Research Article

Scholastic Assessment or *g*?

The Relationship Between the Scholastic Assessment Test and General Cognitive Ability

Meredith C. Frey and Douglas K. Detterman

Case Western Reserve University

ABSTRACT—*There is little evidence showing the relationship between the Scholastic Assessment Test (SAT) and* g *(general intelligence). This research established the relationship between SAT and* g*, as well as the appropriateness of the SAT as a measure of* g*, and examined the SAT as a premorbid measure of intelligence. In Study 1, we used the National Longitudinal Survey of Youth 1979. Measures of* g *were extracted from the Armed Services Vocational Aptitude Battery and correlated with SAT scores of 917 participants. The resulting correlation was .82 (.86 corrected for nonlinearity). Study 2 investigated the correlation between revised and recentered SAT scores and scores on the Raven's Advanced Progressive Matrices among 104 undergraduates. The resulting correlation was .483 (.72 corrected for restricted range). These studies indicate that the SAT is mainly a test of* g*. We provide equations for converting SAT scores to estimated IQs; such conversion could be useful for estimating premorbid IQ or conducting individual difference research with college students.*

The relationship between academic achievement and general intelligence, or *g*, is well established in the field of intelligence (Jensen, 1998). One particular test of academic achievement, however, is conspicuously absent from these studies. The SAT (once called the Scholastic Aptitude Test, and now the Scholastic Assessment Test), an exam that generations of high school juniors and seniors have agonized over and that has spawned innumerable preparatory books and classes, has not, surprisingly enough, been frequently analyzed in conjunction with measures of general cognitive ability.

That is not to say that there is a lack of research in general on the SAT. In fact, to the contrary, the SAT's relation to college admissions and achievement has been the focus of much research. A study of 25,000 applicants at nine private colleges yielded unsurprising results: Two thirds of the variance in predicting admissions decisions could be accounted for by academic factors, split equally between high school record and admissions test scores, including SAT scores (Willingham & Breland, 1982). Furthermore, in 1964, the Educational Testing Service (ETS), the group that administers the SAT along with countless other standardized tests for the College Board, began a program of validity research with the SAT, known as the Validity Study Service (VSS). This service, offered free to interested colleges, computes the average multiple correlation between three predictor variables (SAT math score, SAT verbal score, and high school record) and grade point average (GPA) among college freshmen. Between the years 1964 and 1985, the average correlation between SAT score and freshman GPA was in the range from .38 to .46 (Morgan, 1990). Despite this history of analysis, and the knowledge that *g* is one of the most reliable predictors of academic achievement, ETS refuses to cite any SAT-IQ correlations (Seligman, 1994). The College Board president, Gaston Caperton, was recently quoted in an on-line publication of *U.S. News & World Report* as admitting that "in its original form [the SAT] was an IQ test" (Barnes, 2002). However, the people responsible for developing and administering the SAT insist that, once logic items were replaced by reading comprehension in 1946, the SAT ceased to measure intelligence (Barnes, 2002). Rather, the College Board currently asserts that the SAT measures reasoning ability, which they hold as different from intelligence.

The unusual distinction made by ETS and the College Board notwithstanding, some researchers have voiced support for the SAT as a de facto intelligence test. In his 1994 book, *A Question of Intelligence*, Seligman noted, "Henry Chauncey, a former president of ETS has been quoted (in Klitgaard, *Choosing Elites*, 1985, p. 92) as stating that the SAT is essentially an intelligence test" (pp. 206–207). In fact, the author of the original SAT, Carl Campbell Brigham, developed his test from Army intelligence tests to more accurately differentiate among individuals at the upper end of the scale (Lemann, 1999). Although the College Board asserts that the test is substantially different from its original form, Lemann asserted that, other than having its language updated, the SAT has changed little since its original version.

