

# What's in a Grade? The Multidimensional Nature of What Teacher Assigned Grades Assess in High School<sup>1</sup>

Alex J. Bowers<sup>2</sup>  
 Teachers College, Columbia University<sup>3</sup>  
 Bowers@tc.edu

## ABSTRACT

Historically, teacher assigned grades have been seen as unreliable subjective measures of academic knowledge, since grades and standardized tests have traditionally correlated at about the 0.5 to 0.6 level, and thus explain about 25-35% of each other. However, emerging literature indicates that grades may be a multi-dimensional assessment of both student academic knowledge and a student's ability to negotiate the social processes of schooling, such as behavior, participation, and effort. This study analyzed the high school transcript component of the Education Longitudinal Study of 2002 (ELS:2002) using multi-dimensional scaling (MDS) to describe the relationships between core subject grades, non-core subject grades and standardized test scores in mathematics and reading. The results indicate that when accounting for the academic knowledge component assessed through standardized tests, teacher assigned grades may be a useful assessment of a student's ability at the non-cognitive aspects of school. Implications for practice, research, and policy are discussed.

**Keywords:** Grades (scholastic), grading, academic achievement, standardized tests, multivariate analysis, student evaluation, correlation, high school, core curriculum, non-core curriculum, art, physical education, social studies, grade point average, GPA.

## INTRODUCTION

Currently in K-12 schools in the United States, two co-existing assessment systems are in place, standardized test scores that are reported to school administrators, the community, and state and federal policymakers, and teacher assigned grades, that are reported to students, parents and teachers (Farr, 2000). While increasing attention has been paid to preparing for, proctoring, and analyzing results from standardized testing since the rise of the accountability movement over the past 20 years, the vast industry of assessing and assigning course grades has continued unabated

(Bowers, 2009). Yet, while specific grades or course failures are routinely used by teachers and administrators for remedial or advanced student class assignments or special education identification (Hallinan, 1992), or as an at-risk indicator (Allensworth & Easton, 2005), lost in much of the discussion around accountability and recent calls for data driven decision making (3DM) has been the issue of using teacher assigned grades more systematically to inform decisions in schools and for policymaking (Bowers, 2009). This is reasonable given the policy debates around standardized testing, and that school administrators privilege standardized test score information over other forms of assessment (Guskey, 2007). However, as data already collected daily for the vast majority of students in U.S. schools in each subject and grade level, teacher assigned grades are a potential rich resource of information on student performance in each grade-level and subject area (Bowers, 2009, 2010a). Yet grades have historically been underused as useful data because they have been viewed as overly subjective and not directly tied to academic knowledge.

Much of the research that has asked teachers what they assign a grade for has found that teachers award grades for not only academic knowledge but also a multitude of student behaviors and classroom performance issues, termed "hodgepodge" and "kitchen sink grading" (Brookhart, 1991; Cizek, Fitzgerald, & Rachor, 1995-1996; Cross & Frary, 1999; Frary, Cross, & Weber, 1993; McMillan, 2001; Randall & Englehard, 2009). These include academic performance, classroom participation, effort, behavior, attendance, improvement, and turning in homework, among others. Because of these non-cognitive behavioral aspects of grades, psychometricians and assessment researchers have historically maligned teacher assigned grades as non-useful, due to the perception that grades are unreliable measures of academic knowledge, and they have urged teachers to work to align their grading practices to specific standards and procedures (Brookhart, 2004; Heritage & Yeagley, 2005; Marzano, 2000). Yet, teachers are resistant to aligning grades purely to standardized academic knowledge outcomes (Cizek, 2000). As noted by Cross & Frary (1999):

We must ask, "if hodgepodge grading is so deplorable, why haven't students, parents and administrators, or the general public called for reform?" It may well be that they share a common understanding that grades often do, in fact, represent a hodgepodge of attitude, effort, conduct, growth, and achievement and that is what they expect and endorse. (p.70)

Thus, while administrators and policymakers focus on test scores, and researchers urge teachers to reform grading practices, teachers

<sup>1</sup> This document is a pre-print of this manuscript, published in 2011 in the journal *Educational Research & Evaluation*. Recommended citation:

Bowers, A.J. (2011) What's in a Grade? The Multidimensional Nature of What Teacher Assigned Grades Assess in High School. *Educational Research & Evaluation*, 17(3), 141-159. doi: [10.1080/13803611.2011.597112](https://doi.org/10.1080/13803611.2011.597112)

<sup>2</sup> 525 W. 120<sup>th</sup> Street, New York, New York 10027.

ORCID: 0000-0002-5140-6428, ResearcherID: C-1557-2013

<sup>3</sup> Formerly at The University of Texas at San Antonio at the time of publication.

Note: This document was last updated August 12, 2013.

and schools continue to collect grades in much the same way as they have in the past and report them to students and parents, despite consistent pressure to align grades to academic standards.

While little research has focused on the association of standardized test scores to overall school and life outcomes, teacher assigned grades are well known as predictive of overall schooling outcomes. Over the past forty years, teacher assigned grades have been consistently shown to be strongly associated with the likelihood that a student will drop out or graduate from high school (Allensworth & Easton, 2007; Balfanz, Herzog, & MacIver, 2007; Barclay & Doll, 2001; Barrington & Hendricks, 1989; Bowers, 2009, 2010a, 2010b; Bowers & Sprott, 2012; Ekstrom, Goertz, Pollack, & Rock, 1986; Ensminger & Slusarcick, 1992; Fitzsimmons, Cheever, Leonard, & Macunovich, 1969; Lloyd, 1974, 1978; Rumberger & Palardy, 2005). In addition, high school grades have been shown to be strong predictors of college entrance exams and freshman college performance (Burton & Ramist, 2001; Linn, 1982; Woodruff & Ziomek, 2004; Zwick & Greif Green, 2007). Indeed, these associations have remained consistently strong despite the reputation of grades as poor and subjective measures of academic knowledge. Thus, it appears that while grades have historically been criticized as poor measures of academic knowledge, the association between grades and overall school outcomes indicates that there may be a useful aspect of grades. This is an apparent paradox. Teacher assigned grades are not consistently used for systematic decision making by administrators, central offices and state and federal policymaking due to their subjectivity and incorporation of non-academic knowledge, yet grades have been shown to be strongly associated with overall schooling outcomes. Why? This issue of what it is that grades may be assessing, and how it might be useful, is the focus of the present study.

Historically, for K-12 schools in the U.S., teacher assigned grades have consistently correlated with standardized test scores at about the 0.5 to 0.6 levels (Brennan, Kim, Wenz-Gross, & Siperstein, 2001; Linn, 1982, 2000; Willingham, Pollack, & Lewis, 2002; Woodruff & Ziomek, 2004). As just one example, Brennan et al. (2001) studied a sample of 736 eighth grade students from six Boston schools, and compared grades in mathematics, English and science to scores on the Massachusetts MCAS standardized achievement tests and found the correlations to be 0.54 for mathematics, 0.59 for English, and 0.54 for science. Thus, if grades and test scores correlate at about 0.5 to 0.6, then examining the squared correlation (R-squared) indicates that these two assessments explain about 25-35% of each other. If one accepts that standardized test scores are generally a fairly pure assessment of basic academic knowledge, then by extension about 25-35% of teacher assigned grades is related to academic knowledge. Combining the overall schooling outcomes literature with this relatively small overlap between grades and test scores, if grades are known to be a weak assessment of academic knowledge, but appear to be a strong indicator of overall school outcomes, the question remains as to what the remaining 65-75% of teacher assigned grades actually assesses, and is this assessment useful beyond teachers, parents and students? Much of the “hodgepodge” grading literature cited above indicates that other than academic knowledge, grades represent an assessment of a student’s ability at the process of schooling such as behavior, attendance and participation. This idea that grades may be an assessment of two constructs, academic knowledge and school social processes, has been understood for some time (Cross & Frary, 1999; Parsons, 1959). However, to date grades have been underused as

informative data in schools (Bowers, 2009, 2010a), especially with the rise of state-wide accountability systems over the past 20 years.

