

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. Duckworth
University of Pennsylvania

Patrick D. Quinn
University of Texas at Austin

Eli Tsukayama
University of Pennsylvania

The increasing prominence of standardized testing to assess student learning motivated the current investigation. We propose that standardized achievement test scores assess competencies determined more by intelligence than by self-control, whereas report card grades assess competencies determined more by self-control than by intelligence. In particular, we suggest that intelligence helps students learn and solve problems independent of formal instruction, whereas self-control helps students study, complete homework, and behave positively in the classroom. Two longitudinal, prospective studies of middle school students support predictions from this model. In both samples, IQ predicted changes in standardized achievement test scores over time better than did self-control, whereas self-control predicted changes in report card grades over time better than did IQ. As expected, the effect of self-control on changes in report card grades was mediated in Study 2 by teacher ratings of homework completion and classroom conduct. In a third study, ratings of middle school teachers about the content and purpose of standardized achievement tests and report card grades were consistent with the proposed model. Implications for pedagogy and public policy are discussed.

Keywords: impulsivity, self-control, achievement, success, personality

On January 8, 2002, George W. Bush signed into law the No Child Left Behind (NCLB) Act (2002), legislation which for the first time in U.S. history made federal funding for K–12 public schools contingent upon the use of standardized achievement tests to assess student performance. 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 comparison of students across classrooms and schools. Critics of standardized testing (e.g., Kohn, 2000) have questioned the validity of standardized achievement tests, but such criticisms have been countered by substantial and convincing

empirical evidence to the contrary (Kuncel & Hezlett, 2007; Sackett, Borneman, & Connelly, 2008).

Researchers at Educational Testing Service, 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). Indeed, whereas report card grades and standardized achievement test scores are both designed to gauge students’ academic skills and knowledge, they do not rank students identically (i.e., correlations between these measures are large but do not approach unity; e.g., $r_s = .66$ and $.62$ in Studies 1 and 2, Duckworth & Seligman, 2006; $r = .62$ in Willingham et al., 2002).

We propose that standardized achievement test scores and report card grades differentially reflect student competencies determined by intelligence and self-control, two distinct traits shown to predict successful functioning in—and beyond—the classroom. Our model, described in more detail below, is graphically summarized in Figure 1. The present investigation tests predictions of this model in two independent samples of children followed longitudinally during their middle school years and in a third study in which middle school teachers were asked to compare the content and purpose of standardized achievement tests with report card grades.

This article was published Online First December 19, 2011.

Angela L. Duckworth, Positive Psychology Center, University of Pennsylvania; Patrick D. Quinn, Department of Psychology, University of Texas at Austin; Eli Tsukayama, Positive Psychology Center, University of Pennsylvania.

The research reported here was supported by Grant K01-AG033182 from the National Institute on Aging and Grant R305B090015 from the Institute of Education Sciences, U.S. Department of Education.

Correspondence concerning this article should be directed to Angela L. Duckworth, Positive Psychology Center, University of Pennsylvania, 3701 Market St. Suite 200, Philadelphia, PA 19104. E-mail: duckworth@psych.upenn.edu

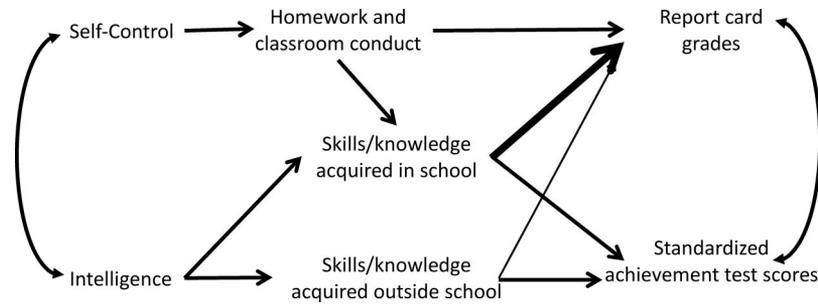


Figure 1. Theoretical model relating self-control and intelligence to competencies differentially related to report card grades and standardized achievement tests. The relative importance of competencies in determining report card grades and standardized achievement scores, respectively, is reflected by the width of corresponding arrows.

Intelligence and Self-Control

Intelligence and self-control are two of the best studied trait predictors of academic performance.¹ *Intelligence* has been defined as the “ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought” (Neisser et al., 1996, p. 77), and *IQ* refers specifically to performance on a variety of cognitive tests specifically designed to measure intelligence. Generally, scores on such cognitive tests are highly correlated, suggesting a domain-general intellectual faculty (i.e., general intelligence; Lubinski, 2004). Individual differences in IQ scores are observable early in life and display substantial rank-order stability ($r = .7$ by middle childhood; Borghans, Duckworth, Heckman, & ter Weel, 2008). A century of empirical evidence has shown that IQ scores predict school success, and this relationship appears to be monotonic, even at the extreme right-tail of the population (Gottfredson, 2004; Kuncel, Ones, & Sackett, 2010; Lubinski, 2009; Neisser et al., 1996).

Self-control refers to the voluntary regulation of attention, emotion, and behavior in the service of personally valued goals and standards (Baumeister, Heatherton, & Tice, 1994). Individual differences in self-control are salient early in life and measurable with rating scales or behavioral tasks (e.g., delay of gratification tasks) specifically developed to assess the ability to inhibit a dominant, maladaptive response in order to execute an adaptive, subdominant response (Duckworth & Kern, 2011; Eisenberg, Smith, Sadovsky, & Spinrad, 2004; Mischel, Shoda, & Rodriguez, 1989; Rothbart, Ellis, & Posner, 2004). Self-control exhibits moderate rank-order stability during childhood (Moffitt et al., 2011) but, like other personality traits, likely does not approach the rank-order stability of intelligence until the fifth decade of life (Roberts & DelVecchio, 2000). While found in some studies to covary with intelligence (e.g., Moffitt et al., 2011), self-control nevertheless prospectively predicts academic performance over and beyond intelligence (Duckworth & Seligman, 2005; Duckworth, Tsukayama, & May, 2010).

Prior research on how self-control and intelligence relate to academic achievement typically conflates course grades and standardized achievement test scores. To our knowledge, however, there has been minimal investigation of how differences in these two forms of academic assessment might give rise to divergent associations with student characteristics (Willingham et al., 2002).

Given the increasing prominence of standardized achievement tests in educational policy and practice, it seems important to examine whether standardized achievement test scores are interchangeable with report card grades or, rather, as we conjecture, these two outcomes reflect distinct underlying student competencies.