Furthermore, empirical evidence suggests a substantial relationship between the SAT and *g*. In a study of 339 undergraduates, Brodnick and Ree (1995) used covariance structure modeling to examine the relationship between psychometric *g*, socioeconomic variables, and achievement-test scores. They found substantial general-factor loadings on both the math (.698) and the verbal (.804) SAT subtests. Because they used the SAT to develop their measure of *g*, it is not clear if this general factor is the same as that obtained from standard intelligence tests. If the general factors are indeed the same, then the

Address correspondence to D.K. Detterman, Department of Psychology, Case Western Reserve University, Cleveland, OH 44106; e-mail: dkd2@po.cwru.edu.

SAT may have been overlooked as a potentially useful measure of general cognitive functioning.

It would be of considerable benefit for researchers and clinicians to have an equation for estimating an individual's IQ from an objective measure such as the SAT. In cases of suspected cognitive decline due to injury or illness, clinicians currently use a series of demographic variables to predict their clients' premorbid IQ. For many of these individuals, the SAT may be the only objective measure of premorbid functioning. The method of predicting IQ with demographic variables carries with it a standard error of prediction (SE_p) of 11.4 (Karzmark, Heaton, Grant, & Matthews, 1985). It appears that there is ample room for improving the accuracy with which clinicians currently estimate premorbid IQ.

In addition to the obvious clinical application, there is a substantial benefit to researchers' use of a regression equation to estimate IQ. Much research in the field of psychology is carried out on undergraduate volunteers. Although there has been a recent push by the University of California system to discontinue using SAT scores as an admission criterion, many colleges and universities continue to require the SAT for admission (Barnes, 2002). Consequently, there is a vast pool of potential subjects who already have a good measure of their general cognitive functioning in their transcripts. If a researcher could estimate IQ with sufficient accuracy from the scores on record, it would eliminate hours of additional testing in cases in which measures of general intelligence are required. To these ends, we examined the relationship between the SAT and general cognitive ability in two studies. We also developed regression equations to estimate IQ from the SAT.

STUDY 1

Method

Participants

Data were extracted from the National Longitudinal Survey of Youth 1979 (NLSY79) data set. The NLSY79 contained a national probability sample of civilian and military subjects aged 14 to 21, with an overrepresentation of certain minority groups. A total of 12,686 men and women participated in the NLSY79.

Procedure

From the NLSY79 data set, we extracted measures of performance on the Armed Services Vocational Aptitude Battery (ASVAB) and the SAT (known as the Scholastic Aptitude Test at that time), as well as a number of intelligence tests that did not have scores available for as many subjects ("small-*N*" tests).

A principal-axis factor analysis was performed on the 10 subtests of the ASVAB in order to derive a measure of *g*. This analysis included 11,878 of the 11,914 subjects who had taken the ASVAB. The remaining 36 subjects were excluded because they were missing scores for 1 or more subtests. ASVAB scores were chosen in place of scores on some of the more traditional intelligence tests because ASVAB scores were available for nearly all of the NLSY79 participants (11,878 of 12,686). Furthermore, prior analysis of the ASVAB confirmed a hierarchical *g* model in which 64% of the variance in the ASVAB was due to a general factor (Ree & Carretta, 1994; see Roberts et al., 2000, for an alternative model). Results of the factor analysis of the ASVAB are shown in Table 1. They indicate a substantial loading of all subtests of the ASVAB on a first factor, *g*.

TABLE 1

Factor Loadings From the Factor Analysis of the Armed Services Vocational Aptitude Battery

Subtest	Factor 1	Factor 2
General Science	.881	−.116
Arithmetic Reasoning	.858	.045
Word Knowledge	.885	.071
Paragraph Comprehension	.825	.175
Numerical Operations	.724	.426
Coding Speed	.657	.417
Auto and Shop Information	.727	−.425
Mathematics Knowledge	.800	.129
Mechanical Comprehension	.794	−.320
Electronics Information	.829	−.320

ASVAB first-factor scores were transformed to an IQ scale using the following equation: $IQ = (z * 15) + 100$. Finally, the IQ scores derived from the ASVAB were correlated with SAT scores for the 917 respondents who had scores on both measures. Simple correlations between both SAT scores and ASVAB IQ scores and scores on the small-*N* intelligence tests were also analyzed.