Recently, this dualistic nature of grades has been explored as useful data as a multi-dimensional assessment that assesses both academic knowledge and non-academic behaviors. In their study of grade 12 students, Willingham, Pollack and Lewis (2002) analyzed the relationship between grades and a standardized test composite in a subset of 8,454 students from the NELS:88 dataset. Interestingly, when they accounted for school grading variations, student characteristics and teacher perceptions of student attendance, class behavior, and motivation using multiple regression, they were able to account for 81% of the variance between grades and the standardized test. They attributed much of this result to the idea that grades are an assessment of both cognitive ability as well as what they termed engagement, or “conative skills” as an assessment of how hard a student endeavored or tried in the classroom. They note that:

A grade represents each teacher’s judgment as to how well a student has fulfilled the implicit local contract between teacher and student....Grades especially reflect engagement, a strength particularly associated with conative skills. Educators recognize that conative skills like volition, habits of inquiry, effort and self-regulation are, in themselves, important goals of schooling. (Willingham, et al., 2002) (p.28-29).

Thus, to these authors, the remaining 75% of grades that is not explained by test scores may be in part associated with a teacher’s assessment of student non-cognitive skills and behaviors.

More recently, using a latent variable framework and structural equation modeling, in a study from Sweden, subject specific grades and standardized assessments in mathematics, English and Swedish were examined for an entire national cohort of students (Klapp Lekholm & Cliffordson, 2008, 2009). Grades were shown to be a multi-dimensional assessment, assessing academic knowledge in a similar fashion to the standardized tests, but also assessing a “common grade dimension” that was not classroom or school dependent and appeared to be a large and consistent component of the assessments that did not assess academic knowledge, but rather some “other” dimension of the schooling process. In a similar study, using multi-dimensional scaling (MDS), 195 student subject specific grades from two small graduating cohorts of U.S. students in 2006 were compared to standardized test scores at the high school level, including the ACT college entrance exam and state-wide grade 10 standardized tests in mathematics, English and science (Bowers, 2009). As with Klapp Lekholm and Cliffordson, grades were multi-dimensional, consisting of an academic knowledge component that aligned with both the ACT and the state standardized tests, and a non-academic component that was distal from standardized tests. Using the past literature on grades, a theory was constructed to argue for a shift in the construct represented by grades. Thus, due to this mounting evidence from the literature, it appears that teacher assigned grades are a multi-dimensional assessment of *both* academic knowledge and a non-academic or non-cognitive component that represents a student’s ability to negotiate the social processes of school, including behavior, participation, attendance, and study skills, among others – termed a Success at School Factor (SSF) (Bowers, 2009). However, these two studies are limited either to a non-U.S. context or to small intact samples. Thus, this evidence of the multi-dimensional nature of grades is tentative and must be replicated and extended in the U.S. context using a large national sample.

### *Framework of the Study*

Nevertheless, from a policy perspective, much of the literature has shown that standardized test scores have only a weak association to overall student schooling outcomes, such as graduating or dropping out, while teacher assigned grades are consistently shown to be a powerful predictor of student schooling success or failure (Bowers, 2009; Bowers & Sprott, 2012; Rumberger & Palardy, 2005). Why? What is it that grades assess, and is it useful not only for students and parents, but for teachers and school leaders as they use student-level data to create a conversation around data driven decision making, as well as for policymakers in the current era of accountability? The majority of the literature to date has left this question unaddressed.

Thus, this study focuses on the intersection of grades as informative as multi-dimensional assessments of both student academic knowledge and non-cognitive behaviors. The aim is to provide a means for school leaders and administrators to use grades more effectively in helping identify student-level data already collected across all schools to help drive decisions on how to direct the limited resources of schools and school districts to the students most in need (Bowers, 2008). This study addresses this aim in the following way. Since the literature on the multi-dimensional nature of grades currently is sparse and limited either to the non-U.S. context, dated samples, or on specific small intact student samples, the goal is to replicate and extend these findings to confirm the multi-dimensional nature of grades comparing teacher assigned grades to standardized assessments in a large national sample. Specifically, the research questions addressed here are:

- To what extent are teacher assigned grades and standardized test scores related?
- To what extent is the relationship multi-dimensional across core and non-core grades versus standardized test scores?
- To what extent does the relationship between test scores and grades vary across different high school course subjects?

## **METHODS**

### *Dataset*

This study is a secondary data analysis of the Education Longitudinal Study of 2002 (ELS:2002) with first follow-up data from 2004 (NCES, n.d.). ELS:2002 is a nationally representative longitudinal sample of about 15,400 students who were in grade 10 in 2002 across about 750 different schools in the U.S. In the base 2001-02 school year, students were asked to fill out an extensive questionnaire, and were tested in multiple subjects. In the first follow-up in the spring of 2004, about 15,000 students were again surveyed and tested, and high school transcripts were collected for a majority of the students (Bozick et al., 2006; Ingles et al., 2004; Ingles et al., 2007). High school transcript, achievement and grading data are historically difficult to collect (Kuncel, Crede, & Thomas, 2005), thus a large national sample of student records throughout high school, such as ELS:2002, is a rare and important database to analyze for overall patterns of student achievement data.

Due to the need for complete data across each variable in the subsequent analysis discussed below, this study analyzed subsets of the ELS:2002 dataset. Namely, this included a sample of

$n=4,520$  students with standardized test and grades data for each of the eight semesters from grade 9 through 12, and  $n=5,230$  for students with standardized test and grades data across each of the main course subject areas included in the analysis discussed below. Due to issues of data confidentiality, some variables have been rounded to the nearest ten.

### *Variables*

Two main types of variables were analyzed from ELS:2002, standardized test scores and teacher assigned grades. For standardized test scores, the ELS:2002 grade 10 standardized test scores in reading and mathematics were included in the analysis. For the base-year when students were in grade 10, students were given an assessment battery in both reading and mathematics that contained questions equated with and adapted from the past nationally normed assessments NAEP, PISA and NELS:1992 (Ingles, et al., 2004). These questions assessed academic knowledge in reading through testing for the reproduction of literary detail, comprehension, inference and evaluation. The questions assessed academic knowledge in mathematics through testing for skills, understanding, comprehension and practical application and problem solving in arithmetic, algebra, geometry, data and probability (Ingles, et al., 2004). To replicate and extend the past research comparing grades to standardized assessments that used national or state-normed standardized assessment scores (Bowers, 2009; Klapp Lekholm & Cliffordson, 2008, 2009), for the analysis discussed below, the standardized T-scores were used as nationally norm-referenced assessments of academic knowledge in both reading and mathematics, ELS:2002 variables BYTXRSTD (reading) and BYTXMSTD (mathematics).

