

Report Card Grades and Educational Outcomes^{1,2}

Alex J. Bowers
Teachers College, Columbia University
Bowers@tc.edu

Introduction:

Over the last 100 years, researchers have criticized teacher-assigned grades as subjective and unreliable measures of student academic achievement (Allen, 2005; Banker, 1927; Carter, 1952; Evans, 1976; Hargis, 1990; Kirschenbaum, Napier, & Simon, 1971; Quann, 1983; Simon & Bellanca, 1976), referring to them as "hodgepodge" (Brookhart, 1991) or "kitchen sink" practices (Cizek, 2000; Cizek, Fitzgerald, & Rachor, 1995–1996). When teachers are asked what they are assessing with their grades, they consistently say not only student academic knowledge and achievement but also student persistence, behavior, participation, and effort (Henke, Chen, Goldman, Rollefson, & Gruber, 1999; Randall & Engelhard, 2009, 2010).

Why Is This Area of Research Important?

Mixing academic and nonacademic information in one grade results in a measure that is hard to interpret. However, as Cross and Frary (1999) note,

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)

Despite 100 years of research on the subjective nature of grades, mixed grading practices continue unabated alongside the rise of standardized testing responsibilities (Busick, 2000; Farr, 2000; Trumbull, 2000). The research shows that grades can be useful indicators of a host of factors besides academic progress (Bisesi,

Farr, Greene, & Haydel, 2000; Folzer-Napier, 1976; Linn, 1982); as Swineford (1947) noted in a study on middle and high school grading practices, "the data clearly show that marks assigned by teachers in this school are reliable measures of *something* but there is apparently a lack of agreement on just what that something should be" (p.47). Indeed, over the past 100 years, a strong line of research has attempted to understand the different components represented by grades as a means to inform decision making in schools and classrooms (Bowers, 2009; Parsons, 1959). Additionally, a persistent finding across this literature is that while standardized test scores have low criterion validity for overall schooling outcomes, such as graduation from high school and admission to post-secondary institutions, grades have consistently been the strongest predictors of K–12 educational persistence, completion, and transition from high school to college (Atkinson & Geiser, 2009; Bowers, Spratt, & Taff, 2013).

In this chapter, I will review the quantitative research over the past 100 years regarding what this "something" is that teacher-assigned grades represent beyond the fundamental academic skills assessed by standardized test scores. I will also examine recent research in this area over the last few decades showing that teacher-assigned grades and marks assess not just student achievement but also persistence, behavior, and substantive engagement in the schooling process. Finally, I will review the research on how grades align to educational outcomes.

What Significant Studies Have Been Conducted in This Area?

To study these issues, I have reviewed studies of the research on K–12 grades as predictors of educational outcomes. My goal was to include studies from the past 100 years that (1) examined the relationship of K–12 grades to other schooling outcomes (e.g., test scores, dropout rates), (2) were quantitative, and (3) analyzed data from actual student assessments rather than focusing on teacher perspectives. For a detailed description of the literature I reviewed, please see Brookhart and colleagues (2016). I reviewed the article titles from the searches and then read the articles that matched the aim of this chapter.

What Have the Results of Those Studies Revealed?

The studies offer insight into both the relationship between grades and test scores and the use of grades as multidimensional assessments.

The Relationship Between Grades and Test Scores

Scholars researching the relationship between grades and test scores have tended to focus on the relationship of average grades

¹ An earlier version of this manuscript appeared as a section in the journal article: Brookhart, S., Guskey, T., Bowers, A.J., McMillan, J. Smith, L. Smith, J., Welsh, M. (2016) A Century of Grading Research: Meaning and Value in the Most Common Educational Measure. *Review of Educational Research*, 86(4), p. 803-848 <http://doi.org/10.3102/0034654316672069>

² This document is a pre-print of a book chapter published in 2019 in the book titled *What We Know About Grading: What Works, What Doesn't, and What's Next*. Citation: Bowers, A.J. (2019) Report Card Grades and Educational Outcomes. In Guskey, T., Brookhart, S. (Eds.) *What We Know About Grading: What Works, What Doesn't, and What's Next*, (p.32-56). Alexandria, VA: Association for Supervision and Curriculum Development.

(usually GPA) to standardized test scores. In the early 20th century, scholars began correlating standardized test scores to grades, first using small intact samples of students and the intelligence tests available at the time (Ross & Hooks, 1930; Unzicker, 1925) and then progressing to larger or more nationally generalizable samples and multiple standardized tests across subjects (Pattison, Grodsky, & Muller, 2013).

Although the perception remains that grades and test scores should logically approach a correlation of 1.0 (Allen, 2005; Duckworth, Quinn, & Tsukayama, 2012; Stanley & Baines, 2004), the research tells us that the correlation is in fact more moderate (closer to 0.5). As noted by Willingham, Pollack and Lewis (2012):

Understanding these characteristics of grades is important for the valid use of test scores as well as grade averages because, in practice, the two measures are often intimately connected... [there is a] tendency to assume that a grade average and a test score are, in some sense, mutual surrogates; that is, measuring much the same thing, even in the face of obvious differences. (p. 2)

The research can be divided into two distinct eras: (1) studies from the early 20th century and (2) studies from the late 20th century and early 21st century, when there was a resurgent interest in the topic. As an example of work in the first era, Unzicker (1925) examined the relationship between the grades of 425 7th, 8th, 9th graders in the same school and their scores on the Otis intelligence test. He found that average grades across English, mathematics, and history had a 0.47 correlation with the tests.