Results

Of the 11,878 subjects who had scores on all ASVAB subtests, 917 also took the SAT and were the primary focus in this study. There was a significant correlation between SAT and ASVAB IQ (r = .820, $p < .001$). However, when a scatter plot of the relationship was examined (see Fig. 1a), a nonlinear relationship was evident. To correct for nonlinearity, we added the squared SAT score ($R = .857$, $p < .001$) to the regression. A cubic component of SAT also added significantly to the prediction of IQ ($p = .011$), but the added variance was so small that we decided to not include it in the regression equation. Figure 1b presents a scatter plot of IQ predicted from a regression equation with the squared and cubic components of SAT versus IQ obtained from the first factor of the ASVAB. Entering math and verbal subtest scores into the regression separately, as opposed to using a composite score, did not improve the prediction of IQ.

The simple correlations between SAT and IQ (or whatever measure was reported for the IQ test) for small-*N* intelligence tests ranged from .53 to .82 for participants who had taken the SAT and also had another IQ test score reported in their school records. All simple correlations between tests are summarized in Table 2. The SAT correlated significantly ($p < .05$) with all six of the traditional intelligence tests examined (see the SAT column in Table 2), although these results must be interpreted with a degree of caution as some of the *n*s are quite small. As the correlations in the ASVAB column indicate, the SAT and the other six intelligence tests all correlated significantly with IQ derived from the first-factor score of the ASVAB (r = .56–.83, $p < .01$). The SAT correlated more highly with the ASVAB first-factor score ($r = .82$, $p < .01$) than with any other measure except for the Coop School & College Test ($r = .83$, $p < .01$). This is strong evidence that the SAT is an intelligence test.

Based on the regression corrected for nonlinearity with the addition of the squared SAT component, we developed the following equation to

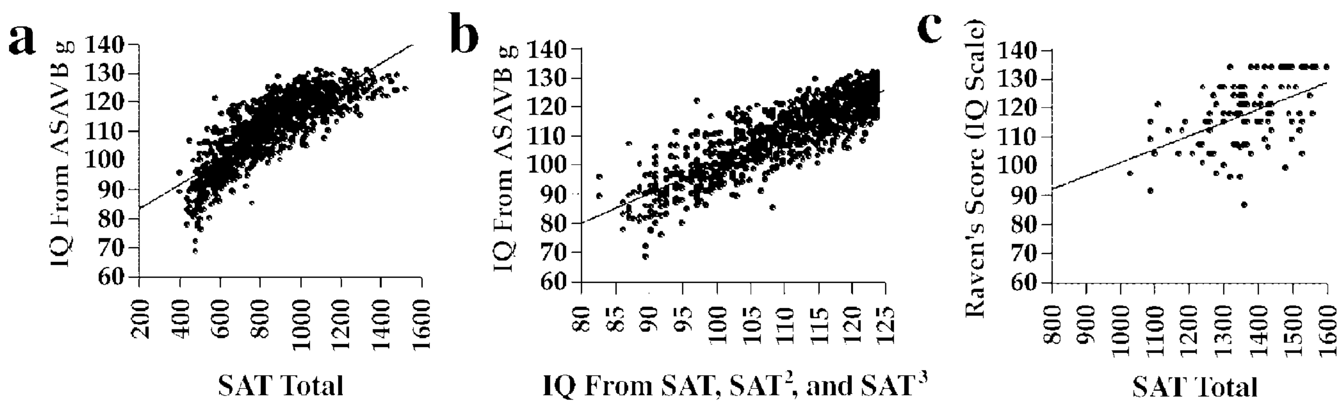

Fig. 1. Scatter plots of Scholastic Assessment Test (SAT) scores and IQ estimates: first-factor score (IQ scale) from the Armed Services Vocational Aptitude Battery (ASVAB) as a function of (a) SAT total score and (b) unstandardized predicted IQ based on SAT total score, SAT^2, and SAT^3 and (c) Raven's Advanced Progressive Matrices score (IQ scale) as a function of SAT total score.

