

Running Head: WHAT NO CHILD LEFT BEHIND LEAVES BEHIND

What *No Child Left Behind* Leaves Behind: The Roles of IQ and Self-Control in Predicting
Standardized Achievement Test Scores and Report Card Grades

Angela L. Duckworth¹
Patrick D. Quinn²
Eli Tsukayama¹

Positive Psychology Center

University of Pennsylvania
3701 Market St. Suite 200
Philadelphia, PA 19104
Phone: 215 898-1339
Fax: 215 573-2188

The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B090015 of the U.S. Department of Education.

Primary contact email: duckworth@psych.upenn.edu

¹ Positive Psychology Center, University of Pennsylvania

² Department of Psychology, The University of Texas at Austin

Abstract

The increasing prominence of standardized achievement tests in educational policy and practice motivated the current investigation. We propose a model in which report card grades differentially assess certain competencies (e.g., completion of homework and long-term assignments and positive classroom behavior) determined more by self-control than by intelligence, whereas standardized achievement tests are more sensitive to other competencies (e.g., capacity to solve novel problems in unfamiliar formats; independent acquisition of skills and knowledge outside of school) determined more by intelligence than by self-control. In three longitudinal, prospective studies of middle school students, self-control was a stronger predictor of changes in report card grades than was IQ. In contrast, IQ was a stronger predictor than self-control of changes in standardized achievement test scores. In a fourth study, a survey of teachers' grading practices provided further support for the proposed model. Implications for pedagogy and public policy are discussed.

Keywords: Impulsivity, Self-Control, Achievement, Success, Personality

What *No Child Left Behind* Leaves Behind: The Roles of IQ and Self-Control in Predicting Standardized Achievement Test Scores and Report Card Grades

On January 8, 2002, George W. Bush signed into law the *No Child Left Behind (NCLB) Act*. For the first time in U.S. history, NCLB made federal funding for K-12 public schools contingent upon the use of standardized achievement tests to assess student performance. As a consequence, all fifty states now administer annual standardized achievement tests to benchmark student performance. Absent any legislative mandate, many private schools are following suit. The crucial advantage of standardized achievement tests—and the *raison d'être* for their increasing importance in American education—is that they enable objective, apples-to-apples comparisons of students across classrooms and schools. Policymakers have for this reason considered standardized achievement tests a valid metric by which to hold schools, teachers, and students accountable for academic progress.

Researchers at Educational Testing Service (ETS), which administers over 50 million standardized achievement tests annually, recently noted the “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*” (emphasis added, Willingham, Pollack, & Lewis, 2002, p. 2). Both report card grades and standardized achievement test scores are designed to gauge students’ academic skills and knowledge. However, report card grades and standardized achievement test scores do not rank students identically (i.e., correlations between these measures are large but do not approach unity) (Geiser & Santelices, 2007; Nofle & Robins, 2007; Sharpley & Edgar, 1986; Willingham, 1985; Willingham, et al., 2002). Perhaps the simplest explanation for less than perfect correlations is that report card grades are affected

by local grading standards and individual teacher bias and, thus, less accurate measures of what students know and can do (Cross & Frary, 1996; Jussim, 1991).

Alternatively, we propose that standardized achievement test scores and report card grades differentially reflect intelligence and self-control, two distinct traits shown in prior studies to predict successful functioning in—and beyond—the classroom. In three prospective, longitudinal studies, we compared the predictive validity of intelligence and self-control for standardized achievement test scores and report card grades in a private middle school (Study 1), two public middle schools (Study 2), and a national sample of private and public school students followed from fourth to ninth grade (Study 3). In Study 4, we surveyed middle school teachers about the content, format, and purpose of assessments that contribute to their students' report card grades and standardized achievement test scores. Collectively, our findings support a model that distinguishes between competencies differentially influenced by self-control and better assessed by report card grades (i.e., completion of homework and long-term projects; positive classroom participation, attitude, effort, and attendance) and, on the other hand, competencies differentially influenced by intelligence and better assessed by standardized achievement test scores (i.e., skills and knowledge acquired outside of the classroom, ability to solve novel problems in novel formats).

Where Report Card Grades and Standardized Achievement Test Scores Differ

Willingham and colleagues (2002) identified four substantive sources of discrepancy between standardized achievement test scores and report card grades. First, the skills and knowledge assessed by standardized achievement tests are unlikely to be perfectly aligned with what teachers have taught their students. Notwithstanding state standards for major academic subjects, developers of standardized achievement tests and local teachers may

disagree about the relative importance of various components of the recommended curriculum. Likewise, standardized achievement tests and textbooks may be misaligned. In an analysis of five standardized achievement tests and four textbooks, Freeman et al. (1983) found that “the proportion of topics presented on a standardized test that received more than cursory treatment in each textbook was never higher than 50%” (p. 509). Because of this misalignment, skills and knowledge students acquire outside of class (i.e., by learning on their own) may disproportionately benefit students on standardized achievement tests (Popham, 1999, 2000). Misalignment should in addition disproportionately benefit students who are more adept at solving novel problems (i.e., problems they have never seen before) (Popham, 1999, 2000).

Second, assessment format may differ. Where there are format differences, students should be more familiar with the particular assessment format their teachers favor for quizzes and tests taken throughout the year. It seems likely to us that teachers tend to write open-ended (e.g., short answer or essay) questions more often than multiple choice questions, whereas standardized achievement tests tend to include more multiple choice questions than open-ended questions. Thus, while contemporary standardized achievement tests tend to include items of both formats, it is nevertheless likely that they are relatively more novel in format than teacher-created assessments.

A third source of discrepancy between report card grades and standardized achievement tests concerns the time period and context in which students are asked to perform. Standardized achievement tests assess knowledge and skills on a single occasion, over the course of a few hours, in a situation designed to minimize distraction and maximize motivation. Report card grades, on the other hand, typically reflect performance on many quizzes and tests for which students have been given repeated opportunities to study, as well as homework assignments and

long-term projects. Given the span of time over which these numerous assessments take place and the fact that preparation for these assessments often takes place at home, it is not possible for teachers to optimize conditions for performance. Rather, students must—without teachers' assistance—regulate their effort to complete the required work, not in a quiet room free of distraction, but in environments in which television, text messages, and other diversions likely compete for students' attention.

Finally, in their calculation of report card grades, teachers may directly factor in aspects of classroom conduct that are distinct from academic knowledge and skills. For instance, teachers might give credit or deductions for attendance, class participation, disruptive classroom behavior, and completion of homework assignments. These aspects of classroom conduct can also influence standardized achievement test scores indirectly via the acquisition of skills and knowledge taught in school, but they cannot directly affect standardized achievement test scores.

In Figure 1, we summarize these four distinct factors and, further, propose relationships with standardized achievement test scores and report card grades. Specifically, our model suggests that skills and knowledge formally taught in school determine both standardized achievement test scores and report card grades, though, as Willingham et al. (2002) suggest, they are better measured by the latter. Second, our model posits that skills and knowledge acquired *outside* of school and, in addition, the ability to solve novel problems in novel formats differentially contribute to standardized achievement test scores. We suggest that, in contrast, the completion of homework and long-term projects as well as classroom participation, attitude, effort, and attendance differentially contribute to report card grades. Finally, we suggest that intelligence and self-control differentially influence these four competencies, which in turn explains the differential impact of these traits on report card grades and standardized

achievement test scores.

We propose that differences between standardized achievement tests and report card grades suit these two forms of assessment to distinct educational purposes. Report card grades seem better suited to providing timely feedback to students about subject matter they have and have not yet mastered. Further, report card grades can motivate students to comply with teacher directives, enforcing an “implicit local contract between teacher and student” whereby desired conduct (e.g., participating productively in class discussion, completing homework thoroughly and on time) are rewarded with higher report card grades (Willingham et al., 2002, p. 28). In contrast, standardized achievement tests seem better suited to sampling what students know and can do in an academic subject, regardless of whether the relevant knowledge and skills were acquired in school (Popham, 1999, 2000). Moreover, because local teachers have no direct control over their design or grading, standardized achievement tests provide an “external standard that is intended to compare performance” of students to one another (Willingham et al., 2002, p. 28).

Intelligence and Standardized Achievement Test Scores

Intelligence, whose central feature is the “ability to learn, reason, and solve problems” (Gottfredson, 2004, p. 35), is the best-documented predictor of standardized achievement test scores (Barton, Dielman, & Cattell, 1972; Brody, 1997; Deary, Strand, Smith, & Fernandes, 2007; Gagné & St Père, 2002; Gottfredson, 2004; Neisser et al., 1996). Correlations between IQ and standardized achievement test scores are usually interpreted as evidence that smarter students learn faster: Given the same opportunity to learn, smarter students presumably accumulate superior levels of skills and knowledge, which in turn produce both higher standardized achievement test scores and higher report card grades (Neisser, et al., 1996). This

sequence is captured in our proposed model (see Figure 1).

We suggest two additional pathways that further increase the positive association between intelligence and standardized achievement test scores. First, *ceteris paribus*, more intelligent children likely acquire skills and knowledge *outside of the classroom* at higher rates than their peers (Gottfredson, 2002). Children read books, interact with other people, and observe the world for more hours outside of the classroom than they do in the classroom. In our model, independent learning outside the context of formal instruction disproportionately improves intelligent students' standardized achievement test scores relative to their report card grades because the former tend to be less aligned with what teachers actually teach in the classroom. Second, more intelligent children are better at solving relatively novel problems in relatively unfamiliar formats (Salthouse & Pink, 2008). In sum, the ability to solve problems not formally taught to them by their teachers and presented in relatively novel formats should disproportionately advantage more intelligent students on standardized achievement tests.

Self-Control and Report Card Grades

Self-control,¹ defined as the ability to regulate behaviors, thoughts, emotions, and attention in the service of valued long-term goals, is among the most robust predictors of consequential life outcomes (Baumeister, Vohs, & Tice, 2007; Mischel, Shoda, & Peake, 1988). Several prospective, longitudinal studies have shown that self-control predicts teacher-assigned course grades at every level of schooling from elementary school to college (Duckworth & Seligman, 2005; Duckworth, Tsukayama, & May, 2010; McDermott, Mordell, & Stoltzfus, 2001; Robbins, Allen, Casillas, Peterson, & Le, 2006; Wulfert, Block, Ana, Rodriguez, & Colsman, 2002; Zhou, Main, & Wang, 2010). In studies employing omnibus measures of personality, self-control has emerged as a stronger predictor than any other trait of course grades

(Robbins, et al., 2006; Wolfe & Johnson, 1995). Moreover, self-control is a facet of Big Five Conscientiousness (Costa & McCrae, 1992), the family of personality traits which meta-analytic analyses have shown to predict GPA far better than other personality traits and about as well as does intelligence (Poropat, 2009).

Why do self-controlled children earn better grades than their more impulsive peers? In a series of lectures addressed to Boston schoolteachers, William James (James, 1899), opined that in “schoolroom work” there is inevitably “a large mass of material that must be dull and unexciting” (pp. 104-105). We agree – while learning is certainly intrinsically rewarding, it is also, at times, boring, tedious, and frustrating relative to rival pursuits (e.g., chatting with friends, daydreaming, watching television). Therefore, the ability to choose long-term rewards over immediate, more pleasant diversions is crucial to learning through formal instruction. Consistent with this view, a study of talented high school students (Wong & Csikszentmihalyi, 1991) found that even students identified as highly talented reported being less happy when studying than when engaged in leisure activities, leading the authors to conclude, “Students study hard not so much because they are intrinsically motivated or happy in their work, but because they want to achieve certain long-term goals such as getting good grades” (p. 563). We reason that because both report card grades and standardized achievement test scores reflect mastery of skills and knowledge taught in class, self-control should be positively related to both forms of assessment. Our model in Figure 1 captures this pathway.

However, prior studies have shown that many teachers also *directly* factor in aspects of student conduct, including punctual and complete homework assignments and class participation (Brookhart, 1994; Cizek, Fitzgerald, & Rachor, 1995; Cross & Frary, 1996; McMillan, Myran, & Workman, 2002). We have argued above, as have Willingham et al. (2002), that these

dimensions are not direct inputs to standardized achievement test scores. To the extent that self-control facilitates positive student conduct, therefore, it may increase report card grades over and beyond the indirect effects of these behaviors on skills and knowledge formally taught in school.

The Current Investigation

In this investigation, we examined whether report card grades and standardized achievement tests are differentially sensitive to self-control and intelligence. In three separate longitudinal studies, we tested the following specific hypotheses: (1) Self-control is a better predictor than IQ of final report card grades and one-year improvements in report card grades (in Studies 1, 2, and 3); (2) IQ is a better predictor than self-control of standardized achievement test scores and improvements in standardized achievement test scores (in Studies 1, 2, and 3); (3) the effect of self-control on improvements in report card grades is mediated by superior teacher ratings of student conduct (including homework completion and classroom behavior) during the school year (in Study 2); and, finally (4) teachers perceive substantial differences between report card grades and standardized achievement tests in terms of their content, format, and purpose (in Study 4).