Where Standardized Achievement Test Scores and Report Card Grades Differ

Willingham and colleagues (2002) have identified several dimensions on which standardized achievement test scores and report card grades differ. Most important, the content assessed by standardized achievement tests diverges at least somewhat from that of the curricula that students are actually exposed to (and then tested on by their own teachers) in the classroom (Popham, 1999, 2000; Willingham et al., 2002). As a consequence, the skills and knowledge acquired outside of formal instruction would be expected to improve standardized achievement test scores more so than report card grades. Conversely, the effort students put forth toward learning teacher-assigned material would be expected to improve report card grades more so than standardized achievement test scores. Second, homework and classroom conduct may be directly factored into report card grades by teachers (Brookhart, 1994; Cizek, Fitzgerald, & Rachor, 1995; Cross & Frary, 1996; McMillan, Myran, & Workman, 2002) but do not directly influence standardized achievement test scores.

¹ For the sake of clarity, we note, as have others (e.g., Block, 1996) that the burgeoning literature on self-control suffers from both the jingle (Kelley, 1927) and jangle fallacies (Thorndike, 1904). That is, diverse terms are used to connote the same construct by researchers working in distinct theoretical traditions and, at the same time, identical terminology is used by different researchers to refer to disparate constructs (Duckworth & Kern, 2011). We use the term *self-control* to refer to a personality trait that is moderately stable across time and situation and, crucially, connotes voluntary regulation of impulses in the service of long-term goals. In our view, metacognitive strategies that facilitate goal pursuit, collectively referred to by some researchers (e.g., Zimmerman & Kitsantas, 2005) as *self-regulation*, likely contribute to individual differences in self-control but are not our explicit focus in this investigation.

The Differential Payoffs of Intelligence and Self-Control

We have so far argued that although report card grades and standardized achievement test scores both reflect formally taught skills and knowledge, they nevertheless differ in important ways. We further suggest that students' intelligence and self-control differentially influence these factors. Specifically, more intelligent students likely acquire skills and knowledge outside of formal instruction at higher rates than do their less intelligent peers (Gottfredson, 2002), in large part because they are better at learning to solve completely novel problems for which they receive no formal instruction (Salthouse & Pink, 2008). Independent learning, in turn, should disproportionately influence standardized achievement test scores because, as argued above, such tests typically include content not formally taught to students by their teachers. In sum, as illustrated in Figure 1, more intelligent students are likely at an advantage in solving problems that they have not been formally taught to solve.

More self-controlled students, on the other hand, should have an advantage studying what is formally taught to them, completing homework, and behaving properly in class. As William James (1899) observed, in "schoolroom work" there is inevitably "a large mass of material that must be dull and unexciting" (pp. 104–105). Likewise, Aristotle suggested in the *Nicomachean Ethics* that "the roots of education are bitter, but the fruit is sweet." Consistent with the speculation that the activities that facilitate learning in formal school settings may not be as immediately rewarding as rival diversions, even high-ability students randomly asked to report about their experiences throughout their day do not report being "motivated, happy, or satisfied about their performance while they study" (Wong & Csikszentmihalyi, 1991, p. 563). On the contrary, "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" (Wong & Csikszentmihalyi, 1991, p. 563).

More generally, it seems reasonable to assume that positive classroom conduct (e.g., concentrating on difficult new concepts rather than daydreaming, participating in teacher-led discussions rather than joking with classmates, arriving promptly to class rather than lingering in the hallway to socialize) as well as studying and homework completion all yield long-term rewards (e.g., entrance to college) at the expense of short-term pleasure. Therefore, the ability to choose long-term rewards over immediate, more pleasant diversions would seem crucial to acquiring skills and knowledge through formal instruction (Ross & Nisbett, 1991). Indeed, at least one study has found that self-control but not IQ predicts the number of hours that middle school students spend on studying and homework (Duckworth & Seligman, 2005). Likewise, self-control has been shown to predict positive classroom conduct (Valiente, Lemery-Chalfant, Swanson, & Reiser, 2008).

The Current Investigation

We propose that intelligence helps students learn outside of formal instruction, whereas self-control helps students overcome temptations that otherwise detract from studying, homework, and positive classroom behavior. In two separate longitudinal studies, we tested the following specific hypotheses generated by our

model: (a) Self-control is a better predictor than IQ of improvements in report card grades over time (in Studies 1 and 2); (b) IQ is a better predictor than self-control of improvements in standardized achievement test scores over time (in Studies 1 and 2); (c) the effect of self-control on improvements in report card grades is mediated by teacher ratings of homework completion and classroom conduct during the school year (in Study 2); and, finally, (d) teachers perceive differences between report card grades and standardized achievement tests in terms of both content and purpose (in Study 3).

We focus on middle school students in the present investigation for several reasons. First, middle 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). This transition in grading practices reflects a more general shift toward rank-ordered comparisons of students (Eccles et al., 1993). Additionally, as children enter middle school, academic performance becomes an increasingly important component of their personally valued goals and overall self-esteem (Galotti, 2005; Harter, 1985); notably, self-esteem, school engagement, and report card grades may all decrease sharply during this transition (Eccles, 2004; Eccles et al., 1993; Simmons & Blyth, 1987). At the same time, children become much more sensitive to the distinction between intelligence and effort, with heightened attention to how they compare with other students (Stipek & Douglas, 1989). In sum, middle school represents an inflection point in the nature, purpose, and interpretative consequence of the assessment of academic performance. Thus, this developmental epoch is the earliest at which we would expect a measurable and consequential rift between standardized achievement test scores and report card grades.

Study 1

In Study 1, we conducted secondary data analysis on a sample of children recruited at birth from 10 sites across the United States by the National Institute of Child Health and Human Development (NICHD). Specifically, we used self-control and IQ data collected from participants when they were in the fourth grade to predict changes in their report card grades and standardized achievement test scores during their middle school years.

Method

Participants. The participants were the 1,364 students in the NICHD Study of Early Child Care and Youth Development (NICHD–SECCYD). Details of study recruitment and data collection protocols are described on the study's website (<https://secc.rti.org/>). Approximately 76% of participants were White, 13% were Black, 6% were Hispanic, 1% were Asian, and 4% were other ethnicities; 48% were girls.

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. The mother (or primary caregiver), father (or other caregiver if the father was not available), and classroom teacher of each participant completed the Social Skills Rating

System (SSRS; Gresham & Elliot, 1990) when participants were in the fourth grade. The SSRS is a widely used inventory of positive child behaviors that 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 as a measure of self-control. This self-control scale has demonstrated strong convergent validity with other questionnaire measures of self-control as well as predictive validity for theoretically relevant outcomes (Tsukayama, Duckworth, & Kim, 2011; Tsukayama, Toomey, Faith, & Duckworth, 2010). The observed internal consistency coefficients were $\alpha = .77, .79,$ and $.87,$ for mother, father, and teacher ratings, respectively.