predict IQ from SAT scores:

$$X'_{IQ} = (0.126 * SAT) + (-4.71E - 5 * SAT^2) + 40.063 \quad (1)$$

The standard error of prediction (SE_p) was 5.94. This process of estimating IQ from SAT scores appears much more accurate than the customary means of predicting IQ using a series of demographic variables. As mentioned earlier, the SE_p using demographic variables is 11.4 (Karzmark et al., 1985), about twice as large as the SE_p using the SAT.

In order to test the equation for consistent prediction, we used a jackknife procedure. A regression equation developed on a random

TABLE 2

Simple Correlations Among Tests

	Test								
Test	California Test of Mental Maturity[a]	Otis-Lennon Mental Ability Test[a]	Lorge-Thorndike Intelligence Test[a]	Henmon-Nelson Test of Mental Maturity[a]	Differential Aptitude Test[a]	Coop School & College Ability Test[a]	Scholastic Aptitude Test	Armed Services Vocational Aptitude Battery[b]	Total *n*
California Test of Mental Maturity[a]		.76** (12)	.77 (6)	.88** (7)	.58** (25)	.89** (19)	.82** (31)	.78** (358)	599
Otis-Lennon Mental Ability Test[a]			.86** (27)	.53 (11)	.74** (85)	.61 (5)	.78** (79)	.76** (572)	1,191
Lorge-Thorndike Intelligence Test[a]				.38 (12)	.55** (64)	.49* (17)	.79** (29)	.56** (295)	691
Henmon-Nelson Test of Mental Maturity[a]					−.53 (7)	.86** (19)	.65** (15)	.69** (166)	201
Differential Aptitude Test[a]						.77** (28)	.78** (68)	.75** (600)	569
Coop School & College Ability Test[a]							.53* (15)	.83** (162)	164
Scholastic Aptitude Test								.82** (917)	948
Armed Services Vocational Aptitude Battery[b]									11,914

Note. Numbers in parentheses indicate the *n* for each correlation.
[a]Correlations performed on percentile score. [b]First-factor scores from factor analysis of the subtests.
$^{*}p < .05$. $^{**}p < .01$.

sample of approximately 50% of the Study 1 (NLSY79) data was used to predict IQ in the remaining half of the sample. A second regression equation was developed on the second half of the sample and used to predict IQ in the first half. The correlation between IQ predicted from the first-half regression equation and IQ extracted from the ASVAB in the second half of the data was .856 ($p < .01$). The correlation between IQ predicted from the regression equation developed on the second half of the sample and IQ extracted from the ASVAB in the first half of the data was .860 ($p < .01$).

Discussion

It is evident from these results that there is a striking relation between SAT scores and measures of general cognitive ability. In fact, when one examines the results in Table 2, especially those in the ASVAB column, it appears that the SAT is a better indicator of g, as defined by the first factor of the ASVAB, than are some of the more traditional intelligence tests. Furthermore, jackknife procedures indicate highly consistent results are obtained with Equation 1.

Since 1979, when the data used in this analysis were collected, the SAT has undergone revision and recentering. Therefore, Equation 1 may not be applicable for a younger subject population. In Study 2, we evaluated the usefulness of this equation for the revised and recentered SAT.

STUDY 2

Method

Participants

One hundred sixteen students from a private university were recruited through the psychology subject pool. Of this total, 65 males and 39 females were included in the analysis; the remaining 12 students did not have valid SAT scores on record at the university.

