INFORMATION CAPSULE

Research Services

Vol. 1613
June 2017

Christie Blazer, Supervisor

COMPARING TEST SCORES AND COURSE GRADES

At a Glance

This Information Capsule examines the relationship between classroom grades and standardized test scores. Studies find there are often discrepancies between the two measures. Reasons why classroom grades and test scores are not more highly correlated are discussed. Research on the strength of the relationship between teacher judgments and test scores, as well as comparisons of course passing rates and test passing rates, is summarized. Finally, studies conducted to determine if students' demographic characteristics influence teacher judgments of academic knowledge and skills are reviewed.

Students' standardized test scores do not always match the grades they earn in school. In most cases, students' course grades suggest higher levels of academic achievement than their test scores. Parents are often confused when students come home with good grades on their report cards, but receive average or below average scores on standardized tests (O'Malley, 2016).

Experts have debated whether course grades or standardized test scores more accurately reflect students' knowledge and skill levels. Some researchers believe that course grades offer a more comprehensive picture of student achievement because they incorporate a broader set of performance factors than standardized tests and provide the ability to gauge whether students are growing intellectually over time (Warsen, 2013; Martínez et al., 2009; Power Score, n.d.).

Other researchers maintain that standardized test scores are a more accurate representation of students' achievement levels because they reflect student performance under the same conditions at a specific point in time. For example, two students scoring 570 on the SAT Mathematics test are believed to have roughly the same aptitude in mathematics, no matter where they live in the U.S. In contrast, it is much more difficult to compare the ability levels of two students earning a "B" in their mathematics course in two different U.S. schools because course grading systems vary widely across districts, schools, and teachers (O'Malley, 2016; Chiekem, 2015; Valdez, 2013; Martínez et al., 2009; Willingham et al., 2000).

Reasons Why Test Scores and Course Grades Do Not Match

Two Different Indicators of Achievement. Researchers have determined that the primary reason course grades and test scores are not more highly correlated is because they measure different indicators of achievement. Standardized test scores measure student knowledge and skills in a specific academic area, such as reading or mathematics, on a particular day. In contrast, students' grades are usually based on a variety of assessments, activities, and behaviors over an extended period of time, such as quizzes, oral and written reports, group projects, homework completion, attendance, classroom participation, diligence, effort, and attitude (O'Malley, 2016; Chiekem, 2015; Warsen, 2013; Godfrey, 2011; Martínez et al., 2009; Gillum et al., 2008; Lekholm & Cliffordson, 2008; Conley, 2000; Willingham et al., 2000; Power Score, n.d.).

Grade Inflation. Many researchers believe that grade inflation is another reason for the mismatch between course grades and standardized test scores. Grade inflation has been defined as the phenomenon of rising grades without a corresponding increase in academic achievement (Zhang & Sanchez, 2013; Godfrey, 2011; Gillum et al., 2008).

Experts have suggested that grade inflation happens for several reasons. First, teachers may want to encourage their students - studies have found that students are more likely to persevere at school when their teachers do not assign them a grade of "F." Second, at the high school level, teachers may not want to hurt students' chances of being accepted to a college or university. Finally, higher grade point averages (GPAs) create the impression that a school's students are achieving at higher levels and that its teachers are more effective compared to other schools in the community (Lindsay, 2015).

Some researchers have presented fairly strong evidence that grade inflation exists in K-12 schools. For example:

- The U.S. Department of Education reported that nationwide GPAs rose steadily from 1990 to 2009 for both male and female high school graduates. The average female GPA increased from 2.8 in 1990 to 3.1 in 2009 (a 0.3 point increase). For males, the average GPA also increased by 0.3 points, from 2.6 in 1990 to 2.9 in 2009 (U.S. News & World Report, 2011).

- Godfrey (2011) compared the SAT scores and cumulative high school GPAs of students from one large state's public school system from 1996 to 2006 to examine general trends in GPAs over time. Findings indicated that students' average GPA rose from 2.6 in 1996 to 2.9 in 2006. Over the same 10-year period, SAT scores remained relatively unchanged. Students' average SAT Verbal score actually decreased by 1.9 points from 1996 to 2006 (from 497.1 to 495.2). On the SAT Mathematics, the average score increased by only 0.7 point (from 497.1 in 1996 to 497.8 in 2006).