Ross and Hooks (1930) provided the most comprehensive review of the research available in 1930, analyzing 20 studies from 1920 through 1929 on the degree of correlation between the intelligence tests and report card grades students receive and their marks across a variety of subjects in 7th, 8th, and 9th grades. Samples across the studies included mostly single-school intact samples of 49 to 157 students and a range of popular intelligence assessments including the Illinois, Otis, and National tests (Brooks, 1929). Ross and Hooks (1930) found the correlation the relationships of intelligence tests and grades given in the same 7th-grade year ranged from 0.38 to 0.44 and drew the following conclusion:

Data from this and other studies indicate that . . . the grade school record affords a more reliable or consistent basis of prediction [of high school achievement] than any other available, the correlations in three widely-scattered school systems showing remarkable stability; and that without question the grade school record of the pupil is the most usable or practical of all bases for prediction, being available wherever cumulative records are kept, without cost and with a minimum expenditure of time and effort. (p. 195)

Subsequent studies compared grades to standardized achievement tests rather than intelligence tests. For example, Moore (1939) directly compared an analysis of about 200 5th and 6th graders in Wyoming and Colorado with similar studies from the time (Dickinson, 1925; Heilman, 1928; Kertes, 1932), identifying the correlation between the students' average grades to their average scores on the new Stanford Achievement Test as 0.61. Similarly, Carter (1952) examined the relationship between 235 student grades from a high school in Pennsylvania and standardized algebra achievement scores, finding a 0.52 correlation. A study by McCandless, Roberts, and Starnes from 1972, between the two dominant eras of research, examined the correlations between the average grades of 433 Atlanta 7th graders and their scores on the Metropolitan Achievement Test, tabulating differences by student socio-economic status, ethnicity, and gender. The researchers found that the overall correlation between grades and the standardized test was 0.31. Clearly, the first 50 or so years of research were marked by a focus on intelligence testing versus achievement testing, reflecting a desire to figure out the "something" that teacher-assigned grades represent (Ross & Hooks, 1930).

Research from the late 20th and early 21st centuries has replicated and extended the early findings using much larger and more representative samples and more modern standardized tests and methods (Brennan, Kim, Wenz-Gross, & Siperstein, 2001; Woodruff & Ziomek, 2004). Using data from 736 8th graders from across six Boston schools, Brennan and colleagues (2001) compared students' scores on the Massachusetts Comprehensive Assessment System reading test to their grades in mathematics, English, and science classes, reporting correlations of 0.54, 0.59, and 0.54 respectively.

In a study using a small intact sample of 140 8th graders, Duckworth and Seligman (2006) compared students' GPAs to their scores on the TerraNova Second Edition California Achievement Test, finding a correlation of 0.66. Subsequently, Duckworth and colleagues (2012) examined the data from 1,364 9th grade students and 510 8th grade students, comparing reading and math scores on standardized tests to GPAs, finding correlations between 0.62 and 0.66. Woodruff and Ziomek (2004) compared the data from all high school students who took the ACT college entrance exam between 1991 and 2003--about 700,000 students per year--and found consistently strong correlations of average GPAs and overall composite ACT scores ranging from 0.56 to 0.58, and specific correlations of math grades to ACT scores between 0.54 to 0.57 and of English scores to ACT scores between 0.45 to 0.50. One critique of this study is that the student self-reported their GPAs (Woodruff & Ziomek, 2004). Pattison and colleagues (2013) examined data from students who completed high school taken from nationally generalizable longitudinal studies from the National Center for Education Statistics, including the National Longitudinal Study of the High School Class of 1972 (NLS72), the High School and Beyond sophomore cohort (HS&B), the National Educational Longitudinal Study of 1988 (NELS), and the Educational Longitudinal Study of 2002 (ELS), comparing high school GPA from reading, mathematics, science, and social studies to the NCES standardized tests in 10th grade reading and mathematics.

Figure 2.1: Studies of K–12 Report Card Grades as Multidimensional Measures of Academic Knowledge, Engagement, and Persistence

Study	Sample	Main Findings
Bowers (2009)	High school students	<ul style="list-style-type: none"> Grades were multidimensional, separating core and noncore grades from state standardized assessments in science, mathematics, and reading.
Bowers (2011)	High school students	<ul style="list-style-type: none"> Three main grading factors were identified: (1) a cognitive factor that describes the relationship between tests and core subject grades, (2) an engagement factor between core subject grades and noncore subject grades, and (3) a factor that described the difference between grades in art and physical education.
Casillas et al. (2012)	7th and 8th graders	<ul style="list-style-type: none"> 25% of the explained variance in GPA was attributable to standardized assessments. Academic discipline and commitment to school were strongly related to GPA.
Farkas, Grobe, Sheehan, and Shaun (1990)	8th graders and their teachers	<ul style="list-style-type: none"> Student work habits were the strongest noncognitive predictors of grades.
Kelly (2008)	6th, 7th, and 8th grade students	<ul style="list-style-type: none"> Positive and significant effects of students' substantive engagement were found on subsequent grades, but no relationship with procedural engagement.
Klapp, Lekholm, and Cliffordson (2008)	Swedish students	<ul style="list-style-type: none"> Grades consisted of two major factors: (1) a cognitive achievement factor and (2) a noncognitive "common grade dimension."
Klapp, Lekholm, and Cliffordson (2009)	Swedish students	<ul style="list-style-type: none"> The cognitive achievement factor of grades consists of student self-perception of competence, self-efficacy, coping strategies, and subject-specific interest. The noncognitive factor consists of motivation and a general interest in school.
Klapp, Lekholm, and Cliffordson (2011)	Swedish students	<ul style="list-style-type: none"> The cognitive achievement factor of grades consists of student self-perception of competence, self-efficacy, coping strategies, and subject-specific interest. The noncognitive factor consists of motivation and a general interest in school.
Miner (1967)	High school students	<ul style="list-style-type: none"> The study examined academic grades in 1st, 3rd, 6th, 9th, and 12th grade; achievement tests in 5th, 6th, and 9th grades; and citizenship grades in 1st, 3rd, and 6th grades. Three factors were found: (1) objective achievement, (2) behavior factor, and (3) high school achievement as measured through grades.
Sobel (1936)	Not reported	<ul style="list-style-type: none"> Students were categorized into three groups based on comparing grades and achievement test levels: (1) grade-superior, (2) middle-group, and (3) mark-superior.
Thorsen and Cliffordson (2012)	9th grade students in Sweden	<ul style="list-style-type: none"> The study generally replicated Klapp, Lekholm, and Cliffordson (2009).
Thorsen (2014)	Students in Sweden	<ul style="list-style-type: none"> The study generally replicated Klapp, Lekholm, and Cliffordson (2009) in examining norm-referenced grades.
Willingham, Pollack, and Lewis (2002)	High school students	<ul style="list-style-type: none"> A moderate relationship between grades and tests was identified as were strong positive relationships between grades and student motivation, engagement, completion of work assigned, and persistence.

