# PERFORMANCE GRADES AS MEASURES OF ACADEMIC ACHIEVEMENT

A Dissertation
by
JED COCKRELL

Submitted to the Graduate School
at Appalachian State University
in partial fulfillment of the requirements for the degree of
DOCTOR OF EDUCATION

May 2016
Educational Leadership Doctoral Program
Reich College of Education

PERFORMANCE GRADES AS MEASURES OF ACADEMIC ACHIEVEMENT

A Dissertation
by
JED COCKRELL
May 2016

APPROVED BY:

George Olson, Ph.D.
Chairperson, Dissertation Committee

Sara Zimmerman, Ph.D.
Member, Dissertation Committee

Roma Angel, Ed.D.
Member, Dissertation Committee

Audrey Dentith, Ph.D.
Director, Educational Leadership Doctoral Program

Max C. Poole, Ph.D.
Dean, Cratis D. Williams School of Graduate Studies

**Abstract**

PERFORMANCE GRADES AS MEASURES OF ACADEMIC ACHIEVEMENT

Jed Cockrell
B.A., University of North Carolina at Charlotte
M.A., Appalachian State University
Ed.D., Appalachian State University

Dissertation Committee Chairperson: Dr. George Olson


Prior research exposes some long-held concerns about the grades teachers assign and what those grades mean (e.g., Starch, 1913; Steele, 1911). Despite an increased effort to improve assessment at the classroom level (e.g., Popham, 2009; Stiggins, 2001), many of the same concerns about the meaning of grades mentioned in earlier research continue to persist. In an effort to connect grades to more objective measures of academic achievement, previous research has examined relationships between students' grades and standardized assessment scores (e.g., Brennan, Kim, Wenz-Gross & Siperstein, 2001; Ross & Kostuch, 2011). However, the relationship between grades and what teachers expect students to score on standardized assessments has not been examined. This study links students' grades, or *performance* grades, to both a teacher-expected EOG/EOC (end-of-grade and end-of-course) achievement level, and an actual EOG/EOC achievement level.

Three years of data linking students' performance grades, standardized assessment scores, and teacher-expected standardized assessment scores for students in grades 3-12 were examined. Correlations between pairs of achievement measures (e.g., performance grades and expected EOG achievement levels) were calculated. While correlations between students' performance grades and standardized assessment scores were similar to those found

in prior studies with respect to students' ethnicity and gender, relationships between those two measures of student achievement and the marks reporting teacher-expected standardized assessment scores indicated that teachers underestimated differences between the performance grades they assigned to students and those students' actual standardized assessment scores. Overestimating or underestimating students' levels of learning has important implications since it affects both students' and parents' understanding of the effectiveness of the learning process (e.g., Ross & Kostuch, 2011; Schneider, Teske, & Marschall, 2000). Just as importantly, misunderstanding or misrepresenting students' levels of learning also directly affects teachers' ability to match appropriate levels of instruction to students' needs in order to maximize learning outcomes (Good, Williams, Peck, & Schmidt, 1969; Herfordt-Stöpel & Hörstermann, 2012).

## Acknowledgements

I would like to thank my chair, Dr. George Olson, for all of his knowledgeable support along the way.  I would also like to thank my committee members, Dr. Sara Zimmerman and Dr. Roma Angel, for their insight and help in putting this study together.

## Dedication

Thank you to my wife and daughter for all of their love and support in this process.

**Table of Contents**

# Introduction

## Grading and Marking Issues

In 1983, the National Commission on Education Excellence published *A Nation at Risk* (NAR), which asserted that K-12 public education in the United States was on a downward trajectory (Gardner, 1983). Among the report's findings regarding expectations, it was noted that students would be responsible for such things as hard work, self-discipline, and motivation and that these expectations would be measured through grades and rigorous examinations. Despite responses questioning the findings and general tone of the NAR report (e.g., Kohn, 2015; Stedman, 1994), the report propelled a movement to judge educational effectiveness by student outcomes, spurred on by follow up legislation such as the No Child Left Behind Act of 2001 (Guthrie & Springer, 2004; Spellings, 2008). Of the educational reforms pushed by NAR, standards based education and standardized assessment programs have grown in strength over the last 30 years.

Even though concerns about levels of student achievement persist, parents continue to express satisfaction with their child's school based on information they receive about their child's progress through grades (Schneider, Teske, & Marschall, 2000; Tuck, 1995; US Department of Education [USDOE], 1992). This reliance upon grades is troubling due to the lack of objective meaning inherent in teachers' grades. For example, a study conducted by the US Department of Education (1994) found that students in high-poverty schools earning grades of an A or B were equivalent academically to students making C's or D's in more affluent schools. The comparison of grading distributions at high-poverty schools and more affluent schools serves as an example of how the assignment of grades is greatly affected by a comparison of a student's against their classmates.

**Inconsistency in the meaning of grades**. Grades, despite their long history of serving as a measure of classroom assessment in American schools, have been shown to be inconsistent measures of student performance. Research citing differences between teachers and teachers' values indicates the varying meanings imbedded in teacher grading practices are not a new phenomenon (Starch, 1913). Other research from the same period appears to validate Starch's assertion by referring to grades as "worthless" and "misleading" (Steele, 1911). Despite decades of research on teacher grading practices, researchers are still asking questions about the merits of grading practices (Allen, 2005; Mansfield, 2001; Waltman & Frisbie, 1994) or whether grades should be used at all (Kohn, 2002, 2015). Given that a common criticism of pre-service teacher training in "educational measurement" courses is that these courses tend to focus more on the technical components of the theory associated with assessment rather than addressing practical application (Stiggins, 2001; Stiggins & Chappuis, 2005; Volante & Fazio, 2007), improvements in teacher training programs to improve teachers' assessment literacy would be expected; however, the same questions about the utility of grades and their inherent subjectivity persist.

**Assessment vs marking and grading**. There is a close relationship between assessment, on the one hand, and marking and grading, on the other hand. Educational assessment is the term typically given to the broad area of "measuring student accomplishment," and applies to any number of techniques used for that purpose, including both formal and informal tests, classroom observation, subjective appraisals of comportment, and so on. Similar, though slightly different, synonyms for assessment include measurement and evaluation. Whatever term is used—and they are often used interchangeably in practice—assessment is used, according to at least one classroom assessment expert (McMillan, 2014), as a basis for diagnosing students' strengths, weaknesses, and other instructional needs, as a basis for teachers' decision-making with respect to both individual students and the classroom as a whole, and, lastly, as a means of communicating students' level of performance or achievement.

**Purpose of Grades.** Over three decades ago, John Hills, in one of the first books on classroom assessment wrote about the purpose of grades: "The primary function of grading and marking is to communicate effectively to a variety of audiences the degree of achievement of academic competence of individual students" (Hills, 1981, p. 283). Later, in 2000, Marzano stated unequivocally that, "The most important purpose for grades is to provide information or feedback to students and parents," (Marzano, 2000). Since Hills, there have been numerous books like Marzano's written on classroom assessment, all of which contain a section or a chapter on marking and grading. Virtually all of those works, to at least some degree, support Hills' and Marzano's statements. It is this use of assessment—specifically, marking and grading—that is the central theme of my study.

Prior research examining the relationship between grades and standardized assessment scores (Ross & Kostuch, 2011) yielded findings reporting that teacher-assigned grades can fulfill multiple roles in that the grades can provide feedback about a student's academics, while also serving to reaffirm a student's self-identity and self-esteem. What had not been examined, prior to this study, was how well teachers understood the degree to which the grades they assign function as marks that advocate for students and simultaneously judge their performance. This study was designed to build from prior studies that directly compared grades and standardized assessment scores, but also to add in the "Expected Achievement Level" (ExpLvl) variable, measuring what teachers expect students to score on EOGs and EOCs, thereby allowing for a comparison of relationships between grades and actual EOG/EOC achievement levels to relationships between grades and expected EOG/EOC levels and relationships between actual EOG/EOC achievement levels to expected EOG/EOC achievement levels. The inclusion of the Expected Achievement Level variable in comparing how well grades align with EOG/EOC achievement levels provides a look into how teachers think the grades they assign will fare as reports of academic achievement to more objective reports of EOG/EOC achievement levels. Given the tendency of teachers to assign grades to students relative to their classmates' performance, student placement in schools and in classes affects learning opportunities and outcomes for all levels of students. The following literature review examines how teachers use student performance in relation to the performance of their peers when assigning grades to students.

**Problem Statement and Research Questions**

An examination of performance grade distributions provides an explanation for the allocation of resources within schools since performance grades often serve as the basis for identifying students needing additional resources, such as time or personnel, to address academic gaps. It is common knowledge that the performance grades teachers assign often do not agree with the more objective measures of performance obtained from standardized tests (e.g., Bowers, 2009; Brennan, Kim, Wenz-Gross, & Siperstein, 2001). My objective is to examine and document those discrepancies to determine to what extent they are a function of factors unrelated to achievement. Questions guiding this research include:

1. What discrepancies exist between performance grades and standardized assessment scores at different levels of schooling (elementary, middle, and high school)?
2. How does subgroup status (gender and race) affect the degree to which performance grades assigned for a given course or grade level differ from standardized measures of achievement?

The questions guiding this research serve as an extension of the dialogue already taking place in research on the topic examining what the role of grades should be (Church, Elliot, & Gable, 2001; Guskey, 2001, 2011), or even if there should be a continuation of the practice of assigning performance grades to students at all (Kohn, 2002).

**Definition of Key Terms**

The literature concerning grading practices and student achievement is relatively accessible in its discourse and terminology; however, a few terms warrant further clarification due to their tendency to vary in meaning depending on the context in which they are used. Other terms, such as Cizek, Fitzgerald, and Rachor's *success bias* (1996), are used to represent common themes found in the literature. For instance, success bias refers to the tendency of teachers to advocate for their students by overestimating students' achievement levels through the assignment of a grade higher than one more representative of their actual academic ability. Some other common terms found in the literature include:

1. Grading Practices: the practices used by teachers in constructing performance grades for students. These practices include, for example, decisions to include homework or class participation as factors in determining performance grades and the extent to

which a teacher takes into account the presumed effect a particular grading criterion will have on a final grade; for instance, does a teacher include class participation as a factor in determining their students' performance grades? If so, to what degree does it count?

2. Standardized Assessment: summative assessments (e.g., end-of-grade or end-of-course tests) given at the end of a course or grade level to assess the amount of knowledge any one student has learned about the subject matter covered in the class or grade.
3. Nonacademic Factors: including, but not limited to, the contribution of factors other than achievement that contribute to performance grades such as teachers' estimations of effort, growth, ability, and student behavior.
4. Performance Grade: any score or mark stemming from a teacher's judgment based on a student's ability to successfully complete work for a given subject area or grade level, e.g., a report card grade.

**Significance of the Study**

Cizek et al. (1996) referred to classroom assessment as the weak link in the move to improve the American public educational system; this conclusion is supported by Stiggins and Chappuis (2005) who claimed that most educators do not understand how to effectively use assessment to improve learning. Research on classroom assessment, and its implications for grading practices, has shown that various nonacademic factors often influence measures of student academic achievement (Brookhart, 1993; Cizek, Fitzgerald, & Rachor, 1996; Cross & Frary, 1996; Willingham, Pollack, & Lewis, 2002). A consistent finding in the research is that factors such as a student's subgroup designation (e.g., socioeconomic status (SES), race, or gender), or even a student's level of school or teacher assignment often influence the performance grades teachers give students. Since these nonacademic factors create unequal access to academic success, the limitation of educational advancement or recognition based upon something other than academic abilities should be a concern to educators. Bowers (2009), for instance, claimed that grades are just as much a function of students' ability to "negotiate the social processes of school" as they are measures of academic achievement (p. 609). The significance of my study is that it will lead to a better understanding of how teachers assign performance grades to students by connecting the performance grades teachers assign to objective measures of student achievement and teacher expectations of student performance on objective measures of achievement. Identifying where teachers' grading practices lose connection with academic content is important to educators who want to be able to use the results gleaned from students' grades to improve learning outcomes for all students.

Given the degree that teachers' grading practices vary, it is relatively safe to assume that the correlation among teachers' grades and their students' standardized assessment scores vary as well. While research exists that compares performance grades to corresponding assessment scores (Brennan, Kim, Wenz-Gross & Siperstein, 2001; McCandless, Roberts, & Starnes, 1972; Olson, 1989; Pedulla, Airsian, & Madaus, 1980; Ross & Kostuch, 2011), the relationship between students' performance grades in comparison to how teachers *expect* students to score on summative assessments has not been examined. It is, therefore, of interest to examine any contrasts of correlations between performance grades and objective measures of student achievement (i.e., end-of-grade and end-of-course tests) against the corresponding correlations between performance grades and the scores teachers expect their students to score on objective measures of academic performance. The comparison of correlations between the two sets of variables (performance grades and actual EOG/EOC achievement level variable and an expected EOG/EOC achievement level and performance grade variable) should determine two things: 1) the degree to which teachers expect the performance grades they assign to vary from standardized test scores when students are sorted by subgroups for gender and ethnicity, and 2) how differences between teachers' expectations of their students' performance on standardized tests compares to their actual performance when sorting students by subgroups for gender and ethnicity.

## Review of Literature

### Creating Meaningful Grades through Teacher Assessment Training

Beziat and Coleman (2015) noted a lack of sound classroom assessment knowledge (including how to mark and grade) by classroom teachers and pre-service teachers despite an increased emphasis being placed on growing knowledge in this realm over the past 30 years. Popham (2009) argued that until pre-service teachers consistently receive training in the field of assessment and measurement, it is necessary that professional development address the need through in service training. Stiggins (2001) wrote that a great deal of the blame for a lack of tangible progress in the development of the field of effective classroom assessment, with the most important component being teachers adopting and implementing effective and valid grading practices, lies with the measurement community itself. Stiggins attributed this lack of progress to the inability of those seeking to effectively bridge accepted theory to the workings of the classroom so that these methods "can be applied efficiently by teachers to the benefit of their students" (p. 7); a claim which Frey and Schmitt (2010) echoed in reporting, "the measurement community must do a better job of training teachers," if teachers are to be able to use assessment in ways that improve student learning (p. 114).

Assessment and measurement training, which informs competent grading practices, is imperative to improving student learning, as Guskey (1994) argued that teachers are not able to bring forth substantive advances in student learning if they are not able to apply appropriate authentic, performance-based assessment to the classroom. However, counter to claims that pre-service teacher training in assessment would help to bring forth more assessment literate educators, Brookhart (1994) expressed doubt that an increase in assessment training would be enough to reconcile grading practice to the recommendations of the measurement community. DeLuca and Bellara (2013) echoed Brookhart's concerns in reporting that, despite an effort to push assessment competency for educators (especially at the pre-service level), beginning teachers continued to lack basic assessment competency skills. The lack of basic competency in assessment skills supports Brookhart's (2015) assertion that validity issues still exist for graded achievement, specifically citing variation in the meaning of grades by teachers.

## Common Bases for Grading

**Subjectivity.** Research on classroom assessment reveals that a large degree of subjectivity in assessing student learning comes from the constructed grading practices of each individual teacher. The variance observed among and within teachers' grading practices (Bowers, 2009; Brookhart, 1993; Cizek et al., 1996; Marzano & Heflebower, 2011; McMillan & Nash, 2000), is supported by Wise, Lukin, and Roos (1991) who found that over half of the teachers surveyed in their research reported that their most substantive training in assessment and measurement had come from trial and error. Cizek et al. (1996) argued that the primary factor influencing teachers' grading schemes is teachers' own trial and error methods. By limiting themselves to their own trials and errors, teachers have little chance of developing grading and assessment philosophies that are not uniquely designed around their own subjective beliefs and experiences. However, the subjectivity in assessment is not only limited to how assessment is constructed; it also plays a role in how results of assessment are reported.

**Contextualization.** Guskey (2001) cited the use of comparative descriptors of student performance such as “above average” and “average” as examples of how traditional student performance appraisal employs a compare and contrast mentality since those terms “reflect norm-referenced examples rather than criterion referenced standards” (p. 25). Students’ performance grades often affect their ability to enroll in classes or even graduate (Bowers, 2009), so it is important to understand how the contextually based inferences influencing grading decisions are made.

**Previous experience as a student**. Other research found that teachers often continue the grading practices they experienced as students. Guskey (2004) reported that “teachers do what was done to them”, (p. 31). Cizek, et al., (1996) in examining teachers’ classroom assessment practices and how those practices are constructed, found that a wide range of factors contribute to the creation of each teacher’s grading scheme within his or her class. The factors cited by Cizek et al. cover teacher grading discretions such as the type of assignments used in each classroom, the frequency with which teachers make those assignments, and the degree to which each assignment factors into a student’s final grade. These factors, along with other factors such as years of experience, the location (urban or rural) in which a teacher works, and the teacher’s grade-level assignment, are relevant to understanding how teachers assign grades (Brookhart, 1993, 1994; Cross & Frary, 1996; Marzano, 2011; McMillan, 1999; Randall & Engelhard, 2009; Resh, 2009).

**Enduring issues with grading and marking.** Concerns about grades and how they are used to communicate students’ performance is an issue that has been examined for many years (e.g., Randall & Engelhard, 2009; Steele, 1911; Starch, 1913). One early examination of teacher-assigned grades and standardized assessment scores comes from a study of Dallas area secondary schools. Olson (1989) found that the grades assigned by teachers and the teacher-created final exams produced low validity coefficients, implying that many characteristics, besides those directly accounting for academic achievement, factored into these scores; for example, incorporating marks for effort and behavior or allowing for extra credit opportunities to students whose grades are not adequate. Olson attributed the low validity of teacher-assigned grades, as well as the low validity of teacher final exams to a lack of adequate teacher preparation in measurement principles. This conclusion is supported through later research confirming a lack of preparedness among teachers and administrators alike in their professional training (Impara, Plake, & Fager, 1993; Popham, 2009; Schafer, 1993).

**Standardized tests and grading**. One method through which educators understand and communicate student academic progress is through the quantification of student achievement results from standardized testing. However, despite an effort to justify the use of standardized testing to assess student learning and teacher effectiveness, there persists a continuing incongruence in how we prepare teachers to properly understand and implement effective grading practices. Prior research (Popham, 2009; Schafer, 1993; Waltman & Frisbie, 1994) noted the effect that a lack of adequate assessment preparation has on teachers; for instance, the tendency of classroom teachers to interpret test scores incorrectly, which, in turn, causes teachers and those with whom they are communicating achievement results to draw erroneous conclusions about a student's academic progress. Schafer reported that a common misconception among teachers is the misreporting of testing results, such as confusing percentiles and percentages. Using Schafer's example, when a student has a percentile rank in the high 60s on a standardized assessment the student is performing at a higher level than approximately two-thirds of his or her classmates; however, if the same score is reported as a percentage the student is understood to be performing poorly. The perils of mistakes in communicating student academic progress are very real since grades, which are often recorded as percentages, serve as the means for the distribution of rewards and access to higher levels of education. The persistence of misinformed and misinterpreted practices, such as these, stems from the lack of assessment training, especially with respect to grading, for both administrators and teachers (Allen, 2005; Trevisan, 1999).

Teachers employ a wide variety of grading practices. Individual grading practices vary so much that, despite common usage of traditional means of communicating grades (typically an "A" through "F" scale), there are still many instances of miscommunication about what these marks really mean when it comes to reporting what students know (e.g., Brookhart, 2003; Cross & Frary, 1996). Cross and Frary (1996) describe the inherent variance in teachers' grading practices as "hodgepodge grading" – a term derived from a Brookhart (1991) reference to teachers' assessment process contributing to a "hodgepodge grade of attitude, effort, and achievement" (p.36). In an attempt to address the hodgepodge contributing to the confusing nature of performance grades, Guskey (2001) separated teacher grading criteria into three categories: *i*, product, which refers specifically to student academic performance; *ii*, process, which includes components enabling students to learn the material being presented (such as student effort and classroom behavior); and *iii*, progress, which entails teachers being able to make judgments about each student's learning potential and how well students achieve desired educational outcomes in relation to those expectations. Guskey cited common themes, such as student motivation and social consequences stemming from the assignment of performance grades, to explain why few teachers apply purely product-referenced grading standards in their classrooms. Most importantly, Guskey noted that the commonly employed practice of combining some form of product, process, and progress ultimately creates a performance grade that is “confounded and impossible to interpret” (p.19). The lack of interpretability of performance grades can be summed up by Cizek, in saying that even as “grades continue to be relied upon to communicate important information about performance and progress… they probably don’t” (1996, p. 104).

