Published in final edited form as:

Pers Individ Dif. 2011 April 1; 50(4): 470–474. doi:10.1016/j.paid.2010.11.009.

SAT predicts GPA better for high ability subjects: Implications for Spearman's Law of Diminishing Returns

Thomas Coyle^{a,*}, Anissa Snyder^a, David Pillow^a, and Peter Kochunov^b

^aUniversity of Texas at San Antonio, Department of Psychology, San Antonio, USA

^bUniversity of Texas Health Science Center at San Antonio, Research Imaging Institute, San Antonio, USA

Abstract

This research examined the predictive validity of the SAT (formerly, the Scholastic Aptitude Test) for high and low ability groups. SAT scores and college GPAs were obtained from the 1997 National Longitudinal Survey of Youth. Subjects were classified as high or low ability by *g* factor scores from the Armed Services Vocational Aptitude Battery. SAT correlations with GPA were higher for high than low ability subjects. SAT *g* loadings (i.e., SAT correlations with *g*) were equivalent for both groups. This is the first study to show that the predictive validity of the SAT varies for ability groups that differ in *g*. The results contradict a presumption, based on Spearman's Law of Diminishing Returns, that a test's predictive validity should be lower for high ability subjects. Further research is needed to identify factors that contribute to the predictive validity of the SAT for groups that differ in *g*.

Keywords

SAT; *g* factor; GPA; Spearman's Law of Diminishing Returns; National Longitudinal Survey of Youth

1. Introduction

This research examined the predictive validity of the SAT (formerly, the Scholastic Aptitude Test) for high and low ability subjects. The SAT is a standardized test for college admissions in the United States. The SAT correlates moderately with college grade-point average (GPA) ($r \approx .35$, Bridgeman, McCamley-Jenkins, & Ervin, 2000, Table 1, p. 4), and strongly with *g* ($r = .82$, Frey & Detterman, 2004; see also, Coyle & Pillow, 2008). *g* measures variance common to cognitive tests. A test's *g* loading (i.e., correlation with *g*) is directly related to its predictive validity at work and school (e.g., Jensen, 1998, pp. 274–294).

The theoretical framework of this research concerns the mental ability hypothesis of Spearman's Law of Diminishing Returns (SLODR; Jensen, 1998, pp. 585–588). SLODR predicts that the mean correlation among cognitive tests, and the *g* loading of such tests, declines as ability level increases. SLODR is based on the theory that mental abilities become more independent, and therefore less *g* loaded (and correlated), as ability increases. SLODR has received substantial, though not unequivocal, support. The *g* loading of

cognitive tests, and the correlation among tests, is generally lower for high ability subjects, though effect sizes (based on average correlations) are small across two standard deviations of ability (effect size $\approx .10$; for a review of SLODR research, see te Nijenhuis & Hartmann, 2006, p. 438).

SLODR has implications for the predictive validity of a particular cognitive test. If the mean correlation among cognitive tests is lower for high ability subjects, then the correlation of a particular test with another test should generally be lower for high ability subjects. Given that the predictive validity of a test is a test's correlation with a criterion (e.g., another test), it follows that a test's predictive validity should generally be lower for high ability subjects. This prediction is central to the present research on the SAT.

Bridgeman et al. (2000) examined the predictive validity of the SAT for high and low ability groups from 23 colleges. Subjects were divided into ability groups based on the mean SAT score of their college. (SAT scores were the sum of math and verbal scores.) Some subjects (high ability) attended colleges with mean SAT scores greater than 1250; others (low ability) attended colleges with mean SAT scores less than 1050. The results contradicted the prediction, based on SLODR, that predictive validity would be lower for high ability subjects. SAT–GPA correlations (adjusted for range restriction) from three SAT scales were modestly *higher* for high ($M_r = .61$) than low ability subjects ($M_r = .50$) (Bridgeman et al., 2000, Table 3, p. 5). Moreover, SAT–GPA correlations from the 23 colleges *increased* with the mean SAT scores of the colleges ($r = .42$, Bridgeman et al., 2000, Table 5, p. 7). Other studies with large samples confirm that the SAT's predictive validity is higher for high ability subjects (e.g., Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2008).