These data consist of tens of thousands of student achievement patterns across four decades (Pattison et al., 2013). The authors found GPA correlations consistent with the past research, including 0.52 to 0.64 for mathematics and 0.46 to 0.54 for reading comprehension.

Over 100 years of research strongly suggest that teacher-assigned grades correlate at about 0.5 to standardized measures of achievement. While there is some variance from year to year and across core subjects, when considering large nationally representative datasets, the correlation is neither very weak (indicating that teacher grades aren't completely subjective) nor very strong (indicating that teacher grades aren't perfect measures of fundamental academic knowledge). Rather, the correlation between grades and tests appears to be consistently moderate. This consistent moderate relationship persists across a significant amount of time and studies and despite large shifts across the educational system, especially in relation to accountability and standardized testing (Linn, 1982). When this moderate correlation is squared, about 25 percent of a teacher-assigned grade appears to address the fundamental academic knowledge measured by standardized tests (Bowers, 2011), with the remaining 75 percent at least partly attributable to separate factors.

Grades as Multidimensional Assessments

The research shows that grades appear to assess not just academic knowledge but substantive engagement and persistence as well. In one early study, Sobel (1936) calculated the difference between student's standardized test scores and classroom grades, categorizing the students as "mark-superior" (high grades and low test scores), "test-superior" (low grades and high test scores), or "middle group" (average grades and test scores). Sobel noted that "mark-superior" students "are apparently outstanding in penmanship, attendance, punctuality, and effort marks. They also rank high in teachers' ratings on certain personality traits--industry, perseverance, dependability, cooperation, and ambition" (p. 239).

Miner (1967) examined 671 student's achievement variables from three high schools in a Midwestern city, including academic grades in 1st, 3rd, 6th, 9th, and 12th grades; achievement tests in 5th, 6th, and 9th grades; and classroom citizenship grades in 1st, 3rd, and 6th grades. She found that the variables separated into three factors: (1) objective achievement as measured through standardized assessments, (2) early classroom citizenship (measuring a behavior factor), and (3) high school achievement as measured through grades. In this study, Miner demonstrated that grades could be identified as a factor separate from other achievement and behavior measures.

In attempting to understand the components of teacher assigned grades, Farkas and colleagues (1990) examined the grades of 486 8th grade students. They also included a teacher survey of measures of each student's basic skills, absenteeism, work habits, disruptiveness, appearance and dress, and coursework mastery as measured by a district curriculum-referenced test on subject-specific skills. The authors show that student work habits and

basic skills were the strongest noncognitive social and emotional predictors of grades:

Most striking is the powerful effect of student work habits upon course grades. This confirms the notion that . . . teacher judgments of student noncognitive characteristics are powerful determinants of course grades, even when student cognitive performance is controlled. (p. 140)

Willingham, Pollack, and Lewis (2002) analyzed the second follow-up to the National Center for Educational Statistics' NELS:88 dataset, examining the data from the full high school transcripts of 8,454 students across 581 schools. The authors examined the relationship between grades and a host of variables, including standardized tests, school skills, initiative, activities such as television watching and socializing, family background, and student attitudes towards school. Beyond the moderate correlation between grades and tests, the authors found strong positive relationships between grades and student motivation, engagement, completion of work assigned, and persistence. The authors found that grades provide a useful assessment of both conative factors (e.g., interest, volition, and self-regulation; see Snow, 1989) and cognitive student factors.

One critique of using grades to assess conative factors is that teachers may award grades based on students *appearing* engaged but just going through the motions (procedural engagement) rather than on legitimate effort and participation that leads to increased learning (substantive engagement). To address this issue, Kelly (2008) examined the data of 1,653 6th, 7th, and 8th graders related to grades, achievement, family background, and student effort. Student engagement was assessed through observation and coding of an extensive set of video-recorded classroom observations. Kelly found a significant correlation between substantive engagement and higher classroom grades but a statistically nonsignificant relationship between procedural engagement and grades, concluding that "most teachers successfully use grades to reward achievement-oriented behavior and promote a widespread growth in achievement" (p. 45). Kelly continued:

The misperception that teachers base grades on the appearance of cooperation, rather than on legitimate effort, lends support to the use of high-stakes tests as the sole criterion for promotion decision by advancing the myth that standardized tests are inherently objective, while teachers' assessments are inherently subjective and likely to be biased. (p. 46)

As a researcher myself, I used multidimensional scaling to examine the relationship between grades and standardized tests, first with a small intact sample of 195 students from two small school districts (Bowers, 2009) and again with a sample from the NCES ELS:2002 dataset of 4,520 students (Bowers, 2011). These studies examined the difference not just between grades and standardized test scores in each semester in high school, but also between grades in core subjects (mathematics, English, science, and social studies) and noncore subjects (foreign/non-

English languages, art and physical education). In my analysis, I found strong evidence of three factors at play: (1) the cognitive factor for the relationship between tests and core subject grades, (2) the conative factor for the relationship between core subject grades and noncore subject grades (termed a "Success at School Factor, SSF"), and (3) a factor related specifically to the difference between grades in art and physical education. I was able to show that much of the variance in grades was not attributable to the moderate correlation between tests and grades, but rather to teachers assessing students' ability to navigate the social processes of schooling using substantive engagement and persistence. Subsequently, Duckworth, Quinn, and Tsukayama (2012) used structural equation modeling of data for 510 New York City 5th–8th graders to show that the engagement and persistence factor is mediated through teacher evaluations of student conduct and homework completion.