Our focus in the current investigation is on middle school. This developmental period is of particular interest for several reasons. First and foremost, middle and high school teachers are much more likely than elementary school teachers to use formal assessments (e.g., paper-and-pencil quizzes and exams), as opposed to informal observation, when determining report card grades (Brookhart, 1994; Gullickson, 1985). Additionally, as children enter middle school, academic performance becomes an increasingly important component of students' personally valued goals and overall self-esteem (Galotti, 2005; Harter, 1985). At the same time, children begin to differentiate between work habits (e.g., self-controlled behavior) and intellectual

competence (Stipek & Douglas, 1989). In sum, middle school represents a developmental epoch during which assessment of student performance transitions from informal to formal and, at the same time, students become more aware of and concerned with their academic performance.

Three strengths of our study design strengthen our ability to make valid inferences and are worth mentioning at the outset. First, following recommended best practices in personality assessment (Eid & Diener, 2006), we increased the reliability and validity of self-control measures in Studies 1, 2, and 3 by using ratings by teachers, parents, and/or the students themselves. In contrast to studies in which students are only rated by their teachers (McDermott, et al., 2001), our multimethod approach to measurement reduced the potential artifact of common method variance (i.e., the inflation of correlations between GPA and self-control ratings by their common rater, the teacher). Second, we maximized internal validity in Studies 1, 2, and 3 by using a prospective longitudinal design and estimating the effect of self-control and IQ on report card grades and standardized achievement test scores when controlling for baseline measures of these outcomes. Third, the children in Studies 1, 2, and 3 were demographically distinct and collectively represented an extremely diverse population in terms of ethnicity, socioeconomic status, geographic location in the United States, and public/private school attendance. Replication across these three samples increases the external validity of our findings.

Study 1

Method

Participants. Participants were sixth through eighth grade students at a private middle school in New York City. About 81% of the school's 247 students elected to participate and were not significantly different from non-participants in terms of gender, race, age at assessment, or household income. Of the 200 consented students, 61 were omitted from analyses due to

listwise deletion for missing data (final $N = 139$ participants, mean age = 12.94 years, $SD = 0.89$). Seventy-three percent of participants were White, 8% were Asian, 6% were Black, 4% were Latino, and 9% were other ethnicities; 53% were female. The estimated median annual household income for this sample was \$76,191.

Procedure and measures. Students, parents, and teachers completed consent forms and questionnaires within the first three months of the school year. Students and teachers completed measures at school. At the conclusion of the school year, student demographic variables and outcome data were collected from school records. We used home addresses in conjunction with U.S. Census Bureau data to estimate household income for each participant.

Self-control. Homeroom teachers, parents, and students completed the Impulsivity Scale for Children with students as targets (ISC; Duckworth, Tsukayama, & Kim, 2010). The ISC questionnaire lists eight behaviors nominated by middle school students and endorsed by public and private teachers as indicating lapses in self-control (e.g., “This student's mind wandered when he or she should have been listening,” “This student interrupted other people while they were talking”). Items were endorsed using a 5-point frequency scale, whose valence we adjusted such that higher scores indicated higher levels of self-control: 1 = *at least once a day*, 2 = *about once a week*, 3 = *about 2 to 3 times a month*, 4 = *about once a month*, and 5 = *almost never*. The observed internal consistency coefficients were $\alpha = .82$, $.83$, and $.87$ for self-report, parent, and teacher ratings, respectively.

IQ. Students completed Raven's Progressive Matrices (Raven, 1948), a widely-used nonverbal test of intelligence. The test comprises a series of 60 matrices, each of which has one element missing. The task in each case is to select from a set of alternatives the piece that completes the pattern correctly. Students were given as much time to finish as they needed; all

finished within 45 minutes. Because published age-related population norms are not available for Raven's Progressive Matrices, we regressed raw scores on participant age and saved the standardized residuals, which we then used as age-corrected IQ scores.

Report card grades. School records included separate fall and spring semester grades for math, science, English, and social studies classes. Confirmatory factor analysis generally indicated good fit for a single-factor model of academic grades, $\chi^2(2) = 10.93$, $p = .004$, CFI = .97, RMSEA = .18, SRMR = .03. Although the RMSEA was greater than expected, this indication of poor fit may have resulted from small model size (Kenny & McCoach, 2003) and large factor loadings (Browne, MacCallum, Kim, Andersen, & Glaser, 2002; Miles & Shevlin, 2007), rather than actual model misspecification. Distinguishing math and science (i.e., quantitative subjects) from English and social studies (i.e., verbal subjects) grades in a two factor model did not significantly improve fit, $\Delta\chi^2(1) = 2.95$, $p = .09$, CFI = .97, RMSEA = .23, SRMR = .03. We therefore calculated each student's grade point average (GPA) for the fall and spring semesters by averaging math, science, English, and social studies grades for each semester, respectively. We calculated cumulative GPA as the average of fall and spring semester GPAs.

Standardized achievement test scores. Students completed the Comprehensive Testing Program test (CTP) in the spring of 2008 and 2009. The CTP is a widely used, nationally-normed assessment of student achievement. We collected scores for all administered subtests (i.e., verbal reasoning, vocabulary, quantitative reasoning, and reading comprehension) from school records.² The CTP verbal reasoning, vocabulary, and quantitative reasoning subtests are entirely multiple-choice in format. The CTP reading comprehension test includes 40 multiple-choice questions answered in 40 minutes and 4 open-ended questions answered in 30 minutes.

In confirmatory factor analysis, a single-factor measurement model with each subtest

as an observed variable fit very well, $\chi^2(2) = 1.63$, $p = .92$, CFI = 1.00, RMSEA = .00, SRMR = .01, and no possible two-factor model resulted in improved fit, all $\Delta\chi^2(1) < 1.32$, $ps > .24$. We therefore averaged the test scores to create composite standardized achievement test scores for 2008 and 2009. Test scores from the spring of 2008 were unavailable for participants ($n = 15$) who joined the school in the fall of 2008. All analyses including 2008 standardized achievement test scores as a covariate therefore have an included sample size of $n = 126$.

Results

Creation of composite self-control scores. Self-report and teacher ratings of self-control were correlated, $r = .25$, $p < .01$. Likewise, parent and teacher ratings of self-control were correlated, $r = .24$, $p < .05$. These associations are comparable to meta-analytic estimates of $r = .20$ between self-report and teacher ratings of children's behavior and $r = .28$ between parent and teacher ratings of behavior (Achenbach, McConaughy, & Howell, 1987). Self-report and parent ratings of self-control were not significantly correlated, $r = .14$, $p = .22$, an association lower than (but within about one standard error of) a meta-analytic estimate of self-report and parent ratings of children's behavior, $r = .25$ (Achenbach, et al., 1987). Generally, moderate cross-informant correlations indicate that behavior demonstrates both consistency and variation across situations (Achenbach, et al., 1987). For instance, it is possible for a child who rarely interrupts her classmates at school to frequently interrupt her siblings during family conversations, but on average, individuals who interrupt others in one context are somewhat more likely to do so in others (Borkenau, Mauer, Riemann, Angleitner, & Spinath, 2004; Epstein, 1979).

Because in the current investigation we were interested in assessing trait-level self-control (i.e., the tendency to act self-controlled, generally, in any given situation), we followed

current recommendations to use information from all three sources (see Achenbach et al., 1987; Eid & Diener, 2006). Specifically, we created a composite measure of self-control by averaging standardized scores for self-report, parent, and teacher ratings. For participants without parent data ($n = 58$), we created composite scores using self- and teacher-report data only. Following the procedures outlined by Nunnally (1978), the internal reliability of the composite was .89. We used this composite measure of self-control in all subsequent analyses.

Examination of continuous variable distributions. All continuous variables met the distributional assumptions of the general linear model, $|\text{skewness}| < .89$, $|\text{kurtosis}| < .78$. Kline (2005) notes that absolute values of the skew index greater than 3.0 is considered extreme and that a kurtosis index greater than 10.0 may suggest a problem.

Self-control but not IQ predicts changes in report card grades. As shown in **Table 1**, self-control and IQ measured in the fall were associated ($r = .23, p < .01$). To estimate their predictive validities for spring semester GPA, we fit a simultaneous multiple regression model with demographic variables as covariates. As shown in **Table 2**, self-control ($\beta = .39, p < .001$) was a slightly better predictor than IQ ($\beta = .33, p < .001$) of spring semester GPA when controlling for demographics. Changes in report card grades over the course of the academic year were predicted only by self-control ($\beta = .11, p = .03$) and not by IQ ($\beta = .06, ns$).

IQ but not self-control predicts standardized achievement test scores. **Table 3** shows that IQ ($\beta = .45, p < .001$) but not self-control ($\beta = .10, ns$) predicted spring standardized achievement test scores when controlling for demographic variables. However, one-year changes in standardized achievement test scores were predicted by neither IQ ($\beta = .07, ns$) nor self-control, $\beta = -.05, ns$. Given the substantial one-year stability of standardized achievement test scores ($r = .89, p < .001$) and modest sample size ($n = 126$), our failure to find a significant

association in this study was not surprising. Indeed, post hoc power analyses suggested that we had 80% power to detect medium sized effects ($r = .25$) at a two-tailed alpha of .05 given the sample size.

Discussion

As hypothesized, self-control measured at the beginning of the school year was a better predictor of spring report card grades than was IQ. Moreover, whereas self-controlled students improved their grades over the school year, students with higher IQ did not.

As expected, IQ but not self-control predicted spring standardized achievement test scores. However, while the standardized coefficient for IQ was .12 greater than for self-control, neither coefficient reached statistical significance. Our post hoc analysis of these findings suggests three reasons for this negative result. First, the modest sample size ($n = 124$) provided statistical power to find only medium-sized effects. Second, the one-year stability of standardized achievement test scores ($r = .89$) suggested that the effect of IQ on changes in standardized achievement test scores would in this population be small in size. Finally, whereas normative data on the IQ test we used (Raven's Progressive Matrices) have not been published, we suspect that there was likely restriction on range in the private school sample employed in Study 1. Collectively, these considerations suggested that rejection of the hypothesized relation between IQ and changes in achievement test scores was premature. We therefore undertook a second study to clarify the predictive validities of self-control and IQ for changes in standardized achievement test scores and, in addition, to replicate the positive findings of Study 1.

Study 2

The objective of Study 2 was to replicate and extend the findings of Study 1 using a much larger and demographically distinct sample. In contrast to the predominantly White,

upper middle-class sample of private school students in Study 1, participants in Study 2 were Black and Latino students from low-income neighborhoods attending public schools in Harlem and the Bronx. By conducting the same longitudinal, prospective study in a community with substantially different ethnic and socioeconomic characteristics, we were able to assess the external validity of our conclusions.

In addition, the two schools involved in Study 2 were able to provide school records on student conduct, an important behavioral outcome that was not available in Study 1. We used quarterly measures of GPA and teacher ratings of student conduct in a longitudinal path analysis model to test our hypothesis that superior student conduct mediates the prediction of grades by self-control. Finally, the sample size of Study 2 was nearly three times as large as the sample size of Study 1, allowing for greater statistical power and an opportunity to test our hypothesis that intelligence is more relevant to standardized achievement test scores than is self-control.

Method

Participants. Participants were fifth through eighth grade students at two public middle schools in New York City. About 94% of the 549 students elected to participate and were not significantly different from non-participants in terms of gender, race, age at assessment, or household income. Of the 516 consented students, 68 were omitted from the analysis due to listwise deletion for missing data, and 3 were excluded based on item response patterns that suggested invalid scores (final $N = 445$ participants, mean age = 11.78, $SD = 1.29$). Sixty-five percent of participants were Latino, and 35% were Black; 53% were female. About 57% of participants were in School 1. The estimated median annual household income was \$23,125 among School 1 participants and \$23,264 among School 2 participants.

Procedure and measures. Students, parents, and homeroom teachers completed consent

forms and measures within the first two months of the school year. Student demographic variables and outcome data were collected from school records at the conclusion of the school year.

Self-control and IQ. We used the same self-control and IQ measures as in Study 1. The observed internal consistency coefficients for the ISC self-control scale were $\alpha = .77, .84$, and $.93$, for self-report, parent, and teacher ratings, respectively.