Bridgeman et al.'s (2000) SAT findings are inconsistent with research on race differences in cognitive ability. Much of this research concerns black-white differences in the population (e.g., Jensen, 1998, pp. 350–417). Given that whites are the higher ability group (as indicated by their test scores), and that predictive validity of cognitive tests is higher for higher ability groups (as in Bridgeman et al.), it follows that predictive validity should be higher for whites. But this is not the case. Black-white differences in predictive validity are typically negligible (Jensen, 1998, pp. 360–362), as are differences in predictive validity for other races that differ in ability (e.g., Rushton, Skuy, & Bons, 2004; te Nijenhuis & van der Flier, 1999; Zeidner, 1987).

The present study examined the predictive validity of the SAT using SLODR as a predictive framework. Ability groups were formed using *g* scores from the computer-adapted Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB consists of 12 diverse cognitive tests (e.g., coding speed, word knowledge, assembling objects). It has a large *g* factor that is comparable (in *g* variance) for high and low ability groups (Hartmann & Reuter, 2006, Table 2, p. 55). *g* scores from the ASVAB correlate strongly with cognitive ability tests (e.g., Frey & Detterman, 2004, Table 2, p. 375).

The study had two aims. The first was to examine the *g* loading of the SAT for high and low ability subjects. *g* loadings were measured as correlations between SAT scores and *g* scores. If the mean *g* loading of cognitive tests is lower for high ability subjects (as SLODR predicts), then the *g* loading of any particular test (including the SAT) should also generally be lower for high ability subjects. This prediction was not tested by Bridgeman et al. (2000), who did not report *g* loadings for the SAT.

The second aim was to examine the predictive validity of the SAT (for GPA) for high and low ability subjects. The predictive validity of a cognitive test has been shown to vary directly with a test's *g* loading: tests with lower *g* loadings generally have lower predictive validity (e.g., Jensen, 1998, Table B.1, p. 590). This has implications for the predictive

validity of the SAT. If the g loading of the SAT is lower for high ability subjects (as SLODR predicts), then the predictive validity of the SAT should also be lower for high ability subjects. This prediction contradicts Bridgeman et al. (2000), who found that the predictive validity of the SAT was *higher* for high ability subjects.

2. Method

2.1. Subjects

Subjects were drawn from the 1997 National Longitudinal Survey of Youth (NLSY97). The NLSY97 is administered by the United States Department of Labor (Hering & McClain, 2003, pp. 1–14). The full NLSY97 consists of 8984 people (4599 males) age 12–16 years in 1996 (M age = 14.00 years, SD = 1.40 years). A total of 737 cases contained all relevant variables (described below), and were extracted from the NLSY97 for analysis.

2.2. Variables

SAT—SAT scores were SAT math (SATm), verbal (SATv), and total scores (SATt). SATt was the sum of SATm and SATv. All SAT scores were from the recentered SAT I, which is strongly related to the current SAT (rs = .95–.97, Kober & Schmidt, 2005).

GPA—GPAs were obtained from student self-reports, which are typically highly correlated with GPAs from college records (r = .90, Kuncel, Credé, & Thomas, 2005, Table 1, p. 73). GPAs were the average number of grade points earned during the first and second terms at colleges with semester schedules (i.e., two terms per academic year). GPAs deviated from normality (skewness = $-.80$, SE = .09; kurtosis = 1.24, SE = .18). Squaring the GPAs normalized the distributions (skewness = $-.08$, SE = .09; kurtosis = $-.66$, SE = .18). Because deviations from normality can distort error estimates, transformed (squared) GPAs were used for analysis.

ASVAB— g scores were estimated using the ASVAB (Hering & McClain, 2003, pp. 82–84). The ASVAB consisted of 12 subtests: Arithmetic Reasoning, Assembling Objects, Auto Information, Coding Speed, Electronics Information, General Science, Mathematics Knowledge, Mechanical Comprehension, Numerical Operations, Paragraph Comprehension, Shop Information, and Word Knowledge. Two subtests (Coding Speed and Numerical Operations) were speed tests; the others were power tests. The power tests were administered using an adaptive procedure, which matched item difficulty to ability level. The speed tests were administered in a non-adaptive format, which presented items in the same order to all subjects. Speed test scores were based on proportion of correct responses, adjusting for guessing and presentation time. Power test scores were based on item response theory, with higher scores indicating better performance.