- Camara and colleagues (2004) compared SAT scores and self-reported GPAs of high school students over a 26-year period. The researchers included data on eight nationwide cohorts of high school seniors and selected a random sample of approximately 100,000 student records per cohort. They found that the average high school GPA of college-bound seniors increased from 1976 to 2002, regardless of subject area. For example, the average English GPA increased by 0.2 of a grade point; the average mathematics GPA increased by 0.3; the average science GPA increased by 0.3; and the average social studies GPA increased by 0.1. Almost one-half of the overall increase in GPAs occurred over the last four years of the study (between 1998 and 2002). The researchers found that students' SAT Mathematics scores also increased from 1976 to 2002 (from 507 to 516); however, SAT Verbal scores decreased (from 514 to 506) over the same time period.

- Woodruff and Ziomek (2004) studied grade inflation by comparing students' self-reported GPAs and ACT scores. Data included students who graduated from public high schools from 1991 to 2003 and took the ACT in the eleventh or twelfth grades during those years. Sample sizes ranged from approximately 619,000 to 777,000 students each year. The researchers found increases in high school students' self-reported English and mathematics GPAs from 1991 to 2003, while their ACT scores remained relatively stable. Students' average mathematics GPA rose from 2.8 to 3.0, while the average ACT Math score increased by only 0.6 point (on a 36-point scale), from 20.0 in 1991 to 20.6 in 2003. Similarly, the average English GPA rose from 3.0 to 3.3, whereas the average ACT English score only increased by 0.2 point (from 20.2 to 20.4).

In contrast to the studies summarized above, Zhang and Sanchez (2013) found no evidence of systematic grade inflation in U.S. high schools. The researchers examined the self-reported GPAs of public high school graduates who took the ACT in the eleventh or twelfth grade from 2004 to 2011. Students from approximately 12,000 high schools per year were included in the sample. Analyses found no annual change in students' average GPA from 2004 to 2011 on a nationwide basis; however, the researchers reported that grade inflation had occurred to varying degrees at individual high schools.

Some researchers argue that decreases in college entrance exam scores and concomitant increases in GPAs are not evidence of grade inflation. They attribute decreases in average SAT and ACT scores to increases in the number of students taking these tests. From 1986 to 2012, the percentage of students taking the SAT increased by 66% and the percentage of ACT test takers increased by 128%. Generally, when a greater proportion of high school juniors and seniors sit for these exams – in some cases because of state requirements rather than being college-ready - scores decrease (Jaschik, 2016; Lewin, 2013; Kohn, 2002).

According to Alfie Kohn (2002), rising grades don't prove grade inflation exists. He pointed out that no study has definitively shown that the higher grades are undeserved and suggested instead that students may be turning in higher quality assignments or that more recent assessments allow students to better demonstrate their knowledge and skills. Kohn (2002) stated, "No one has ever demonstrated that students today get A's for the same work that used to receive B's or C's. We simply do not have the data to support such a claim."

Additional Factors. Researchers have identified several other factors that may contribute to lack of correspondence between GPAs and standardized test scores.

- Some students suffer from test anxiety during high-stakes assessments. Consequently, their test scores are not an accurate reflection of their achievement levels.
- Some students don't take standardized tests seriously. Their test scores are not a true measure of their knowledge and skills.
- Minority and low-income students may perform at lower levels on standardized tests than their white and higher-income peers due to the racial and class bias that may be present in some tests.
- Teachers may not be aware of the content covered on standardized exams, especially when the tests are new, and fail to cover all parts of the curriculum that will be tested.
- Some classes are not rigorous enough to prepare students adequately for the exams.

