Intelligence 54 (2016) 90–104

Contents lists available at ScienceDirect

Intelligence

Role of mental abilities and mental tests in explaining high-school grades☆

Jeffrey M. Cucina *, Sharron T. Peyton, Chihwei Su, Kevin A. Byle

U.S. Customs and Border Protection, 1400 L Street, NW, 7th Floor, Washington, DC 20229-1145, United States

ARTICLE INFO

Article history:
Received 22 February 2015
Received in revised form 27 November 2015
Accepted 29 November 2015
Available online xxxx

Keywords:
Mental abilities
Academic performance
Grade point average
Conscientiousness
Specific aptitude theory

ABSTRACT

It is well-known that some students earn higher grades than others; however, published research on the mental abilities that are correlated with high school grades is sparse. Two studies examined the relationship between different mental abilities and high school grades. Study 1 showed that the personality trait conscientiousness predicted high school grades ($r = .32$) almost as well as g ($r = .37$ to .40). In Study 2, the relationship between general mental ability (g) and high school grades was linear and fairness analyses indicated slight overprediction for Hispanics and Blacks and underprediction for females. Validity was lowered slightly by group preferences. With the exception of mathematical knowledge, the correlation between mental abilities and high school grades in both studies was largely attributable to g rather than specific abilities (s) measured by each test. Additional analyses showed that grade point averages are reliable and conscientiousness and g do not interact when predicting high school grades.

1. Introduction

Research on predictors of job and training performance paints a clear picture: general mental ability (g) is primarily responsible for the predictive power of standardized tests of mental abilities, with specific mental abilities (s) adding little or nothing to the prediction of *job* and *training* performance (Brown, Le, & Schmidt, 2006; Gottfredson, 2002; Hunter, 1983a, 1983b, 1984, 1985, 1986; Jensen, 1986; Olea & Ree, 1994; Ree & Earles, 1991; Ree, Earles, & Teachout, 1994; Sackett & Wilk, 1994; Schmidt, 1988, 2011; Schmidt & Hunter, 1996 [see Scenario 6], Schmidt & Hunter, 2004; Schmidt, Hunter, & Caplan, 1981; Schmidt, Ones, & Hunter, 1992; Thorndike, 1985, 1986). However, there is a lack of research on the role of specific mental abilities in the relationship between standardized tests and *academic* performance using large samples. Instead, much of the large sample research focuses on the validity of specific operational tests (e.g., the SAT, the ACT), which only include a small number of subtests. The studies that have been conducted on the role of s have used small sample sizes. Conry and Plant (1965) and Anderson (1971) examined the criterion-related validity of scores from test batteries measuring multiple abilities for high school (HS) performance, but only used 98 and 127 students, respectively. Jensen (1998) noted the lack of research, stating "Surprisingly little of this applied literature on test validity, however, examines the degree to which g itself, as compared to other factors and specificity, contributes to tests' validity. Fortunately, the few studies that focus specifically on this question have been conducted by the armed forces and by the U.S. Employment Service of the Department of Labor [i.e., Hunter & Schmidt]. These studies, based on huge samples, are technically excellent" (p. 271).

There has been research in the mental abilities' literature which suggests that specific abilities might be related to academic performance (above and beyond g). Reeve (2004) has shown that narrow latent abilities can add incremental validity over g when predicting scores on achievement tests. Additionally, there is some evidence that perceptual speed adds incremental validity for predicting clerical job performance and that spatial ability adds incremental validity for predicting job performance for certain technical jobs (Gottfredson, 2002; Jensen, 1998; Johnson & Bouchard, 2005; Schmidt, 1988). Past research also has demonstrated the importance of spatial ability for scientific, technical, engineering, and math (STEM) jobs (Kell, Lubinski, Benbow, & Steiger, 2013; Lubinski, 2010; Wai, Lubinski, & Benbow, 2009; Webb, Lubinski, & Benbow, 2007). Additionally, Carroll's (1993) meaningful memory factor has emerged as a unique predictor of training performance (Cucina, Su, Busciglio, & Thompson Peyton, 2015) and Coyle and Pillow (2008) have shown that the non-g residuals of the SAT and ACT predict undergraduate academic performance.

Furthermore, most criterion-related validation work in academic settings has focused on college (Hezlett et al., 2001) and graduate-level performance (e.g., Kuncel & Hezlett, 2007; Kuncel, Hezlett, & Ones, 2001, 2004); few studies exist on HS performance. Grigorenko et al. (2009) stated that "In contrast to [the] rich literature on college-

☆ The views expressed in this paper are those of the authors and do not necessarily reflect the views of U.S. Customs and Border Protection or the U.S. Federal Government. Portions of this research have been presented at the annual meetings of the Society for Industrial and Organizational Psychology (Division 14 of the American Psychological Association) in 2014 and will be presented at the American Educational Research Association in 2016.

* Corresponding author.
E-mail address: jcucina@gmail.com (J.M. Cucina).

http://dx.doi.org/10.1016/j.intell.2015.11.007
0160-2896/

level tests, there is a much smaller body of literature on the predictive validity of secondary-level standardized tests" (p. 964). This is somewhat surprising given that admissions-testing programs do exist for high schools. For example, the New York City (2010) school system uses the Specialized High Schools Admissions Test (SHSAT), which consists of verbal and mathematical tests, and many private schools use the Secondary School Admission Test (SSAT; SSAT Board, 2012), which consists of reading comprehension, quantitative, verbal, and essay tests. According to a critique by Feinman (2008), no criterion-related validation studies have been conducted on the SHSAT; however, Grigorenko et al. (2009) noted that there is some research on the validity of the SSAT (they reported a validity of .377 in their largest sample). Some studies have shown that HS grades are related to college entrance examination scores (e.g., Mattern & Patterson, 2013) and a study by Chamorro-Premuzic, Quiroga, and Colom (2009) reported that HS grades were related to college entrance examination scores but not to measures of fluid, crystallized, and visual intelligence. Although there are some studies examining the relationship between mental abilities and achievement tests (e.g., Furnham & Monsen, 2009; Furnham, Rinaldelli-Tabaton, & Chamorro-Premuzic, 2011; Reeve, 2004) and teacher ratings (Chamorro-Premuzic, Harlaar, Greven, & Plomin, 2010) for high school students, these studies did not use overall course grades (e.g., grade point average) as a criterion.