2.3. Estimating g and forming ability groups

Jensen (2003) has argued that comparisons among groups that differ in g should (a) estimate g from large and diverse batteries of cognitive tests, and (b) divide subjects into two, rather than several, ability groups. Estimating g from large and diverse (as opposed to truncated) batteries of cognitive tests increases the likelihood of obtaining a good g (i.e., variance common to the theoretical population of cognitive tests). Dividing subjects into two rather than several ability groups minimizes the problem of range restriction (of test scores) in ability groups. It also minimizes artificial negative correlations among tests, which can distort estimates of g .

The approach used here followed Jensen's (2003) recommendations. g was estimated using principal factor analysis (PFA) of all ASVAB subtests, and was defined as the first principal

factor (PF1). g scores were estimated as the sum of standardized ASVAB scores, weighted by g factor score coefficients (Tabachnick & Fidell, 2007, pp. 622–625). Subjects were classified as high or low ability based on the mean g score of the entire sample ($M = 0$). Subjects with g scores above or below the mean were classified as high ($n = 373$) or low ability ($n = 364$), respectively. SAT g loadings were correlations between SAT scores and g scores from the ASVAB.

2.4. Statistics

Independent sample t -tests were used to examine group differences in means. Fisher's r to z difference test was used to examine group differences in correlations. Cohen's d was used as a measure of between-groups effect size. d s around .20, .50, and .80, were labeled small, medium, and large, respectively. Significant effects are reported at $p < .05$ (two-tailed tests).

3. Results

3.1. Preliminary analyses

Table 1 reports the means and standard deviations of all variables, by ability group. SAT scores, g scores, and GPAs were significantly higher for high than low ability subjects. The between-group effect size was medium for GPA ($d = .51$) and very large for SAT and g scores ($M_d = 1.70$). The distributional characteristics of variables in each group signaled no problems with deviations from normality (skewness ± 1 ; kurtosis ± 1).

Because group differences in standard deviations (SD s) can produce spurious differences in correlations, tests of group differences in SD s were performed. Group differences in SD s were tested using F ratios. The F ratios were computed using the formula, (largest SD /smallest SD)² (Hartley, 1950). The F ratios were small and nonsignificant for all variables (F s = 1.11, 1.12, 1.10, 1.14, and 1.03 for SATt, SATm, SATv, g , and GPA, respectively).

3.2. Does g represent the same construct for high and low ability subjects?

Comparing groups that differ in g is appropriate only if g represents the same construct across groups. Jensen (2003) argued that if g represents the same construct across groups, then (a) the g loadings of cognitive tests should be similar across groups, and (b) the g factors should account for similar variance across groups.

The g loadings of ASVAB subtests were estimated for each ability group using PFA (Table 2). The mean g loading was very similar for high ($M = .49$) and low ability subjects ($M = .51$). The pattern of g loadings across groups was examined using the congruence coefficient (CC). The CC was computed using the formula, $\sum XY / \sqrt{\sum X^2 \sum Y^2}$, where X and Y are subtest g loadings for each ability group. CCs range from -1 to 1 ; CCs higher than .95 suggest that the g factor is nearly identical between groups. The CC approached unity (CC = .98).

g factor variance among ASVAB subtests was estimated for each ability group using PFA. g factor variance was measured using g factor eigenvalues. The eigenvalues were very similar for high (3.12) and low ability groups (3.35). Following Jensen (2003), the F ratio of g factor eigenvalues (low ability/high ability) was used to test for group differences in g variance. The F ratio (1.07) was not significant.

3.3. Are SAT g loadings lower for high ability subjects, as SLODR implies?

SAT g loadings were measured as correlations between SAT scores and g scores from the ASVAB (Table 3, top). Contra SLODR, SAT g loadings were equivalent for high ($M_r = .50$) and low ability subjects ($M_r = .51$). GPA g loadings (i.e., GPA- g correlations) were also equivalent for both ability groups (r s = .13, Table 3, top). The equivalent g loadings indicate

that group differences in the predictive validity of the SAT cannot be attributed to group differences in SAT (or GPA) g loadings.