The second type of variable included in the analysis were grades across subjects in high school. Teacher assigned grades in each subject for each student at each grade level from grades 9 through 12 were included in the analysis. Historically, much of the literature on grades has relied either on teacher perception of how they grade students (Cizek, et al., 1995-1996; Cross & Frary, 1999; McMillan, 2003) or on student's self reported grades (Woodruff & Ziomek, 2004), which recently have been shown to be problematic as useful data (Kuncel, et al., 2005). Thus, this study relied on the grades reported on school transcripts within each semester at each grade level 9-12 for each student. Student transcripts for all courses taken from grades 9 through 12 were collected for ELS:2002, which included the subject, grade level, semester and the grade awarded (Bozick, et al., 2006; Ingles, et al., 2004). This resulted in the extraction of about 638,000 individual course grades across 1,600 different course subject titles. The ELS:2002 variable FICGRADE reported letter grades on a standard A+ to F scale. These grades were converted to a numeric scale in which A=4.0, A- = 3.666, B+ = 3.333, B = 3.0, B- = 2.666, C+ = 2.333, C = 2.0, C- = 1.666, D+ = 1.333, D = 1.0, D- = 0.666, E or F = 0. Each course was coded within ELS with the FICCSSC variable according to the Classification of Secondary School Courses (CSSC) (for a review of CSSC please see Bozick et al., 2006). To decrease issues with missing data across different course subjects, subjects were aggregated to twelve top-level course domains for either core subject courses (mathematics, English, science, social studies) or non-core subject courses (foreign/non-English languages, family and consumer science, military science, physical education (PE) health and recreation, religion, art, general skills, and labor preparation).

**Table 1:** Descriptives for variables included in the multi-dimensional scaling models

<i>Variable</i>	<i>Min</i>	<i>Max</i>	<i>Mean</i>	<i>SD</i>
<i>Grade-Level GPA for Core Courses n=4520</i>				
9 Semester 1 (9S1)	0	4	2.686	0.884
9 Semester 2 (9S2)	0	4	2.652	0.925
10 Semester 1 (10S1)	0	4	2.630	0.900
10 Semester 2 (10S2)	0	4	2.622	0.924
11 Semester 1 (11S1)	0	4	2.623	0.891
11 Semester 2 (11S2)	0	4	2.607	0.933
12 Semester 1 (12S1)	0	4	2.699	0.902
12 Semester 2 (12S2)	0	4	2.693	0.901
<i>Grade-Level GPA for Non-Core Courses n=4520</i>				
9 Semester 1 (9S1)	0	4	3.191	0.737
9 Semester 2 (9S2)	0	4	3.168	0.784
10 Semester 1 (10S1)	0	4	3.167	0.781
10 Semester 2 (10S2)	0	4	3.139	0.822
11 Semester 1 (11S1)	0	4	3.207	0.780
11 Semester 2 (11S2)	0	4	3.177	0.831
12 Semester 1 (12S1)	0	4	3.342	0.748
12 Semester 2 (12S2)	0	4	3.268	0.822
<i>Subject-Level GPA for grades 9-12 n=5230</i>				
Mathematics	0	4	2.382	0.951
English Language Arts	0	4	2.657	0.881
Science	0	4	2.528	0.921
Social Studies	0	4	2.709	0.893
Foreign/Non-English Languages	0	4	2.592	1.069
Physical Education Health & Recreation (PE)	0	4	3.319	0.762
Art	0	4	3.275	0.853
Mathematics standardized test (Grade 10)	19.38	84.85	51.988	9.278
Reading standardized test (Grade 10)	24.29	78.76	51.733	9.318

Because of the wide variety of course selection, to increase the total number of non-missing data, two sets of non-cumulative grade point averages (GPA) were calculated. First, mean non-cumulative GPA for each of the eight high school semesters from grade 9 through 12 was calculated for each student for both core and non-core courses to examine the relationship between core courses, non-core courses and standardized tests. In an effort to maximize the grade and subject level information available for the distance measure calculations discussed below and to ensure coverage across each semester and grade level, the analysis focused only on students who had attended high schools on either a semester or quarter term system and who had valid transcript data for each of the eight semesters. This resulted in  $n=4,520$  students (see Table 1).

Second, to examine cross-subject relationships between grades and test scores, the mean grade in each of the twelve course domains for each student was calculated. However, since complete data was required in the subsequent analysis across each of these course subject means, the subjects family and consumer science, military science, religion, general skills and labor prep were removed from the course subject means analysis due to either low enrollment (each of the courses family and consumer science, military science, religion and general skills represented less than 3% of the course enrollment), or missing data in comparison with the included subjects. This resulted in  $n=5230$  students (see Table 1).

#### *Analytic Technique: Multi-dimensional Scaling (MDS)*

To examine the underlying structure of the relationship between teacher assigned grades and standardized test scores, multi-dimensional scaling was used (MDS). Similar to factor analysis, MDS is a form of data reduction, estimating the latent variables underlying the relationships between a set of variables (Borg & Goren, 1997; Mead, 1992; Subkoviak, 1975). MDS is a descriptive statistic that relies on a set of distance measures to calculate a similarity or dissimilarity matrix, and then graphically “projects” the most significant latent variable dimensions into a two or three dimensional plot in which variables are separated across the dimensions based on their similarity or dissimilarity. The classic example of MDS is the recreation of a map of the relative distances of U.S. cities. If one takes a straight-line lower-triangular distance matrix of the distances between the major cities in the U.S., as is found on most driving atlases, the distance between any two cities can be found by looking up one city as a row, and the other as a column, and then finding the intersecting cell with the distance between the two cities. If MDS is applied to such a distance matrix, the relative location of all of the cities is recreated on a two-dimensional plot, in which Boston and New York are in the upper right, Miami in the lower right, Seattle in the upper left, Los Angeles in the lower left, and cities like Detroit, St. Louis and Houston all aligned where they would be expected (Borg & Goren, 1997; Davidson, 1983). MDS has been used in many research fields, including global health and medicine, higher education, education finance and policy, and K-12 education

(Bowers, 2009; Enserink, 2008; Glass, 1967; Pincus & Schmelkin, 2003; Russell et al., 2008; Smith et al., 2004; Song & Miskel, 2007). Here, the data were analyzed using ordinal MDS with Euclidean distance as the distance measure. To prevent overweighting in the distance matrix by the differences in scales between the standardized test scores and grades, all variables were z-scored. The MDS analysis was calculated using the ALSCAL procedure in SPSS/PASW version 18.0, and subsequent distance relationships were plotted in either two or three dimensions.

## RESULTS

To date, a pervasive question in the literature on the difference between teacher assigned grades and standardized testing is what do grades actually assess and can that assessment be useful for school and policy decision making and continuous improvement? The goal of this study is to delineate the commonalities and differences between grades and standardized tests in high school. Despite the strong association between grades and overall schooling outcomes, researchers continuously urge teachers to align their grading practices with standardized curricula and grading practices to change grades into a more accurate assessment of academic knowledge. However, teachers are resistant to these recommendations as they report that they award grades for a multitude of behaviors, from academic knowledge, to attendance, participation, effort and overall positive classroom behaviors. Since historically grades and test scores have correlated around 0.5 to 0.6, and thus test scores explain about 25-35% of grades, the central question of this study is to describe what the other approximately 65-75% of grades may be assessing, and is it an assessment that is useful. The recent research reviewed above indicates that this other dimension of grades may be aligned with the hodgepodge grading literature, in that teachers are awarding grades as a multidimensional assessment, one that encompasses both academic knowledge and student affective classroom behaviors. To test this hypothesis using a large national sample of high school students from the U.S., I first examine the overall correlation from the ELS:2002 dataset between tests and grades, then outline a theoretical framework for the difference between grades and tests. I next describe two multi-dimensional scaling (MDS) analyses, one that compares core and non-core GPA across all eight semesters of high school to standardized tests in mathematics and reading, and a second that compares grades in each main subject area in high school to the same standardized tests. I then turn to a discussion of the usefulness of grades as a multi-dimensional assessment in schools and for policy.

