Journal of Educational Psychology
2006, Vol. 98, No. 1, 198–208

0022-0663/06/$12.00 DOI: 10.1037/0022-0663.98.1.198

Self-Discipline Gives Girls the Edge: Gender in Self-Discipline, Grades, and Achievement Test Scores

Angela Lee Duckworth and Martin E. P. Seligman
University of Pennsylvania

Throughout elementary, middle, and high school, girls earn higher grades than boys in all major subjects. Girls, however, do not outperform boys on achievement or IQ tests. To date, explanations for the underprediction of girls' GPAs by standardized tests have focused on gender differences favoring boys on such tests. The authors' investigation suggests an additional explanation: Girls are more self-disciplined, and this advantage is more relevant to report card grades than to achievement or aptitude tests. Eighth-grade girls at an urban magnet school were more self-disciplined than their male counterparts according to delay of gratification measures and self-report, teacher, and parent ratings. Whereas girls earned higher grades in all courses, they did only marginally better on an achievement test and worse on an IQ test. Mediation analyses suggested girls earned higher GPAs at least in part because they were more self-disciplined.

Keywords: gender differences, self-control, self-discipline, academic achievement, standardized tests

Why do girls get better grades than boys? Throughout elementary, middle, and high school, girls earn higher grades than boys in all major subjects, including math and science (American Association of University Women Educational Foundation [AAUWEF], 1998; N. S. Cole, 1997; Pomerantz, Altermatt, & Saxon, 2002), and girls graduate from high school with higher overall GPAs than their male counterparts (Perkins, Kleiner, Roey, & Brown, 2004). In most colleges and in most subjects, women continue to outperform men (Clark & Grandy, 1984; Kimball, 1989; Mau & Lynn, 2001; Willingham & Cole, 1997). However, girls do not have higher IQs, and they score lower on some (but not all) standardized tests, including the SAT, ACT, and AP exams (AAUWEF, 1998). Women typically earn higher grades than predicted by their performance on ability or achievement tests, and men earn lower than expected grades; these phenomena have been termed *underprediction* and *overprediction*, respectively. Underprediction has been extensively investigated among undergraduates (e.g., Hunter, Schmidt, & Rauschenberger, 1984; Jensen, 1980; Linn, 1973; Stricker, Rock, & Burton, 1993) but not in younger populations.

Several studies have shown that the underprediction of college grades based on SAT or ACT scores is small in size but consistent. For instance, in a sample of 41 colleges and universities, Clark and Grandy (1984) found that a common regression equation predicting freshman GPA on a standard 4.0 scale from SAT scores and high school GPA or class rank underpredicted the performance of female students by 0.05 points and overpredicted the performance of male students by 0.04 points. Similarly, at a large state university, Stricker et al. (1993) found that SAT scores underpredicted the first-semester GPA of female students by 0.10 and overpredicted the first-semester GPA of male students by 0.11.

Although a full explanation of underprediction must address both the female advantage in grades and the male advantage on standardized tests, historically, more attention has been given to the male advantage. For instance, Steele and colleagues have suggested that because men are expected to outperform women on standardized tests, women experience *stereotype threat*, or added pressure that interferes with test performance (e.g., Steele, 1997; Spencer, Steele, & Quinn, 1999). Loewen, Rosser, and Katzman (1988) have argued that SAT math questions refer to situations more familiar to boys (e.g., sports teams and male recreational activities). Willingham and Cole (1997) have pointed out that boys tend to excel on multiple-choice questions, the most prevalent format for standardized tests, whereas girls outperform boys on free-response (e.g., essay) assessments. This explanation is consistent with the finding that girls surpass boys on tests of grammar and spelling (Feingold, 1988). Finally, it is possible that female students are confident answering questions about familiar material but discouraged by the novel problems presented on standardized tests (Kimball, 1989). Indeed, the gender gap favoring boys increases with the complexity of the standardized test (AAUWEF, 1998).

The above explanations do not address why female students bring home better report cards, however. Elliott and Strenta (1988) and Young (1991) have suggested that underprediction may result from female college students' choosing easier courses with more lenient grading systems. However, Leonard and Jiang (1999)

Angela Lee Duckworth and Martin E. P. Seligman, Department of Psychology, University of Pennsylvania.

This work was supported by National Institute of Mental Health Grants MH 52270 and MH 63430, the John Templeton Foundation, and U.S. Department of Education Grant R215S020045.

John Sabini passed away on July 15, 2005. We gratefully acknowledge his contributions to our research and to our field. We thank Dianne Chambless, Michelle Goldwin, Sara Jaffee, Edgar Knizhnik, and Lawrence Stricker for their helpful comments.

Correspondence concerning this article should be addressed to Angela Lee Duckworth, Department of Psychology, University of Pennsylvania, 3701 Market Street, Suite 200, Philadelphia, PA 19104. E-mail: angela_duckworth@yahoo.com

found that underprediction of female grades at the University of California at Berkeley by SAT scores persisted even when they corrected for gender differences in fields of study and for sample selection bias. Similarly, Stricker and colleagues (1993) found that adjusting GPA for course-by-course differences in grading standards did not affect underprediction at a large state university. Relevantly, Stricker et al. found that adding self-report variables tapping academic preparation and studiousness—behaviors related to self-discipline, which we examine in detail in this investigation—significantly reduced the amount of over- and underprediction. Further, the preference for easier courses does not explain why girls earn higher grades than boys as early as elementary school, where most students take identical courses.

This article examines whether underprediction might be explained, at least in part, by gender differences in self-discipline. To explore this possibility, one must have a clear definition of self-discipline and a reliable method of measuring it. Many authors have pointed out that the construct of self-control is often loosely defined and that definitions and measurement approaches vary widely (e.g., Evenden, 1999; Gerbing, Ahadi, & Patton, 1987; Parker & Bagby, 1997; Zaparniuk & Taylor, 1997). Block (1996) and Gwin (1997) have noted that the self-control literature is particularly afflicted with both the jangle (Kelley, 1927) and jingle (Thorndike, 1904) fallacies. In the former case, different terms are used to describe common underlying conceptions (e.g., *self-control* and *self-regulation*). In the latter, common terms refer to different underlying conceptions (e.g., *self-control*, defined as the ability to delay gratification, vs. *self-control*, defined as risk aversion).

We use the terms *self-discipline* and *self-control* interchangeably, defining both as the ability to suppress prepotent responses in the service of a higher goal and further specifying that such a choice is not automatic but rather requires conscious effort. Examples of self-discipline include deliberately modulating one's anger rather than having a temper tantrum, reading test instructions before proceeding to the questions, paying attention to a teacher rather than daydreaming, saving money so that it can accumulate interest in the bank, choosing homework over TV, and persisting on long-term assignments despite boredom and frustration.