Report card grades. School records included quarterly report card grades for math, science, English, writing, and social studies classes. As in Study 1, a single-factor measurement model generally fit the report card grade data well, with the exception of one fit index, $\chi^2(5) = 76.29, p < .001$, CFI = .95, RMSEA = .21, SRMR = .04. Although a two-factor model with separate factors for verbal (English, writing, and social studies) and quantitative (math and science) grades fit the data significantly better, $\Delta\chi^2(1) = 50.80, p < .001$, CFI = .99, RMSEA = .13, SRMR = .02, the two factors were very highly correlated, $r = .90, p < .001$. Given the large proportion of shared variance between the two factors—and in the interest of consistency with Study 1 and prior studies of personality and academic achievement (e.g., Duckworth & Seligman, 2006; Nofle & Robins, 2007)—we calculated GPA for the fall and spring semester as the mean of all subject grades for each semester, respectively. We computed cumulative GPA as the average of grades across all four quarters.³

Standardized achievement test scores. We obtained 2008 and 2009 scores from the English/Language Arts and Mathematics standardized achievement tests. The New York State Education Department uses these scores to assess yearly progress in accordance with *No Child Left Behind* legislation. About 75% of questions on this test are multiple-choice and 25% are short-answer or extended written response in format. Language arts and math scores were

strongly correlated in both 2008 ($r = .53, p < .001$) and 2009, $r = .41, p < .001$. We therefore averaged language arts and math scores to create composite standardized achievement test scores for 2008 and 2009.

Student conduct ratings. As part of regular school practice, academic subject teachers rated student conduct in each class using a 5-point scale, where 1 = *unsatisfactory*, 2 = *needs improvement*, 3 = *satisfactory*, 4 = *good*, and 5 = *excellent*. Conduct ratings reflect teachers' overall assessment of student's comportment and effort, including evidence of preparation for class, behavior in class, and completion of homework. A single-factor measurement model largely fit the conduct data well, with the exception of one fit index, $\chi^2(5) = 42.01, p < .001$, CFI = .97, RMSEA = .15, SRMR = .03, and a two-factor model with separate verbal and quantitative conduct factors did not significantly improve fit, $\Delta\chi^2(1) = 0.78, p = .38$, CFI = .97, RMSEA = .17, SRMR = .03. For each student, we averaged conduct grades from all teachers.

Results

Creation of composite self-control score. Correlations among self-report, parent, and teacher ratings of self-control ranged from $r = .33$ to $.45$ and averaged $r = .38, ps < .001$.⁴ The internal reliability of the composite score, calculated as the average of standardized self-report, parent, and teacher ratings, was .91. This composite score was used to index self-control in all subsequent analyses.

Examination of continuous variable distributions. IQ, log-transformed income, as well as 2008 and 2009 standardized achievement test scores were slightly leptokurtic: 1.01, 2.83, 3.95, and 3.77, respectively. All other continuous variables had absolute skew and kurtosis indices less than 1.00. Removing two outliers from the log-transformed income distribution, one outlier from the IQ distribution, and 14 outliers from the 2008 and 2009 standardized achievement test

score distributions reduced the kurtosis indices to .05, .31, .58, and .59, respectively. We decided to include in our final analyses participants who were outliers for IQ or log-transformed income because results were virtually identical when these participants were excluded. However, we did exclude participants with outlying standardized achievement test scores from analyses that included standardized achievement test scores because their inclusion slightly decreased the association between IQ and standardized achievement test scores.

Self-control but not IQ predicts changes in report card grades. As shown in **Table 4**, self-control and IQ measured in the fall were associated ($r = .16, p < .01$). To estimate their predictive validities for spring semester GPA, we fit a simultaneous multiple regression model with demographic variables as covariates. As shown in **Table 2**, self-control ($\beta = .50, p < .001$) predicted spring semester GPA better than did IQ ($\beta = .32, p < .001$) when controlling for demographics. Moreover, *changes* in report card grades over the course of the academic year were predicted only by self-control ($\beta = .08, p < .01$) and not by IQ ($\beta = .03, ns$).

IQ predicts changes in standardized achievement test scores better than does self-control. As shown in **Table 3**, IQ ($\beta = .39, p < .001$) predicted standardized achievement scores better than did self-control ($\beta = .27, p < .001$) when controlling for demographics. In contrast to Study 1, however, in which neither predictor accounted for significant variance in changes in standardized achievement test scores, both self-control ($\beta = .16, p < .001$) and IQ ($\beta = .21, p < .001$) emerged as significant predictors of one-year changes in standardized achievement test scores.

Improvements in student conduct mediate the relationship between self-control and report card grades. To test our hypothesis that superior student conduct mediates the relationship between self-control and improvements in report card grades, we fit a longitudinal

mediation model using path analysis. Following Cole and Maxwell's (2003) recommendations⁵ for testing mediation using longitudinal data, we specified an autoregressive model in which self-control and first-quarter conduct ratings and grades predicted second- and third-quarter conduct and grades, which in turn predicted fourth-quarter conduct and grades. We also included a path from self-control to fourth-quarter grades to assess the direct, unmediated effect of self-control on grades. The mediation model fit the data well, $\chi^2(4) = 13.2, p = .01$, CFI = 1.00, RMSEA = .072, SRMR = .004. As shown in **Figure 2**, our hypothesis was supported. Self-control measured in the first-quarter predicted increases in conduct ratings from the first quarter to the second and third quarters, $\beta = 0.22, p < .001$. Second- and third-quarter conduct predicted increases from prior to fourth-quarter report card grades, $\beta = 0.07, p < .01$. The Sobel (1982) test confirmed that the indirect effect of self-control on increases in grades via increases in conduct was significant, $z = 2.41, p = .02$. In support of full mediation, when taking into account the effects of conduct on grades, self-control did not directly predict increases in grades through fourth quarter, $\beta = 0.02, ns$.

Discussion

In Study 2 we replicated the major findings of Study 1 in a demographically distinct sample of public middle school students. Specifically, self-control predicted improvements in report card grades from the fall to spring semester, whereas IQ did not. Conversely, IQ was a better predictor than self-control of standardized achievement test scores. In Study 2, teacher ratings of student conduct throughout the year enabled us to fit a longitudinal mediation model in which we found that improvements in student conduct fully mediated the prospective association between self-control and improvements in report card grades.

As hypothesized, IQ predicted improvements in standardized achievement test scores

better than did self-control. Several factors may account for the positive findings regarding this specific hypothesis in Study 2 and the equivocal results in Study 1. First, the sample size in Study 2 was more than three times greater than the sample size in Study 1. Second, standardized achievement test scores were markedly less stable over a one-year period in Study 2 ($r = .64$) than in Study 1 ($r = .89$). Third, when comparing the observed raw scores (not reported) for our IQ measure in Studies 1 and 2, we found that there was more than twice as much variance in the public school sample in Study 2 as in the private school sample in Study 1. In sum, compared to Study 1, there was in Study 2 greater statistical power and more variance in both IQ and changes in standardized achievement test scores. A more substantive, alternative explanation is that school type (public vs. private) or socioeconomic status moderates the effects of self-control and IQ on standardized achievement test scores. We used a large national sample in Study 3 to test this possibility.

Study 3

Studies 1 and 2 used convenience samples of students recruited from one private and two New York City public middle schools, respectively. To extend the external validity of our findings and test for moderating variables such as school type and socioeconomic status, we used in Study 3 a national dataset administered by the National Institute of Child Health and Human Development (NICHD). Specifically, we used self-control and IQ data on fourth grade students originally recruited at birth from 10 sites across the United States to predict their ninth grade report card grades and standardized achievement test scores, as well as changes in these outcomes over time. Notably, the specific standardized achievement and intelligence tests and self-control questionnaires in Study 3 differed from those used in Studies 1 and 2. Thus, in Study 3, we were able to test the generalizability of our findings using alternative measures

and a large national sample of students. Moreover, we were able to test whether school type or socioeconomic status moderated the effects of self-control and IQ on GPA and standardized achievement test scores.

Method

Participants. The participants were the 1,364 students in the NICHD Study of Early Child Care and Youth Development (NICHD-SECCYD), a longitudinal study originally designed to examine the effects of child care on development in a diverse sample of children. Details of study recruitment and data collection protocols are described on the study's Web site (<https://secc.rti.org/>).

Approximately 76% of participants were White, 1% were Asian, 13% were Black, 6% were Hispanic, and 4% were other ethnicities; 48% were female. The median household income-to-needs ratio (assessed in terms of income compared with the US Census Bureau–defined poverty line) for this sample was 3.4, indicating that the median household in this sample reported income of more than three times the federal poverty level.

Procedure and measures. Data collection was approved by the appropriate institutional review boards for each of 10 U.S. study sites in the NICHD-SECCYD, and written informed consent was received from each family.

Self-control. When participants were in fourth grade, mothers (or primary caregiver), fathers (or another caregiver if the father was not available), and classroom teachers completed the Social Skills Rating System (SSRS; Gresham & Elliot, 1990), a widely used inventory of positive child behaviors which caregivers rate on a 3-point frequency scale ranging from 0 = *never* to 2 = *very often*. Our own factor analyses as well as independent research on separate samples (Whiteside, McCarthy, & Miller, 2007) failed to replicate the original published factor

structure of the SSRS. Therefore, we used 9 face-valid self-control items (e.g., “controls temper in conflict situations,” “attends to your instructions”) from the parent version of the SSRS and 10 items from the teacher version of the SSRS shown in a separate study to correlate with theoretically predicted outcomes (see Tsukayama, Toomey, Faith, & Duckworth, 2010). See the Appendix for the complete list of SSRS items used. The observed internal consistency coefficients were $\alpha = .77$, $.79$, and $.87$, for mother, father, and teacher ratings, respectively.

IQ. We used the full-scale IQ score ($M = 107$, $SD = 14$) from the Wechsler Abbreviated Scale of Intelligence (WASI) administered when students were in fourth grade. The WASI is an individually administered test of intelligence, includes four subscales (Vocabulary, Block Design, Similarities, and Matrix Reasoning), and is highly correlated with the longer Wechsler Intelligence Scale for Children–Third Edition ($r = .87$; PsychCorp, 1999).

Report card grades. Principals or their designated staff members reported final grades for math, English, science, and social studies for participants at the end of eighth grade. Schools provided official student transcripts at the end of ninth grade. Final grades for math, science, English, and social studies were converted by NICHD-SECCYD staff to a numeric scale where $A+ = 4.33$ to $F = 0.00$.

Confirmatory factor analysis indicated good fit for a single-factor model of academic grades, $\chi^2(2) = 1.08$, $p = .58$, CFI = 1.00, RMSEA = .00. Distinguishing math and science (i.e., quantitative subjects) from English and social studies (i.e., verbal subjects) grades in a two factor model did not significantly improve fit, $\Delta\chi^2(1) = 1.01$, $p = .32$, CFI = 1.00, RMSEA = .00. We therefore calculated each student’s grade point average (GPA) for the eighth and ninth grade by averaging math, science, English, and social studies grades for each grade, respectively.

Standardized achievement test scores. In fifth and ninth grade, students completed

the Passage Comprehension and Applied Problems achievement subscales of the Woodcock-Johnson Psycho-Educational Battery-Revised (WJ-R). The WJ-R includes separate tests of cognitive ability and achievement, the latter of which are designed to assess academic skills and knowledge (Mather, 1991). All items in the WJ-R are individually administered by a trained examiner. Passage Comprehension and Applied Problems were strongly correlated at both fifth ($r = .61, p < .001$) and ninth grade, $r = .67, p < .001$. We therefore averaged these scales to create composite standardized achievement test scores for fifth and ninth grade.

Results

Examination of continuous variable distributions. Standardized achievement test scores at fifth grade (2.67) and log-transformed income-to-needs were somewhat leptokurtic (1.69). Removing two outliers from the fifth grade standardized achievement test distribution and six outliers from the log income-to-needs distribution reduced the kurtosis indices to .70 and .83, respectively. However, analyses excluding these scores produced results virtually identical to those using the full sample, and we therefore report results using the full sample below. All other continuous variables had absolute skew and kurtosis indices less than 1.00.

Structural equation model justification and fit statistics. Unlike Studies 1 and 2, the sample size of the dataset used for Study 3 met the 10:1 to 20:1 subject-to-parameter ratio recommended for structural equation modeling (SEM; Kline, 2005). We used SEM rather than ordinary least squares regression models in Study 3 for two reasons. First, SEM allowed us to create a latent variable for self-control rather than a linear composite as in Studies 1 and 2. Latent variables enable correction for measurement error and produce less-biased estimates of coefficients (Kline, 2005).

A second advantage of SEM was that maximum likelihood procedures allowed for the

retention of participants with missing data using full information maximum likelihood (FIML). This feature was important because about 81% of participants were missing data on at least one of the variables in our analyses. FIML is less biased and more efficient than traditional missing data techniques (Enders & Bandalos, 2001; Peters & Enders, 2002). We note that our predicted results were significant when using listwise deletion (not reported) and were similar to the presented results using FIML.