1.1. Conscientiousness and HS grades

In contrast to the mental abilities' literature, there is more extensive research on the factors of personality that predict academic performance. Factor analytic work indicates that five large factors underlie personality variables (Digman, 1990; Digman & Takemoto-Chock, 1981; Norman, 1963; Tupes & Christal, 1961, 1992). These factors include Neuroticism (i.e., negative emotionality), Extraversion (i.e., sociability and energy level), Openness to Experience (i.e., imagination, intellect and culture), Agreeableness (i.e., cooperation, friendliness and consideration), and Conscientiousness (i.e., dutifulness and achievement-striving). The latter factor has been given prominence for predicting academic performance in a review by de Raad & Schouwenburg (1996); similar conclusions were drawn from the results of a meta-analysis by McAbee and Oswald (2013). This factor also predicts job and training performance (Barrick & Mount, 1991). Conscientiousness taps aspects of personality such as dutifulness and achievement-striving. Digman and Takemoto-Chock (1981) interpreted this dimension as a person's will to achieve, whereas Cattell (1957, 1973) interpreted it as the strength of the superego. This dimension is quite similar to Webb's (1915) factor of volition or will. Individuals who are high in Conscientiousness are persevering, responsible, dependable, ambitious and scrupulous. Whereas individuals low in Conscientiousness can be described as fickle, likely to quit, unambitious, undependable, careless, impulsive, lazy, and lacking in self-control.

1.2. The importance of studying HS grades

Studying HS grades as a criterion is important for several reasons. First, the use of standardized testing for college admissions remains controversial, with many critics arguing that more weight be given to HS grades or that HS grades be used in lieu of standardized test scores. Even proponents of the SAT and ACT (e.g., Kobrin, Sinharay, Haberman, & Chajewski, 2011) suggest that test scores and HS grade point average (GPA) should be used in combination (e.g., using a regression equation). However, there is relatively little research on what psychological constructs HS grades measure. Since both HS grades and training performance are learning activities, we hypothesize that HS grades will be predicted by *g* and conscientiousness, just as training performance is (Schmidt & Hunter, 1998). Second, recent reforms in the education system (e.g., No Child Left Behind) have given rise to alternative approaches to developing curricula for public schools. For example, several schools (see Seider & Geiger, 2009) have adopted Gardner's (1983) multiple intelligences framework into their curricula, customizing instruction in different modes to match students' strengths and weaknesses on different specific abilities (Armstrong, 1994; Blythe & Gardner, 1990; Dastgoshadeh & Jalilzadeh, 2011; Gardner & Hatch, 1989; Howard Gardner Multiple Intelligence School, 2010; Su, 2012). This approach is partially predicated on the assumption that different abilities contribute to academic success. Third, some specialized public high schools, and many private high schools, use standardized tests as part of their admissions process (see the discussion above); however, the criterion-related validity of mental-abilities tests in this setting remains largely unexamined. Fourth, the College Board President announced plans to revamp the SAT (Strauss, 2013); however, there is relatively little large-scale research on the correlates of academic performance above and beyond the SAT and ACT. Fifth, there is very little published research on the reliability of HSGPA. Camara and Michaelides (2005, p. 2) suggested that HSGPA might be "unreliable" (p. 2) and Gesier and Santelices (2007) indicated that HSGPA has a "reputation for 'unreliability'" (p. 27); however, neither examined data to estimate the reliability of HSGPA. There is evidence that similar criteria have reasonably adequate reliability coefficients in the .80s—Ramist, Lewis, and McCamley (1990) conducted a multi-sample study (with 40,622 students in 38 colleges) which estimated the reliability of Freshman undergraduate GPA to be .82.

In this paper, we present two empirical studies examining the relationship between mental abilities and academic performance. Study 1 also examined the relationship between the conscientiousness personality factor and academic performance (including the possibility of an interaction between *g* and conscientiousness). Study 2 added analyses examining the reliability of HSGPA, the presence of predictive bias, and the impact of minority preferences on the criterion-related validity of mental ability test scores.

2. Study 1

The first study examined the criterion-related validity of a large battery of mental abilities tests and a measure of conscientiousness in the prediction of HS grades. Past research suggests that *g*, not *s*, predicts job/training performance, thus we predict the same finding for our

Table 1
Demographic Statistics.

Variable	Frequency	Percent
Gender		
Male	155,109	48.2
Female	166,479	51.8
Missing	1	<.1
Race/ethnicity[a]		
White (non-Hispanic)	132,822	41.3
Black (non-Hispanic)	4612	1.4
Hispanic (non-Black)[b]	301	0.1
Native American[c]	209	0.1
Asian	935	0.3
Missing/other	182,710	56.8
Grade		
Ninth	84,526	26.3
Tenth	84,457	26.3
Eleventh	80,848	25.1
Twelfth	71,757	22.3
Missing	1	<.1
	Mean	SD
Age[d]	15.79	1.26
Missing	7183	2.2%
Total	321,589	

[a] We use modern terminology to describe the races/ethnicities. When Project TALENT was begun (in the 1960s) other terms were used for these groups.
[b] Includes cases labeled as "Mexican American," "Puerto Rican American," and "Cuban".
[c] Includes cases labeled as "American Indian" and "Eskimo".
[d] Age was missing for 7183 cases (2.2% of the sample).

study (Hypothesis 1-1). We also hypothesize that conscientiousness will predict HS grades (Hypothesis 1-2), based on meta-analytic research showing that it predicts college GPA (McAbee & Oswald, 2013). Hypothesis 1-3 states that conscientiousness and *g* will interact when predicting performance, following the classic equation (Maier, 1955; Vroom, 1960, 1964):

$$Performance = motivation \times ability.$$

Table 2
Names, Abbreviations, and Factors for Mental Abilities Tests.