- Most teachers are capable of accurately judging their students' performance, but not all teachers are able to assess their students' performance with equal precision and validity. In fact, it has been reported that the majority of teachers in the U.S. receive little or no pre-service training on appropriate grading practices (O'Malley, 2016; North Cross School, 2013; Warsen, 2013; Martínez et al., 2009; Gillum et al., 2008; Shapira, 2006).

Correlations Between Test Scores and Course Grades/Teacher Ratings

As summarized below, most studies have reported moderately positive correlations between standardized assessments and teacher judgments; however, the strength of the relationship has been found to vary considerably among teachers.

- Willingham and colleagues (2000) examined the correlation between tests administered as part of the National Education Longitudinal Study of 1988 (NELS) and students' average GPA in four subjects combined (reading, mathematics, science, and social studies). They found high positive correlations between NELS tests and students' average GPA - 0.79 in standard academic programs and 0.80 in rigorous academic programs. Analysis of NELS test scores and GPAs by individual subject area found the highest correlations in the area of mathematics. Correlations between NELS test scores and GPAs were 0.77 in mathematics, 0.68 in reading, and 0.66 in both science and social studies.

  Willingham and colleagues' (2000) examination of the relationship between NELS test scores and combined GPAs by ethnicity found the highest correlation between the two measures among Asian students (0.84), followed by 0.80 for White students, 0.76 for Black students, and 0.72 for Hispanic students.

- Warsen (2013) examined the relationship between course grades and ACT test scores for four years of graduating students in two Michigan high schools. He found moderately positive correlations between GPAs and ACT scores for each graduating class studied over a four-year period - 0.69 in mathematics and 0.58 in English at one school, and 0.54 in mathematics and 0.60 in English at the other school.

- The College Board conducted a study to investigate the correlations between Advanced Placement (AP) course grades and AP exam scores in five subject areas at the five largest high schools in one large state public school system. Results indicated that correlations varied widely, depending on the school and subject area. Correlations between AP course grades and AP exam scores were highest in U.S. History (ranging from 0.65 to 0.77, depending on the school) and lowest in English Literature (0.23 to 0.56). Correlations between AP course grades and AP exam scores ranged from 0.29 to 0.77 in Biology, 0.45 to 0.63 in Calculus AB, and 0.35 to 0.60 in English Language (Godfrey, 2011).

- Südkamp and colleagues (2012) conducted a meta-analysis of 75 studies that reported correlational data on the relationship between teachers' judgments of students' academic achievement and students' performance on standardized achievement tests.

The researchers found that the correlation between teachers' judgments of students' academic achievement and students' actual test performance in language arts and mathematics was moderately positive (0.63).

- Valdez (2013) obtained data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K), a study that followed over 22,000 nationally representative kindergarten children through eighth grade. Valdez' (2013) study tracked students in the ECLS-K dataset through third grade. Teacher judgments on the Language and Literacy items of the ECLS-K's Academic Rating Scale were compared to students' scores on the ECLS-K reading assessment. Results indicated that teacher judgments of students' language and literacy skills were moderately to highly associated with students' reading test scores at all four waves of data collection (0.58 for fall kindergarten, 0.62 for spring kindergarten, 0.71 for spring first grade, and 0.64 for spring third grade).

- Hinnant and colleagues (2009) followed children born at 10 different geographic sites from birth to fifth grade. Teachers rated students' reading and mathematics ability at grades 1, 3, and 5 using two Academic Skills questionnaires adapted from the ECLS-K – the Language and Literacy scale and the Mathematical Thinking scale. Students were administered two subtests from the Woodcock-Johnson Psycho-Educational Battery – Letter-Word Identification and Applied Problems. The researchers found that teacher ratings were moderately related to Woodcock-Johnson test scores for both Letter-Word Identification (0.64 for first grade, 0.67 for third grade, and 0.53 for fifth grade) and Applied Problems (0.54 for first grade, 0.57 for third grade, and 0.56 for fifth grade).

- Martínez and colleagues (2009) also analyzed data from the ECLS-K. Their study focused on mathematics achievement in grades 3 and 5. The researchers found that the correlations between the ECLS-K mathematics assessment and teacher judgments on the Mathematical Thinking items of the Academic Rating Scale were moderately positive in both grades 3 and 5 (0.58 and 0.64, respectively).