Procedure

Subjects were administered the Raven's Advanced Progressive Matrices (APM; 1962 revision; J.C. Raven, 1978), a test of nonverbal reasoning skills. The test was administered to subjects in groups of 10 to 25 individuals, and the sessions were untimed. Following test administration, all participants' SAT scores were obtained from college admissions records.

Raven's APM scores were converted to percentiles using 1993 smoothed detailed U.S. norms (J. Raven, Raven, & Court, 1998). Z scores were computed from the percentiles, then transformed to an IQ scale with the same equation used in Study 1. IQs obtained in this manner were identical (within rounding error) to those obtained using Table APM36 of the Raven's APM manual, which converts percentiles to IQ (J. Raven et al., 1998, p. APM 102). Following the conversion to an IQ scale, we calculated a simple correlation between the Raven's APM and SAT total scores for the 104 subjects who had valid scores on both measures.

Results

Preliminary box plots revealed an outlier in the data set, which was excluded from further analysis. The correlation between the remaining 103 SAT and APM scores, although significant, was lower than the correlation obtained in Study 1 ($r = .483$, $p < .001$). The scatter plot in Figure 1c shows the simple correlation between SAT total scores and IQ as obtained from the Raven's APM. The mean SAT score of the sample in Study 1 was 854, whereas the mean SAT score of the sample in Study 2 was 1372. Furthermore, the standard deviation of the sample in Study 2 was 119, and the standard deviation obtained in Study 1 was 226 (the standardization of the SAT had a standard deviation of 200). These comparisons indicated a restricted range of the Study 2 participants. We therefore corrected the Study 2 simple correlation between SAT and Raven's IQ ($r = .483$) for restriction of range, simply to obtain an estimate of the correlation in a less restricted sample of college students. The resulting corrected correlation was .72. In addition, it is evident from Figure 1c that there is a ceiling on the Raven's scores, which may have also suppressed the correlation.

In contrast to what we found in Study 1, entering SAT math (SAT-M) and SAT verbal (SAT-V) scores into the regression equation in place of a composite, SAT total score, slightly improved the prediction of IQ. The multiple R for SAT-M and SAT-V entered separately was .554 ($p < .001$), or .745 when corrected for restriction of range.

We developed an equation for predicting IQ from SAT scores in the Study 2 data set. The resulting equation includes SAT-M and SAT-V scores, in place of the total score, as neither the squared nor the cubic component of SAT added significantly to the prediction of IQ ($p > .05$). The following is the equation obtained:

$$X'_{\mathrm{IQ}} = (0.095 * SAT\text{-}M) + (-0.003 * SAT\text{-}V) + 50.241 \qquad (2)$$

The standard error of prediction (SE_{p}) was 9.76. Although this standard error is higher than what was obtained in Study 1, it is nonetheless more accurate than the 11.4 one would expect estimating IQ using demographic variables. Part of the reason for the larger standard error is the restricted range of Study 2.

Repeating the jackknife procedure used in Study 1 on the Study 2 data set (i.e., using equations developed on one half of the data to predict IQ in the second half) yielded correlations of .523 (equation developed on first half) and .542 (equation developed on second half), respectively (both significant at the .01 level). These correlations indicate that results obtained using Equation 2 are stable.

Cross-Validation

Both studies showed highly consistent results even when a jackknife procedure was used. There was very little shrinkage in the correlations when regression equations were applied to new samples. This finding raised the question of whether it would be possible to develop a single equation that would apply to both samples. To test this, we applied the equation from Study 1 to the data of Study 2 and the equation from Study 2 to the data of Study 1. If one equation can apply to both data sets, then there should be very little shrinkage in correlations when regression equations are applied across the studies.