**Teachers’ Contribution to the Confounding of Performance Grades**

In an attempt to understand the inclusion of nonacademic factors affecting teachers’ grading practices, Brookhart (1993) identified a potential conflict faced by each teacher whose primary duty is to serve as an advocate for the student. Although teachers are responsible for assessing a student’s work, teachers face a difficult choice of balancing the interpretability of the grade assigned against the consequences attached to the assigned grades faced by each student. Brookhart’s contention that teachers take into account how their assessment practices affect students beyond the simple assignment of a performance grade is interesting because it acknowledges the role of nonacademic factors as an essential part of grade construction. McMillan and Nash (2000) identified several influences as core components of teacher grading and assessment practice; among these components is the need that teachers have to “pull for students” in ways that assist students to achieve success that teachers feel they normally would not be able to achieve through the use of more standard grading techniques. This finding is supported through research demonstrating that teachers have difficulty separating judgments about students’ academic ability from other factors (Brackett, Floman, Ashton-James, Cherkasskiy, & Salovey, 2013; Pedulla, Airasian, & Madaus, 1980), due in no small part to teachers’ inability to balance their roles as both "coach and judge" (Bishop, 1992, p. 2). A primary method through which subjectivity becomes evident in teacher evaluation of student progress is through teacher overestimation of student ability; Cizek et al. (1996, p.170) refer to this phenomenon as a "success bias" that teachers have in assessing the achievement of their own students. The tendency of teachers to advocate for their students through assigning inflated performance grades confuses the role teachers are required to play when it comes to assessing achievement objectively (Cross & Frary, 1996).

**Parental misunderstandings of grades.** The issue of misinterpreting student performance persists when teachers and parents discuss grades. Waltman and Frisbie (1994), using a questionnaire, compared the meanings parents interpreted from math grades assigned to their fourth grade child with the meanings inferred by the teachers. A common misunderstanding of parents was their belief that most students in the teachers' classes were assigned grades in the "C" range while teachers reported their average assigned grade to be a "B". This misunderstanding poses a problem for a parent whose child receives a grade of "C", who then believes his or her child is performing at an average level while, in actuality, their child is receiving one of the lower grades in the class. Cross and Frary (1996) cited the tendency of teachers to assign grades higher than academically warranted due to the professional pressure to report certain levels of student achievement. Cross and Frary found that this pressure was understood by teachers as either an indicator of one's own professional abilities or a way of avoiding excessive numbers of failing grades that might suggest some sort of bias against any one student group.

**Hodgepodge grading**. Cross and Frary (1996) reported that the subjectivity embedded within teachers' grading practices exists in large part due to the professional and social consequences attached to performance grades. Although intertwining performance grades and nonacademic factors only contributes to confusion about student academic performance (Nitko, 2004), Bonner and Chen (2009) found that social factors play a large part in the assignment of grades with some teachers becoming more flexible with grades as a response to parental involvement. When parental pressure influences the assignment of performance grades there is a danger that the grades will be misinterpreted and there is likely to be confusion about a student's academic ability or achievement (Brookhart, 1993). The inclusion of nonacademic factors not only affects the academic validity of the grades given by each teacher, but the practice also keeps teachers from being able to match appropriate levels of student ability and task difficulty in order to maximize learning outcomes (Good, Williams, Peck, & Schmidt, 1969; Herfordt-Stöpel & Hörstermann, 2012).

Parents are not the only stakeholders who believe that performance grades should be negotiable. Cross and Frary (1996) found students to be proponents of including nonacademic factors, such as teacher estimates of ability, class participation, growth, and effort, into performance grades. The fact that students consider the inclusion of nonacademic factors in assessing their academic performances to be a fair practice tends to be in agreement with Brookhart's assertion that classroom grading practices function as a type of "academic token economy" through which grades are exchanged for behavior and other nonacademic issues (1993, p. 139). While this practice is at odds with recommended grading practices (O'Connor, 2007; Stiggins, Frisbie, & Griswold, 1989), it appears that the use of nonstandard grading practices are not only prevalent but are also expected.

The relationships teachers and their students build act as a powerful variable in influencing how teachers define and identify successful students (Bishop, 1992; Brookhart, 1993, 2003; Cizek, Fitzgerald, & Rachor, 1996; McMillan & Nash, 2000). One explanation of the role social norms play in defining student success is Bowers' finding that the subjective construction of grading schemes and classroom assessment practices is affected by the degree to which students are able to "negotiate the social processes of school", (2009, p. 609). Bowers described this phenomenon as a way in which the child being assessed is rewarded for a myriad of reasons including his or her capabilities in the "behavioral, attention, social, and academic" realms, (2009, p. 623). Brookhart (2003) suggested that there is a psychosocial context in classroom assessment that affects how expectations are set, at least in part, through the teacher's perceptions of students and the assessment environment. Pairing Brookhart's claim with Bowers' finding concerning the effect of social influences lends support to the idea that performance grades are influenced through students' relationships with their teacher and other students within the classroom.

**Influence of Level of Schooling**

Resh (2009) used a sample of high school language, math, and science teachers to determine how teachers allocate grades for such factors as effort, behavior, and academic success. Resh noted two important reasons for identifying the respondents by subject area: first, the separation of subject areas in high school creates pockets of contextualized knowledge and pedagogical practice based on socialization and professional development patterns; secondly, the "closed" nature of the sciences requires a more prescribed method for learning compared to the more "open" nature of the humanities, where learning can take on a more flexible manner allowing for more "pedagogical variations" to play out (p. 318). Resh's claims about differences in how teachers in different subject areas assess student performance agree with previous research noting that teachers' assigned subject area affects their method of assigning grades (Deutsch, 1985; McMillan & Nash, 2000), thus affecting the degree to which items such as effort or tests count towards an overall grade.

The high school setting and middle school setting, where students switch classes and teachers for different subject areas, is a direct contrast to the elementary setting where teachers are responsible for teaching every core subject to every student. Randall and Engelhard (2009), examined differences between the grading practices of individual teachers at the elementary and middle school levels and found that elementary teachers assign higher performance grades than their middle school counterparts, which is consistent with Brookhart (1994) who noted the tendency for elementary teachers to assign more lenient performance grades since elementary teachers are more likely to include nonachievement related factors in grading. Randall and Engelhard found that one of the issues causing a discrepancy between the grading practices of elementary and middle school teachers is that elementary teachers spend more time with their students and therefore feel compelled “to nurture and protect the self-esteem of their students” (p. 184). Randall and Engelhard’s conclusion, that the subjective nature of performance grades leads students to be confused about the meaning of grades, paralleled findings from Nitko (2004) and Brookhart (1993) who reported the use of nonacademic factors in performance grades caused confusion when reporting a student’s level of academic performance.

**Student-Level Variables Affecting Achievement Measures**

Brennan, Kim, Wenz-Gross, and Siperstein (2001) examined the relationship between standardized test scores and teacher-assigned grades, using a two-level hierarchical linear model (HLM) with one level establishing the “measurement model” being employed and the second level representing the race/ethnicity and gender of each student. This study yielded two important findings: first, although boys tended to outperform girls on standardized assessment scores, girls typically outperformed boys in terms of performance grades; and secondly, Brennan et al. (2001) noted a larger achievement gap between Black and White students and Hispanic and White students using results from the Massachusetts Comprehensive Assessment System (MCAS) when compared to the use of performance grades. These findings served as the foundation for Brennan's et al. (2001) comment that performance grades "usually produce more equitable achievement results than standardized tests" (p. 209), which is a socially desirable result since, as Cross and Frary (1996) noted, teachers do not want their grades to suggest a possible bias against a student or student group. Brennan et al. concluded that it is likely that performance grades, which include a mixture of academic and nonacademic factors, may allow students to compensate for academic struggles by meeting other teacher-imposed criteria, e.g., rewarding students for their ability to successfully "negotiate the social processes of school" (Bowers, 2009).

Martinez, Stecher, and Borko (2009) confirmed the concept of teachers using grades as a method of establishing performance equity through finding that teachers achievement ratings where higher for minority students than should have been expected from their test scores. Martinez et al. supported this finding with the explanation "that teachers compensate for perceived disadvantages faced by these groups by adjusting ratings up – or, alternatively, adjusting their criteria and expectations down" (p. 97). Hochweber, Hosenfeld, and Klieme (2014) cited Martinez et al. (2009) and Brookhart (1993) in noting that teachers do tend to care about the social consequences of the grades they assign, and therefore tend to use varying criteria for assigning grades to different groups of students.

Cornwell, Mustard, and Van Pary (2011) addressed gender differences in grade and test score relationships for students in kindergarten using reading, math, and science scores from the Early Childhood Longitudinal Study (ECLS). Their study yielded findings showing that differences between teachers' assessment of student performance compared to students' performance on the ECLS assessment favored females in every subject area. Even in math and science, where male test scores were higher than female test scores, females received higher grades from teachers. Cornwell et al. found that the female-male gap in reading grades was over 300 % larger than the gap between white and black students in reading, and the female-male gap in math and science grades was about 40 % larger than the corresponding gap in white and black students for those same subject areas.

**Goal Orientations**

Church, Elliot, and Gable (2001) noted two distinct goal orientations at play when considering the meaning of grades: a standards-based approach, which considers the students' level of performance relative to the standards being taught, and a normative approach, which emphasizes a student's performance relative to that of other students. Guskey (2011) explained the difference between the two approaches in terms of whether it is a teacher's job to "select talent or develop it" (p.16). If teachers believe it is their job to "select talent," Guskey explained that teachers work to maximize differences between student achievement. The results of maximizing these differences would result in a grade distribution resembling a normal distribution "of randomly occurring events when nothing intervenes," (p. 17). Assessments designed for selection purposes, such as the American College Testing (ACT) exam and the Scholastic Aptitude Test (SAT), are, as Popham (2007) described, "instructionally insensitive," thus allowing students to be more easily sorted. The distribution of achievement looks different in a standards-based approach where all students are expected to reach identified academic goals (Hershberg, 2005), since the job of the teacher is to identify what students are and are not able to do and then design instruction to address students' academic deficiencies. If teachers believe it is their job to develop talent, teachers must clarify the standards they want their students to accomplish, then grade that performance against those standards. Whether or not a student is able to master the standards taught would be a testament to how effectively the teacher provided instructional intervention enabling the student to reach the desired goal.

Chappuis, Stiggins, Chappuis, and Arter (2012) addressed the issue of teachers being able to clearly and effectively identify what students need to know and building from that as the difference of designing assessment *for* learning and designing assessments *of* learning. Chappuis et al.'s "assessment for learning" designation is a formative measurement taken by a teacher indicating where the student is in the learning process, thereby allowing the teacher to design instruction appropriate to that level of learning, while the "assessment of learning" designation is a summative measure of student learning used to make broader decisions such as a student's quarterly grade or to determine whether teachers or schools are doing a good job. Chappuis et al.  noted that the traditional method of aggregating assessments of learning (e.g., grades) has been to include factors such as participation and effort. The inclusion of these affective factors into students' grades dilutes the ability of the grade to report what it was designed to measure (assessment of students' learning), when the two factors can be used to tell a teacher a lot more about how a student is learning (assessment for student learning).

## Grading Confounds Relating to Self-Efficacy

Ross and Kostuch (2011) acknowledged that teachers consider the role of self-efficacy and its relationship to achievement, both positively and negatively, when assigning grades to their students. In support of Ross and Kostuch's premise that teachers use compensatory grading practices for minority students, Martinez, Stecher, and Borko (2009) claimed that compensatory grading mitigates the effects of racial, SES, and gender differences in grading distributions – a finding which supported the claims of Brennan et al. (2001) that found that grades produced more equitable results amongst groups of students than do standardized assessment results. Ross and Kostuch summed up their findings by suggesting that the discrepancies between performance grades and standardized assessment scores were small enough that report card grades can be both positively reaffirming for students, through what the authors call a "modest inflation of self-efficacy arising from report card generosity" (p. 175), while also contributing some useful information regarding a student's mastery in a given subject area. Even so, Ross and Kostuch commented that given the variability between the performance grades and standardized assessment scores, both of which purport to measure student academic achievement, there exists a large enough discrepancy between the two measures to warrant questioning the validity of one or even both of the measures (p. 175).

The issue of interpretability in grades is a theme often cited in research (Brookhart, 1993; Cross & Frary, 1996; Guskey, 2011; USDOE, 1994), which leads to confusion on the parts of parents, students, and even educators (Schafer, 1993; Waltman & Frisbie, 1994). With little to no inherent meaning beyond the class or task to which they are assigned, performance grades serve as arbitrary measures of student performance consisting of a hodgepodge of influences (Dornbusch, Ritter, Leiderman, Roberts, & Fraleigh, 1987). The lack of any standardization in grading practices is problematic, considering grades serve as the basis from which students are selected for academic honors, enabled to enroll in certain classes, or even accepted into post-secondary education. While it is simple enough to look at students' transcripts and determine that one student's A is better than another student's C, the story that is not told is how the teachers of the given courses arrive at the grades they assign.

## Methodology

### Methodological Approach and Research Questions

This study examined the relationships between student achievement measures and is, therefore, correlational in nature. Correlations between achievement measures were examined to address two research questions:

1. What discrepancies exist between performance grades and standardized assessment scores at different levels of schooling (elementary, middle, and high school)?
2. How does subgroup status (gender and race) affect the degree to which performance grades assigned for a given course or grade level differ from standardized measures of achievement?

### Data Sources and Data Collection

This study used 80,247 student records from reading, math, and science courses spanning three years covering grades 3 through 12 from a school district in western North Carolina. The following information was collected for each student: the performance grade the teacher anticipated assigning to the student (AntGrd), the expected achievement level for each student on the North Carolina (EOG) or End-of-Course (EOC) assessment (ExpLvl), and the actual achievement level each student scored on his or her EOG/EOC assessment (ActLvl). AntGrds assigned by each teacher were used in place of students' actual grades because the latter was not available from the district. All information used for the study was provided by the district's accountability department.

Anticipated performance grades should function as an acceptable substitute for actual performance grades for two reasons: 1) the AntGrd is assigned by the same teacher assigning the actual performance grade, and 2) the AntGrd is recorded immediately following administration of the EOG/EOC, which is at the end of the grade level or course from which the performance grade is assigned. At the conclusion of EOG/EOC test administration, teachers code students' AntGrds and ExpLvl onto student EOG/EOC answer sheets. EOG/EOC test administration manuals instruct teachers to code AntGrds to reflect the "best estimation of what the student will earn" and not what the student has the ability to earn (NCDPI, 2009, p.87). While the EOG/EOC test administrator's manual states that teachers may elect to use students' AntGrds as a factor in determining the ExpLvl, the manual acknowledges that "grades are often influenced by factors other than pure achievement" and that the teacher is to "provide information that reflects only the achievement of each student in the subject matter tested" in order to determine a student's ExpLvl (NCDPI, 2011, p. 85).

**Data Coding**

Data regarding AntGrds were coded F = 0, D = 1, C = 2, B = 3, and A = 4. Data pertaining to ExpLvl and ActLvl were numerically coded 1, 2, 3, and 4. The numerical codes assigned to ExpLvl and ActLvl use the scale provided by North Carolina Department of Public Instruction (NCDPI) to indicate whether student mastery of knowledge and skills in the tested subject area is deemed to be insufficient (level 1), inconsistent (level 2), consistent (level 3), or superior (level 4) (NCDPI, 2009).

**Data Analysis**

The first part of this study examined correlations between AntGrds, ExpLvl, and ActLvl across elementary, middle, and high schools. An examination of the correlations between the three student achievement variables determined which levels of schooling assign performance grades that more closely correlate with standardized assessment scores. Examining a range of grades spanning elementary, middle, and high school allowed for comparisons of performance grades and standardized assessment scores to be made in three subject areas that span all three levels of schooling: math, reading, and science. The study was correlational in nature, so rather than independent and dependent variables, my study used correlated variables (i.e., test scores and performance grades). For the first part of the study, Kendall's tau b (Agresti, 2010; Kendall, 1938) was used to determine the statistical significance between variables when examining the relationship of student achievement variables at different grade levels, e.g., AntGrds and ActLvls. Kendall's tau was chosen over the more widely used Spearman's rank correlation because the Kendall's tau statistic provides a direct interpretation of the probabilities of observing concordant and discordant pairs, (Conover, 1980).

The second part of this study examined how a student's subgroup status (gender and race) affected correlations between AntGrds, ExpLvls, and ActLvl. While some data were available for students of all ethnicities, there were too few students in the Asian, American Indian, Multiracial, and Pacific Islander subgroups to provide meaningful data for this study. AntGrd, ExpLvl, and ActLvl data for White, Black, and Hispanic subgroups were used since the number of male and female students within these subgroups was sufficiently large enough (i.e., $n > 400$) to examine for this study.

## Findings

To examine the relationship between students' performance grades and both their expected EOG/EOC achievement levels and actual EOG/EOC achievement levels, students were cross-classified by gender and ethnicity. Contingency tables were generated for male and female comparisons and for ethnicity comparisons. The contingency tables examined three new variables, "discrepancies," which were constructed by comparing pairs of achievement measures: ExpLvl and ActLvl, AntGrd and ActLvl, and ExpLvl and AntGrd. The discrepancy for each individual was computed by subtracting the second member of each pair from the first. Then, when a discrepancy was less than zero it was labeled "-1," indicating that the second member of a pair was larger than the first and vice versa (if the second member was less than the first, it was labeled "+1"). When there was no difference, the discrepancy was labeled "0." The number (frequency) of the different types of discrepancies was then used as the body of the contingency tables. Hence, in Table 1, the first entry, 343, gives the frequency of times that the actual reading EOG scores for 3rd Grade males was *lower* than their expected scores.

If teachers were neutral, or unbiased, in their assessment of students' performances, we would expect, in the best of situations, the discrepancies for male and female students, and for White, Black, and Hispanic students to be independent of gender or race. In other words, even though teachers might evidence discrepancies between performance grades and standardized test scores, if no bias exists, then those discrepancies should be more or less uniform across gender and ethnicity.

Table 1

*Observed Discrepancies between Achievement Measures for Students Classified by Gender (Reading)*

| | ExpLvl, ActLvl | | | AntGrd, ActLvl | | | ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | TOTAL | Male | Female | TOTAL |
| | | | | Grade 3 | | | | | |
| -1 | 343 | 270 | 613 | 478 | 343 | 821 | 350 | 390 | 740 |
| 0 | 1197 | 1131 | 2328 | 1053 | 971 | 2024 | 1514 | 1475 | 2989 |
| 1 | 780 | 767 | 1547 | 794 | 859 | 1653 | 477 | 311 | 788 |
| TOTAL | 2320 | 2168 | 4488 | 2325 | 2173 | 4498 | 2341 | 2176 | 4517 |
| | | | | Grade 4 | | | | | |
| -1 | 321 | 291 | 612 | 503 | 373 | 876 | 323 | 407 | 730 |
| 0 | 1170 | 1138 | 2308 | 985 | 939 | 1924 | 1506 | 1488 | 2994 |
| 1 | 813 | 795 | 1608 | 819 | 913 | 1732 | 498 | 342 | 840 |
| TOTAL | 2304 | 2224 | 4528 | 2307 | 2225 | 4532 | 2327 | 2237 | 4564 |
| | | | | Grade 5 | | | | | |
| -1 | 211 | 204 | 415 | 387 | 277 | 664 | 288 | 377 | 665 |
| 0 | 1156 | 1025 | 2181 | 1042 | 891 | 1933 | 1519 | 1555 | 3074 |
| 1 | 919 | 1010 | 1929 | 858 | 1072 | 1930 | 512 | 330 | 842 |
| TOTAL | 2286 | 2239 | 4525 | 2287 | 2240 | 4527 | 2319 | 2262 | 4581 |
| | | | | Grade 6 | | | | | |
| -1 | 538 | 100 | 638 | 551 | 261 | 812 | 509 | 656 | 1165 |
| 0 | 1467 | 384 | 1851 | 904 | 871 | 1775 | 1245 | 1230 | 2475 |
| 1 | 840 | 334 | 1174 | 773 | 1010 | 1783 | 506 | 272 | 778 |
| TOTAL | 2845 | 818 | 3663 | 2228 | 2142 | 4370 | 2260 | 2158 | 4418 |
| | | | | Grade 7 | | | | | |
| -1 | 355 | 287 | 642 | 630 | 368 | 998 | 385 | 578 | 963 |
| 0 | 1035 | 1025 | 2060 | 834 | 844 | 1678 | 1168 | 1157 | 2325 |
| 1 | 844 | 867 | 1711 | 766 | 966 | 1732 | 700 | 452 | 1152 |
| TOTAL | 2234 | 2179 | 4413 | 2230 | 2178 | 4408 | 2253 | 2187 | 4440 |
| | | | | Grade 8 | | | | | |
| -1 | 289 | 209 | 498 | 451 | 256 | 707 | 497 | 630 | 1127 |
| 0 | 1018 | 988 | 2006 | 784 | 736 | 1520 | 1216 | 1250 | 2466 |
| 1 | 936 | 1030 | 1966 | 1007 | 1236 | 2243 | 554 | 357 | 911 |
| TOTAL | 2243 | 2227 | 4470 | 2242 | 2228 | 4470 | 2267 | 2237 | 4504 |
| | | | | ENG1 | | | | | |
| -1 | 684 | 636 | 1320 | 1163 | 849 | 2012 | 301 | 430 | 731 |
| 0 | 1178 | 1180 | 2358 | 865 | 959 | 1824 | 1194 | 1333 | 2527 |
| 1 | 625 | 587 | 1212 | 497 | 625 | 1122 | 1003 | 651 | 1654 |
| TOTAL | 2487 | 2403 | 4890 | 2525 | 2433 | 4958 | 2498 | 2414 | 4912 |

Table 2
*Observed and Expected Discrepancies between Achievement Measures for Students Classified by Ethnicity*

| | Observed Frequencies Grade 3 Math AntGrd and ActLvl | | | | Expected Frequencies Grade 3 Math AntGrd and ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 755 | 240 | 193 | 1188 | 782.502 | 220.010 | 185.488 |
| 0 | 1377 | 341 | 277 | 1995 | 1314.050 | 369.461 | 311.489 |
| +1 | 724 | 222 | 207 | 1153 | 759.448 | 213.528 | 180.023 |
| TOTAL | 2856 | 803 | 677 | 4336 | | | |

$\chi^2(4)=18.15$; $p$ = .0012
V = .0457

Chi square tests were used to test this assumption of independence. Because there were so many subgroups, a large number of chi square tests were computed. Hence, the probability of obtaining one or more statistically significant results at the conventional alpha level of .05 would have been greatly inflated. For instance, for all 17 AntGrd and ActLvl comparisons across all grade levels and subject areas, 17 chi square tests were computed. For this many tests, if each used an alpha level of .05, the actual probability of a Type I error is given as 1 – (1-0.05)17 (Uitenbroek, 1997), or .58. A typical solution to this problem (and the one I used) is the Bonferroni correction (Napierala, 2012), which adjusts the nominal *p* value by dividing it by the number of tests. This yielded a new alpha level of .0029. Thus, in this study, any chi square test statistic having a *p* value less than .003 was considered statistically significant. In Table 2, the *p* value was .0012, which is less than .003; hence, significant (i.e., the null hypothesis of independence was rejected).