  Martínez and colleagues' (2009) study also found that the strength of the correlation between mathematics test scores and teacher ratings differed substantially between teachers. Higher correlations were found when teachers:

  - Reported placing more importance on student effort (grade 3); and
  - Reported assigning more importance to the completion of homework (grade 5).

  Lower correlations were found when teachers:

  - Had higher proportions of students who were non-native English speakers;
  - Assigned more importance to assessing student achievement relative to state or local standards;
  - Placed more emphasis on student behavior or conduct;
  - Valued classroom participation more highly;
  - Conducted frequent classroom assessments, such as teacher-made quizzes and textbook chapter-end tests; and

- Spent more time on test preparation activities.

Comparison of Course Passing Rates and Test Passing Rates

Studies comparing course passing rates and test passing rates have produced inconsistent results. In some cases, course passing rates have far exceeded passing rates on corresponding subject area standardized tests, while strong alignment between passing scores on courses and subject area exams has also been reported.

- The Texas Education Agency (2009) compared the passing rates of grade 10 students on Texas Assessment of Knowledge and Skills (TAKS) mathematics tests with their passing rates in Algebra 1 and geometry courses. They found that students passed their mathematics courses at higher rates than they passed the TAKS mathematics test – 88% of students enrolled in Algebra 1 passed the course, while only 56% passed the TAKS mathematics test. Similarly, 86% of students enrolled in geometry passed the course, yet only 67% passed the TAKS mathematics test.

- The Governor’s Office of Student Achievement in Georgia compared students’ state test scores with their corresponding course grades. The study found that more students failed the state’s End of Course Tests (EOCT) than failed their courses. For example, in algebra, 39% of students failed the EOCT, but only 20% failed the course; in geometry, 39% failed the EOCT, but only 14% failed the course; and in ninth grade English Literature, 32% failed the EOCT, but only 16% failed the course (Gelpi, 2009; Georgia Governor’s Office of Student Achievement, 2009).

- Compared to the two studies summarized above, Boykin (2010) reported stronger alignment between course and standardized exam passing rates. She examined the relationship between End of Course (EOC) test results and course grades in North Carolina’s Wake County Public School System. All ninth grade students with both an EOC test score and a course grade in Algebra I and/or English I during 2004-2005 through 2007-2008 were included in the study. For both subjects, findings indicated that the majority of students passed both the EOC exam and the course (82% to 92% of students in Algebra I, depending on the school year; 83% to 91% of students in English 1, depending on the school year). Only 3% to 4% of students passed the Algebra course, but failed the corresponding EOC exam; only 2% to 7% of students passed the English course, but failed the EOC exam.

Relationship Between Student Characteristics and Course Grades Assigned by Teachers

Studies have produced mixed findings on whether students’ demographic characteristics affect teachers’ judgments of their academic proficiency.

- Lekholm and Cliffordson’s (2008) study of almost 100,000 Swedish ninth graders found that students attending schools where the majority of students were from families with lower educational backgrounds received somewhat higher grades. The researchers hypothesized that teachers tried to raise students' grades because of the negative consequences of low grades both for the individual student and the school.

- Hinnant and colleagues (2009) followed close to 3,000 students from 10 different geographic sites from birth to fifth grade. The researchers found that students' gender and social skills were consistent predictors of teacher expectations of reading ability and, to a lesser extent, mathematics ability. Results indicated that teachers judged girls and socially skilled students as more academically competent than they judged boys and students who were less socially skilled.

- Valdez (2013) obtained data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99, and followed the nationally representative group of students through their third year of elementary school. The researchers found that teachers' judgments of students' language and literacy skills were affected by students' socioeconomic status (SES), but that SES biased teacher judgments differently for early (kindergarten/first grade) and later (third grade) elementary levels of instruction. At kindergarten and first grade, teacher judgments of reading performance for highly proficient students were more influenced by students' SES; judgments of reading performance for low proficiency students were less influenced by students' SES. However, at grade 3, the reverse was found – SES had a greater influence on teachers' judgments of students who were at lower reading proficiency levels. Valdez (2013) also found that years of teaching experience did not affect teachers' ability to reliably judge students' language and literacy skills.