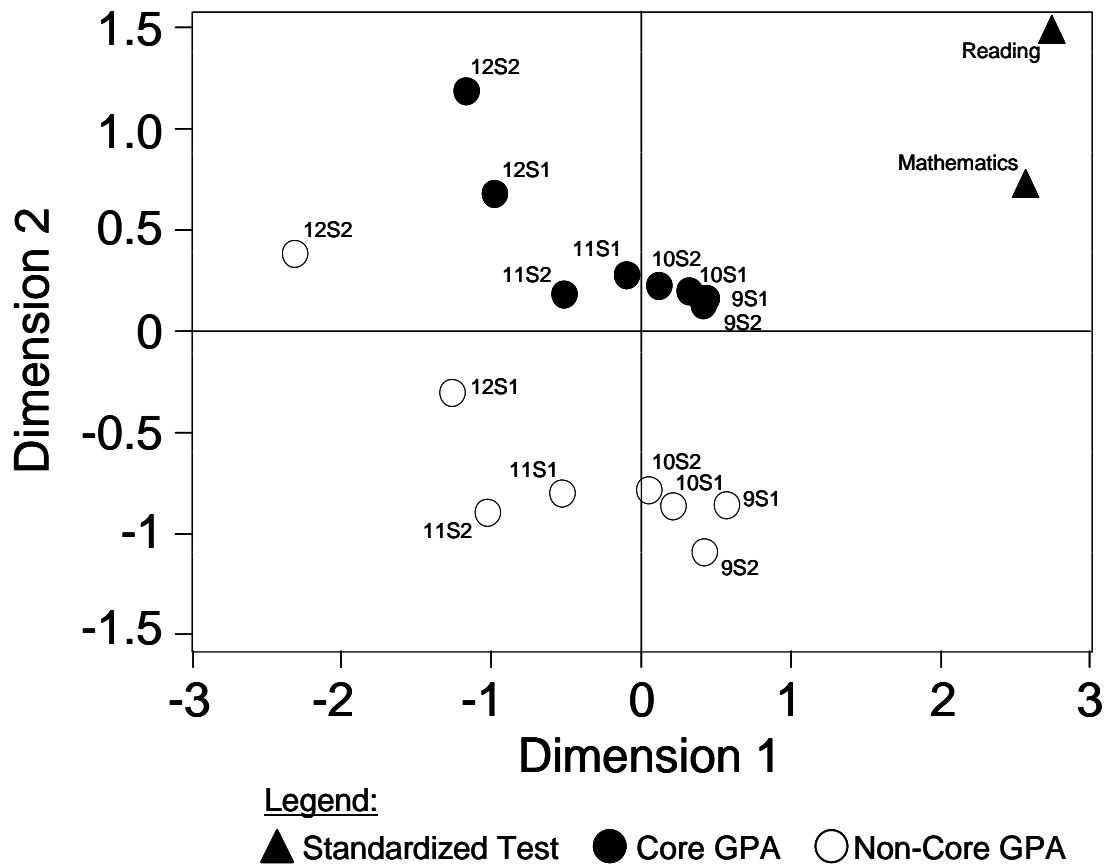
### *The Relationship between Grades and Tests*

Overall, for the ELS:2002 sample, the Pearson product moment correlation comparing overall high school GPA to the ELS:2002 mathematics and reading composite test score was  $r(14,520)=0.572$ . This replicates the long history of studies comparing teacher assigned grades and standardized tests reviewed above (Brennan, et al., 2001; Linn, 1982, 2000; Willingham, et al., 2002). For this dataset, the overall variance explained between grades and standardized tests is 0.327 (R-squared), such that these two assessments appear to be explaining about 32.7% of each other, replicating and extending the past research to the ELS:2002 dataset. Thus, this correlation suggests that one dimension of teacher assigned grades for these students is highly related to core academic knowledge, such as the knowledge assessed by the standardized mathematics and reading assessments. However, this appears to be only a minor proportion of what grades are assessing. What does the remaining component of grades assess?

### *Framing a Hypothesis for the Difference between Grades and Tests*

To address this question, this study explores the difference between core grades, non-core grades and test scores. As discussed above in the recent literature on the “common grades dimension” (Klapp Lekholm & Cliffordson, 2008, 2009), which was also termed “conative” engagement and effort behaviors (Willingham, et al., 2002) or a success at school factor (SSF) (Bowers, 2009), multiple studies have described this “other” dimension of grades. Controlling for academic knowledge in the past models, this literature indicates that the hodgepodge grading literature may be correct, in that in large measure, teachers award grades for effort, engagement, and participation. These studies indicate however that rather than an inconsistent, unreliable or subjective assessment, it may be that teacher assigned grades are more systematic than previously suggested. Due to this, Bowers (2009) argued for a re-conception of the construct validity of grades, in that grades should be considered a multi-dimensional assessment of both academic knowledge and a student’s ability to negotiate the social processes and demands of schooling.

This study is in part a replication and extension of this research domain, especially Bowers (2009) which also used MDS to describe this grades/test score relationship. However, the past research has been limited either to dated or non-U.S. samples, or to small limited and intact samples of less than 200 students. In addition, other than Bowers (2009), the majority of the research has focused exclusively on grades in core subjects, such as mathematics and English. Left out of the discussion has been grades in non-core subjects, such as art, foreign/non-English languages and physical education (PE). Indeed, as noted above, teachers award and record grades in each of these subjects for students in high schools, yet this data has historically not been viewed as relevant assessments of student progress in schools for overall school decision and policymaking. However, grades in non-core subjects provide a novel means to explore this multi-dimensional relationship between grades and standardized tests. If the 25-35% overlap between grades and standardized tests is due to an academic knowledge/cognitive ability dimension, as hypothesized by Klapp Lekholm (2008, 2009), Willingham et al. (2002), and Bowers (2009), then it stands to reason that grades in core subjects, such as mathematics, English, science and social studies, would align in some part with standardized tests in core subjects. Thus, for students, high or low performance in the academic knowledge dimension would contribute similarly to both test scores and grades. Said another way, what it takes to do well in a core class, such as mathematics or science, is similar in the academic knowledge dimension to what it takes to do well on a standardized test on core subjects. Conversely, non-core subjects have less to do with academic knowledge and thus should not align with standardized tests. However, if the current hypothesis from the above literature is correct, then similar affective classroom behaviors, such as participation, attendance, and effort, should help a student do well in both core and non-core subject courses. If this is correct, teacher assigned grades then, as an assessment of both of these cognitive and non-cognitive behaviors could be a very useful assessment of a student’s ability across the spectrum of school performance expectations, and be put to better use than they currently are.



