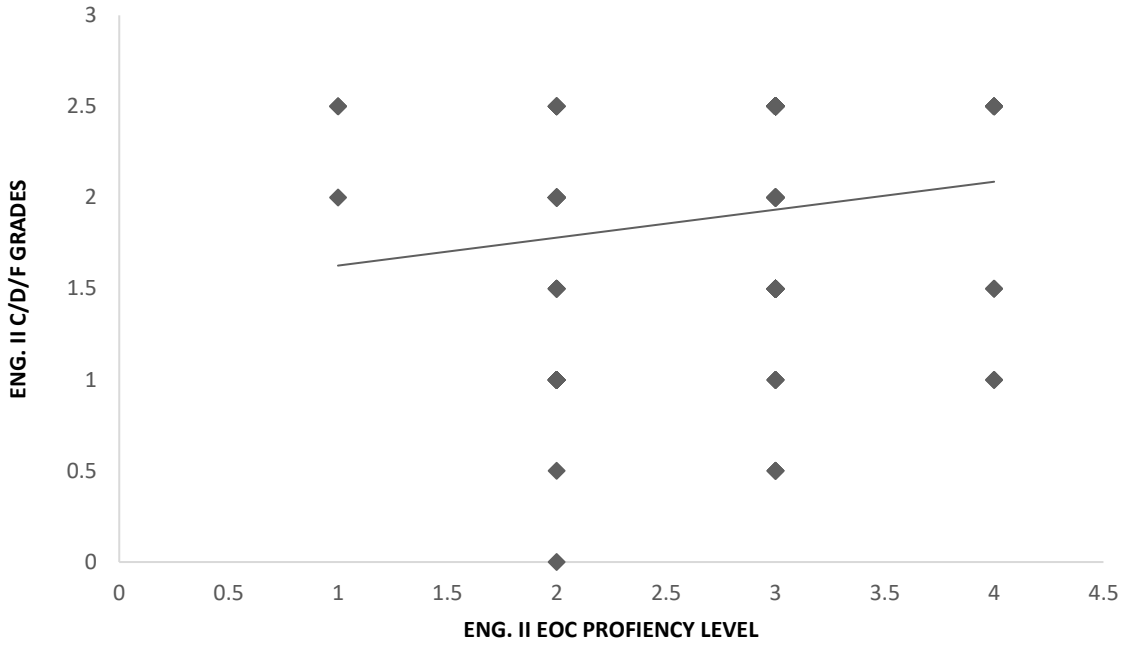


Figure 21. Correlation between three-year C/D/F grades and EOC proficiency level.

Correlation Female English II EOC and Grade

The next PPMC was performed comparing the female students' EOC proficiency levels on the English II end-of-course assessment and the course grade point averages for English II for the 2014 school year. The value of r was 0.5986 (n = 45), and the value of r^2, the coefficient of determination, was 0.3583. Although technically a positive correlation, the relationship was on the lower range of strong as seen in Figure 22.

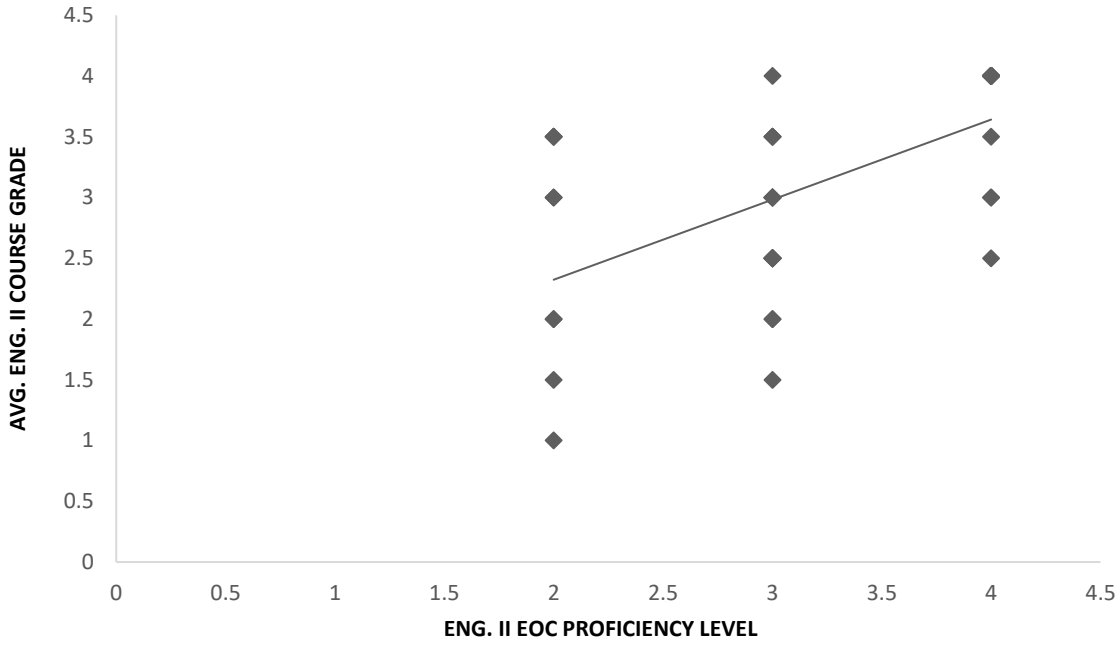


Figure 22. Correlation between 2014 female EOC proficiency level and grade.

Furthermore, the PPMC was performed comparing the female students' EOC proficiency levels on the English II end-of-course assessment and the course grade point averages for English II for the 2015 school year. The value of r was 0.4488 (n = 50), and the value of r^2, the coefficient of determination, was 0.2027. Although technically a positive correlation, the relationship was of medium strength as seen in Figure 23.

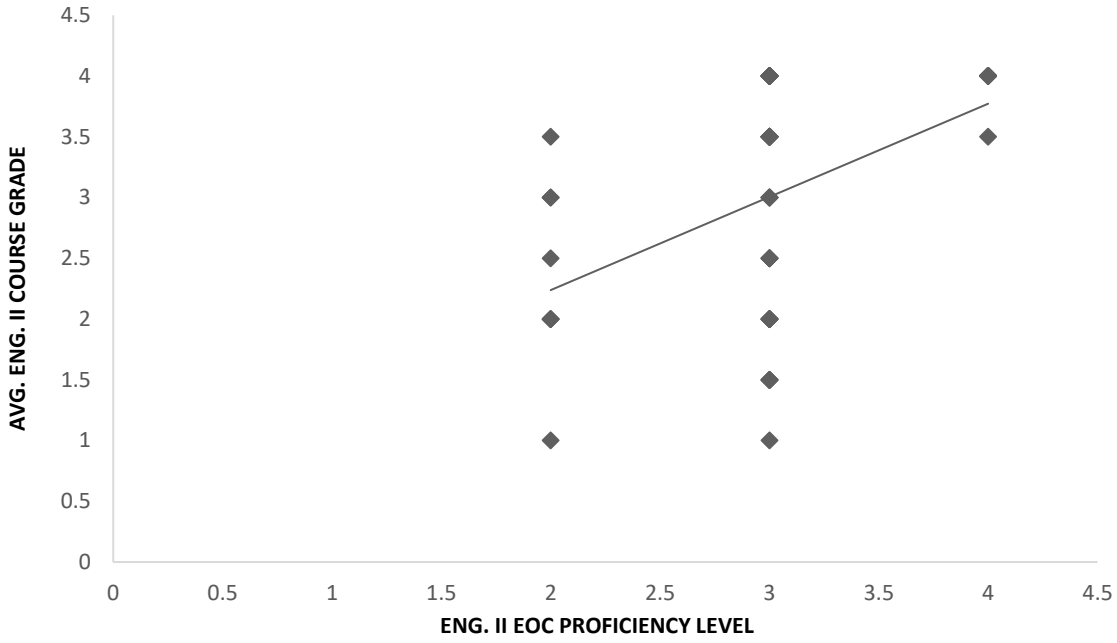


Figure 23. Correlation between 2015 female EOC proficiency level and grade.

Next, the PPMC was performed comparing the female students' EOC proficiency level on the English II end-of-course assessment and the course grade point averages for English II for the 2016 school year. The value of r was 0.4502 (n = 46), and the value of r^2, the coefficient of determination, was 0.2027. Although still a positive correlation, the relationship was of medium strength as seen in Figure 24.

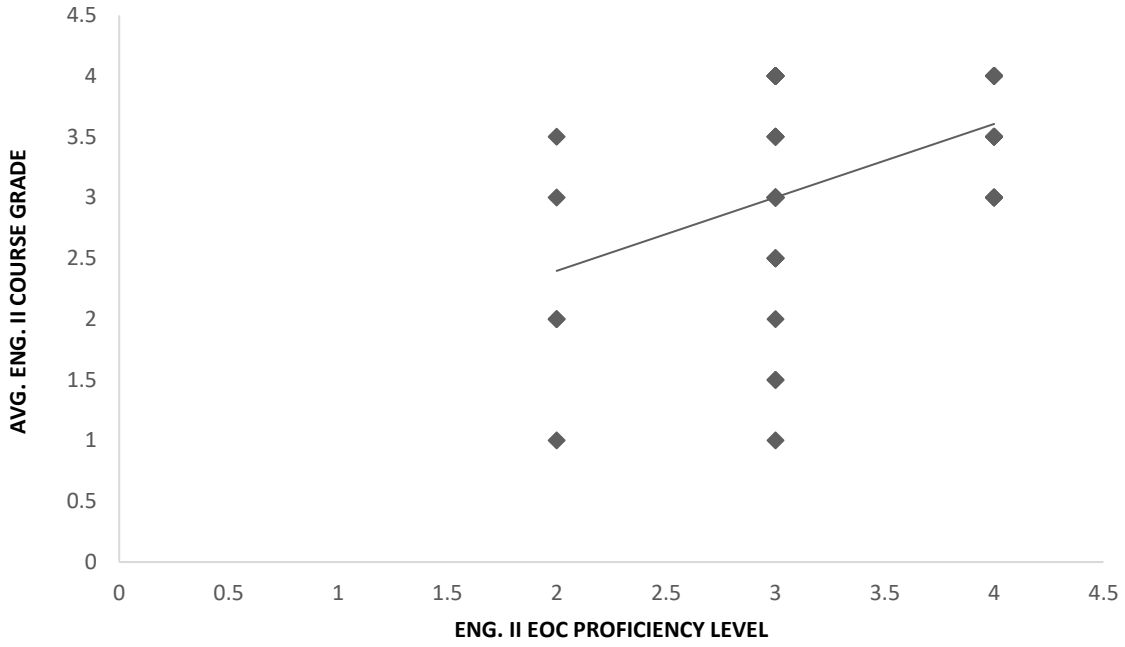


Figure 24. Correlation between 2016 female EOC proficiency level and grade.

Finally, the PPMC was performed comparing the female students' EOC proficiency levels on the English II end-of-course assessment and the course grade point averages for English II for all three school years combined. The value of r was 0.4977 (n = 141), and the value of r^2, the coefficient of determination, was 0.2477. Although still a positive correlation, the relationship was of medium strength as seen in Figure 25.

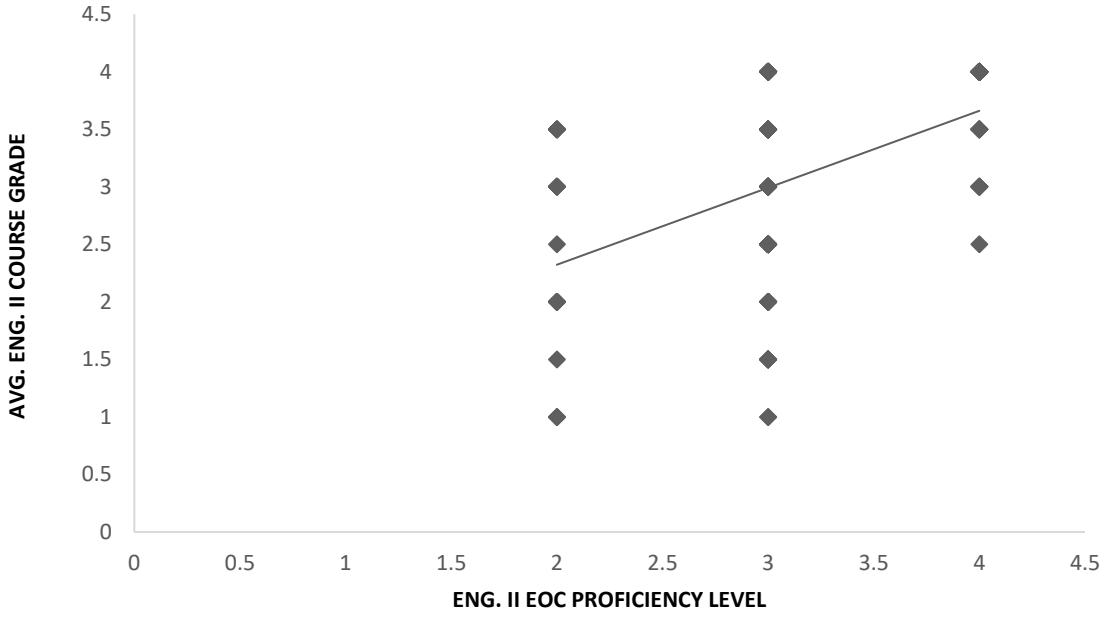


Figure 25. Correlation between three-year female EOC proficiency level and grade.

Correlation Male English II EOC and Grade

Inversely, the PPMC was performed comparing the male students' EOC proficiency levels on the English II end-of-course assessment and the course grade point averages for English II for the 2014 school year. The value of r was 0.364 (n = 50), and the value of r^2, the coefficient of determination, was 0.133. Although technically a positive correlation, the relationship was of medium strength as seen in Figure 26.

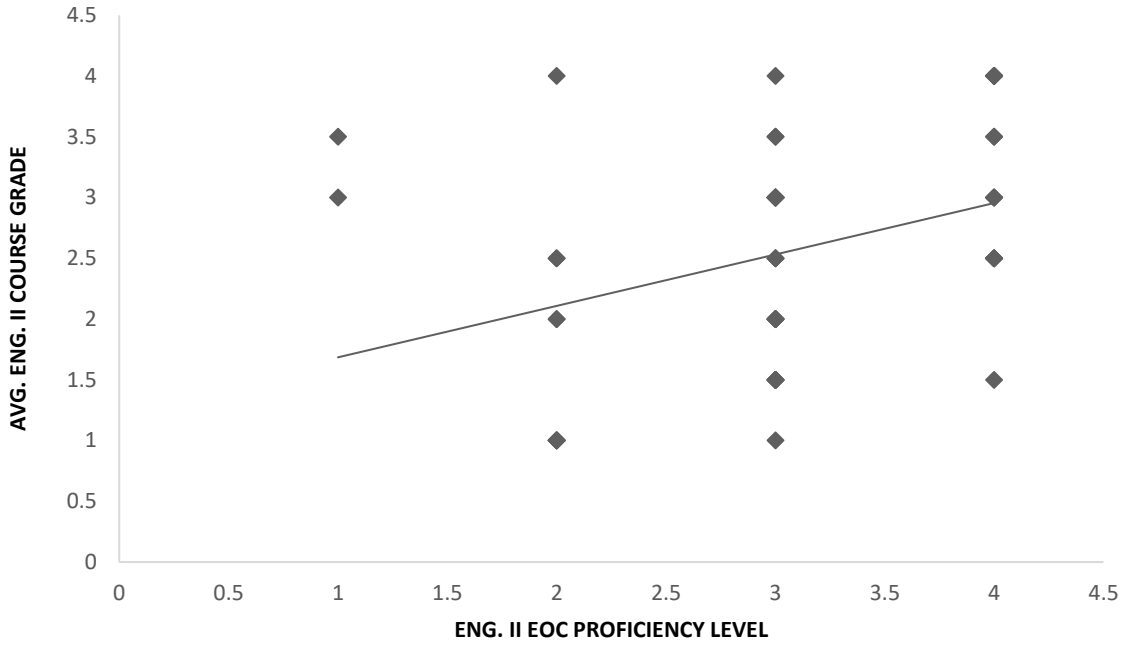


Figure 26. Correlation between 2014 male EOC proficiency level and grade.