Analyzing a sample of 4,660 7th and 8th graders across 24 middle schools from the Midwest and South, Casillas and colleagues (2012) examined the relationship between students' classroom grades and scores on ACT's middle-grades English, mathematics, science, and reading assessments and a range of psychosocial characteristics (student motivation through academic discipline and commitment to school, social control such as positive family and personal relationships, self-regulation and orderly conduct, and behaviors such as absenteeism and amount of time spent on homework). The authors found that 25 percent of students' GPA was attributable to standardized assessments, with 30 percent attributable to prior grades, 23 percent to psychosocial factors, 10 percent to behavioral indicators, 9 percent to demographics, and 3 percent to school factors. Of the psychosocial factors, academic discipline and commitment to school had the strongest relationship to GPA.

A strong set of recent studies hails from Sweden (see, for example, Cliffordson, 2008; Klapp Lekholm, 2011; Klapp Lekholm & Cliffordson, 2008, 2009; Thorsen, 2014; Thorsen & Cliffordson, 2012) and analyzes datasets of about 100,000 Swedish students. Klapp, Lekholm, and Cliffordson (2008) examined data for the entire population of 99,070 Swedish students who left compulsory school in 2003 across 1,246 schools and showed that grades in mathematics, English, and Swedish consisted of two major factors: (1) a cognitive achievement factor and (2) a noncognitive "common grade dimension." In a follow-up study (Klapp, Lekholm, & Cliffordson, 2009), the authors reanalyzed the same with a focus on multiple student and school characteristics that influenced both factors. For the cognitive achievement factor of grades, the following characteristics were most important: student self-perception of competence, self-efficacy, coping strategies, and subject-specific interest. By contrast, for the noncognitive "common grade dimension" related to higher grades across all three subjects, the most important student variables were motivation and a general interest in school. These results were then replicated across three full population-level cohorts in Sweden representing all 9th grade students in the years 2003 (99,085), 2004 (105,697), and 2005 (108,753) (Thorsen & Cliffordson, 2012), as well as in comparison to both norm-

Bowers (2019)

referenced and criterion-referenced grading systems using a data set of 3,855 students (Thorsen, 2014). Klapp, Lekholm, and Cliffordson (2009) noted:

The relation between general interest or motivation and the common grade dimension seems to recognize that students who are motivated often possess both specific and general goals and approach new phenomena with the goal of understanding them, which is a student characteristic awarded in grades. (p. 19)

These findings provide strong evidence in replication of Kelly's (2008), Bowers' (2009, 2011) and Casillas and colleagues' (2012) findings that substantive engagement in the process of schooling is an important component of grades unrelated to the component of grades that is measured well by standardized tests.

Thus, grades are not and have never been a valid measure of academic achievement. Rather, they are a multidimensional assessment of *both* academic achievement (cognitive factors) *and* substantive engagement in the schooling process (noncognitive/conative factors). This makes grading a very useful assessment, particularly when augmented by standardized test score data. As the goals of education are far broader than acquiring the fundamental academic knowledge and skills represented by scores on standardized achievement tests (Cusick, 1983; Labaree, 1997, 2012; Nichols & Berliner, 2007), it is a strength and a benefit of the system that schools already possess both tests and grades for assessing multiple and sometimes conflicting goals of schooling. Grades appear to be very useful as assessments of noncognitive social and emotional behavior factors that are well-known to predict educational outcomes (Heckman & Rubinstein, 2001; Levin, 2013). This is important, as contemporary researchers have postulated that while noncognitive skills help to build cognitive skills, the reverse may not be the case (Cunha & Heckman, 2008).

Grades as Predictors of Educational Outcomes

Teacher-assigned grades are known predictors of high school graduation (Bowers, 2014) and college attendance (Atkinson & Geiser, 2009; Cliffordson, 2008). This is not surprising, as satisfactory grades historically have been one of the main criteria for a high school diploma (Rumberger, 2011). Many studies have used grades in early elementary school to identify students categorized as "at-risk" (Gleason & Dynarski, 2002; Pallas, 1989). Early studies in this domain (Fitzsimmons, Cheever, Leonard, & Macunovich, 1969; Lloyd, 1974, 1978; Voss, Wendling, & Elliott, 1966) identified teacher-assigned grades as among the strongest predictors of whether or not a student would reach graduation. Subsequent studies combined these findings with multiple other variables, such as absences and misbehavior; however, grades remained a strong predictor (Barrington & Hendricks, 1989; Cairns, Cairns, & Neckerman, 1989; Ekstrom, Goertz, Pollack, & Rock, 1986; Ensminger & Slusarcick, 1992; Finn, 1989; Hargis, 1990; Morris, Ehren, & Lenz, 1991; Rumberger, 1987; Troob, 1985). More recent research regards low or failing grades as having a cumulative effect over years on students deciding to drop out of school (Alexander, Entwisle, & Kabbani, 2001; Jimerson, Egeland, Sroufe, & Carlson, 2000; Pallas, 2003; Roderick & Camburn, 1999).