Not surprisingly, the trait of doing what one intends to do when a more pleasurable alternative beckons has been shown to predict academic performance (Hogan & Weiss, 1974; Tangney, Baumeister, & Boone, 2004; Wolfe & Johnson, 1995). We reason that the extent to which self-discipline predicts academic performance depends on how one measures performance. Our logic is as follows: The reliable variance of any academic performance assessment comprises both common factor variance, corresponding to skills and knowledge, and unique variance, corresponding to the particularities of each assessment method. So, it is true that report card grades, achievement tests, and aptitude tests all evaluate skills and knowledge, but it is also true that there is significant method variance unique to each assessment modality.

In particular, these methods differ in the degree to which self-discipline affects performance. Report card grades typically reflect the ability to study for exams, complete homework assignments and long-term projects on time, and prepare for class discussions. Thus, grades depend heavily on the ability to sustain effort and concentration despite boredom, fatigue, and innumerable distractions over the course of an academic year. In contrast, achievement tests require sustained effort for only a few hours in a testing situation specifically designed to minimize distractions. Finally, paper-and-pencil IQ tests such as the one used in this study require sustained effort for less than 1 hr. We imagine, therefore, that self-discipline should help a student most on report card grades, somewhat on achievement tests, and minimally on brief aptitude tests.

Are girls more self-disciplined than boys? In a series of meta-analyses of 33 delay of gratification studies, Silverman (2003) found a small but reliable advantage favoring women and girls ($r = .06$). It is important to note that the female advantage was almost twice as large when continuous ($r = .10$) rather than dichotomous ($r = .02$) measures were used. Silverman cautioned, "This difference may go undetected when measurement instruments lack precision and when sample sizes are small" (pp. 456–457). Consistent with Silverman's conclusions, Walsh (1967) found a dichotomous measure failed to detect a gender difference in self-discipline that, among the same participants, was evident when a more precise, continuous measure was used. Specifically, when left alone in a room, 6- to 8-year-old girls were as likely as same-aged boys to touch forbidden toys; however, a significant gender difference favoring girls was evident in the number of minutes that elapsed before succumbing to temptation.

In a series of meta-analyses of gender differences in personality, Feingold (1994) cautioned that "different measures of the same trait do not assess exactly the same latent dimension" (p. 447). This caution is particularly relevant for the construct of self-control, for which there is yet no consensual definition. For example, Feingold found that males scored higher than females (ds = .17 to .22, comparable with $r = .10$ found by Silverman, 2003) on impulsiveness on both the NEO Personality Inventory (Costa & McCrae, 1989, 1992) and the Personality Research Form (Jackson, 1974), but he found no gender differences ($d = .03$, comparable with $r = .015$) on the Gordon Personality Inventory Cautiousness subscale (Gordon, 1978). Thus, to detect a gender difference in self-control, if indeed it exists, one needs a defensible definition of the construct of self-control, measures that operationalize the construct in a consistent way, and a strategy for assessment that maximizes reliability.

In the current investigation, we increased the validity of our self-discipline measure by creating a composite self-discipline score from parent, teacher, and self-report questionnaires and delay of gratification measures. Given the range of behaviors that require self-discipline and the limitations of any single measurement tool, we agree with Lucas and Baird (2006) that a comprehensive, multimethod battery of assessments is more reliable and meaningful than any component measure alone. Another distinctive aspect of our research design was our research sample. Unlike other studies of underprediction in college populations, we conducted our investigation with eighth graders at a school where all students take identical courses except math, which is offered at an advanced and basic level. Thus, we could rule out the possibility that girls earned higher grades than boys simply because they chose easier courses. Because this academically competitive school recruits students likely to attend 4-year colleges, the sample is comparable, except in age, with those used in prior studies of underprediction. Finally, by replicating our study with a second cohort of eighth graders recruited the following year from the same school, we were able to confirm the findings of our first study. In

summary, our design addressed whether underprediction extends to college-bound adolescents, whether the effect is maintained across subjects and regardless of course difficulty, and whether underprediction can be explained by a female advantage in trait self-discipline.

Specifically, Study 1 tested the following hypotheses:

Hypothesis 1: Adolescent girls earn higher grades in Algebra I, Algebra II, English, and social studies than boys.

Hypothesis 2: The female advantage in report card grades is higher than the female advantage in achievement test scores.

Hypothesis 3: Achievement test scores underpredict the report card grades of girls.

Hypothesis 4: Girls are more self-disciplined than boys.

Hypothesis 5: Self-discipline predicts report card grades better than it predicts achievement test scores.

Hypothesis 6: Self-discipline mediates the relationship between gender and report card grades, even when controlling for achievement test scores.

In Study 2, we replicated Study 1, improving our composite measure by adding a behavioral measure of delay of gratification and, in addition, administering an IQ test to determine whether the relationships in Hypotheses 2, 3, and 5 were even more dramatic for aptitude than for achievement test scores.

Study 1

Method

Participants

The participants were eighth-grade students from a socioeconomically and ethnically diverse magnet public school in a city in the northeast. Fifth-grade students are admitted to this middle school on the basis of prior grades and standardized test scores.

About 70% of the school's 198 eighth-grade students ($n = 140$) elected to participate in a longitudinal study of character strengths and academic achievement in children. Signed child assent and parent consent forms, which assured participants of the confidentiality of their data, were received for all participants. In mid-November 2002, when self-discipline measures were administered, the mean age of participants was 13.4 years ($SD = 0.4$). Fifty-five percent of participants were Caucasian, 32.1% were Black, 8.6% were Asian, 3.6% were Latino, and 0.7% were American Indian. Fifty-six percent of participants were female. Thirteen percent of participants were from low-income families, as indicated by participation in the federal lunch program. About 16% ($n = 13$) of girls and 23% ($n = 14$) of boys were enrolled in Algebra II; about 84% ($n = 66$) of girls and 77% ($n = 47$) of boys were enrolled in Algebra I.

Procedure

In Fall 2002, we collected self-report, parent, and teacher questionnaires in addition to delay of gratification data. We received completed questionnaires for 96%, 100%, and 96% of students, teachers, and parents, respectively. In Spring 2003, we recorded report card grades, school attendance, and standardized achievement test scores from school records.

Measures

Our battery of self-discipline measures included previously validated measures that operationalize self-discipline as the ability to choose successfully among conflicting impulses. Internal consistency coefficients were good for all measures and are provided in Table 1. Scores for some measures were recoded for statistical analyses, such that for all measures, higher scores indicated higher self-discipline.