Ability	Abbreviation	Carroll broad factor			Carroll narrow factor		
		1	2	3	1	2	3
Information tests							
Vocabulary	T102VOCA	2C			V	VL	
Literature	T103LITE	2C			V	K2	
Music	T104MUSI	2C			V	K2	
Social studies	T105SOCI	2C			V	K2	
Mathematics	T106MATH	2C			KM	N	A3
Physical science	T107PHYS	2C			V	A7	
Biological science	T108BIOL	2C			V	A7	
Scientific attitude	T109SCIE	2F	2R		RG/I	SP	
Aeronautics and space	T110AERO	2C			V	A7	
Electricity and electronics	T111ELEC	2C			V	MK	A7
Mechanical	T112MECH	2C			V	MK	
Farming	T113FARM	2C			V	MK	
Home economics	T114HOME	2C			V		K2
Sports	T115SPOR	2C			V		K2
Art	T131ART	2C			V	K2	
Law	T132LAW	2C			V	K2	
Health	T133HEAL	2C			V	A7	
Engineering	T134ENGI	2C			V	A7	
Architecture	T135ARCH	2C			V	K2	
Journalism	T136JOUR	2C			V	K2	
Foreign travel	T137FORE	2C			V	K2	
Military	T138MILI	2C			V	K2	
Accounting, business, sales	T139ACCO	2C			V	K2	
Practical knowledge	T140PRAC	2C			V	K2	
Clerical	T141CLER	2C			V	K2	
Bible	T142BIBL	2C			V	K2	
Colors	T143COLO	2C			V	K2	
Etiquette	T144ETIQ	2C			V	K2	
Hunting	T145HUNT	2C			V	K2	
Fishing	T146FISH	2C			V	K2	
Other outdoor activities	T147OTHE	2C			V	K2	
Photography	T148PHOT	2C			V	K2	
Games (sedentary)	T149GAME	2C			V	K2	
Theater and ballet	T150THEA	2C			V	K2	
Food	T151FOOD	2C			V	K2	
Miscellaneous test score	T152MISC	2C			V		
Non-information tests							
Vocabulary	T162VOCA	2C			V	VL	
Memory for sentences	T211MEMO	2Y			MA		
Memory for words	T212MEMO	2Y			MA		
Disguised words	T220DISG	2R	2F	2C	FW	RG/I	V
Spelling	T231SPEL	2C			SG	V	
Capitalization	T232CAPI	2C			MY	SG	V
Punctuation	T233PUNC	2C			MY	V	
English usage	T234ENGL	2C			MY	V	
Effective expression	T235EFFE	2C			MY	V	
Word functions in sentences	T240WORD	2C			MY	V	
Reading comprehension	T250READ	2C			RC	V	
Creativity	T260CREA	2R			FI	FE	
Mechanical reasoning	T270MECH	2V			VZ	MK	
Visualization in 2 dimensions	T281VISU	2V			VZ	SR	
Visualization in 3 dimensions	T282VISU	2V			VZ		
Abstract reasoning	T290ABST	2F			RG/I		
Math I. Arithmetic reasoning	T311MTH1	2F			RQ		
Math II. Introductory HS mathematics	T312MTH2	2C			KM	N	A3
Math III. Advanced HS Mathematics	T333MTHe	2C			KM	N	A3
Arithmetic computation (R − 3W)	T410ARIT	2S			N		
Table reading (R − W)	T420TABL	2S			P		
Clerical checking (R − 3W)	T430CLER	2S			P		
Object inspection (R − W)	T440OBJE	2S			P		

Notes: Carroll's (1993) broad abilities are as follows: 2C = Crystallized Intelligence; 2F = Fluid Intelligence; 2R = Broad Retrieval Ability; 2Y = General Memory & Learning; 2 V = Broad Visual Perception; 2S = Broad Cognitive Speediness. Carroll's narrow abilities are as follows: V = Verbal (Printed) Language Comprehension; VL = Lexical Knowledge (Vocabulary); K2 = Cultural Information (e.g., art, history; see Carroll, page 521); KM = Knowledge of Mathematics (see page 523–4); N = Numerical Facility; A3 = Tested Math Achievement (see page 532); A7 = Tested Science Achievement (see page 524); RG/I = General Reasoning Sequential Reasoning/Induction (these were combined per Colberg, 1985; Colberg, Nester, & Trattner, 1985); SP = Sensitivity to Problems; MK = Mechanical Knowledge (see page 525); MA = Associative Memory; FW = Word Fluency; SG = Spelling Ability; MY = Grammatical Sensitivity; RC = Reading Comprehension; FI = Ideation Fluency; FE = Expressional Fluency; VZ = Visualization; SR = Spatial Relations; RQ = Quantitative Reasoning; P = Perceptual Speed.

We used the dataset for the Project TALENT study which was conducted by Flanagan and associates (1964). Although it was conducted some time ago, it makes up for its age by virtue of the sheer number of cases and tests administered.