Intelligence. Students completed the Wechsler Abbreviated Scale of Intelligence (WASI) when they were in the fourth grade. The WASI is an individually administered test of intelligence that includes four subscales (Vocabulary, Block Design, Similarities, and Matrix Reasoning) and is highly correlated with the longer Wechsler Intelligence Scale for Children—Third Edition ($r = .87$; Psychological Corp., 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, $\chi^2(2) = 1.08, p = .58,$ confirmatory fit index (CFI) = 1.00, root mean square error of approximation (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. The Woodcock–Johnson Psycho-Educational Battery—Revised (WJ–R) is an individually administered test battery that, in addition to cognitive ability tests, includes standardized achievement tests (Mather, 1991). At fifth and ninth grade, participants completed both the Passage Comprehension and Applied Problems achievement tests of the WJ–R. Because reading and math achievement test scores were strongly correlated at both fifth ($r = .61, p < .001$) and ninth grade ($r = .67, p < .001$), we averaged them to create composite standardized achievement test scores for fifth and ninth grade, respectively.

Socioeconomic status. The median household income-to-needs ratio (assessed in terms of income compared with the U.S. 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.

School type. Principals completed a school demographics survey when participants were in the ninth grade. School type

included public (84.7%), private, nonreligious (1.9%), and private, religious (13.4%). We created a binary variable to indicate private school (0 = public; 1 = private).

Results and Discussion

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, because analyses excluding these scores produced results virtually identical to those using the full sample, we report results using the full sample below. All other continuous variables had absolute skew and kurtosis indices less than 1.

Structural equation model. We used structural equation modeling (SEM) to test our hypotheses for two reasons. First, SEM allowed us to create a latent variable for self-control using self-report and parent and teacher ratings. 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 is important because about 20% of the data were missing (see Table 1). FIML is less biased and more efficient than traditional missing data techniques (Enders & Bandalos, 2001; Peters & Enders, 2002).

Predictive validities of self-control and IQ for GPA and standardized achievement test scores. To estimate the predictive validities of self-control and IQ for ninth grade GPA and standardized achievement test scores, we fit an SEM model with demographic variables (gender, age, ethnicity, socioeconomic status, and school type), eighth grade GPA, and fifth-grade standardized achievement test scores as covariates. Self-control and IQ were allowed to covary, as were ninth grade GPA and standardized achievement test scores. We used mother-, father-, and teacher-report self-control measures as indicators of a latent self-control variable 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 parent and teacher ratings of self-control ranged from $.48$ to $.73, ps < .001$. All other constructs were treated as observed variables. The model fit the data well: $\chi^2(23) = 50.21, p < .001$; CFI = .99, RMSEA = .03.

As shown in Table 2, changes in report card grades from eighth to ninth grade were predicted by both self-control ($\beta = .20, p = .002$) and IQ ($\beta = .09, p = .044$). Conversely, longitudinal changes in standardized achievement test scores were predicted by IQ ($\beta = .29, p < .001$) but not self-control ($\beta = .01, p = .88$).

Study 2

The findings of Study 1 were consistent with our first two hypotheses. Specifically, self-control predicted longitudinal changes in report card grades better than did IQ, and IQ predicted longitudinal changes in standardized achievement test scores better than did self-control. In Study 2, we sought to replicate the same pattern of prospective associations and, further, examine evidence

Table 1
Summary Statistics and Bivariate Correlations in Study 1

Variable	M	SD	n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1. Mother-report self-control	1.23	0.35	1,022	—															
2. Father-report self-control	1.22	0.35	703	.52***	—														
3. Teacher-report self-control	1.52	0.40	915	.36***	.34***	—													
4. IQ	106.86	14.44	1,012	.25***	.25***	.33***	—												
5. Fifth grade achievement test	107.35	11.62	993	.23***	.18***	.31***	.77***	—											
6. Ninth grade achievement test	105.29	13.72	892	.21***	.22***	.31***	.74***	.78***	—										
7. Eighth grade GPA	2.95	0.92	691	.27***	.32***	.44***	.44***	.42***	.51***	—									
8. Ninth grade GPA	2.88	0.82	721	.26***	.31***	.42***	.44***	.47***	.53***	.71***	—								
9. Income-to-needs ratio ^a	4.50	3.88	985	.29***	.24***	.32***	.44***	.44***	.42***	.37***	.39***	—							
10. Private school (15%)	687	1.17	985	.08	.08	.05	.19***	.18***	.15***	.14***	.11***	.25***	—						
11. Female (48%)	1,364	0.09	1,364	.09***	.13***	.23***	.03	.02	-.02	.21***	.20***	.01	.02	—					
12. White (76%)	1,364	0.13	1,364	.13***	.09***	.24***	.29***	.30***	.29***	.29***	.32***	.30***	.07	-.01	—				
13. Asian (1%)	1,364	0.04	1,364	.04	.00	.06	.08*	.03	.03	.04	.04	.04	-.04	.02	-.21***	—			
14. Black (13%)	1,364	0.13	1,364	-.13***	-.08*	-.30***	-.35***	-.34***	-.33***	-.22***	-.28***	-.33***	-.09*	.00	-.69***	-.05	—		
15. Hispanic (6%)	1,364	0.04	1,364	-.04	.00	-.03	-.06	-.06	-.06	-.11***	-.10***	-.07*	.03	.00	-.46***	-.03	-.10***	—	
16. Other (4%)	1,364	0.05	1,364	-.05	-.07	-.04	-.01	-.04	-.03	-.16***	-.13***	-.05	-.04	-.01	-.34***	-.02	-.07**	-.05	—

Note. N = 1,364. GPA = grade point average.
^a Log-transformed income-to-needs ratio is used for correlations.
 * p < .05. ** p < .01. *** p < .001.

Table 2
Summary of Structural Equation Models Predicting Time 2 GPA and Standardized Achievement Test Scores

Outcome/Predictor	Study 1		Study 2	
	β	R ²	β	R ²
GPA		.62		.86
Control variables ^a				
Self-control	.20**		.22***	
IQ	.09*		.01	
Time 1 GPA	.52***		.71***	
Standardized achievement test scores		.69		.58
Control variables ^a				
Self-control	.01		-.05	
IQ	.29***		.12***	
Time 1 achievement test	.46***		.29***	

Note. Study 1 N = 1,364; Study 2 N = 510. GPA = grade point average.
^a Control variables included gender, ethnicity, and socioeconomic status in all studies. Study 1 included school type. Study 2 included grade level and school.
 * p < .05. ** p < .01. *** p < .001.

for our third hypothesis that homework completion and classroom conduct mediate the relation between self-control and report card grades. To this end, we partnered with two urban middle schools whose students are rated weekly by teachers on homework completion and classroom conduct.