To estimate the predictive validities of self-control and IQ for ninth grade GPA and standardized achievement test scores, we fit structural equation models that were analogous to the hierarchical regression models in Studies 1 and 2. We used mother, father, and teacher-report self-control measures as indicators of a latent variable of self-control and allowed the parent-report measurement errors to covary. Correlations among ratings were medium-to-large in size (ranging from $r = .34$ to $.52$; average $r = .41$, $ps < .001$), and factor loadings for self-report, parent, and teacher ratings of self-control ranged from $.48$ to $.73$, $ps < .001$. All models fit the data well: $\chi^2(17 \text{ to } 19) = 45.06 \text{ to } 48.01$ for models without moderators and $\chi^2(87 \text{ to } 93) = 160.15 \text{ to } 203.28$ for models testing moderation; all $ps < .001$, CFIs $\geq .98$, RMSEAs $\leq .035$.

Self-control predicts changes in report card grades better than does IQ. As shown in **Table 2**, self-control ($\beta = .40$, $p < .001$) was a better predictor of ninth grade GPA than was IQ ($\beta = .28$, $p < .001$) when controlling for demographic variables. Similarly, *changes* in report card grades from eighth to ninth grade were better predicted by self-control ($\beta = .19$, $p < .01$) than by IQ ($\beta = .14$, $p < .001$).

IQ predicts changes in standardized achievement test scores better than does self-control. As shown in **Table 3**, IQ ($\beta = .64$, $p < .001$) was a better predictor of standardized achievement scores than was self-control ($\beta = .11$, $p < .05$) when controlling for demographic

variables. Consistent with Study 2, both self-control ($\beta = .08, p < .05$) and IQ ($\beta = .31, p < .001$) emerged as significant predictors of changes in standardized achievement test scores from fifth to ninth grade, and IQ predicted more unique variance than did self-control.

Moderation analyses suggest generality of effects across school type and socioeconomic status. Following procedures outlined by Little, Bovoid, and Widamen (2006) to test moderation in structural equation models with latent variables, we found no evidence that type of school (public or private) or socioeconomic status (i.e., log-transformed income-to-needs ratio) moderated the effects of self-control and IQ on GPA or standardized achievement test scores, $ps > .10$.

Discussion

As hypothesized, in a large, national sample of public and private school students, self-control in fourth grade was a better predictor than IQ of ninth grade report card grades. Likewise, self-control was a better predictor than IQ of longitudinal improvements in report card grades. Conversely, IQ was a better predictor than self-control of later standardized achievement test scores as well as longitudinal changes in standardized achievement test scores.

The findings in Study 3 helped clarify the relationship between IQ and standardized achievement test scores. In all samples examined, IQ predicted later standardized achievement test scores better than did self-control. However, in Study 1, the association between IQ and longitudinal *changes* in standardized achievement test scores failed to reach statistical significance. In both Studies 2 and 3, which employed larger samples with more variance in IQ, we did find, as expected, that IQ predicted changes in standardized achievement test scores, and furthermore, that these predictive validities surpassed those of self-control. Our view, therefore, is that the null findings in Study 1 were a consequence of small sample size and restriction on

range. We found no support for the alternative explanation that school type (public vs. private) or socioeconomic status moderates the effects of IQ or self-control on standardized achievement test scores.

Study 4

Collectively, studies 1, 2, and 3 supported our hypotheses that intelligence disproportionately determines standardized achievement test scores whereas self-control disproportionately determines report card grades. In Study 4, we returned to the same middle schools attended by students in Studies 1 and 2 to survey teachers about the content, format, and purpose of these two forms of assessment. Our purpose was to elucidate the proposed mediating competencies differentially assessed by report card grades and standardized achievement test scores.

Method

Participants and procedure. Participants were 57 teachers from the private ($n = 17$) and two public middle schools ($n = 23$ and $n = 17$) at which we conducted Studies 1 and 2. About two-thirds of participants taught humanities (e.g., reading, writing, social studies) and about one-third of participants taught science and math. The average number of years of teaching experience was 7.32 years ($SD = 4.97$). At regularly scheduled faculty meetings in late September 2010, teachers were asked to complete an anonymous questionnaire about “similarities and differences between report card grades and standardized achievement tests.” To preserve anonymity, teachers did not report their ethnicity or gender.

Measures. Based on questionnaires used in prior survey research on grading practices (Cross & Frary, 1996; Gullickson, 1985), we developed items to assess teachers’ judgments of the content, format, and purpose of “academic grades you assign to students (not effort

or conduct grades)” and “standardized achievement tests.” Our questionnaire included four categories of items, labeled *subject material*, *factors unrelated to subject mastery*, *assessment format*, and *purpose of assessment*. Teachers responded to each item using a 5-point Likert-type scale ranging from 1 = *not at all important* to 5 = *very important*.

Subject mastery. Teachers rated separately the relevance to report card grades and standardized achievement test scores of “mastery of specific skills and knowledge taught in my class” and “mastery of skills and knowledge in this subject that were not directly taught (e.g., learned outside of or before my class).”

Factors unrelated to subject material. Teachers rated separately the relevance of four dimensions of student conduct to report card grades, including “prompt and reliable attendance in my class,” “positive class participation (as opposed to disruptive behavior),” “positive attitude and effort,” and “prompt and thorough completion of homework.”

Assessment format. Teachers rated separately the relevance to report card grades and standardized achievement test scores of “multiple choice test questions,” “short answer or fill-in the-blank questions,” and “essay-type test questions.”

Purpose of assessment. Teachers rated separately the relevance to report card grades and standardized achievement test scores of four functions: “to summarize mastery of skills and knowledge,” “to provide feedback to students about areas of mastery and weakness,” “to motivate students to work hard,” and “to enable a comparison of an individual student’s performance to that of other students.”

Results

As shown in **Figure 3**, teachers judged the mastery of skills and knowledge taught in class to be much more relevant to report card grades ($M = 4.73$, $SD = 0.49$) relative to the

mastery of skills and knowledge *not* taught in class ($M = 3.11$, $SD = 1.09$), $t(55) = 9.25$, $p < .001$, $d = 1.92$. In contrast, teachers felt that standardized achievement test scores were equally determined by skills and knowledge acquired in ($M = 3.76$, $SD = 1.16$) and outside of school ($M = 3.67$, $SD = 1.11$), $t(48) = .47$, ns , $d = .08$.

As shown in **Figure 4**, the relevance of homework completion ($M = 4.32$, $SD = 0.72$) and other aspects of student conduct, including class participation ($M = 4.14$, $SD = 1.02$), general attitude and effort ($M = 4.00$, $SD = 1.03$), and attendance ($M = 3.75$, $SD = 1.28$), were judged as intermediate in relevance to report card grades. Specifically, teachers judged skills and knowledge taught in class to be more important ($t(55) = 4.00$, $p < .001$, $d = .60$) and skills and knowledge not taught in class to be less important ($t(53) = 2.88$, $p = .006$, $d = .50$) than the most closely ranked student conduct factor.

Figure 5 shows that when assessing students for the purpose of report card grades, teachers favored short-answer or fill-in-the-blank questions ($M = 4.04$, $SD = 1.11$) over essay-type questions ($M = 3.57$, $SD = 1.56$; $t(55) = 1.97$, $p = .05$, $d = .35$), which in turn were favored over multiple choice questions ($M = 3.00$, $SD = 1.41$; $t(54) = 2.15$, $p = .04$, $d = .37$). In contrast, teachers judged the importance of multiple-choice questions ($M = 4.33$, $SD = 1.01$) to standardized achievement tests as higher than that of short answer or fill-in-the-blank questions ($t(48) = 3.61$, $p = .001$; $M = 3.61$, $SD = 1.37$, $d = .60$), which in turn were judged marginally more important than essay-type questions ($M = 3.18$, $SD = 1.52$), $t(48) = 1.77$, $p = .083$, $d = .30$. For each of these three formats, differences in the relative importance to report card grades vs. standardized achievement tests were significant, $ps < .03$, $ds \geq .30$.

As shown in **Figure 6**, teachers considered the primary purpose of report card grades to be the provision of feedback to the student about areas of mastery and weakness ($M = 4.46$, $SD =$

0.85) and the summary of skills and knowledge, $M = 4.44$, $SD = 0.76$; $t(56) = .13$, ns , $d = .02$. For report card grades, these two functions were more important than motivating students to work hard, $M = 3.98$, $SD = 1.02$; $t(55) = 2.92$, $p = .005$, $d = .52$, which in turn was rated substantially more important than the comparison of a student's performance to that of other students, $M = 2.64$, $SD = 1.15$; $t(55) = 7.07$, $p < .001$, $d = 1.23$.

In contrast, **Figure 6** shows that teachers judged the most important purpose of standardized achievement tests to be the comparison of a student's performance to that of other students ($M = 3.80$, $SD = 1.13$), which was rated marginally higher than the most closely ranked factor, the summary of mastery of skills and knowledge, $M = 3.47$, $SD = 1.27$; $t(50) = 1.69$, $p = .10$, $d = .27$. Teachers judged the *least* important functions of standardized achievement tests to be the provision of feedback to students about areas of mastery and weakness ($M = 2.59$, $SD = 1.24$) and motivation to work hard, $M = 2.49$, $SD = 1.33$, which were not significantly different from each other, $t(50) = .48$, ns , $d = .08$. For each of these four purposes, differences in the relative importance to report card grades vs. standardized achievement tests was significant, $ps < .001$, $ds \geq .93$.

Discussion

Teacher ratings in Study 4 support our theorized differences in the content, format, and purpose of report card grades on the one hand, and standardized achievement test scores on the other. Consistent with the model in Figure 1, teachers considered skills and knowledge acquired inside and outside of school to be equally important to performance on standardized achievement tests. For report card grades, in contrast, what students are formally taught in school was judged dramatically more relevant than what they might pick up outside of school.

Furthermore, multiple-choice questions, the format of assessment *least* favored by

teachers when determining report card grades, were considered the *most* important format for standardized achievement tests. It is worth noting that the standardized achievement tests in Studies 1 and 2 indeed mostly comprised multiple-choice items, providing evidence for the accuracy of teacher perceptions of standardized achievement tests. The standardized achievement test used in Study 3 was an individually administered oral examination of reading and math skills.

Finally, teachers judged the most important purpose of standardized achievement tests and the least important function of report card grades to be student-to-student comparisons in academic performance. Providing feedback to students about areas of mastery and weakness was judged among the most important functions of report card grades but among the least important functions of standardized achievement tests. Similarly, whereas teachers reported using report card grades to motivate students to work hard, increasing student motivation was among the least important functions of standardized achievement tests.

General Discussion

We proposed a theoretical model distinguishing between competencies better assessed by report card grades and influenced by self-control and, in contrast, competencies better assessed by standardized achievement tests and influenced by intelligence (see Figure 1). Using demographically distinct samples of middle school students in three prospective, longitudinal studies, we tested our model by comparing the predictive validities of self-control and IQ for report card grades and standardized achievement test scores. As hypothesized, self-control was a better predictor than IQ of later report card grades. Moreover, self-control reliably predicted changes in report card grades over time, whereas IQ did not. Increases in report card grades were, in Study 2, mediated by mid-year improvements in student conduct. Contrariwise, measures of

intelligence were better predictors than self-control of standardized achievement test scores, and in Studies 2 and 3, also were better predictors of changes in standardized achievement test scores.

Teacher judgments in Study 4 supported our proposition that report card grades and standardized achievement test scores differentially reflect the student competencies illustrated in Figure 1. Specifically, middle school teachers indicated that when determining academic report card grades, they considered aspects of student conduct, including completion of homework assignments, class participation, effort, and attendance. These factors were not rated as important as mastery of skills and knowledge formally taught in class but were rated more important than skills and knowledge acquired outside of class. In contrast, teachers perceived skills and knowledge acquired outside and inside the classroom to be *equally* relevant to performance on standardized achievement tests. In addition, teachers confirmed our supposition that multiple-choice, the most common format for standardized achievement test items, is rarely used in everyday quizzes and tests. Furthermore, the one-on-one oral administration format of the WJ-R is presumably more novel than what is commonly encountered in the classroom. Thus, to a degree, in Studies 1, 2, and 3, standardized tests challenged students to solve relatively novel problems in relatively novel formats.