2.1. Method

We used the base-year dataset from Project TALENT (American Institutes for Research, 1960a,b). Project TALENT has data on over 300,000 high school students on a battery of 59 mental abilities tests as well as personality tests and other measures (e.g., interests, background questionnaires). More information on the dataset and study can be found in Flanagan et al. (1961, 1964, Wise, McLaughlin, & Steel, 1979). This dataset has been used in a large number of studies, primarily conducted in the 1960s–1970s (e.g., Campbell, 1979; Cureton, 1968); however, it continues to be used in research today (e.g., Arneson, Sackett, & Beatty, 2011; Major, Johnson, & Deary, 2012; Reeve, 2001, 2004; Reeve, Meyer, & Bonaccio, 2006; Waters, 2007). Despite a current bibliography of over 356 publications to date, no study has investigated the relationship between the 59 mental abilities tests, conscientiousness, and HS grades using the full dataset. We suspect that this is due to the relative difficulty of conducting such analyses prior to the advent of modern computers. Indeed, only one study examined HS grades, and it used only four of the 59 tests, a subsample of approximately 3000 of the 300,000 participants and was focused on the relationship between "cultural capital" and HS grades, rather than mental abilities and personality (DiMaggio, 1982). Other studies focused on the relationships between the tests and college grades for a subset of the participants (e.g., Arneson et al., 2011) or between different Project TALENT tests (e.g., Reeve, 2001, 2004; Waters, 2007). For example, Reeve (2004) used the ability tests as a predictor of scores on the achievement tests; our study uses both to predict HS grades. Indeed, there is some disagreement in the literature about the relationship between individual differences in the general factor for ability and the general factor for achievement and whether achievement is a predictor or criterion. Reeve (2004) notes that whether a variable is a predictor or criterion depends on the hypotheses being tested and our hypotheses identified the external variables of HS grades and GPA as the criteria. Readers who view achievement as a criterion variable and not as a predictor variable can focus on the results for the *g* scores we computed without using the achievement/information tests. These readers could also refer to the analyses that were conducted by Reeve (2004) who used achievement tests as the criterion (as opposed to HS grades).

The Project TALENT researchers provided a credibility index for each participant and suggested that cases with an unacceptable index should be removed before conducting analyses. These cases were identified using a screening test (which consisted of very easy items designed to measure whether or not the participant was motivated to cooperate with the researchers and provide accurate data) and measures of reading (to ensure literacy), slowness, and inaccuracy. We removed these cases (comprising 8.1% of the cases) from further analyses. In addition, we also removed cases that had missing data on one or more of the abilities tests. It was necessary to remove these cases because scores on all of the tests were needed to compute variables for *g*. Table 1 presents the demographics for the cases in our final dataset. According to analyses conducted using G*POWER (Faul, Erdfelder, Lang, & Buchner, 2007; Faul, Erdfelder, Buchner, & Lang, 2009) a sample of this size yields 99% power for detecting a significant correlation coefficient of 0.01 (a 99% confidence interval around a correlation of 0.01 is 0.005 to 0.015 at this sample size). Note that we used an archival dataset for this study; none of the authors

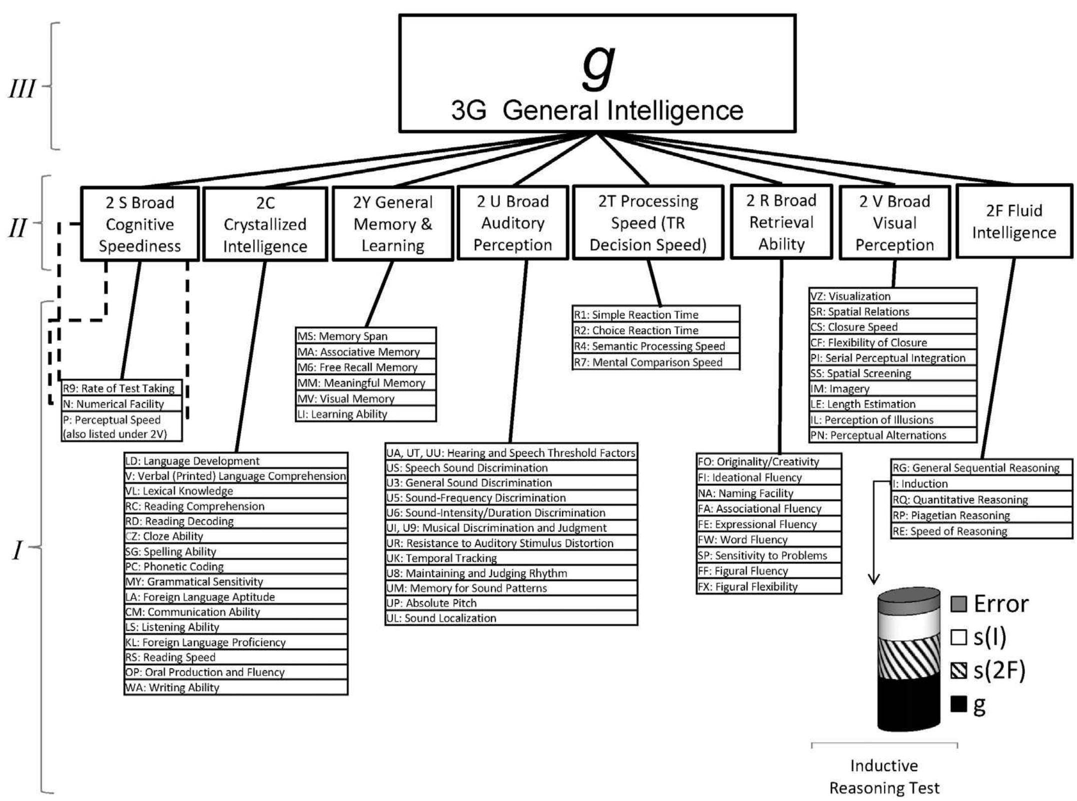

Fig. 1. This is a representation of Carroll's (1993) Three Stratum Theory of intelligence, also known as the Carroll–Horn–Cattell model. Note that general mental ability (*g*) lies at the highest stratum, with 8 broad abilities at the second stratum, and 69 specific narrow abilities at the first stratum. Technically, there should be lines from the second stratum abilities to each of the specific abilities falling under a broad ability. We present dashed versions of these lines for the Broad Cognitive Speediness ability, but leave them out for the remaining 7 broad abilities. Also, note that the abilities presented here represent factors obtained through factor analyses and do not represent test scores. To explain how variance in a test score could be explained under this theory, we present a hypothetical representation of an inductive reasoning test score in the bottom right corner. Note that this score includes variance attributable to 3G (*g*; general intelligence), 2F (Fluid Intelligence), I (Induction), and error variance. In essence, this is an extension of Spearman's (1904 and 1927) two-factor model of intelligence and other models can be subsumed under it (Keith & Reynolds, 2010; McGrew, 2005, 2009; McGrew & Evans, 2004).