To assess the strength of the association between gender and performance and between ethnicity and performance, I computed a Cramer's *V* (1946) correlation coefficient. Cramer's V was used because it allows for a measurement of association between variables in all rectangular-shaped tables (i.e., when there are not equal numbers of rows and columns), whereas another comparable test of association, the phi statistic, only applies to square-shaped tables. Additionally, percentage deviations (Lowry, 1998) were calculated to show, as a percent, how often the observed discrepancies between achievement measures were greater (or smaller) than expected. Table 3 is an example of how observed discrepancies and expected discrepancies were used to determine percentage deviations. For example, in the lower right of Table 3, White male third-graders in math had negative discrepancies (where

Table 3
*Method for Calculating Percentage Deviations from Observed and Expected Discrepancies*

| | Observed Discrepancies | | | | Expected Discrepancies | | |
|---|---|---|---|---|---|---|---|
| | Grade 3 Math AntGrd and ActLvl | | | | Grade 3 Math AntGrd and ActLvl | | |
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 755 | 240 | 193 | 1188 | 782.502 | 220.010 | 185.488 |
| 0 | 1377 | 341 | 277 | 1995 | 1314.050 | 369.461 | 311.489 |
| +1 | 724 | 222 | 207 | 1153 | 759.448 | 213.528 | 180.023 |
| TOTAL | 2856 | 803 | 677 | 4336 | | | |

| | Observed & Expected Discrepancies | | | Percentage Deviations | | |
|---|---|---|---|---|---|---|
| | Grade 3 Math AntGrd and ActLvl | | | Grade 3 Math AntGrd and ActLvl | | |
| | White | Black | Hisp | White | Black | Hisp |
| -1 | -27.502 | 19.990 | 7.512 | -3.515% | 9.086% | 4.050% |
| 0 | 62.950 | -28.461 | -34.489 | 4.791% | -7.704% | -11.072% |
| +1 | -35.448 | 8.472 | 26.977 | -4.668% | 3.967% | 14.985% |

AntGrd was lower than ActLvl) meaning that White male third-graders had lower instances than expected (3.5% lower) of AntGrds lower than their ActLvls, while positive discrepancies (where AntGrd exceeded ActLvl) were also lower than expected (4.7% lower).

Of all of the Cramer's *V* values calculated for relationships found to be significant, all but Grade 8 science were below .2, indicating a weak relationship between the variables; the Cramer's *V* value, for Grade 8 science for the AntGrd and ActLvl relationship was .2136, indicating a moderate relationship between the variables. According to Dattalo (2009), the more unequal the marginal distributions between the variables, the more likely it is that *V* will be less than 1. Furthermore, one of the assumptions of *V* is that the relationship between the variables is *monotonic*, meaning that both of the variables being measured increase concurrently. The lack of a monotonic relationship as well as unequal marginal distributions may have acted to suppress *V* values in this study.

## Results

**Influence of Level of Schooling on Relationships among Achievement Measures.** The first part of this study examined the relationships between a student's level of schooling, i.e., elementary, middle, and high school, and the corresponding achievement measures. Correlations between AntGrd and ActLvl and between ExpLvl and AntGrd are displayed in Tables 4, 5, and 6 for elementary, middle, and high school students, respectively. Algebra 1 (ALG 1), English 1 (ENG1), and Biology (BIO) are coded as classes instead of grade levels because students of varying grade level classifications were able to enroll in the courses being offered at this level of schooling.

Since both ExpLvl and AntGrd are marks that represent teacher estimates of achievement, it was not surprising that the correlation between ExpLvl and AntGrd was stronger than the correlation between AntGrd and ActLvl at every grade level for all three

Table 4

*Math Tau-b Correlation Coefficients for AntGrd* and *ActLvl and ExpLvl and AntGrd*

| Grade | AntGrd and ActLvl | ExpLvl and AntGrd |
|---|---|---|
| 3 | 0.570 | 0.763 |
| 4 | 0.494 | 0.725 |
| 5 | 0.552 | 0.751 |
| 6 | 0.476 | 0.671 |
| 7 | 0.485 | 0.683 |
| 8 | 0.489 | 0.701 |
| ALG1 | 0.366 | 0.768 |

$p < .0005$

Table 5

*Reading Tau-b Correlation Coefficients for AntGrd and ActLvl and ExpLvl and AntGrd*

| Grade | AntGrd and ActLvl | ExpLvl and AntGrd |
|---|---|---|
| 3 | 0.600 | 0.760 |
| 4 | 0.527 | 0.735 |
| 5 | 0.557 | 0.753 |
| 6 | 0.499 | 0.643 |
| 7 | 0.456 | 0.614 |
| 8 | 0.470 | 0.624 |
| ENG1 | 0.414 | 0.594 |

$p < .0005$

Table 6

*Science Tau-b Correlation Coefficients for AntGrd and ActLvl and ExpLvl and AntGrd*

| Grade | AntGrd and ActLvl | ExpLvl and AntGrd |
|---|---|---|
| 5 | 0.509 | 0.704 |
| 8 | 0.428 | 0.667 |
| BIO | 0.103 | 0.858 |

$p < .0005$

areas. The ExpLvls and AntGrds correlations suggest that teachers anticipated assigning performance grades that reflected their students' academic performance. However, when ExpLvl and AntGrd correlations are compared to the AntGrd and ActLvl correlations, there is an apparent disconnect between the degree to which teachers *believed* AntGrds reflected their students' academic performance and how AntGrds measured performance as measured by EOG/EOC results. Given prior research noting the tendency of teachers to assign performance grades using a "hodgepodge" of various factors (Cross & Frary, 1996), and the research of Brennan et al. (2001) noting that performance grades "usually produce more equitable achievement results than standardized tests" (p. 209), it was not altogether unexpected to see AntGrds with a stronger correlation to ExpLvl than with ActLvl.

Correlations between AntGrd and ActLvl attenuated, somewhat, from elementary school (grades 3, 4, and 5) to middle school (grades 6, 7, and 8), in both reading and math (in math, the correlation between AntGrd and ActLvl for ALG1—a high school course—was considerably lower). Also, the AntGrd and ActLvl correlations in science were lower in Grade 8 than Grade5 and were even lower in Biology (a high school course). Furthermore, while the correlations between ExpLvl and AntGrd were higher than those between AntGrd and ActLvl, there was still a tendency for the correlations to attenuate with increasing grade level. Prior research regarding the performance grades assigned to students and how the assignment of performance grades changes as students progress through levels of schooling indicated that elementary teachers were more prone to assigning performance grades that were less likely to reflect academic achievement (Brookhart, 1994; Randall & Engelhard, 2009). In contrast, the results from this study indicate that teachers assign AntGrds that are less reflective of academic achievement in middle school than in elementary. These findings would seem to indicate that as students matriculated through higher levels of schooling, teachers either found ways to mitigate performance grades as students progressed through levels of school, or they increasingly graded students on content different from the content being assessed by EOGs or EOCs.

One interesting note is that, at the high school level, math (ALG1) and science (BIO) correlations between ExpLvl and AntGrd were higher than for any other grade levels; while at the same time, the correlations between AntGrd and ActLvl for these two courses were the two weakest in the study. Discrepancies between ExpLvl and AntGrd correlations and AntGrd and ActLvl correlations for Algebra 1 and Biology indicate a need to further examine the grading policies at the teacher and subject area levels, since teachers apparently believed they were assigning grades that reflected what their students learned.

While there were some fluctuations among individual grade levels, the general pattern in the correlations between AntGrd and ActLvl across elementary, middle, and high school levels in math, reading, and science indicated that teachers became less reliable, as compared to standardized tests, in assigning performance grades. When coupled with attenuating correlations between ExpLvl and AntGrd as grade levels increased (from elementary to high school in reading, and through middle school in all subject areas), it appears obvious that, at higher grade levels, teachers graded students independently of standardized levels of performance.

**Influence of Gender on Relationships among Achievement Measures.** Prior research regarding the influence of a student's gender on the relationship between performance grades and standardized assessment scores noted that performance grades often produced more equitable results than standardized assessment (Brennan et al., 2001) and that compensatory grading mitigated differences in student subgroup performance (Ross &

Kostuch, 2011). The results from this study not only agreed with those findings, but, by examining the relationship of anticipated performance grades to expected EOG/EOC achievement levels, it also produced evidence that teachers were aware of the influence students' gender had on the differences between the anticipated performance grades and EOG/EOC achievement levels.

Table 7
*Percentage Deviations of Discrepancies between Math Achievement Measures for Students Classified by Gender*

| | | All Males | | | All Females | | |
|---|---|---|---|---|---|---|---|
| | | ExpLvl and AntGrd | ExpLvl and ActLvl | AntGrd and ActLvl | ExpLvl and AntGrd | ExpLvl and ActLvl | AntGrd and ActLvl |
| Grade 6 | -1 | -19.5 | 17 | 27.2 | 20.4 | -17.7 | -28.3 |
| | 0 | -5.4 | 3.3 | 0.6 | 5.7 | -3.4 | -0.7 |
| | +1 | 22.7 | -11.6 | -19.9 | -23.8 | 12.1 | 20.8 |
| | | All Males | | | All Females | | |
| Grade 7 | -1 | -13.7 | 10.9 | 24.2 | 14.1 | -11.2 | -25.1 |
| | 0 | -5 | 0.5 | -2.8 | 5.1 | -0.5 | 2.9 |
| | +1 | 21.5 | -5.1 | -11.8 | -22.2 | 5.3 | 12.3 |
| | | All Males | | | All Females | | |
| | -1 | -11.8 | 13 | 17.1 | 12.1 | -13.2 | -17.4 |
| | 0 | -2.5 | -0.9 | 1.9 | 2.6 | 0.9 | -1.9 |
| | +1 | 15.7 | -5 | -13.4 | -16.1 | 5.1 | 13.6 |

In the tables throughout this section, color-coding was used to highlight discrepancies in percentage deviations between the following achievement measures: AntGrd, ExpLvl, and ActLvl. The pink values indicate instances where ExpLvl and AntGrd discrepancies were at least five percent higher than expected; yellow values indicate instances where ExpLvl and ActLvl discrepancies were at least five percent higher than expected; and, green values indicate instances where AntGrds and ActLvl discrepancies were at least five percent higher than expected. Using Table 7 as an example, the pink percentage deviation entry of 22.7 for Grade 6 math males indicates that the discrepancies in the "+1" category for the ExpLvl and AntGrd column were 22.7% greater than expected. The green percentage deviation entry of 13.6 for Grade 8 math females indicates that the discrepancies in the "+1" category for the AntGrd and ActLvl column were 13.6% greater than expected.

Tables 8, 9, and 10 yielded percentage deviations indicating consistent relationships between AntGrd and ActLvl and those found between ExpLvl and AntGrd for male and female students. AntGrds for male students were lower than ExpLvls and were also lower than their (male students') ActLvls, while AntGrds for female students exceeded ExpLvls and also exceeded their (female students') ActLvls. Every grade level within each subject area with reportable data yielded positive "+1" ExpLvl and AntGrd percentage deviationsfor males, while yielding positive "-1" ExpLvl and AntGrd percentage deviations for females; these patterns are highlighted red and green in their respective categories.  Even in grade levels lacking a red or green code (e.g., ALG1 in math), indicating that no percentage deviation equaled or exceeded five percent, positive percentage deviations indicated that the same pattern in relationships between AntGrd and ActLvls and those found between ExpLvl and AntGrd in other subject areas and grade levels remained consistent.

The ExpLvl and ActLvl discrepancies yielded percentage deviations indicating that teachers tended to under-predict male EOG/EOC performance, while over-predicting female EOG/EOC performance. These ExpLvl and ActLvl discrepancies consistently produced "-1" ExpLvl and ActLvl percentage deviations for males and "+1" ExpLvl and ActLvl percentage deviations for females that were coded yellow, indicating that totals in these categories equaled or exceeded 5 %. Again, even in grade levels lacking a yellow percentage deviation (e.g., Grade 4 males in math), positive percentage deviations indicated the relationship between ExpLvl and ActLvl remained consistent.

Given prior research indicating that females score higher than males in terms of performance grades, despite males scoring higher than females when it comes to standardized assessment results (Brennan et al., 2001; Ross & Kostuch, 2011), it was not surprising to find the same to be true in this study. However, the ExpLvl variable in this study provided a new context for viewing the relationship between a student's AntGrd and the student's ActLvl because it allowed for a direct comparison between what the student's teacher *thought* the student would achieve on an EOG or EOC (ExpLvl) and the AntGrd the teacher assigned to that student.

Table 8

*Percentage Deviations of Discrepancies between Achievement Measures for Students Classified by Gender (Reading)*

| | All Males | | | All Females | | |
|---|---|---|---|---|---|---|
| | ExpLvl and AntGrd | ExpLvl and ActLvl | AntGrd and ActLvl | ExpLvl and AntGrd | ExpLvl and ActLvl | AntGrd and ActLvl |
| Grade 3 | | | | | | |
| -1 | -8.4 | | 11.2 | 9 | | -12.1 |
| 0 | -2.1 | | 0.2 | 2.3 | | -0.2 |
| +1 | 15.7 | | -5.9 | -17 | | 6.3 |
| Grade 4 | | | | | | |
| -1 | -13.7 | | 12 | 14.3 | | -12.4 |
| 0 | -1 | | 0.7 | 1.1 | | -0.7 |
| +1 | 15.4 | | -7 | -16 | | 7.2 |
| Grade 5 | | | | | | |
| -1 | -14.7 | 0.2 | 15.2 | 15.3 | -0.2 | -15.7 |
| 0 | -2.2 | 4.9 | 6.4 | 2.3 | -5 | -6.6 |
| +1 | 19.6 | -5.5 | -11.7 | -20.3 | 5.7 | 12.1 |
| Grade 6 | | | | | | |
| -1 | -13.8 | 11.6 | 32.3 | 14.4 | -12 | -33.6 |
| 0 | -2.2 | 1.4 | 0 | 2.3 | -1.4 | 0 |
| +1 | 27.5 | -8.1 | -15.1 | -28.7 | 8.4 | 15.6 |
| Grade 7 | | | | | | |
| -1 | -20.6 | | 23.6 | 21.3 | | -24.4 |
| 0 | -0.9 | | -1.2 | 0.9 | | 1.3 |
| +1 | 19.1 | | -12.4 | -19.8 | | 12.8 |
| Grade 8 | | | | | | |
| -1 | -12.5 | 14.9 | 26.1 | 12.8 | -15.2 | -26.6 |
| 0 | -1.7 | 0.7 | 2.8 | 1.8 | -0.8 | -2.8 |
| +1 | 20 | -4.6 | -10.3 | -20.5 | 4.7 | 10.5 |
| ENG1 | | | | | | |
| -1 | -18.4 | | 13.1 | 19.3 | | -13.8 |
| 0 | -7.2 | | -7.7 | 7.5 | | 8.1 |
| +1 | 18.7 | | -11.4 | -19.6 | | 12 |

Table 9

*Percentage Deviations of Discrepancies between Achievement Measures for Students Classified by Gender (Math)*

| | All Males | | | All Females | | |
|---|---|---|---|---|---|---|
| | ExpLvl and AntGrd | ExpLvl and ActLvl | AntGrd and ActLvl | ExpLvl and AntGrd | ExpLvl and ActLvl | AntGrd and ActLvl |
| Grade 3 | | | | | | |
| -1 | | | 7.4 | | | -8 |
| 0 | | | 0.3 | | | -0.3 |
| +1 | | | -8.1 | | | 8.7 |
| Grade 4 | | | | | | |
| -1 | -6.9 | 2.7 | 9.9 | 7.2 | -2.9 | -10.3 |
| 0 | -2.6 | 2.2 | -1.3 | 2.7 | -2.3 | 1.4 |
| +1 | 10.4 | -5.8 | -7.4 | -10.8 | 6 | 7.7 |
| Grade 5 | | | | | | |
| -1 | -16 | | 10.4 | 16.6 | | -10.8 |
| 0 | -0.3 | | 0.5 | 0.3 | | -0.5 |
| +1 | 11.2 | | -9.7 | -11.6 | | 10 |
| Grade 6 | | | | | | |
| -1 | -19.5 | 17 | 27.2 | 20.4 | -17.7 | -28.3 |
| 0 | -5.4 | 3.3 | 0.6 | 5.7 | -3.4 | -0.7 |
| +1 | 22.7 | -11.6 | -19.9 | -23.8 | 12.1 | 20.8 |
| Grade 7 | | | | | | |
| -1 | -13.7 | 10.9 | 24.2 | 14.1 | -11.2 | -25.1 |
| 0 | -5 | 0.5 | -2.8 | 5.1 | -0.5 | 2.9 |
| +1 | 21.5 | -5.1 | -11.8 | -22.2 | 5.3 | 12.3 |
| Grade 8 | | | | | | |
| -1 | -11.8 | 13 | 17.1 | 12.1 | -13.2 | -17.4 |
| 0 | -2.5 | -0.9 | 1.9 | 2.6 | 0.9 | -1.9 |
| +1 | 15.7 | -5 | -13.4 | -16.1 | 5.1 | 13.6 |
| ALG1 | | | | | | |
| -1 | -2.5 | 6.2 | 12.4 | 2.5 | -6.6 | -13.3 |
| 0 | -3.2 | 2.5 | -3.2 | 3.3 | -2.7 | 3.5 |
| +1 | 4.3 | -7.4 | -17.9 | -4.4 | 8 | 19.2 |

Table 10
*Percentage Deviations of Discrepancies between Achievement Measures for Students Classified by Gender (Science)*

| | All Males | | | All Females | | |
|---|---|---|---|---|---|---|
| | ExpLvl and AntGrd | ExpLvl and ActLvl | AntGrd and ActLvl | ExpLvl and AntGrd | ExpLvl and ActLvl | AntGrd and ActLvl |
| | | | Grade 5 | | | |
| -1 | -19.2 | 18.4 | 24.5 | 20 | -19.1 | -25.4 |
| 0 | 0.7 | 0.4 | 5 | -0.7 | -0.5 | -5.2 |
| +1 | 19.1 | -10.6 | -19.9 | -19.9 | 11 | 20.6 |
| | | | Grade 8 | | | |
| -1 | -21.6 | 18.8 | 30.2 | 22 | -19.2 | -30.7 |
| 0 | -3.7 | -0.2 | -0.6 | 3.7 | 0.2 | 0.6 |
| +1 | 31.1 | -15 | -24.6 | -31.7 | 15.3 | 24.9 |
| | | | BIO | | | |
| -1 | -24.4 | 16.4 | 15.6 | 23.5 | -15.7 | -14.9 |
| 0 | -4.4 | -0.7 | 0.3 | 4.3 | 0.6 | -0.3 |
| +1 | 14.8 | -9.3 | -22.2 | -14.2 | 8.9 | 21.3 |

Given the percentage deviation patterns from discrepancies in Tables 8, 9, and 10, there appears to be support for the following inferences:

I. males obtained higher ActLvls than they were predicted to obtain by their teachers (ExpLvls), while females obtained ActLvls lower than ExpLvls;

II. teachers predicted that males would score ExpLvls higher than the AntGrds teachers assigned to them (male students); on the other hand, teachers predicted that females would score ExpLvls lower than the AntGrds teachers assigned to them (female students); and

III. teachers assigned AntGrds to males that were lower than what should have been expected given their ActLvls, while females were expected to obtain AntGrds from teachers that were higher than their ActLvls. This finding is consistent with prior research (Brookhart, 1993; Cross & Frary, 1996;

McMillan & Nash, 2000) on performance grade and standardized achievement test patterns, which indicated that teachers' role as advocates makes them consider certain factors, such as student gender, when assigning performance grades because there is an effort to avoid assigning an excessive number of failing performance grades that might suggest a bias against any one student group (Cross & Frary, 1996). The significance of the AntGrd and ActLvl discrepancies found in this study, especially in context of ExpLvl and AntGrd discrepancies and ExpLvl and ActLvl discrepancies, is addressed more thoroughly in Chapter 5.