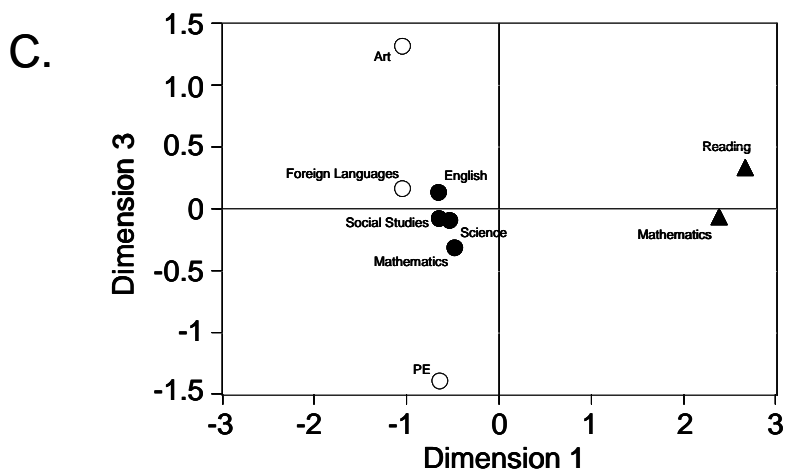
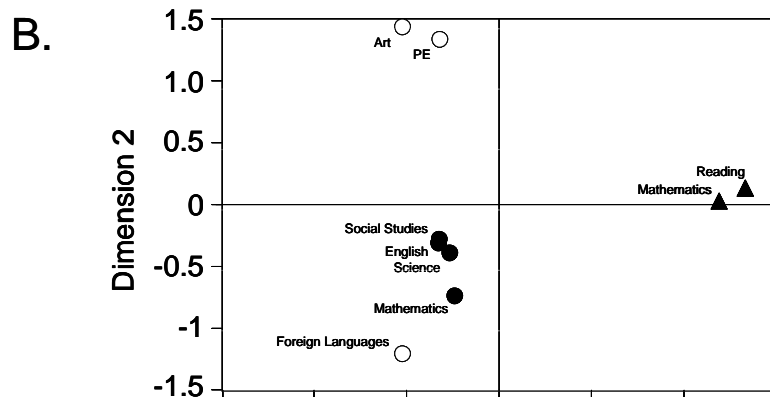
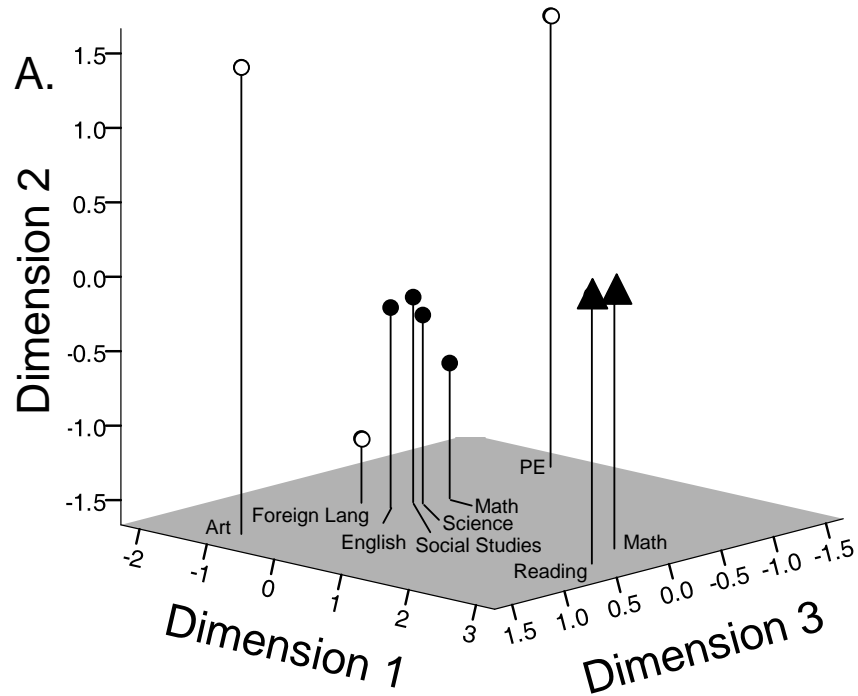
**Figure 1:** MDS of standardized tests and core and non-core GPA,  $n=4520$ .

#### *Assessing the Multi-dimensional Relationship between Grades and Tests*

To address this issue, using multi-dimensional scaling (MDS), the relationship between standardized tests in mathematics and reading were compared from the ELS:2002 dataset to student non-cumulative GPA in each of the eight semesters of high school (grades 9-12) in either core or non-core subjects (see Figure 1). The MDS fit the data well, separating grades and test scores in two dimensions, with a low stress value of 0.097, explaining 96.6% of the variance. As noted in the methods, much like other latent variable analysis techniques, MDS fits the data to the significant dimensions in the data-space. Here, the standardized tests (Figure 1, shaded triangles) and grades in core (shaded circles) and non-core subjects (open circles) separated along two dimensions. Grade level and semester are indicated with a number and either S1 or S2 for semester 1 and semester 2. Dimension 2 appears to be the difference between core and non-core subjects, with the standardized tests and the majority of the core subject grades across the semesters to the north of the plot, and non-core grades to the south. Dimension 1 appears to be the difference between grades and tests, with the standardized tests to the east of the plot, while grades are more to the west. This replicates and extends the findings from Bowers (2009) to a large recent national U.S. sample. Interestingly, grades across the semesters also appear to be distributed chronologically along dimension 1, with grades in both core and non-core subjects in grade levels 9 and 10 closer to the

standardized tests than for grade levels 11 and 12. This may be due to the tests being given in grade 10, as well as the tendency of student enrollment in core subjects to begin to disperse generally after grade 10 (Bowers, 2010a) which may align less with the standardized tests. Thus, dimension 2 appears to describe a more academic knowledge/cognitive behavior dimension, while dimension 1 appears to be the difference between tests and grades.

Nevertheless, aggregating all subjects by grade and semester level removes the ability to examine these relationships by course subject. While students in U.S. high schools are able to select from a broad variety of courses and overall subjects (Bowers, 2010a; Powell, Farrar, & Cohen, 1985), rarely in education research has the relationship between different core and non-core course subjects been compared to standardized tests. However, examining the extent of the relationship between core and non-core subjects and standardized tests is the focus of this study. Here, using the transcript study component of ELS:2002, and following the guidelines of the Classification of Secondary School Courses (CSSC) used to classify course topic in ELS:2002 (Bozick, Lauff, & Wirt, 2007), high school subject-specific grade point averages were calculated for the core subject courses mathematics, English, science and social studies, as well as for the non-core subject courses art, foreign/non-English languages, and physical education (PE) (see methods). These were analyzed along with standardized test scores in mathematics and reading using MDS (see Figure 2).



Legend:

▲ Standardized Test ● Core Subject GPA ○ Non-Core Subject GPA

**Figure 2:** Three-dimensional MDS of standardized tests, grades in core subjects, and grades in non-core subjects,  $n=5230$ . Panel A displays the full MDS solution in three dimensions, while Panel B and C display only two dimensions at a time, holding dimension 1 constant.

Here, the three dimensional solution fit the data well, with a stress value of 0.012, explaining 99.9% of the variance across the different assessments. Figure 2 displays the three dimensional MDS in two ways. First, the standardized tests (shaded triangles), core subjects (shaded circles) and non-core subjects (open circles) are plotted in three dimensions (Figure 2A). Second, to allow the reader to visualize the multiple dimensions, Figure 2 plots two additional two-dimensional plots of the same MDS with dimension 1 held constant along the horizontal axis and the front plane of dimension 1 by dimension 2 plotted as length by height (Figure 2B), with dimension 1 by dimension 3 plotted as length by depth, or the “floor” of the figure (Figure 2C).