Additionally, the PPMC was performed comparing the male students' EOC proficiency levels on the English II end-of-course assessment and the course grade point averages for English II for the 2015 school year. The value of r was 0.4934 (n = 42), and the value of r^2, the coefficient of determination, was 0.2434. Although technically a positive correlation, the relationship was of medium strength as seen in Figure 27.

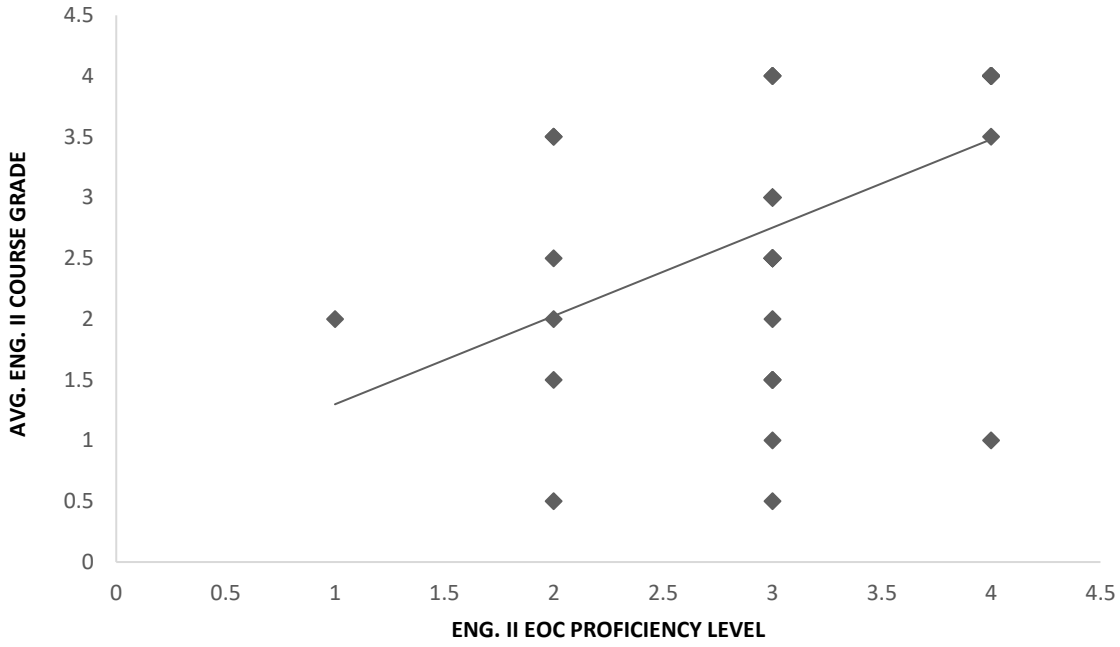


Figure 27. Correlation between 2015 male EOC proficiency level and grade.

Next, the PPMC was performed comparing the male students' EOC proficiency levels on the English II end-of-course assessment and the course grade point averages for English II for the 2016 school year. The value of r was 0.3466 (n = 68), and the value of r^2, the coefficient of determination, was 0.1201. Although still a positive correlation, the relationship was of medium strength as seen in Figure 28.

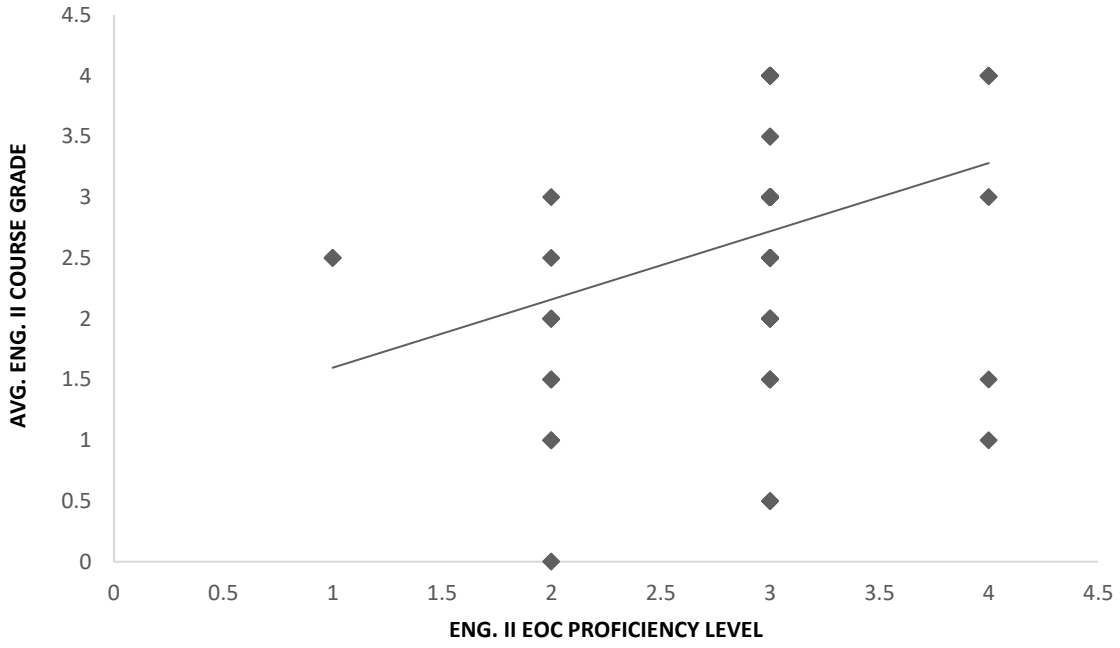


Figure 28. Correlation between 2016 male EOC proficiency level and grade.

Lastly, the PPMC was performed comparing the male students' EOC proficiency levels on the English II end-of-course assessment and the course grade point averages for English II for all three school years combined. The value of r was 0.3913 (n = 160), and the value of r^2, the coefficient of determination, was 0.1531. Although still a positive correlation, the relationship was of medium strength as seen in Figure 29.

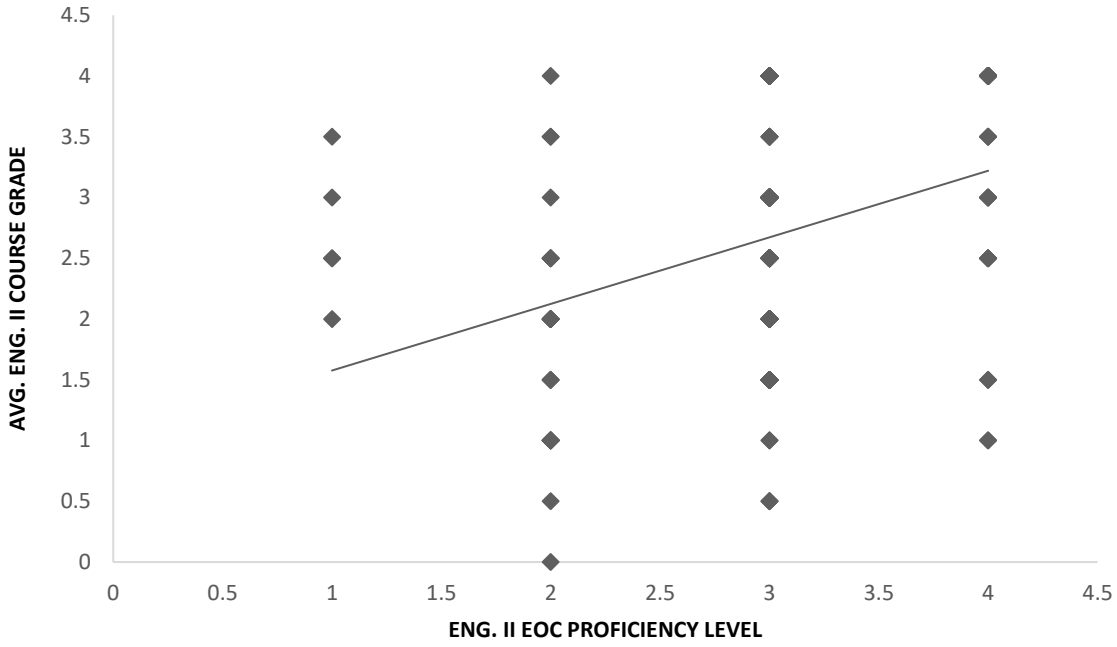


Figure 29. Correlation between three-year male EOC proficiency level and grade.

Analysis of Quantitative Data–Descriptive Statistics

Descriptive statistics were used to analyze the quantitative data from each gender's English II end-of-course exam scores and student grade point averages in English II. According to Ary et al. (2014), descriptive statistics help describe data or summarize data in a meaningful way. Graphs are utilized to display data in an organized and meaningful way.

EOC Proficiency Level and Gender

Figure 30 illustrates the 2014 school year distribution of students scoring in each of the four English II EOC proficiency levels by gender. Both the median and mode for females was three, which was equivalent to the proficient level on the English II EOC. Likewise, the median and mode for males was three. The standard deviation for females was .7474 with 80% being advanced or proficient, while the standard deviation for males

was .8204 with 76% being advanced or proficient. In comparison, Figure 31 shows the 2014 school year distribution of students scoring in each of the four English II EOC proficiency levels by gender across the state with 78.8% of females and 70.5% males advanced or proficient.

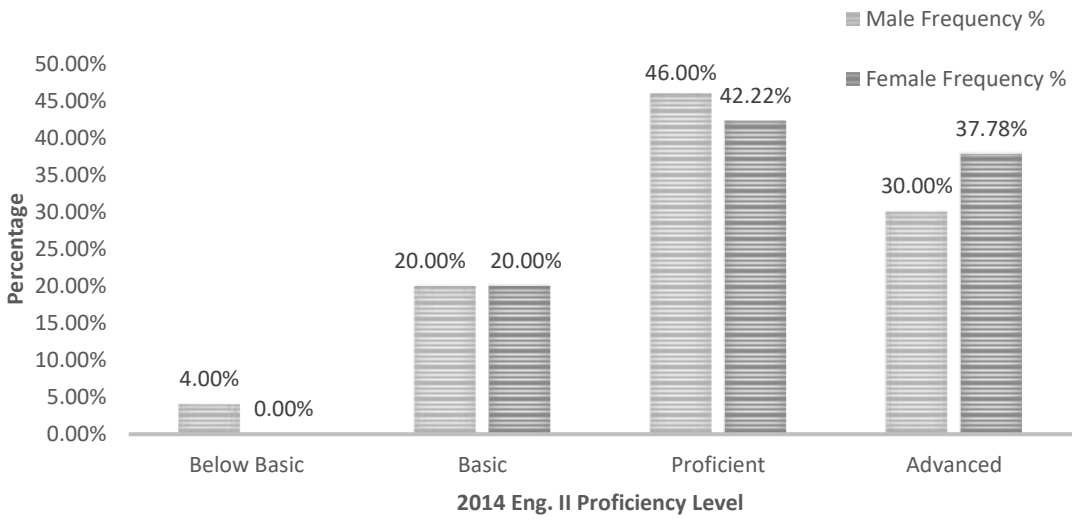


Figure 30. Percentage of English II proficiency level by gender for 2014.

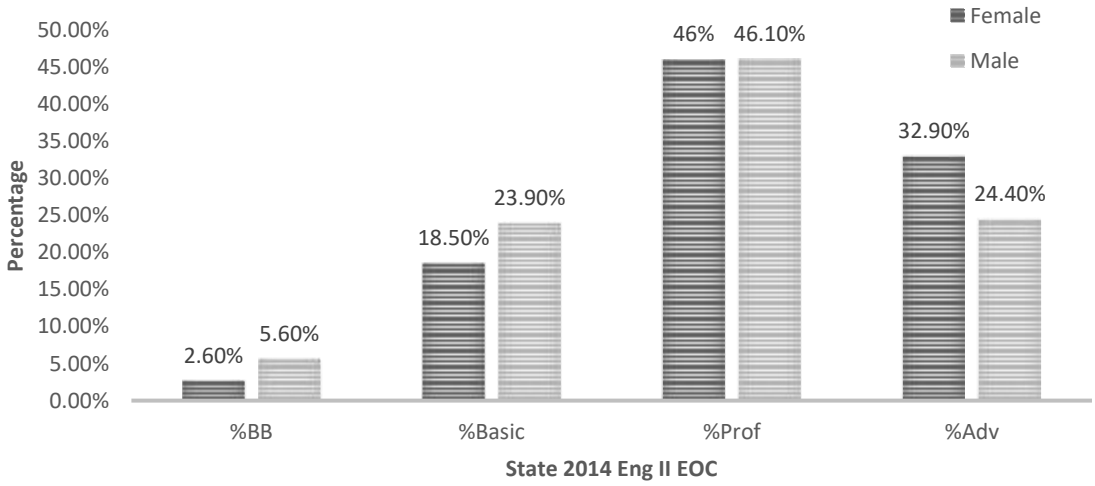


Figure 31. State percentage of English II proficiency level by gender for 2014.

Figure 32 displays the 2015 school year distribution of students scoring in each of the four English II EOC proficiency levels by gender. Both the median and mode for females was three, which was equivalent to proficient on the English II EOC. Similarly, the median and mode for males was three. The standard deviation for females was .5529 with 84.0% scoring advanced or proficient, while the standard deviation for males was .7262 with 83.33% scoring advanced or proficient. In comparison, Figure 33 shows 2015 school year distribution of students scoring in each of the four English II EOC proficiency levels by gender across the state with 78.0% of females and 69.3% males advanced or proficient.

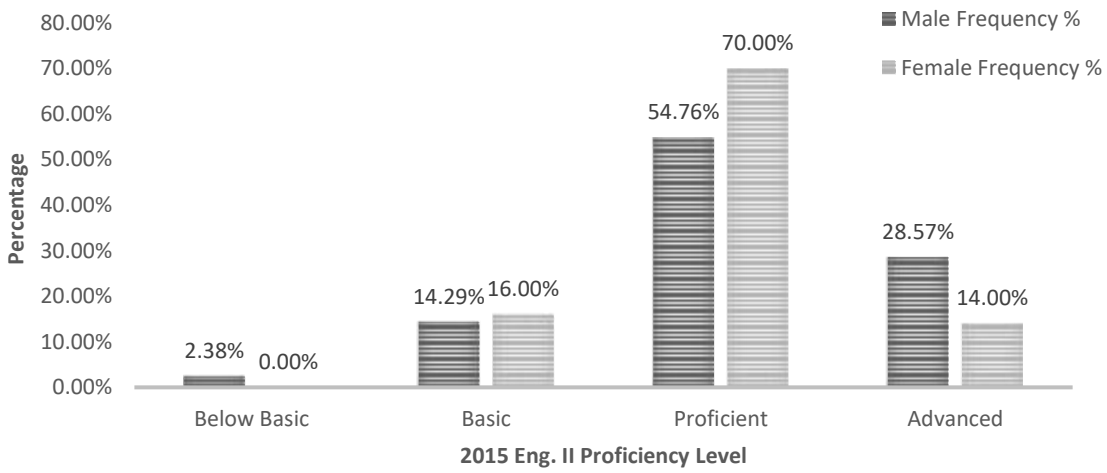


Figure 32. Percentage of English II proficiency level by gender for 2015.

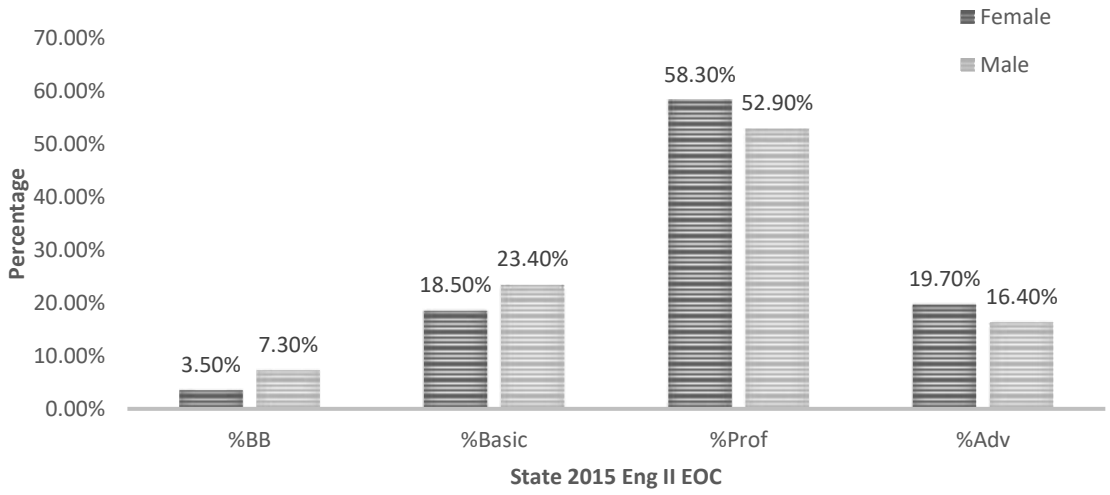


Figure 33. State percentage of English II proficiency level by gender for 2015.