Method

Participants. Participants were fifth through eighth grade students at two public schools in New York City. About 93% of the 549 students elected to participate and were not significantly different from nonparticipants in terms of gender, race, age at assessment, or household income. Of the 513 consented students, 3 were excluded on the basis of questionnaire response patterns that upon visual inspection suggested invalid scores (final N = 510 participants, n = 286 from School 1 and n = 224 from School 2, mean age = 11.74, SD = 1.28). Sixty-four percent of participants were Latino, 35% were Black, and 1% was Asian; 52% were girls. (See Table 3 for more details.)

Procedure and measures. Students, parents, and homeroom teachers completed consent forms and self-control and IQ measures within the first 2 months of the school year. 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; Tsukayama et al., 2011). 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 was 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

Table 3
 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	11
Self-control													
1. Self-report	3.79	0.85	—										
2. Parent-report	4.04	0.84	.32***	—									
3. Teacher-report	3.77	1.12	.36***	.45***	—								
4. IQ	0.01	1.00	.06	.11*	.23***	—							
5. Prior-year achievement test	677.84	25.10	.03	.14*	.27***	.45***	—						
6. Spring semester achievement test	683.72	20.58	.16**	.19**	.33***	.46***	.64***	—					
7. Fall semester GPA	80.86	8.42	.31***	.36***	.53***	.42***	.62***	.68***	—				
8. Spring semester GPA	81.26	8.61	.30***	.43***	.56***	.40***	.57***	.66***	.91***	—			
9. Income ^a	\$24,759	\$10,282	-.01	.00	.01	.01	.05	.04	.06	.07	—		
10. Female (52%)			.06	.11*	.28***	.04	.07	.10*	.16***	.17***	.01	—	
11. Black (35%)			-.11*	-.22***	-.19***	-.12**	-.06	-.05	-.04	-.07	-.03	-.03	—
12. Asian (1%)			-.05	.02	.05	.04	.04	.03	.02	.03	.00	.00	-.07

Note. GPA = grade point average.

^a Log-transformed income is used for correlations.

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

month, and 5 = almost never. The observed internal consistency coefficients for the ISC self-control scale were $\alpha = .78$, $.85$, and $.93$, for self-report, parent, and teacher ratings, respectively.

In a validation study (Tsukayama et al., 2011), the ISC demonstrated convergent validity with the SSRS self-control measure used in Study 1 ($r = .62$, $p < .001$) as well as with a widely used trait measure of self-control, the Brief Self-Control Scale (BSCS; Tangney, Baumeister, & Boone, 2004), $r = .71$, $p < .001$. The correlation between the SSRS and BSCS measures was $r = .64$, $p < .001$. To test discriminant validity, Tsukayama et al. (2011) examined correlations between the Openness to Experience subscale of the Big Five Inventory (John & Srivastava, 1999) and the SSRS ($r = .37$, $p < .001$), ISC ($r = .30$, $p < .001$), and BSCS ($r = .40$, $p < .001$) measures of self-control. Following procedures outlined by Meng, Rosenthal, and Rubin (1992), Tsukayama et al. (2011) confirmed that correlations among self-control measures were significantly higher than were correlations between measures of self-control and openness to experience, $ps < .001$.

Intelligence. 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 min. 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 quarterly report card grades for math, science, English, writing, and social studies classes. A single-factor measurement model generally fit the report card grade data well, $\chi^2(5) = 97.30$, $p < .001$, CFI = .95, RMSEA = .19. 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. 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, $\Delta\chi^2(1) = 61.55$, $p < .001$, CFI = .98, RMSEA = .13, but the two factors were very highly correlated, $r = .88$, $p < .001$. Given the large proportion of shared variance between the two factors—and in the interest of consistency with Study 1 as well as prior studies of personality and academic achievement (e.g., Duckworth & Seligman, 2006; Nofle & Robins, 2007)—we calculated GPA for the fall semester as the mean of all subject grades for the first and second quarters, and GPA for the spring semester as the mean of all subject grades for the third and fourth quarters. Likewise, for mediation analyses, we calculated quarterly GPAs as the mean of all subject grades for each quarter.

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 NCLB 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.

Homework completion and classroom conduct. As part of regular school practice, academic subject teachers rated student homework and conduct in each class using a single 5-point scale, where 1 = unsatisfactory, 2 = needs improvement, 3 = satisfactory, 4 = good, and 5 = excellent. A single-factor measurement model largely fit the conduct data well, with the exception of one fit index, $\chi^2(5) = 34.76$, $p < .001$, CFI = .98, RMSEA = .11, and a two-factor model with separate verbal and quantitative conduct factors did not significantly improve fit, $\Delta\chi^2(1) = 0.01$, $p = .93$, CFI = .98, RMSEA = .12. For each student, we averaged conduct grades from all teachers.

Socioeconomic status. Using home addresses in conjunction with U.S. Census Bureau figures, we calculated the estimated median neighborhood household income for each participant. The median household income was \$23,125 among School 1 partici-

pants and \$24,536 among School 2 participants. To reduce skew, we log-transformed income for all subsequent analyses.

Results and Discussion

Examination of continuous variable distributions. Log-transformed income and 2008 and 2009 standardized achievement test scores were slightly leptokurtic: 2.59, 3.87, and 3.54, respectively. All other continuous variables had absolute skew and kurtosis indices less than 1.00. Removing two outliers from the log-transformed income distribution and 17 outliers from the 2008 and 2009 standardized achievement test score distributions reduced the kurtosis indices to .09, $-.01$, and $.95$, respectively. However, because results were virtually identical when these participants were excluded, we included in our final analyses participants with outlying household incomes or test scores.

Predictive validities of self-control and IQ for report card grades and standardized achievement test scores. To estimate the predictive validities of self-control and IQ for spring-semester GPA and standardized achievement test scores, we fit a model with demographic variables, fall semester GPA, and prior-year standardized achievement test scores as covariates. Self-control and IQ were allowed to covary, as were spring-semester GPA and standardized achievement test scores. Correlations among self-report and parent and teacher ratings of self-control ranged from $r = .32$ to $.45$, $ps < .001$, and averaged $r = .38$. Loadings for self-control scores were $.47$ (self-report), $.62$ (parent report), and $.77$ (teacher report). All other constructs, including IQ, were treated as observed variables. The model fit the data well, $\chi^2(22) = 62.37$, $p < .001$, CFI = $.98$, RMSEA = $.06$.

As shown in Table 2, self-control ($\beta = .22$, $p < .001$) but not IQ ($\beta = .01$, $p = .60$) predicted spring semester GPA when controlling for fall semester GPA. In contrast, IQ ($\beta = .12$, $p < .001$) but not self-control ($\beta = -.05$, $p = .44$) predicted standardized achievement test scores when controlling for prior-year standardized achievement test scores.