The results of the MDS plotted in Figure 2 describes the relationships between standardized tests in mathematics and reading and core and non-core subject specific grades. As in Figure 1, dimension 1 of Figure 2 appears to represent the difference between standardized tests and grades. The tests are to the far east, while the subject grades are to the west (Figure 2B). If one collapses Figure 2B into a single horizontal line along dimension 1, the different subjects cluster very closely together, as do the tests. As will be discussed below, the results of the MDS along dimension 1 suggests that this is the non-cognitive classroom behavior dimension, or the difference between what it takes to get high marks on tests versus grades, accounting for an academic knowledge dimension. The academic knowledge dimension appears to fall along dimension 2. As plotted in Figures 2A and 2B, foreign language, art and PE appear to be distal along dimension 2 from core courses and the standardized tests, with art and PE clustering at the “top” of the three dimensions along dimension 2 with foreign language closer to the “bottom”. Thus, dimension 2 appears to represent the difference between core and non-core subjects. Interestingly, foreign language appears to differ substantially from the other types of courses along dimension 2, indicating that this subject is a different type of non-core subject from art and PE. Dimension 3 plots a novel finding of the Figure 2 MDS that appears to be the difference between art and PE, with core subjects, the standardized tests, and foreign language clustering between the two. Interestingly, there appears to be a continuum along dimension 3, from art, to foreign languages, the reading standardized test, and English to social studies, science the mathematics standardized test, mathematics grades and PE. This result, for the first time, delineates the multi-dimensional aspects of the differences in what is assessed across these different types of high school course subject assessments, and will be discussed in greater detail below.

## DISCUSSION

This study is the first to examine the multi-dimensional relationship between teacher assigned grades in core high school subjects, non-core subjects and standardized tests as a measure of academic knowledge through the analysis of a large U.S. sample of high school students. The results of this study suggest that teacher assigned grades in high school are a multi-dimensional assessment of *both* student academic knowledge and student non-cognitive behaviors. In opposition to the pejorative “hodgepodge” and “kitchen sink” grading literature that asserts that teachers are not reliable sources of information about the performance of their students, such as the literature urging teachers to reform grading practices and align grades to academic knowledge and more standardized assessments (Brookhart, 1991, 2004; Cizek, et al., 1995-1996; Cross & Frary, 1999; Marzano, 2000), the results of this study argue for a change in the conception of grades. This study extends the recent literature that has identified a common grades dimension (Klapp Lekholm & Cliffordson, 2008) that

appears to assess a student’s effort and engagement (Willingham, et al., 2002) and ultimately their success at being schooled (Bowers, 2009). Rather than malign teachers for continuing to report that they award grades for the “hodgepodge” of factors, such as attendance, participation, effort and behavior, the results of this line of research suggest that grades be reconsidered as valuable and useful assessments for not only students, teachers and parents, but also for administrators, researchers, and policymakers. While teacher assigned grades are not a pure assessment of academic knowledge, it appears that teachers may be adept at assessing a student’s ability to perform at the social processes of the institution of schooling, in which academic knowledge is just one component of a much broader array of behaviors required by a student’s community and school. It may be this component of grades that continues to give grades their strong association with overall schooling outcomes, such as graduating or dropping out.

The idea that teacher assigned grades are a multi-dimensional assessment is not new. In writing about elementary school grades over 50 years ago, Talcott Parsons (1959) noted this dual assessment issue between cognitive abilities and behavior when examining teacher assessments:

The pupil is evaluated in diffusely general terms; a *good* pupil is defined in terms of a fusion of the cognitive and the moral components, in which varying weight is given to one or the other. Broadly speaking, then, we may say that the “high achievers” of the elementary school are both the “bright” pupils, who catch on easily to their more strictly intellectual tasks, and the more “responsible” pupils, who “behave well” and on whom the teacher can “count” in her difficult problems of managing the class. (p.304)

A good student gets high grades, but not necessarily high test scores. In the end, teachers award grades for not only academic knowledge, but in the terms of Parsons, for responsibility and behavior. The results of this study, along with the recent research on grades discussed above, extends this to a broader conception of grades as assessments of effort, participation and behavior.

Thus, this study comes to what may appear to be an obvious conclusion, teachers award grades based on how well a student performs across all of the expectations in the classroom. While this point has been known for quite some time, this study, along with the other recent research discussed above on this topic, has worked to describe this relationship between grades and test scores. Because of the perception that grades are not useful beyond the classroom due to a perception that they are unreliable measures of academic knowledge, grades have not been privileged as useful data in the accountability and data driven decision making movements. Yet, schools collect grades continually and the results here support the conclusion that grades are an important assessment of a student’s ability at schooling.

The results of the multi-dimensional scaling between core and non-core grades and the reading and mathematics standardized tests indicate that there is a substantial component of grades that aligns across different subjects when accounting for the academic knowledge component represented by the standardized tests. The three dimensional MDS plotted in Figure 2 above indicates that along dimension 1, grades in all of the different subjects aligned very closely, with tests distal to the course subject grades.



Dimension 1 of both MDS analyses appears to represent a latent factor that describes the difference between standardized tests and grades. One alternative explanation is that this could be just the difference between a student's natural ability at test taking strategies on standardized tests versus teacher created assessments, or a student's test wiseness. Alternatively, dimension 1 could be related to academic knowledge in some way, but dimension 2 appears to relate to academic knowledge in both MDS analyses. The novel aspect of both MDS analyses presented here is the use of non-core subject grades to help interpret dimension 1. Across the literature, non-core subject grades have received little attention in relation to standardized tests. Here, in both MDS analyses, dimension 2 appears to describe the difference between core and non-core subjects, in which the standardized tests align with the core subjects, while grades align together along dimension 1. Thus, if dimension 1 describes the assessment of a student's ability to perform at the social processes of school, termed previously a Success at School Factor (SSF) (Bowers, 2009), then as hypothesized here, grades in core and non-core subjects would align along dimension 1 when academic knowledge is accounted for in the model, here along dimension 2, since non-core subject grades lack the academic knowledge component assessed in the standardized tests, but contain the same SSF component as core subject grades. This is especially evident in Figure 2 when examining specific course subject grades.

However, the distribution along dimension 1 in Figure 1 appears to also include time, with teacher assigned grades in grade levels 9 and 10 proximal to the test scores, and grades in both core and non-core subjects distributing away from test scores as the grade levels increase. As one explanation, the standardized tests were conducted in grade 10, and so this distribution along dimension 1 in Figure 1 may be due to grade levels that were aligned more closely in overall content up until grade 10, and then subjects began to disperse after that point. Previous research on the distribution of course subject enrollment in high school indicates that after grade 10, student enrollment in core course subjects begins to decrease (Bowers, 2010a). Thus for this analysis, teacher assigned grades in the grade levels 11 and 12 may diverge as students select different types of courses, with some selecting core courses that align with college aspirations, and others selecting non-core courses. The second MDS that examined grades by subject area in Figure 2 was included in the study in part to help control for this issue. When examined without grades calculated by time period, but rather by subject area, the grades aligned extremely closely along dimension 1 (*see Figure 2*).

Dimension 3 in Figure 2 is a novel finding of this study. Accounting for the core to non-core dimension (dimension 2) as well as the difference between tests and grades (dimension 1), dimension 3 appears to describe the difference between art and PE. Interestingly, subject area topics align along dimension 3 from art to languages and reading, then to social studies, science, mathematics and PE. This finding was unexpected, but is included here to aid in future research on the differences between grades in core and non-core courses in high school. Because of the order of the subject alignment along dimension 3, these results suggest that grades in art were more related to grades and test scores in languages, while grades in physical education were more related to mathematics. However, the distance along dimension 3 between the subject areas in the core subject cluster in the center is much smaller than the overall distance from the cluster to either art or PE. This means that the relationship along dimension 3 between English and mathematics is greater than the relationship between English and art or mathematics and PE. While this finding is

tentative and in need of replication, dimension 3 may represent the difference between assessments of creative works versus strategic, analytic, or physical ability. Future research should work to replicate and examine this potential relationship further.