- In contrast to the studies summarized above, Willingham and colleagues (2000) reported that school grading standards were unrelated to the personal or background characteristics of students. Their analysis of data from the National Education Longitudinal Study of 1988 found no differences in the strictness of grading policies between schools with students from mostly lower-income and minority families and schools comprised of mostly higher-income and white families.

- Similarly, Zhang and Sanchez (2013) analyzed data from a nationwide sample of eleventh and twelfth grade students who took the ACT college admissions test. Although they found evidence of variations in grading standards across high schools, these variations were not attributable to the school's percentage of students who were eligible for free or reduced price lunch or the proportion of minority students.

Summary

This Information Capsule examined the relationship between course grades and standardized test scores. Researchers have found that there are often discrepancies between the two measures. In most cases, students' course grades suggest higher levels of academic achievement than their test scores.

Researchers have determined that the primary reason course grades and test scores are not more highly correlated is that they measure different indicators of achievement. Grade inflation is another reason for the frequent mismatch between course grades and standardized test scores. Other contributing factors may include test anxiety that artificially deflates test scores, instruction that is not rigorous enough to prepare students adequately for standardized tests, and variability in teachers' capacity to accurately judge students' knowledge and skills.

Most studies have found moderately positive correlations between standardized test scores and classroom grades. However, due to inconsistency in grading systems across districts, schools, and teachers, the strength of the relationship varies considerably.

Studies comparing course passing rates and test passing rates have reported inconsistent results. In some cases, course passing rates have far exceeded passing rates on corresponding subject area standardized tests, but other studies have found strong alignment between passing scores on courses and subject area exams.

Finally, studies have produced mixed findings on whether students' demographic characteristics affect teachers' judgments of academic knowledge and skills. Some studies have found that students' income level, gender, social skills, and parents' education level affect teacher judgments of academic proficiency, while other studies have concluded that teacher judgments are unrelated to students' personal or background characteristics.

References

Boykin, A-S.M. (2010). *The Relationship Between High School Course Grades and Exam Scores.* Evaluation and Research Department Report No. 09.39, Wake County Public School System, Cary, NC. ERIC Document Reproduction Service No. ED564393.

Camara, W., Kimmel, E., Scheuneman, J., & Sawtell, E.A. (2004). *Whose Grades Are Inflated?* College Board Research Report No. 2003-4, College Entrance Examination Board, New York, NY. ERIC Document Reproduction Service No. ED563203.

Chiekem, E. (2015). Grading Practice as Valid Measures of Academic Achievement of Secondary Schools Students for National Development. *Journal of Education and Practice, 6*(26), 24-28. ERIC Document Reproduction Service No. EJ1077389.

Conley, D.T. (2000). *Who is Proficient: The Relationship between Proficiency Scores and Grades.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 2000. ERIC Document Reproduction Service No. ED445025.

Gelpi, G. (2009). Test Scores Don't Match Class Grades. *The Augusta Chronicle,* February 4, 2009. Retrieved from http://chronicle.augusta.com/metro/2009-02-04/test-scores-dont-match-class-grades.

Georgia Governor's Office of Student Achievement. (2009). *Comparison of End-of-Course Test Results and Course Grades – State Level Summary.* Retrieved from http://gosa.georgia.gov/sites/gosa.georgia.gov/files/related_files/document/EOCT_Grading_Alignment_State_Results.pdf.

Gillum, J., Rivera, A., Sánchez, G.B., & Younger, J. (2008). Grade Inflation Adds to Woes, Especially in Middle Schools. *Arizona Daily Star,* May 12, 2008. Retrieved from http://tucson.com.

Godfrey, K.E. (2011). *Investigating Grade Inflation and Non-Equivalence.* College Board Research Report No. 2011-2, College Entrance Examination Board, New York, NY. Retrieved from https://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2011-2-investigating-grade-inflation-non-equivalence.pdf.