**Influence of Ethnicity on Relationships among Achievement Measures.** Percentage deviations were calculated to determine if a student's ethnicity affected relationships between AntGrds, ExpLvls, and ActLvls using the same student achievement variable pairings from the gender part of this study: ExpLvl and ActLvl, AntGrd and ActLvl, and ExpLvl and AntGrd. Percentage deviations for students sorted by ethnicity are reported in Table 11 for reading, Table 12 for math, and Table 13 for science.

The relationship between ExpLvl and ActLvl produced consistent patterns in discrepancies across every subject area. For White students, ActLvls were higher than ExpLvls, as indicated by the positive percentage deviations in the "-1" (ExpLvl and ActLvl) categories. For instance, in Table 11, the "-1" percentage deviation was 10.4 for Grade 3, which indicated that White students scoring ActLvls exceeding their ExpLvls in reading were 10.4% higher than expected. The relationship between ExpLvl and ActLvl produced a pattern of discrepancies for Black and Hispanic groups that were opposite of the White discrepancies pattern for the same category; percentage deviations indicated that Black and

Hispanic students scored ActLvls higher than their ExpLvls less often than expected. The ExpLvl and ActLvl comparison between White, Black, and Hispanic percentage deviations is important to note because it indicated that teachers underestimated White achievement while they overestimated Black and Hispanic achievement.

Given the opposite ExpLvl and ActLvl discrepancy patterns for White and for Black and Hispanic groups, the differences between individual ethnic groups are also differences between White and Nonwhite groups. Restating the findings through a White and Nonwhite perspective, teachers underestimated White EOG/EOC achievement and overestimated Nonwhite EOG/EOC achievement. The relationships found between White and Nonwhite ExpLvls and White and Nonwhite ActLvls mirrors the relationships between White to Nonwhite AntGrds and ActLvls, which supports Brennan et al.'s (2001) assertion that teachers seek to produce more equitable results than those found with standardized tests. The AntGrd and ActLvl relationship yielded a similar pattern of discrepancies to those found with the ExpLvl and ActLvl relationship across subject areas. White AntGrd and ActLvl discrepancies yielded patterns, which indicated that White students received AntGrds that were lower than their ActLvls more often than expected. For instance, in Table 11, the " +1" percentage deviation was -8.8 for Grade 3, which indicated that White students in Grade 3 received AntGrds exceeding their ActLvls 8.8% less often than expected. AntGrd and ActLvl discrepancy patterns for Black and Hispanic "+1" categories yielded positive percentage deviations, which indicated that Black and Hispanic students received AntGrds that were higher than their ActLvls more often than expected. The similarity between AntGrd and ActLvl percentage deviation patterns and ExpLvl and ActLvl percentage deviation patterns across ethnicities indicated that, as far as their relationship to ActLvl, ExpLvls and AntGrds were similarly affected by ethnicity.

Two new patterns emerged when examining ExpLvl and AntGrd relationships: differences between Black and Hispanic groups not seen in AntGrd and ActLvl relationships nor in ExpLvl and ActLvl relationships, and differences in reading and math relationships patterns and those found in science. While there were consistent patterns within ExpLvl and ActLvl relationships and AntGrd and ActLvl relationships for White and Nonwhite discrepancies, different patterns emerged in the ExpLvl and AntGrd relationship with Black and Hispanic discrepancies. Although White categories did not yield any patterns of positive percentage deviations, ExpLvl and AntGrd discrepancies did yield patterns with negative percentage deviations in grade levels with reportable data. For instance, in Table 11, the "-1" percentage deviation for Grade 4 was -9.1, which indicated that instances of White students with AntGrds exceeding their ExpLvls was 9.1% lower than expected. All discrepancies for White categories reported that instances of AntGrds exceeding ExpLvls were lower than expected, with the exception of Grade 7 math. Discrepancies for Black percentage deviations yielded patterns indicating that instances of Black students receiving AntGrds exceeding ExpLvls were higher than expected. The difference in ExpLvl and AntGrd discrepancy patterns for White students and Black students is important to note because it shows that teachers assign AntGrds differently to each group of students compared to what the teachers predict each group of students will be able to do on their EOG or EOC assessment (ExpLvl). By assigning AntGrds to Black students that are higher than ExpLvls, and not to White students, teachers are signaling that they use a method of assigning AntGrds that is unique to

Black students - a method which assigns AntGrds that are higher than teachers believe to be academically warranted (as indicated by ExpLvl) to Black students.

In contrast to the discrepancy patterns that emerged from reading and math, the Hispanic ExpLvl and AntGrd discrepancy patterns differed from Black discrepancy patterns, with percentage deviations indicating that Hispanic ExpLvls exceeded AntGrds. Although ExpLvls and AntGrds exceeded ActLvls for Black students and Hispanic students in all subject areas, the beneficial increase in marks Black students get with their AntGrds relative to their ExpLvls did not appear to be happening with Hispanic students, which is important to note because, for other achievement measure relationships, i.e., ExpLvl and ActLvl discrepancies and AntGrd and ActLvl discrepancies, patterns of discrepancies for Black and Hispanic percentage deviations were the same, running counter to White patterns.

ExpLvl and AntGrd discrepancy patterns found in science were different from those found with reading and math. With reading and math, White ExpLvl and AntGrd discrepancy patterns indicated that observations of White students with AntGrds lower than their ExpLvls were higher than expected, but this discrepancy pattern was only observed in Grade 5 science. Black and Hispanic discrepancy patterns in reading and math indicated that observations of Black and Hispanic students with AntGrds exceeding than their ExpLvls were higher than found for all other subject areas. Given that science only provided data from three grade levels, it was more difficult to make generalizations based on patterns found in science, especially when these findings are inconsistent with other subject areas that provide a greater number of grade levels. However, since Grade 8 and Biology (BIO), two of the three total grade levels under science, yielded Black and Hispanic ExpLvl and AntGrd discrepancies indicating that Black and Hispanic students received AntGrds lower than their

ExpLvls, it appears that science teachers assigned AntGrds differently to these groups of students than do reading and math teachers.

Given the percentage deviation patterns from discrepancies in Tables 11, 12, and 13, there appears to be support the following inferences:

1. White students obtained higher ActLvls than they were predicted to obtain by their teachers (ExpLvls), while Black and Hispanic students obtained ActLvls lower than ExpLvls;

2. for reading and math, teachers predicted that White students would score ExpLvls higher than the AntGrds teachers assigned to them (White students); on the other hand, teachers predicted that Black and Hispanic students would score ExpLvls lower than the AntGrds teachers assigned to them (Black and Hispanic students); however, these patterns were not consistent through science; and,

3. teachers assigned AntGrds to White students that were lower than should have been expected given their ActLvls, while Black and Hispanic students were expected to obtain AntGrds from teachers that were higher than their ActLvls.

Table 11

*Percentage Deviations of Discrepancies between Achievement Measures for Students Classified by Ethnicity (Reading)*

| | ExpLvl and ActLvl | | | AntGrd and ActLvl | | | ExpLvl and AntGrd | | |
|---|---|---|---|---|---|---|---|---|---|
| **Grade 3** | | | | | | | | | |
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | 10.4 | -22.2 | -18.1 | 0.1 | -0.5 | 0.1 | | | |
| 0 | 5.2 | -7.5 | -13 | 7 | -9.9 | -18.2 | | | |
| +1 | -12 | 20.3 | 27 | -8.8 | 12.6 | 22.6 | | | |
| **Grade 4** | | | | | | | | | |
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | | | | -1.6 | 10.4 | -5.7 | -9.1 | 16.9 | 19.2 |
| 0 | | | | 5.6 | -15 | -6.4 | 4.4 | -9.5 | -7.5 |
| +1 | | | | -5.4 | 11.3 | 10.1 | -7.4 | 19.4 | 8.8 |
| **Grade 5** | | | | | | | | | |
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | -2.1 | 16.8 | -13 | -6.9 | 32.5 | -10.9 | -9.5 | 33.4 | -0.2 |
| 0 | 3.2 | -2 | -13.2 | 3 | 0.3 | -15.3 | 2.1 | -6.3 | -1.3 |
| +1 | -3.2 | -1.3 | 17.8 | -0.6 | -11.5 | 19.2 | 2.6 | -13.5 | 6.2 |
| **Grade 6** | | | | | | | | | |
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | 10.1 | -28.9 | -9.4 | 9.2 | -23 | -13.5 | | | |
| 0 | 3.3 | -5.9 | -8.6 | 3.9 | -8.3 | -8.3 | | | |
| +1 | -10.3 | 24.1 | 18 | -8.2 | 19 | 14.5 | | | |
| **Grade 7** | | | | | | | | | |
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | 7.9 | -22.2 | -10 | | | | | | |
| 0 | 2.5 | -5.8 | -4.9 | | | | | | |
| +1 | -6 | 15.5 | 9.7 | | | | | | |
| **Grade 8** | | | | | | | | | |
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | 1.4 | -3.9 | -1.8 | 6.1 | -8.2 | -23.9 | -1.2 | 5.8 | -3.3 |
| 0 | 3.5 | -15.9 | 7.4 | 2.5 | -7.6 | -2.2 | 2.3 | 1.6 | -17.4 |
| +1 | -3.9 | 17.5 | -7.2 | -3.7 | 7.8 | 9.2 | -4 | -8.3 | 40.7 |
| **ENG1** | | | | | | | | | |
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | 3.5 | -6.5 | -9.9 | | | | | | |
| 0 | 1.2 | -2.9 | -0.2 | | | | | | |
| +1 | -6.2 | 12.9 | 14.8 | | | | | | |

Table 12

*Percentage Deviations of Discrepancies between Achievement Measures for Students Classified by Ethnicity* (Math)

Grade 3

| | ExpLvl and ActLvl | | | AntGrd and ActLvl | | | ExpLvl and AntGrd | | |
|---|---|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | 3.6 | -6.2 | -7.8 | -3.5 | 9.1 | 4 | -10.9 | 23.8 | 17.8 |
| 0 | 3.2 | -4.5 | -8.4 | 4.8 | -7.7 | -11.1 | 2 | -5 | -2.6 |
| +1 | -10.7 | 15.9 | 26.5 | -4.7 | 4 | 15 | 6.7 | -12 | -14 |

Grade 4

| | ExpLvl and ActLvl | | | AntGrd and ActLvl | | | ExpLvl and AntGrd | | |
|---|---|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | -2.4 | -8 | 20.1 | -5 | 12.3 | 6.9 | -7.8 | 25.8 | 2.7 |
| 0 | 5 | -9.3 | -10.4 | 7.7 | -15.9 | -14.1 | 3.4 | -7.4 | -5.6 |
| +1 | -7.8 | 21.6 | 7.6 | -7 | 12.7 | 14.9 | -2.4 | -11.4 | 24.1 |

Grade 5

| | ExpLvl and ActLvl | | | AntGrd and ActLvl | | | ExpLvl and AntGrd | | |
|---|---|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | | | | -7.3 | 25.2 | 0.9 | -7.7 | 25.4 | 2 |
| 0 | | | | 3.7 | -12.2 | -1.4 | 2.2 | -5.6 | -3 |
| +1 | | | | 0.3 | -2 | 1.5 | 1 | -12.1 | 11.9 |

Grade 6

| | ExpLvl and ActLvl | | | AntGrd and ActLvl | | | ExpLvl and AntGrd | | |
|---|---|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | 12.1 | -30.9 | -16.7 | 5.9 | -11 | -14.3 | | | |
| 0 | 2.8 | -8.5 | -1.8 | 1.7 | -5.7 | -0.2 | | | |
| +1 | -8.9 | 24.8 | 9.2 | -6.3 | 14.9 | 10.4 | | | |

Grade 7

| | ExpLvl and ActLvl | | | AntGrd and ActLvl | | | ExpLvl and AntGrd | | |
|---|---|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | 2.9 | -20.8 | 16.7 | 10.3 | -24.9 | -19.5 | 6.9 | -6.8 | -28.9 |
| 0 | 6.7 | -18.6 | -8.6 | 2.7 | -9.8 | 0.2 | -1.9 | 5.7 | 1.8 |
| +1 | -11 | 36 | 6 | -9.5 | 26.7 | 11.9 | -3.2 | -10.2 | 35.2 |

Grade 8

| | ExpLvl and ActLvl | | | AntGrd and ActLvl | | | ExpLvl and AntGrd | | |
|---|---|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | 0.9 | -0.3 | -5 | 1.5 | -3.6 | -2.6 | -5.4 | 19.6 | -2 |
| 0 | 4.6 | -14.3 | -3 | 5.9 | -9.9 | -18.9 | 4 | -8.9 | -9.1 |
| +1 | -7 | 20.3 | 6.7 | -7.3 | 13.1 | 22.2 | -7.2 | 5.6 | 34.8 |

ALG1

| | ExpLvl and ActLvl | | | AntGrd and ActLvl | | | ExpLvl and AntGrd | | |
|---|---|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | -7.9 | 18 | 8.9 | | | | -4.8 | 12.6 | 0.1 |
| 0 | 6.2 | -13.6 | -7.8 | | | | 2.5 | -7.5 | 1.8 |
| +1 | 7.5 | -19.9 | -2.6 | | | | 5.4 | -12.6 | -3.6 |

Table 13
*Percentage Deviations of Discrepancies between Achievement Measures for Students Classified by Ethnicity* (Science)

Grade 5

| | ExpLvl and ActLvl | | | AntGrd and ActLvl | | | ExpLvl and AntGrd | | |
|---|---|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | 8.1 | -18.2 | -14.8 | -4.5 | 13.4 | 3.4 | -11.2 | 26.2 | 18.7 |
| 0 | 3.2 | -2.6 | -12.4 | 7.5 | -14.8 | -16.6 | 1.4 | -4 | -1 |
| +1 | -9.5 | 13.9 | 27.6 | -5.9 | 8.9 | 16.8 | 6.3 | -12.3 | -13.9 |

Grade 8

| | ExpLvl and ActLvl | | | AntGrd and ActLvl | | | ExpLvl and AntGrd | | |
|---|---|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | 8.7 | -31.2 | 2.2 | 8.3 | -25.9 | -5 | 3 | -4.9 | -9.8 |
| 0 | 5.1 | -14.8 | -5 | 6.9 | -17 | -12.6 | 1.7 | -2 | -6.7 |
| +1 | -15.7 | 50.5 | 6.7 | -15.3 | 42 | 19.3 | -7.7 | 10.7 | 28.9 |

BIO

| | ExpLvl and ActLvl | | | AntGrd and ActLvl | | | ExpLvl and AntGrd | | |
|---|---|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | White | Black | Hisp | White | Black | Hisp |
| -1 | 9.3 | -21.8 | -16.3 | | | | 2.5 | -3.5 | -9.9 |
| 0 | 3.5 | -7.5 | -7.5 | | | | 3.7 | -8.6 | -6.4 |
| +1 | -11 | 24.8 | 21.3 | | | | -8.4 | 17.4 | 19.4 |

## Discussion

The goal of this study was to determine how a student's subgroup status (male or female, and White, Black, or Hispanic) and level of schooling (elementary, middle, or high school) affected the relationship between a student's anticipated performance grade and a corresponding standardized measures of achievement. Anticipated performance grades were compared to students' expected EOG/EOC achievement levels (both marks were coded by the teacher) and the students' actual EOG/EOC levels. After students were grouped by subgroup classification, discrepancy variable relationships were determined by calculating differences between achievement variables. For instance, when I compared anticipated performance grades and actual EOG/EOC levels, I coded a "-1" when EOG/EOC scores were higher than anticipated grades, a "0" when the two levels were equal, and a "+1" when anticipated grades were higher than EOG/EOC scores.

Observed discrepancies between achievement variables followed a general pattern, indicating that anticipated performance grades often mitigated performance differences between genders and ethnicities in a way that provided higher performance grade marks (when compared to EOG/EOC achievement levels) to females and to Black and Hispanic students, a finding that is consistent with prior research (Brennan et al., 2001; Martinez et al., 2009). Differences across gender and ethnic subgroups showed that the discrepancies comparing anticipated performance grades and actual EOG/EOC achievement levels, (AntGrd and ActLvl), followed the same pattern as the discrepancies comparing expected EOG/EOC achievement levels and anticipated performance grades, (ExpLvl and AntGrd), indicating that teachers were good predictors of how a student's teacher-anticipated performance grade would relate to his or her (the student's) EOG and EOC achievement level. However, comparisons of expected EOG/EOC achievement levels to actual EOG and EOC achievement levels, as measured by the ExpLvl and ActLvl discrepancy, showed that gaps between anticipated performance grades and actual EOG and EOC achievement scores across all grade levels and subject areas are wider than teachers acknowledge for both genders and all three ethnicities.

The findings from the student subgroup part of the study indicate that individual teachers usually have a good idea of how the performance grade they assign to a student will fare against that same student's EOG/EOC achievement level – however, teachers are prone to underestimate discrepancies between the performance grades they assign and students' EOG/EOC achievement levels. For example, teachers acknowledge that they tend to assign performance grades to female, Black, and Hispanic students that are higher than academically warranted when compared to the expected EOG/EOC achievement levels for the same grade level or subject area. Given the prior research noting the tendency of teachers to assign performance grades to groups of students that are higher than academically warranted (Brennan et al., 2001; Cizek et al., 1996; Martinez et al., 2009; Randall & Engelhard, 2009; Ross & Kostuch, 2011), it is reasonable to suspect that academic standards have been manipulated in order to create a more equitable level of achievement in the classroom. Teachers' desires for equitable achievement in the classroom is a byproduct of grading practices adopted by teachers stemming from their need to pull for their students (McMillan & Nash, 2000); or, as Brookhart (1993) noted, teachers simultaneously serve in conflicted roles as "judge and advocate" which affect their ability to assess student performance objectively.

This study's findings also produced a pattern of attenuating correlations between performance grades and standardized measures of achievement across students' levels of schooling that was inconsistent with prior research, which noted that elementary teachers tend to be more lenient graders due to their tendency to protect the self-esteem of their students and, thus, assign performance grades that are less academic in nature than their middle and high school colleagues, who spend less of the day with their students and, thus, feel less inclined to include nonacademic factors into the grades they assign (Brookhart, 1994; Randall & Engelhard, 2009). For this study, correlations between anticipated performance grades and actual EOG/EOC achievement levels were strongest at the elementary level, weakened in middle school, and were weakest at the high school level across reading, math, and science. While inconsistent with prior research, the relationship between anticipated performance grades and EOG/EOC achievement levels (AntGrd and ActLvl) was consistent with the relationship between anticipated performance grades and expected EOG/EOC achievement levels (AntGrd and ExpLvl), with exceptions in high school math and science. One possible explanation for this inconsistency could be grading policies adopted at the elementary, middle, and high school levels within the district. Since the data used for the study came from only one district, issues such as a district-wide grading policy could make a big difference in how well performance grades relate to academic performance – especially academic performance as measured by more objective measures of academic achievement, i.e., EOGs and EOCs. An examination of district-mandated grading policies used at all three levels of schooling could provide more insight into how those directives may have influenced any differences in how performance grades were assigned to students in this study and those in other studies. One other possible explanation for performance grades to become less associated with EOG/EOC scores as students move into middle and high school could be that performance grades often serve as a "gatekeeping" mechanism for access to athletics and extracurricular activities to which students at the elementary level do not have access. If this explanation factored into the differences seen from the relationships between performance grades and standardized test scores as students, moved through levels of schooling, one would expect to see performance grades rise relative to standardized test scores as they did in this study in order to help students maintain their eligibility for extracurricular events as students moved into middle and high school; however, this possibility was not examined in this study.