Self-report questionnaires. We administered two widely used self-report measures of self-discipline. The Impulsivity subscale of the Eysenck I_6 Junior Questionnaire (Eysenck, Easting, & Pearson, 1984) was designed exclusively for children and includes 23 *yes–no* questions about doing and saying things impulsively (e.g., "Do you save regularly?" "Do you mostly speak before thinking things out?"). The Brief Self-Control Scale (Tangney et al., 2004) is a 13-item questionnaire designed for adults, but it is face valid for adolescents. Items are endorsed on a 5-point scale, where 1 = *not like me at all* and 5 = *very much like me* (e.g., "I have a hard time breaking bad habits" and "I do certain things that are bad for me, if they are fun").

Teacher and parent questionnaires. Parents and teachers completed the Self-Control Rating Scale (Kendall & Wilcox, 1979): This 33-item questionnaire asks the rater to assess children on a 7-point scale, where 1 = *maximally self-controlled*, 4 = *the average child*, and 7 = *maximally impulsive*. Items tap the ability to inhibit behavior, follow rules, and control impulsive reactions. To avoid confounding teacher ratings and teacher-determined grades, students' homeroom advisors, rather than course teachers, completed the questionnaires. To accommodate for different interpretations of an average student, we standardized each teacher's scores about his or her own mean prior to all statistical analyses.

Delay of gratification questionnaire. Students completed a paper-and-pencil measure called the Kirby Delay-Discounting Rate Monetary Choice

Table 1
Summary Statistics for Self-Discipline Measures in Study 1

Measure	α	Girls		Boys		
		M	*SD*	*M*	*SD*	*d*
Eysenck I_6 Junior Questionnaire Impulsivity subscale[a]	.80	10.49	4.46	12.72	4.75	.49**
Brief Self-Control Scale self-report	.83	3.26	0.71	2.88	0.73	.52**
Self-Control Rating Scale teacher report	.99	5.92	1.20	5.13	1.33	.78***
Self-Control Rating Scale parent report	.96	5.42	0.98	5.12	0.84	.32**
Kirby Delay-Discounting Rate Monetary Choice Questionnaire[a]	.98	0.02	0.02	0.04	0.07	.49
Composite self-discipline score	.91	0.19	0.62	−0.29	0.73	.71***

[a] Higher scores for these measures indicate lower self-discipline.
** $p < .01$. *** $p < .001$.

Questionnaire (Kirby, Petry, & Bickel, 1999). This questionnaire contains 27 questions posing hypothetical choices between smaller, immediate rewards and larger, delayed rewards (e.g., "Would you prefer $55 today or $75 in 61 days?"). From these responses, we calculated a discounting rate (k), a parameter that reflects the degree to which future rewards are diminished in value as a function of the delay that must be endured to receive them. Specifically, the diminished subjective value of the delayed reward is expressed by the following equation,

$$V = \frac{A}{1 + kd}$$

where V is the subjective value of the delayed reward, A is the subjective value of the immediate reward, k is the discount rate, and d is the delay. Internal consistency was calculated as the proportion of responses that were consistent with calculated k values. To normalize the distribution of scores, we used a natural log transformation of k for all statistical analyses.

Report card grades, school attendance, and standardized achievement test scores. From school records, we recorded math course level (Algebra I or Algebra II) and final report card grades for English, math, and social studies classes. For comparison with previously published studies, we converted GPAs from a 100-point scale to a standard 4.0 scale (i.e., Grades 97 to 100 = 4.3; Grades 93 to 96.99 = 4.0; Grades 90 to 92.99 = 3.7, etc.). We also recorded year-end attendance and Spring 2003 TerraNova Second Edition California Achievement Test (CTB/McGraw-Hill, 2001) normal curve equivalent scores in reading, language arts, and math. The TerraNova Second Edition California Achievement Test is a widely used standardized achievement test that includes selected-response and constructed-response items; scores are referenced against a nationwide normative sample of students.

Results

Gender Differences in Report Card Grades and Achievement Test Scores

As shown in Table 2, girls earned significantly higher final grades in Algebra I, English, and social studies than did boys. Girls also earned higher final grades in Algebra II, but because of the reduced sample size in the math subgroups, this difference did not reach statistical significance. Effect sizes ranged from $d = .48$ for Algebra II to $d = .70$ for English. Girls also outperformed boys on the standardized achievement test ($d = .30$), but, consistent with our predictions, this advantage was half that for overall GPA ($d = .66$) and not statistically significant.

Achievement Test Scores Underpredict Female GPA

Over- and underprediction were quantified as the difference between actual GPA and GPA predicted by test scores. We calculated over- and underprediction using five separate equations for Algebra II, Algebra I, English, social studies, or cumulative GPA as the predicted variable and achievement test score as the predictor. Then, to quantify the magnitude of over- or underprediction, we substituted into these regression equations the appropriate achievement test means for boys or girls, respectively. Over- and underprediction were quantified as the difference between the predicted mean and the actual mean for boys and girls, respectively. As summarized in Table 3, achievement test scores underpredicted female report card grades. Underprediction ranged from −0.11 for Algebra I to −0.18 for Algebra II, and for overall GPA, it was −0.10.

Girls Are More Self-Disciplined Than Boys

Girls were more self-disciplined than boys on every measure, and for four of the five measures, this difference was statistically significant. As shown in Table 1, gender differences were greatest on teacher ratings of self-discipline ($d = .78$) and smallest on parent ratings ($d = .32$). As shown in Table 4, correlations among self-discipline measures were positive and ranged from $r = .12$ to .66, with an average of $r = .31$.

Given that self-discipline measures converged, we created a composite self-discipline score to increase validity and reduce multicollinearity. This composite self-discipline score was used in all subsequent analyses. We began by creating a composite self-reported self-discipline score as the mean of standardized scores for the Impulsivity subscale of the Eysenck I_6 Junior Questionnaire and the Brief Self-Control Scale. We then standardized this score and averaged it with standardized scores for teacher, parent, and delay of gratification measures. The internal reliability of this linear combination was $r = .96$, according to a formula specific to linear combinations of standardized scores (Nunnally, 1978). Comparing composite scores of self-discipline, girls were more self-disciplined than boys, $t(138) = 4.12$, $p < .001$, $d = .71$. As shown in Table 5, composite self-discipline correlated significantly with overall GPA ($r = .57$, $p < .001$) and less robustly with achievement test scores ($r = .29$, $p = .001$).

Table 2
Summary Statistics for Academic Performance and IQ Variables in Study 1

	Girls		Boys		
Measure	*M*	*SD*	*M*	*SD*	*d*
Algebra II final grade ($n = 27$)	3.49	0.66	3.14	0.69	.48
Algebra I final grade ($n = 111$)	3.13	0.96	2.58	0.92	.60**
English final grade	3.46	0.71	2.92	0.78	.70***
Social studies final grade	3.65	0.46	3.34	0.56	.60**
Overall GPA	3.37	0.65	2.94	0.69	.66***
TerraNova Second Edition California Achievement Test normal curve equivalent score	77.35	9.78	74.18	11.38	.30

Note. $n = 140$, unless otherwise indicated.
** $p < .01$. *** $p < .001$.