Hinnant, J.B., O'Brien, M., & Ghazarian, S.R. (2009). The Longitudinal Relations of Teacher Expectations to Achievement in the Early School Years. *Journal of Educational Psychology, 101*(3), 662-670.

Jaschik, S. (2016). ACT Scores Drop as More Take Test. *Inside Higher Ed*, August 24, 2016. Retrieved from https://www.insidehighered.com/news/2016/08/24/average-act-scores-drop-more-people-take-test.

Kohn, A. (2002). *The Dangerous Myth of Grade Inflation.* Retrieved from http://www.alfiekohn.org/article/dangerous-myth-grade-inflation/.

Lekholm, A.K., & 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.

Lewin, T. (2013). More Students Are Taking Both the ACT and SAT. *The New York Times,* August 2, 2013.

Lindsay, S. (2015). SAT/ACT Prep Online Guides and Tips: Is Grade Inflation in High School Real? *Prep Scholar,* October 20, 2015. Retrieved from http://blog.prepscholar.com/grade-inflation-high-school.

Martínez, J.F., Stecher, B., & Borko, H. (2009). Classroom Assessment Practices, Teacher Judgments, and Student Achievement in Mathematics: Evidence from the ECLS. *Educational Assessment, 14*(2), 78-102.

North Cross School. (2013). *When Standardized Tests and Grades Don't Match Up.* Retrieved from http://www.northcross.org/crossties-blog/when-standardized-tests-and-grades-dont-match-up.

O'Malley, K. (2016). Why Good Grades Don't Always Match Good Test Scores, *Noodle,* January 7, 2016. Retrieved from http://www.noodle.com/articles/why-good-grades-dont-always-match-good-test-scores.

Power Score. (n.d.). *SAT vs. GPA.* Retrieved from https://www.powerscore.com/sat/help/sat_vs_gpa.cfm.

Shapira, I. (2006). Those Who Pass Classes But Fail Tests Cry Foul. *The Washington Post,* November 21, 2006.

Südkamp, A., Kaiser, J., & Möller, J. (2012). Accuracy of Teachers' Judgments of Students' Academic Achievement: A Meta-Analysis. *Journal of Educational Psychology, 104*(3), 743-762.

Texas Education Agency. (2009). *A Study of the Correlation Between Grade 10 Math Performance and Course Performance.* Retrieved from http://tea.texas.gov/WorkArea/DownloadAsset.aspx?id=2147494123.

U.S. News & World Report. (2011). *Average High School GPAs Increased Since 1990.* April 19, 2011. Retrieved from http://www.usnews.com/opinion/articles/2011/04/19/average-high-school-gpas-increased-since-1990.

Valdez, A. (2013). Teacher Judgment of Reading Achievement: Cross-Sectional and Longitudinal Perspective. *Journal of Education and Learning, 2*(4), 186-200. ERIC Document Reproduction Service No. EJ1077262.

Warsen, G.D. (2013). *Making Grades Matter: Connections Between Teacher Grading Practices and Attention to State Assessment.* Doctoral Dissertation, Western Michigan University, Kalamazoo, MI. Retrieved from http://scholarworks.wmich.edu/cgi/viewcontent.cgi?article=1158&context=dissertations.

Willingham, W.W., Pollack, J.M., & Lewis, C. (2000). *Grades and Test Scores: Accounting for Observed Differences.* Educational Testing Service Research Report, September 2000. Retrieved from http://www.ets.org/Media/Research/pdf/RR-00-15-Willingham.pdf.

Woodruff, D.J., & Ziomek, R.L. (2004). *High School Grade Inflation from 1991 to 2003.* ACT Research Report Series 2004-4. ERIC Document Reproduction Service No. ED484784.

Zhang, Q., & Sanchez, E.I. (2013). *High School Grade Inflation from 2004 to 2011.* ACT Research Report Series 2013(3). ERIC Document Reproduction Service No. ED555577.