## Implications

Helson's Adaptation Level Theory (1964) suggests that a manipulation of standards affects the group to whom it is applied, i.e., when teachers assign performance grades using different standards to subgroups, (such as female, Black, and Hispanic students), teachers are communicating to these students that whatever they are doing to earn their A or B performance grade is above average or excellent work, and, therefore, defines what these groups of students view as "above average" or "excellent." The students in these subgroups have their spectrums of performance possibility stunted by being subjected to lowered standards compared to their peers (in this case, male or White students) who are being held to higher standards. When groups of students are contrasted (male and females or White, Black, and Hispanic students) and are held to different standards within the classroom despite ultimately taking the same standardized achievement test, such as the EOG or EOC, differences in performance can — and often do — result. Consistent differences in EOG and EOC performance between subgroups is why targets for closing subgroup achievement gaps are often set as a part of school and district performance goals.

What complicates the issue of reporting students' academic progress even more is that, even with teachers expecting differences between EOG/EOC achievement levels and performance grades, teachers underestimate discrepancies between performance grades and how students actually perform on EOGs and EOCs. While expected EOG/EOC achievement levels and actual EOG/EOC performance levels often relate to anticipated performance grades in the same manner for each subgroup of students, expected EOG/EOC achievement levels and anticipated performance grade discrepancies often underestimate discrepancies between anticipated performance grades and actual EOG/EOG achievement levels. This makes the aforementioned discrepancies between expected EOG/EOC achievement levels and anticipated performance grades even more noteworthy, since the performance grades assigned to groups of students have even less of a connection to EOG and EOC performance than teachers expect.

The ostensible goal of the No Child Left Behind Act (2002) (NCLB) was to ensure that all children had access to sound education regardless of their backgrounds. Standardized assessments administered to students were designed to assess students' knowledge of grade level or course material at the end of the learning sequence. However, despite the time, effort, and money invested in the implementation of NCLB, and despite performance grades that report a more relatively equitable distribution of academic performance across subgroups, the findings from this study indicate that standardized test scores are not reporting a similarly shrinking academic achievement gap; and the results from this study indicate that teachers are aware of this discrepancy when they assign expected EOG/EOC achievement levels that are not aligned with anticipated performance grades. Furthermore, the gaps that teachers are trying to close through assigning performance grades that do not align with teacher's expectation of EOG/EOC achievement levels are even wider than teachers are aware of when comparing students' anticipated performance grades to their actual EOG/EOC achievement levels. The lack of awareness on teachers' parts of how wide the actual differences in academic performance is between subgroups makes it harder, if not impossible, to adequately address the performance gaps between subgroups; or, as Good et al. (1969) noted nearly fifty years ago, assigning performance grades that do not reflect academic performance makes it more difficult to match student ability to appropriate levels of instruction in order to maximize student outcomes.

Prior studies have noted the conflicts that teachers face when assigning performance grades to students (Brookhart, 2003; Cross & Frary, 1996; Guskey, 2011; Nitko, 2004; Pedulla, Airasian, & Madaus, 1980; Randall & Engelhard, 2009). It should be noted that some of the studies predate NCLB and its testing mandate, while other studies noting similar issues have been published since the implementation of NCLB. However, many of the same concerns and conversations about how to improve teachers' grading practices remain unchanged despite increased training in classroom assessment being offered to pre-service and current teachers. One constant is that performance grades have a profound level of implications for students, ranging from access to personal privileges at home (Olson, 1989) to class rank and honor roll determinations at school (Bowers, 2009), the latter of which can affect post-secondary options available to students. The assignment of performance grades has implications for teachers as well, since teachers are faced with professional pressures to assign performance grades reporting certain levels of achievement in order to avoid grading distributions that suggest possible biases against particular groups of students (Cross & Frary, 1996). Teachers' recognition of how their performance grades are perceived helps explain why prior research cited that performance grade distribution produced more equitable achievement results than standardized assessments (Brennan et al., 2001).

The assignment of performance grades affects the day-to-day or semester-to-semester options faced by students and teachers, with the most important affected being the opportunity to inform the instructional sequence for each student in order to close achievement gaps. However, if teachers continue to avoid reporting performance grades that reflect the academic performance they expect from a student on his or her EOG or EOC, real opportunities to make meaningful decisions about student learning are being lost throughout the school year. While achievement gaps between subgroups of students are being "closed" in the classroom, there is evidence, as found in this study, that this same gap is not being closed when more objective assessments of academic achievement are used.

**Limitations and Suggestions for Further Research**

This study examined student achievement data from one school district. Findings related to the relationships between measures of student achievement are, therefore, impacted by the school district's policies guiding grading practices at all levels of schooling. An example of how district-level influence might affect achievement variable relationships would be if teachers were required to count certain percentages of homework or classwork as a part of a student's overall grade for a class or course. District-level grading policies, or any changes to these policies over the three years for which data were collected, were not examined within this study.

School-level variables, other than grade levels, were not examined for influence on the relationship between student achievement measures. School-level variables such as school leadership, student demographic makeup, and individual school classroom conditions are factors that were not examined from this study's model. These issues work, sometimes in concert, to determine factors such as students' placement in classes, which research has indicated to have an effect on student achievement (LaPrade, 2011; Oakes & Wells, 1997).

Finally, one student-level variable that was not examined, but has been found to have an influence on student achievement (Willingham et al., 2002; Zwick & Sklar, 2005) is a student's socio-economic status (SES). The district did not make this information available.

Studies intending to build from this study's findings should consider the aforementioned influences not examined in this study. Extending the database of students beyond the boundaries of one school district would help put findings from this study that were inconsistent with prior studies of this nature (i.e., attenuating agreement between report card grades and standardized test scores as levels of schooling increased) into context. In addition, addressing some of the school-level variables that were not controlled for in this study might lend some further explanation for the findings of this study as well, since controlling for these school-level variables might expose some achievement variable relationships impacting larger findings from district-level achievement variable relationships.

## References


Agresti, A. (1984). *Analysis of ordinal categorical data*. New York: Wiley.

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *Clearing House*, *78*(5), 218-223.

Beziat, T. L. R., & Coleman, B. K. (2015). Classroom assessment literacy: Evaluating pre-service teachers. *The Researcher, 27*(1), 25-30.

Bishop, J. H. (1992). Why US students need incentives to learn. *Educational leadership*, *49*(6), 15-18.

Bonner, S., & Chen, P. (2009). Teacher candidates' perceptions about grading and constructivist teaching. *Educational Assessment*, *14*(2), 57-77.

Bowers, A. J. (2009). Reconsidering grades as data for decision making: More than just academic knowledge. *Journal of Educational Administration*, *47*(5), 609-629.

Brackett, M. A., Floman, J. L., Ashton-James, C., Cherkasskiy, L., & Salovey, P. (2013). The influence of teacher emotion on grading practices: a preliminary look at the evaluation of student writing. *Teachers and Teaching: Theory and practice, 19*, (6), 634-646.

Brennan, R. T., Kim, J., Wenz-Gross, M., & Siperstein, G. N. (2001). The relative equitability of high-stakes testing versus teacher-assigned grades: An analysis of the Massachusetts Comprehensive Assessment System (MCAS). *Harvard Educational Review*, *71*(2), 173-217.

Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues & Practice*, *10*(1), 35. doi:10.1111/j.1745-3992.1991.tb00182.x

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, *30*(2), 123-142.

Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, *7*(4), 279-301.

Brookhart, S. M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice*, *22*(4), 5-12.

Brookhart, S. M. (2015). *Performance assessment: Showing what students know and can do.* West Palm Beach, FL: Learning Sciences International.

Chappuis, J., Stiggins, R., Chappuis, S., & Arter, J. (2012). *Classroom assessment for student learning doing it right – using it well* (2nd ed.). Boston, MA.: Pearson.

Church, M., Elliot, A., & Gable, S. (2001). Perceptions of classroom environment, achievement goals, and achievement outcomes. *Journal of Educational Psychology, 93*(1), 43-54.

Cizek, G. J. (1996). Grades: The final frontier in assessment reform. *NASSP Bulletin*, *80* (584), 103–10.

Cizek, G., Fitzgerald, S., & Rachor, R. (1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, *3*(2), 159-179.

Conover W.J., (1980). Practical non-parametric statistics (2nd ed.). New York: John Wiley and Sons.

Cornwell, C., Mustard, D. B., & 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.

Cramér, H., (1946). *Mathematical methods of statistics*. Princeton: Princeton University Press.

Cross, L. H., & Frary, R. B. (1996). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, *12*(1), 53-72.

Dattalo, P. (2009). A review of software for sample size determination. *Evaluation & the Health Professions*, *32*, 229-248. doi:10.1177/0163278709338556

DeLuca, C., & Bellara, A. (2013). The current state of assessment education: aligning policy, standards, and teacher education curriculum. *Journal of Teacher Education*, *64*(4), 356–372.

Deutsch, M. (1985). *Distributive justice: A social-psychological perspective*. New Haven: Yale University Press.

Dornbusch, S. M., Ritter, P. L., Leiderman, P. H., Roberts, D. F., & Fraleigh, M. J. (1987). The relation of parenting style to adolescent school performance. *Child Development*, (5), 1244.

Frey, B. B., & Schmitt, V. L. (2010). Teachers' classroom assessment practices. *Middle Grades Research Journal, 5*(3), 107-117.

Gardner, D. P. (1983). *A nation at risk*. Washington, DC: The National Commission on Excellence in Education, US Department of Education.

Good, T., Williams, D., Peck, R., & Schmidt, L. (1969). *Teacher assessment of pupil potential*. Report Series No. 33. University of Texas at Austin, The Research and Development Center for Teacher Education.

Guskey, T. (1994). What you assess may not be what you get. *Educational Leadership*, *51*(6), 51-54.

Guskey, T. (2001). Helping standards make the grade. *Educational Leadership*, *59*(1), 20-27.

Guskey, T. (2004). Are zeros your ultimate weapon? *Education Digest*, *70*(3), 31-35.

Guskey, T. (2011). Five obstacles to grading reform. *Educational Leadership*, *69*(3), 16-21.

Guthrie, J. W., & Springer, M. G. (2004). 'A Nation at Risk' revisited: Did 'wrong' reasoning result in 'right' results? At what cost? *Peabody Journal of Education*, *1*, 7.

Helson, H. (1964). *Adaptation level theory: An experimental and systematic approach to behavior.* New York, NY: Harper and Row.

Herfordt-Stöpel, J. E., & Hörstermann, T. (2012). The influence of sccountability on teachers' decision making in a simulated classroom. *International Journal for Cross-Disciplinary Subjects in Education (IJCDSE), 2*(2)*,* 985-992.

Hershberg, T. (2005). Value-added assessment and systemic reform: A response to the challenge of human capital development. *Phi Delta Kappan, 87*(4), 276-283.

Hills, J. R. (1981). *Measurement and evaluation in the classroom* (2nd ed). Columbus, OH: Merrill.

Hochweber, J., Hosenfeld, I., & Klieme, E. (2014). Classroom composition, classroom management, and the relationship between student attributes and grades. *Journal Of Educational Psychology*, *106*(1), 289-300. doi:10.1037/a0033829

Impara, J., Plake, B., & Fager, J. (1993). Teachers' assessment background and attitudes toward testing. *Theory into Practice*, *32*(2), 113-117.

Kendall, M. G. (1938). A New Measure of Rank Correlation. *Biometrika*, *1/2,* 81.

Kohn, A. (2002). The dangerous myth of grade inflation. *Chronicle of Higher Education*, *49*(11), B7.

Kohn, A. (2015). The war on American public education. In *Harnessing the dynamics of public education: Preparing for a return to greatness*. Lanham, MD: Rowman & Littlefield.

LaPrade, K. (2011). Removing instructional barriers: One track at a time. *Education*, *131*(4), 740-752.

Lowry, R. (1998). *VassarStats: Website for statistical computation*. Retrieved from http://vassarstats.net/

Mansfield, H. C. (2001). Grade inflation: It's time to face the facts. *Chronicle of Higher Education*, *47*(30), B24.

Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: ASCD.

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

Martínez, J., Stecher, B., & Borko, H. (2009). Classroom assessment practices, teacher judgments, and student achievement in mathematics: Evidence from the ECLS. *Educational Assessment*, *14*(2), 78-102.

McCandless, B. R., Roberts, A., & Starnes, T. (1972). Teachers' marks, achievement test scores, and aptitude relations with respect to social class, race, and sex. *Journal of Educational Psychology*, *63*(2), 153-159. doi:10.1037/h0032646

McMillan, J. H. (1999). The devastating effect of zeros on grades: What can be done? *Practice*, *10*, 36.

McMillan, J. H. (2014). *Classroom assessment: Principles and practice for effective standards-based Instruction*, (6th ed). New York, NY: Pearson Education.

McMillan, J. H., & Nash, S. (2000, April). *Teachers' classroom assessment and grading decision making*. Paper presented at the Annual Meeting of the National Council of Measurement in Education, New Orleans.

Muñoz, M. A., & Guskey, T. R. (2015). Standards-Based Grading and Reporting Will Improve Education. *Phi Delta Kappan*, *96*(7), 64-68.

Napierala, M. A. (2012). What is the Bonferroni correction. *AAOS Now*, *6*(4), 40-40.

Nitko, A. J., (2004). *The educational assessment of students* (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

North Carolina Department of Public Instruction (NCDPI). (2009). *North Carolina reading comprehension tests technical report*. Raleigh, NC: Academic Services and Instructional Support.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Oakes, J., Wells, A. S., Jones, M., & Datnow, A. (1997). Detracking: The social construction of ability, cultural politics, and resistance to reform. *Teachers College Record*, (3), 482.

O'Connor, K. (2007). *A repair kit for grading*. Portland, OR: Educational Testing Service.

Olson, G. H. (1989, March). *On the validity of performance grades: The relationship between teacher-assigned grades and standard measures of subject matter acquisition*. Paper presented at the meeting of the National Council on Measurement in Education, San Francisco, CA.

Pedulla, J. J., Airasian, P. W., & Madaus, G. F. (1980). Do teacher ratings and standardized test results of students yield the same information? *American Educational Research Journal*, *17*(3), 303-307.

Popham, W. J. (2007). Instructional insensitivity of tests: Accountability's dire drawback. *Phi Delta Kappan, 89*(2), 146–150.

Popham, W. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory Into Practice*, *48*(1), 4-11.

Randall, J., & Engelhard, G. (2009). Differences between teachers' grading practices in elementary and middle schools. *Journal of Educational Research*, *102*(3), 175-185.

Resh, N. (2009). Justice in grades allocation: Teachers' perspective. *Social Psychology of Education: An International Journal*, *12*(3), 315-325.

Ross, J. A., & Kostuch, L. (2011). Consistency of report card grades and external assessments in a Canadian province. *Educational Assessment, Evaluation And Accountability*, *23*(2), 159-180. doi:10.1007/s11092-011-9117-3

Schafer, W. (1993). Assessment literacy for teachers. *Theory into Practice*, *32*(2), 118-26.

Schneider, M., Teske, P., & Marschall, M. (2000). *Choosing schools: Consumer choice and the quality of American schools*. Princeton, N.J. : Princeton University Press, c2000.

Spellings, M. (2008). A nation accountable: Twenty five years after a nation at risk. Retrieved from http://www2.ed.gov/rschstat/research/pubs/accountable/accountable.pdf

Starch, D. D. (1913). Reliability and distribution of grades. *Science*, *38*(983), 630–636. doi:10.1126/science.38.983.630

Stedman, L. C.. (1994). The Sandia Report and U.S. Achievement: An assessment. *The Journal of Educational Research*, *87*(3), 133–146. Retrieved from http://www.jstor.org/stable/27541911

Steele, A. G. (1911). Training teachers to grade. *The Pedagogical Seminary*, *18*(4), 523-532. doi: 10.1080/08919402.1911.10532798

Stiggins, R. J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice*, *20*(3), 5-14.

Stiggins, R. J., & Chappuis, J. (2005). Using student-involved classroom assessment to close achievement gaps. *Theory Into Practice*, *44*(1), 11-18.

Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice*, *8*(2), 5-14.

Trevisan, M. S. (1999). Administrator certification requirements for student assessment competence. *Applied Measurement in Education*, *12*(1), 1-11.

Tuck, K. D. (1995). *Parent satisfaction and information: A customer satisfaction survey*. Retrieved from http://www.eric. ed.gov/PDFS/ED401326.pdf

Uitenbroek, D. G. (1997). *SISA Simple Interactive Statistical Analysis*. Retrieved from http://www.quantitativeskills.com/sisa/

U.S. Department of Education. (November 1992). *Parental satisfaction with schools and the need for standards*. Washington, DC: Office of Educational Research and Improvement.

U.S. Department of Education. (1994). *Research report: What do student grades mean? Differences across schools*. Washington, DC: Office of Educational Research and Improvement.

Volante, L., & Fazio, X. (2007). Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education*, *30*(3), 749-770.

Waltman, K. K., & Frisbie, D. A. (1994). Parents' understanding of their children's report card grades. *Applied Measurement in Education*, *7*(3), 223-40.

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, *39*(1), 1-37.

Wise, S. L., Lukin, L. E., & Roos, L. L. (1991). Teacher beliefs about training in testing and measurement. *Journal of Teacher Education*, (1), 37.

Zwick, R., & Sklar, J. C. (2005). Predicting college grades and degree completion using high school and SAT scores: the role of student ethnicity and first language. *American Educational Research Journal*, *42*(3), 439–464.