### *Limitations*

This study is limited in three main ways. First, due to the use of the ELS:2002 dataset, the analysis is limited only to students who were enrolled in grade 10 in 2002 across the sample. Students and student data from earlier grade levels or who dropped out of school before grade 10 are not included. It may be that the relationships between standardized tests and teacher assigned grades differs for students at earlier grade levels, or by different background factors that may lead to students dropping out before grade 10. Indeed, recent work on teacher assigned grades in Sweden indicates that grades, and especially the common grades dimension discussed here, may vary by student background such as gender (Klapp Lekholm & Cliffordson, 2009), which is also known to be associated with differing rates of dropping out of school in the U.S. context (Rumberger, 2004). Thus, the results here can only be interpreted as relating to students who completed grade 10. Second, the need for complete data across the variables in the ordinal MDS required the use of a subset of students from the full ELS:2002 dataset, limiting the sample size and the ability to generalize to the entire sample, and therefore to the entire population of students who were in grade 10 in 2002. Due to this issue, the results of this study should be considered descriptive rather than generalizable. Third, the requirement for complete data for the MDS resulted in differing sample sizes for the two MDS analyses. In the first MDS in Figure 1, students had to have taken both standardized tests, and had recorded both core and non-core subject grades for all eight semesters of high school, grades 9-12. This means that no students who dropped out of school are included in the first MDS. However, to help address this issue, the second MDS was conducted that required that students had taken both standardized tests and had at least one grade in each of the seven subject areas included. This helped to include students who had dropped out, but students who did not have a grade recorded for all seven subject areas were excluded from this second MDS. Because the first MDS averaged grades across all subjects for core and non-core, it did not have this limitation. Thus, while care was taken to keep the sample sizes as large and comprehensive as possible, both samples should be considered biased, but with each addressing issues from the other. For future work in this area, one way to address these issues would be to analyze entire intact cohorts from large school districts in which all of the data was collected across the different subject areas and tests.

### *Conclusions*

In conclusion, this study extends the past work on what teacher assigned grades actually assess by describing the relationships between core and non-core subject grades in high school in comparison to standardized tests in mathematics and reading. As has been argued extensively over the past few decades, educational research must strive to replicate past findings from specific intact contexts and extend those findings into new domains to help build theory around the core knowledge in education (Berliner, 2002; Shavelson & Towne, 2002). The findings here replicate and extend the work that indicates that teacher assigned grades assess academic knowledge as well as student engagement, effort, participation, and behavior. Teachers award grades based on how well students conform to both the academic and social process demands of the institution of schooling. In one way, it could be

said that grades represent an assessment of how well students play the game of being schooled. Some students are academically gifted and thus can score well on a standardized test, yet for some reason do not conform to the rules and expectations of the schooling process, and thus are given low grades. These low grades, in turn, are strongly associated with students dropping out of school. In contrast, a student may have difficulty with the academic and cognitive aspects of schooling, and thus do poorly on academic tests, but may do well in the SSF or conative engagement, participation, behavior and effort aspects of the schooling process. These scenarios are conceivable given the 0.5 to 0.6 correlation relationship shown here and historically between grades and tests.

The issue for administrators, researchers and policymakers is that teacher assigned grades have not been systematically used for accountability reporting, despite the strong association with schooling outcomes. The results of this study indicate that for data driven decision making and continuous school improvement, teacher assigned grades should be used more systematically to identify how to direct the limited resources of a school district towards the students most in need. It may be that a student is academically challenged, and standardized assessments, be they either periodic assessments or state accountability tests, would most likely show this. These students would benefit from tutoring in specific academic subjects. However, it could be just as likely that a student is not performing well within the social and institutional expectations of the school and classroom. This would be shown by decreasing or failing grades. While grades also indicate academic performance, a student may need tutoring in the social and engagement processes of school, such as how to study, take notes, attend, behave well, participate, and engage in the process. Teacher assigned grades appear to assess precisely this aspect of schooling. In the end, the results of this line of research urge for a reconception of the utility of teacher assigned grades such that interventions should differ based on student cognitive versus behavioral and engagement needs.

## ACKNOWLEDGEMENT

This research was supported by a grant from the American Educational Research Association which receives funds for its "AERA Grants Program" from the National Science Foundation and the National Center for Education Statistics of the Institute of Education Sciences (U.S. Department of Education) under NSF Grant #DRL-0634035. Opinions reflect those of the author(s) and do not necessarily reflect those of the granting agencies.

## RECOMMENDED CITATION FORMAT

Bowers, A.J. (2011) What's in a Grade? The Multidimensional Nature of What Teacher Assigned Grades Assess in High School. *Educational Research & Evaluation*, 17(3), 141-159. [doi: 10.1080/13803611.2011.597112](https://doi.org/10.1080/13803611.2011.597112)

## REFERENCES

- Allensworth, E. M., & Easton, J. Q. (2005). The on-track indicator as a predictor of High School graduation (Vol. 2006): Consortium on Chicago School Research at the University of Chicago.
- Allensworth, E. M., & Easton, J. Q. (2007). What matters for staying on-track and graduating in Chicago public high schools: A close look at course grades, failures, and attendance in the freshman year. Chicago: Consortium on Chicago School Research.
- Balfanz, R., Herzog, L., & MacIver, D. J. (2007). Preventing student disengagement and keeping students on the

- graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, 42(4), 223-235.
- Barclay, J. R., & Doll, B. (2001). Early prospective studies of the high school dropout. *School Psychology Quarterly*, 16(4), 357-370.
- Barrington, B. L., & Hendricks, B. (1989). Differentiating characteristics of high school graduates, dropouts, and nongraduates. *Journal of Educational Research*, 82(6), 309-319.
- Berliner, D. C. (2002). Educational Research: The Hardest Science of All. *Educational Researcher*, 31(8), 18-20.
- Borg, I., & Goren, P. (1997). *Modern multidimensional scaling: Theory and applications*. New York: Springer.
- Bowers, A. J. (2008). Promoting Excellence: Good to great, NYC's district 2, and the case of a high performing school district. *Leadership and Policy in Schools*, 7(2), 154-177. [doi:10.1080/15700760701681108](https://doi.org/10.1080/15700760701681108)
- Bowers, A. J. (2009). Reconsidering grades as data for decision making: More than just academic knowledge. *Journal of Educational Administration*, 47(5), 609-629. [doi:10.1108/09578230910981080](https://doi.org/10.1108/09578230910981080)
- Bowers, A. J. (2010a). Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierarchical cluster analysis. *Practical Assessment Research and Evaluation*, 15(7), 1-18. <http://pareonline.net/pdf/v15n7.pdf>
- Bowers, A. J. (2010b). Grades and Graduation: A Longitudinal Risk Perspective to Identify Student Dropouts. *Journal of Educational Research*, 103(3), 191-207. [doi:10.1080/00220670903382970](https://doi.org/10.1080/00220670903382970)
- Bowers, A. J., & Sprott, R. (2012) Examining the Multiple Trajectories Associated with Dropping Out of High School: A Growth Mixture Model Analysis. *The Journal of Educational Research*, 105(3), 176-195. [doi:10.1080/00220671.2011.552075](https://doi.org/10.1080/00220671.2011.552075)
- Bozick, R., Lauff, E., & Wirt, J. (2007). *Education longitudinal study of 2002 (ELS:2002): A first look at the initial postsecondary experiences of the high school sophomore class of 2002*. Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Bozick, R., Lyttle, T., Siegel, P. H., Ingels, S. J., Rogers, J. E., Lauff, E., & Planty, M. (2006). Education Longitudinal Study of 2002: First follow-up transcript component data file documentation. Washington, DC: National Center for Education Statistics, U.S. Department of Education.
- 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), 173-215.
- Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, 10(1), 35-36.
- Brookhart, S. M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers College Record*, 106(3), 429-458.
- Burton, N. W., & Ramist, L. (2001). Predicting success in college: SAT studies of classes graduating since 1980. New York, NY: The College Board.
- Cizek, G. J. (2000). Pockets of resistance in the assessment revolution. *Educational Measurement: Issues and Practice*, 19(2), 16-23.