Table 3
Over- and Underprediction of Final Course Grades and GPA From Achievement Test Scores in Study 1

Measure	Girls	Boys
Algebra II grade	−0.18	0.16
Algebra I grade	−0.11	0.15
English grade	−0.14	0.25
Social studies grade	−0.12	0.09
Overall GPA	−0.10	0.20

Table 5
Intercorrelations Among Gender, Composite Self-Discipline, Achievement Test Scores, and Overall GPA in Study 1

Measure	1	2	3	4
1. Gender	—	.33***	.15	.32***
2. Composite self-discipline score		—	.29**	.57***
3. Achievement test score			—	.66***
4. Overall GPA				—

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

Self-Discipline as a Mediator for Gender and Grades

Three criteria must be met for a variable to be considered a mediator: The independent variable must predict the mediator, the independent variable must predict the dependent variable, and the mediator must predict the dependent variable when the independent variable is held constant. Baron and Kenny (1986) have outlined procedures for confirming these three relationships and then determining whether the relationship between the independent variable and the dependent variable is significantly diminished when controlling for the mediator. We conducted two separate mediation analyses to determine whether girls earned higher report card grades because they were more self-disciplined and, second, whether this advantage in self-discipline accounted for underprediction of grades by achievement test scores. Where appropriate, we report part correlations, which describe the relationship between predictors to outcome, controlling for the effects of other variables in the equation.

Composite self-discipline mediated the relationship between gender and final GPA. Regressing self-discipline on gender, we confirmed that gender predicted self-discipline ($\beta = .33$, $p < .001$). Second, we found that self-discipline predicted overall GPA when controlling for gender ($\beta = .50$, part $r = .47$, $p < .001$). Finally, in a hierarchical multiple regression predicting overall GPA, gender was entered in Step 1 and found to be a significant predictor ($\beta = .31$, $p < .001$), but when self-discipline was added in Step 2, the β for gender fell 55% and was no longer significant ($\beta = .14$, part $r = .13$, $p = .08$). A test for the significance of the decrement in the gender regression coefficient yielded $t(135) = 3.53$, $p < .001$ (see Table 6).

Similarly, a female advantage in self-discipline explained underprediction of GPA by achievement test scores. Gender predicted self-discipline when controlling for achievement test scores ($\beta = .31$, $p < .001$). Second, self-discipline predicted overall GPA when controlling for gender and achievement test scores ($\beta = .36$, part $r = .32$, $p < .001$). Finally, in a hierarchical multiple regression predicting overall GPA, gender was a significant predictor when controlling for achievement test scores ($\beta = .22$, part $r = .22$, $p < .001$) in Step 1, but when self-discipline was added in Step 2, the β for gender fell 50% and was no longer significant ($\beta = .11$, part $r = .10$, $p = .08$). A test for the significance of the decrement in the gender regression coefficient yielded $t(135) = 3.35$, $p < .001$ (see Table 7).

Discussion

Study 1 results supported Hypotheses 1–6. Adolescent girls earned substantially higher course grades than their male classmates, and the gender difference in overall GPA was more than twice that in achievement test scores. Consequently, achievement test scores underpredicted female GPAs and overpredicted male GPAs; the magnitude of under- and overprediction in Study 1 was small but comparable with prior studies showing SAT scores do underpredict the first-year GPAs of college students. According to parents, teachers, and the students themselves, girls were more self-disciplined than boys, and composite self-discipline scores mediated the relationship between gender and overall GPA. Thus, as predicted, the female advantage in self-discipline made a bigger difference in report card grades than in achievement test scores.

To confirm these findings, we replicated our research design in Study 2 with a consecutive cohort of students from the same middle school. We added an IQ test in order to compare the relationships among aptitude and achievement tests, gender, self-discipline, and report card grades. In addition, we considered the possibility that the girls in Study 1 were rated higher in self-discipline only because of societal norms suggesting that girls

Table 4
Intercorrelations Among Self-Discipline Measures in Study 1

Measure of self-discipline	1	2	3	4	5
1. Eysenck I_6 Junior Questionnaire Impulsivity subscale	—	.66**	.27**	.20*	.13
2. Brief Self-Control Scale self-report		—	.38**	.37**	.16
3. Brief Self-Control Scale teacher report			—	.48**	.32**
4. Brief Self-Control Scale parent report				—	.12
5. Kirby Delay-Discounting Rate Monetary Choice Questionnaire					—

Note. Scores for some measures were recoded such that for all measures, higher scores indicated higher self-discipline.
* $p < .05$. ** $p < .01$.

Table 6
Summary of Hierarchical Multiple Regression Predicting Overall GPA From Gender and Self-Discipline in Study 1

Variable	*B*	*SE B*	β	part *r*
Step 1				
Gender	0.44	0.11	.31	.31***
Step 2				
Gender	0.19	0.11	.14	.13
Composite self-discipline	0.46	0.07	.50	.47***

Note. Step 1: $R^2 = .10$, $p < .001$; Step 2: $\Delta R^2 = .22$, $p < .001$.
*** $p < .001$.

ought to be more disciplined. That is, we admitted the possibility—however unlikely—that parents, teachers, and the students themselves all systematically overrated the ability of girls (or underrated the ability of boys) to break bad habits, resist temptation, concentrate, work effectively toward long-term goals, and so on. Thus, as an objective measure of self-discipline, we included a behavioral delay of gratification task of our own design in which participants chose to receive either $1 immediately or $2 a week later as a gift for their participation in our study.

Study 2

Method

Participants

Participants in Study 2 were eighth-grade students from the same school as in Study 1, recruited 1 academic year after Study 1. About 83% of students ($n = 164$) elected to participate. In mid-October 2003, when self-discipline measures were administered, the mean age of participants was 13.8 years ($SD = 0.5$). Fifty-two percent of participants were Caucasian, 31.1% were Black, 12.2% were Asian, 4.3% were Latino, and 0.6% were American Indian. Fifty-four percent of participants were female, and 21% of participants were from low-income families. About 17% ($n = 15$) of girls and 21% ($n = 16$) of boys were enrolled in Algebra II; about 83% ($n = 74$) of girls and 79% ($n = 59$) of boys were enrolled in Algebra I.

Procedure

In Fall 2003, we collected self-report, parent, and teacher questionnaires; delay of gratification data; and IQ scores. Completed questionnaires were received for 96%, 98%, and 93% of students, teachers, and parents, respectively. IQ scores were recorded for 98% of participants. In Spring 2004, we recorded academic performance variables from school records. Simultaneously, we administered the delay of gratification task for the second time and a brief survey of study and lifestyle habits, which was returned by 98% of participants.