3.4. Are SAT predictive validities lower for high ability subjects, as SLODR implies?

SAT predictive validities (for GPA) were computed separately for high and low ability subjects (Table 3, bottom). Contra SLODR, SAT predictive validities were significantly higher for high ($M_r = .32$) than low ability subjects ($M_r = .18$) ($ps = .019, .047$, and $.041$ for SATt, SATm, and SATv, respectively).

3.5. Can the results be attributed to differential reliability between groups?

Differential reliabilities between groups (on the variables correlated) can create spurious differences in correlations. Differential reliabilities can be inferred by the pattern of correlations among variables. If a variable is differentially *unreliable* for one group, then correlations with that variable should be uniformly (and significantly) lower for that group. No such pattern was found for SAT and GPA correlations with g . Although SAT relations with GPA were higher for high ability subjects (Table 3, bottom), SAT relations with g were equivalent for both ability groups (Table 3, top). Similarly, GPA relations with g were identical for both ability groups (Table 3, top).

3.6. Supplemental analyses

The preceding analyses examined the predictive validity of the SAT for ability groups based on a dichotomy of g scores, as is typical in SLODR research (e.g., Hartmann & Reuter, 2006). This dichotomy may have distorted SAT–GPA relations over the full range of g (e.g., DeCoster, Iselin, & Gallucci, 2009).

Supplemental regressions examined the predictive validity of the SAT over the full range of g . The analyses predicted GPA using SAT, g , and the interaction between SAT and g . (SAT and g were centered prior to creating the interaction term.) The interaction tested whether SAT–GPA relations varied over the full range of g . Three analyses were performed, one for each SAT score (SATt, SATm, SATv).

Table 4 reports the results of the analyses. SAT main effects were significant for all SAT scores. g main effects were significant only for SATm. The main effects were qualified by significant (for SATv) and marginally significant (for SATt) interactions. Post-hoc analyses of these interactions examined SAT–GPA relations at low (1 SD below average), average, and high (1 SD above average) levels of g (e.g., Cohen, Cohen, West, & Aiken, 2003). Consistent with the prior analyses, SAT–GPA relations increased with levels of g (low, average, high) for SATt ($\beta = .30, .35, .41$) and SATv ($\beta = .24, .32, .40$). Parallel analysis of SATm indicated that SAT–GPA relations also increased, albeit not significantly, with g ($\beta = .21, .23, .26$). SAT–GPA relations at different levels of g did not include a significant quadratic component for any SAT score.

4. Discussion

This is the first research to show that the predictive validity of the SAT varies for ability groups that differ in g . The SAT predicted college GPA better for high than low ability subjects. This pattern was replicated for all SAT scores (SATt, SATm, SATv). The pattern could not be attributed to SAT or GPA g loadings (i.e., correlations with g), which were equivalent for both ability groups. The pattern also could not be attributed to group differences in g variance (based on the ASVAB). g variance was comparable for the ability groups, replicating prior research (e.g., Hartmann & Reuter, 2006) and contradicting SLODR.

The findings confirm those of Bridgeman et al. (2000), who also found that SAT scores predicted GPA better for high ability subjects. The findings also extend those of Bridgeman et al., who divided subjects into ability groups based on the mean SAT scores of their college. Subjects in the current study were divided into groups based on *g* scores, which were needed to test SLODR (a hypothesis of *g*). The findings reported here and those of Bridgeman et al. were based on large and representative samples, which decrease the likelihood of sampling error.

The findings contradict SLODR and its implications for predictive validity. Cognitive tests with lower predictive validities typically have lower *g* loadings (e.g., Jensen, 1998, p. 590). If the *g* loading of a test is generally lower for high ability subjects (as SLODR predicts), then the predictive validity of a test should also be lower for high ability subjects. The results reported here did not support this prediction. The predictive validity of the SAT was *higher* for high ability subjects, and the *g* loading of the SAT was *equivalent* for both ability groups.

