Intelligence 53 (2015) 16–22

Contents lists available at ScienceDirect

Intelligence

Relations among general intelligence (*g*), aptitude tests, and GPA: Linear effects dominate

Thomas R. Coyle

University of Texas at San Antonio, Department of Psychology, 1 UTSA Circle, San Antonio, TX 78249, United States

ARTICLE INFO

Article history:
Received 22 February 2015
Received in revised form 26 June 2015
Accepted 13 August 2015
Available online 29 August 2015

Keywords:
General intelligence (*g*)
Spearman's Law of Diminishing Returns
Threshold theories
Aptitude tests (SAT ACT, PSAT, ASVAB)
Grade point average (GPA)

ABSTRACT

This research examined linear and nonlinear (quadratic) relations among general intelligence (*g*), aptitude tests (SAT, ACT, PSAT), and college GPAs. Test scores and GPAs were obtained from the National Longitudinal Survey of Youth ($N = 1950$) and the College Board Validity Study ($N = 160670$). Regressions estimated linear and quadratic relations among *g*, based on the Armed Services Vocational Aptitude Battery, composite and subtest scores of aptitude tests, and college GPAs. Linear effects explained almost all the variance in relations among variables. In contrast, quadratic effects explained trivial additional variance among variables (less than 1%, on average). The results do not support theories of intelligence (threshold theories or Spearman's Law of Diminishing Returns), which predict that test scores lose predictive power with increases in ability level or at a certain threshold.

1. Introduction

Suppose a large group of students takes an academic aptitude test (SAT or ACT). Further suppose that the ability level of the group ranges from very low (IQ < 70) to very high (IQ > 130). Should relations between the aptitude test and academic outcomes (e.g., college GPAs) decrease at higher levels of ability or beyond a certain ability threshold (e.g., IQ > 120)? The purpose of the current study was to address this question and its implications for intelligence.

This study examined linear and nonlinear relations among general intelligence (*g*), college grade point average (GPA), and aptitude tests. The aptitude tests included the SAT (formerly, Scholastic Aptitude Test), ACT (formerly, American College Test), Preliminary SAT (PSAT), and Armed Services Vocational Aptitude Battery (ASVAB). These tests are used for selection in college admissions and the armed services. In addition, the tests are strongly related to *g* (e.g., $\beta \approx .78$, Coyle & Pillow, 2008, Fig. 2), which reflects variance common to mental tests. A test's *g* loading (correlation with *g*) predicts outcomes in everyday life: Tests with strong *g* loadings strongly predict work and school performance, and tests with weak *g* loadings weakly predict such outcomes (Jensen, 1998, pp. 274–294).

Predictions were guided in part by threshold theories of cognitive ability (e.g., Robertson, Smeets, Lubinski, & Benbow, 2010). Such theories have been promoted by Malcolm Gladwell, who asserted in *Outliers* (Gladwell, 2008, p. 79), "The relationship between success and IQ works only up to a point. Once someone has an IQ of somewhere around 120, having additional IQ points doesn't seem to translate into any measureable real-world advantage." The important point for the current study is that threshold theories predict that test scores lose (linear) predictive power beyond a certain level, which produces nonlinear effects. These nonlinear effects should manifest in regression as quadratic effects, as tests lose predictive power beyond the threshold. (In a scatterplot of GPA against IQ, quadratic effects would appear as a function with positive slope and a concave bend [apex at top, curve opens down] around the IQ threshold.)

Similar predictions follow from Spearman's Law of Diminishing Returns (SLODR; Jensen, 1998, pp. 585–588; te Nijenhuis & Hartmann, 2006, p. 438). SLODR is based on Spearman's (1932) observation that correlations among tests (which estimate *g* loadings) decrease at higher levels of mental ability. In Spearman's (1932, p. 219) words, "The correlations [among tests] always become smaller—showing the influence of *g* on any ability to grow less—in just those classes of person which, on the whole, possess this *g* more abundantly. The rule is, then, that the more 'energy' [i.e., *g*] a person has available already, the less advantage accrues to his ability from further increments of it."