Figure 34 demonstrates the 2016 school year distribution of students scoring in each of the four English II EOC proficiency levels by gender. Both the median and mode for females was three, which was equivalent to proficient on the English II EOC. Likewise, the median and mode for males was three. The standard deviation for females was .6189 with 89.13% advanced or proficient, while the standard deviation for males was .602 82.35% advanced or proficient. In comparison, Figure 35 shows 2016 school year distribution of students scoring in each of the four English II EOC proficiency levels by gender across the state with 83.9% of females and 74.5% males advanced or proficient.

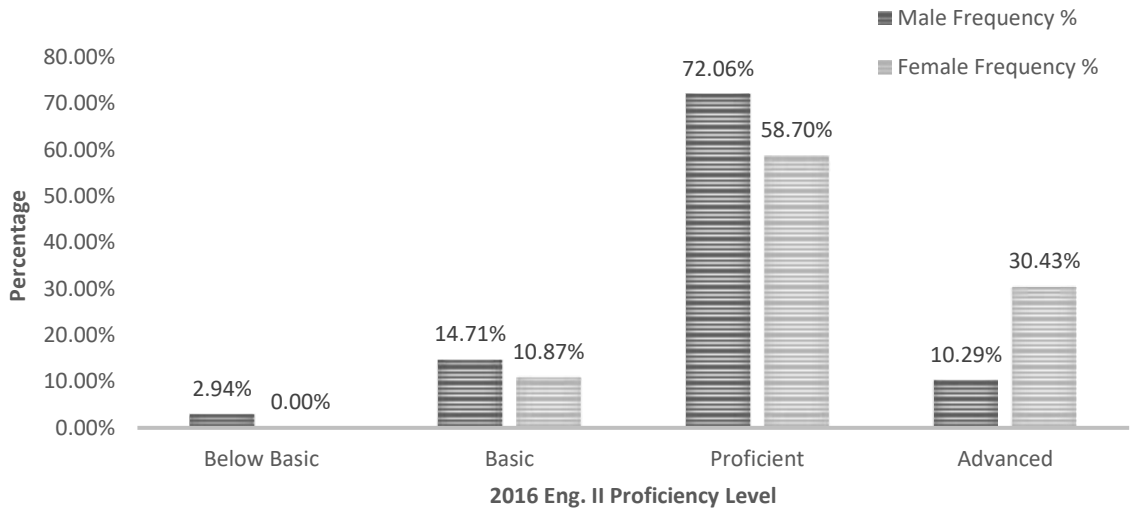


Figure 34. Percentage of English II proficiency level by gender for 2016.

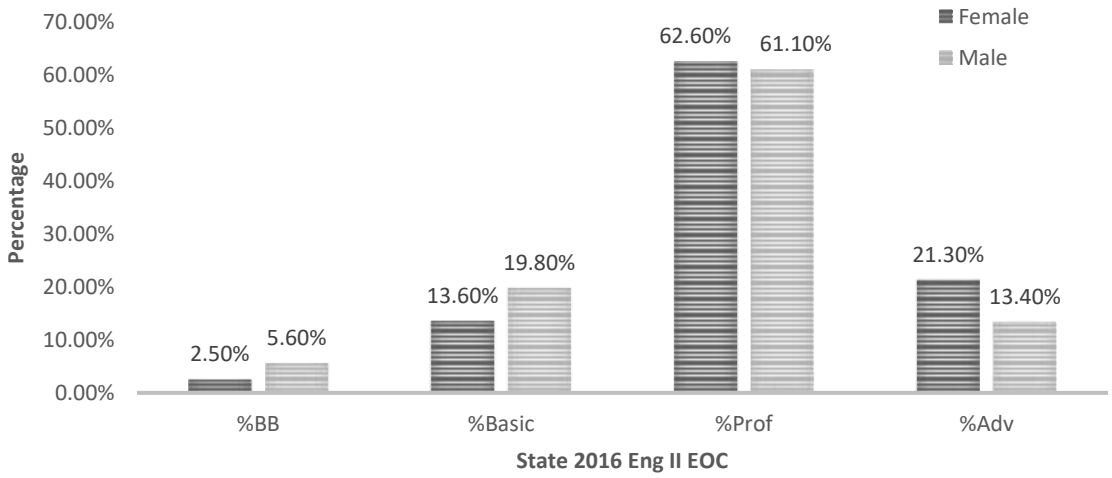


Figure 35. Percentage of English II proficiency level by gender for 2016.

Figure 36 represents three years of combined data showing the distribution of students scoring in each of the four English II EOC proficiency levels by gender. Both the median and mode for females was three, which was equivalent to proficient on the

English II EOC. Equally, the median and mode for males was three. The standard deviation for females was .6447 with 84.40% advanced or proficient, while the standard deviation for males was .7092 with 80.63% advanced or proficient.

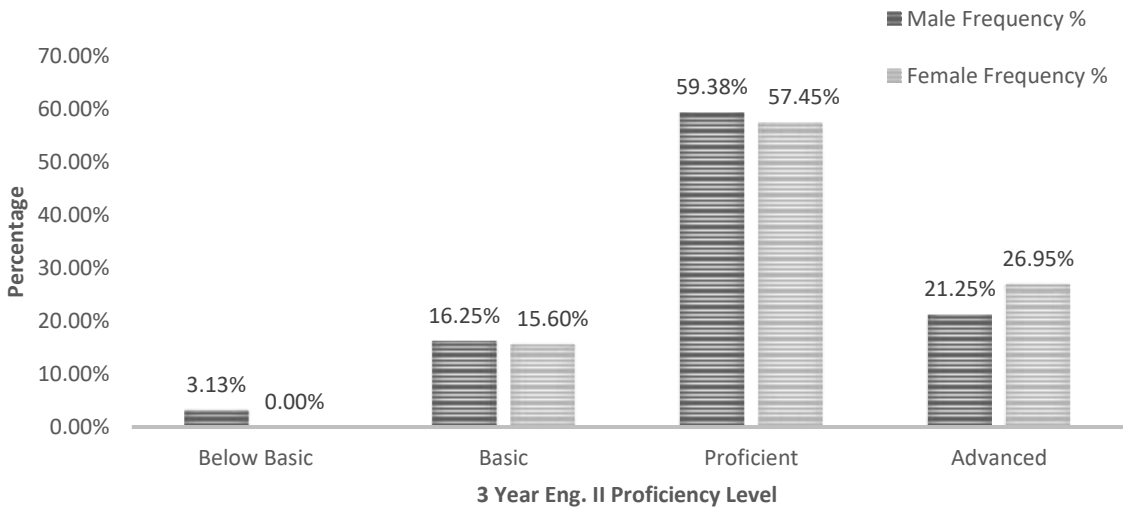


Figure 36. Percentage of English II proficiency level by gender for three years combined.

Grades and Gender

Figure 37 displays the 2014 school year distribution of students' average GPAs in English II by gender. The median for females was three and mode for females was four with 31.11% earning a 4.0 GPA. Similarly, the median for male students was 2.5 and mode for males was three with 14.00% earning a 4.0 GPA. The standard deviation for females was .823, while the standard deviation for males was .9521.

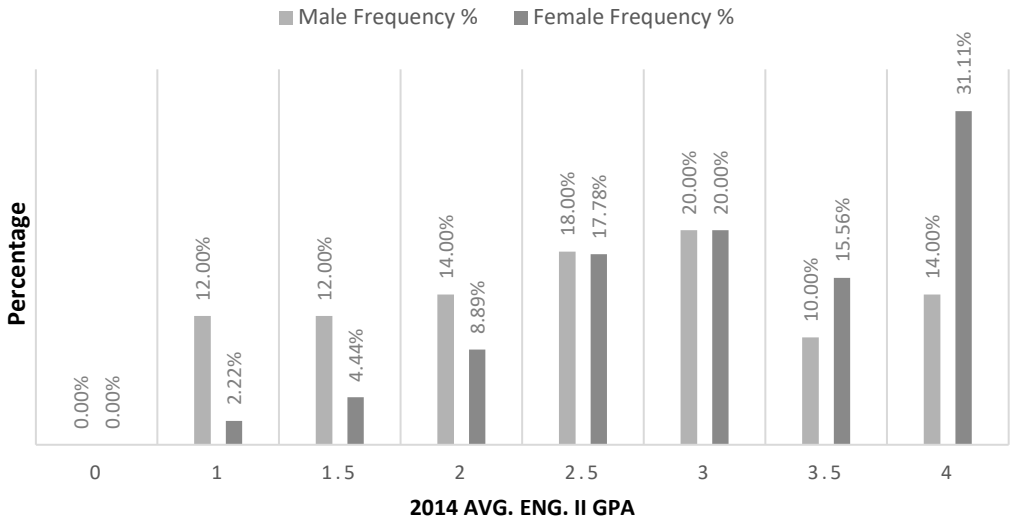


Figure 37. Percentage of English II average GPA by gender for 2014.

Figure 38 illustrates the 2015 school year distribution of students' average GPAs in English II by gender. The median for females was 3.25 and mode for females was four with 30.95% earning a 4.0 GPA. Correspondingly, the median for male students was three and mode for males was four with 30.00% earning a 4.0 GPA. The standard deviation for females was .823, while the standard deviation for males was 1.07.

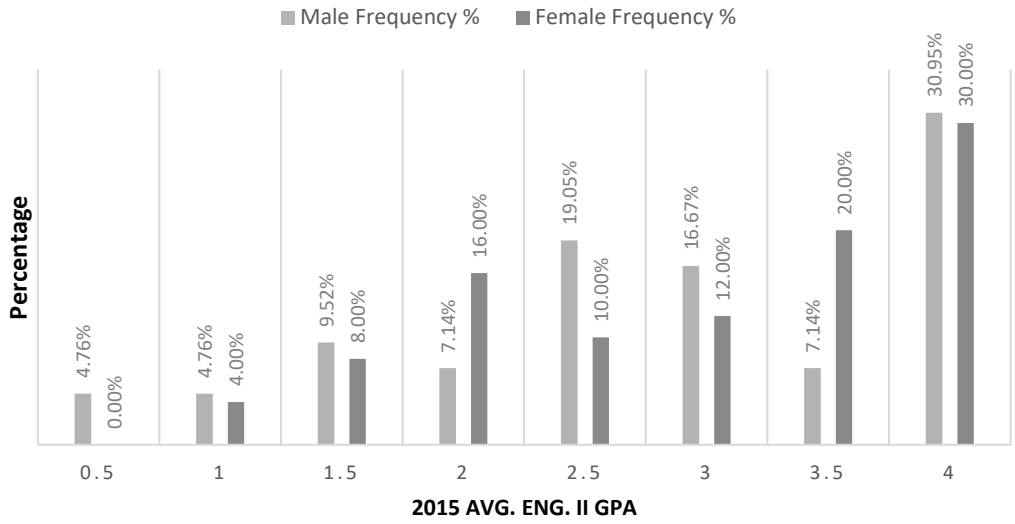


Figure 38. Percentage of English II average GPA by gender for 2015.

Figure 39 elucidates the 2016 school year distribution of students' average GPAs in English II by gender. The median for females was three and mode for females was four with 28.26% earning a 4.0 GPA. Likewise, the median for male students was three and mode for males was three with 19.12% earning a 4.0 GPA. The standard deviation for females was .8312, while the standard deviation for males was .9752.

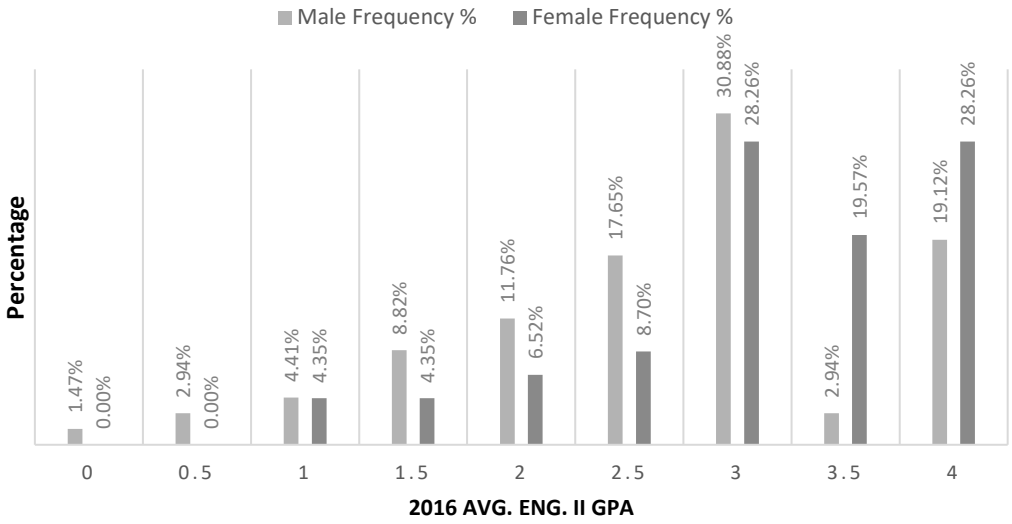


Figure 39. Percentage of English II average GPA by gender for 2016.

Finally, Figure 40 represents three years of combined data showing the distribution of students' average GPAs in English II by gender. The median for females was three and mode for females was four with 29.79% earning a 4.0 GPA. Similarly, the median for male students was three and mode for males was three with 20.63% earning a 4.0 GPA. The standard deviation for females was .8665, while the standard deviation for males was .9933.

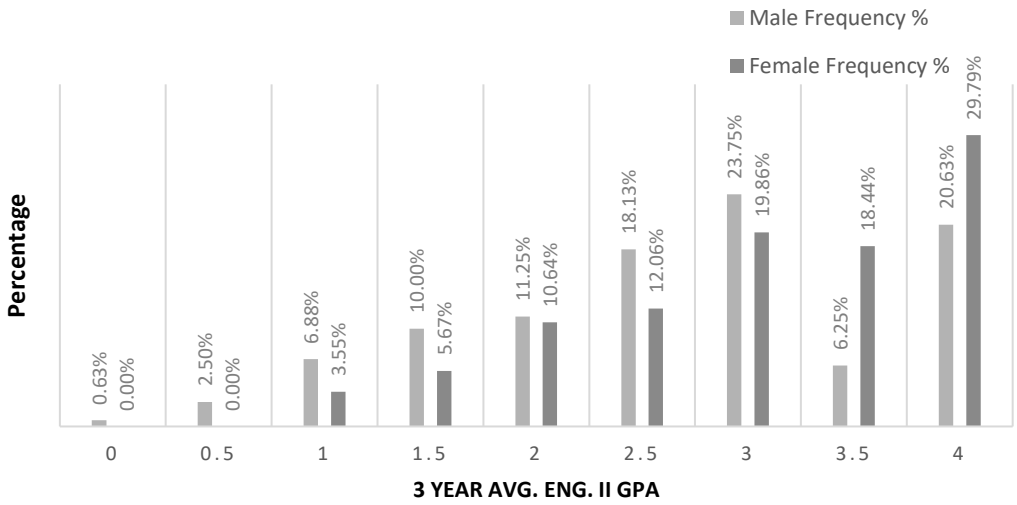


Figure 40. Percentage of English II average GPA by gender for three years combined.

Analysis of Survey Data

An online survey was administered to participating high school principals. The survey participants were given two weeks to complete the online survey. Principals from six of the seven schools participated in the survey. The survey utilized a Likert-type scale composed of a series of four trait items that are combined into a single composite score to provide a quantitative measure (Boone & Boone, 2012). Responses from principals surveyed were tabulated and shown in tables corresponding to each of the 10 survey questions/statements.