It could be argued that the absence of SLODR effects for the SAT was attributable to the absence of SLODR effects for *g* (as *g* did not differentiate the ability groups). However, the two effects did not have to coincide, as they were based on different aspects of *g*. SLODR effects for *g* were based on *g* variance from the 12 diverse ASVAB subtests, which were likely to produce a good *g* (cf. Johnson, te Nijenhuis, & Bouchard, 2008). SLODR effects for the SAT were based on the *g* loading of the SAT. The SAT (like any test) is an imperfect indicator of *g*, as indicated by its *g* loading (Table 3, top). Because the *g* loading of an imperfect indicator of *g* can decline (with ability) without a general decline in *g* variance (for a supporting trend, see Hartmann & Reuter, 2006, p. 56), SLODR effects for the SAT could occur without SLODR effects for *g*. In the present study, SLODR effects were not found for the *g* loading of the SAT or for *g*, as neither variable differentiated the ability groups. The results suggest that the predictive validity of the SAT (which did differentiate the groups) was attributable to non-*g* factors.

The predictive validity of the SAT was examined further in supplemental analyses, which treated ability as a continuous variable (Table 4). The predictive validity of the SAT still increased with ability for all SAT scores, but reliably for SATv only. These results partially confirm the analyses treating ability as a dichotomy, which found that predictive validity increased with ability for all SAT scores. (Differences between the analyses may be attributed to small group differences in predictive validity, which make effects sensitive to different analytical techniques [DeCoster et al., 2009].) The findings suggest that the differential validity of the SAT is attributed more to SATv than SATm when the full range of ability is considered. Perhaps this is because SATv is less *g* loaded than SATm for high ability subjects (Table 3, top) and is more likely to tap less *g*-loaded traits that contribute to GPA (as discussed below).

The analyses treating *g* as a continuous variable complement, but differ from, the analyses treating *g* as a dichotomous variable. The latter analyses established *g* factor invariance between ability groups, so that group differences in SAT–GPA correlations could not be attributed to *g*. The analyses treating *g* as a continuous variable could not establish *g* factor invariance at all levels of *g*. This is because *g* factor variance cannot be estimated at each unique value of *g*, but only within a range of *g*. It is an open question whether the SAT results reported here would be replicated if *g* factor invariance does not hold across all possible ranges of *g*.

The findings reported here have implications for research on race differences and the predictive validity of cognitive tests. Much of this research has focused on black-white

differences (e.g., Jensen, 1998, pp. 360–362). Although black-white differences in the predictive validity of cognitive tests are typically negligible, such differences may be found for the SAT. Assuming that the SAT predicts GPA better for high ability subjects, and that whites perform better than blacks on the SAT, black-white differences in predictive validity on the SAT (favoring whites) would be expected. The College Board reports data consistent with this hypothesis. SAT scores are consistently higher for whites than blacks, and SAT–GPA correlations are consistently (though modestly) higher for whites ($M_r \approx .52$) than blacks ($M_r \approx .44$) (Bridgeman, Pollack, & Burton, 2008, Table 3, p. 4; Mattern, Patterson, Shaw, Kobrin, & Barbuti, 2008, Table 2, p. 7). (Mean correlations are computed using all SAT scores and are collapsed across gender.)

The differential validity of the SAT for high and low ability subjects has implications for using the SAT as an admissions test. The SAT is used as an admissions test at selective and less selective colleges. If the SAT predicts GPA better for high ability subjects, then its predictive validity should be higher at selective colleges (whose students have higher SAT scores). This hypothesis can be tested with data from Bridgeman et al. (2000), who reported SAT–GPA correlations, along with mean SAT scores (SATv + SATm), from 23 colleges. If the predictive validity of the SAT increases with ability, then the magnitude of SAT–GPA correlations should increase with the mean SAT scores from the colleges. In fact, SAT–GPA correlations did increase with the mean SAT scores ($r = .42$, Bridgeman et al., 2000, Table 5, p. 7).

The g loadings (i.e., correlations with g) of the SAT and GPA were comparable between ability groups. The comparable g loadings suggest that the differential validity of the SAT might be related to non- g factors. Two such factors are noncognitive factors and specific cognitive abilities. Noncognitive factors include proxies of academic motivation such as study time and effort, which predict college grades (Stinebrickner & Stinebrickner, 2007). They also include personality traits such as openness and conscientiousness, which also predict college grades (O'Connor and Paunonen, 2007). Specific cognitive abilities include narrow math and verbal abilities, which are modestly g loaded but predict school achievement (Park, Lubinski, & Benbow, 2007). If specific cognitive abilities or noncognitive factors differentiate high ability subjects more than low ability subjects (and are related to the SAT and GPA), then such factors may contribute to the differential validity of the SAT for high and low ability groups.