The equations developed from both studies were validated across the data sets. Equation 1, developed on the NLSY79 data set, was used to estimate IQ from the Study 2 SAT scores. These estimated IQs were then correlated with the Raven's IQs of the Study 2 subjects, as well as the IQs we estimated from Equation 2. IQs predicted from Equation 1 using SAT scores from Study 2 correlated significantly with both the Raven's IQs ($r = .481$, $p < .01$) and the IQs predicted from Equation 2 ($r = .862$, $p < .01$). Equation 2 was then used to estimate IQ from the SAT scores of Study 1. These estimates were then correlated with both ASVAB IQs ($r = .482$, $p < .01$) and IQs predicted from Equation 1 ($r = .526$, $p < .01$). Not surprisingly, there was substantial shrinkage in the cross-validation. This shrinkage could indicate one of three things: First, the equation developed using the NLSY79 data set may not be valid, in any sample, for predicting IQ with the new recentered SAT scores. Second, the characteristics of the Study 2 sample (i.e., much higher mean and restricted range of the SAT scores compared with the Study 1 sample) may have resulted in decreased effectiveness of the Study 1 equation in predicting IQ, but the Study 1 equation might be applicable to other, less restricted samples. Third, the estimate of *g* used in Study 1 may be substantively different from the measure of *g* in Study 2.

Discussion

The comparison between Study 1 and Study 2 samples presents an interesting problem. Can one assume that the revised, recentered SAT predicts *g* in the same manner that the previous version (used in the NLSY79 data set) did? Despite the differences in the samples and the best regression equations, there was still a strong correlation between the IQs predicted by the two equations. Unfortunately, the results of the jackknife procedures indicate that the substantial shrinkage in cross-validation correlations across studies is due, at least in part, to the inability of Equation 1 to predict IQ using the recentered SAT scores, although this equation is completely acceptable to use with archived data sets or individuals who took the SAT prior to the 1994 recentering. Ultimately, Equation 2 is more appropriate for use with recentered scores. However, Equation 2 is likely to be most useful in predicting IQ at the high end of the distribution, and therefore must be used with caution.

Although Equation 2 is better suited than Equation 1 for predicting high-end IQ, it is still a valuable tool. Individual difference researchers at colleges and universities will likely encounter a population similar to the one on which Equation 1 was based. Therefore, this equation will provide an alternative to time-consuming intelligence testing if only an estimate of IQ is required.

An alternative explanation of the differences between Equations 1 and 2 can be based on the measures of general intelligence used. The ASVAB first factor, although often used as a surrogate for *g*, can be thought of as primarily measuring what individuals have learned, or crystallized intelligence (Gc; Roberts et al., 2000). In contrast, the Raven's APM is typically characterized as a test of reasoning ability, or fluid intelligence (Gf). Therefore, Equation 1 may be likely to best reproduce IQ as measured in tests of Gc, most notably the widely used Wechsler scales, whereas Equation 2 may be more useful in predicting IQ as measured by tests of fluid reasoning. Further support for the Gf-SAT correlation is evident in an article by Raz, Willerman, Ingmundson, and Hanlon (1983), who reported a correlation of .81 between SAT total score and another nonverbal measure of Gf, the Cattell Culture Fair Intelligence Test; however, the extreme-groups design of the study may have resulted in an artificially high correlation.

Unfortunately, the usefulness of Equation 2 may not last long. Currently, the SAT is being revised to shift the focus from general reasoning ability to academic achievement. The objective is to test content knowledge rather than intelligence (Barnes, 2002). After this overhaul of the SAT is completed in 2005, another examination of the relationship between SAT scores and general cognitive ability will be required to determine if the SAT will still be an adequate measure of general intelligence. We expect that it will.

GENERAL DISCUSSION

Overall, the results of these studies support two major findings. First, the SAT is an adequate measure of general intelligence, and second, it is a useful tool in predicting cognitive functioning when other estimates of intelligence are unavailable, too time-consuming, or too costly. One implication of these results is that it might be more useful if the SAT were reported as a score on a general factor, plus separate math and verbal subscale scores.[1] Using the regression equations presented here, SAT scores can be converted to estimates of IQ. These estimates are especially useful in studies of college students when a rough measure of *g* is needed. Although it would be perfectly acceptable to use SAT scores without conversion, conversion to an IQ score provides a basis for comparing studies.