Survey question one. A student's course grade accurately shows that student's mastery of the course.

The weighted average Likert-type scale score for this question was 2.67. While the majority agreed with this statement, as shown in Figure 41, there were no strongly agree responses. Likewise, none of the principals strongly disagreed a student's grade demonstrates mastery of the course.

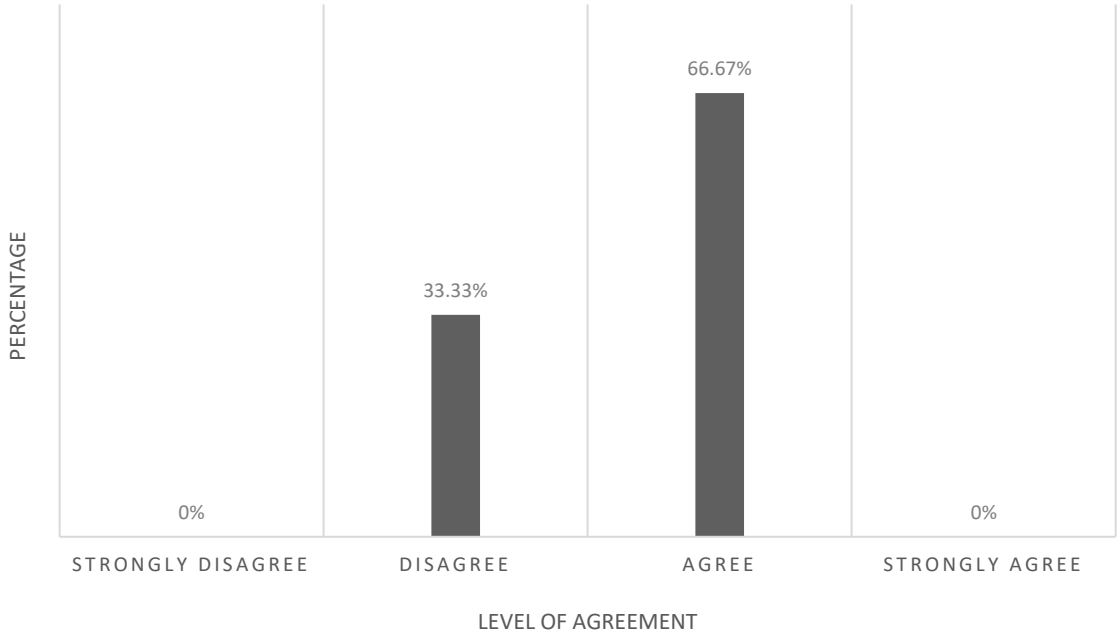


Figure 41. Survey question one results. Grade shows mastery.

Survey question two. The best way to calculate a course grade is to average the points of all assignments.

Figure 42 shows a similar distribution of responses to Figure 38. Four of the six principals agreed the best way to figure a grade is by averaging the points of all assignments. The weighted average Likert-type scale score for this question was 2.67.

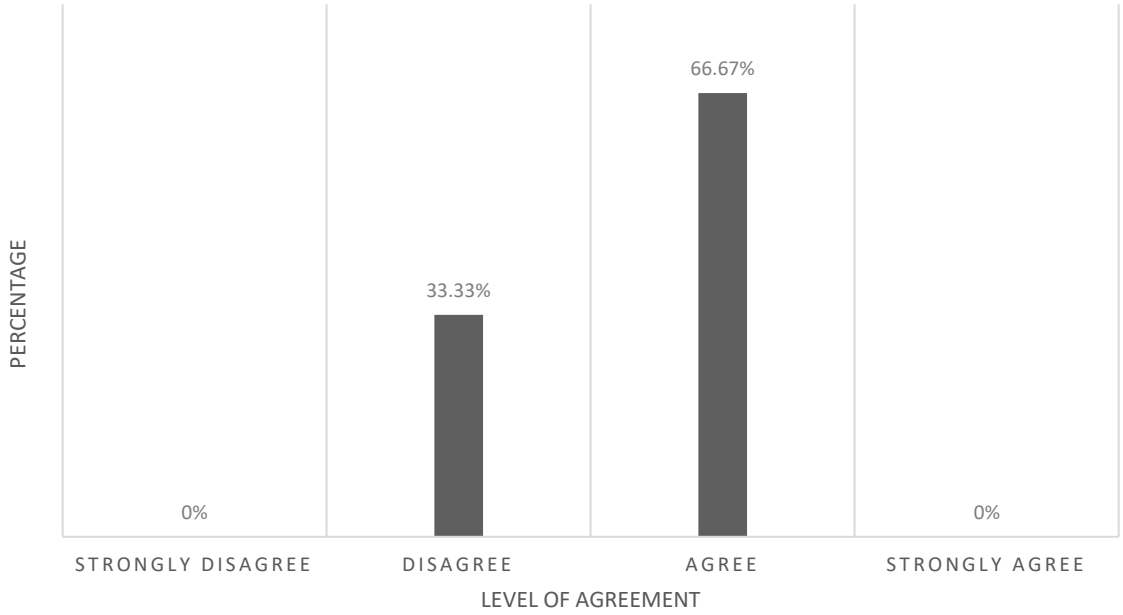


Figure 42. Survey question two results. Grade calculation.

Survey question three. Course grades are free of gender bias.

As shown in Figure 43, a strong majority of principals surveyed believed course grades are free of gender bias. With only one principal disagreeing, the weighted average Likert-type scale score for this question was 2.83. Additionally, none of the principals surveyed gave strong responses either for or against the presence of gender bias in course grades.

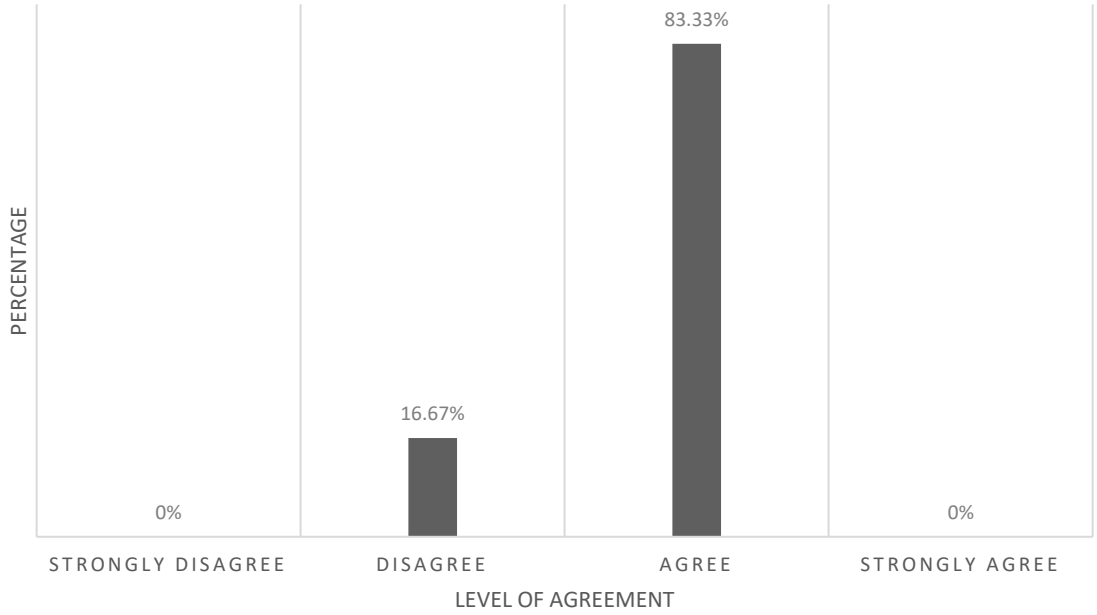


Figure 43. Survey question three results. Grades are free of gender bias.

Survey question four. As a result of high-stakes testing, teachers use standardized test data to help guide and improve instruction.

The weighted average Likert-type scale score for this question was 3.17. Figure 44 shows while half of the respondents agreed high-stakes testing helps guide teachers to improve instruction, one-third strongly agreed. None of the participants strongly disagreed.

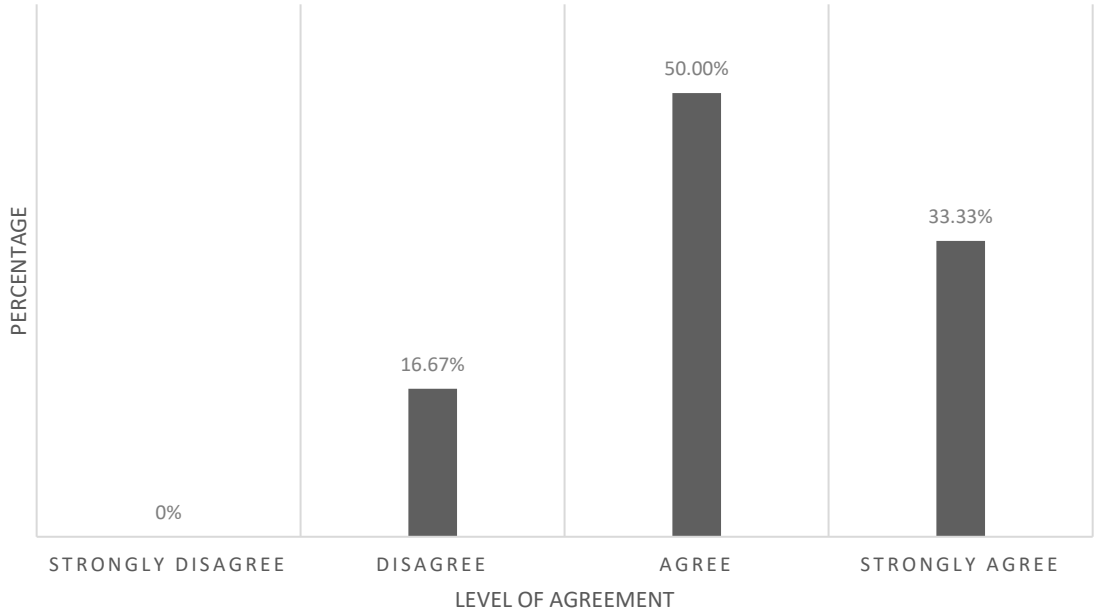


Figure 44. Survey question four results. Teachers use test data to guide and improve instruction.

Survey question five. Teachers need to be held accountable through high-stakes tests to motivate them to teach the content students need to know.

As seen in Figure 45, over half of the principals surveyed disagreed or strongly disagreed with teachers being held accountable through high-stakes tests in order to motivate teaching of content. However, two of the six principals agreed with high-stakes testing to hold teachers accountable. The weighted average Likert-type scale score for this question was 2.17.

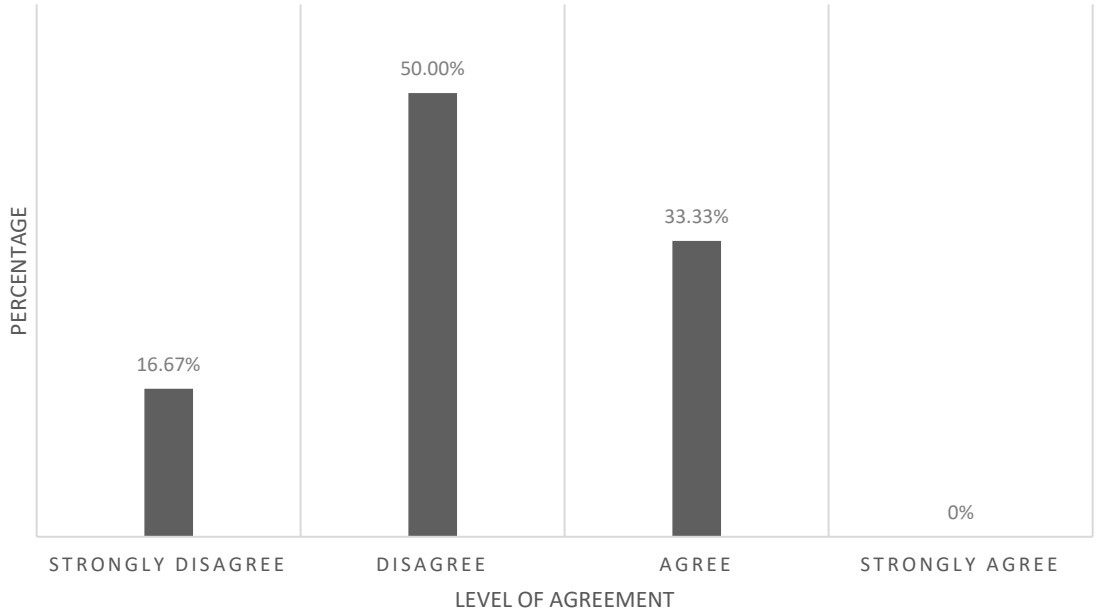


Figure 45. Survey question five results. Teacher's accountability through high-stakes tests.

Survey question six. A student's score on standardized assessments accurately shows the student's mastery of the subject matter.

The weighted average Likert-type scale score for this question was 1.67. As seen in Figure 46, none of the participants agreed standardized assessment scores show the student's mastery of the subject. Inversely, two of the six strongly disagreed.

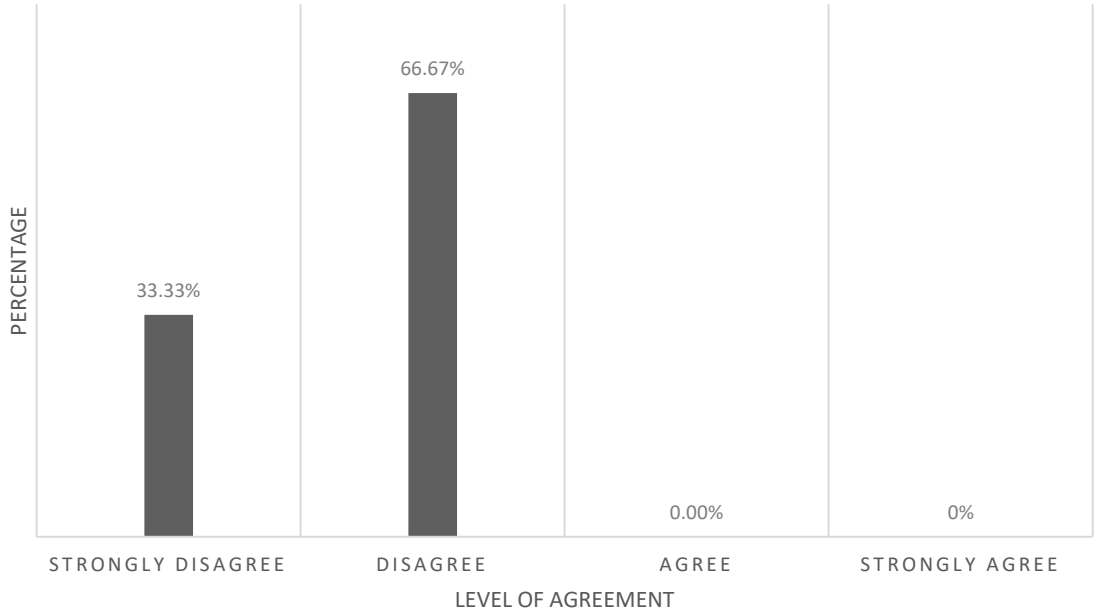


Figure 46. Survey question six results. Assessments show students' mastery.

Survey question seven. Standardized state assessments are free of gender bias.

The weighted average Likert-type scale score for this question was 2.83. While the majority agreed with this statement, as shown in Figure 47, there were no strongly agree responses. Likewise, none of the principals strongly disagreed state assessments are free of gender bias.