In sum, this is the first study to show that the predictive validity of the SAT varies with cognitive ability and g . SAT–GPA correlations were generally higher for high ability subjects, while SAT g loadings were equivalent for both ability groups. The pattern of SAT–GPA correlations contradicts SLODR, which predicted the opposite pattern. The pattern also confirms Bridgeman et al. (2000), who obtained similar results using different methodology. Further research is needed to identify factors that contribute to the predictive validity of the SAT for high and low ability subjects.

Acknowledgments

This research was supported by the San Antonio Area Foundation; a University of Texas at San Antonio Faculty Development Leave; the University of Texas Health Science Center – General Clinical Research Center (M01-RR-01346); and the National Institute of Biomedical Imaging and Bioengineering (K01-EB006395). We thank Deborah Coyle, Annette Fields, Jason Purcell, Tony Vernon, and three anonymous reviewers for comments.

References

- Bridgeman, B.; McCamley-Jenkins, L.; Ervin, N. Predictions of freshman grade-point average from the revised and recentered SAT I: Reasoning Test. College Board; New York, NY: 2000. (College Board Research Report No. 2000-1)
- Bridgeman, B.; Pollack, J.; Burton, N. Predicting grades in different types of college courses. College Board; New York, NY: 2008. (College Board Research Report No. 2008-1)
- Cohen, J.; Cohen, P.; West, S.G.; Aiken, L.S. Applied multiple regression/correlation analysis for the behavioral sciences. 3rd ed. Erlbaum; Mahwah, NJ: 2003.
- Coyle TR, Pillow DR. SAT and ACT predict college GPA after removing *g*. *Intelligence*. 2008; 36:719–729.
- DeCoster J, Iselin AR, Gallucci M. A conceptual and empirical examination of justifications for dichotomization. *Psychological Methods*. 2009; 14:349–366. [PubMed: 19968397]
- Frey MC, Detterman DK. Scholastic assessment or *g*? the relationship between the scholastic assessment test and general cognitive ability. *Psychological Science*. 2004; 15:373–378. [PubMed: 15147489]
- Hartley HO. The maximum *F*-ratio as a short-cut test for heterogeneity of variance. *Biometrika*. 1950; 37:308–312. [PubMed: 14801057]
- Hering, J.; McClain, A., editors. NLSY97 user's guide: A guide to rounds 1–5 data. Center for Human Resource Research, Ohio State University; Columbus, OH: 2003.
- Hartmann P, Reuter M. Spearman's "law of diminishing returns" tested with two methods. *Intelligence*. 2006; 34:47–62.
- Jensen, AR. The *g* factor: The science of mental ability. Praeger; Westport, CT: 1998.
- Jensen AR. Regularities in Spearman's law of diminishing returns. *Intelligence*. 2003; 31:95–105.
- Johnson W, te Nijenhuis J, Bouchard TJ Jr. Still just 1 *g*: Consistent results from five test batteries. *Intelligence*. 2008; 36:81–95.
- Kobrin, J.L.; Patterson, B.F.; Shaw, E.J.; Mattern, K.D.; Barbuti, S.M. Validity of the SAT for predicting first-year college grade point average. College Board; New York, NY: 2008. (College Board Research Report No. 2008-5)
- Kobrin, J.L.; Schmidt, A.E. The research behind the new SAT. College Board; New York, NY: 2005. (College Board Research Summary RS-11)
- Kuncel NR, Credé M, Thomas LL. The validity of self-reported grade point averages, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*. 2005; 75:63–82.
- Mattern, K.D.; Patterson, B.F.; Shaw, E.J.; Kobrin, J.L.; Barbuti, S.M. Differential validity and prediction of the SAT. College Board; New York, NY: 2008. (College Board Research Report No. 2008-4)
- O'Connor MC, Paunonen SV. Big Five personality predictors of post-secondary academic performance. *Personality and Individual Differences*. 2007; 43:971–990.
- Park G, Lubinski D, Benbow CP. Contrasting intellectual patterns predict creativity in the arts and sciences: Tracking intellectually precocious youth over 25 years. *Psychological Science*. 2007; 18:948–952. [PubMed: 17958707]
- Rushton JP, Skuy M, Bons TA. Construct validity of Raven's advanced progressive matrices for African and non-African engineering students in South Africa. *International Journal of Selection and Assessment*. 2004; 12:220–229.
- Stinebrickner, T.R.; Stinebrickner, R. The causal effect of studying on academic performance. National Bureau of Economic Research; Cambridge, MA: 2007. (NBER Working Paper No. 13341)
- Tabachnick, B.G.; Fidell, L.S. Using multivariate statistics. 5th ed. Allyn and Bacon; Boston, MA: 2007.
- te Nijenhuis J, Hartmann P. Spearman's "law of diminishing returns" in samples of Dutch and immigrant children and adults. *Intelligence*. 2006; 34:437–447.
- te Nijenhuis J, van der Flier H. Bias research in the Netherlands: Review and implications. *European Journal of Psychological Assessment*. 1999; 15:165–175.