Consideration of homework and other aspects of student conduct, over and beyond their contribution to academic mastery, has been interpreted as a source of bias in report card grades (e.g., Cross & Frary, 1996; Jussim, 1989). Is it unfair or misleading for teachers to consider anything other than academic skills and knowledge when determining report card grades? Or, is it justifiable, as Willingham et al. (2002) have suggested, for teachers to use grades to enforce a local contract with students? That is, considering students of equivalent academic mastery, are teachers justified in awarding higher grades to those who come to class more promptly and

better prepared, who have their homework completed more thoroughly and punctually, and who participate more positively in class discussions? Arguably, the question of what report card grades are good for is philosophical, not empirical. Nevertheless, there are empirically verifiable assumptions upon which philosophical positions are built. In a review of grading practices and the student motivation literature, Brookhart (1994) concluded that “rewarding students who follow the rules is one way teachers can demonstrate that achievement can be *under a student’s control*” (p. 294, italics added). Research from a variety of traditions confirms that individuals who believe their futures are largely shaped by their own actions, as opposed to factors beyond their control, are more likely to sustain effort toward long-term goals (Bandura, 1977; Judge & Hurst, 2007; Rotter, 1966; Seligman, 2007).

A second argument can be made for explicitly rewarding behaviors that demand students to practice self-control: the benefits of self-control are ubiquitous and profound. More self-controlled children are less likely to abuse drugs and alcohol (Wills & Stoolmiller, 2002), are protected from unhealthy weight gain (Duckworth, Tsukayama, & Geier, 2010; Tsukayama, et al., 2010), are more likely to refrain from delinquent and criminal acts (Caspi et al., 1994; Lynam et al., 2000), are less likely to develop externalizing symptomology (Eisenberg et al., 2009), enjoy higher levels of positive emotion and life satisfaction and lower levels of negative emotion (Duckworth, Tsukayama, & Kim, 2010), and have more adaptive relationships with other people (Mischel, 1989; Mischel, Shoda, & Rodriguez, 1989). James (1899) speculated that practicing self-control encourages its development, an idea that has found support in more recent empirical studies (Baumeister, Gailliot, DeWall, & Oaten, 2006; Muraven, Baumeister, & Tice, 1999).

. If the goals of formal education extend to setting children on paths toward more productive and happier lives (Brighouse, 2008), then, in our view, there is good argument to explicitly

encourage self-regulation of attention, behavior, and emotion in the service of long-term goals.

Limitations and Future Directions

To our knowledge, the current investigation is the first to compare directly standardized achievement test scores and report card grades in terms of their relative weighting of intelligence and self-control. Like any empirical investigation, ours had strengths and weaknesses, which we consider here and use to suggest directions for future research. First, notwithstanding the fact that our samples collectively represented both private and public school students from a wide range of socioeconomic and ethnic backgrounds, further studies are needed to confirm these findings and verify the degree to which they can be generalized to, for instance, older and younger students and students in non-US countries. We hope that such replication studies would follow a similar multi-source approach to the measurement of self-control, a methodology that increases reliability and therefore optimizes the predictive validity of non-IQ measures (Duckworth & Seligman, 2005).

A second limitation concerns the non-experimental nature of this investigation. Studies 1, 2, and 3 employed longitudinal, prospective designs and included demographic control variables. Nevertheless, we cannot rule out all possible third-variable confounds.⁶ Random-assignment, placebo-controlled experimental research would most clearly expiate the causal role of self-control and intelligence in determining report card grades and standardized achievement test scores, respectively. Traits such as self-control and intelligence demonstrate substantial but far from perfect rank-order stability over time (Borghans, Duckworth, Heckman, & Ter Weel, 2008). Recent advances demonstrating that both self-control (Diamond, Barnett, Thomas, & Munro, 2007; Duckworth, Grant, Loew, Oettingen, & Gollwitzer, 2010) and intelligence (Jaeggi, Buschkuhl, Jonides, & Perrig, 2008; Nisbett, 2009) may respond to deliberate intervention

suggests that such experimental research may soon be within the realm of possibility.

Conclusion

Because high stakes standardized achievement tests are playing an increasingly prominent role in policy and practice, with schools devoting more formal instruction time to test preparation (McMurrer, 2007), it seems imperative to balance awareness of the strengths of standardized achievement tests with a nuanced understanding of their inherent limitations (Popham, 1999, 2000). Our findings suggest that report card grades reflect dimensions of student competence related more to self-control than to intelligence. Standardized achievement tests, in contrast, reflect dimensions of student competence related more to intelligence than to self-control.

These results may help explain why, in a recent study of almost 80,000 students admitted to the University of California, high school GPA was a better predictor than SAT test scores of cumulative college GPA (Geiser & Santelices, 2007). Likewise, in a study of 21 U.S. universities of varying size and selectivity, high school GPA predicted successful graduation better than did SAT or ACT test scores, even without controlling for high school quality or rigor of local grading standards (Bowen, Chingos, & McPherson, 2009). The superior incremental predictive validity of report card grades relative to these widely used standardized achievement tests suggests that “grades measure a student’s ability to ‘get it done’ in a more powerful way than do SAT scores...grades reveal much more than mastery of content...Getting good grades in high school, however demanding (or not) the high school, is evidence that a student consistently met a certain standard of performance. It is hardly surprising that doing well on a single standardized achievement test is less likely to predict the myriad qualities a student needs to ‘cross the finish line’ and graduate from college.” (Bowen, et al., 2009, pp.123-124).

The current investigation raises two sets of policy questions: First, what are the implications of more intelligent students performing better on standardized achievement tests for reasons other than learning more in the classroom? Value-added analyses (VAA) have recently been proposed to gauge the efficacy of a school or teacher by gains in student standardized achievement test scores over time (e.g., departures from predicted trajectories). If such gains not only reflect what children are formally taught, but also knowledge acquired outside of school and facility with novel tasks, do value-added analyses in fact perform their intended accountability function? In particular, do value-added analyses advantage teachers and schools with more intelligent students? Second, what are the implications of report card grades reflecting more than just academic skills and knowledge? Is it fair – or useful – for teachers to combine assessments of academic competence and student conduct in the calculation of report card grades? Or, are so-called “hodgepodge” grading practices (Cross & Frary, 1996) detrimental?

We offer two specific suggestions for policy and practice that address these concerns. First, curriculum and standardized assessment should be as closely aligned, in both content and format, as possible. Willingham et al. (2002) have pointed out that no standardized achievement test can, in a few hours, sample in as much detail the skills and knowledge acquired throughout an entire year of formal instruction. Nevertheless, better alignment in both content and format should reduce unintended g-loading of standardized achievement tests and, at the same time, increase the importance of skills and knowledge formally taught in class. Recent reforms aimed at simplifying and clarifying academic standards are, we hope, one step in this direction (Gates Foundation, 2010).

Second, we see drawbacks to the current practice of mixing student conduct with

academic competence. Namely, the signal sent by report card grades is ambiguous. Does an A grade indicate superior academic mastery, superior student conduct, or, given that teachers vary in their grading practices (McMillan, et al., 2002), an unknowable amalgam of both? It is important not to overstate the direct influence of student conduct on report card grades. Teachers in Study 4 identified skills and knowledge taught in class as significantly more important to report card grades than any aspect of student conduct. Nevertheless, we suggest reviving – and standardizing – the practice of grading students on discrete aspects of student conduct. Teachers could separately indicate on report cards, for example, estimates of the percentage of homework assignments students completed, the percentage of classes to which students arrived on time and prepared, the estimated percentage of time students paid attention in class, and the number of positive contributions students made to classroom discussion. If these aspects of student conduct requiring self-control were separately specified on report cards, teachers could then base students' academic grades solely on measured academic skills and knowledge. In this way, teachers could preserve the motivational function of report card grades while also accurately summarizing academic competence.

Public policies emphasizing any single type of assessment, such as standardized achievement test scores, to the exclusion of all others necessarily overlook aspects of student competence that are important, not only instrumentally insofar as they predict future achievement in and beyond the classroom, but, arguably, as ends in themselves. We suggest that the *No Child Left Behind* policy, in its singular focus on standardized achievement test scores as the metric of student performance, inadvertently devalues complementary sources of information. In particular, leaving behind report card grades in an effort to standardize assessments across teachers, schools, and regions leaves behind essential information about self-

control competence.

Footnotes

¹The constructs of self-control, self-regulation, and self-discipline, as well as impulsiveness and impulsivity, are defined differently by different authors (Duckworth & Schulze, 2009; Evenden, 1999; Olson, Schilling, & Bates, 1999). There is, however, considerable overlap. Most definitions of self-control and synonymous constructs refer to the ability to regulate impulses in the service of long-term goals and standards; most definitions of impulsiveness and impulsivity connote the obverse.

²This school as a policy administers the CTP quantitative reasoning subtest rather than the CTP math subtest to its middle school students.

³In analyses examining individual subject grades separately, we replicated our findings for overall GPA, with the single exception that self-control did not account for significantly greater variance in math grades than did IQ. The results of these analyses are available from the first author upon request.

⁴We cannot confirm why cross-informant correlations were higher in Study 2 than in Study 1. A comparison of the variance in raw scores in both studies (results not shown) suggested comparable range. One explanation is students in the private school in Study 1 demonstrated more cross-domain variability in self-controlled behavior than did the public school students in Study 2. Another possibility is that the larger sample in Study 2 reduced sampling error.

⁵In accordance with these recommendations, we did not include in our estimate of indirect effects of self-control on fourth-quarter grades any paths involving synchronous measures (e.g., first quarter self-control → first-quarter conduct → second and third quarter

grades → fourth quarter grades). Our estimate of the indirect effects is conservative in this respect.

⁶Findings not discussed here indicate that our multi-source index of self-control aligned most closely with the Big Five dimension of Conscientiousness (as measured by teacher and self-report). Self-control, however, continued to predict change in grades and conduct over and beyond all Big Five dimensions. In contrast, Big Five Openness to Experience was a stronger predictor of standardized achievement tests. Thus, although the present investigation compared the role of a single, empirically and theoretically predicted trait (i.e., self-control) to that of IQ in predicting academic outcomes, these additional analyses add confidence that self-control is particularly important relative to other dimensions of personality regarding school conduct and report card grades.

References

- Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, 101, 213-232.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Barton, K., Dielman, T. E., & Cattell, R. B. (1972). Personality and IQ measures as predictors of school achievement. *Journal of Educational Psychology*, 63, 398-404.
- Baumeister, R. F., Gailliot, M., DeWall, C. N., & Oaten, M. (2006). Self-regulation and personality: How interventions increase regulatory success, and how depletion moderates the effects of traits on behavior. *Journal of Personality*, 74, 1773-1801.
- Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science*, 16, 351-355.
- Borghans, L., Duckworth, A. L., Heckman, J. J., & Ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources*.
- Borkenau, P., Mauer, N., Riemann, R., Angleitner, A., & Spinath, F. M. (2004). Thin Slices of Behavior as Cues of Personality and Intelligence. *Journal of Personality and Social Psychology*, 86, 599-614.
- Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). Test scores and high school grades as predictors *Crossing the finish line: Completing college at America's public universities* (pp. 112-133). Princeton, NJ: Princeton University Press.
- Brighouse, H. (2008). Education for a flourishing life. In D. L. Coulter & J. r. Wiens (Eds.), *Why*

- do we educate? Renewing the conversation: The 107th yearbook of the National Society for the Study of Education. In G. D. Fenstermacher (Series Ed.) (Vol. 1, pp. 58-71). New York, NY: Wiley-Blackwell.
- Brody, N. (1997). Intelligence, schooling, and society. *American Psychologist. Special Issue: Intelligence & Lifelong Learning*, 52, 1046-1050.
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7, 279-301.
- Browne, M. W., MacCallum, R. C., Kim, C.-T., Andersen, B. L., & Glaser, R. (2002). When fit indices and residuals are incompatible. *Psychological Methods*, 7, 403-421.
- Caspi, A., Moffit, T. E., Silva, P. A., Stouthamer-Loeber, M., Krueger, R. F., & Schmutte, P. S. (1994). Are some people crime-prone? Replications of the personality-crime relationship across countries, genders, races, and methods. *Criminology*, 32, 163-195.
- Cizek, G. J., Fitzgerald, S. M., & Rachor, R. A. (1995). Teachers' assessment of practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, 3, 159-179.
- Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of Structural Equation Modeling. *Journal of Abnormal Psychology*, 112, 558-577.
- Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Cross, L. H., & Frary, R. B. (1996). *Hodgepodge grading; Endorsed by students and teachers alike*. Paper presented at the Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.

- Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13-21.
- Diamond, A., Barnett, S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, 318, 1387-1388.
- Duckworth, A. L., Grant, H., Loew, B., Oettingen, G., & Gollwitzer, P. M. (2010). Self-regulation strategies improve self-discipline in adolescents: Benefits of mental contrasting and implementation intention. *Educational Psychology*.
- Duckworth, A. L., & Schulze, R. (2009). *Jingle Jangle: A meta-analytic investigation of the convergent validity of self-control*. manuscript.
- Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, 16, 939-944.
- 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, 198-208.
- Duckworth, A. L., Tsukayama, E., & Geier, A. (2010). Self-controlled children stay leaner in the transition to adolescence. *Appetite*, 54, 304-308.
- Duckworth, A. L., Tsukayama, E., & Kim, B. E. (2010). Domain-specific impulsivity in school-age children. Manuscript submitted for publication.
- Duckworth, A. L., Tsukayama, E., & May, H. (2010). Establishing causality using longitudinal hierarchical linear modeling: An illustration predicting achievement from self-control. *Social Psychology and Personality Science*, 1, 311-317.
- Eid, M., & Diener, E. (2006). Handbook of multimethod measurement in psychology. Washington, DC, US: American Psychological Association, 553.

- Eisenberg, N., Valiente, C., Spinrad, T. L., Cumberland, A., Liew, J., Reiser, M., . . . Losoya, S. H. (2009). Longitudinal relations of children's effortful control, impulsivity, and negative emotionality to their externalizing, internalizing, and co-occurring behavior problems. *Developmental Psychology, 45*, 988-1008.
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. [Journal Article]. *Structural Equation Modeling, 8*, 430-457.
- Epstein, S. (1979). The stability of behavior: I. On predicting most of the people much of the time. *Journal of Personality and Social Psychology, 37*, 1097-1126.
- Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology. Special Issue: Impulsivity, 146*, 348-361.
- Gagné, F., & St Père, F. (2002). When IQ is controlled, does motivation still predict achievement? *Intelligence, 30*, 71-100.
- Galotti, K. M. (2005). Setting goals and making plans: How children and adolescents frame their decisions. In J. E. Jacobs & P. A. Klaczynski (Eds.), *The development of judgment and decision making in children and adolescents*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Geiser, S., & Santelices, M. V. (2007). Validity of high school grades in predicting student success beyond the freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes. *Research and Occasional Papers Series from the Center for Studies in Higher Education at the University of California, Berkeley, CSHE 2007*(CSHE.6.07). Retrieved from <http://cshe.berkeley.edu/publications/publications.php?id=265>

- Gottfredson, L. S. (2002). g: Highly general and highly practical. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 331-380). Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
- Gottfredson, L. S. (2004). Schools and the g factor. *The Wilson Quarterly* 34-35.
- Gresham, F. M., & Elliot, S. N. (1990). *Social Skills Rating Scale Manual*. Circle Pines, MN: American Guidance Service.
- Gullickson, A. R. (1985). Student evaluation techniques and their relationship to grade and curriculum. *Journal of Educational Research*, 79, 96-100.
- Harter, S. (1985). *Self-perception profile for children*: University of Denver.
- Jaeggi, S. M., Buschkuhl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Science USA*, 105, 6829-6833.
- James, W. (1899). *Talks to teachers on psychology and to students on some of life's ideals*. New York: Holt and Company.
- Judge, T. A., & Hurst, C. (2007). Capitalizing on one's advantages: Role of core self-evaluations. *Journal of Applied Psychology*, 92, 1212-1227.
- Jussim, L. (1989). Teacher expectations: Self-fulfilling prophecies, perceptual biases, and accuracy. *Journal of Personality and Social Psychology*, 57, 469-480.
- Jussim, L. (1991). Grades may reflect more than performance: Comment on Wentzel (1989). *Journal of Educational Psychology*, 83, 153-155.
- Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in Structural Equation Modeling. *Structural Equation Modeling*, 10, 333-351.
- Kline, R. B. (2005). *Principles and Practice of Structural Equation Modeling* (2nd ed.). New

- York: Guilford.
- Little, T. D., Bovaird, J. A., & Widaman, K. F. (2006). On the merits of orthogonalizing powered and product terms: Implications for modeling interactions among latent variables. *Structural Equation Modeling, 13*, 497-519.
- Lynam, D. R., Caspi, A., Moffit, T. E., Wikstrom, P.-O., Loeber, R., & Novak, S. (2000). The interaction between impulsivity and neighborhood context on offending: The effects of impulsivity are stronger in poorer neighborhoods. *Journal of Abnormal Psychology, 109*, 563-574.
- Mather, N. (1991). *An instructional guide to the Woodcock-Johnson Psycho-Educational Battery-Revised*. New York: John Wiley & Sons.
- McDermott, P. A., Mordell, M., & Stoltzfus, J. C. (2001). The organization of student performance in American schools: Discipline, motivation, verbal learning, nonverbal learning. *Journal of Educational Psychology, 93*, 65-76.
- McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research, 95*, 203-213.
- McMurrer, J. (2007). *Choices, changes, and challenges: Curriculum and instruction in the NCLB era*. Washington, D.C.: Center on Education Policy.
- Miles, J., & Shevlin, M. (2007). A time and a place for incremental fit indices. *Personality and Individual Differences, 42*.
- Mischel, W. (1989, November 15). *Delay of gratification in children*. Paper presented at the Columbia University.
- Mischel, W., Shoda, Y., & Peake, P. K. (1988). The nature of adolescent competencies predicted by preschool delay of gratification. *Journal of Personality and Social Psychology, 54*,

- 687-696.
- Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science*, 244, 933-938.
- Muraven, M., Baumeister, R. F., & Tice, D. M. (1999). Longitudinal improvement of self-regulation through practice: Building self-control strength through repeated exercise. *Journal of Social Psychology*, 139, 446-457.
- Neisser, U., Boodoo, G., Bouchard, T. J., Jr., Boykin, A. W., Brody, N., Ceci, S. J., . . . Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51, 77-101.
- Nisbett, R. E. (2009, February 8). Education is all in your mind, *The New York Times*.
- Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology*, 93, 116-130.
- Nunnally, J. C. (1978). *Psychometric theory* (Vol. 640). New York: McGraw-Hill.
- Olson, S. L., Schilling, E. M., & Bates, J. E. (1999). Measurement of impulsivity: Construct coherence, longitudinal stability, and relationship with externalizing problems in middle childhood and adolescence. *Journal of Abnormal Child Psychology*, 27, 151-165.
- Peters, C. L. O., & Enders, C. (2002). A primer for the estimation of structural equation models in the presence of missing data: Maximum likelihood algorithms. *Journal of Targeting, Measurement and Analysis for Marketing*, 11, 81-95.
- Popham, W. (1999). Why standardized test scores don't measure educational quality. *Educational Leadership*, 56, 8-15.
- Popham, W. (2000). *Modern Educational Measurement* (3rd ed.). Boston: Allyn and Bacon.
- Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic

- performance. *Psychological Bulletin*, 135, 322-338.
- PsychCorp. (1999). *Wechsler Abbreviated Scale of Intelligence manual*. San Antonio, TX: Author.
- Raven, J. C. (1948). The comparative assessment of intellectual ability. *British Journal of Psychology*, 39, 12-19.
- Robbins, S. B., Allen, J., Casillas, A., Peterson, C. H., & Le, H. (2006). Unraveling the differential effects of motivational and skills, social, and self-management measures from traditional predictors of college outcomes. *Journal of Educational Psychology*, 98, 598-616.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 80, 1-28.
- Salthouse, T. A., & Pink, J. E. (2008). Why is working memory related to fluid intelligence? *Psychonomic Bulletin & Review*, 15, 364-371.
- Seligman, M. E. (2007). *The optimistic child: A proven program to safeguard children against depression and build lifelong resilience*. New York: Houghton Mifflin.
- Sharpley, C., F., & Edgar, E. (1986). Teachers' ratings vs standardized tests: An empirical investigation of agreement between two indices of achievement. *Psychology in the Schools*, 23, 106-111.
- Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhardt (Ed.), *Sociological methodology 1982* (pp. 290-312). San Francisco: Jossey-Bass.
- Stipek, D., & Douglas, M. (1989). Developmental change in children's assessment of intellectual competence. *Child Development*, 60, 521-538.

- Tsukayama, E., Toomey, S. L., Faith, M. S., & Duckworth, A. L. (2010). Self-control protects against overweight status in the transition from childhood to adolescences. *Archives of Pediatrics and Adolescent Medicine* 164, 631-635.
- Whiteside, S. P., McCarthy, D. M., & Miller, J. D. (2007). An examination of the factor structure of the social skills rating system parent elementary form. *Assessment*, 14, 246-254.
- Willingham, W. W. (1985). *Success in college: The role of personal qualities and academic ability*. New York: College Entrance Examination Board.
- Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, 39, 1-37.
- Wills, T. A., & Stoolmiller, M. (2002). The role of self-control in early escalation of substance use: A time-varying analysis. *Journal of Consulting and Clinical Psychology*, 70, 986-997.
- Wolfe, R. N., & Johnson, S. D. (1995). Personality as a predictor of college performance. *Educational and Psychological Measurement*, 55, 177-185.
- Wong, M. M.-h., & Csikszentmihalyi, M. (1991). Motivation and academic achievement: The effects of personality traits and the duality of experience. *Journal of Personality*, 59, 539-574.
- Wulfert, E., Block, J. A., Ana, E. S., Rodriguez, M. L., & Colsman, M. (2002). Delay of gratification: Impulsive choices and problem behaviors in early and late adolescence. *Journal of Personality*, 70, 533-552.
- Zhou, Q., Main, A., & Wang, Y. (2010). The relations of temperamental effortful control and anger/frustration to Chinese children's academic achievement and social adjustment: A longitudinal study. *Journal of Educational Psychology*, 102, 180-196.

Table 1

Summary Statistics and Bivariate Correlations in Study 1

Variable	M	SD	1	2	3	4	5	6	7	8	9	10	11	12
1. Self-control	0.06	0.71	-											
2. IQ	0.00	1.00	.23**	-										
3. 2008 Achievement	359.31	14.78	.16	.57**	-									
4. 2009 Achievement	365.68	16.67	.07	.48**	.89**	-								
5. First-term GPA	3.18	0.51	.45**	.45**	.52**	.40**	-							
6. Second-term GPA	3.25	0.50	.48**	.43**	.45**	.37**	.86**	-						
7. Income ^a	\$83,387	\$40,857	*	*	*	*	*							
8. Female	53%		-.03	-.03	-.02	.06	.13	.12	-					
9. White	73%		.15	-.05	-.09	-.05	-.03	.02	-.08	-				
10. Asian	8%		.06	-.13	-.23**	-.12	.07	.04	.43***	-.28**	-			
11. Black	6%		-.02	.19*	.24**	.19*	.20*	.20*	-.03	.17*	-.49**	-		
12. Hispanic	4%		.02	-.08	-.08	-.13	-.19*	-.15	-.40***	.05	-.41**	-.0	-	
13. Other	9%		.00	.11	.10	.00	-.20*	-.21	*	-.29**	-.35**	-.0	-.0	-
			-.08	.01	.13	.11	.01	.02	-.12	.14	-.51**	-.0	-.0	-.0
											*	9	8	7

Note. $N = 139$. Summary statistics and correlations including 2008 Achievement are based on $n = 124$.

^aLog-transformed income is used for correlations

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 2

Hierarchical Linear Regression Models (Studies 1 and 2) and Analogous Structural Equation Models (Study 3) Predicting Time 2 GPA

	Study 1		Study 2		Study 3	
Predictor	ΔR^2	β	ΔR^2	β	ΔR^2	β
Step 1	.44***		.41***		.45***	
Control variables ^a						
Self-control		.39***		.50***		.40***
IQ		.33***		.32***		.28***
Step 2	.32***		.42***		.16***	
Control variables ^a						
Self-control		.11*		.08**		.19**
IQ		.06		.03		.14***
Time 1 GPA		.76***		.86***		.53***
Total R^2	.76***		.84***		.61***	

Note. Study 1 $N = 139$; Study 2 $N = 445$; Study 3 $N = 1,364$.

^aControl variables included gender, ethnicity, and socioeconomic status in all studies. Studies 1 and 2 included grade level. Study 2 included school. Study 3 included school type.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3

Hierarchical Linear Regression Models (Studies 1 and 2) and Analogous Structural Equation Models (Study 3) Predicting Time 2 Standardized Achievement Test Scores

	Study 1		Study 2		Study 3	
Predictor	ΔR^2	β	ΔR^2	β	ΔR^2	β
Step 1	.49***		.30***		.58***	
Control variables ^a						
Self-control		.10		.27***		.11*
IQ		.45***		.39***		.65***
Step 2	.34***		.19***		.10***	
Control variables ^a						
Self-control		-.05		.16***		.08*
IQ		.07		.21***		.31***
Time 1 achievement test		.82***		.54***		.49***
Total R^2	.84***		.48***		.68***	

Note. Study 1 $n = 124$; Study 2 $n = 431$; Study 3 $N = 1,364$.

^aControl variables included gender, ethnicity, and socioeconomic status in all studies. Studies 1 and 2 included grade level. Study 2 included school. Study 3 included school type (i.e., public or private).