SLODR is based on the theory that higher levels of ability predict cognitive specialization (e.g., verbal and math). This specialization reduces a test's *g* loading, which is directly related to predictive validity (Jensen, 1998, pp. 274–294). Because *g* loadings decrease with ability level, tests should lose predictive power with increases in ability level. This pattern should manifest in regression as quadratic effects, as tests lose predictive power with increases in ability level. (In a scatterplot of GPA against IQ, quadratic effects would appear as a function with a positive slope

E-mail address: thomas.coyle@utsa.edu.

http://dx.doi.org/10.1016/j.intell.2015.08.005
0160-2896/

and a concave bend, which would reflect the loss of IQ's predictive power.)

SLODR and threshold theories have practical implications for the predictive power of *g*-loaded tests (e.g., SAT or ACT). First, if the predictive power of *g* decreases as the ability level of subjects increases (as SLODR implies), then *g*-loaded tests should predict outcomes (e.g., GPA) more weakly at higher levels of ability. Second, if the predictive power of *g* is neutralized beyond a certain ability threshold (as threshold theories predict), then *g*-loaded tests should have no predictive power beyond that threshold (e.g., IQ 120). Both predictions suggest that *g*-loaded tests should show non-trivial quadratic effects, due to the loss of linear effects at higher levels of ability.

Detterman and colleagues provide preliminary data relevant to SLODR and threshold theories. SAT (Frey & Detterman, 2004) and ACT (Koenig, Frey, & Detterman, 2008) scores were correlated with *g* scores, based on the ASVAB. Contra SLODR and threshold theories, linear effects (between each test and *g*) were dominant ($r = .82$ and .77, SAT and ACT, respectively). In contrast, quadratic effects were trivial ($\Delta R^2 = .05$ and .00, SAT and ACT, respectively), and only the SAT confirmed the predicted pattern (positive slope, concave bend).

The current study is the first systematic and comprehensive analysis of linear and quadratic effects among *g*, aptitude tests, and GPAs. Test scores were obtained from the 1997 National Longitudinal Survey of Youth (NLSY, $N = 8984$) and the College Board Validity Study (CBVS, $N = 192{,}467$), two large and representative samples of students in the United States. The NLSY and CBVS included college GPAs and aptitude tests (SAT, ACT, PSAT, ASVAB), which estimated *g* and tested the predictions of threshold theories and SLODR. In addition, the NLSY and CBVS sampled ability levels over 3 *SD*s above average (IQ > 145, where $M = 100$, $SD = 15$), which was sufficient to detect quadratic effects beyond the assumed threshold (IQ ≈ 120).

The current study differed from prior studies of aptitude tests and *g* (Coyle & Pillow, 2008; Coyle, Purcell, Snyder, & Kochunov, 2013; Frey & Detterman, 2004; Koenig et al., 2008). First, whereas prior studies analyzed the SAT and ACT, the current study also analyzed the PSAT. The PSAT is taken a year before the SAT and ACT and determines eligibility for National Merit Scholarships. Unlike the SAT and ACT, the PSAT has not been correlated with *g*. Together, the PSAT and other tests provide the first comprehensive analysis of linear and nonlinear relations among *g*, aptitude tests, and GPAs.

Second, whereas prior studies examined nonlinear effects of *composite* scores, which reflect performance on multiple tests (e.g., Frey & Detterman, 2004), the current study also examined nonlinear effects of *subtest* scores. Subtest scores are based on individual tests (e.g., math or verbal), which are loaded with specific variance and have lower *g* loadings. Because a test's predictive power is related to its *g* loading (Jensen, 1998, pp. 274–294), the lower *g* loadings of subtests should decrease predictive power. This decrease in predictive power should manifest as stronger quadratic effects, especially at high ability levels, as subtests become loaded with more specificity (due to cognitive specialization) and less *g*.