Using meta-analytic techniques, we compared the predictive validity of self-control for report card grades, controlling for prior report card grades and demographic covariates, across Studies 1 and 2. Specifically, for each study, we converted change in chi-square estimates (when paths from self-control and IQ were constrained to be equal) to Pearson r effect size estimates, converted to the Fisher Z -transformed rs , weighted by sample size, and then combined these and converted back to r . As hypothesized, the predictive validity of self-control across Study 1 and Study 2 was greater than that of IQ, $r = .11$, 95% confidence interval (CI) $[.06, .15]$, $p < .001$. Similarly, the predictive validity of IQ for standardized achievement test scores, controlling for prior standardized achievement test scores, was greater than that of self-control, $r = .14$, 95% CI $[.10, .19]$, $p < .001$.

Improvements in homework completion and classroom conduct mediate the relationship between self-control and report card grades. Following Cole and Maxwell's (2003) recommendations² for testing mediation using longitudinal data, we specified a second order autoregressive model in which self-control and fall semester conduct ratings and grades predicted third-quarter conduct and grades, which in turn predicted fourth-quarter conduct and grades (see Figure 2). 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, $\chi^2(15) = 28.82$, $p = .02$, CFI = $.997$, RMSEA = $.04$. As shown in Figure 2, our hypothesis was supported. Self-control measured in the fall semester predicted increases in conduct ratings from the fall semester to the third quarter, $\beta = .23$, $p = .005$. Third-quarter conduct predicted increases from third- to fourth-quarter report card grades, $\beta = .13$, $p < .001$. The Sobel (1982) test confirmed that the indirect effect of self-control on increases in grades via increases in conduct was significant ($z = 2.31$, $p = .02$). Consistent with full mediation, when taking into account the effects of conduct on grades, self-control did not directly predict increases in grades through the fourth quarter, $\beta = -.04$, $p = .35$.

Study 3

Studies 1 and 2 supported our hypotheses that intelligence disproportionately determines standardized achievement test scores, whereas self-control disproportionately determines report card grades. In Study 3, we surveyed teachers about the content and purpose of these two forms of assessment. Our purpose was to gather further evidence for the competencies differentially assessed by report card grades and standardized achievement test scores.

Method

Participants and procedure. Participants were $N = 57$ teachers from one private ($n = 17$) and two public middle schools ($n = 23$ and $n = 17$). 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, 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. On the basis of questionnaires used in prior survey research on grading practices (Cross & Frary, 1996; Gullickson, 1985), we developed items to assess teachers' judgments of "academic grades you assign to students (not effort or conduct grades)" and "standardized achievement tests." Our questionnaire included three categories of items, labeled *subject material*, *factors unrelated to subject mastery*, 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*.

Content. 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 to report card grades of the following: (a) "prompt and thorough completion of homework," (b) "prompt and reliable

² In 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., fall semester self-control \rightarrow fall semester conduct \rightarrow third-quarter grades \rightarrow fourth-quarter grades). Our estimate of the indirect effects is conservative in this respect. Correlation matrices for mediation model analyses are available upon request.

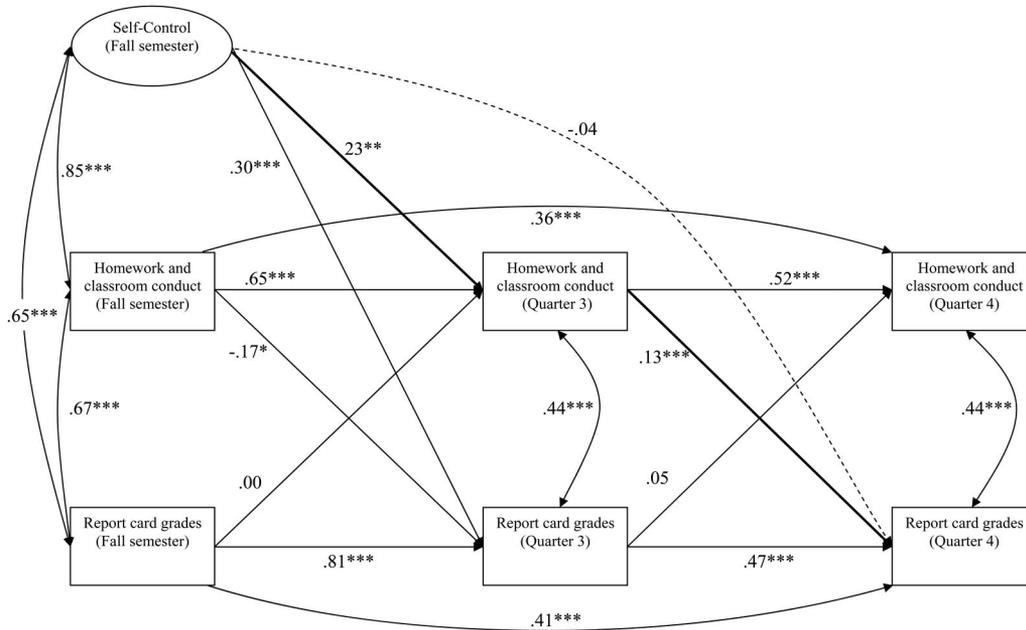


Figure 2. Conduct ratings mediate the relation between self-control and report card grades. Bolded lines represent the indirect effect of self-control on fourth-quarter grades. The dashed line represents the direct effect of self-control on grades accounting for conduct ratings. * $p < .05$. ** $p < .01$. *** $p < .001$.

attendance in my class,” (c) “positive class participation (as opposed to disruptive behavior),” and (d) “positive attitude and effort.”

Purpose of assessment. Teachers rated separately the relevance to report card grades and standardized achievement test scores of four functions: (a) “to summarize mastery of skills and knowledge,” (b) “to provide feedback to students about areas of mastery and weakness,” (c) “to motivate students to work hard,” and (d) “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 ($M = 4.73, SD = 0.49$) much more relevant than mastery of skills and knowledge not taught in class ($M = 3.11, SD = 1.09$) to report card grades, $t(55) = 9.25, p < .001, d = 1.92$. In contrast, teachers felt that standardized achievement test scores were equally determined by skills and knowledge acquired in school ($M = 3.76, SD = 1.16$) and outside of school ($M = 3.67, SD = 1.11$), $t(48) = 0.47, ns, d = 0.08$.

As shown in Figure 4, the relevance of homework completion ($M = 4.32, SD = 0.72$) and 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, $t(55) = 4.00, p < .001, d = 0.60$ and skills and knowledge not taught in class to be less important, $t(53) = 2.88, p = .006, d = 0.50$ than the most closely ranked classroom conduct factor.