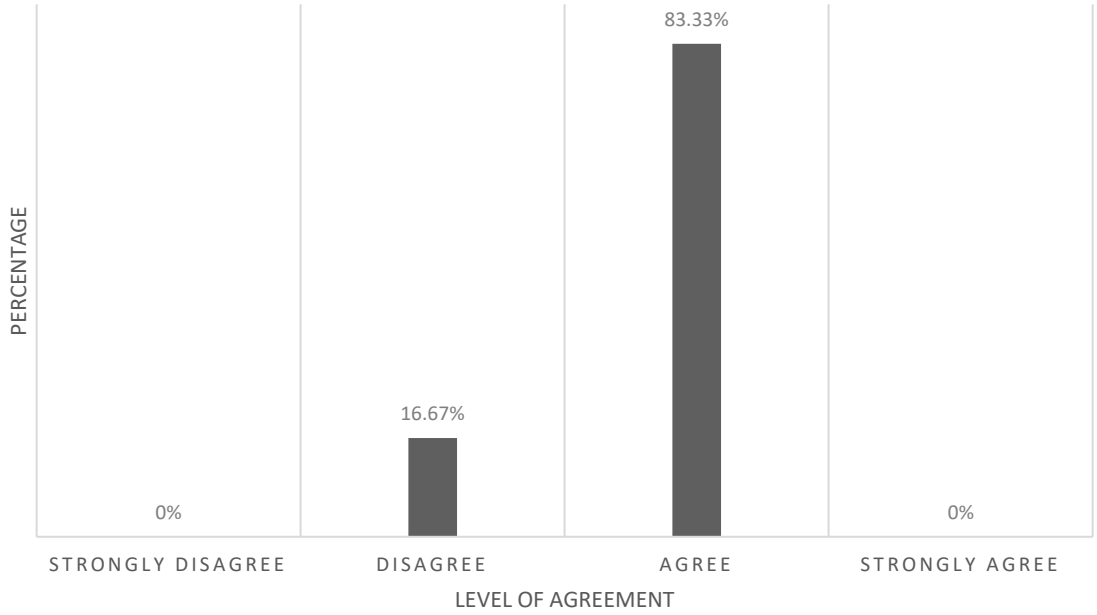


Figure 47. Survey question seven results. Assessments free of gender bias.

Survey question eight. Standardized state assessment scores motivate administrators to ensure standards on which the tests are based are part of the curriculum being taught.

As shown in Figure 48, 83.34% of the principals agreed or strongly agreed state assessments motivate administrators to ensure assessments standards are being taught. The weighted average Likert-type scale score for this question was 3.0. Only one participant disagreed standardized state assessment scores motivate administrators to ensure standards are being taught.

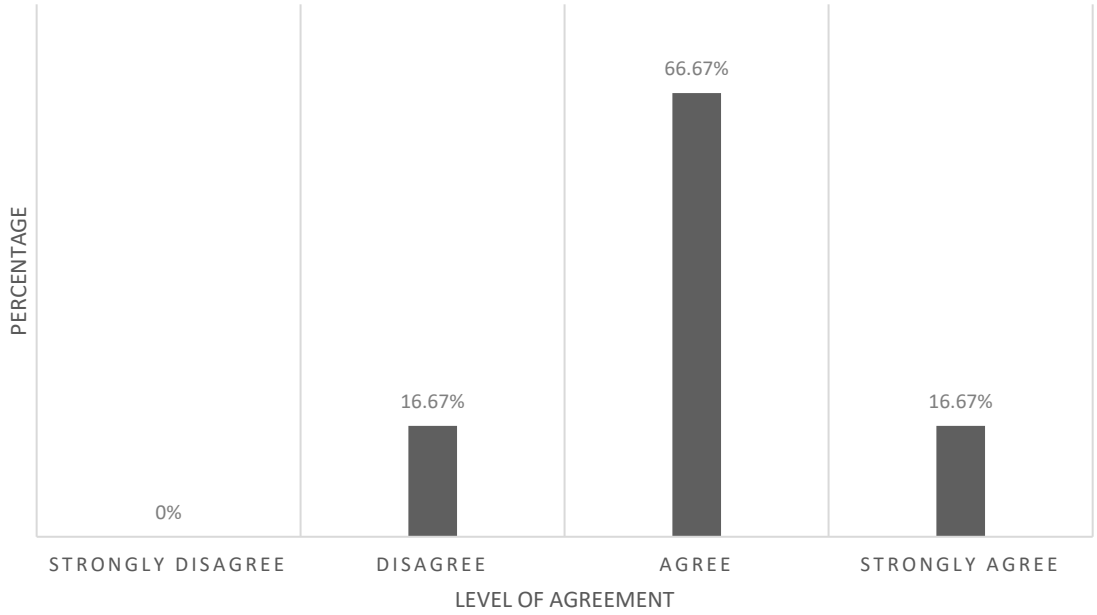


Figure 48. Survey question eight results. Assessment scores motivate administrators to ensure standards are taught.

Survey question nine. Administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their teachers.

While Figure 49 shows the majority of principals surveyed agreed state assessments motivate administrators to ensure standards are taught, Figure 46 illustrates differing data. Half of the responses demonstrated strong disagreement administrators need to be held accountable through high-stakes tests to be more effective in supervising teachers. The weighted average Likert-type scale score for this question was 1.83.

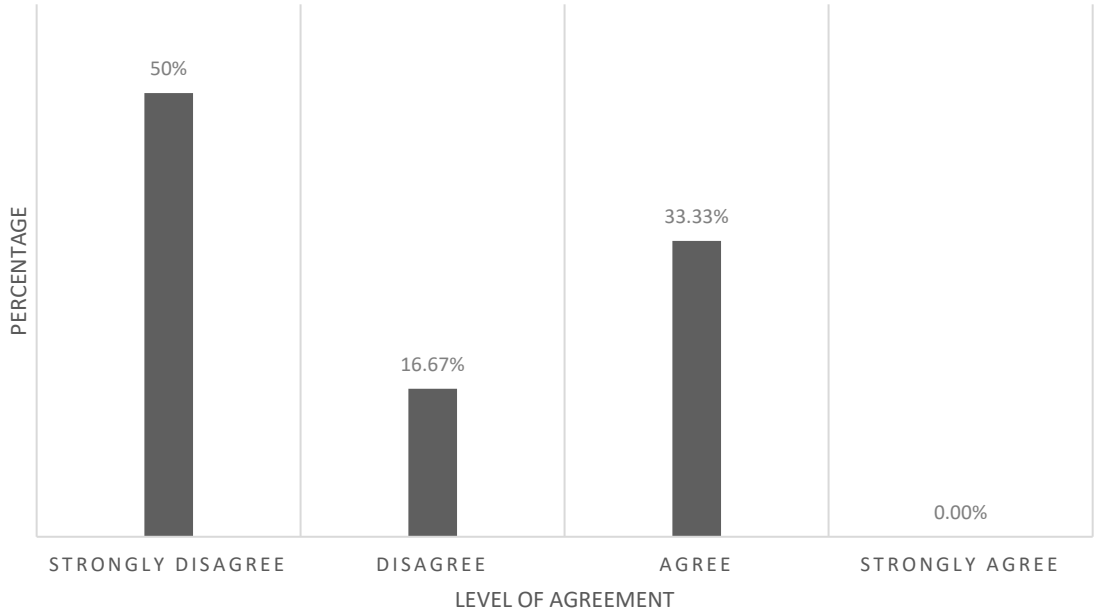


Figure 49. Survey question nine results. Administrator's accountability through high-stakes tests.

Survey question 10. There should be a direct correlation between the grade earned in a course and the corresponding standardized state assessment score.

As shown in Figure 50, all but one of the principals surveyed agreed there should be a correlation between the student's grade earned and the standardized state assessment score that corresponds with the course. The weighted average Likert-type scale score for this question was 2.83. Additionally, none of the principals surveyed expressed strong responses either for or against the need for a correlation between state assessments and course grades.

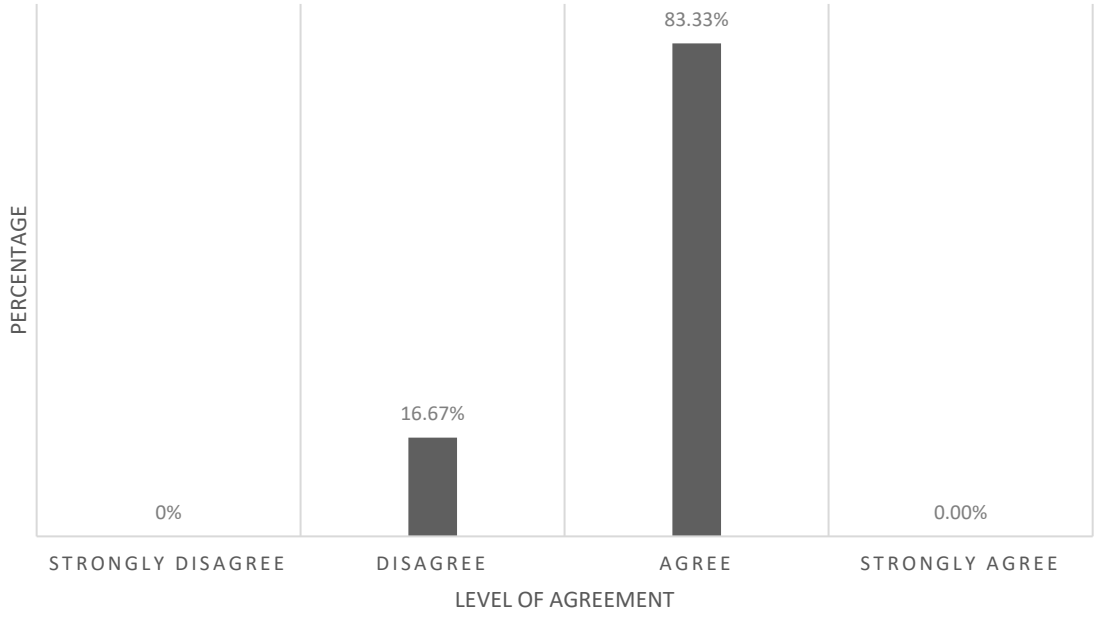


Figure 50. Survey question 10 results. Correlation between grade earned and assessment.

Summary

This chapter included findings from the analysis of data collected through principal surveys and secondary data collected from schools. Additionally, demographic data for the participating schools were represented. Survey data represented principals' perceptions on course grades to demonstrate mastery of content, standardized assessments to demonstrate mastery of content, use of standardized assessments for teacher and principal accountability, alignment of curriculum, correlation of course grades and standardized test results, and the existence of gender bias in course grades or assessments. The secondary data collected from participating schools included three

years of English II grades, English II EOC proficiency levels for each student, and students' gender.

Both the survey data and a portion of the secondary data concerning gender in relation to English II grades and English II EOC proficiency levels were analyzed using descriptive statistics. In each case, graphs were utilized to illustrate a percentage of responses or students in corresponding categories. Additionally, a Pearson product-moment correlation coefficient and the coefficient of determination was found for each of the following categories:

- A in English II and proficiency level
- B in English II and proficiency level
- C in English II and proficiency level
- D in English II and proficiency level
- advanced and grade in English II
- proficient and grade in English II
- basic and grade in English II
- below basic and grade in English II
- overall gender and proficiency level
- overall gender and grade

Finally, scatter plots were employed to demonstrate the magnitude and direction of the correlation between each of the variable pairs.

Chapter Five includes a summary of findings from the analysis of data. Furthermore, the research questions are reexamined and conclusions are discussed. Additionally, implications for practice in the area of grades and standardized assessments

are examined. Finally, recommendations for future research are suggested along with a summary of the study.

Chapter Five: Summary and Conclusions

A crucial task for educational leaders is to find the best practices for motivating students and accurately reporting progress (Popham, 2011). The Every Student Succeeds Act, a successor to the No Child Left Behind Act, allows states to decide if teacher evaluations will be tied to student performance on standardized tests (Klein, 2016). Nonetheless, districts under the ESSA are still required to test students, and states are required to identify and transform poor-performing schools (Klein, 2016). As a result, teachers and administrators feel intense pressure to ensure test scores improve (Styron & Styron, 2012). Ultimately, one effect of high-stakes testing is the narrowing of curriculum to focus more intently on what is tested (Chatterji, 2013). Godfrey (2011) showed some schools inflate grades and have lower levels of proficiency on Advanced Placement tests than other schools do (Godfrey, 2011). Additionally, there is a negative correlation between grade inflation and student performance on standardized tests (Laurie, 2007). Furthermore, an increase in the gender gap has occurred, and female students are receiving higher grades (Voyer & Voyer, 2014).

This quantitative study was designed to determine if a significant positive correlation exists between Missouri English II end-of-course exam scores and GPAs earned in the English II course. Additionally, data were examined to determine if a correlation exists among student gender, Missouri English II end-of-course exam scores, and GPAs earned in the course. Finally, perceptions from high school principals concerning grades, standardized assessments, and gender bias were analyzed. This chapter includes a review of the research questions that guided this work, a summary of

data to support the findings, a discussion of the significance and implications for practice, and recommendations for future research in the areas of grades and assessments.

Findings

Research question one. What is the correlation between Missouri English II end-of-course exam scores and corresponding end-of-year GPAs earned in English II?

Null hypothesis one. There is no correlation between Missouri English II end-of-course exam scores and corresponding end-of-year GPAs earned in English II.

Alternate hypothesis one. There is a correlation between Missouri English II end-of-course exam scores and corresponding end-of-year GPAs earned in English II.

A PPMC was performed using proficiency levels on the English II end-of-course assessment and course grade point averages for two semesters of English II. The PPMC coefficient, *r*, is useful in statistics because it indicates both the magnitude and direction of the relationship between two variables (Ary et al., 2014). According to Lund and Lund (2013), an *r* value of .1-.3 indicates a small correlation, .3-.5 shows a medium correlation, and .5-1.0 signifies a large correlation. Furthermore, for an *r* value to be considered statistically significant, the size of the sample, *n*, must be evaluated to determine the critical value required (Illowsky, 2016). Three years of data and the combined data of all three years were analyzed to determine the magnitude and direction of the correlation.

The initial comparison was a correlation between all proficiency levels on the English II end-of-course exam and English II course grades. For the 2014 school year, $n = 95$, requiring r to exceed the critical value of +/-.205. In 2014, $r = .4681$, indicating

statistical significance. Furthermore, the magnitude and direction of *r* validates a medium-strength positive correlation.

The 2015 school year yielded a value *n* = 92, requiring *r* to exceed the critical value of +/-.205. In 2015, *r* = .461, indicating statistical significance. In addition, the magnitude and direction of *r* validates a medium-strength positive correlation.

In the 2016 school year, *n* = 114, requiring *r* to exceed the critical value of +/-.195. In 2016, *r* = .4182, indicating statistical significance. Furthermore, the magnitude and direction of *r* validates a medium-strength positive correlation.

For all three years combined, *n* = 301, requiring *r* to exceed the critical value of +/-.195. Additionally, *r* = .446, indicating statistical significance for the three combined school years. Likewise, the magnitude and direction of *r* validates a medium-strength positive correlation.

Secondly, an evaluation of the correlation between advanced and proficient scores on the English II end-of-course exam and English II course grades was performed. For the 2014 school year, *n* = 74, requiring *r* to exceed the critical value of +/-.232. In 2014, *r* = .5446, indicating statistical significance. Furthermore, the magnitude and direction of *r* validates a large positive correlation.

The 2015 school year yielded a value *n* = 77, requiring *r* to exceed the critical value of +/-.232. In 2015, *r* = .4445, indicating statistical significance. In addition, the magnitude and direction of *r* validates a medium-strength positive correlation.

In the 2016 school year, *n* = 97, requiring *r* to exceed the critical value of +/-.205. In 2016, *r* = .2442, indicating statistical significance. However, the magnitude and direction of *r* demonstrates a weak positive correlation.

For all three years combined, n = 248, requiring r to exceed the critical value of +/-.195. Additionally, r = .3882, indicating statistical significance for the three combined school years. Likewise, the magnitude and direction of r validates a medium-strength positive correlation.

Subsequently, an evaluation of the correlation between basic and below basic scores on the English II end-of-course exam and English II course grades was performed. For the 2014 school year, n = 21, requiring r to exceed the critical value of +/-.433. In 2014, r = .-3182, indicating no statistical significance. Notwithstanding, the magnitude and direction of r validates a medium negative correlation.

The 2015 school year yielded a value n = 15, requiring r to exceed the critical value of +/-.514. In 2015, r = .092, indicating no statistical significance. In addition, the magnitude and direction of r validates a weak-strength positive correlation.