Another application of the SAT-IQ conversion is as an estimate of premorbid intelligence, as often SAT scores are the only objective measure of premorbid intellectual functioning available, and the resulting estimate of IQ is much more accurate than standard estimates based on demographic variables. We have provided two equations that can be used to estimate IQ from SAT scores, depending on whether the scores are from before or after the 1994 recentering. The evidence presented here strongly suggests that estimates of general intellectual functioning obtained from SAT scores are accurate and acceptable, and that the SAT-IQ conversion is to be encouraged, whether for clinical application or in a research setting.

Acknowledgments—Parts of this work were supported by Grant No. HD07176 from the National Institute of Child Health and Human Development, Office of Mental Retardation.

REFERENCES

Barnes, J.E. (2002, November 11). The SAT revolution: The new test spells the end of IQ- and big changes for American education. Retrieved November 18, 2002, from *U.S. News & World Report* Web site: http://www.usnews.com/usnews/issue/021111/education/11sat.htm

Brodnick, R.J., & Ree, M.J. (1995). A structural model of academic performance, socioeconomic status, and Spearman's *g*. *Educational and Psychological Measurement*, *55*, 583–594.

Jensen, A.R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Karzmark, P., Heaton, R.K., Grant, I., & Matthews, C.G. (1985). Use of demographic variables to predict full scale IQ: A replication and extension. *Journal of Clinical and Experimental Neuropsychology*, *7*, 412–420.

Lemann, N. (1999). *The big test: The secret history of the American meritocracy.* New York: Farrar, Straus, & Giroux.

Morgan, R. (1990). Analyses of the predictive validity of the SAT and high school grades from 1976 to 1985. In W.W. Willingham, C. Lewis, R. Morgan, & L. Ramist (Eds.), *Predicting college grades: An analysis of institutional trends over two decades* (pp. 195–212). Princeton, NJ: Educational Testing Service.

Raven, J., Raven, J.C., & Court, J.H. (1998). *Raven Manual: Sec. 4. Advanced Progressive Matrices (1998 ed.).* Oxford, England: Oxford Psychologists Press.

Raven, J.C. (1978). *Advanced Progressive Matrices: Set II (1962 revision).* London: H.K. Lewis & Co.

[1]We would like to thank Nathan Brody for this suggestion.

Raz, N., Willerman, L., Ingmundson, P., & Hanlon, M. (1983). Aptitude-related differences in auditory recognition masking. *Intelligence*, *7*(2), 71–90.

Ree, M.J., & Carretta, T.R. (1994). Factor analysis of the ASVAB: Confirming a Vernon-like structure. *Educational & Psychological Measurement*, *54*, 459–463.

Roberts, R.D., Goff, G.N., Anjoul, F., Kyllonen, P.C., Pallier, G., & Stankov, L. (2000). The Armed Services Vocational Aptitude Battery (ASVAB): Little more than acculturated learning (Gc)!? *Learning and Individual Differences*, *12*(1), 81–103.

Seligman, D. (1994). *A question of intelligence*. New York: Citadel Press.

Willingham, W.W., & Breland, H.M. (1982). *Personal qualities and college admissions*. New York: College Entrance Examination Board.

(Received 2/19/03; Revision accepted 5/2/03)

Erratum

Meredith C. Frey and Douglas K. Detterman, "Scholastic Assessment or *g*? The Relationship Between the Scholastic Assessment Test and General Cognitive Ability." *Psychological Science*, Volume 15, Number 6, June 2004, pp. 373–378.

There is an error in Equation 2 on page 376. All of the coefficients in the equation ought to be positive. The equation line should read:

$$X'_{\mathrm{IQ}} = (0.095 * SAT\text{-}M) + (0.003 * SAT\text{-}V) + 50.241 \qquad (2)$$

The authors would like to thank Wendy Johnson and Arthur Jensen for drawing their attention to this error.