As shown in Figure 5, 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, $M = 4.44, SD = 0.76$; $t(56) = 0.13, ns, d = 0.02$. For report card grades, these two functions were more important than motivating students to work hard, $M = 3.98, SD = 1.02$; $t(55) = 2.92, p = .005, d = 0.52$,

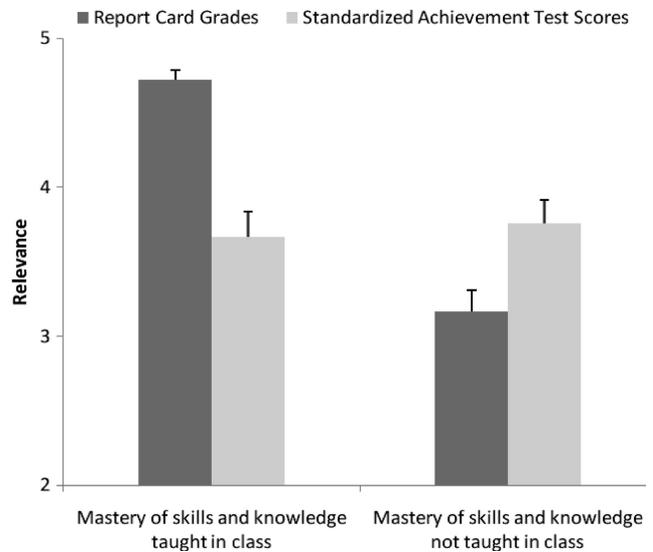


Figure 3. Teacher-reported relevance of skills and knowledge taught in class versus skills and knowledge not taught in class ratings for report card grades and standardized achievement test scores. Average teacher ratings (vertical axis) on a Likert-type scale ranging from 1 (not at all important) to 5 (very important). Error bars signify standard error.

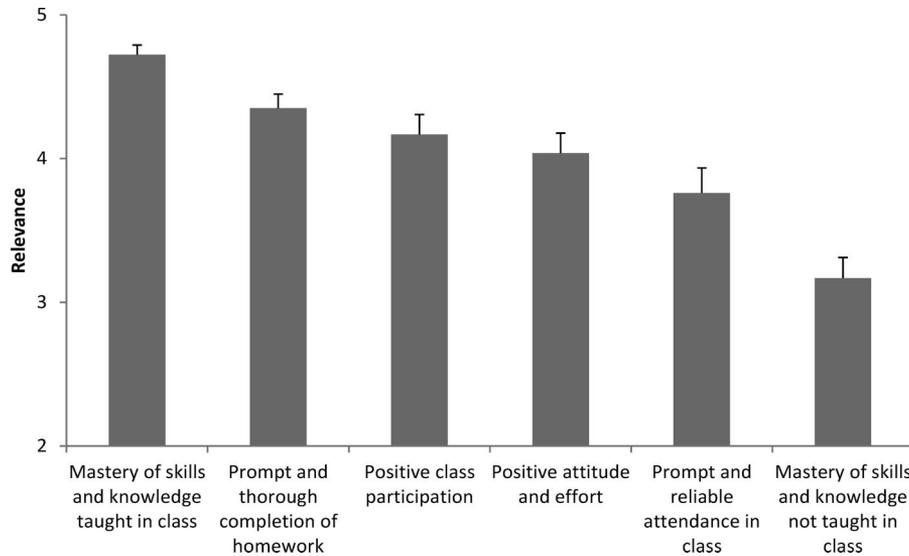


Figure 4. Teacher-reported relevance of subject mastery and student conduct ratings for report card grades. Error bars signify standard error. Average teacher ratings (vertical axis) on a Likert-type scale ranging from 1 (*not at all important*) to 5 (*very important*).

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 5 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, $M = 3.47$, $SD = 1.27$; $t(50) = 1.69$, $p = .10$,

$d = 0.27$. 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, $t(50) = 0.48$, ns , $d = 0.08$. For each of these four purposes, differences in the relative importance to report card grades versus standardized achievement tests were significant, $ps < .001$, $ds \geq .93$.

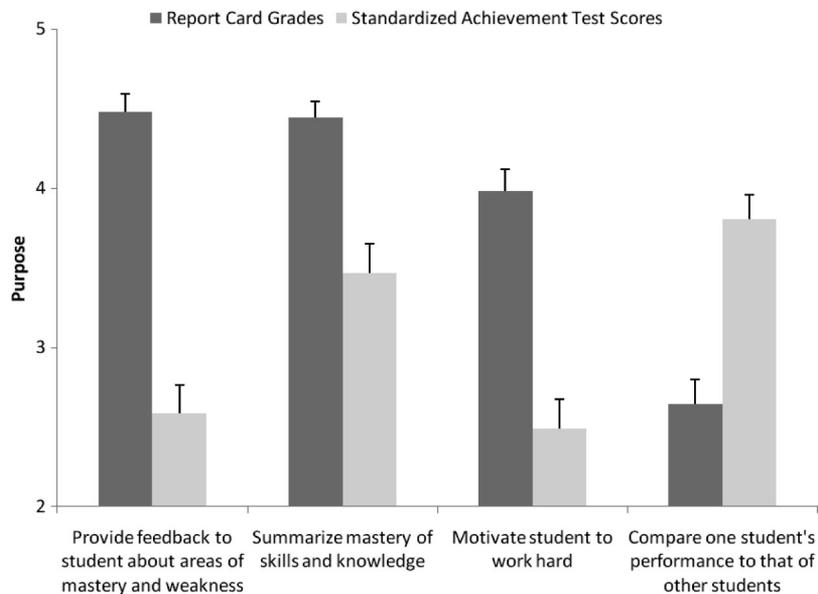


Figure 5. Teacher-reported purpose of assessment ratings for report card grades and standardized achievement tests. Error bars signify standard error. Average teacher ratings (vertical axis) on a Likert-type scale ranging from 1 (*not at all important*) to 5 (*very important*).

General Discussion

We proposed a theoretical model (summarized in Figure 1) 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. Two prospective, longitudinal studies of middle school students supported predictions from this model: Self-control predicted changes in report card grades over time better than did IQ, whereas IQ predicted changes in standardized achievement test scores better than did self-control. As expected, increases in report card grades were mediated in Study 2 by midyear improvements in homework completion and classroom conduct. Teacher judgments in Study 3 provided further support for our model. Specifically, middle school teachers indicated that when determining academic report card grades, they factored in completion of homework assignments, class participation, effort, and attendance. Notably, homework completion and classroom conduct were rated less important to report card grades than were the skills and knowledge that teachers formally taught their students. In contrast, teachers perceived skills and knowledge acquired outside and inside the classroom to be equally relevant to performance on standardized achievement tests.

Teachers surveyed in Study 3 considered standardized achievement tests and report card grades as serving distinct, but complementary, educational purposes. Their perspective resonates with our own view. In particular, we agree that report card grades are better suited to providing timely feedback to students about their level of mastery over the formal curriculum and, further, that report card grades can motivate students to comply with teacher directives, enforcing what Willingham et al. (2002) called the “implicit local contract between teacher and student” (p. 28). Standardized achievement tests, on the other hand, enable administrators and policy makers to sample what students can do in an academic domain, regardless of whether the relevant knowledge and skills were acquired in school (Popham, 1999, 2000). Importantly, 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).