## Appendix A

*Observed and Expected Discrepancies between AntGrd and ActLvl for Students Classified by Ethnicity in Math*

| | Observed Frequencies Grade 3 Math AntGrd, ActLvl | | | | Expected Frequencies Grade 3 Math AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 755 | 240 | 193 | 1188 | 782.502 | 220.01 | 185.488 |
| 0 | 1377 | 341 | 277 | 1995 | 1314.05 | 369.461 | 311.489 |
| 1 | 724 | 222 | 207 | 1153 | 759.448 | 213.528 | 180.023 |
| TOTAL | 2856 | 803 | 677 | 4336 | | | |

$\chi^2(4)=18.15$; $p = .0012$
V = .0457

| | Observed Frequencies Grade 4 Math AntGrd, ActLvl | | | | Expected Frequencies Grade 4 Math AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 745 | 245 | 193 | 1183 | 782.25 | 214.865 | 179.683 |
| 0 | 1377 | 299 | 253 | 1929 | 1270.971 | 346.541 | 288.673 |
| 1 | 784 | 264 | 223 | 1271 | 838.88 | 230.472 | 189.773 |
| TOTAL | 2906 | 808 | 669 | 4383 | | | |

$\chi^2(4)=40.7$; $p = .0001$
V = .0681

| | Observed Frequencies Grade 5 Math AntGrd, ActLvl | | | | Expected Frequencies Grade 5 Math AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 672 | 258 | 149 | 1079 | 721.056 | 192.984 | 147.659 |
| 0 | 1410 | 339 | 273 | 2022 | 1357.83 | 380.358 | 276.822 |
| 1 | 854 | 237 | 176 | 1267 | 851.438 | 241.74 | 173.36 |
| TOTAL | 2936 | 834 | 598 | 4368 | | | |

$\chi^2(4)=24.88$; $p = .0001$
V = .0534

| | Observed Frequencies Grade 6 Math AntGrd, ActLvl | | | | Expected Frequencies Grade 6 Math AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 720 | 175 | 111 | 1006 | 677.52 | 194.25 | 126.873 |
| 0 | 1231 | 330 | 230 | 1791 | 1210.073 | 348.81 | 230.46 |
| 1 | 905 | 321 | 203 | 1429 | 962.015 | 273.171 | 181.888 |
| TOTAL | 2856 | 826 | 544 | 4226 | | | |

$\chi^2(4)=56.84$; $p = .0001$
V = .0829

## Appendix A - Continued

| | Observed Frequencies Grade 7 Math AntGrd, ActLvl | | | | Expected Frequencies Grade 7 Math AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 723 | 137 | 91 | 951 | 648.531 | 171.113 | 108.745 |
| 0 | 1248 | 305 | 210 | 1763 | 1214.304 | 334.89 | 209.58 |
| 1 | 957 | 373 | 204 | 1534 | 1047.915 | 273.409 | 179.724 |
| TOTAL | 2928 | 815 | 505 | 4248 | | | |

$\chi^2(4)=59.86$; $p = .0001$
V = .0839

| | Observed Frequencies Grade 8 Math AntGrd, ActLvl | | | | Expected Frequencies Grade 8 Math AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 502 | 144 | 75 | 721 | 494.47 | 149.184 | 76.95 |
| 0 | 1523 | 362 | 224 | 2109 | 1433.143 | 397.838 | 266.336 |
| 1 | 962 | 361 | 175 | 1498 | 1032.226 | 313.709 | 136.15 |
| TOTAL | 2987 | 867 | 474 | 4328 | | | |

$\chi^2(4)=30.32$; $p = .0001$
V = .0592

| | Observed Frequencies ALG1 AntGrd, ActLvl | | | | Expected Frequencies ALG1 AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 1441 | 551 | 246 | 2238 | 1469.82 | 515.185 | 250.674 |
| 0 | 1182 | 405 | 207 | 1794 | 1178.454 | 414.72 | 200.79 |
| 1 | 830 | 260 | 136 | 1226 | 804.27 | 281.58 | 137.36 |
| TOTAL | 3453 | 1216 | 589 | 5258 | | | |

$\chi^2(4)=5.97$; $p = .2014$
V = .0238

# Appendix B

*Observed and Expected Discrepancies between ExpLvl and ActLvl for Students Classified by Ethnicity in Math*

| | Observed Frequencies Grade 3 Math ExpLvl, ActLvl | | | | Expected Frequencies Grade 3 Math ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 607 | 154 | 128 | 889 | 585.148 | 163.548 | 137.984 |
| 0 | 1643 | 426 | 345 | 2414 | 1590.424 | 445.17 | 373.98 |
| 1 | 605 | 220 | 203 | 1028 | 669.735 | 185.02 | 149.205 |
| TOTAL | 2855 | 800 | 676 | 4331 | | | |

$\chi^2(4)=31.32$; $p = .0001$
V = .0601

| | Observed Frequencies Grade 4 Math ExpLvl, ActLvl | | | | Expected Frequencies Grade 4 Math ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 477 | 125 | 135 | 737 | 488.448 | 135 | 107.865 |
| 0 | 1640 | 394 | 322 | 2356 | 1558 | 430.642 | 355.488 |
| 1 | 786 | 288 | 211 | 1285 | 847.308 | 225.792 | 194.964 |
| TOTAL | 2903 | 807 | 668 | 4378 | | | |

$\chi^2(4)=34.48$; $p = .0001$
V = .0628

| | Observed Frequencies Grade 5 Math ExpLvl, ActLvl | | | | Expected Frequencies Grade 5 Math ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 458 | 153 | 113 | 724 | 485.022 | 136.323 | 97.293 |
| 0 | 1682 | 442 | 316 | 2440 | 1639.95 | 464.1 | 333.38 |
| 1 | 795 | 237 | 169 | 1201 | 807.72 | 228.705 | 164.437 |
| TOTAL | 2935 | 832 | 598 | 4365 | | | |

$\chi^2(4)=9.05$; $p = .0599$
V = .0322

| | Observed Frequencies Grade 6 Math ExpLvl, ActLvl | | | | Expected Frequencies Grade 6 Math ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 438 | 78 | 62 | 578 | 385.002 | 102.102 | 72.354 |
| 0 | 1516 | 390 | 276 | 2182 | 1473.552 | 423.15 | 280.968 |
| 1 | 902 | 357 | 206 | 1465 | 982.278 | 268.464 | 187.048 |
| TOTAL | 2856 | 825 | 544 | 4225 | | | |

$\chi^2(4)=49.91$; $p = .0001$
V = .0769

## Appendix B - Continued

| | Observed Frequencies<br>Grade 7 Math ExpLvl, ActLvl | | | | Expected Frequencies<br>Grade 7 Math ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 421 | 90 | 82 | 593 | 408.791 | 108.72 | 68.306 |
| 0 | 1604 | 340 | 236 | 2180 | 1496.532 | 403.24 | 256.296 |
| 1 | 904 | 384 | 185 | 1473 | 1003.44 | 245.76 | 173.9 |
| TOTAL | 2929 | 814 | 503 | 4246 | | | |

$\chi^2(4)=79.93$; $p = .0001$
V = .097

| | Observed Frequencies<br>Grade 8 Math ExpLvl, ActLvl | | | | Expected Frequencies<br>Grade 8 Math ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 502 | 144 | 75 | 721 | 497.482 | 144.432 | 78.75 |
| 0 | 1523 | 362 | 224 | 2109 | 1452.942 | 413.766 | 230.72 |
| 1 | 962 | 361 | 175 | 1498 | 1029.34 | 287.717 | 163.275 |
| TOTAL | 2987 | 867 | 474 | 4328 | | | |

$\chi^2(4)=30.32$; $p = .0001$
V = .0592

| | Observed Frequencies<br>ALG1 ExpLvl, ActLvl | | | | Expected Frequencies<br>ALG1 ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 747 | 192 | 104 | 1043 | 806.013 | 157.44 | 94.744 |
| 0 | 1663 | 576 | 284 | 2523 | 1559.894 | 654.336 | 306.152 |
| 1 | 1042 | 463 | 205 | 1710 | 963.85 | 555.137 | 199.67 |
| TOTAL | 3452 | 1231 | 593 | 5276 | | | |

$\chi^2(4)=35.21$; $p = .0001$
V = .0578

## Appendix C

*Observed and Expected Discrepancies between ExpLvl and AntGrd for Students Classified by Ethnicity in Math*

| Observed Frequencies Grade 3 Math ExpLvl, AntGrd | | | | | Expected Frequencies Grade 3 Math ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 500 | 199 | 158 | 857 | 554.5 | 151.638 | 129.876 |
| 0 | 1888 | 495 | 428 | 2811 | 1850.24 | 519.75 | 439.128 |
| 1 | 476 | 112 | 94 | 682 | 444.108 | 125.44 | 107.16 |
| TOTAL | 2864 | 806 | 680 | 4350 | | | |

$\chi^2(4)=28.91$; $p = .0001$
V = .0576

| Observed Frequencies Grade 4 Math ExpLvl, AntGrd | | | | | Expected Frequencies Grade 4 Math ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 571 | 208 | 149 | 928 | 615.538 | 154.336 | 144.977 |
| 0 | 1927 | 478 | 407 | 2812 | 1861.482 | 513.372 | 429.792 |
| 1 | 420 | 122 | 119 | 661 | 430.08 | 135.908 | 90.321 |
| TOTAL | 2918 | 808 | 675 | 4401 | | | |

$\chi^2(4)=29.71$; $p = .0001$
V = .0581

| Observed Frequencies Grade 5 Math ExpLvl, AntGrd | | | | | Expected Frequencies Grade 5 Math ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 514 | 201 | 113 | 828 | 553.578 | 149.946 | 110.74 |
| 0 | 1989 | 524 | 388 | 2901 | 1945.242 | 553.344 | 399.64 |
| 1 | 449 | 117 | 106 | 672 | 444.51 | 131.157 | 93.386 |
| TOTAL | 2952 | 842 | 607 | 4401 | | | |

$\chi^2(4)=20.86$; $p = .0003$
V = .0487

| Observed Frequencies Grade 6 Math ExpLvl, AntGrd | | | | | Expected Frequencies Grade 6 Math ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 652 | 219 | 133 | 1004 | 676.776 | 193.596 | 129.675 |
| 0 | 1730 | 480 | 329 | 2539 | 1714.43 | 495.84 | 328.013 |
| 1 | 493 | 133 | 88 | 714 | 482.154 | 139.251 | 92.048 |
| TOTAL | 2875 | 832 | 550 | 4257 | | | |

$\chi^2(4)=5.14$; $p = .2732$
V = .0246

## Appendix C - Continued

| | Observed Frequencies<br>Grade 7 Math ExpLvl, AntGrd | | | |
|---|---|---|---|---|
| | White | Black | Hisp | TOTAL |
| -1 | 766 | 187 | 88 | 1041 |
| 0 | 1696 | 512 | 304 | 2512 |
| 1 | 481 | 125 | 116 | 722 |
| TOTAL | 2943 | 824 | 508 | 4275 |

$\chi^2(4)=29.57$; $p = .0001$
V = .0588

| | Expected Frequencies<br>Grade 7 Math ExpLvl, AntGrd | |
|---|---|---|
| White | Black | Hisp |
| 713.146 | 199.716 | 113.432 |
| 1728.224 | 482.816 | 298.528 |
| 496.392 | 137.75 | 75.168 |

| | Observed Frequencies<br>Grade 8 Math ExpLvl, AntGrd | | | |
|---|---|---|---|---|
| | White | Black | Hisp | TOTAL |
| -1 | 640 | 235 | 105 | 980 |
| 0 | 1886 | 480 | 261 | 2627 |
| 1 | 478 | 158 | 110 | 746 |
| TOTAL | 3004 | 873 | 476 | 4353 |

$\chi^2(4)=32.04$; $p = .0001$
V = .0607

| | Expected Frequencies<br>Grade 8 Math ExpLvl, AntGrd | |
|---|---|---|
| White | Black | Hisp |
| 674.56 | 188.94 | 107.1 |
| 1810.56 | 522.72 | 284.751 |
| 512.416 | 149.152 | 71.72 |

| | Observed Frequencies<br>ALG1 Math ExpLvl, AntGrd | | | |
|---|---|---|---|---|
| | White | Black | Hisp | TOTAL |
| -1 | 265 | 70 | 42 | 377 |
| 0 | 1790 | 516 | 272 | 2578 |
| 1 | 1425 | 647 | 295 | 2367 |
| TOTAL | 3480 | 1233 | 609 | 5322 |

$\chi^2(4)=53.78$; $p = .0001$
V = .0711

| | Expected Frequencies<br>ALG1 ExpLvl, AntGrd | |
|---|---|---|
| White | Black | Hisp |
| 245.125 | 83.93 | 43.092 |
| 1679.02 | 586.176 | 293.216 |
| 1537.575 | 530.54 | 268.745 |

## Appendix D

*Observed and Expected Discrepancies between AntGrd and ActLvl for Students Classified by Ethnicity in Reading*

| | Observed Frequencies | | | | Expected Frequencies | | |
|---|---|---|---|---|---|---|---|
| | Grade 3 Read AntGrd, ActLvl | | | | Grade 3 Read AntGrd, ActLvl | | |
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 518 | 144 | 122 | 784 | 517.482 | 144.72 | 121.878 |
| 0 | 1389 | 327 | 250 | 1966 | 1291.77 | 359.373 | 295.5 |
| 1 | 947 | 327 | 300 | 1574 | 1030.336 | 285.798 | 232.2 |
| TOTAL | 2854 | 798 | 672 | 4324 | | | |

$\chi^2(4)=45.33$; $p = .0001$
V = .0724

| | Observed Frequencies | | | | Expected Frequencies | | |
|---|---|---|---|---|---|---|---|
| | Grade 4 Read AntGrd, ActLvl | | | | Grade 4 Read AntGrd, ActLvl | | |
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 557 | 172 | 121 | 850 | 565.912 | 154.112 | 127.897 |
| 0 | 1298 | 288 | 261 | 1847 | 1225.312 | 331.2 | 277.704 |
| 1 | 1046 | 339 | 276 | 1661 | 1102.484 | 300.693 | 248.124 |
| TOTAL | 2901 | 799 | 658 | 4358 | | | |

$\chi^2(4)=24.42$; $p = .0001$
V = .0529

| | Observed Frequencies | | | | Expected Frequencies | | |
|---|---|---|---|---|---|---|---|
| | Grade 5 Read AntGrd, ActLvl | | | | Grade 5 Read AntGrd, ActLvl | | |
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 399 | 160 | 77 | 636 | 426.531 | 108 | 85.393 |
| 0 | 1291 | 354 | 214 | 1859 | 1252.27 | 355.062 | 246.742 |
| 1 | 1236 | 310 | 299 | 1845 | 1243.416 | 274.35 | 241.592 |
| TOTAL | 2926 | 824 | 590 | 4340 | | | |

$\chi^2(4)=36.87$; $p = .0001$
V = .0652

| | Observed Frequencies | | | | Expected Frequencies | | |
|---|---|---|---|---|---|---|---|
| | Grade 6 Read AntGrd, ActLvl | | | | Grade 6 Read AntGrd, ActLvl | | |
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 586 | 119 | 87 | 792 | 532.088 | 147.56 | 98.745 |
| 0 | 1197 | 304 | 198 | 1699 | 1150.317 | 329.232 | 214.434 |
| 1 | 1061 | 396 | 248 | 1705 | 1148.002 | 320.76 | 212.04 |
| TOTAL | 2844 | 819 | 533 | 4196 | | | |

$\chi^2(4)=44.42$; $p = .0001$
V = .0728

## Appendix D - Continued

| | Observed Frequencies<br>Grade 7 Read AntGrd, ActLvl | | | |
|---|---|---|---|---|
| | White | Black | Hisp | TOTAL |
| -1 | 709 | 163 | 89 | 961 |
| 0 | 1102 | 313 | 197 | 1612 |
| 1 | 1112 | 338 | 213 | 1663 |
| TOTAL | 2923 | 814 | 499 | 4236 |

$\chi^2(4)=14.94$; $p = .0048$
V = .042

| Expected Frequencies<br>Grade 7 Read AntGrd, ActLvl | | |
|---|---|---|
| White | Black | Hisp |
| 660.079 | 182.071 | 108.046 |
| 1111.918 | 309.87 | 189.711 |
| 1146.472 | 318.396 | 194.469 |

| | Observed Frequencies<br>Grade 8 Read AntGrd, ActLvl | | | |
|---|---|---|---|---|
| | White | Black | Hisp | TOTAL |
| -1 | 506 | 127 | 57 | 690 |
| 0 | 1048 | 274 | 157 | 1479 |
| 1 | 1431 | 465 | 255 | 2151 |
| TOTAL | 2985 | 866 | 469 | 4320 |

$\chi^2(4)=16.13$; $p = .0028$
V = .0432

| Expected Frequencies<br>Grade 8 Read AntGrd, ActLvl | | |
|---|---|---|
| White | Black | Hisp |
| 475.134 | 137.414 | 70.623 |
| 1021.8 | 294.824 | 160.454 |
| 1483.947 | 428.73 | 231.54 |

| | Observed Frequencies<br>ENG1 Read AntGrd, ActLvl | | | |
|---|---|---|---|---|
| | White | Black | Hisp | TOTAL |
| -1 | 1376 | 388 | 192 | 1956 |
| 0 | 1209 | 391 | 176 | 1776 |
| 1 | 679 | 252 | 121 | 1052 |
| TOTAL | 3264 | 1031 | 489 | 4784 |

$\chi^2(4)=11.22$; $p = .0242$
V = .0342

| Expected Frequencies<br>ENG1 Read AntGrd, ActLvl | | |
|---|---|---|
| White | Black | Hisp |
| 1333.344 | 419.04 | 199.68 |
| 1211.418 | 382.398 | 181.28 |
| 715.666 | 223.776 | 105.875 |

# Appendix E

*Observed and Expected Discrepancies between ExpLvl and ActLvl for Students Classified by Ethnicity in Read*

| | Observed Frequencies Grade 3 Read ExpLvl, ActLvl | | | | Expected Frequencies Grade 3 Read ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 429 | 84 | 75 | 588 | 384.384 | 102.648 | 88.575 |
| 0 | 1563 | 382 | 305 | 2250 | 1481.724 | 410.65 | 344.65 |
| 1 | 858 | 326 | 292 | 1476 | 960.96 | 259.822 | 213.16 |
| TOTAL | 2850 | 792 | 672 | 4314 | | | |

$\chi^2(4)=66.73$; $p = .0001$
V = .0879

| | Observed Frequencies Grade 4 Read ExpLvl, ActLvl | | | | Expected Frequencies Grade 4 Read ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 420 | 109 | 68 | 597 | 396.06 | 109.545 | 84.728 |
| 0 | 1495 | 389 | 330 | 2214 | 1472.575 | 405.727 | 334.62 |
| 1 | 982 | 301 | 260 | 1543 | 1024.226 | 282.037 | 230.1 |
| TOTAL | 2897 | 799 | 658 | 4354 | | | |

$\chi^2(4)=14.06$; $p = .0071$
V = .0402

| | Observed Frequencies Grade 5 Read ExpLvl, ActLvl | | | | Expected Frequencies Grade 5 Read ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 262 | 88 | 47 | 397 | 267.502 | 73.216 | 53.11 |
| 0 | 1457 | 389 | 247 | 2093 | 1410.376 | 396.78 | 279.604 |
| 1 | 1206 | 346 | 296 | 1848 | 1244.592 | 350.498 | 243.312 |
| TOTAL | 2925 | 823 | 590 | 4338 | | | |

$\chi^2(4)=19.08$; $p = .0008$
V = .0469

| | Observed Frequencies Grade 6 Read ExpLvl, ActLvl | | | | Expected Frequencies Grade 6 Read ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 538 | 100 | 83 | 721 | 483.662 | 128.9 | 90.802 |
| 0 | 1467 | 384 | 243 | 2094 | 1418.589 | 406.656 | 263.898 |
| 1 | 840 | 334 | 207 | 1381 | 926.52 | 253.506 | 169.74 |
| TOTAL | 2845 | 818 | 533 | 4196 | | | |

$\chi^2(4)=53.63$; $p = .0001$
V = .0799

## Appendix E - Continued

| | Observed Frequencies Grade 7 Read ExpLvl, ActLvl | | | |
|---|---|---|---|---|
| | White | Black | Hisp | TOTAL |
| -1 | 459 | 92 | 65 | 616 |
| 0 | 1410 | 360 | 222 | 1992 |
| 1 | 1061 | 362 | 210 | 1633 |
| TOTAL | 2930 | 814 | 497 | 4241 |

$\chi^2(4)=25.2$; $p = .0001$
V = .0545

| Expected Frequencies Grade 7 Read ExpLvl, ActLvl | | |
|---|---|---|
| White | Black | Hisp |
| 422.739 | 112.424 | 71.5 |
| 1374.75 | 380.88 | 232.878 |
| 1124.66 | 305.89 | 189.63 |

| | Observed Frequencies Grade 8 Read ExpLvl, ActLvl | | | |
|---|---|---|---|---|
| | White | Black | Hisp | TOTAL |
| -1 | 342 | 94 | 52 | 488 |
| 0 | 1392 | 328 | 227 | 1947 |
| 1 | 1252 | 444 | 190 | 1886 |
| TOTAL | 2986 | 866 | 469 | 4321 |

$\chi^2(4)=27.53$; $p = .0001$
V = .0564

| Expected Frequencies Grade 8 Read ExpLvl, ActLvl | | |
|---|---|---|
| White | Black | Hisp |
| 337.212 | 97.666 | 52.936 |
| 1343.28 | 380.152 | 210.202 |
| 1300.828 | 366.3 | 203.68 |

| | Observed Frequencies ENG1 Read ExpLvl, ActLvl | | | |
|---|---|---|---|---|
| | White | Black | Hisp | TOTAL |
| -1 | 912 | 257 | 118 | 1287 |
| 0 | 1571 | 470 | 226 | 2267 |
| 1 | 748 | 281 | 136 | 1165 |
| TOTAL | 3231 | 1008 | 480 | 4719 |

$\chi^2(4)=14.09$; $p = .0007$
V = .0386

| Expected Frequencies ENG1 Read ExpLvl, ActLvl | | |
|---|---|---|
| White | Black | Hisp |
| 880.08 | 273.705 | 129.682 |
| 1552.148 | 483.63 | 226.452 |
| 794.376 | 244.751 | 115.872 |

# Appendix F

*Observed and Expected Discrepancies between ExpLvl and AntGrd for Students Classified by Ethnicity in Read*

| | Observed Frequencies Grade 3 Read ExpLvl, AntGrd | | | | Expected Frequencies Grade 3 Read ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 442 | 164 | 127 | 733 | 479.57 | 129.724 | 113.157 |
| 0 | 1923 | 502 | 445 | 2870 | 1890.309 | 529.61 | 448.115 |
| 1 | 494 | 137 | 106 | 737 | 485.108 | 136.315 | 114.374 |
| TOTAL | 2859 | 803 | 678 | 4340 | | | |