Measures

We used the same measures as in Study 1, with four exceptions: We changed the parent and teacher questionnaires and added both a behavioral delay of gratification task and a group-administered IQ test. Internal consistency statistics are provided in Table 8.

Teacher and parent questionnaires. Because several teachers and parents in Study 1 had complained that the format of the Self-Control Rating Scale questionnaire confused them, we omitted it from Study 2. Instead, we asked teachers and parents to complete a version of the Brief Self-Control Scale written in the third person, with the student as target (e.g., "This student/child has a hard time breaking bad habits").

Delay of gratification task. Recognizing that one's stated intention to wait for future rewards may differ from what one would do given an actual choice between immediate and delayed gratification, we designed the delay choice task. As a gift of appreciation for completing self-report questionnaires, we gave each participant an envelope that held a $1 bill. We then asked participants either to take the dollar at that moment or to return it to us in exchange for $2 dollars exactly 1 week later, coding the choice to take the dollar immediately as 0 and the choice to wait as 1. Seven-month test–retest stability ($r = .41$) was good, considering the length of the retest interval.

IQ. As a measure of intelligence, we used the Otis–Lennon School Ability Test—Seventh Edition (Harcourt Brace Educational Measurement, 1997) Level G. This 40-min group-administered, paper-and-pencil test measures verbal, quantitative, and figural reasoning skills. The school ability index for this test is a standard score normalized according to the student's age in months, with a mean of 100 and a standard deviation of 16. Normal curve equivalent scores were derived from percentile ranks for use in all statistical analyses.

Results

Gender Differences in Report Card Grades, Achievement Test Scores, and Measured IQ

As shown in Table 8, girls earned significantly higher final grades in Algebra II, English, and social studies than did boys. Girls also earned higher final grades in Algebra I, though this difference failed to reach statistical significance. The effect sizes for the female advantage in course grades ranged from .25 for Algebra I to .80 for Algebra II and was .54 for overall GPA. Girls also outperformed boys on the standardized achievement test, but, in contrast, the difference was small ($d = .15$) and did not reach statistical significance. Surprisingly, boys earned significantly higher IQ scores than did girls, $t(158) = 3.03$, $p = .003$, $d = .48$.

Achievement Test and IQ Scores Each Underpredict Female GPA

As summarized in Table 9, report card grades of female students were underpredicted in all courses by both achievement test and IQ scores. Underprediction by IQ scores ranged from −0.13 for social studies to −0.27 for Algebra II and, in every course, was larger than underprediction by achievement test scores, which ranged

Table 7
Summary of Hierarchical Multiple Regression Predicting Overall GPA From Gender, Achievement Test Score, and Self-Discipline in Study 1

Variable	*B*	*SE B*	β	part *r*
Step 1				
Gender	0.31	0.09	.22	.22***
Achievement test	0.04	0.00	.63	.62***
Step 2				
Gender	0.15	0.08	.11	.10
Achievement test	0.04	0.00	.54	.52***
Composite self-discipline	0.32	0.06	.36	.32***

Note. Step 1: $R^2 = .48$, $p < .001$; Step 2: $\Delta R^2 = .10$, $p < .001$.
*** $p < .001$.

Table 8
Summary Statistics for Academic Performance and IQ Variables in Study 2

Measure	Girls		Boys		
	M	*SD*	*M*	*SD*	*d*
Algebra II final grade ($n = 31$)	3.53	0.82	2.98	0.83	.80*
Algebra I final grade ($n = 133$)	3.25	0.78	2.98	0.83	.25
English final grade	3.52	0.65	3.04	0.68	.69***
Social studies final grade	3.64	0.49	3.41	0.47	.53**
Overall GPA	3.47	0.63	3.13	0.57	.54***
TerraNova Second Edition Achievement Test normal curve equivalent score	77.89	9.72	76.37	10.81	.15
Otis–Lennon School Ability Test—Seventh Edition	106.94	9.95	111.21	9.43	.48**

Note. $n = 164$, unless otherwise indicated. Effect size and statistical significance of independent-sample *t* tests were conducted with normal curve equivalent, not school ability index, scores.
* $p < .05$. ** $p < .01$. *** $p < .001$.

from −0.04 for Algebra I to −0.22 for Algebra II. Underprediction of overall GPA was −0.20 for IQ and −0.13 for achievement test scores.

Girls Are More Self-Disciplined Than Boys

Girls were more self-disciplined than boys on every measure, and, for two of the six measures, this difference was statistically significant. As shown in Table 10, gender differences were greatest on teacher ratings of self-discipline ($d = .61$) and smallest on the dichotomous delay choice task ($d = .08$). As shown in Table 11, correlations among self-discipline measures were positive and ranged from .06 to .73, with an average of .32.

Using the same procedures as in Study 1, we created composite self-discipline scores. The internal reliability of this linear combination was .90. Comparing composite scores of self-discipline, we found that girls were more self-disciplined than boys, $t(162) = 2.64$, $p = .01$, $d = .41$. Affirming our predictions, composite self-discipline correlated robustly with overall GPA ($r = .67$, $p < .001$), moderately with achievement test scores ($r = .43$, $p < .001$), and only weakly with IQ ($r = .13$, $p = .10$; see Table 12).

Consistent with the observed gender differences in self-discipline, over the course of the school year, girls reported devoting an average of 1 hr per day on homework, almost twice the time that boys spent on homework, $t(157) = 4.37$, $p < .001$, $d = .71$. Girls also began their homework about 20 min earlier in the day, $t(153) = 1.96$, $p = .05$, $d = .31$. Boys watched an average of 8 min more of TV per day than did girls ($d = .18$), though this difference failed to reach significance.

Table 9
Over- and Underprediction of Final Course Grades and GPA From Achievement Test and IQ Scores in Study 2

Measure	Predicted minus actual GPA			
	IQ scores as predictor		Achievement test scores as predictor	
	Girls	Boys	Girls	Boys
Algebra II grade	−0.27	0.26	−0.22	0.28
Algebra I grade	−0.21	0.23	−0.04	0.12
English grade	−0.27	0.28	−0.19	0.23
Social studies grade	−0.13	0.17	−0.08	0.10
Overall GPA	−0.20	0.23	−0.13	0.16

Note. As described in Study 1, grades were converted from a 0–100 scale to a 4.0 scale for convenient comparison with prior studies on underprediction.

Self-Discipline as a Mediator for Gender and Grades

Following the same mediation procedures as in Study 1, we found composite self-discipline partially mediated the relationship between gender and final GPA. Gender predicted self-discipline ($\beta = .20$, $p = .01$). Self-discipline predicted overall GPA when controlling for gender ($\beta = .64$, part $r = .63$, $p < .001$). Finally, in a hierarchical multiple regression predicting overall GPA, gender was entered in Step 1 and found to be a significant predictor ($\beta = .26$, $p = .001$), but when self-discipline was added in Step 2, the β for gender fell 54% ($\beta = .12$, part $r = .12$, $p = .05$). A test for the significance of the decrement in the gender regression coefficient yielded $t(160) = 2.51$, $p = .01$. Thus, controlling for self-discipline, we found that the association between gender and grades was greatly diminished but remained statistically significant, indicating partial mediation (see Table 13).