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. As in any empirical investigation, ours had strengths and weaknesses, which suggest directions for future research. First, although we were able to test the hypothesized mediating role of homework completion and classroom conduct for prospective associations between self-control and changes in report card grades, data were not available to confirm that the benefits of intelligence for increases in standardized achievement test scores are mediated by superior performance on problems that require skills and knowledge acquired outside of formal instruction. To directly test this idea would require independent measures of general knowledge and facility with completely novel tasks.

A second limitation concerns the nonexperimental nature of this investigation. Studies 1 and 2 employed longitudinal, prospective

designs and included demographic control variables. Nevertheless, 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 et al., 2008). Recent advances demonstrating that both self-control (Diamond, Barnett, Thomas, & Munro, 2007; Duckworth, Grant, Loew, Oettingen, & Gollwitzer, 2011) and intelligence (Jaeggi, Buschkuhl, Jonides, & Perrig, 2008; Nisbett, 2009) may respond to deliberate intervention suggest that such experimental research may soon be within the realm of possibility.

Finally, 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 verify the degree to which our conclusions generalize to, for instance, older students and students in non-U.S. countries. We hope that such replication studies would follow a similar multisource 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).

Implications

Because high-stakes standardized achievement tests are playing an increasingly prominent role in policy and practice, with schools devoting more and more 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, whereas standardized achievement tests 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 per-

forming 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, 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 as possible. Willingham et al. (2002) have pointed out that no standardized achievement test can, in a few hours, sample in adequate detail the skills and knowledge acquired throughout an entire year of formal instruction. Nevertheless, better alignment in both content and format should reduce unintended effects of general intelligence on standardized achievement test performance 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 (Bill and Melinda Gates Foundation, 2010).

Second, we suggest reviving—and standardizing—the practice of separately grading student effort. At present, the signal sent by report card grades is ambiguous: Does an A grade indicate superior academic mastery, superior effort on homework and classroom conduct, or, given that teachers vary in their grading practices (McMillan et al., 2002), an unknowable amalgam of both? One remedy for the mixed signal sent by report card grades is for teachers to indicate separately 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 objectively measurable effort-related behaviors were separately described on report cards, or if teachers gave a subjective rating of overall student effort, academic grades could then be based solely on demonstrated academic skills and knowledge. In this way, teachers could preserve the motivational function of report card grades while providing accurate information about student mastery of academic skills and knowledge.

Considering success beyond the classroom, a compelling argument can be made for providing feedback about—and thereby encouraging—self-controlled behavior. 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 symptomatology (Eisenberg et al., 2009), enjoy higher levels of positive emotion and life satisfaction and lower levels of negative emotion (Tsukayama et al., 2011) and have more adaptive

relationships with other people (Mischel, Shoda, & Rodriguez, 1989). James (1899) and Aristotle speculated that practicing self-control encourages its development, an idea that has found support in 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 (Brighthouse, 2008), then, in our view, there is good reason for explicitly encouraging self-regulation of attention, behavior, and emotion in the service of long-term goals.

In closing, we suggest that the NCLB 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 report card grades behind in an effort to standardize assessments across teachers, schools, and regions also leaves behind essential information about self-control and the competencies it enables.

References

- 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*, 1773–1802. doi:10.1111/j.1467-6494.2006.00428.x
- Baumeister, R. F., Heatherton, T. F., & Tice, D. M. (1994). *Losing control: How and why people fail at self-regulation*. San Diego, CA: Academic Press.
- Bill and Melinda Gates Foundation. (2010). *Fewer, clearer, higher: Moving forward with consistent, rigorous standards for all students*. Seattle, WA: Author.
- Block, J. (1996). Some jangly remarks on Baumeister and Heatherton. *Psychological Inquiry, 7*, 28–32. doi:10.1207/s15327965pli0701_5
- Borghans, L., Duckworth, A. L., Heckman, J. J., & ter Weel, B. (2008). The economics and psychology of personality traits. *Journal of Human Resources, 43*, 972–1059.
- Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). Test scores and high school grades as predictors. In W. G. Bowne, M. M. Chingos, & M. S. McPherson (Eds.), *Crossing the finish line: Completing college at America's public universities* (pp. 112–133). Princeton, NJ: Princeton University Press.
- Brighthouse, 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* (Vol. 1, pp. 58–71). New York, NY: Wiley-Blackwell. doi:10.1002/9781444307214.ch4
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education, 7*, 279–301. doi:10.1207/s15324818ame0704_2
- 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. doi:10.1037/1082-989X.7.4.403
- 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*, 163–196. doi:10.1111/j.1745-9125.1994.tb01151.x
- 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. doi:10.1207/s15326977ea0302_3
- 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. doi:10.1037/0021-843X.112.4.558