Table 3
Incremental criterion-related validity of mental abilities tests for standardized HS grades criterion controlling for g score from non-Information tests.

Ability	Bivariate				Incremental validity over g									Our linkage				Major et al.'s linkage			
	r_{obs}	p	ρ_{OV}	ρ_{TS}	$r_{obs \cdot g}$	p	β_{obs}	p	ΔR^2_{obs}	ΔR_{obs}	β_{TS}	ΔR^2_{TS}	ΔR_{TS}	SEM_{path}	p	$r_{SEM\text{-}implied}$	$r_{SEM\text{-}implied}$ - r_{obs}	SEM_{path}	p	$r_{SEM\text{-}implied}$	$r_{SEM\text{-}implied}$ - r_{obs}
g-All tests	.37	<.01	.39	.40																	
g-No info	.40	<.01	.43	.43																	
T102VOCA	.30	<.01	.32	.38	.01	<.01	.01	<.01	<.01	<.01	−.02	<.01	<.01	−.03	<.01	.28	.02	−.16	<.01	.30	.00
T103LITE	.28	<.01	.30	.36	.01	<.01	.02	<.01	<.01	<.01	<.01	<.01	<.01	−.01	<.01	.26	.02	−.01	<.01	.29	−.01
T104MUSI	.29	<.01	.31	.38	.06	<.01	.07	<.01	<.01	<.01	.12	.01	.01	.02	<.01	.23	.06	.02	<.01	.25	.04
T105SOCI	.31	<.01	.33	.37	.06	<.01	.07	<.01	<.01	<.01	.09	<.01	<.01	.02	<.01	.27	.04	.01	<.01	.28	.03
T106MATH	.36	<.01	.39	.44	.12	<.01	.15	<.01	.01	.01	.24	.02	.02	.01	<.01	.26	.10	.02	<.01	.31	.05
T107PHYS	.29	<.01	.31	.36	.07	<.01	.08	<.01	<.01	<.01	.11	.01	.01	.02	<.01	.23	.06	.10	<.01	.27	.02
T108BIOL	.23	<.01	.25	.34	.02	<.01	.02	<.01	<.01	<.01	.02	<.01	<.01	−.03	<.01	.22	.01	.06	<.01	.24	.00
T109SCIE	.22	<.01	.23	.37	−.02	<.01	−.02	<.01	<.01	<.01	−.45	.02	.02	−.05	<.01	.23	−.01				
T110AERO	.11	<.01	.12	.17	−.06	<.01	−.06	<.01	<.01	<.01	−.14	.01	.01	−.10	<.01	.19	−.08	.01	<.01	.20	−.09
T111ELEC	.11	<.01	.12	.16	−.05	<.01	−.05	<.01	<.01	<.01	−.10	.01	.01	−.08	<.01	.18	−.07	.02	<.01	.18	−.07
T112MECH	.04	<.01	.05	.06	−.13	<.01	−.13	<.01	.01	.02	−.23	.04	.04	−.17	<.01	.18	−.14	−.06	<.01	.17	−.13
T113FARM	.17	<.01	.18	.24	−.02	<.01	−.02	<.01	<.01	<.01	−.06	<.01	<.01	−.06	<.01	.18	−.01				
T114HOME	.21	<.01	.23	.32	.06	<.01	.06	<.01	<.01	<.01	.10	.01	.01	.03	<.01	.12	.09				
T115SPOR	.17	<.01	.18	.24	−.02	<.01	−.02	<.01	<.01	<.01	−.06	<.01	<.01	−.07	<.01	.19	−.02				
T131ART	.24	<.01	.26	.32	−.01	<.01	−.01	<.01	<.01	<.01	−.04	<.01	<.01	−.03	<.01	.24	.00	−.02	<.01	.26	−.02
T132LAW	.20	<.01	.21	.31	−.02	<.01	−.02	<.01	<.01	<.01	−.10	<.01	<.01	−.05	<.01	.22	−.02	−.05	<.01	.23	−.03
T133HEAL	.24	<.01	.26	.35	.02	<.01	.02	<.01	<.01	<.01	.03	<.01	<.01	.00	0.02	.21	.03	.06	<.01	.24	.00
T134ENGI	.13	<.01	.14	.24	−.05	<.01	−.05	<.01	<.01	<.01	−.24	.02	.03	−.08	<.01	.18	−.05				
T135ARCH	.21	<.01	.22	.38	.04	<.01	.04	<.01	<.01	<.01	.11	<.01	.01	.01	<.01	.17	.04				
T136JOUR	.22	<.01	.24	.38	.01	<.01	.01	<.01	<.01	<.01	.01	<.01	<.01	−.01	<.01	.21	.01				
T137FORE	.21	<.01	.22	.33	<.01	.90	<.01	.90	<.01	<.01	−.03	<.01	<.01	−.03	<.01	.22	−.01				
T138MILI	.15	<.01	.16	.27	−.01	<.01	−.01	<.01	<.01	<.01	−.06	<.01	<.01	−.04	<.01	.17	−.02				
T139ACCO	.22	<.01	.24	.34	<.01	.60	<.01	.60	<.01	<.01	−.03	<.01	<.01	−.02	<.01	.22	.00				
T140PRAC	.14	<.01	.15	.24	−.07	<.01	−.07	<.01	<.01	<.01	−.28	.03	.03	−.08	<.01	.18	−.04				
T141CLER	.16	<.01	.17	.32	<.01	.19	<.01	.19	<.01	<.01	−.02	<.01	<.01	−.01	<.01	.13	.03				
T142BIBL	.27	<.01	.29	.34	.08	<.01	.09	<.01	.01	.01	.12	.01	.01	.07	<.01	.19	.08	.07	<.01	.22	.06
T143COLO	.18	<.01	.20	.40	.05	<.01	.05	<.01	<.01	<.01	.17	.01	.01	.03	<.01	.11	.07				
T144ETIQ	.16	<.01	.17	.55	.02	<.01	.02	<.01	<.01	<.01	−.21	.04	.05	.02	<.01	.10	.06				
T145HUNT	−.05	<.01	−.05	−.11	−.09	<.01	−.09	<.01	.01	.01	−.12	.01	.02	−.11	<.01	.07	−.12				
T146FISH	.01	<.01	.01	.02	−.05	<.01	−.05	<.01	<.01	<.01	−.16	.01	.01	−.07	<.01	.08	−.07				
T147OTHE	.17	<.01	.19	.27	−.04	<.01	−.05	<.01	<.01	<.01	−.07	<.01	<.01	−.08	<.01	.21	−.04				
T148PHOT	.14	<.01	.15	.38	−.03	<.01	−.03	<.01	<.01	<.01	−.02	<.01	<.01	−.04	<.01	.15	−.01				
T149GAME	.14	<.01	.15	.28	−.01	<.01	−.01	<.01	<.01	<.01	−.02	<.01	<.01	−.02	<.01	.15	−.01				
T150THEA	.24	<.01	.26	.34	<.01	.08	<.01	.08	<.01	<.01	.04	<.01	<.01	−.01	<.01	.22	.02	−.01	<.01	.24	.00
T151FOOD	.16	<.01	.17	.28	<.01	>.99	<.01	>.99	<.01	<.01	−.07	<.01	<.01	−.02	<.01	.15	.01				
T152MISC	.24	<.01	.26	.38	.02	<.01	.02	<.01	<.01	<.01	.05	<.01	<.01	−.01	<.01	.22	.02	−.01	<.01	.23	.01