As in Study 1, a female advantage in self-discipline partially explained underprediction of GPA by achievement test scores. Gender predicted self-discipline when controlling for achievement test scores ($\beta = .19$, part $r = .19$, $p = .01$), and self-discipline predicted overall GPA when controlling for achievement test scores and gender ($\beta = .46$, part $r = .41$, $p < .001$). Finally, in a hierarchical multiple regression predicting overall GPA, gender was a significant predictor when controlling for achievement test scores ($\beta = .20$, part $r = .20$, $p = .001$) in Step 1; when self-discipline was added in Step 2, the β for gender fell 40% but remained significant ($\beta = .12$, part $r = .11$, $p = .03$). A test for the significance of the decrement in the gender regression coefficient yielded $t(155) = 2.40$, $p = .02$ (see Table 14).

Similarly, gender differences in self-discipline partially explained underprediction of GPA by IQ. Gender predicted self-discipline when controlling for IQ ($\beta = .26$, part $r = .25$, $p = .001$). Self-discipline predicted overall GPA when controlling for self-discipline and IQ ($\beta = .60$, part $r = .57$, $p < .001$). Finally,

Table 10
Summary Statistics for Self-Discipline Measures in Study 2

Measure	Internal consistency	Girls		Boys		
		M	*SD*	*M*	*SD*	*d*
Eysenck I_6 Junior Questionnaire Impulsivity subscale[a]	.83	10.56	4.86	11.63	5.19	.21
Brief Self-Control Scale						
Self-report	.86	3.40	0.65	3.10	0.79	.42**
Teacher report	.97	4.38	0.81	3.79	1.09	.61***
Parent report	.91	3.98	0.74	3.83	0.75	.19
Kirby Delay-Discounting Rate Monetary Choice Questionnaire[a]	.98	0.02	0.03	0.03	0.04	.14
Delay choice task		84%		80%		.08
Composite self-discipline score	.90	0.13	0.64	−0.14	0.67	.41***

Note. For the dichotomous delay choice task, effect size was first calculated as Φ, according to Cohen (1988), and then converted to *d*.
[a] Higher scores for these measures indicate lower self-discipline.
** $p < .01$. *** $p < .001$.

in a hierarchical multiple regression predicting overall GPA, gender was a significant predictor when controlling for IQ scores ($\beta = .35$, part $r = .34$, $p < .001$) in Step 1; when self-discipline was added in Step 2, the β for gender fell 46% but remained significant ($\beta = .19$, part $r = .18$, $p = .001$). A test for the significance of the decrement in the gender regression coefficient yielded $t(155) = 3.10$, $p = .002$ (see Table 15).

Discussion

The findings in Study 2 supported the conclusions made in Study 1. There were, in addition, two unexpected results. First, we had predicted that girls and boys in Study 2 would do equally well on a measure of aptitude, yet boys earned substantially higher IQ scores than their female counterparts. As a consequence, underprediction of female GPAs by IQ scores was larger than underprediction by achievement test scores. Second, on the delay choice task, girls were only marginally more likely than boys to choose a $2 reward 1 week later over a $1 reward immediately.

General Discussion

In the present investigation, we suggested that girls outdo boys on report card grades partly because they are more self-disciplined. We further proposed that superior self-discipline helps girls less on achievement tests and minimally on tests of intellectual aptitude. Thus, the enigma of gender differences in the underprediction of grades by standardized tests might be explained, at least in part, by gender differences in trait self-discipline.

In both Study 1 and Study 2, girls ended the school year with GPAs that were more than half a standard deviation above those of their male classmates. Notably, girls outperformed boys in every course subject, including both basic and advanced math. In contrast, gender differences favoring girls on a standardized achievement test were more modest and not statistically significant. And, contrary to our expectation that girls and boys would do equally well on an IQ test, the mean IQ score for girls was about half a standard deviation lower than that for boys. So, although these three measures of academic performance were correlated (*r*s = .32–.62), a significant proportion of each measure's reliable variance seemed attributable to self-discipline-related method variance.

The observed magnitude of underprediction of overall GPA by IQ (−0.20) and for achievement test scores (−0.10 in Study 1 and −0.13 in Study 2) for girls was consistent with studies of underprediction among college students. Mediation analyses suggested gender differences in self-discipline at least partially accounted for the female advantage in course grades. That is, girls earned higher grades at least in part because they were more self-disciplined. Moreover, self-discipline partly corrected for underprediction of course grades by either IQ or achievement test scores.

Composite scores accounting for all self-discipline measures in our battery revealed a significant self-discipline difference favor-

Table 11
Intercorrelations Among Self-Discipline Measures in Study 2

Measure of self-discipline	1	2	3	4	5	6
1. Eysenck I_6 Junior Questionnaire Impulsivity subscale	—	.73***	.33***	.44***	.39***	.07
2. Brief Self-Control Scale self-report		—	.45***	.52***	.29**	.06
3. Brief Self-Control Scale teacher report			—	.36***	.32***	.24**
4. Brief Self-Control Scale parent report				—	.31***	.10
5. Kirby Delay-Discounting Rate Monetary Choice Questionnaire					—	.23**
6. Delayed choice						—

Note. Scores for some measures were recoded such that for all measures, higher scores indicated higher self-discipline.
** $p < .01$. *** $p < .001$.

Table 12
Intercorrelations Among Gender, Composite Self-Discipline, IQ, Achievement Test Scores, and Overall GPA in Study 2

Measure	1	2	3	4	5
1. Gender	—	.20**	−.23**	.07	.26***
2. Composite self-discipline score		—	.13	.43***	.66***
3. IQ			—	.36***	.32***
4. Achievement test score				—	.62***
5. Overall GPA					—

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

ing girls ($d = .71$ in Study 1 and $d = .41$ in Study 2). Consistent with this finding, girls started their homework earlier in the day and spent almost twice as much time completing it. However, although girls were more self-disciplined than boys on all measures, we point out that gender differences were smallest for the only objective measure in our battery ($d = .08$ for the delay choice task in Study 2) and did not always reach significance. To confirm our conclusion that adolescent girls are more self-disciplined than boys, we plan to develop more precise, objective measures of self-discipline (e.g., a delay of gratification task that yields a continuous rather than dichotomous score.)