In the 2016 school year, n = 17, requiring r to exceed the critical value of +/-.482. In 2016, r = -.2385, indicating no statistical significance. Furthermore, the magnitude and direction of r demonstrates a weak negative correlation.

For all three years combined, n = 53, requiring r to exceed the critical value of +/-.273. Additionally, r = -.1827, indicating no statistical significance for the three combined school years. Likewise, the magnitude and direction of r validates a weak strength and a negative correlation.

Inversely, an evaluation of the correlation between students with an A or B in English II and English II end-of-course exam proficiency levels was performed. For the 2014 school year, n = 52, requiring r to exceed the critical value of +/-.273. In 2014, r =

.4111, indicating statistical significance. Furthermore, the magnitude and direction of *r* validates a medium-strength positive correlation.

The 2015 school year yielded a value *n* = 54, requiring *r* to exceed the critical value of +/-.273. In 2015, *r* = .5319, indicating statistical significance. In addition, the magnitude and direction of *r* validates a large positive correlation.

In the 2016 school year, *n* = 71, requiring *r* to exceed the critical value of +/-.250. In 2016, *r* = .267, indicating statistical significance. However, the magnitude and direction of *r* demonstrates a weak positive correlation.

For all three years combined, *n* = 177, requiring *r* to exceed the critical value of +/-.195. Additionally, *r* = .3807, indicating statistical significance for the three combined school years. Likewise, the magnitude and direction of *r* validates a medium-strength positive correlation.

Contrariwise, an evaluation of the correlation between students with a C or below in English II and English II end-of-course exam proficiency levels was performed. For the 2014 school year, *n* = 43, requiring *r* to exceed the critical value of +/-.304. In 2014, *r* = .4636, indicating statistical significance. Furthermore, the magnitude and direction of *r* validates a medium positive correlation.

The 2015 school year yielded a value *n* = 38, requiring *r* to exceed the critical value of +/-.349. In 2015, *r* = -.015, indicating no statistical significance. In addition, the magnitude and direction of *r* validates a weak-strength negative correlation.

In the 2016 school year, *n* = 43, requiring *r* to exceed the critical value of +/-.304. In 2016, *r* = .0026, indicating no statistical significance. Furthermore, the magnitude and direction of *r* demonstrates a weak positive correlation.

For all three years combined, $n = 124$, requiring r to exceed the critical value of +/-.195. Additionally, $r = .157$, indicating no statistical significance for the three combined school years. Likewise, the magnitude and direction of r validates a weak strength and a positive correlation.

Philipp (2014) concluded there is a small to moderate positive correlation between ninth-grade literature end-of-course test scores and end-of-course grades. Data collected from the schools participating in this study indicated similar results. Overall comparisons between English II end-of-course exam proficiency levels and English II course grades demonstrated statistical significance showing a positive correlation of medium strength. The data collected provided sufficient evidence to reject the null hypothesis.

Research question two. What is the correlation between Missouri English II end-of-course exam scores and gender?

Null hypothesis two. There is no correlation between Missouri English II end-of-course exam scores and gender.

Alternate hypothesis two. There is a correlation between Missouri English II end-of-course exam scores and gender.

A PPMC was performed using proficiency levels on the English II end-of-course assessment and course grade point averages for two semesters of English II for female and male students. The PPMC coefficient, r, was calculated to illustrate both the magnitude and direction of the relationship between two variables (Ary et al., 2014). According to Lund and Lund (2013), an r value of .1-.3 indicates a small correlation, .3-.5 shows a medium correlation, and .5-1.0 signifies a large correlation. Illowsky (2016)

suggested for an *r* value to be considered statistically significant, the size of the sample, *n*, must be evaluated to determine the critical value required. Three years of data and the combined data of all three years were analyzed to determine the magnitude and direction of the correlation for each gender.

An evaluation of the correlation between all proficiency levels on the English II end-of-course exam and English II course grades was performed for female students. For the 2014 school year, $n = 45$, requiring *r* to exceed the critical value of +/-.304. In 2014, $r = .5986$, indicating statistical significance. Furthermore, the magnitude and direction of *r* validates a large positive correlation.

The 2015 school year yielded a value $n = 50$, requiring *r* to exceed the critical value of +/-.304. In 2015, $r = .4488$, indicating statistical significance. In addition, the magnitude and direction of *r* validates a medium-strength positive correlation.

In the 2016 school year, $n = 46$, requiring *r* to exceed the critical value of +/-.304. In 2016, $r = .4502$, indicating statistical significance. However, the magnitude and direction of *r* demonstrates a medium positive correlation.

For all three years combined, $n = 141$, requiring *r* to exceed the critical value of +/-.195. Additionally, $r = .4977$, indicating statistical significance for the three combined school years. Likewise, the magnitude and direction of *r* validates a medium-strength positive correlation.

Subsequently, an evaluation of the correlation between all proficiency levels on the English II end-of-course exam and English II course grades was performed for male students. For the 2014 school year, $n = 50$, requiring *r* to exceed the critical value of +/-

.304. In 2014, r = .3647, indicating statistical significance. Additionally, the magnitude and direction of r validates a medium positive correlation.

The 2015 school year yielded a value n = 42, requiring r to exceed the critical value of +/-.304. In 2015, r = .4934, indicating statistical significance. In addition, the magnitude and direction of r validates a medium-strength positive correlation.

In the 2016 school year, n = 68, requiring r to exceed the critical value of +/-.250. In 2016, r = .3466, indicating statistical significance. Furthermore, the magnitude and direction of r demonstrates a medium positive correlation.

Lastly, for all three years combined, n = 160, requiring r to exceed the critical value of +/-.195. Additionally, r = .3913, indicating a statistical significance for the three combined school years. Likewise, the magnitude and direction of r validates a medium strength and a positive correlation.

In addition to the PPMC, descriptive statistics were performed using proficiency levels on the English II end-of-course assessment and the gender of the students. A two-tailed *t*-test was conducted to determine if there was a statistically significant difference between the frequency of male and female proficiency levels (Murphy et al., 2014). The observed t value was compared with the critical t value to determine statistical significance (Murphy et al., 2014).

For the 2014 school year, 80.00% of the females scored in the advanced or proficient levels of the English II end-of-course assessment. Males had 76.00% scoring in the advanced or proficient levels of the English II end-of-course assessment. Comparatively, the percentage of female students scoring in the advanced or proficient levels on the English II EOC in 2014 across the state was 78.90%. Additionally, 70.50%

of male students in 2014 scored in the advanced or proficient levels of the English II end-of-course assessment across the state.

For the 2015 school year, 84.00% of females scored in the advanced or proficient levels of the English II end-of-course assessment. Males had 83.33% scoring in the advanced or proficient levels of the English II end-of-course assessment. Comparatively, female students scoring in the advanced or proficient levels on the English II EOC in 2015 across the state was 78.00%. Additionally, 69.30% of male students in 2015 scored in the advanced or proficient levels of the English II end-of-course assessment across the state.

For the 2016 school year, 89.13% of females scored in the advanced or proficient levels of the English II end-of-course assessment. Males had 82.35% scoring in the advanced or proficient levels of the English II end-of-course assessment. Comparatively, female students scoring in the advanced or proficient levels on the English II EOC in 2016 across the state was 83.90%. Additionally, 75.50% of male students in 2016 scored in the advanced or proficient levels of the English II end-of-course assessment across the state.

Finally, for all three years combined, 84.40% of the females scored in the advanced or proficient levels of the English II end-of-course assessment. Males had 80.63% scoring in the advanced or proficient levels of the English II end-of-course assessment.

A two-tailed *t*-test was conducted to compare proficiency levels on the English II end-of-course assessment and the gender of the students for all three years combined. There was no statistically significant difference in the scores for females ($M = 3.113$, *SD*

= 0.6447) and males (M = 2.988, SD = 0.792); $t(299) = -1.592$, $p = 0.1124$. Therefore, based on descriptive statistics and the *t*-test, the results indicate there was not sufficient evidence to reject the null hypothesis.

Research question three. What is the correlation between gender and end-of-year GPAs earned in English II?

Null hypothesis three. There is no correlation between gender and end-of-year GPAs earned in English II.

Alternate hypothesis three. There is a correlation between gender and end-of-year GPAs earned in English II.

A commonality in education research proclaims females have an advantage in school and achieve higher marks (Voyer & Voyer, 2014). Additionally, Voyer and Voyer (2014) stated a significant advantage for females in language courses. Additionally, from the 1980s to the 2000s, the average high school letter grade for females shifted from "B" to "A," while the average high school letter grade for males remained at "B" (Fortin et al., 2013).

Descriptive statistics were performed using the average GPAs for English II and the gender of the students. A two-tailed *t*-test was conducted to determine if there was a statistically significant difference between the frequency of male and female proficiency levels (Murphy et al., 2014). The observed *t* value was compared with the critical *t* value to determine statistical significance (Murphy et al., 2014).

For the 2014 school year, 66.67% of the females earned a 3.0 or above average GPA for English II. Males had 44.00% earning a 3.0 or above average GPA for English

II. Comparatively, 2014 female students' mean average GPA for English II was 3.1. Additionally, male students' 2014 mean average GPA for English II was 2.54.

For the 2015 school year, 62.00% of the females earned a 3.0 or above average GPA for English II. Males had 54.76% earning a 3.0 or above average GPA for English II. Subsequently, 2015 female students' mean average GPA for English II was 2.99. Furthermore, male students' 2015 mean average GPA for English II was 2.821.

For the 2016 school year, 76.09% of the females earned a 3.0 or above average GPA for English II. Males had 52.94% earning a 3.0 or above average GPA for English II. Subsequently, 2016 female students' mean average GPA for English II was 3.12. Likewise, male students' 2016 mean average GPA for English II was 2.662.

Finally, for all three years combined, 68.09% of the females earned a 3.0 or above average GPA for English II. Males had 50.63% earning a 3.0 or above average GPA for English II. Comparatively, for all three years combined, female students' mean average GPA for English II was 3.067. Additionally, male students' three-year combined mean average GPA for English II was 2.666.

A two-tailed *t*-test was conducted to compare the average GPAs earned in the English II course and the gender of the students for all three years combined. There was a statistically significant difference in the scores for females ($M = 3.067$, $SD = 0.8665$) and males ($M = 2.666$, $SD = 0.9933$); $t(299) = -3.7087$, $p = 0.0002$. Therefore, based on descriptive statistics and the *t*-test, the results indicate there was sufficient evidence to reject the null hypothesis.

Research question four. What are the perceptions of high school principals regarding best practices for accurately reporting student progress?

Descriptive statistics were used to analyze the quantitative data from the surveys. According to Ary et al. (2014), "Descriptive statistics serve to describe and summarize observations" (p. 154). Tables were created to display the descriptive data and to view a summary of the responses as a whole.

The majority of principals surveyed agreed course grades accurately show a student's mastery of a course. Two of the six principals disagreed course grades accurately show a student's mastery of a course. Inversely, none of the principals surveyed agreed standardized assessments accurately show a student's mastery of the subject matter. Moreover, two of the six principals surveyed strongly disagreed standardized assessments accurately show a student's mastery of the subject matter.

Additionally, the final grade for a course might include a percentage for homework, projects, quizzes, unit tests, participation, and attendance (Reeves, 2011). However, a majority of the principals surveyed agreed the best way to calculate a course grade is to average the points of all assignments. Conversely, two of the six principals surveyed disagreed the best way to calculate a course grade is to average the points of all assignments. Grades given by teachers should be a reliable measurement of the students' knowledge in a particular subject; however, there are critics who state teachers give high grades for mediocre work (Goodwin, 2011).

Furthermore, 83.33% of the principals surveyed agreed course grades are free of gender bias. This perception was contrary to Voyer and Voyer's (2014) assertion suggesting a significant advantage for females in language courses. Additionally, as reported previously, there was a statistically significant difference between females and males in the average GPA earned in English II.

Likewise, 83.33% of the principals surveyed agreed standardized state assessments are free of gender bias. As reported previously, there was no statistically significant correlation between females and Missouri English II end-of-course exam scores. Cornwell et al. (2013) stated while girls tend to outperform boys on reading tests at the primary level, boys and girls tend to score the same on science and math assessments.

Research question five. What are the perceptions of high school principals regarding high-stakes testing?

The quantitative data from the surveys concerning principals' perceptions regarding high-stakes testing were analyzed utilizing descriptive statistics. Descriptive statistics are a convenient way to describe and summarize data (Ary et al., 2014). Tables were created to display the descriptive data and to view a summary of the responses as a whole.

The Every Student Succeeds Act ensures districts are held accountable when students are not making progress (U.S. Department of Education, 2016). Districts can be assessed penalties and federal funding can be withheld for inadequate educational performance (Kaufman & Blewett, 2012; Meier, 2012). As a result, teachers and administrators feel intense pressure to ensure test scores improve (Styron & Styron, 2012).

Schools use standardized tests as a tool to measure how students compare to other students (Hout et al., 2012). The majority of principals surveyed agreed or strongly agreed as a result of high-stakes testing, teachers use standardized test data to help guide and improve instruction. Only one of the six principals surveyed disagreed with teachers

using standardized test data to help guide and improve instruction as a result of high-stakes testing. Despite the majority agreement of high-stakes testing being a motivating factor for teachers to use data to improve instruction, a majority of 76.67% disagreed or strongly disagreed with teachers needing to be held accountable through high-stakes tests to motivate them to teach the content students need to know.

Additionally, 83.34% of principals surveyed agreed or strongly agreed standardized state assessment scores motivate administrators to ensure standards on which the tests are based are part of the curriculum being taught. Only one of the six principals surveyed disagreed with assessment scores motivating administrators to ensure the tested standards are part of the curriculum being taught. Despite the majority agreement of high-stakes testing being a motivating factor for administrators to ensure tested standards are part of the taught curriculum, half of the principals surveyed strongly disagreed and 16.67% disagreed with standardized state assessment scores motivating administrators to ensure standards on which the tests are based are part of the curriculum being taught.

Conclusions

The data from this study support a positive correlation between Missouri English II end-of-course exam scores and corresponding end-of-year GPAs earned in English II. The initial statistical analysis in Chapter Four included an examination of all proficiency levels each year and the combined data for all three years. Subsequently, each year examined and the combined data for all three years demonstrated the existence of a moderate positive correlation. Furthermore, both male and female subgroups shared

similar results with respect to a positive correlation between Missouri English II end-of-course exam scores and corresponding end-of-year GPAs earned in English II.

Additionally, an evaluation of the correlation between advanced and proficient scores on the English II end-of-course exam and English II course grades was performed. While showing a positive correlation of statistical significance each year studied and for the combined data from all three years, the strength of the correlation varied. In 2014 there was a strong correlation, in 2015 there was a medium correlation, and in 2016 there was a weak correlation. The correlation between advanced and proficient scores on the English II end-of-course exam and English II course grades for all three years of data combined produced a medium-strength positive correlation.

Inversely, an evaluation of the correlation between students with an A or B in English II and English II end-of-course exam proficiency levels was performed. The correlation analysis produced similar results to the advanced and proficient correlation. In 2014 there was a medium correlation, in 2015 there was a strong correlation, and in 2016 there was a weak correlation. The correlation between students with an A or B in English II and English II end-of-course exam proficiency levels for all three years of data combined produced a medium-strength positive correlation.

An evaluation of the correlation between basic and below basic scores on the English II end-of-course exam and English II course grades was performed. The correlation for each year was examined, and the combined data for all three years varied in magnitude and direction. Additionally, the statistical analysis in Chapter Four provided an absence of statistical significance.

Likewise, an evaluation of the correlation between students with a C or below in English II and English II end-of-course exam proficiency levels was performed. The correlation for each year examined and the combined data for all three years varied in magnitude and direction. The only data set from this group to produce a statistically significant result was the 2014 school year with a medium-strength positive correlation. Additionally, the statistical analysis in Chapter Four provided an absence of statistical significance and a weak correlation for the other two school years and for combined data from all three years.

Similarly, Philipp (2014) determined there is a small-to-moderate positive correlation between ninth-grade literature end-of-course test scores and end-of-course grades. Furthermore, data collected revealed there was a stronger positive correlation between students scoring advanced and proficient on the English II end-of-course exam and students earning an A or B in English II. However, insufficient evidence exists to determine if a correlation exists between students scoring basic or below basic on the English II end-of-course exam and students in the English II course earning a C or lower grade.