Figure 2.1: Studies of Grades as Predictors of Educational Outcomes

Study	Sample	Main Findings
Alexander, Entwisle, and Kabbani (2001)	9th grade students	<ul style="list-style-type: none"> Student background, grade retention, academic performance and behavior strongly related to dropping out of school.
Allensworth and Easton (2007)	9th grades students in Chicago	<ul style="list-style-type: none"> GPA and failing a course in early high school strongly predicted dropping out of school.
Allensworth, Gwynne, Moore, and de la Torre (2014)	8th grade Chicago students	<ul style="list-style-type: none"> Middle school grades and attendance are stronger predictors of high school performance in comparison to test scores, and middle school grades are a strong predictor of students on or off track for high school success.
Balfanz, Herzog, and MacIver (2007)	6th grade students from Philadelphia	<ul style="list-style-type: none"> Predictors of dropping out of high school included failing mathematics or English, low attendance, and poor behavior.
Barrington and Hendricks (1989)	High school students	<ul style="list-style-type: none"> GPA, number of low grades, intelligence test scores, and student mobility significantly predicted dropout.
Bowers (2010a)	Students tracked from 1st grade through high school	<ul style="list-style-type: none"> Having low grades over time across all types of course subjects correlated with dropping out and not taking the ACT.
Bowers (2010b)	Students tracked from 1st grade through high school	<ul style="list-style-type: none"> Receiving low grades (<i>D</i> or <i>F</i>) and being held back in a grade strongly related to dropping out.
Bowers and Sprott (2012)	10th grade students	<ul style="list-style-type: none"> Noncumulative GPA trajectories in early high school were strongly predictive of dropping out.
Bowers, Sprott, and Taff (2013)	Review of 36 previous studies	<ul style="list-style-type: none"> Dropout flags focusing on GPA were some of the most accurate dropout flags across the literature.
Cairns, Cairns, and Neckerman (1989)	7th grade students	<ul style="list-style-type: none"> Beyond student demographics, student aggressiveness and low levels of academic performance were associated with dropping out.
Cliffordson (2008)	Swedish students	<ul style="list-style-type: none"> Grades predicted achievement in higher education more strongly than the Swedish Scholastic Aptitude Test) and criterion-referenced grades were slightly better predictors than norm-referenced grades.
Ekstrom, Goertz, Pollack, and Rock (1986)	10th grade students	<ul style="list-style-type: none"> Grades and problem behavior were identified as the most important variables for identifying dropping out, even higher than test scores.
Ensminger and Slusarcick (1992)	1st graders from historically disadvantaged communities	<ul style="list-style-type: none"> Low grades and aggressive behavior were related to eventually dropping out, with low SES skewing relationships negatively.
Fitzsimmons, Cheever, Leonard, and Macunovich (1969)	High school students	<ul style="list-style-type: none"> Students receiving low grades (<i>D</i> or <i>F</i>) in elementary or middle school were at much higher risk of dropping out.
Jimerson, Egeland, Sroufe, and Carlson (2000)	Children tracked from birth through age 19	<ul style="list-style-type: none"> Home environment, quality of parent caregiving, academic achievement, student problem behaviors, peer competence, and intelligence test scores were significantly related with dropping out.
Lloyd (1978)	3rd grade students	<ul style="list-style-type: none"> Grades and marks significantly correlated with dropping out.

Bowers (2019)

Morris, Ehren, and Lenz (1991)	Students in 7th through 12th grade	<ul style="list-style-type: none"> • Dropping out was predicted by absences, low grades (<i>D</i> or <i>F</i>), mobility.
Roderick and Camburn (1999)	Chicago 9th graders	<ul style="list-style-type: none"> • Significant predictors of course failure included low attendance and found failure rates varied significantly at the school level.
Troob (1985)	New York City high school students	<ul style="list-style-type: none"> • Low grades and high absences corresponded to higher levels of dropping out.

The more recent research focuses on the influence of low grades and of a continuous scale of grades (such as GPA) on students deciding to drop out. For example, studies of students in Chicago have shown that failing a core subject course in 9th grade is highly correlated with dropping out of school (Allensworth, 2013; Allensworth & Easton, 2005, 2007) and at the middle school level, there is a correlation between middle school grades and transitioning from middle school to high school (Allensworth, Gwynne, Moore, & de la Torre, 2014). Using data from Philadelphia, Balfanz, Herzog, and McIver (2007) showed a strong relationship between failing core courses in 6th grade and dropping out. In my own work, I have found the strongest predictor of dropping out, after grade retention, to be receipt of *Ds* and *Fs* (Bowers, 2010b).

Many studies also consider the full GPA scale in predicting school completion (Rumberger & Palardy, 2005). However, few studies have focused on grades alone as the only predictor of graduation or dropping out, rather examining patterns in grades (Bowers, 2010a; Bowers & Sprott, 2012). A recent review of the research on the accuracy of dropout flags and predictors showed that longitudinal GPA trajectories were among of the most accurate predictors in the literature to date (Bowers et al., 2013).

What Are the Implications of These Research Findings for Improvement in Grading Policy and Practice?

A century of quantitative studies on K–12 classroom grades shows that teacher-assigned grades are a multidimensional assessment of student cognitive and noncognitive/conative factors. Grades represent both the academic knowledge represented in standardized test scores as well as substantive engagement, persistence, and positive school behaviors. Grades and standardized tests are moderately correlated, and the assumption that grades and tests should have a strong relationship is misplaced, as these two assessments have never been shown to have a strong relationship. Rather, grades are a useful assessment of multiple factors that teachers value in student work, and thus useful in identifying students who may face either academic or socio-behavioral challenges in the future. The research, especially over the last two decades, suggests that when combined with standardized tests, teacher-assigned grades provide a rich multidimensional assessment of student performance. From a policy perspective, then, both grades and standardized test scores should be considered when making major decisions about students.

Unfortunately, policy and practice discussions often confuse standardized tests and grades. Do they measure the same thing? *Bowers (2019)*

Are grades more or less subjective than tests? I started this chapter by relating early research that maligns teacher-assigned grades as subjective and unreliable while accepting standardized test scores as objective measures of fundamental academic knowledge. This perspective, promulgated by the testing industry, can lead teachers, principals, and districts to think there is something wrong with grades and focus on standardized assessments. However, just from a logic and efficiency standpoint, if a school already has tests to assess fundamental academic knowledge, why would it need to create another one by aligning grades and tests? As I have described throughout this chapter, grades are not subjective and unreliable; the research is very clear that teachers are quite good at assessing student engagement and persistence through grades. Rather, grades are multidimensional, with about 25 percent of any grade assessing fundamental knowledge and the balance assessing engagement. It is this remaining 75 percent that correlates to overall life outcomes such as graduating from high school and college.

Of course, grades are not perfect, and there is room for improvement in the system. However, as we attempt to clarify the signal and meaning of grades to students and parents, let us remember that a century of research shows assessing engagement to be an important component of grading that is valued by students, parents, schools, employers, and communities.