- Cizek, G. J., Fitzgerald, S. M., & Rachor, R. E. (1995-1996). Teachers' assessment practices: Preparation, isolation and the kitchen sink. *Educational Assessment, 3*(2), 159-179.
- Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education, 12*(1), 53-72.
- Davidson, M. L. (1983). *Multidimensional scaling*. New York: Wiley.
- Ekstrom, R. B., Goertz, M. E., Pollack, J. M., & Rock, D. A. (1986). Who drops out of high school and why? Findings from a national study. *Teachers College Record, 87*(3), 356-373.
- Enserink, M. (2008). Mapmaker for the world of influenza. *Science, 320*(5874), 310-311.
- Ensminger, M. E., & Slusarcick, A. L. (1992). Paths to high school graduation or dropout: A longitudinal study of a first-grade cohort. *Sociology of Education, 65*(2), 91-113.
- Farr, B. P. (2000). Grading practices: An overview of the issues. In E. Trumbull & B. Farr (Eds.), *Grading and reporting student progress in an age of standards* (pp. 1-22). Norwood: Christopher-Gordon Publishers.
- Fitzsimmons, S. J., Cheever, J., Leonard, E., & Macunovich, D. (1969). School failures: Now and tomorrow. *Developmental Psychology, 1*(134-146).
- Frary, R. B., Cross, L. H., & Weber, L. J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. *Educational Measurement: Issues and Practice, 12*(3), 23-30.
- Glass, G. V. (1967). Factors in teachers' perceptions of students. *Journal of Educational Measurement, 4*(2), 87-93.
- Guskey, T. R. (2007). Multiple sources of evidence: An analysis of stakeholder perceptions of various indicators of student learning. *Educational Measurement: Issues and Practice, 26*(1), 19-27.
- Hallinan, M. T. (1992). The organization of students for instruction in middle school. *Sociology of Education, 65*(2), 114-127.
- Heritage, M., & Yeagley, R. (2005). Data use and school improvement: Challenges and prospects. In J. L. Herman & E. H. Haertel (Eds.), *Uses and misuses of data for educational accountability and improvement: 104th yearbook of the National Society for the Study of Education* (pp. 320-339). Malden, Mass.: Blackwell Publishing.
- Ingles, S. J., Pratt, D. J., Rogers, J. E., Siegel, P. H., Stutts, E. S., & Owings, J. A. (2004). *Education longitudinal study of 2002: Base year data file user's manual*. Washington, D.C.: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Ingles, S. J., Pratt, D. J., Wilson, D., Burns, L. J., Currivan, D., Rogers, J. E., & Hubbard-Bednasz, S. (2007). *Education longitudinal study of 2002: Base-year to second follow-up data file documentation*. Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.
- Klapp Lekholm, A., & Cliffordson, C. (2008). Discrepancies between school grades and test scores at individual and school level: effects of gender and family background. *Educational Research and Evaluation, 14*(2), 181-199.
- Klapp Lekholm, A., & Cliffordson, C. (2009). Effects of student characteristics on grades in compulsory school. *Educational Research and Evaluation, 15*(1), 1-23. doi: 10.1080/13803610802470425
- Kuncel, N. R., Crede, M., & Thomas, L. L. (2005). The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research, 75*(1), 63-83.
- Linn, R. L. (1982). Ability testing: Individual differences, prediction, and differential prediction. In A. K. Wigdor & W. R. Garner (Eds.), *Ability testing: Uses, consequences, and controversies* (pp. 335-388). Washington DC: National Academy Press.
- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher, 29*(2), 4-16.
- Lloyd, D. N. (1974). Analysis of sixth grade characteristics predicting high school dropout or graduation. *JSAS Catalog of Selected Documents in Psychology, 4*, 90.
- Lloyd, D. N. (1978). Prediction of school failure from third-grade data. *Educational and Psychological Measurement, 38*(4), 1193-1200.
- Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development (ASCD).
- McMillan, J. H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice, 20*(1), 20-32.
- McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice, 22*(4), 34-43.
- Mead, A. (1992). Review of the development of multidimensional scaling. *The Statistician, 41*(1), 27-39.
- NCES. (n.d.). Education longitudinal study of 2002 (ELS:2002), from <http://www.nces.ed.gov/surveys/els2002/>
- Parsons, T. (1959). The school class as a social system: Some of its functions in American society. *Harvard Educational Review, 29*(4), 297-318.
- Pincus, H. S., & Schmelkin, L. P. (2003). Faculty perceptions of academic dishonesty: A multidimensional scaling analysis. *The Journal of Higher Education, 74*(2), 196-209.
- Powell, A. G., Farrar, E., & Cohen, D. K. (1985). *The shopping mall high school: Winners and losers in the educational marketplace*. Boston: Houghton Mifflin.
- Randall, J., & Englehard, G. (2009). Examining teacher grades using rasch measurement theory. *Journal of Educational Measurement, 46*(1), 1-18. doi: 10.1111/j.1745-3984.2009.01066.x
- Rumberger, R. W. (2004). Why students drop out of school. In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 131-156). Cambridge, MA: Harvard Education Press.
- Rumberger, R. W., & Palardy, G. J. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Educational Research Journal, 42*(1), 3-42.
- Russell, C. A., Jones, T. C., Barr, I. G., Cox, N. J., Garten, R. J., Gregory, V., . . . Smith, D. J. (2008). The global circulation of seasonal influenza A (H3N2) viruses. *Science, 320*(5874), 340-346.
- Shavelson, R. J., & Towne, L. (2002). *Scientific Research in Education*. Washington, DC: National Academy Press.
- Smith, D. J., Lapedes, A. S., deJong, J. C., Bestebroer, T. M., Rimmelzwaan, G. F., Osterhaus, A. D. M. E., & Fouchier, R. A. M. (2004). Mapping the antigenic and genetic evolution of influenza virus. *Science, 305*(5682), 371-376.

- Song, M., & Miskel, C. G. (2007). Exploring the structural properties of the state reading policy domain using network visualization technique. *Educational Policy*, 21(4), 589-614.
- Subkoviak, M. J. (1975). The use of multidimensional scaling in educational research. *Review of Educational Research*, 45(3), 387-423.
- 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.
- Woodruff, D. J., & Ziomek, R. L. (2004). High school grade inflation from 1991 to 2003. Research report series 2004-04. Iowa City, IA: ACT, Inc.
- Zwick, R., & Greif Green, J. (2007). New perspectives on the correlation of SAT scores, high school grades and socioeconomic factors. *Journal of Educational Measurement*, 44(1), 23-45.10.1111/j.1745-3984.2007.00025.x.