Third, whereas prior studies estimated effects between aptitude tests and *g* (Coyle et al., 2013; Frey & Detterman, 2004; Koenig et al., 2008), the current study also estimated effects between aptitude tests and college GPA. College GPA is moderately related to the SAT and ACT (β ≈ .43) and to *g* (β ≈ .30), which strongly predicts the SAT and ACT (β ≈ .78) (Coyle & Pillow, 2008, Fig. 2A, 2B). The current study provides the first test of linear and quadratic effects of the SAT and ACT with *g* and GPA, the latter being the usual criterion in validity studies. In addition, the study also estimates linear and quadratic effects for all possible pairs of aptitude tests (using composite and subtest scores), making it the most comprehensive study of its kind.

Regressions estimated linear effects among variables, and also estimated additional variance explained by quadratic effects. Predictions were based on SLODR and threshold theories. First, if the predictive power of cognitive tests decreases with increases in ability level or at high thresholds, then quadratic effects should explain non-trivial variance beyond linear effects. Second, if quadratic effects are greater for test scores loaded with more specificity (verbal and math), then such effects should be greater for subtest scores than composite scores. Third, if these predictions apply to any cognitive variable, then the results should replicate for *g*, aptitude tests, and GPAs.

2. Method

2.1. Subjects

Subjects were drawn from the NLSY ($N = 8984$) and the CBVS ($N = 192{,}467$). In the NLSY, subjects were selected if they had ASVAB scores and either SAT or ACT scores ($N = 1950$), which are used in college admissions. In the CBVS, subjects were selected if they had scores for all three SAT subtests (math, reading, writing) ($N = 160{,}670$).

2.2. NLSY variables

Table 1 reports *M*s and *SD*s for variables from the NLSY. SAT subtest scores were the math (SATm) and verbal (SATv) scores (range = 200 to 800). ACT subtest scores were the math (ACTm) and reading (ACTv) scores, with verbal ability based on reading (range = 1 to 36). PSAT subtest scores were the math (PSATm) and verbal (PSATv) scores (range = 20 to 80).

SAT composite scores (SATc) were the sum of SATm and SATv scores (range = 400 to 1600). ACT composite scores (ACTc) were the average of the four ACT subtest scores (math, reading, English, science) (range = 1 to 36). PSAT composite scores (PSATc) were the sum of PSATm and PSATv scores (range = 40 to 160).

GPA measured the average grade points earned during the first year of college (range = 0 to 4.0). GPAs were based on self-reports, which correlate strongly with actual GPAs ($r = .90$, Kuncel, Credé, & Thomas, 2005, Table 1, p. 73). The distribution of GPAs showed elevated kurtosis (2.85) and skewness (−1.16). GPAs were squared to normalize the distributions (kurtosis = −.46; skewness = −.16). The squared GPAs were used in analyses.

ASVAB scores were from 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. ASVAB scores were based on item response theory statistics, with higher scores indicating better performance.

g was based on the first unrotated factor in principal axis factoring of the ASVAB, which yields a strong *g* (Ree & Carretta, 1994). *g* accounted for most of the variance in test scores (variance = 55%; eigenvalue = 6.54). *g* factor scores were estimated using regression, which weighted ASVAB scores by *g* factor coefficients (Tabachnick & Fidell, 2007, pp. 622–625). To facilitate interpretation, *g* factor scores were standardized ($M = 0$, $SD = 1$) prior to analyses.

2.3. CBVS variables

Table 2 reports *M*s and *SD*s for variables from the CBVS. SAT subtest scores were from the math (SATm), reading (SATv), and writing (SATw) subtests (range = 200 to 800). SAT composite scores (SATc) were the sum of all three subtest scores (range = 600 to 2400). GPA measured the average grade points earned during the first year of college (range = 0 to 4.33).

2.4. Statistical analyses

Regressions estimated linear effects in an initial step, followed by quadratic (nonlinear) effects in a second step. Each regression involved a predictor (e.g., *g*) and a criterion (e.g., SAT or GPA). To facilitate interpretation, predictors were standardized and centered prior to analysis