Recommended Citation:

Bowers, A.J. (2019) Report Card Grades and Educational Outcomes. In Guskey, T., Brookhart, S. (Eds.) *What We Know About Grading: What Works, What Doesn't, and What's Next*, (p.32-56). Alexandria, VA: Association for Supervision and Curriculum Development.

References

- Alexander, K. L., Entwisle, D. R., & Kabbani, N. S. (2001). The dropout process in life course perspective: Early risk factors at home and school. *The Teachers College Record*, 103(5), 760–822.
- Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House*, 78(5), 218-223.
- Allensworth, E. M. (2013). The Use of Ninth-Grade Early Warning Indicators to Improve Chicago Schools. *Journal of Education for Students Placed at Risk (JESPAR)*, 18(1), 68-83. doi: 10.1080/10824669.2013.745181

- Allensworth, E. M., & Easton, J. Q. (2005). *The on-track indicator as a predictor of High School graduation* (Vol. 2006). Chicago: 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.
- Allensworth, E. M., Gwynne, J. A., Moore, P., & de la Torre, M. (2014). *Looking forward to high school and college: Middle grade indicators of readiness in Chicago Public Schools*. Chicago: The University of Chicago Consortium on Chicago School Research.
- Atkinson, R. C., & Geiser, S. (2009). Reflections on a century of college admissions tests. *Educational Researcher*, 38(9), 665, IL 676. doi: 10.3102/0013189x09351981
- 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, IL 235. doi: 10.1080/00461520701621079
- Banker, H. J. (1927). The significance of teachers' marks. *The Journal of Educational Research*, 16(3), 159–171. doi: 10.1080/00220671.1927.10879778
- Barrington, B. L., & Hendricks, B. (1989). Differentiating characteristics of high school graduates, dropouts, and nongraduates. *Journal of Educational Research*, 82(6), 309–319.
- Bisesi, T., Farr, R., Greene, B., & Haydel, E. (2000). Reporting to parents and the community. In E. Trumbull & B. Farr (Eds.), *Grading and reporting student progress in an age of standards* (pp. 157–183). Norwood: Christopher-Gordon Publishers.
- 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
- 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.
- 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
- Bowers, A. J. (2011). What's in a grade? The multidimensional nature of what teacher-assigned grades assess in high school. *Educational Research and Evaluation*, 17(3), 141–159. doi: 10.1080/13803611.2011.597112
- Bowers (2019)
- Bowers, A. J. (2014). Student risk factors. In D. J. Brewer & L. O. Picus (Eds.), *Encyclopedia of Education Economics and Finance*. Thousand Oaks, CA: Sage Publications.
- Bowers, A. J., & Sprott, R. (2012). Examining the multiple trajectories associated with dropping out of high school: A growth mixture model analysis. *Journal of Educational Research*, 105(3), 176–195. doi: 10.1080/00220671.2011.552075
- Bowers, A. J., Sprott, R., & Taff, S. (2013). Do we know who will drop out? A review of the predictors of dropping out of high school: Precision, sensitivity and specificity. *The High School Journal*, 96(2), 77–100.
- 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). Letter: Grading practices and validity. *Educational Measurement: Issues and Practice*, 10(1), 35–36.
- Brooks, F. D. (1929). *The psychology of adolescence*. Oxford, England: Houghton Mifflin.
- Busick, K. (2000). Grading and standards-based assessment. In E. Trumbull & B. Farr (Eds.), *Grading and reporting student progress in an age of standards* (pp. 71–86). Norwood: Christopher-Gordon Publishers.
- Cairns, R. B., Cairns, B. D., & Neckerman, H. J. (1989). Early school dropout: Configurations and determinants. *Child Development*, 60, 1437–1452.
- Carter, R. S. (1952). How invalid are marks assigned by teachers? *Journal of Educational Psychology*, 43(4), 218–228. doi: 10.1037/h0061688
- Casillas, A., Robbins, S., Allen, J., Kuo, Y.-L., Hanson, M. A., & Schmeiser, C. (2012). Predicting early academic failure in high school from prior academic achievement, psychosocial characteristics, and behavior. *Journal of Educational Psychology*, 104(2), 407–420. doi: 10.1037/a0027180
- 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.
- Cliffordson, C. (2008). Differential prediction of study success across academic programs in the Swedish context: The validity of grades and tests as selection instruments for higher education.

Educational Assessment, 13(1), 56–75. doi: 10.1080/10627190801968240

Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, 12(1), 53–72.

Cunha, F., & Heckman, J. J. (2008). Formulating, identifying, and estimating the technology of cognitive and noncognitive skill formation. *Journal of Human Resources*, 43(4), 738–782. doi: 10.3368/jhr.43.4.738

Cusick, P. A. (1983). *The egalitarian ideal and the American high school: Studies of three schools*. New York: Longman.

Dickinson, C. E. (1925). A study of the relation of reading ability to scholastic achievement. *The School Review*, 33(8), 616–626. doi: 10.2307/1078923

Duckworth, A. L., Quinn, P. D., & Tsukayama, E. (2012). What *No Child Left Behind* leaves behind: The roles of IQ and self-control in predicting standardized achievement test scores and report card grades. *Journal of Educational Psychology*, 104(2), 439–451. doi: 10.1037/a0026280

Duckworth, A. L., & Seligman, M. E. P. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, 98(1), 198–208. doi: 10.1037/0022-0663.98.1.198

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.

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.

Evans, F. B. (1976). What research says about grading. In S. B. Simon & J. A. Bellanca (Eds.), *Degrading the grading myths: A primer of alternatives to grades and marks* (pp. 30–50). Washington, DC: ASCD.

Farkas, G., Grobe, R. P., Sheehan, D., & Shuan, Y. (1990). Cultural Resources and School Success: Gender, Ethnicity, and Poverty Groups within an Urban School District. *American Sociological Review*, 55(1), 127–142. doi: 10.2307/2095708

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.