The data collected from participating schools revealed each of the years studied and the combined data for all three years resulted in a higher percentage of females than males scoring advanced or proficient on the English II end-of-course assessment. However, *t*-tests refuted the existence of a statistical significance. Comparatively, descriptive statistics in Chapter Four presented similar results when comparing English II end-of-course exam proficiency levels for male and female students. Data for the 2014, 2015, and 2016 school years state-wide showed 80.27% of female students scored

advanced or proficient on the Missouri English II end-of-course exam compared to 71.43% of the male students.

Voyer and Voyer (2014) stated female students often achieve higher marks in school and have an advantage in language courses. Likewise, data collected from participating schools revealed each of the years studied and the combined data for all three years resulted in a higher percentage of females than males earning a 3.0 or higher average GPA in English II. Furthermore, the average GPA for females was 3.067, while the average GPA for males was 2.666.

The data collected from participating high school principals demonstrate a majority agree with course grades being the best practice for accurately reporting student progress. Reeves (2011) suggested a final grade for a course might include a percentage for homework, projects, quizzes, unit tests, participation, and attendance. However, analysis of data in Chapter Four demonstrates a majority of the principals surveyed agree the best way to calculate a course grade is to average the points of all assignments. Contrary to the findings in Chapter Four showing a correlation between gender and grades, 83.33% of the principals surveyed agreed course grades are free of gender bias.

Conversely, none of the principals surveyed agreed standardized assessments accurately show a student's mastery of the subject matter. Furthermore, two of the six principals surveyed strongly disagreed standardized assessments accurately show a student's mastery of the subject matter. Additionally, the majority of principals surveyed disagreed with teachers or principals being held accountable for high-stakes testing. However, findings in Chapter Four demonstrate a majority of principals surveyed agreed high-stakes testing helps motivate teachers and principals to ensure tested content is a

part of the taught curriculum. As a result, teachers and administrators feel intense pressure to ensure test scores improve (Styron & Styron, 2012).

Implications for Practice

The positive correlation between Missouri English II end-of-course exam scores and GPAs earned in the course suggested several implications for practice. Administrators and classroom teachers can use this information to ensure students have acquired the intended content knowledge. Furthermore, administrators and classroom teachers can accurately convey what students have learned. Although student grades and standardized assessment scores reflect noticeably differing achievement constructs, an emphasis on content standards and standardized assessments may increase alignment between what is taught and assessed in the classroom and on standardized assessments (Dickinson & Adelson, 2015). Improved alignment to assessment standards should then increase students' performance on standardized assessments.

Finally, course grades should be less punitive and more a reflection of what the students have learned. Boys who perform the same as girls on assessments are often given lower grades by their teachers (Cornwell et al., 2013). In addition, males tend to receive more attention from their teachers (Stevens, 2015). The negative behavior may be reflected in the students' grades (Reeves, 2011). Therefore, teachers aligning course grades with what the students have learned and not with behavior should decrease the number of low grades and the discrepancy between male and female grades.

Recommendations for Future Research

The research from this study can lead to numerous future research projects. First, it would be beneficial to expand the sample size to include a larger group of students.

With a larger sampling size, the results would have a greater statistical significance. Additionally, results might be different if the demographics were varied to include larger districts or urban districts. Furthermore, the results from the study would be more applicable to a variety of districts. Likewise, using ACT test scores and course grades could make the study more relevant on a national level.

Next, this study could be replicated to include other subject areas. By incorporating other subject areas, comparisons with subject areas and gender stereotypes can be made. Males tend to require more attention from their teachers (Stevens, 2015). Discipline data could be included with a correlational study between gender and grades. The results from the study could help narrow the gap between male and female GPAs.

Despite alignment of curriculum to state standards, there are often many components to a classroom grade (Goodwin, 2011; Tomlinson, 2014). Therefore, another study could correlate standardized assessment scores with schools that use standards-based grading and with schools using traditional grading methods. Additional data could be gathered on post-secondary success from schools that use standards-based grading and from schools using traditional grading methods. With the addition of post-secondary success, the three variables could be triangulated to better demonstrate successful learner outcomes.

The final future research recommendation deals with the correlation of perceptions with course grades and performance on standardized assessments. The first would correlate principals' perceptions with the alignment of curriculum between tested standards and grades and the students' assessment levels and course grades. The next would correlate teachers' perceptions with the alignment of curriculum between tested

standards and grades and the students' assessment levels and course grades. Results could demonstrate the impact of teachers and administrators on student outcomes.

Summary

Educational leaders are tasked with finding best practices for motivating students and accurately reporting progress (Popham, 2011). Additionally, the Every Student Succeeds Act ensures districts are held accountable when students are not making progress (U.S. Department of Education, 2016). Districts can be assessed penalties and federal funding can be withheld for inadequate educational performance (Kaufman & Blewett, 2012; Meier, 2012). As a result, teachers and administrators feel intense pressure to ensure test scores improve (Styron & Styron, 2012).

The alignment of curriculum, assessment, and pedagogy can facilitate changes in education (Klenowski & Wyatt-Smith, 2014). Therefore, as districts align curriculum to assessment standards, students' scores on assessments should increase (Squires, 2012). In this study, determination of a significant positive correlation existing between Missouri English II end-of-course exam scores and GPAs earned in the English II course was examined. Moreover, data were examined to determine if a correlation exists among student gender, Missouri English II end-of-course exam scores, and GPAs earned in the course.

Appendix A

Interview Questions

The survey will use a Likert-type scale:

(1) Strongly Disagree, (2) Disagree, (3) Agree, or (4) Strongly Agree

Principal Perceptions of Course Grades, Standardized Tests, Gender Equity, and Accountability

1. A student's course grade accurately shows that student's mastery of the course.
2. The best way to calculate a course grade is to average the points of all assignments.
3. Course grades are free of gender bias.
4. As a result of high-stakes testing, teachers use standardized test data to help guide and improve instruction.
5. Teachers need to be held accountable through high-stakes tests to motivate them to teach the content students need to know.
6. A student's score on standardized assessments accurately shows the student's mastery of the subject matter.
7. Standardized state assessments are free of gender bias.
8. Standardized state assessment scores motivate administrators to ensure standards on which the tests are based are part of the curriculum being taught.
9. Administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising teachers.
10. There should be a direct correlation between the grade earned in a course and the corresponding standardized state assessment score.

Appendix B

Approval Letter

Institutional Review Board

LINDENWOOD

LINDENWOOD UNIVERSITY ST. CHARLES, MISSOURI

DATE:	September 7, 2016
TO:	Rodney Delmont
FROM:	Lindenwood University Institutional Review Board
STUDY TITLE:	[924835-1] Comparison of Classroom Grades and Missouri End-of-Course Exam Scores
IRB REFERENCE #:	
SUBMISSION TYPE:	New Project
ACTION:	DETERMINATION OF EXEMPT STATUS
DECISION DATE:	August 22, 2016
REVIEW CATEGORY:	Exemption category # *2*

Thank you for your submission of New Project materials for this research study. Lindenwood University Institutional Review Board has determined this project is EXEMPT FROM IRB REVIEW according to federal regulations.

We will put a copy of this correspondence on file in our office.

If you have any questions, please send them to IRB@lindenwood.edu. Please include your project title and reference number in all correspondence with this committee.

This letter has been electronically signed in accordance with all applicable regulations, and a copy is retained within Lindenwood University Institutional Review Board's records.

Appendix C

Dear Superintendent,

My name is Rodney Delmont, and I am a doctoral student at Lindenwood University. As part of my program requirements, I am conducting a study which will be documented in a dissertation titled, *Comparison of Classroom Grades and Missouri End of Course Exam Scores*. The purpose of this study is to explore the correlation between English II course grades and the proficiency level achieved on the English II End-of-Course Exam in rural southwest and west central Missouri.

As a participant in this study, high school principals will be asked to complete a brief online survey. The survey questions are focused on their perceptions regarding standardized assessments, course grades, accountability, and gender equity. The principals will be informed of the research, and a survey link will be attached to the email. Through the link, the principals will agree to participate in the research and will answer the survey questions. No one will be forced to participate, and the survey will be anonymous. No cost will be incurred, and the amount of time required to take the survey is approximately five minutes. Additionally, English II course grades, English II EOC proficiency levels, and students' gender will be requested. No personally identifiable information will be requested.

Your district's identity and the identity of your employees and students will be protected, and any documents will be locked in a cabinet or password-secured by the researcher. Three years after the completion of the study, all electronic data and paper documents will be deleted or destroyed using secure shred.

If you are willing to participate in the survey and supply the requested data, please sign the attached permission letter. Once signed, the letter can be scanned and emailed or faxed (fax: 417-428-XXXX) to me. Please let me know, at any time, if you have questions about the survey or research project. Approval to conduct this study will be greatly appreciated. Please do not hesitate to contact me with any questions or concerns about participation at 417-298-XXXX or rodney.delmont@weaubleau.k12.mo.us. You may also contact Dr. Shelly Fransen at 417-337-XXXX or sfransen@lindenwood.edu. A copy of this letter and your written consent should be retained by you for future reference.

Thanks you for your consideration,

Rodney Delmont, Researcher
Doctoral Student
School of Education
Lindenwood University

Appendix D

LINDENWOOD

INFORMED CONSENT FOR PARTICIPATION IN RESEARCH ACTIVITIES

"Comparison of Classroom Grades and Missouri End-of-Course Exam Scores"

Principal Investigator Rodney Delmont
Telephone: 417-298-XXXX E-mail: XXXX

Participant _____ Contact info_____

_____

1. You are invited to participate in a research study conducted by Rodney Delmont under the guidance of Dr. Shelly Fransen. The purpose of this research is to determine if there is a correlation between genders, Missouri English II End-of-Course Exam scores, and the GPAs earned in the English II.

2. a) Your participation will involve

   - Principals will complete a short survey through SurveyMonkey utilizing a Likert-type scale.

   - A designated staff member responsible for student data will be ask to export the following data: a list of students' gender, their course grades for each semester of English II, and their English II End-of-Course Exam performance levels. MOSIS numbers will be kept confidential and not reported in the findings. Data will be collected for the 2013-2014, 2014-2015, and 2015-2016 school years.

   b) The amount of time involved in your participation will be five minutes for the survey and 15-20 minutes to download and export the data from your student records program.

   Approximately seven principals and 450 students will be involved in this research.

3. There are no anticipated risks associated with this research.
4. There are no direct benefits for you participating in this study. However, your participation will contribute to the knowledge about relationships between curriculum, grading practices, and standardized test results.
5. Your participation is voluntary, and you may choose not to participate in this research study or to withdraw your consent at any time. You may choose not to answer any

questions you do not want to answer. You will NOT be penalized in any way should you choose not to participate or to withdraw.

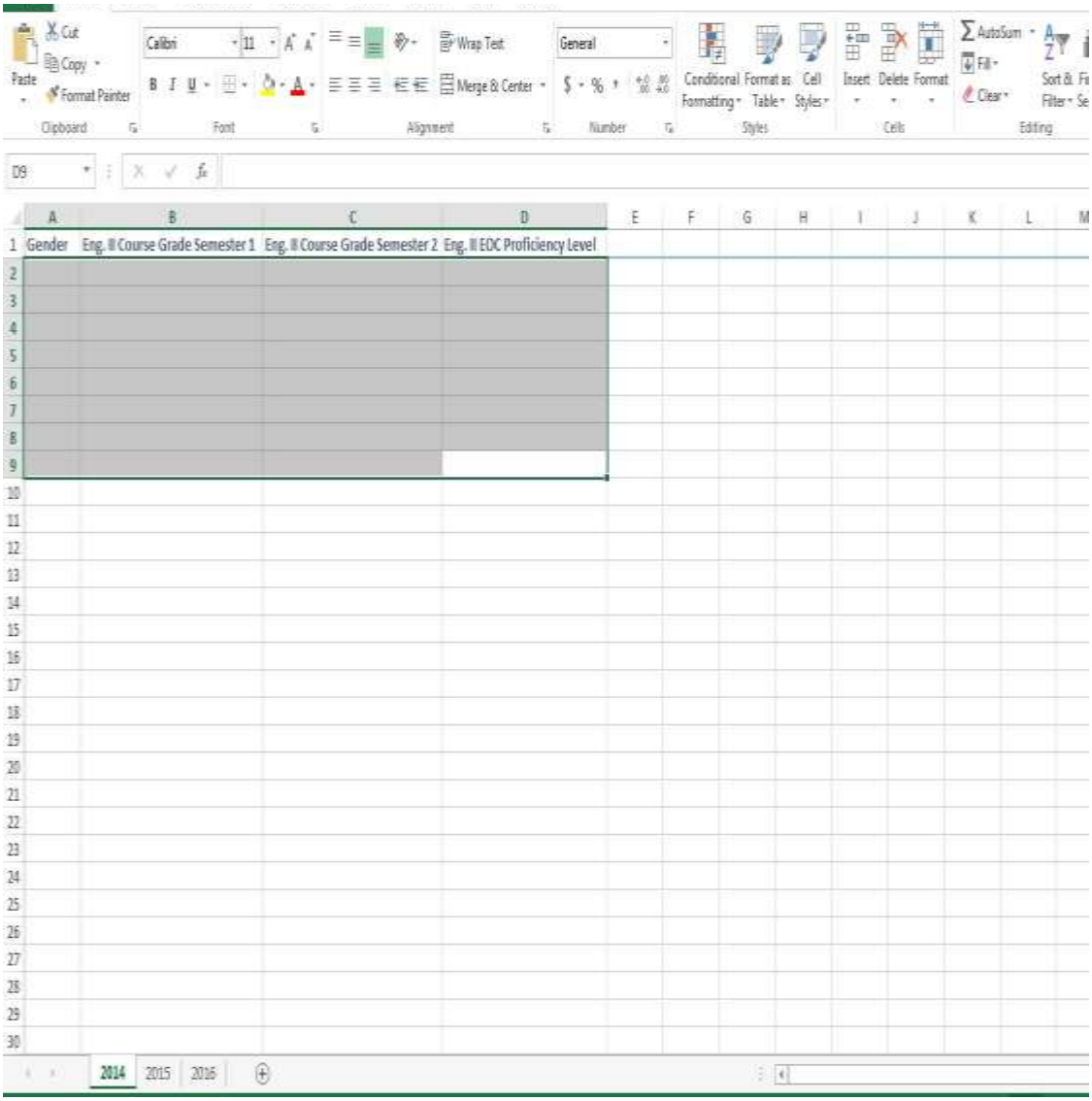
6. We will do everything we can to protect your privacy. As part of this effort, your identity will not be revealed in any publication or presentation that may result from this study, and the information collected will remain in the possession of the investigator in a safe location.

7. If you have any questions or concerns regarding this study, or if any problems arise, you may call the Investigator, Rodney Delmont, at 417-298-XXXX or the Supervising Faculty, Dr. Shelly Fransen, at 417-337-XXXX. You may also ask questions of or state concerns regarding your participation to the Lindenwood Institutional Review Board (IRB) through contacting Dr. Marilyn Abbott, Interim Provost, at mabbott@lindenwood.edu or 636-949-4912.

I have read this consent form and have been given the opportunity to ask questions. I will also be given a copy of this consent form for my records. By completing this survey, I consent to my participation in the research described above.

Appendix E

Secondary Data Spreadsheet



References

Adams, K. A., & Lawrence, E. K. (2015). *Research methods, statistics, and applications*. Thousand Oaks, CA: SAGE Publications.

Almy, S. (2011). *Fair to everyone: Building the balanced teacher evaluations that educators and students deserve.* Washington, DC: Education Trust. Retrieved from https://www.edtrust.org/resource fair-to-everyone-building-the-balanced-teacher-evaluations-that-educators-and-students-deserve

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Lancaster, PA: Author.

Ary, D., Jacobs, L. C., Sorensen, C. K., & Walker, D. A. (2014). *Introduction to research in education* (9th ed.). Belmont, CA: Wadsworth, Cengage Learning.

Astin, A. W. (2013). The promise and peril of outcomes assessment. *Chronicle Of Higher Education*, *60*(1), A53-A54.