Zeidner M. Test of the cultural bias hypothesis: Some Israeli findings. *Journal of Applied Psychology*. 1987; 72:38–48.

Table 1

Means, standard deviations, and between-group statistics for high and low ability groups.

Variable ^a	High ability (<i>n</i> = 373)		Low ability (<i>n</i> = 364)		Between-group statistic	
	M	SD	M	SD	<i>F</i> ^b	<i>d</i>
SATt	1157.77	157.21	921.24	149.45	20.93	1.54
SATm	580.00	91.10	459.31	85.99	18.49	1.36
SATv	577.77	85.94	461.92	82.05	18.71	1.38
<i>g</i>	.75	.58	-.77	.62	34.30	2.53
GPA	3.17	.59	2.87	.58	6.84	.51

^a SATt = SATm + SATv; SATm = SAT math; SATv = SAT verbal; *g* = PFI scores using ASVAB; GPA = college freshman GPA on a 4.0 scale.

^b *F*s are significant at *p* < .001.

Table 2

g loadings of ASVAB subtests for high and low ability groups.

ASVAB subtest	High ability	Low ability
Arithmetic reasoning	.60	.63
Assembling objects	.35	.42
Auto information	.46	.39
Coding speed	.24	.35
Electronics information	.60	.56
General science	.71	.67
Math knowledge	.54	.61
Mechanical comprehension	.64	.60
Numerical operations	.25	.28
Paragraph comprehension	.32	.58
Shop information	.56	.42
Word knowledge	.56	.66

Table 3Correlations of SAT with *g* and GPA for high and low ability groups.

Correlation	High ability	Low ability	<i>z</i>
<i>Correlation with g</i>			
SATt	.54**	.55**	-.19 ^{ns}
SATm	.55**	.50**	.93 ^{ns}
SATv	.41**	.47**	-1.01 ^{ns}
GPA	.13*	.13*	.00 ^{ns}
<i>Correlation with GPA</i>			
SATt	.35**	.19**	2.34*
SATm	.28**	.14**	1.98*
SATv	.34**	.20**	2.05*

Note: ns = not significant. *z* = *z*-score differences (high-low) using Fisher's *r* to *z* transformation. Correlations with *g* = *g* loadings of SAT and GPA.

*
p < .05.

**
p < .01.

Table 4

Supplemental analyses: regressions of GPA on SAT and *g*, using the full range of *g* scores.

Analysis	β	<i>t</i>	<i>F</i>	<i>R</i> ²
<i>SATt</i>			41.98**	.15
SAT	.35	6.89**		
<i>g</i>	.03	.61 ^{ns}		
SAT $\times$ <i>g</i>	.06	1.80 ⁺		
<i>SATm</i>			31.54**	.11
SAT	.23	4.71**		
<i>g</i>	.13	2.55*		
SAT $\times$ <i>g</i>	.03	.94 ^{ns}		
<i>SATv</i>			43.77**	.15
SAT	.32	6.96**		
<i>g</i>	.09	1.86 ⁺		
SAT $\times$ <i>g</i>	.09	2.62**		

Note. ns = not significant.

⁺ $p < .10$.

* $p < .05$.

** $p < .01$.