Finn, J. D. (1989). Withdrawing from school. *Review of Educational Research*, 59(2), 117–142.

Fitzsimmons, S. J., Cheever, J., Leonard, E., & Macunovich, D. (1969). School failures: Now and tomorrow. *Developmental Psychology*, 1(134–146).

Bowers (2019)

Folzer-Napier, S. (1976). Grading and young children. In S. B. Simon & J. A. Bellanca (Eds.), *Degrading the grading myths: A primer of alternatives to grades and marks* (pp. 23–27). Washington DC: ASCD.

Gleason, P., & Dynarski, M. (2002). Do we know whom to serve? Issues in using risk factors to identify dropouts. *Journal of Education for Students Placed at Risk*, 7(1), 25–41. doi: 10.1207/S15327671ESPR0701_3

Hargis, C. H. (1990). *Grades and grading practices: Obstacles to improving education and helping at-risk students*. Springfield: Charles C. Thomas.

Heckman, J. J., & Rubinstein, Y. (2001). The importance of noncognitive skills: Lessons from the GED testing program. *The American Economic Review*, 91(2), 145–149. doi: 10.2307/2677749

Heilman, J. D. (1928). The relative influence upon educational achievement of some hereditary and environmental factors *The twenty-seventh yearbook of the National Society for the Study of Education: Nature and Nurture, Part II--Their Influence upon Achievement* (pp. 35–65). Bloomington, IL: Public School Publishing Co.

Henke, R. R., Chen, X., Goldman, G., Rollefson, M., & Gruber, K. (1999). What happens in classrooms? Instructional practices in elementary and secondary schools, 1994–1995. Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Jimerson, S. R., Egeland, B., Sroufe, L. A., & Carlson, B. (2000). A prospective longitudinal study of high school dropouts examining multiple predictors across development. *Journal of School Psychology*, 38(6), 525–549.

Kelly, S. (2008). What types of students' effort are rewarded with high marks? *Sociology of Education*, 81(1), 32–52. doi: 10.1177/003804070808100102

Kertes, F. (1932). Ability grouping in the high school. *The Mathematics Teacher*, 25(1), 5–16.

Kirschenbaum, H., Napier, R., & Simon, S. B. (1971). *Wad-jaget? The grading game in American education*. New York: Hart Publishing Company.

Klapp Lekholm, A. (2011). Effects of school characteristics on grades in compulsory school. *Scandinavian Journal of Educational Research*, 55(6), 587–608. doi: 10.1080/00313831.2011.555923