T162VOCA	.27	<.01	.29	.36	−.01	<.01	−.01	<.01	<.01	<.01	.11	.01	.01	−.03	<.01	.26	.02	.05	<.01	.11	.16
T211MEMO	.17	<.01	.19	.24	.04	<.01	.04	<.01	<.01	<.01	−.03	<.01	<.01	.03	<.01	.11	.06	.06	<.01	.19	−.02
T212MEMO	.29	<.01	.31	.34	.08	<.01	.09	<.01	.01	.01	.27	.02	.03	.02	<.01	.19	.10				
T220DISG	.27	<.01	.29	.31	−.01	<.01	−.01	<.01	<.01	<.01	−.09	<.01	.01	.01	<.01	.24	.03	.01	<.01	.24	.03
T231SPEL	.31	<.01	.34	.44	.09	<.01	.11	<.01	.01	.01	.50	.03	.03	.10	<.01	.20	.11	.06	<.01	.22	.09
T232CAPI	.19	<.01	.20	.22	−.05	<.01	−.05	<.01	<.01	<.01	−.48	.01	.02	−.02	<.01	.18	.01	−.05	<.01	.19	.00
T233PUNC	.37	<.01	.40	.47	.10	<.01	.15	<.01	.01	.01	−.27	.02	.02	.07	<.01	.26	.11	.02	<.01	.27	.10
T234ENGL	.26	<.01	.28	.39	−.02	<.01	−.02	<.01	<.01	<.01	.41	.05	.05	.01	<.01	.23	.03	−.03	<.01	.23	.03
T235EFFE	.20	<.01	.22	.29	−.06	<.01	−.07	<.01	<.01	<.01	−.15	<.01	<.01	.01	<.01	.20	.00	−.06	<.01	.21	−.01
T240WORD	.41	<.01	.44	.49	.19	<.01	.26	<.01	.03	.04	−.24	.02	.02	.12	<.01	.24	.17	.13	<.01	.25	.16
T250READ	.32	<.01	.35	.38	−.02	<.01	−.03	<.01	<.01	<.01	−.29	.05	.05	−.02	<.01	.31	.01	.06	0.01	.32	.01
T260CREA	.23	<.01	.25	.30	−.07	<.01	−.09	<.01	<.01	<.01	−.14	.01	.02	−.06	<.01	.25	−.02	.06	<.01	.25	−.02
T270MECH	.11	<.01	.12	.14	−.15	<.01	−.17	<.01	.02	.02	−.23	.02	.03	−.06	<.01	.20	−.09	.02	<.01	.23	−.12
T281VISU	.11	<.01	.12	.14	−.09	<.01	−.10	<.01	.01	.01	−.37	.03	.03	.00	0.06	.16	−.05	.01	<.01	.16	−.05
T282VISU	.17	<.01	.18	.23	−.09	<.01	−.10	<.01	.01	.01	.07	<.01	<.01	.01	<.01	.19	−.02	.00	0.95	.19	−.02
T290ABST	.24	<.01	.26	.32	−.07	<.01	−.09	<.01	<.01	.01	.45	.04	.04	−.04	<.01	.24	.00	.06	<.01	.23	.01
T311MTH1	.32	<.01	.35	.41	.03	<.01	.05	<.01	<.01	<.01	.10	<.01	<.01	.02	<.01	.27	.05	.03	<.01	.28	.05
T312MTH2	.38	<.01	.41	.48	.13	<.01	.19	<.01	.01	.02	−.01	<.01	<.01	.03	<.01	.26	.13	−.02	<.01	.31	.07
T333MTHe	.24	<.01	.26	.38	.04	<.01	.04	<.01	<.01	<.01	−.08	.01	.01	−.01	<.01	.20	.05				
T410ARIT	.22	<.01	.23	.25	<.01	.04	<.01	.04	<.01	<.01	−.01	<.01	<.01	.02	<.01	.10	.12	.01	<.01	.17	.05
T420TABL	.07	<.01	.08	.08	−.07	<.01	−.06	<.01	<.01	<.01	−.08	.01	.01	−.02	<.01	.12	−.05	−.02	<.01	.08	−.01
T430CLER	.15	<.01	.16	.17	−.01	<.01	−.01	<.01	<.01	<.01	−.02	<.01	<.01	.03	<.01	.12	.03	.03	<.01	.09	.06
T440OBJE	.08	<.01	.09	.10	−.06	<.01	−.06	<.01	<.01	<.01	<.01	<.01	<.01	−.03	<.01	.11	−.03	.01	<.01	.11	−.03
2C														.00	0.03	.28	.02				
2C & KM														.07	<.01	.26	.02				
2C & MY														.05	<.01	.23	.06				
2F														.00	0.37	.27	.04				
2R														.00	0.74	.26	.10				
2S														.01	<.01	.23	.06				
2V														−.10	<.01	.22	.01				
2Y														.08	<.01	.23	−.01				
A6																		.06	<.01		
K0																		−.02	<.01		
K1																		−.16	<.01		
KM																		.03	<.01		
P																		.01	<.01		
VZ																		−.10	<.01		
g														.36	<.01	.19	−.08	.39	<.01		