$\chi^2(4)=13.8$; $p = .008$
V = .0399

| | Observed Frequencies Grade 4 Read ExpLvl, AntGrd | | | | Expected Frequencies Grade 4 Read ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 469 | 160 | 139 | 768 | 511.679 | 132.96 | 112.312 |
| 0 | 1994 | 476 | 408 | 2878 | 1906.264 | 521.22 | 438.6 |
| 1 | 451 | 166 | 126 | 743 | 484.374 | 133.796 | 114.912 |
| TOTAL | 2914 | 802 | 673 | 4389 | | | |

$\chi^2(4)=32.3$; $p = .0001$
V = .0607

| | Observed Frequencies Grade 5 Read ExpLvl, AntGrd | | | | Expected Frequencies Grade 5 Read ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 452 | 191 | 102 | 745 | 494.94 | 127.206 | 102.204 |
| 0 | 2013 | 527 | 401 | 2941 | 1970.727 | 560.201 | 406.213 |
| 1 | 482 | 122 | 104 | 708 | 469.468 | 138.47 | 97.552 |
| TOTAL | 2947 | 840 | 607 | 4394 | | | |

$\chi^2(4)=28.12$; $p = .0001$
V = .0566

| | Observed Frequencies Grade 6 Read ExpLvl, AntGrd | | | | Expected Frequencies Grade 6 Read ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 473 | 152 | 83 | 708 | 478.676 | 137.104 | 90.221 |
| 0 | 1602 | 465 | 312 | 2379 | 1608.408 | 465.465 | 305.448 |
| 1 | 793 | 213 | 150 | 1156 | 781.105 | 225.354 | 148.5 |
| TOTAL | 2868 | 830 | 545 | 4243 | | | |

$\chi^2(4)=3.18$; $p = .5282$
V = .0194

## Appendix F - Continued

Observed Frequencies

Grade 7 Read ExpLvl, AntGrd

| | White | Black | Hisp | TOTAL |
|---|---|---|---|---|
| -1 | 764 | 231 | 113 | 1108 |
| 0 | 1538 | 429 | 262 | 2229 |
| 1 | 641 | 164 | 127 | 932 |
| TOTAL | 2943 | 824 | 502 | 4269 |

$\chi^2(4)=7.84$; $p$ = .0976

V = .0303

Expected Frequencies

Grade 7 Read ExpLvl, AntGrd

| White | Black | Hisp |
|---|---|---|
| 764 | 212.52 | 128.029 |
| 1536.462 | 430.287 | 262 |
| 642.282 | 178.432 | 106.807 |

Observed Frequencies

Grade 8 Read ExpLvl, AntGrd

| | White | Black | Hisp | TOTAL |
|---|---|---|---|---|
| -1 | 606 | 189 | 94 | 889 |
| 0 | 1679 | 486 | 215 | 2380 |
| 1 | 719 | 200 | 167 | 1086 |
| TOTAL | 3004 | 875 | 476 | 4355 |

$\chi^2(4)=31.98$; $p$ = .0001

V = .0606

Expected Frequencies

Grade 8 Read ExpLvl, AntGrd

| White | Black | Hisp |
|---|---|---|
| 613.272 | 178.038 | 97.102 |
| 1640.383 | 478.224 | 252.41 |
| 747.76 | 216.6 | 99.031 |

Observed Frequencies

ENG1 Read ExpLvl, AntGrd

| | White | Black | Hisp | TOTAL |
|---|---|---|---|---|
| -1 | 894 | 241 | 125 | 1260 |
| 0 | 1643 | 557 | 237 | 2437 |
| 1 | 706 | 216 | 121 | 1043 |
| TOTAL | 3243 | 1014 | 483 | 4740 |

$\chi^2(4)=9.95$; $p$ = .0413

V = .0324

Expected Frequencies

ENG1 Read ExpLvl, AntGrd

| White | Black | Hisp |
|---|---|---|
| 904.728 | 227.022 | 129.125 |
| 1605.211 | 548.088 | 278.238 |
| 734.24 | 233.928 | 71.753 |

## Appendix G

*Observed and Expected Discrepancies between AntGrd and ActLvl for Students Classified by Ethnicity in Science*

| | Observed Frequencies Grade 5 Science AntGrd, ActLvl | | | | Expected Frequencies Grade 5 Science AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 602 | 202 | 132 | 936 | 629.09 | 174.932 | 127.512 |
| 0 | 1317 | 295 | 207 | 1819 | 1218.225 | 338.66 | 241.362 |
| 1 | 1017 | 333 | 256 | 1606 | 1077.003 | 303.363 | 212.992 |
| TOTAL | 2936 | 830 | 595 | 4361 | | | |

$\chi^2(4)=38.44$; $p = .0001$

V = .0664

| | Observed Frequencies Grade 8 Science AntGrd, ActLvl | | | | Expected Frequencies Grade 8 Science AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 886 | 176 | 123 | 1185 | 812.462 | 221.584 | 129.15 |
| 0 | 1259 | 284 | 163 | 1706 | 1172.129 | 332.28 | 183.538 |
| 1 | 830 | 404 | 185 | 1419 | 956.99 | 234.32 | 149.295 |
| TOTAL | 2975 | 864 | 471 | 4310 | | | |

$\chi^2(4)=119.17$; $p = .0001$

V = .1176

| | Observed Frequencies BIO Science AntGrd, ActLvl | | | | Expected Frequencies BIO Science AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 767 | 261 | 104 | 1132 | 773.903 | 249.516 | 107.744 |
| 0 | 774 | 217 | 101 | 1092 | 745.362 | 238.7 | 103.929 |
| 1 | 533 | 192 | 84 | 809 | 552.721 | 177.792 | 76.44 |
| TOTAL | 2074 | 670 | 289 | 3033 | | | |

$\chi^2(4)=6.55$; $p = .1617$

V = .0329

## Appendix H

*Observed and Expected Discrepancies between ExpLvl and ActLvl for Students Classified by Ethnicity in Science*

| | Observed Frequencies Grade 5 Science ExpLvl, ActLvl | | | | Expected Frequencies Grade 5 Science ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 545 | 116 | 87 | 748 | 500.855 | 137.112 | 99.876 |
| 0 | 1541 | 409 | 265 | 2215 | 1491.688 | 419.634 | 297.86 |
| 1 | 850 | 301 | 243 | 1394 | 930.75 | 259.161 | 175.932 |
| TOTAL | 2936 | 826 | 595 | 4357 | | | |

$\chi^2(4)=44.92$; $p = .0001$
V = .0718

| | Observed Frequencies Grade 8 Science ExpLvl, ActLvl | | | | Expected Frequencies Grade 8 Science ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 748 | 137 | 111 | 996 | 682.924 | 179.744 | 108.558 |
| 0 | 1521 | 357 | 217 | 2095 | 1443.429 | 409.836 | 227.85 |
| 1 | 716 | 370 | 143 | 1229 | 828.412 | 183.15 | 133.419 |
| TOTAL | 2985 | 864 | 471 | 4320 | | | |

$\chi^2(4)=122.39$; $p = .0001$
V = .119

| | Observed Frequencies BIO Science ExpLvl, ActLvl | | | | Expected Frequencies BIO Science ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 459 | 106 | 49 | 614 | 416.313 | 129.108 | 56.987 |
| 0 | 1020 | 294 | 127 | 1441 | 984.3 | 316.05 | 136.525 |
| 1 | 595 | 269 | 113 | 977 | 660.45 | 202.288 | 88.931 |
| TOTAL | 2074 | 669 | 289 | 3032 | | | |

$\chi^2(4)=40.89$; $p = .0001$
V = .0821

## Appendix I

*Observed and Expected Discrepancies between ExpLvl and AntGrd for Students Classified by Ethnicity in Science*

| | Observed Frequencies Grade 5 Science ExpLvl, AntGrd | | | | Expected Frequencies Grade 5 Science ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 602 | 202 | 132 | 936 | 669.424 | 149.076 | 107.316 |
| 0 | 1317 | 295 | 207 | 1819 | 1298.562 | 306.8 | 209.07 |
| 1 | 1017 | 333 | 256 | 1606 | 952.929 | 373.959 | 291.584 |
| TOTAL | 2936 | 830 | 595 | 4361 | | | |

$\chi^2(4)=38.44$; $p = .0001$

V = .0664

| | Observed Frequencies Grade 8 Science ExpLvl, AntGrd | | | | Expected Frequencies Grade 8 Science ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 886 | 176 | 123 | 1185 | 859.42 | 184.624 | 135.054 |
| 0 | 1259 | 284 | 163 | 1706 | 1237.597 | 289.68 | 173.921 |
| 1 | 830 | 404 | 185 | 1419 | 893.91 | 360.772 | 131.535 |
| TOTAL | 2975 | 864 | 471 | 4310 | | | |

$\chi^2(4)=119.17$; $p = .0001$

V = .1176

| | Observed Frequencies BIO Science ExpLvl, AntGrd | | | | Expected Frequencies BIO Science ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|---|
| | White | Black | Hisp | TOTAL | White | Black | Hisp |
| -1 | 767 | 261 | 104 | 1132 | 747.825 | 270.135 | 114.296 |
| 0 | 774 | 217 | 101 | 1092 | 745.362 | 235.662 | 107.464 |
| 1 | 533 | 192 | 84 | 809 | 577.772 | 158.592 | 67.704 |
| TOTAL | 2074 | 670 | 289 | 3033 | | | |

$\chi^2(4)=6.55$; $p = .1617$

V = .0329

# Appendix J

*Observed and Expected Discrepancies between AntGrd and ActLvl for Students Classified by Gender in Math*

| | Observed Frequencies Grade 3 Math AntGrd, ActLvl | | | Expected Frequencies Grade 3 Math AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 695 | 541 | 1236 | 643.57 | 584.28 | 1227.85 |
| 0 | 1073 | 1003 | 2076 | 1069.781 | 1006.009 | 2075.79 |
| 1 | 567 | 632 | 1199 | 612.927 | 577.016 | 1189.943 |
| TOTAL | 2335 | 2176 | 4511 | | | |

$\chi^2(4)=19.49$; $p = .0001$
V = .0657

| | Observed Frequencies Grade 4 Math AntGrd, ActLvl | | | Expected Frequencies Grade 4 Math AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 685 | 533 | 1218 | 617.185 | 587.899 | 1205.084 |
| 0 | 1008 | 1002 | 2010 | 1021.104 | 987.972 | 2009.076 |
| 1 | 630 | 700 | 1330 | 676.62 | 646.1 | 1322.72 |
| TOTAL | 2323 | 2235 | 4558 | | | |

$\chi^2(4)=20.98$; $p = .0001$
V = .0678

| | Observed Frequencies Grade 5 Math AntGrd, ActLvl | | | Expected Frequencies Grade 5 Math AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 598 | 455 | 1053 | 535.808 | 504.14 | 1039.948 |
| 0 | 1528 | 1506 | 3034 | 1520.36 | 1513.53 | 3033.89 |
| 1 | 344 | 474 | 818 | 377.368 | 426.6 | 803.968 |
| TOTAL | 2470 | 2435 | 4905 | | | |

$\chi^2(4)=39.99$; $p = .0001$
V = .0903

| | Observed Frequencies Grade 6 Math AntGrd, ActLvl | | | Expected Frequencies Grade 6 Math AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 675 | 364 | 1039 | 491.4 | 467.012 | 958.412 |
| 0 | 958 | 906 | 1864 | 952.252 | 912.342 | 1864.594 |
| 1 | 615 | 883 | 1498 | 737.385 | 699.336 | 1436.721 |
| TOTAL | 2248 | 2153 | 4401 | | | |

$\chi^2(4)=140.5$; $p = .0001$
V = .1787

## Appendix J - Continued

Observed Frequencies
Grade 7 Math AntGrd, ActLvl

| | Male | Female | TOTAL |
|---|---|---|---|
| -1 | 619 | 362 | 981 |
| 0 | 906 | 933 | 1839 |
| 1 | 717 | 883 | 1600 |
| TOTAL | 2242 | 2178 | 4420 |

$\chi^2(4)=84.04$; $p = .0001$
V = .1379

Expected Frequencies
Grade 7 Math AntGrd, ActLvl

| Male | Female | Total |
|---|---|---|
| 469.202 | 452.862 | 922.064 |
| 931.368 | 905.943 | 1837.311 |
| 801.606 | 774.391 | 1575.997 |

Observed Frequencies
Grade 8 Math AntGrd, ActLvl

| | Male | Female | TOTAL |
|---|---|---|---|
| -1 | 634 | 437 | 1071 |
| 0 | 896 | 877 | 1773 |
| 1 | 717 | 920 | 1637 |
| TOTAL | 2247 | 2234 | 4481 |

$\chi^2(4)=61.58$; $p = .0001$
V = .1172

Expected Frequencies
Grade 8 Math AntGrd, ActLvl

| Male | Female | Total |
|---|---|---|
| 525.586 | 513.038 | 1038.624 |
| 878.976 | 893.663 | 1772.639 |
| 813.078 | 794.88 | 1607.958 |

Observed Frequencies
ALG1 Math AntGrd, ActLvl

| | Male | Female | TOTAL |
|---|---|---|---|
| -1 | 1337 | 978 | 2315 |
| 0 | 917 | 946 | 1863 |
| 1 | 537 | 739 | 1276 |
| TOTAL | 2791 | 2663 | 5454 |

$\chi^2(4)=85.14$; $p = .0001$
V = .1249

Expected Frequencies
ALG1 Math AntGrd, ActLvl

| Male | Female | Total |
|---|---|---|
| 1171.212 | 1108.074 | 2279.286 |
| 946.344 | 912.89 | 1859.234 |
| 633.123 | 597.112 | 1230.235 |

## Appendix K

*Observed and Expected Discrepancies between ExpLvl and ActLvl for Students Classified by Gender in Math*

| | Observed Frequencies Grade 3 Math ExpLvl, ActLvl | | | Expected Frequencies Grade 3 Math ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 514 | 415 | 929 | 478.534 | 445.71 | 924.244 |
| 0 | 1288 | 1212 | 2500 | 1294.44 | 1205.94 | 2500.38 |
| 1 | 531 | 547 | 1078 | 556.488 | 518.556 | 1075.044 |
| TOTAL | 2333 | 2174 | 4507 | | | |

$\chi^2(4)=7.5$; $p = .0235$
V = .0408

| | Observed Frequencies Grade 4 Math ExpLvl, ActLvl | | | Expected Frequencies Grade 4 Math ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 477 | 125 | 602 | 464.121 | 128.625 | 592.746 |
| 0 | 1640 | 394 | 2034 | 1603.92 | 403.062 | 2006.982 |
| 1 | 786 | 288 | 1074 | 831.588 | 270.72 | 1102.308 |
| TOTAL | 2903 | 807 | 3710 | | | |

$\chi^2(4)=23.3$; $p = .0001$
V = .0792

| | Observed Frequencies Grade 5 Math ExpLvl, ActLvl | | | Expected Frequencies Grade 5 Math ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 408 | 352 | 760 | 383.52 | 373.824 | 757.344 |
| 0 | 1290 | 1248 | 2538 | 1284.84 | 1252.992 | 2537.832 |
| 1 | 608 | 647 | 1255 | 634.144 | 617.885 | 1252.029 |
| TOTAL | 2306 | 2247 | 4553 | | | |

$\chi^2(4)=5.27$; $p = .0717$
V = .034

| | Observed Frequencies Grade 6 Math ExpLvl, ActLvl | | | Expected Frequencies Grade 6 Math ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 354 | 241 | 595 | 293.82 | 283.657 | 577.477 |
| 0 | 1199 | 1076 | 2275 | 1159.433 | 1112.584 | 2272.017 |
| 1 | 693 | 836 | 1529 | 773.388 | 734.844 | 1508.232 |
| TOTAL | 2246 | 2153 | 4399 | | | |

$\chi^2(4)=39.54$; $p = .0001$
V = .0948

## Appendix K - Continued

Observed Frequencies
Grade 7 Math ExpLvl, ActLvl

| | Male | Female | TOTAL |
|---|---|---|---|
| -1 | 344 | 266 | 610 |
| 0 | 1150 | 1112 | 2262 |
| 1 | 744 | 802 | 1546 |
| TOTAL | 2238 | 2180 | 4418 |

$\chi^2(4)=12.03$; $p = .0024$
V = .0522

Expected Frequencies
Grade 7 Math ExpLvl, ActLvl

| Male | Female | Total |
|---|---|---|
| 306.504 | 295.792 | 602.296 |
| 1144.25 | 1117.56 | 2261.81 |
| 781.944 | 759.494 | 1541.438 |

Observed Frequencies
Grade 8 Math ExpLvl, ActLvl

| | Male | Female | TOTAL |
|---|---|---|---|
| -1 | 417 | 320 | 737 |
| 0 | 1082 | 1109 | 2191 |
| 1 | 747 | 802 | 1549 |
| TOTAL | 2246 | 2231 | 4477 |

$\chi^2(4)=15$; $p = .0006$
V = .0579

Expected Frequencies
Grade 8 Math ExpLvl, ActLvl

| Male | Female | Total |
|---|---|---|
| 362.79 | 362.24 | 725.03 |
| 1091.738 | 1099.019 | 2190.757 |
| 784.35 | 761.098 | 1545.448 |

Observed Frequencies
ALG1 Math ExpLvl, ActLvl

| | Male | Female | TOTAL |
|---|---|---|---|
| -1 | 589 | 488 | 1077 |
| 0 | 1371 | 1240 | 2611 |
| 1 | 844 | 940 | 1784 |
| TOTAL | 2804 | 2668 | 5472 |

$\chi^2(4)=17.84$; $p = .0001$
V = .0571

Expected Frequencies
ALG1 Math ExpLvl, ActLvl

| Male | Female | Total |
|---|---|---|
| 552.482 | 520.208 | 1072.69 |
| 1336.725 | 1273.48 | 2610.205 |
| 906.456 | 864.8 | 1771.256 |

## Appendix L

*Observed and Expected Discrepancies between ExpLvl and AntGrd for Students Classified by Gender in Math*

| | Observed Frequencies Grade 3 Math ExpLvl, AntGrd | | | Expected Frequencies Grade 3 Math ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 498 | 395 | 893 | 460.65 | 426.995 | 887.645 |
| 0 | 1494 | 1427 | 2921 | 1514.916 | 1405.595 | 2920.511 |
| 1 | 356 | 357 | 713 | 369.172 | 342.72 | 711.892 |
| TOTAL | 2348 | 2179 | 4527 | | | |

$\chi^2(4)=7.12$; $p = .0284$
V = .0397

| | Observed Frequencies Grade 4 Math ExpLvl, AntGrd | | | Expected Frequencies Grade 4 Math ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 267 | 299 | 566 | 285.423 | 277.472 | 562.895 |
| 0 | 1454 | 1470 | 2924 | 1491.804 | 1430.31 | 2922.114 |
| 1 | 612 | 473 | 1085 | 548.352 | 524.084 | 1072.436 |
| TOTAL | 2333 | 2242 | 4575 | | | |

$\chi^2(4)=17.9$; $p = .0001$
V = .0626

| | Observed Frequencies Grade 5 Math ExpLvl, AntGrd | | | Expected Frequencies Grade 5 Math ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 256 | 348 | 604 | 296.96 | 290.232 | 587.192 |
| 0 | 1528 | 1506 | 3034 | 1532.584 | 1501.482 | 3034.066 |
| 1 | 541 | 411 | 952 | 480.408 | 458.676 | 939.084 |
| TOTAL | 2325 | 2265 | 4590 | | | |

$\chi^2(4)=31.15$; $p = .0001$
V = .0824

| | Observed Frequencies Grade 6 Math ExpLvl, AntGrd | | | Expected Frequencies Grade 6 Math ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 257 | 367 | 624 | 307.115 | 292.132 | 599.247 |
| 0 | 1283 | 1366 | 2649 | 1352.282 | 1288.138 | 2640.42 |
| 1 | 730 | 429 | 1159 | 564.29 | 531.102 | 1095.392 |
| TOTAL | 2270 | 2162 | 4432 | | | |

$\chi^2(4)=97.59$; $p = .0001$
V = .1484

## Appendix L - Continued

| | Observed Frequencies<br>Grade 7 Math ExpLvl, AntGrd | | |
|---|---|---|---|
| | Male | Female | TOTAL |
| -1 | 385 | 578 | 963 |
| 0 | 1168 | 1157 | 2325 |
| 1 | 700 | 452 | 1152 |
| TOTAL | 2253 | 2187 | 4440 |