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
- Labaree, D. F. (1997). Public goods, private goods: The American struggle over educational goals. *American Educational Research Journal*, 34(1), 39–81. doi: 10.3102/00028312034001039
- Labaree, D. F. (2012). *Someone Has to Fail: The Zero-Sum Game of Public Schooling*. Cambridge, MA: Harvard University Press.
- Levin, H. M. (2013). The utility and need for incorporating noncognitive skills into large-scale Educational Assessments. In M. von Davier, E. Gonzalez, I. Kirsch & K. Yamamoto (Eds.), *The role of international large-scale assessments: Perspectives from technology, economy, and educational research* (pp. 67–86). The Netherlands: Springer.
- 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.
- 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.
- 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
- Miner, B. C. (1967). Three factors of school achievement. *The Journal of Educational Research*, 60(8), 370–376. doi: 10.2307/27531890
- Moore, C. C. (1939). The elementary school mark. *The Pedagogical Seminary and Journal of Genetic Psychology*, 54(2), 285–294. doi: 10.1080/08856559.1939.10534336
- Morris, J. D., Ehren, B. J., & Lenz, B. K. (1991). Building a model to predict which fourth through eighth graders will drop out in high school. *Journal of Experimental Education*, 59(3), 286–293.
- Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: How high stakes testing corrupts America's schools*. Cambridge: Harvard Education Press.
- Pallas, A. M. (1989). Conceptual and measurement issues in the study of school dropouts. In K. Namboodiri & R. G. Corwin (Eds.), *Research in the sociology of education and socialization* (Vol. 8, pp. 87–116). Greenwich, CT: JAI Press Inc.
- Pallas, A. M. (2003). Educational transitions, trajectories, and pathways. In J. T. Mortimer & M. J. Shanahan (Eds.), *Handbook of the life course*. New York: Kluwer Academic/Plenum Publishers.
- Parsons, T. (1959). The school class as a social system: Some of its functions in American society. *Harvard Educational Review*, 29(4), 297–318.
- Pattison, E., Grodsky, E., & Muller, C. (2013). Is the sky falling? Grade inflation and the signaling power of grades. *Educational Researcher*, 42(5), 259–265. doi: 10.3102/0013189x13481382
- Quann, C. J. (1983). *Grades and grading: Historical perspectives and the 1982 AACRAO study* (pp. 75). Washington, DC: American Association of Collegiate Registrars and Admissions Officers.
- Randall, J., & Engelhard, 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
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26(7), 1372–1380. doi: <http://dx.doi.org/10.1016/j.tate.2010.03.008>
- Roderick, M., & Camburn, E. (1999). Risk and recovery from course failure in the early years of High School. *American Educational Research Journal*, 36(2), 303–343. doi: 10.3102/00028312036002303
- Ross, C. C., & Hooks, N. T. (1930). How shall we predict high-school achievement? *The Journal of Educational Research*, 22(3), 184–196. doi: 10.2307/27525222
- Rumberger, R. W. (1987). High school dropouts: A review of issues and evidence. *Review of Educational Research*, 57(2), 101–121.
- Rumberger, R. W. (2011). *Dropping out: Why students drop out of high school and what can be done about it*. Cambridge, MA: Harvard University 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. doi: 10.3102/00028312042001003
- Simon, S. B., & Bellanca, J. A. (1976). *Degrading the grading myths: A primer of alternatives to grades and marks*. Washington DC: Association for Supervision and Curriculum Development.
- Pallas, A. M. (1989). Conceptual and measurement issues in the study of school dropouts. In K. Namboodiri & R. G. Corwin (Eds.), *Research in the sociology of education and socialization* (Vol. 8, pp. 87–116). Greenwich, CT: JAI Press Inc.
- Pallas, A. M. (2003). Educational transitions, trajectories, and pathways. In J. T. Mortimer & M. J. Shanahan (Eds.), *Handbook of the life course*. New York: Kluwer Academic/Plenum Publishers.
- Parsons, T. (1959). The school class as a social system: Some of its functions in American society. *Harvard Educational Review*, 29(4), 297–318.
- Pattison, E., Grodsky, E., & Muller, C. (2013). Is the sky falling? Grade inflation and the signaling power of grades. *Educational Researcher*, 42(5), 259–265. doi: 10.3102/0013189x13481382
- Quann, C. J. (1983). *Grades and grading: Historical perspectives and the 1982 AACRAO study* (pp. 75). Washington, DC: American Association of Collegiate Registrars and Admissions Officers.
- Randall, J., & Engelhard, 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
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26(7), 1372–1380. doi: <http://dx.doi.org/10.1016/j.tate.2010.03.008>
- Roderick, M., & Camburn, E. (1999). Risk and recovery from course failure in the early years of High School. *American Educational Research Journal*, 36(2), 303–343. doi: 10.3102/00028312036002303
- Ross, C. C., & Hooks, N. T. (1930). How shall we predict high-school achievement? *The Journal of Educational Research*, 22(3), 184–196. doi: 10.2307/27525222
- Rumberger, R. W. (1987). High school dropouts: A review of issues and evidence. *Review of Educational Research*, 57(2), 101–121.
- Rumberger, R. W. (2011). *Dropping out: Why students drop out of high school and what can be done about it*. Cambridge, MA: Harvard University 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. doi: 10.3102/00028312042001003
- Simon, S. B., & Bellanca, J. A. (1976). *Degrading the grading myths: A primer of alternatives to grades and marks*. Washington DC: Association for Supervision and Curriculum Development.
- Pallas, A. M. (1989). Conceptual and measurement issues in the study of school dropouts. In K. Namboodiri & R. G. Corwin (Eds.), *Research in the sociology of education and socialization* (Vol. 8, pp. 87–116). Greenwich, CT: JAI Press Inc.
- Pallas, A. M. (2003). Educational transitions, trajectories, and pathways. In J. T. Mortimer & M. J. Shanahan (Eds.), *Handbook of the life course*. New York: Kluwer Academic/Plenum Publishers.
- Parsons, T. (1959). The school class as a social system: Some of its functions in American society. *Harvard Educational Review*, 29(4), 297–318.
- Pattison, E., Grodsky, E., & Muller, C. (2013). Is the sky falling? Grade inflation and the signaling power of grades. *Educational Researcher*, 42(5), 259–265. doi: 10.3102/0013189x13481382
- Quann, C. J. (1983). *Grades and grading: Historical perspectives and the 1982 AACRAO study* (pp. 75). Washington, DC: American Association of Collegiate Registrars and Admissions Officers.
- Randall, J., & Engelhard, 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
- Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education*, 26(7), 1372–1380. doi: <http://dx.doi.org/10.1016/j.tate.2010.03.008>
- Roderick, M., & Camburn, E. (1999). Risk and recovery from course failure in the early years of High School. *American Educational Research Journal*, 36(2), 303–343. doi: 10.3102/00028312036002303
- Ross, C. C., & Hooks, N. T. (1930). How shall we predict high-school achievement? *The Journal of Educational Research*, 22(3), 184–196. doi: 10.2307/27525222
- Rumberger, R. W. (1987). High school dropouts: A review of issues and evidence. *Review of Educational Research*, 57(2), 101–121.
- Rumberger, R. W. (2011). *Dropping out: Why students drop out of high school and what can be done about it*. Cambridge, MA: Harvard University 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. doi: 10.3102/00028312042001003
- Simon, S. B., & Bellanca, J. A. (1976). *Degrading the grading myths: A primer of alternatives to grades and marks*. Washington DC: Association for Supervision and Curriculum Development.

Snow, R. E. (1989). Toward assessment of cognitive and conative structures in learning. *Educational Researcher*, 18(9), 8–14. doi: 10.3102/0013189x018009008

Sobel, F. S. (1936). Teachers' marks and objective tests as indices of adjustment. *Teachers College Record*, 38(3), 239–240.

Stanley, G., & Baines, L. (2004). No more shopping for grades at B-Mart: Re-establishing grades as indicators of academic performance. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 77(3), 101–104. doi: 10.1080/00098650409601237

Swineford, F. (1947). Examination of the purported unreliability of teachers' marks. *The Elementary School Journal*, 47(9), 516–521. doi: 10.2307/3203007

Thorsen, C. (2014). Dimensions of norm-referenced compulsory school grades and their relative importance for the prediction of upper secondary school grades. *Scandinavian Journal of Educational Research*, 58(2), 127–146. doi: 10.1080/00313831.2012.705322

Thorsen, C., & Cliffordson, C. (2012). Teachers' grade assignment and the predictive validity of criterion-referenced grades. *Educational Research and Evaluation*, 18(2), 153–172. doi: 10.1080/13803611.2012.659929

Troob, C. (1985). *Longitudinal study of students entering high school in 1979: The relationship between first term performance and school completion*. New York: New York City Board of Education.

Trumbull, E. (2000). Why do we grade--and should we? In E. Trumbull & B. Farr (Eds.), *Grading and reporting student progress in an age of standards* (pp. 23–44). Norwood: Christopher-Gordon Publishers.

Unzicker, S. P. (1925). Teachers' Marks and Intelligence. *The Journal of Educational Research*, 11(2), 123–131. doi: 10.1080/00220671.1925.10879537

Voss, H. L., Wendling, A., & Elliott, D. S. (1966). Some types of high school dropouts. *The Journal of Educational Research*, 59(8), 363–368.

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 2004–05*. Iowa City, IA: ACT, Inc.