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 4

Summary Statistics and Bivariate Correlations in Study 2

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10
1. Self-control	-0.03	0.79	-									
2. IQ	0.00	1.00	.16**	-								
3. 2008 Achievement	677.85	22.93	.19***	.42**	-							
4. 2009 Achievement	682.61	17.00	.28***	.46**	.64**	-						
5. First-term GPA	81.14	8.34	.53***	.42**	.61**	.68**	-					
6. Second-term GPA	81.46	8.63	.55***	.40**	.56**	.66**	.91**	-				
7. First-term conduct grades	3.83	0.75	.66***	.25**	.46**	.43**	.68**	.66**	-			
8. Second-term conduct grades	3.93	0.77	.63***	.23**	.43**	.41**	.61**	.71**	*	-		
9. Income ^a	\$24,479.16	\$9,917.16	.03	.04	.05	.09	.06	.08	.06	.05	-	
10. Female	53%		.17***	.00	.04	.07	.14**	.15**	.30***	.32***	.01	-
11. Black	35%		-.21***	-.11*	-.05	-.06	-.04	-.07	-.19**	-.21**	-.04	-.04

Note. $N = 445$. Summary statistics and correlations including 2008 and 2009 Achievement are based on $n = 441$ and $n = 434$,

respectively.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5

Summary Statistics and Bivariate Correlations in Study 3

Variable	M	SD	<i>n</i>	1	2	3	4	5	6	7	8
			1,02								
1. Mother-report self-control	1.23	0.35	2	-							
2. Father-report self-control	1.22	0.35	703	.52***	-						
					.34**						
3. Teacher-report self-control	1.52	0.40	915	.36***	*	-					
	106.8	14.4	1,01		.25**						
4. IQ	6	4	2	.25***	*	.33***	-				
	107.3	11.6			.18**						
5. Fifth-grade achievement	5	2	993	.23***	*	.31***	.77***	-			
	105.2	13.7			.22**						
6. Ninth-grade achievement	9	2	892	.21***	*	.31***	.74***	.78***	-		
					.32**						
7. Eighth-grade GPA	2.95	0.92	691	.27***	*	.44***	.44***	.42***	.51***	-	
					.32**						
8. Ninth-grade GPA	3.03	0.71	721	.27***	*	.42***	.51***	.48***	.54***	.71***	-
					.24**						
9. Income-to-Needs Ratio	4.50	3.88	985	.29***	*	.32***	.44***	.44***	.42***	.37***	.40***
			1,36								
10. Female	48%		4	.09**	.13**	.23***	.03	.02	-.02	.21***	.23***
			1,36								
11. White	76%		4	.13***	.09*	.24***	.29***	.30***	.29***	.29***	.31***
			1,36								
12. Asian	1%		4	.04	.00	.06	.08**	.08*	.03	.04	.05
			1,36			-.30**		-.34**			-.28**
13. Black	13%		4	-.13***	-.08*	*	-.35***	*	-.33***	-.22***	*
			1,36								
14. Hispanic	6%		4	-.04	.00	-.03	-.06*	-.06	-.06	-.11**	-.09*

15. Other	4%	1,364								
		4	-.05	-.07	-.04	-.01	-.04	-.03	-.16***	-.13**

Note. $N = 1,364$.

^aLog-transformed income-to-needs ratio is used for correlations.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5, Continued

Summary Statistics and Bivariate Correlations in Study 3

Variable	9	10	11	12	13	14
1. Mother-report self-control						
2. Father-report self-control						
3. Teacher-report self-control						
4. IQ						
5. Fifth-grade achievement						
6. Ninth-grade achievement						
7. Eighth-grade GPA						
8. Ninth-grade GPA						
9. Income-to-Needs Ratio	-					
10. Female	.01	-				
11. White	.30***	-.01	-			
			-.21**			
12. Asian	.04	.02	*	-		
			-.69**			
13. Black	-.33***	.00	*	-.05	-	
			-.46**		-.10**	
14. Hispanic	-.07*	.00	*	-.03	*	-
			-.34**			-.0
15. Other	-.05	-.01	*	-.02	-.07**	5

Note. $N = 1,364$.

^aLog-transformed income-to-needs ratio is used for correlations.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Figure Captions

Figure 1. Theoretical model relating self-control and intelligence to competencies differentially relevant to report card grades and standardized achievement test scores.

Figure 2. Conduct ratings mediate the relation between self-control and report-card grades.

Bolded lines represent the indirect effect of self-control on fourth-quarter grades. The dashed line represents the direct effect of self-control on grades accounting for conduct ratings. * $p < .05$. ** $p < .01$. *** $p < .001$.

Figure 3. Relevance of skills and knowledge taught in class versus skills and knowledge not taught in class ratings for report card grades and standardized achievement test scores.

Figure 4. Relevance of subject mastery and student conduct ratings for report card grades.

Figure 5. Importance of assessment formats ratings for report card grades and standardized achievement tests.

Figure 6. Purpose of assessment ratings for report card grades and standardized achievement tests.