We do not know why girls in Study 2 did worse on an IQ test, which, according to its publishers, has been designed to exclude items that favor either girls or boys. Moreover, gender differences in general mental ability are very small and seem to be vanishing (Feingold, 1988). Perhaps this is just a local property of this magnet school. Alternatively, it is possible that although the mean IQs of boys and girls do not differ, the variance among boys is greater than that among girls. In that case, given that the adolescents in this study represent the upper tail of the IQ distribution, we would expect a comparatively higher mean IQ for boys than for girls. Still, this explanation begs the question of why there should be a larger spread in the distribution of male IQs than female IQs. Future replication studies are planned to clarify whether the observed pattern of self-discipline, aptitude and achievement test scores, and grades is unique to college-bound adolescents or generally true for adolescents at all levels of achievement.

Whereas prior studies have highlighted factors that favor boys on standardized tests, the current investigation suggests a dimension of greater real-world significance on which girls surpass boys. A recent article in *The Philadelphia Inquirer* entitled "Girls Edge Boys at Head of the Class" reported that local female valedictorians outnumbered male valedictorians nearly 2 to 1 (Snyder, 2003). When interviewed, educators, researchers, and the valedictorians themselves attributed this to noncognitive factors. Said one teacher, "It has to do with input or energy toward studies or lack of it" (p. A1).

Table 13
Summary of Hierarchical Multiple Regression Predicting Overall GPA From Gender and Self-Discipline

Variable	*B*	*SE B*	β	part *r*
Step 1				
Gender	3.25	0.95	.26	.26**
Step 2				
Gender	1.47	0.74	.12	.16*
Composite self-discipline	6.15	0.57	.64	.63***

Note. Step 1: $R^2 = .07$, $p = .001$; Step 2: $\Delta R^2 = .40$, $p < .001$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 14
Summary of Hierarchical Multiple Regression Predicting Overall GPA From Gender, Achievement Test Score, and Self-Discipline

Variable	*B*	*SE B*	β	part *r*
Step 1				
Gender	2.55	0.76	.20	.20**
Achievement test	0.37	0.04	.61	.61***
Step 2				
Gender	1.45	0.66	.12	.11*
Achievement test	0.26	0.03	.42	.38***
Composite self-discipline	4.41	0.55	.46	.41***

Note. Step 1: $R^2 = .43$, $p < .001$; Step 2: $\Delta R^2 = .17$, $p < .001$.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Our findings are consistent with several other studies showing girls to be more self-disciplined than boys. For example, P. M. Cole (1986), Davis (1995), and Sarni (1984) found preschool and school-age girls to be better at regulating emotional expression (e.g., the expression of disappointment) than boys. Humphrey (1982) found that according to both teacher- and self-report questionnaires, fourth-grade girls were more self-controlled than boys. Similarly, Kendall and Wilcox (1979) found that elementary school teachers rated their female students as more self-controlled than their male students. If girls are indeed more self-disciplined than boys, further research is needed to explain how, when, and why this gender difference emerges.

Another implication of our findings is methodological. We concur with Silverman (2003) that inadequate sample sizes and imprecise measures (particularly a problem when a single self-discipline measure is used) may explain the failure of some studies to detect gender differences. For example, Eysenck et al. (1984) found girls aged 7 to 15 years to be slightly more self-disciplined than boys, but this difference failed to achieve significance. Logue, Forzano, and Ackerman (1996) found no gender differences when using a single delay of gratification task with young children. And, in their studies of the factors that influence delay of gratification

Table 15
Summary of Hierarchical Multiple Regression Predicting Overall GPA From Gender, IQ, and Self-Discipline

Variable	*B*	*SE B*	β	part *r*
Step 1				
Gender	4.45	0.92	.35	.34***
IQ	0.19	0.04	.40	.39***
Step 2				
Gender	2.44	0.73	.19	.18**
IQ	0.14	0.03	.30	.29***
Composite self-discipline	5.69	0.54	.60	.57***

Note. Step 1: $R^2 = .22$, $p < .001$; Step 2: $\Delta R^2 = .33$, $p < .001$.
** $p < .01$. *** $p < .001$.

behavior in preschool children, most of which involved relatively small samples, Mischel and colleagues never reported gender as a significant predictor variable (e.g., Mischel, Shoda, & Peake, 1988).

It is not clear whether gender differences in self-discipline persist into adulthood. Some studies of adults have reported no gender differences in self-discipline (e.g., Feingold, 1994; Tangney et al., 2004), but others have found women to be slightly better at delaying gratification than men (e.g., Kirby & Marakovic, 1996; Silverman, 2003). Certainly, the most obvious gender differences in self-discipline among adults relate to the domains in which self-discipline problems emerge. For example, more women than men suffer from binge eating, whereas more men than women are alcohol or drug abusers (Nolen-Hoeksema & Corte, 2004). Baumeister and Vohs (2004) have pointed out that in adulthood, this pattern may arise from gender differences in impulse strength rather than in trait self-discipline.

If female students earn higher grades than male students at every grade level through college, why do more men than women earn medical, law, and other first-professional degrees (Freeman, 2004)? And why do men earn higher salaries than women in equivalent occupations (Freeman, 2004)? In a review of gender differences in professional achievement, Reis (1991) concluded that relative to gifted men, gifted women have lower aspirations, fewer mentors, more pressure to assume family responsibilities, and lower self-esteem. There is also experimental evidence suggesting that women are not as motivated as men in competitive, winner-take-all environments (Gneezy, Niederle, & Rustichini, 2003). Thus, the fact that men enjoy greater professional success than women does not necessarily imply that gender differences in self-discipline diminish or reverse with age. Indeed, it seems at least as plausible that women are more self-disciplined than men, but that beyond college, other psychological, social, and cultural factors swamp the self-discipline edge.

References

American Association of University Women Educational Foundation. (1998). *Gender gaps: Where schools still fail our children.* Washington, DC: Author.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51,* 1173–1182.

Baumeister, R. F., & Vohs, K. D. (2004). Self-regulation. In C. Peterson & M. E. P. Seligman (Eds.), *Character strengths and virtues: A handbook and classification* (pp. 499–516). New York: Oxford University Press.

Block, J. (1996). Some jangly remarks on Baumeister and Heatherton. *Psychological Inquiry, 7,* 28–32.

Clark, M. J., & Grandy, J. (1984). *Sex differences in the academic performance of Scholastic Aptitude Test takers* (College Board Rep. No. 84–8; Educational Testing Science Research Rep. No. 84–43). New York: College Entrance Examination Board.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cole, N. S. (1997). *The ETS gender study: How males and females perform in educational settings.* Princeton, NJ: Educational Testing Service.

Cole, P. M. (1986). Children's spontaneous control of facial expression. *Child Development, 57,* 1309–1321.

Costa, P. T., Jr., & McCrae, R. R. (1989). *NEO PI/FFI manual supplement*. Odessa, FL: Psychological Assessment Resources.