Ballentine, S. (2016, April 20). St. Louis Post-Dispatch. *Missouri education officials replace Common Core standards | Education | stltoday.com.* Retrieved November 1, 2016, from http://www.stltoday.com/news/local/education/missouri-education-officials-replace-common-core-standards/article_050fbd0a-5dce-54f8-a502-c655ab409fe7.html

Bartlett, J. C. (2012). Attitudes for excellence. *Principal Leadership, 13*(1), 60-63.

Bluman, A. G. (2012). *Elementary statistics: A step by step approach* (8th ed.). New York: McGraw-Hill.

Boone, H. N., Jr., & Boone, D. A. (2012). Analyzing likert data. Retrieved from http://www.joe.org/joe/2012april/tt2.php

Boundless (2016). *Collaborative statistics*. Retrieved from https://www.boundless.com/users/235422/textbooks/collaborative-statistics/linear-regression-and-correlation-13/95-critical-values-of-the-sample-correlation-coefficient-table-184/95-critical-values-of-the-sample-correlation-coefficient-table-428-15976/

Brookhart, S. M. (2011). Starting the conversation about grading. *Educational Leadership (69)*3. 10-14.

Bryman, A. (2012). *Social research methods* (4th ed.). Oxford: Oxford University Press.

Bushaw, W., & Lopez, S. (2013, September). Which way do we go? *Kappan*, *95*(N1), 9-25.

Butin, D. W. (2010). *The education dissertation: A guide for practitioner scholars.* Thousand Oaks, CA: Corwin.

Carey, T., & Carifo, J. (2011). Minimum grading, maximum learning. *Principal Leadership, 11*(7), 42-46.

Chappuis, J., Stiggins, R., Chappuis, S., & Arter, J. (2012). *Classroom assessment for student learning: Doing it right – using it well.* Boston, MA: Allyn & Bacon.

Chappuis, S., Commodore, C., & Stiggins, R. J. (2016). Balanced assessment systems: Leadership, quality, and the role of classroom assessment. Thousand Oaks, CA: Corwin, a SAGE Publishing Company.

Chatterji, M. (2013). Bad tests or bad test use? Why we need stakeholder conversations on validity. *Teachers College Record, 115*(9), 1.

Clark, I. (2015). Formative assessment: Translating high-level curriculum principles into classroom practice. *Curriculum Journal, 26*(1), 91-114. doi:10.1080/09585176.2014.990911

Cohen, J., Germuska, J., & Rado, D. (2011, August 31). Public high school grads struggle at college. *Chicago Tribune*, pp. 1-3.

Cornwell, C., Mustard, D., & Van Parys, J. (2013). Noncognitive skills and the gender disparities in test scores and teacher assessments: Evidence from primary school. *Journal of Human Resources, 48*(1), 236-264.

Cox, K. (2011). Putting classroom grading on the table: A reform in progress. *American Secondary Education*, *40*(1), 67-87.

Creswell, J. W. (2013). *Qualitative inquiry and research design choosing among five approaches* (3rd ed.). Los Angeles, CA: SAGE Publications.

Creswell, J. W. (2014). *Educational research: Planning, conducting and evaluating quantitative and qualitative research.* Harlow, Essex: Pearson.

Dickinson, E. R., & Adelson, J. L. (2015). Choosing among multiple achievement measures: Applying multitrait-multimethod confirmatory factor analysis to state assessment, act, and student GPA data. *Journal of Advanced Academics, 27*(1), 4-22.

Doorey, N. A. (2012). Coming soon: A new generation of assessments. *Educational Leadership, 70*(4), 28-34.

Doubet, K. J. (2012). Formative assessment jump-starts a middle grades differentiation initiative. *Middle School Journal, 43*(3), 32-38.

Double Take. (2011). *Educational Leadership*, *69*(3), 8-9.

Eakle, A. J. (2012). *Curriculum and instruction*. Thousand Oaks, Calif: SAGE Publications, Inc.

Earl, L. M. (2004). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press.

Eastwood, M., D'Agostino, J. V., & Welsh, M. E. (2014). Conceptualizing teaching to the test under standards-based reform. *Applied Measurement in Education, 27*(2), 98-114.

Edwards, S. (2013). Preparing new teachers for contemporary middle grades schools. *Middle School Journal, 44*(4), 14-20.

Emanuel, E. J. (2013). Test our children well. *New Republic, 244*(16), 9-10.

Erickson, J. A. (2011). How grading reform changed our school. *Educational Leadership, 69*(3), 66-70.

Fortin, N., Oreopoulos, P., & Phipps, S. (2013). Leaving boys behind: Gender disparities in high academic achievement. *Journal of Human Resources, 50*(3), 549-579. doi:10.3386/w19331

Frey, N., & Fisher, D. (2011). *The formative assessment action plan : Practical steps to more successful teaching and learning*. Alexandria, Va: Assoc. for Supervision and Curriculum Development.

Geiser, S., & Santelices, M. (2007). *Validity of high-school record vs. standardized tests as indicators of four-year college outcomes.* Berkeley, CA: Center for Studies in Higher Education.

Glenn, D. (2011). Presidents are divided on best ways to measure quality. *Chronicle Of Higher Education, 57*(37), A10.

Godfrey, K. (2011). *Investigating grade inflation and non-equivalence* (Research Report 2011-2). New York, NY: The College Board. Retrieved from https://professionals.collegeboard.com/profdownload/pdf/11b_3181_RR_InvestGradeInfl_WEB_110419.pdf

Goodwin, B. (2011). Grade inflation: Killing with kindness? *Educational Leadership, 69*(3), 80-81.

Grow, A., & Van Bavel, J. (2015). Assortative mating and the reversal of gender inequality in education in Europe: An agent-based model. *PloS ONE, 10*(6), 124.

Hess, F. M., & Mcshane, M. Q. (2013, October 29). Common Core in the Real World. Phi Delta Kappan, 95(3), 61-66. doi:10.1177/003172171309500313

Hidden curriculum (2014, August 26). In S. Abbott (Ed.), The glossary of education reform. Retrieved from http://edglossary.org/hidden-curriculum

Hout, M., & Elliott, S. (2011). *Incentives and test-based accountability in education.* Washington, DC: National Academies Press. Retrieved from http://journalistsresource.org/wp-content/uploads/2011/07/Test-Incentives Education-NAS-report1.pdf

Hout, M., Elliott, S., & Frueh, S. (2012). Do high-stakes tests improve learning? *Issues in Science and Technology, 29*(1), 33-38.

Humphrey, C. (2013). A paradigmatic map of professional education research. *Social Work Education, 32*(1), 3-16.

Hutt, E., & Schneider, J. (2014). Making the grade: A history of the A-F marking scheme. *Journal of Curriculum Studies, 46*(2), 201-224.

Ikenberry, S. O., Jankowski, N., Kinzie, J., & Kuh, G. D. (2014). *Knowing what students know and can do: The current state of student learning outcomes assessment in U.S. colleges and universities*. Urbana, IL: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment (NILOA).

Illowsky, B. (2016). 95% critical values of the sample correlation coefficient table. *Collaborative Statistics*. Retrieved from https://www.boundless.com/users/ 235422 /textbooks/collaborative-statistics/linear-regression-and-correlation-13/95-critical-values-of-the-sample-correlation-coefficient-table-184/95-critical-values-of-the-sample-correlation-coefficient-table-428-15976/

Jacobsen, E. (2016). A (mostly) brief history of the SAT and ACT tests. Retrieved from http://www.erikthered.com/tutor/sat-act-history-printable.html

Johnson, D. (2013). Technology skills every teacher needs. *Educational Leadership, 70*(6), 84-85.

Kaufman, A., & Blewett, E. (2012). When good enough is no longer good enough: How the high stakes nature of the No Child Left Behind Act supplanted the Rowley definition of a free appropriate public education. *Journal of Law & Education, 41*(1), 5-23.

Kaufman, J. C. (2012). Self-estimates of general, crystallized, and fluid intelligences in an ethnically diverse population. *Learning and Individual Differences, 22*(1), 118-222.

Klein, A. (2016). New law, fresh challenges. *Education Week, 35*(15), 10-11.

Klenowski, V., & Wyatt-Smith, C. (2014). *Assessment for education: Standards, judgement and moderation.* Thousand Oaks, CA: SAGE Publications.

Laurie, R. (2007). *Setting them up to fail? Excellent school marks don't necessarily lead to excellent exam marks*. Halifax, Nova Scotia: Atlantic Institute for Market Studies.

Laurie, R. (2009). Raising the bar: A data-driven discussion on grade inflation. *Education Canada, 49*(4), 32-34.

Lund, A., & Lund, M. (2013). Pearson product-moment correlation. Retrieved from https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php

Marzano, J. R., & Heflebower, T. (2011). Grades that show what students know. *Educational Leadership, 69*(3), 34-39.

McNeil, M. (2013). Duncan's clout may be sorely tested as incentives dwindle. *Education Week, 33*(7), 1-23.

Meier, D. (2012). *Schooling of and for democracy* (Occasional paper 27). New York, NY: Bank Street College of Education. Retrieved from https://www.bankstreet.edu/occasional-paper-series/27/part-ii/schooling-democracy/

Missouri Department of Elementary and Secondary Education. (2016a). Assessment. Retrieved from http://dese.mo.gov/college-career-readiness/assessment

Missouri Department of Elementary and Secondary Education. (2016b). Missouri comprehensive data system − District info. Retrieved from https://mcds.dese.mo.gov/quickfacts/SitePages/DistrictInfo.aspx

Murphy, P. (2016). Engaging with educational research. Retrieved from http://www.open.edu/openlearn/education/educational-technology-and-practice/educational-practice/engaging-educational-research/content-section-3.3

Murphy, K. R., Myors, B., & Wolach, A. (2014). *Statistical power analysis: A simple and general model for traditional and modern hypothesis tests* (4th ed.). London, United Kingdom: Routledge.

Murphy, P. (2016). Engaging with educational research. Retrieved March 31, 2016, from http://www.open.edu/openlearn/education/educational-technology-and-practice/educational-practice/engaging-educational-research/content-section-3.3

O'Connor, K., & Wormeli, R. (2011). Reporting student learning. *Educational Leadership*, *69*(3), 41-44.

Oleinik, A. (2009). Does education corrupt? Theories of grade inflation. *Educational Research Review*, *4*(2), 156-164.

Philipp, J. (2014). *End of course grades and end of course tests in the virtual environment: A study of correlation* (Unpublished doctoral dissertation). Liberty University, Lynchburg, VA.

Picciano, A. G. (2015). Paradigms, methodologies, methods, and the role of theory in online and blended learning research. In C. D. Dziuban, A. G. Picciano, C. R. Graham, & P. D. Moskal (Eds.), *Conducting research in online and blended learning environments: New pedagogical frontiers* (pp. 12-27). New York, NY: Routledge.

Pinar, W. F. (2012). *What is curriculum theory?* (2nd ed.). New York, NY: Routledge.

Popham, W. J. (2011). *Classroom assessment: What teachers need to know*. Boston, MA: Pearson.

Porter, E. (2015, March 11). Gender gap in education cuts both ways. *New York Times*. Retrieved from http://www.nytimes.com/2015/03/11/business/gender-gap-in-education-cuts-both-ways.html?_r = 0

Punch, K. (2014). *Introduction to social research: Quantitative and qualitative approaches* (3rd ed.). London: SAGE Publications.

Reeves, B. D. (2011). Taking the grading conversation public. *Educational Leadership, 69*(3), 76-79.

Roberts, C. M. (2010). *The dissertation journey* (2nd ed.). Thousand Oaks, CA: Corwin.

Sawyer, R. (2013). Beyond correlations: Usefulness of high school GPA and test scores in making college admissions decisions. *Applied Measurement in Education*, *26*(2), 89–112. doi:10.1080/08957347.2013.765433

Schimmer, T. (2012). *Ten things that matter from assessment to grading.* Toronto, Ontario: Pearson Canada, Inc.

Schmidt, P. (2007). High-school students aim higher without learning more, federal studies find. *Chronicle of Higher Education, 53*(27), A32.

Schmitt, N. (2012). Development of rationale and measures of noncognitive college student potential. *Educational Psychologist, 47*(1), 18-29. doi:10.1080/00461520.2011.610680

Sindelar, N. W. (2011). *Assessment-powered teaching*. Thousand Oaks, Calif: Corwin.

Slavin, R. E. (2014). *Classroom management and assessment*. Thousand Oaks, Calif: Corwin Publications, Inc.

Slon, S. (2013). Teaching to the test gets an 'F'. *Saturday Evening Post, 285*(5), 47-49.

Soares, J. A. (2012). The future of college admissions: Discussion. *Educational Psychologist, 47*(1), 66-70. doi:10.1080/00461520.2011.638902

Soh, K. C. (2011). Grade point average: What's wrong and what's the alternative? *Journal of Higher Education Policy and Management, 33*(1), 27-36.

Spencer, K. (2012). Standards-Based Grading. *Education Digest, 78*(3), 4-10.

Squires, D. (2012). Curriculum alignment research suggests that alignment can improve student achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 85*(4), 129-135. doi:10.1080/00098655.2012.657723

Stevens, K. (2015). *Gender bias in teacher interactions with students* (Master's thesis). Retrieved on from http://digitalcollections.dordt.edu/med_theses/90

Strauss, V. (2014, February 21). Study: High school grades best predictor of college success – Not SAT/ACT scores. *The Washington Post.* Retrieved from http://www.washingtonpost.com/blogs/answer-sheet/wp/2014/02/21/a-telling-study-about-act-sat-scores/

Stumm, S. (2014). Intelligence, gender, and assessment method affect the accuracy of self-estimated intelligence. *British Journal of Psychology, 105*(2), 243-253.

Styron, J. L., & Styron, R. A., Jr. (2012). Teaching to the test: A controversial issue in quantitative measurement. *Journal of Systemics, Cybernetics, & Informatics, 10*(5), 22-25.

Taylor, P. C., & Medina, M. N. D. (2013). Educational research paradigms: From positivism to multiparadigmatic. *Meaning Centered Education, 1*(1). Retrieved

from http//www.meaningcentered.org/educational-research-paradigms-from-positivism-to-multiparadigmatic/

Tomlinson, C. A. (2014). *The differentiated classroom: Responding to the needs of all learners* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Tomlinson, C. A., Moon, T. R. (2013). *Assessment and student success in a differentiated classroom*. Alexandria, Va: ASCD.

United States Department of Education. (2016). ESEA flexibility. Retrieved from http://www2.ed.gov/about/inits/ed/implementation-support-unit/tech-assist/evaluations-teacher-effectiveness.pdf

University of Texas at Austin, Faculty Innovation Center. (2015). Methods of assessment. Retrieved from https://learningsciences.utexas.edu/teaching/assess-learning/methods-overview

Vatterott, C. (2015). *Rethinking grading: Meaningful assessment for standards-based learning*. Alexandria, Virginia: ASCD.

Voyer, D., & Voyer, S. D. (2014). Gender differences in scholastic achievement: A meta-analysis. *Psychological Bulletin, 140*(4), 1174-1204.

Wellington, J. (2015). *Educational research: Contemporary issues and practical approaches* (2nd ed.). New York, NY: Bloomsbury Academic.

Wells, J. (2015, December 1). How do we close the education gender gap? *The Telegraph*. Retrieved from http://www.telegraph.co.uk/men/thinking-man/how-do-we-close-the-education-gender-gap/

Wormeli, R. (2011). Redos and retakes done right. *Educational Leadership*, *69*(3), 22-26.

Yesbeck, D. M. (2011). *Grading practices: Teachers' considerations of academic and non-academic factors* (Doctoral dissertation). Retrieved from ERIC database. (ED 913076079)

Vita

Rodney O. Delmont completed his undergraduate studies in Biology at Southwest Baptist University in 1997. He earned his Master's Degree in Educational Administration from William Woods University and his Specialist's Degree in Educational Administration from Missouri State University.

Rodney began his teaching career at Halfway R-3 School District where he taught seventh- through 12th-grade science. In 2006, he accepted the position as secondary principal at Weaubleau R-3 School District. He is currently in his 11th year as principal at Weaubleau R-3 School District.