Figure 1

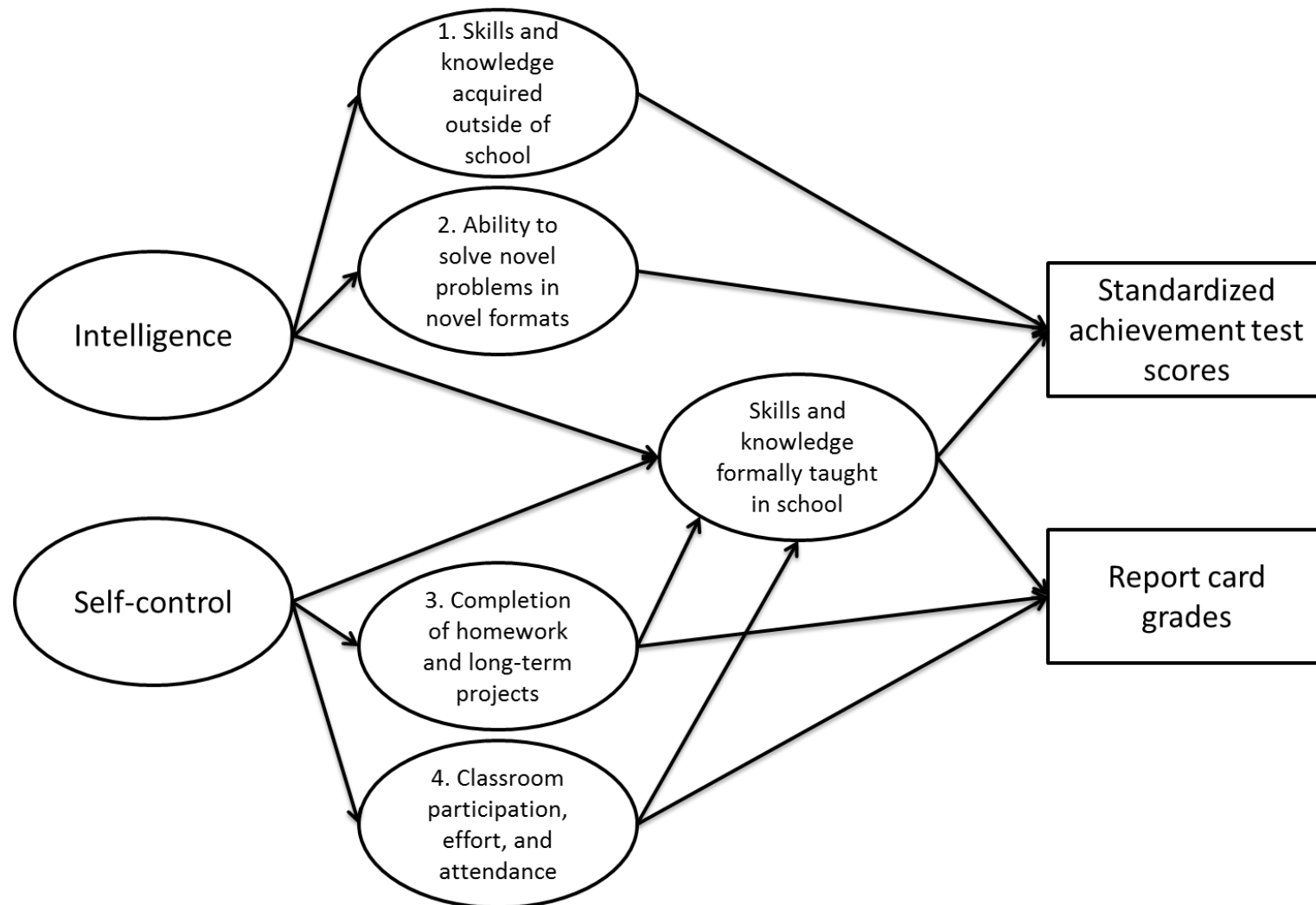


Figure 2

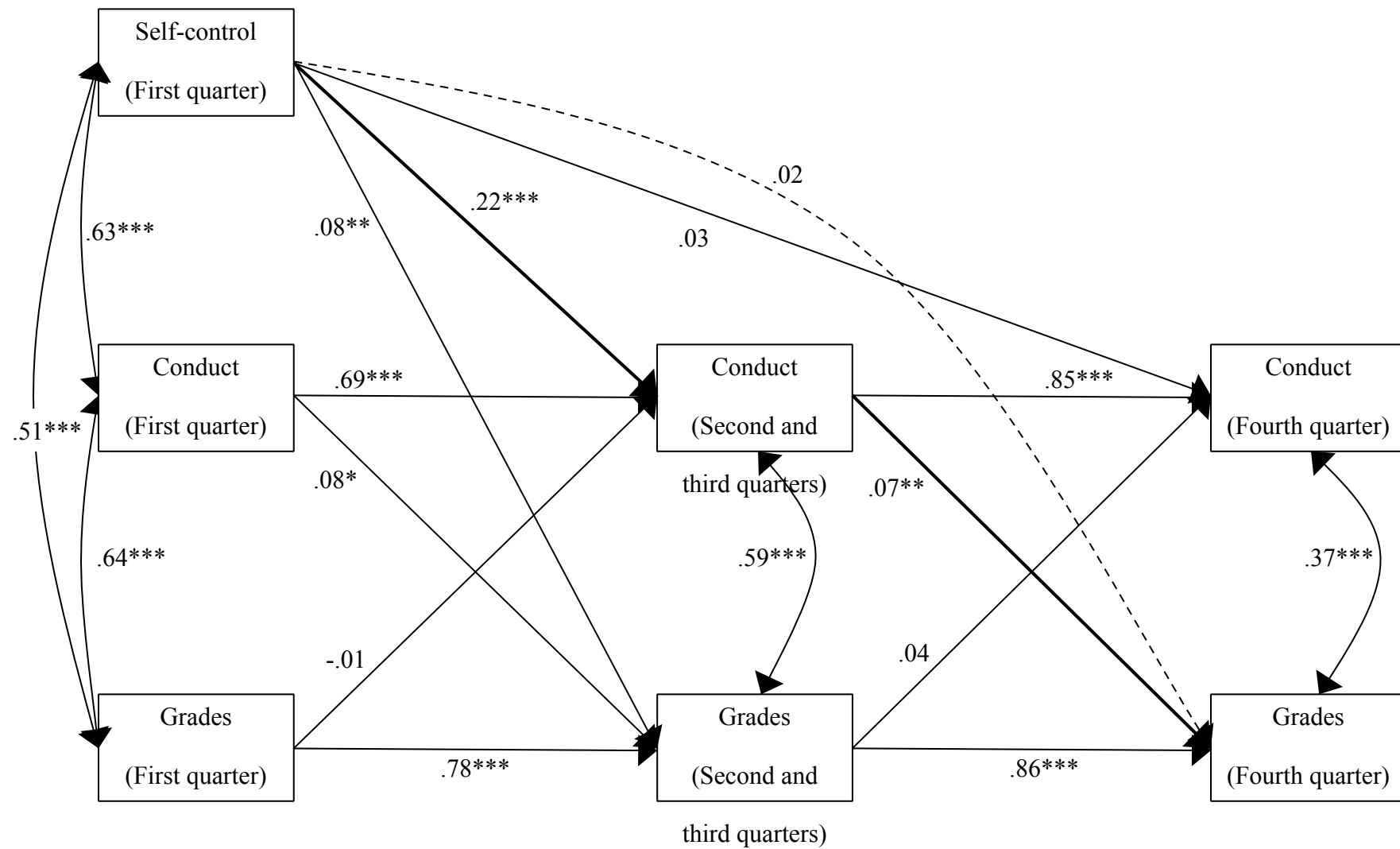


Figure 3

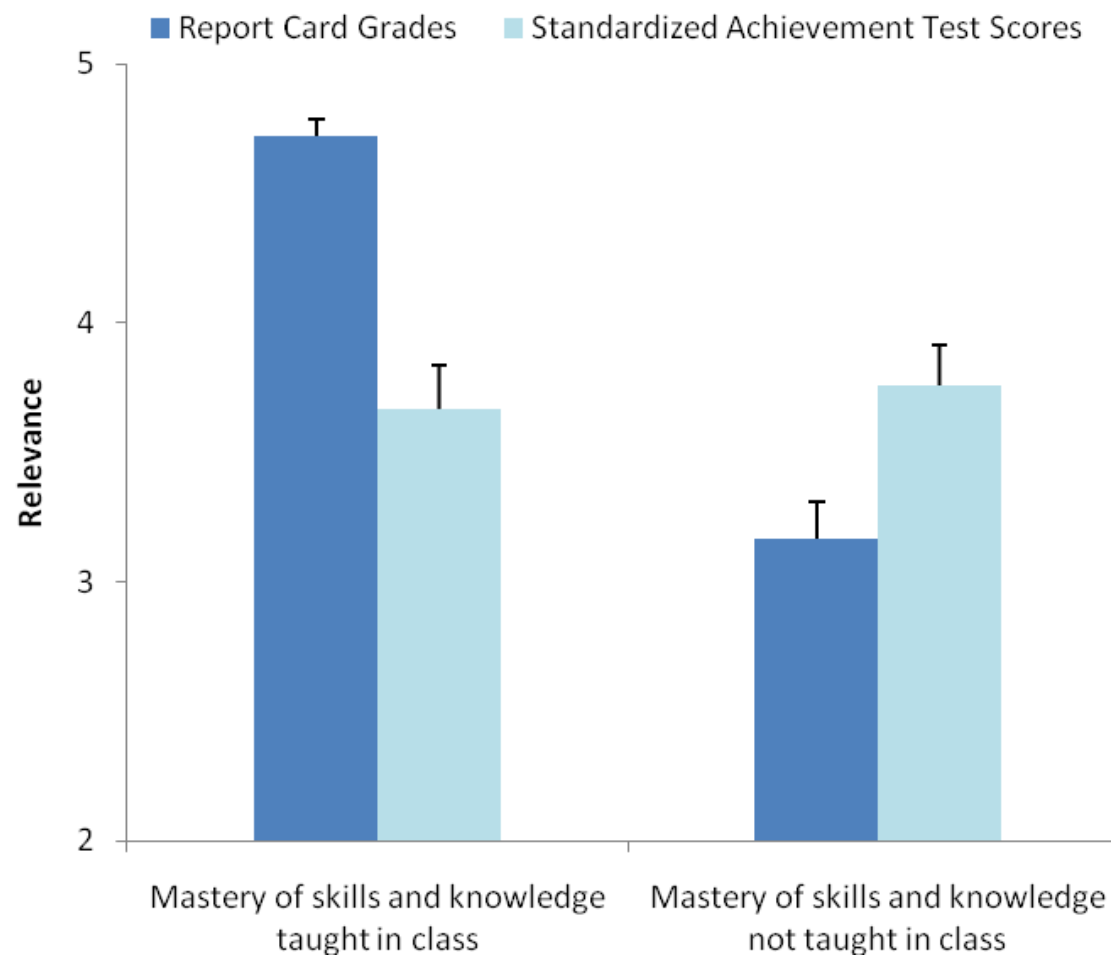


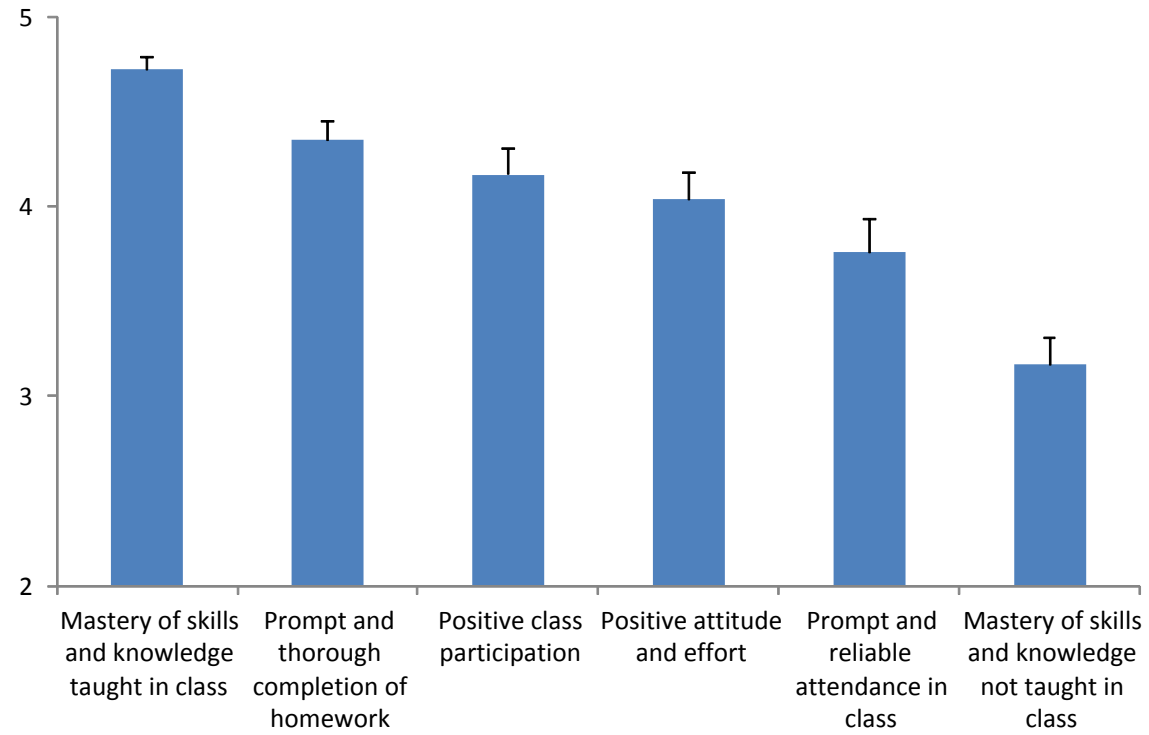
Figure 4

Figure 5

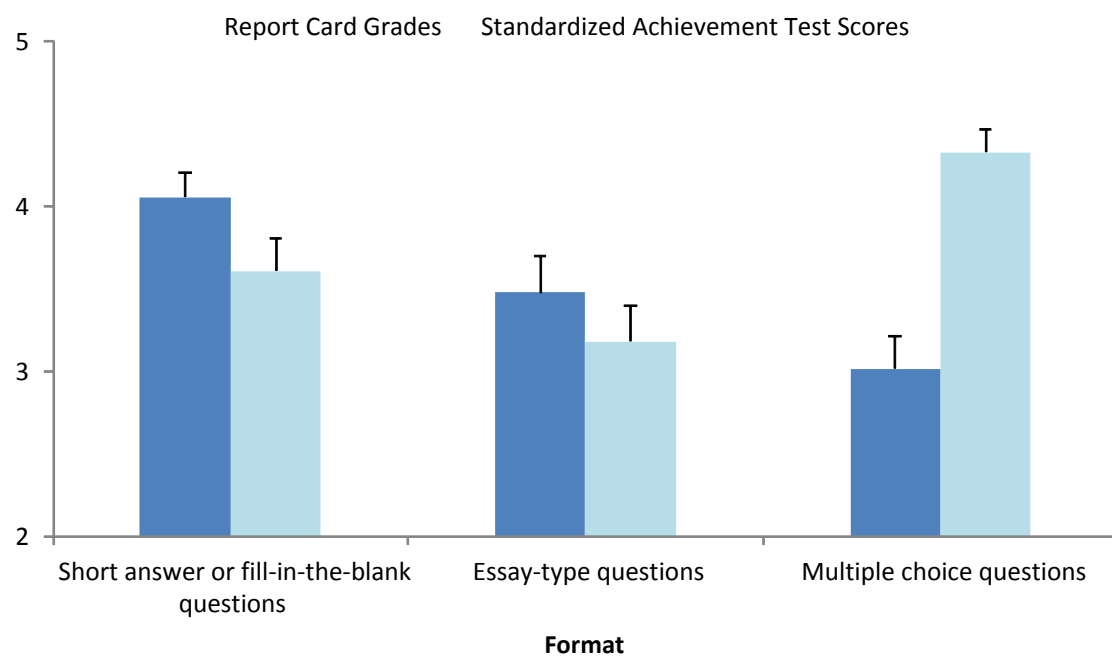
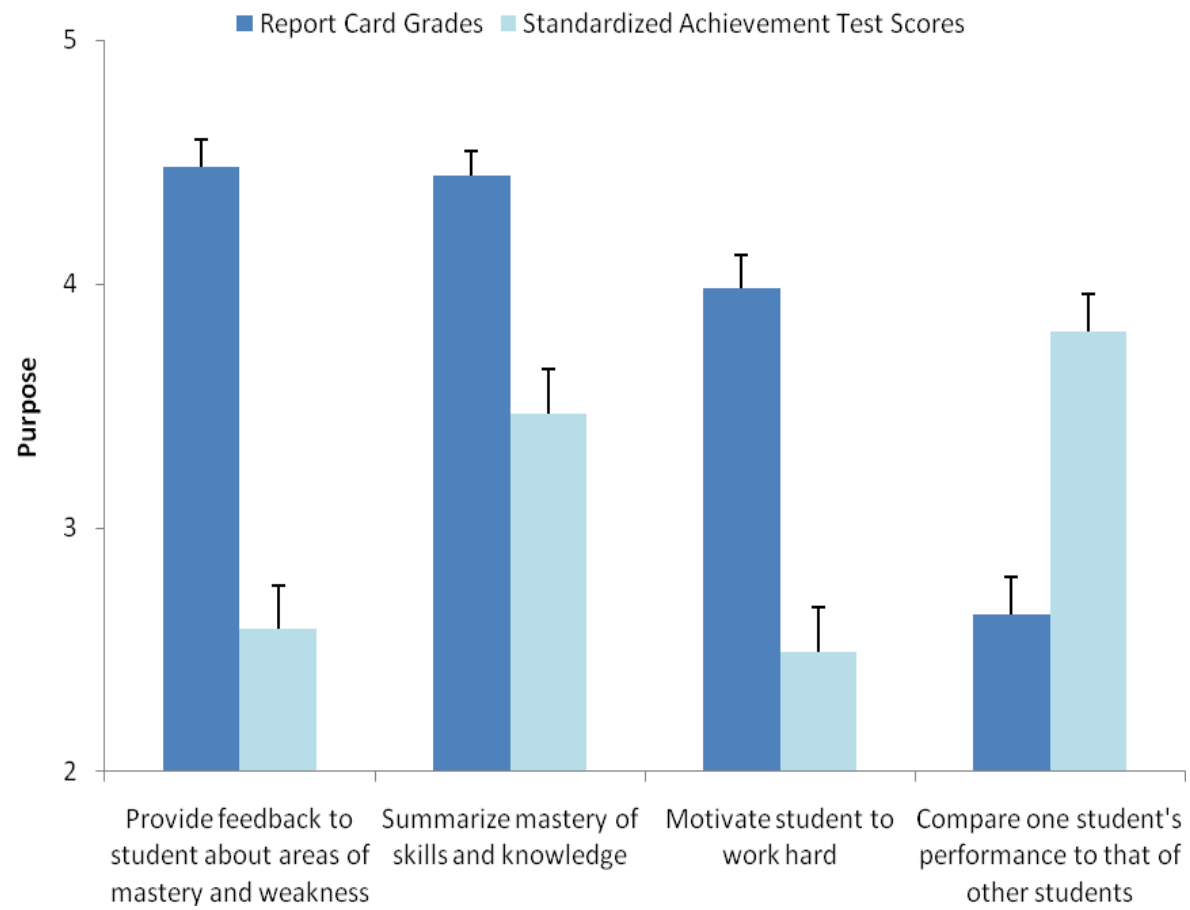


Figure 6



Appendix

Social Skill Rating System (SSRS) Items^a Used to Assess Self-Control at Fourth Grade in Study 3

Keeps room/desk clean and neat without being reminded
Responds appropriately when pushed or hit by others
Controls temper when arguing with other children
Finishes tasks within a reasonable amount of time
Receives criticism well
Responds appropriately to teasing by peers
Ends disagreements with you calmly^b
Controls temper in conflict situations with you^b
Attends to speakers at meetings^b
Controls temper in conflict situations with adults^c
Follows your directions^c
Attends to your instructions^c
Ignores peer distraction when doing classwork^c

Note.

^aItems were paraphrased to generalize across rater (i.e., parent or teacher) and for brevity

^bParent-report only.

^cTeacher-report only.