$\chi^2(4)$=91.16; $p$ = .0001
V = .1433

| Expected Frequencies<br>Grade 7 Math ExpLvl, AntGrd | | |
|---|---|---|
| Male | Female | Total |
| 437.745 | 496.502 | 934.247 |
| 1226.4 | 1097.993 | 2324.393 |
| 549.5 | 552.344 | 1101.844 |

| | Observed Frequencies<br>Grade 8 Math ExpLvl, AntGrd | | |
|---|---|---|---|
| | Male | Female | TOTAL |
| -1 | 497 | 630 | 1127 |
| 0 | 1216 | 1250 | 2466 |
| 1 | 554 | 357 | 911 |
| TOTAL | 2267 | 2237 | 4504 |

$\chi^2(4)$=58.57; $p$ = .0001
V = .114

| Expected Frequencies<br>Grade 8 Math ExpLvl, AntGrd | | |
|---|---|---|
| Male | Female | Total |
| 555.646 | 553.77 | 1109.416 |
| 1246.4 | 1217.5 | 2463.9 |
| 467.022 | 414.477 | 881.499 |

| | Observed Frequencies<br>ALG1 Math ExpLvl, AntGrd | | |
|---|---|---|---|
| | Male | Female | TOTAL |
| -1 | 169 | 225 | 394 |
| 0 | 1251 | 1426 | 2677 |
| 1 | 1412 | 1040 | 2452 |
| TOTAL | 2832 | 2691 | 5523 |

$\chi^2(4)$=72.28; $p$ = .0001
V = .1144

| Expected Frequencies<br>ALG1 Math ExpLvl, AntGrd | | |
|---|---|---|
| Male | Female | Total |
| 173.225 | 219.375 | 392.6 |
| 1291.032 | 1378.942 | 2669.974 |
| 1351.284 | 1085.76 | 2437.044 |

## Appendix M

*Observed and Expected Discrepancies between AntGrd and ActLvl for Students Classified by Gender in Reading*

| | Observed Frequencies Grade 3 Reading AntGrd, ActLvl | | | Expected Frequencies Grade 3 Reading AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 478 | 343 | 821 | 424.464 | 384.503 | 808.967 |
| 0 | 1053 | 971 | 2024 | 1050.894 | 972.942 | 2023.836 |
| 1 | 794 | 859 | 1653 | 840.846 | 804.883 | 1645.729 |
| TOTAL | 2325 | 2173 | 4498 | | | |

$\chi^2(4)=22.97$; $p = .0001$
V = .0715

| | Observed Frequencies Grade 4 Reading AntGrd, ActLvl | | | Expected Frequencies Grade 4 Reading AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 503 | 373 | 876 | 442.64 | 419.252 | 861.892 |
| 0 | 985 | 939 | 1924 | 978.105 | 945.573 | 1923.678 |
| 1 | 819 | 913 | 1732 | 876.33 | 847.264 | 1723.594 |
| TOTAL | 2307 | 2225 | 4532 | | | |

$\chi^2(4)=24.02$; $p = .0001$
V = .728

| | Observed Frequencies Grade 5 Reading AntGrd, ActLvl | | | Expected Frequencies Grade 5 Reading AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 387 | 277 | 664 | 328.176 | 320.489 | 648.665 |
| 0 | 1042 | 891 | 1933 | 975.312 | 949.806 | 1925.118 |
| 1 | 858 | 1072 | 1930 | 958.386 | 942.288 | 1900.674 |
| TOTAL | 2287 | 2240 | 4527 | | | |

$\chi^2(4)=53.26$; $p = .0001$
V = .1085

| | Observed Frequencies Grade 6 Reading AntGrd, ActLvl | | | Expected Frequencies Grade 6 Reading AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 551 | 261 | 812 | 373.027 | 348.696 | 721.723 |
| 0 | 904 | 871 | 1775 | 904 | 871 | 1775 |
| 1 | 773 | 1010 | 1783 | 889.723 | 852.44 | 1742.163 |
| TOTAL | 2228 | 2142 | 4370 | | | |

$\chi^2(4)=134.05$; $p = .0001$
V = .1751

## Appendix M - Continued

Observed Frequencies
Grade 7 Reading AntGrd, ActLvl

| | Male | Female | TOTAL |
|---|---|---|---|
| -1 | 630 | 368 | 998 |
| 0 | 834 | 844 | 1678 |
| 1 | 766 | 966 | 1732 |
| TOTAL | 2230 | 2178 | 4408 |

$\chi^2(4)=91.34$; $p$ = .0001
V = .1439

Expected Frequencies
Grade 7 Reading AntGrd, ActLvl

| Male | Female | Total |
|---|---|---|
| 481.32 | 457.792 | 939.112 |
| 844.008 | 833.028 | 1677.036 |
| 860.984 | 842.352 | 1703.336 |

Observed Frequencies
Grade 8 Reading AntGrd, ActLvl

| | Male | Female | TOTAL |
|---|---|---|---|
| -1 | 451 | 256 | 707 |
| 0 | 784 | 736 | 1520 |
| 1 | 1007 | 1236 | 2243 |
| TOTAL | 2242 | 2228 | 4470 |

$\chi^2(4)=78.64$; $p$ = .0001
V = .1326

Expected Frequencies
Grade 8 Reading AntGrd, ActLvl

| Male | Female | Total |
|---|---|---|
| 333.289 | 324.096 | 657.385 |
| 762.048 | 756.608 | 1518.656 |
| 1110.721 | 1106.22 | 2216.941 |

Observed Frequencies
ENG1 Reading AntGrd, ActLvl

| | Male | Female | TOTAL |
|---|---|---|---|
| -1 | 1163 | 849 | 2012 |
| 0 | 865 | 959 | 1824 |
| 1 | 497 | 625 | 1122 |
| TOTAL | 2525 | 2433 | 4958 |

$\chi^2(4)=66.77$; $p$ = .0001
V = .116

Expected Frequencies
ENG1 Reading AntGrd, ActLvl

| Male | Female | Total |
|---|---|---|
| 1010.647 | 966.162 | 1976.809 |
| 931.605 | 881.321 | 1812.926 |
| 553.658 | 550 | 1103.658 |

# Appendix N

*Observed and Expected Discrepancies between ExpLvl and ActLvl for Students Classified by Gender in Read*

| | Observed Frequencies Grade 3 Reading ExpLvl, ActLvl | | | Expected Frequencies Grade 3 Reading ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 514 | 415 | 929 | 473.394 | 445.71 | 919.104 |
| 0 | 1288 | 1212 | 2500 | 1294.44 | 1205.94 | 2500.38 |
| 1 | 531 | 547 | 1078 | 556.488 | 518.556 | 1075.044 |
| TOTAL | 2333 | 2174 | 4507 | | | |

$\chi^2(4)=7.5$; $p = .0235$
V = .0408

| | Observed Frequencies Grade 4 Reading ExpLvl, ActLvl | | | Expected Frequencies Grade 4 Reading ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 477 | 125 | 602 | 470.799 | 130.625 | 601.424 |
| 0 | 1640 | 394 | 2034 | 1590.8 | 436.946 | 2027.746 |
| 1 | 786 | 288 | 1074 | 837.09 | 220.896 | 1057.986 |
| TOTAL | 2903 | 807 | 3710 | | | |

$\chi^2(4)=23.3$; $p = .0001$
V = .0792

| | Observed Frequencies Grade 5 Reading ExpLvl, ActLvl | | | Expected Frequencies Grade 5 Reading ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 408 | 352 | 760 | 4080.816 | 352.704 | 4433.52 |
| 0 | 1290 | 1248 | 2538 | 1226.79 | 1310.4 | 2537.19 |
| 1 | 608 | 647 | 1255 | 641.44 | 610.121 | 1251.561 |
| TOTAL | 2306 | 2247 | 4553 | | | |

$\chi^2(4)=5.27$; $p = .0717$
V = .034

| | Observed Frequencies Grade 6 Reading ExpLvl, ActLvl | | | Expected Frequencies Grade 6 Reading ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 354 | 241 | 595 | 312.936 | 269.92 | 582.856 |
| 0 | 1199 | 1076 | 2275 | 1182.214 | 1091.064 | 2273.278 |
| 1 | 693 | 836 | 1529 | 749.133 | 765.776 | 1514.909 |
| TOTAL | 2246 | 2153 | 4399 | | | |

$\chi^2(4)=39.54$; $p = .0001$
V = .0948

## Appendix N - Continued

| | Observed Frequencies Grade 7 Reading ExpLvl, ActLvl | | |
|---|---|---|---|
| | Male | Female | TOTAL |
| -1 | 355 | 287 | 642 |
| 0 | 1035 | 1025 | 2060 |
| 1 | 844 | 867 | 1711 |
| TOTAL | 2234 | 2179 | 4413 |

$\chi^2(4)=6.88$; $p = .0321$
V = .0395

| Expected Frequencies Grade 7 Reading ExpLvl, ActLvl | | |
|---|---|---|
| Male | Female | Total |
| 322.34 | 314.265 | 636.605 |
| 1043.28 | 1016.8 | 2060.08 |
| 8461.944 | 844.458 | 9306.402 |

| | Observed Frequencies Grade 8 Reading ExpLvl, ActLvl | | |
|---|---|---|---|
| | Male | Female | TOTAL |
| -1 | 289 | 209 | 498 |
| 0 | 1018 | 988 | 2006 |
| 1 | 936 | 1030 | 1966 |
| TOTAL | 2243 | 2227 | 4470 |

$\chi^2(4)=15.7$; $p = .0001$
V = .063

| Expected Frequencies Grade 8 Reading ExpLvl, ActLvl | | |
|---|---|---|
| Male | Female | Total |
| 245.939 | 240.768 | 486.707 |
| 1010.874 | 995.904 | 2006.778 |
| 979.056 | 981.59 | 1960.646 |

| | Observed Frequencies ENG1 Reading ExpLvl, ActLvl | | |
|---|---|---|---|
| | Male | Female | TOTAL |
| -1 | 684 | 636 | 1320 |
| 0 | 1178 | 1180 | 2358 |
| 1 | 625 | 587 | 1212 |
| TOTAL | 2487 | 2403 | 4890 |

$\chi^2(4)=1.5$; $p = .4724$
V = .0175

| Expected Frequencies ENG1 Reading ExpLvl, ActLvl | | |
|---|---|---|
| Male | Female | Total |
| 671.004 | 648.72 | 1319.724 |
| 1199.204 | 1158.76 | 2357.964 |
| 616.25 | 595.218 | 1211.468 |

## Appendix O

*Observed and Expected Discrepancies between ExpLvl and AntGrd for Students Classified by Gender in Read*

| | Observed Frequencies Grade 3 Reading ExpLvl, AntGrd | | | Expected Frequencies Grade 3 Reading ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 350 | 390 | 740 | 379.4 | 354.9 | 734.3 |
| 0 | 1514 | 1475 | 2989 | 1545.794 | 1441.075 | 2986.869 |
| 1 | 477 | 311 | 788 | 402.111 | 363.87 | 765.981 |
| TOTAL | 2341 | 2176 | 4517 | | | |

$\chi^2(4)=31.66$; $p = .0001$
V = .0837

| | Observed Frequencies Grade 4 Reading ExpLvl, AntGrd | | | Expected Frequencies Grade 4 Reading ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 323 | 407 | 730 | 367.251 | 348.799 | 716.05 |
| 0 | 1506 | 1488 | 2994 | 1521.06 | 1471.632 | 2992.692 |
| 1 | 498 | 342 | 840 | 421.308 | 396.72 | 818.028 |
| TOTAL | 2327 | 2237 | 4564 | | | |

$\chi^2(4)=36.99$; $p = .0001$
V = .009

| | Observed Frequencies Grade 5 Reading ExpLvl, AntGrd | | | Expected Frequencies Grade 5 Reading ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 288 | 377 | 665 | 330.336 | 319.319 | 649.655 |
| 0 | 1519 | 1555 | 3074 | 1552.418 | 1519.235 | 3071.653 |
| 1 | 512 | 330 | 842 | 411.648 | 396.99 | 808.638 |
| TOTAL | 2319 | 2262 | 4581 | | | |

$\chi^2(4)=50.97$; $p = .0001$
V = .1055

| | Observed Frequencies Grade 6 Reading ExpLvl, AntGrd | | | Expected Frequencies Grade 6 Reading ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 509 | 656 | 1165 | 579.242 | 561.536 | 1140.778 |
| 0 | 1245 | 1230 | 2475 | 1272.39 | 1201.71 | 2474.1 |
| 1 | 506 | 272 | 778 | 366.85 | 350.064 | 716.914 |
| TOTAL | 2260 | 2158 | 4418 | | | |

$\chi^2(4)=86.71$; $p = .0001$
V = .1401

## Appendix O - Continued

Observed Frequencies
Grade 7 Reading ExpLvl, AntGrd

| | Male | Female | TOTAL |
|---|---|---|---|
| -1 | 385 | 578 | 963 |
| 0 | 1168 | 1157 | 2325 |
| 1 | 700 | 452 | 1152 |
| TOTAL | 2253 | 2187 | 4440 |

$\chi^2(4)=91.16$; $p = .0001$
V = .1433

Expected Frequencies
Graded 7 Reading ExpLvl, AntGrd

| Male | Female | Total |
|---|---|---|
| 464.31 | 454.886 | 919.196 |
| 1178.512 | 1146.587 | 2325.099 |
| 566.3 | 541.496 | 1107.796 |

Observed Frequencies
Grade 8 Reading ExpLvl, AntGrd

| | Male | Female | TOTAL |
|---|---|---|---|
| -1 | 497 | 630 | 1127 |
| 0 | 1216 | 1250 | 2466 |
| 1 | 554 | 357 | 911 |
| TOTAL | 2267 | 2237 | 4504 |

$\chi^2(4)=58.57$; $p = .0001$
V = .114

Expected Frequencies
Graded 8 Reading ExpLvl, AntGrd

| Male | Female | Total |
|---|---|---|
| 559.125 | 549.36 | 1108.485 |
| 1236.672 | 1227.5 | 2464.172 |
| 443.2 | 430.185 | 873.385 |

Observed Frequencies
ENG1 Reading ExpLvl, AntGrd

| | Male | Female | TOTAL |
|---|---|---|---|
| -1 | 301 | 430 | 731 |
| 0 | 1194 | 1333 | 2527 |
| 1 | 1003 | 651 | 1654 |
| TOTAL | 2498 | 2414 | 4912 |

$\chi^2(4)=103.92$; $p = .0001$
V = .1455

Expected Frequencies
ENG1 Reading ExpLvl, AntGrd

| Male | Female | Total |
|---|---|---|
| 356.384 | 347.01 | 703.394 |
| 1279.968 | 1233.025 | 2512.993 |
| 815.439 | 778.596 | 1594.035 |

## Appendix P

*Observed and Expected Discrepancies between AntGrd and ActLvl for Students Classified by Gender in Science*

| | Observed Frequencies Grade 5 Reading AntGrd, ActLvl | | | Expected Frequencies Grade 5 Reading AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 612 | 361 | 973 | 462.06 | 452.694 | 914.754 |
| 0 | 1006 | 883 | 1889 | 955.7 | 928.916 | 1884.616 |
| 1 | 687 | 1000 | 1687 | 823.713 | 794 | 1617.713 |
| TOTAL | 2305 | 2244 | 4549 | | | |

$\chi^2(4)=130.04$; $p = .0001$
V = .1691

| | Observed Frequencies Grade 8 Reading AntGrd, ActLvl | | | Expected Frequencies Grade 8 Reading AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 796 | 419 | 1215 | 555.608 | 547.633 | 1103.241 |
| 0 | 873 | 892 | 1765 | 878.238 | 886.648 | 1764.886 |
| 1 | 562 | 916 | 1478 | 700.252 | 687.916 | 1388.168 |
| TOTAL | 2231 | 2227 | 4458 | | | |

$\chi^2(4)=201.97$; $p = .0001$
V = .2128

| | Observed Frequencies BIO Reading AntGrd, ActLvl | | | Expected Frequencies BIO Reading AntGrd, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 666 | 503 | 1169 | 562.104 | 577.947 | 1140.051 |
| 0 | 562 | 581 | 1143 | 560.314 | 582.743 | 1143.057 |
| 1 | 318 | 524 | 842 | 388.596 | 412.388 | 800.984 |
| TOTAL | 1546 | 1608 | 3154 | | | |

$\chi^2(4)=72.25$; $p = .0001$
V = .1514

# Appendix Q

*Observed and Expected Discrepancies between ExpLvl and ActLvl for Students Classified by Gender in Science*

| | Observed Frequencies Grade 5 Reading ExpLvl, ActLvl | | | Expected Frequencies Grade 5 Reading ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 468 | 305 | 773 | 381.888 | 363.255 | 745.143 |
| 0 | 1165 | 1143 | 2308 | 1160.34 | 1148.715 | 2309.055 |
| 1 | 671 | 793 | 1464 | 742.126 | 705.77 | 1447.896 |
| TOTAL | 2304 | 2241 | 4545 | | | |

$\chi^2(4)=43.88$; $p = .0001$
V = .0983

| | Observed Frequencies Grade 8 Reading ExpLvl, ActLvl | | | Expected Frequencies Grade 8 Reading ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 610 | 407 | 1017 | 495.32 | 485.144 | 980.464 |
| 0 | 1082 | 1087 | 2169 | 1084.164 | 1084.826 | 2168.99 |
| 1 | 549 | 734 | 1283 | 631.35 | 621.698 | 1253.048 |
| TOTAL | 2241 | 2228 | 4469 | | | |

$\chi^2(4)=67.17$; $p = .0001$
V = .1226

| | Observed Frequencies BIO Reading ExpLvl, ActLvl | | | Expected Frequencies BIO Reading ExpLvl, ActLvl | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 365 | 272 | 637 | 305.14 | 314.704 | 619.844 |
| 0 | 730 | 764 | 1494 | 735.11 | 759.416 | 1494.526 |
| 1 | 452 | 571 | 1023 | 494.036 | 520.181 | 1014.217 |
| TOTAL | 1547 | 1607 | 3154 | | | |

$\chi^2(4)=27.06$; $p = .0001$
V = .0926

## Appendix R

*Observed and Expected Discrepancies between ExpLvl and AntGrd for Students Classified by Gender in Science*

| | Observed Frequencies Grade 5 Reading ExpLvl, AntGrd | | | Expected Frequencies Grade 5 Reading ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 393 | 570 | 963 | 468.456 | 456 | 924.456 |
| 0 | 1395 | 1333 | 2728 | 1385.235 | 1323.669 | 2708.904 |
| 1 | 526 | 345 | 871 | 425.534 | 413.655 | 839.189 |
| TOTAL | 2314 | 2248 | 4562 | | | |

$\chi^2(4)=70.61$; $p = .0001$
V = .1244

| | Observed Frequencies Grade 8 Reading ExpLvl, AntGrd | | | Expected Frequencies Grade 8 Reading ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 368 | 563 | 931 | 447.488 | 439.14 | 886.628 |
| 0 | 1254 | 1350 | 2604 | 1300.398 | 1300.05 | 2600.448 |
| 1 | 630 | 324 | 954 | 434.07 | 426.708 | 860.778 |
| TOTAL | 2252 | 2237 | 4489 | | | |

$\chi^2(4)=142.48$; $p = .0001$
V = .1782

| | Observed Frequencies BIO Reading ExpLvl, AntGrd | | | Expected Frequencies BIO Reading ExpLvl, AntGrd | | |
|---|---|---|---|---|---|---|
| | Male | Female | TOTAL | Male | Female | Total |
| -1 | 147 | 247 | 394 | 182.868 | 188.955 | 371.823 |
| 0 | 776 | 876 | 1652 | 810.144 | 838.332 | 1648.476 |
| 1 | 647 | 502 | 1149 | 551.244 | 573.284 | 1124.528 |
| TOTAL | 1570 | 1625 | 3195 | | | |

$\chi^2(4)=48.8$; $p = .0001$
V = .1236

## Vita

Jed Cockrell was born in Statesville, North Carolina, to Eric and Linda Cockrell. He graduated from the University of North Carolina at Charlotte in 2002, where he earned a Bachelor of Arts degree in History. The following autumn, he enrolled in Gardner Webb University, and in 2004 he earned his certification in Elementary Education. Dr. Cockrell enrolled in Appalachian State University in the fall of 2006, where he would eventually graduate with degrees in Curriculum and Instruction and School Administration. In May of 2016, Dr. Cockrell graduated with his Ed.D. in Educational Leadership.

Dr. Cockrell works as a public educator and advocates for causes supporting public education. He resides in North Wilkesboro, NC with his wife, Jenny, and daughter, Mallory.