Notes: For the analyses conducted using Major et al.'s (2012) linkage, we use the same abbreviations as the authors: K0 = General Verbal Information; K1 = Science Knowledge; A6 = English achievement; KM = Math achievement; P = Perceptual Speed; VZ = Visualization.

were responsible for, or had control over, the data collection. Therefore, we did not employ our own data-collection stopping rules.

2.1.1. HS grades

In addition to completing the mental abilities tests, participants also completed a "Student Information Blank" which consisted of a number of background questions. Eight of these questions dealt with HS grades and the Project TALENT researchers created a weighted composite using seven of these questions. The items and weights are provided in Table S1 of the Supplementary Online Materials (SOM) and were obtained from pages 36 and R-4 of Wise et al. (1979). The Project TALENT participants collectively attended 1214 schools, and there were 1095 schools that had 30 or more cases in the data. To address the possibility that different schools might have more lenient or severe grading standards, we created a second criterion by converting HS grades (for schools with 30 or more cases) to *z*-scores.

The use of self-reported HS grades is, admittedly, a limitation of the current study; however, Kuncel, Crede, and Thomas (2005) found that self-reported HS grades have an observed correlation of .82 with grades from school records. They stated that self-reported grades "generally predict outcomes to a similar extent as actual grades" (p. 76). They also recommend using self-reported grades "as replications when there is clear evidence that the findings agree with other research that used school-reported grades." Note that in this paper, we corroborate the results for the self-reported grades (in Study 1) with school-reported grades (in Study 2).

2.1.2. Conscientiousness

Ten personality scales were administered with the Project TALENT mental abilities tests. Reeve et al. (2006) re-administered the personality scales with an inventory from Goldberg's (1999) International Personality Item Pool and reported salient loadings on each of the big five personality factors. Using their results, we assigned each of the ten personality scales to the big five factor (or factors in the case of self-confidence, which had two loadings of .60) on which it had the highest salient loading. Two scales were used to compute Conscientiousness: Tidiness (example items provided by Reeve et al. include "I am never sloppy in my personal appearance" and "My work suffers from lack of neatness") and Maturity (example items provided by Reeve et al. include "It bothers me to leave a task half done" and "I do things the best I know how, even if no one checks up on me"). According to data reported by Reeve et al., these two scales had loadings of .79 and .63, respectively, on the Conscientiousness factor formed using both the ten personality scales and scales from the International Personality Item Pool (Goldberg, 1999; Goldberg et al., 2006).

2.1.3. Mental abilities tests

Table 2 lists the names/abbreviations for all tests. Each test was rationally linked to Carroll's (1993) mental abilities (see Fig. 1); Table 2 presents these linkages. Major et al. (2012) independently linked a subset of the tests to Carroll's abilities; we also used this linkage in additional analyses. In general, the first half of the tests measure crystallized intelligence and Flanagan et al. (1962) included these "Information" tests as measures of general knowledge, specifically, knowledge that is gained through general life rather than through formal schooling. The remaining tests measure specific abilities such as verbal, quantitative, spatial, perceptual speed, and memory. An analog to the Raven's Progressive Matrices (Raven, Raven, & Court, 2004), Test 290 Abstract Reasoning, was also administered.

2.2. Results

Descriptive statistics and reliabilities for the research variables are provided in Table S2 in the SOM. To address Hypothesis 1-1, we examined the criterion-related validity of the mental abilities tests. These analyses are summarized in Table 3 for the standardized HS grades criterion (additional validities using unstandardized grades are in Table S3).

We computed the bivariate observed correlation (r_{obs}) between each of the tests and HS grades. We also computed the operational validity (ρ_{ov}), which corrects for criterion unreliability (which will be described later) and the true score validity (ρ_{TS}), which corrects for predictor and criterion unreliability. Next, we conducted a Principal Components Analysis (PCA) on the tests and computed scores on the first unrotated principal component to serve as an estimate of *g* (see Jensen & Weng, 1994 for a justification). Two types of *g* scores were computed, one using all of the 59 tests, and the other omitting the crystallized knowledge information tests. The *g* scores using all 59 tests had an observed criterion-related validity of .32 for unstandardized HS grades and .37 for the standardized HS grades. The *g* scores that omitted crystallized knowledge had observed validities of .35 and .40, respectively. We focused our analysis on the *g* score omitting crystallized knowledge since it had slightly higher criterion-related validity. We computed the partial correlations ($r_{obs\cdot g}$) between each subtest and the HS grades criterion, controlling for the *g* scores, and we ran a series of regression analyses and recorded the incremental validity (ΔR_{obs}) and observed standardized regression weights (β_{obs}) for both *g* and each subtest. The latter weight estimates the relationship between each subtest and HS grades controlling for *g*.