Costa, P. T., Jr., & McCrae, R. R. (1992). *NEO PI-R: Professional manual*. Odessa, FL: Psychological Assessment Resources.

CTB/McGraw-Hill. (2001). *TerraNova, the Second Edition: Technical Quality.* Monterey, CA: Author.

Davis, T. L. (1995). Gender differences in masking negative emotions: Ability or motivation? *Developmental Psychology, 31,* 660–667.

Elliott, R., & Strenta, A. C. (1988). Effects of improving the reliability of the GPA on prediction generally and on comparative predictions for gender and race particularly. *Journal of Educational Measurement, 25,* 333–347.

Evenden, J. L. (1999). Varieties of impulsivity. *Psychopharmacology, 146,* 348–361.

Eysenck, S. B., Easting, G., & Pearson, P. (1984). Age norms for impulsiveness, venturesomeness and empathy in children. *Personality & Individual Differences, 5,* 315–321.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43,* 95–103.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116,* 429–456.

Freeman, C. E. (2004). *Trends in educational equity of girls & women: 2004*. Washington, DC: U.S. Government Printing Office.

Gerbing, D. W., Ahadi, S. A., & Patton, J. H. (1987). Toward a conceptualization of impulsivity: Components across the behavioral and self-report domains. *Multivariate Behavioral Research, 22,* 357–379.

Gneezy, U., Niederle, M., & Rustichini, A. (2003). Performance in competitive environments: Gender differences. *Quarterly Journal of Economics, 118,* 1049–1073.

Gordon, L. V. (1978). *Manual for the Gordon Personal Profile*. New York: Psychological Corporation.

Gwin, R. D. (1997). *Relation among neuropsychology and personality measures of impulsivity.* Unpublished doctoral dissertation, Pacific Graduate School of Psychology.

Harcourt Brace Educational Measurement. (1997). *Otis–Lennon School Ability Test—Seventh edition technical manual.* San Antonio, TX: Author.

Hogan, R., & Weiss, D. S. (1974). Personality correlates of superior academic achievement. *Journal of Counseling Psychology, 21,* 144–149.

Humphrey, L. L. (1982). Children's and teachers' perspectives on children's self-control: The development of two rating scales. *Journal of Consulting and Clinical Psychology, 50,* 624–633.

Hunter, J. E., Schmidt, F. L., & Rauschenberger, J. (1984). Methodological, statistical, and ethical issues in the study of bias in psychological tests. In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing* (pp. 41–99). New York: Plenum Press.

Jackson, D. N. (1974). *Personality Research Form manual*. Port Huron, MI: Research Psychologists Press.

Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.

Kelley, T. L. (1927). Interpretation of educational measurements. Yonkers, NY: World Book.

Kendall, P. C., & Wilcox, L. E. (1979). Self-control in children: Development of a rating scale. *Journal of Consulting and Clinical Psychology, 47,* 1020–1029.

Kimball, M. M. (1989). A new perspective on women's math achievement. *Psychological Bulletin, 105,* 198–214.

Kirby, K. N., & Marakovic, N. N. (1996). Delay-discounting probabilistic rewards: Rates decrease as amounts increase. *Psychonomic Bulletin & Review, 3,* 100–104.

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General, 128,* 78–87.

Leonard, D. K., & Jiang, J. (1999). Gender bias and the college predictions of the SATs: A cry of despair. *Research in Higher Education, 40,* 375–407.

Linn, R. L. (1973). Fair test use in selection. *Review of Educational Research, 43,* 139–161.

Loewen, J. W., Rosser, P., & Katzman, J. (1988). *Gender bias in SAT items.* New Orleans, LO: American Educational Research Association.

Logue, A., Forzano, L., & Ackerman, K. (1996). Self-control in children: Age, preference for reinforcer amount and delay, and language ability. *Learning and Motivation, 27,* 260–277.

Lucas, R. E., & Baird, B. M. (2006). Global self assessment. In M. Eid & E. Diener (Eds.), *Handbook of psychological measurement: A multimethod perspective* (pp. 29–42). Washington, DC: American Psychological Association.

Mau, W., & Lynn, R. (2001). Gender differences on the Scholastic Aptitude Test, the American College Test and college grades. *Educational Psychology, 21,* 133–136.

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.

Nolen-Hoeksema, S., & Corte, C. (2004). Gender and self-regulation. In R. F. Baumeister & K. D. Vohs (Eds.), *Handbook of self-regulation: Research, theory, and applications* (pp. 411–421). New York: Guilford Press.

Nunnally, J. (1978). *Psychometric theory.* New York: McGraw-Hill.

Parker, J. D. A., & Bagby, R. M. (1997). The measurement of impulsivity. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: A foundation for the understanding of psychopathy and violence* (pp. 142–157). New York: Guilford Press.

Perkins, R., Kleiner, B., Roey, S., & Brown, J. (2004). *The High School Transcript Study: A decade of change in curricula and achievement, 1990–2000* (Report No. NCES 2004455). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance and internal distress. *Journal of Educational Psychology, 94,* 396–404.

Reis, S. M. (1991). The need for clarification in research designed to examine gender differences in achievement and accomplishment. *Roeper Review, 13,* 193–198.

Sarni, C. (1984). An observational study of children's attempts to monitor their expressive behavior. *Child Development, 55,* 1504–1513.

Silverman, I. W. (2003). Gender differences in delay of gratification: A meta-analysis. *Sex Roles, 49,* 451–463.

Snyder, S. (2003, July 14). Girls edge boys at head of the class. *Philadelphia Inquirer,* p. A1.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology, 35,* 4–28.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identify and performance. *American Psychologist, 52,* 613–629.

Stricker, L. J., Rock, D. A., & Burton, N. W. (1993). Sex differences in predictions of college grades from Scholastic Aptitude Test scores. *Journal of Educational Psychology, 85,* 710–718.

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,* 271–322.

Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurements.* New York: Teachers College, Columbia University.

Walsh, R. P. (1967). Sex, age, and temptation. *Psychological Reports, 21,* 625–629.

Willingham, W., & Cole, N. (1997). *Gender and fair assessment*. Mahwah, NJ: Erlbaum.

Wolfe, R. N., & Johnson, S. D. (1995). Personality as a predictor of college performance. *Educational and Psychological Measurement, 55,* 177–185.

Young, J. W. (1991). Gender bias in predicting college academic performance: A new approach using item response theory. *Journal of Educational Measurement, 28,* 37–47.

Zaparniuk, J., & Taylor, S. (1997). Impulsivity in children and adolescents. In C. D. Webster & M. A. Jackson (Eds.), *Impulsivity: Theory, assessment, and treatment* (pp. 158–179). New York: Guilford Press.

Received January 11, 2005
Revision received August 10, 2005
Accepted August 10, 2005 ■