- Cross, L. H., & Frary, R. B. (1996, April). *Hodgepodge grading; Endorsed by students and teachers alike*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York, NY.
- Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007, November 30). Preschool program improves cognitive control. *Science*, *318*, 1387–1388. doi:10.1126/science.1151148
- Duckworth, A. L., Grant, H., Loew, B., Oettingen, G., & Gollwitzer, P. M. (2011). Self-regulation strategies improve self-discipline in adolescents: Benefits of mental contrasting and implementation intention. *Educational Psychology*, *31*, 17–26. doi:10.1080/01443410.2010.506003
- Duckworth, A. L., & Kern, M. (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, *45*, 259–268. doi:10.1016/j.jrp.2011.02.004
- Duckworth, A. L., & Seligman, M. E. P. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological Science*, *16*, 939–944. doi:10.1111/j.1467-9280.2005.01641.x
- 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. doi:10.1037/0022-0663.98.1.198
- Duckworth, A. L., Tsukayama, E., & Geier, A. (2010). Self-controlled children stay leaner in the transition to adolescence. *Appetite*, *54*, 304–308. doi:10.1016/j.appet.2009.11.016
- Duckworth, A. L., Tsukayama, E., & May, H. (2010). Establishing causality using longitudinal hierarchical linear modeling: An illustration predicting achievement from self-control. *Social Psychological & Personality Science*, *1*, 311–317. doi:10.1177/1948550609359707
- Eccles, J. (2004). Schools, academic motivation, and stage–environment fit. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (2nd ed., pp. 125–153). Hoboken, NJ: Wiley.
- Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & Mac Iver, D. (1993). Development during adolescence: The impact of stage–environment fit on young adolescents' experiences in schools and in families. *American Psychologist*, *48*, 90–101. doi:10.1037/0003-066X.48.2.90
- Eisenberg, N., Smith, C. L., Sadovsky, A., & Spinrad, T. (Eds.). (2004). *Effortful control: Relations with emotion regulation, adjustment, and socialization in childhood*. New York, NY: Guilford Press.
- 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. doi:10.1037/a0016213
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum-likelihood estimation for missing data in structural equation models. *Structural Equation Modeling*, *8*, 430–457. doi:10.1207/S15328007SEM0803_5
- 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* (pp. 303–326). Mahwah, NJ: Erlbaum.
- 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* (No. 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: Erlbaum.
- Gottfredson, L. S. (2004, Summer). 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. *The Journal of Educational Research*, *79*, 96–100.
- Harter, S. (1985). *Self-Perception Profile for Children*. Unpublished manuscript, University of Denver, Denver, CO.
- 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 Sciences, USA*, *105*, 6829–6833. doi:10.1073/pnas.0801268105
- James, W. (1899). *Talks to teachers on psychology and to students on some of life's ideals*. New York, NY: Holt. doi:10.1037/10814-000
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). New York, NY: Guilford Press.
- Kelley, T. L. (1927). *Interpretation of educational measurements*. Oxford, England: World Book.
- 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. doi:10.1207/S15328007SEM1003_1
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.
- Kohn, A. (2000, September 27). Standardized testing and its victims. *Education Week*. Retrieved from <http://www.alfiekohn.org/teaching/edweek/staiv.htm>
- Kuncel, N. R., & Hezlett, S. A. (2007, February 23). Standardized tests predict graduate students' success. *Science*, *315*, 1080–1081. doi:10.1126/science.1136618
- Kuncel, N. R., Ones, D. S., & Sackett, P. R. (2010). Individual differences as predictors of work, educational, and broad life outcomes. *Personality and Individual Differences*, *49*, 331–336. doi:10.1016/j.paid.2010.03.042
- Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904). "General intelligence," Objectively determined and measured." *Journal of Personality and Social Psychology*, *86*, 96–111. doi:10.1037/0022-3514.86.1.96
- Lubinski, D. (2009). Exceptional cognitive ability: The phenotype. *Behavior Genetics*, *39*, 350–358. doi:10.1007/s10519-009-9273-0
- Lynam, D. R., Caspi, A., Moffitt, T. E., Wikstrom, P., 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. doi:10.1037/0021-843X.109.4.563
- Mather, N. (1991). *An instructional guide to the Woodcock–Johnson Psycho-Educational Battery—Revised*. New York, NY: Wiley.
- McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary teachers' classroom assessment and grading practices. *The Journal of Educational Research*, *95*, 203–213. doi:10.1080/00220670209596593
- McMurrer, J. (2007). *Choices, changes, and challenges: Curriculum and instruction in the NCLB era*. Washington, DC: Center on Education Policy.
- Meng, X., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*, 172–175. doi:10.1037/0033-2909.111.1.172
- Miles, J., & Shevlin, M. (2007). A time and a place for incremental fit indices. *Personality and Individual Differences*, *42*, 869–874. doi:10.1016/j.paid.2006.09.022
- Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989, May 26). Delay of gratification in children. *Science*, *244*, 933–938. doi:10.1126/science.2658056
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H. L., . . . Caspi, A. (2011). A gradient of childhood self-

- control predicts health, wealth, and public safety. *PNAS: Proceedings of the National Academy of Sciences, USA*, *108*, 2693–2698. doi:10.1073/pnas.1010076108
- Muraven, M., Baumeister, R. F., & Tice, D. M. (1999). Longitudinal improvement of self-regulation through practice: Building self-control strength through repeated exercise. *The Journal of Social Psychology*, *139*, 446–457. doi:10.1080/00224549909598404
- 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. doi:10.1037/0003-066X.51.2.77
- Nisbett, R. E. (2009, February 7). Education is all in your mind. *The New York Times*. Retrieved from <http://www.nytimes.com/2009/02/08/opinion/08nisbett.html>
- No Child Left Behind Act of 2001, Pub. L. No. 107–110 (2002).
- 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. doi:10.1037/0022-3514.93.1.116
- 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. doi:10.1057/palgrave.jt.5740069
- 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, MA: Allyn & Bacon.
- Psychological Corporation. (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.
- Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological Bulletin*, *126*, 3–25. doi:10.1037/0033-2909.126.1.3
- Ross, L., & Nisbett, R. E. (1991). *The person and the situation: Perspectives of social psychology*. Philadelphia, PA: Temple University Press.
- Rothbart, M. K., Ellis, L. K., & Posner, M. I. (2004). Temperament and self-regulation. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation* (pp. 357–370). New York, NY: The Guilford Press.
- Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High-stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist*, *63*, 215–227. doi:10.1037/0003-066X.63.4.215
- Salthouse, T. A., & Pink, J. E. (2008). Why is working memory related to fluid intelligence? *Psychonomic Bulletin & Review*, *15*, 364–371. doi:10.3758/PBR.15.2.364
- Simmons, R. G., & Blyth, D. A. (1987). *Moving into adolescence: The impact of pubertal change and school context*. New York, NY: Aldine De Gruyter.
- Sobel, M. E. (1982). Asymptotic intervals for indirect effects in structural equations models. In S. Leinhardt (Ed.), *Sociological methodology* (pp. 290–312). San Francisco, CA: Jossey-Bass.
- Stipek, D., & Douglas, M. (1989). Developmental change in children's assessment of intellectual competence. *Child Development*, *60*, 521–538. doi:10.2307/1130719
- Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, *72*, 271–324. doi:10.1111/j.0022-3506.2004.00263.x
- Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurements*. Oxford, England: Science Press. doi:10.1037/13283-000
- Tsukayama, E., Duckworth, A. L., & Kim, B. E. (2011). *Domain-specific impulsivity in school-age children*. Manuscript submitted for publication.
- 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. doi:10.1001/archpediatrics.2010.97
- Valiente, C., Lemery-Chalfant, K., Swanson, J., & Reiser, M. (2008). Prediction of children's academic competence from their effortful control, relationships, and classroom participation. *Journal of Educational Psychology*, *100*, 67–77. doi:10.1037/0022-0663.100.1.67
- 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. doi:10.1177/1073191107302062
- Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, *39*, 1–37. doi:10.1111/j.1745-3984.2002.tb01133.x
- 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. doi:10.1037/0022-006X.70.4.986
- Wong, M. M., & Csikszentmihalyi, M. (1991). Motivation and academic achievement: The effects of personality traits and the duality of experience. *Journal of Personality*, *59*, 539–574. doi:10.1111/j.1467-6494.1991.tb00259.x
- Zimmerman, B. J., & Kitsantas, A. (2005). The hidden dimension of personal competence. In A. J. Elliot & C. S. Dweck (Eds.), *Handbook of competence and motivation* (pp. 509–526). New York, NY: Guilford Press.

Received January 26, 2010

Revision received September 24, 2011

Accepted October 17, 2011 ■