Following recommendations by Schmidt et al. (1981), Schmidt and Hunter (1996, scenario 6), and Schmidt (2011), we also corrected the correlation matrix of predictors and the criterion for unreliability and ran a true-score regression analysis to obtain true score incremental validities (ΔR_{TS}) and standardize regression weights (β_{TS}). Finally, we conducted three types of Structural Equation Models (SEM). We first estimated the standardized regression path coefficients from *g* to the criterion. Next, leaving the path from *g* to the criterion intact (and fixing it to the first value), we added a path from a second factor (e.g., 2C) to the criterion and recorded its estimate. This serves as an estimate of the unique contribution of the *s* factor over and above *g*. We repeated this analysis for each of the remaining factors; however, we only included two paths (one of which was from *g*, the other from the target factor) to the criterion in each model. We then conducted this analysis at the test score level; we conducted 59 SEMs, all of which had a path from *g* to the criterion and each of which had a different subtest-to-criterion path. For example, one model had paths from *g*, 2F, and Test 290 to the criterion (i.e., three paths to the criterion) and another had paths from *g*, 2Y, and Test 211 to the criterion. The SEMs were conducted twice, once using our linkage of the tests to Carroll's (1993) factors and once using Major et al.'s (2012) linkage. To provide a better modeling of the factor structure, we included data from 14,278 cases that had scores on all tests but lacked criterion data (using AMOS's estimate means and intercepts option to handle the missing data). Table S4 presents the fit statistics for all of the models. We computed the model implied correlation between each subtest and HS grades using the model that only allowed *g* to load onto HS grades. The model implied correlations are shown in Table 3 along with the difference between them and the observed validities.

The test scores and the *g* scores were significantly related to HS grades. However, after controlling for *g*, the unique variance of the test scores had either negative validities or small validities that were less than .10 (i.e., "Irish validities" Kunin, 1965; Seberhagen, 2007). Furthermore, the observed validities were very close to those implied by the SEM model that only allowed *g* to predict HS grades, suggesting that much of the correlation between the tests and HS grades is attributable to the relationship between *g* and HS grades. Thus, we found strong support for Hypothesis 1-1.

Hypothesis 1-2 was also supported as the correlation between conscientiousness and HS grades was .32 for both the unstandardized and standardized criteria. Although the other big five factors predicted the criteria, their validity appears to be due to shared variance with Conscientiousness rather than a unique contribution. Conscientiousness added incremental validity (for predicting HS grades) over the big five scores scales for Neuroticism ($\Delta R^2 = .05^{***}$), Extraversion ($\Delta R^2 = .06^{***}$),

Table 4
Description of GPAs used as criteria.

GPA variable	Abbreviation	Comments
HSGPA-final	HSGPA-F	This is the final HSGPA over a student's high school (e.g., four-year) career. It was a credit-weighted average using the Carnegie credit system. The NLSY97 researchers standardized the coding system for HSGPAs so that when schools had differing grading systems (e.g., percentage systems vs. letter grade systems), the coded HSGPAs would be on a comparable scale.
HSGPAs for each year	HSGPAY1 HSGPAY2 HSGPAY3 HSGPAY4	The NLSY97 dataset also includes HSGPAs for each academic year (e.g., 1996–1997). To compute the reliability of HSGPA, we isolated the 3670 cases that had data for only 4 sequential academic years. To create the HSGPAs for years 1 (i.e., Freshman year) to 4 (i.e., Senior year), we manually reassigned the first year's GPA to the Freshman HSGPA (HSPGA-Y1), the second year's GPA to the Sophomore HSGPA (HSGPA-2), and so forth for each case.
HSGPA-English	GPAeng	This is the credit-weighted GPA for each student's grades in English courses in high school.
HSGPA-foreign language	GPAflng	This is the credit-weighted GPA for each student's grades in Foreign Language courses in high school.
HSGPA-Math	GPAmath	This is the credit-weighted GPA for each student's grades in Math courses in high school.
HSGPA-Social Sciences	GPASocS	This is the credit-weighted GPA for each student's grades in Social Studies courses in high school.
HSGPA-Life and Physical Sciences	GPAlpSci	This is the credit-weighted GPA for each student's grades in Life and Physical Science courses in high school.

Openness ($\Delta R^2 = .05^{***}$), and Agreeableness ($\Delta R^2 = .07^{***}$); however, none of these four personality scales added appreciable incremental validity over conscientiousness (all ΔR^2 were <.01 and were all statistically significant except for Agreeableness). Conscientiousness added incremental validity over *g* for both the unstandardized and standardized criteria ($\Delta R^2 = .06^{***}$ and $.06^{***}$, respectively) and vice versa ($\Delta R^2 = .09^{***}$ and $.11^{***}$, respectively) yielding an adjusted multiple *R*s of $.431^{***}$ and $.463^{***}$, respectively. To address Hypothesis 1-3, we computed interaction terms between standardized scores for Conscientiousness and the *g* scores. Although statistically significant, this interaction term did not add appreciable incremental validity over the main effects ($\Delta R^2 < .01$) for either of the criteria. Given the power associated with detecting moderating effect with our sample size (>.99) Type II error can be ruled out and Hypothesis 1-3 is not supported.

We also conducted two exploratory analyses. First, we examined the validities for different areas of coursework and found similar results (see Table S5). Second, we examined the validities by the four grades (i.e., ninth–twelfth) and again found similar results (see Table S6).

3. Study 2

Study 2 replicates the results of Study 1, while answering additional research questions and addressing three limitations of Study 1: age of the data, use of a self-report criterion, and reliability. To explain, Study 1 used data which is several decades old and it is possible that the increases over time in mean levels of *g* (Flynn, 1984, 1987, 2007, 2012) and grades (e.g., Kostal, Kuncel, & Sackett, 2014) could limit the generalizability of the results from Study 1 to present-day generations. Additionally, the median reliability of Study 1's 59 tests was .58, which falls below Nunnally and Bernstein's (1994) cutoff of .70 for research purposes; Study 2 includes more reliable measures (median reliability = .86).

Study 2 addressed five research questions/hypotheses: (1) what is the criterion-related validity of different mental abilities for HSGPA?, (2) are the tests that have criterion-related validity fair under the Cleary (1968) model?, (3) what is the impact of racial preferences on the validity of the valid tests?, (4) what is the *g*-loading of HSGPA?, and (5) what is the reliability of HSGPA?

The first research question (and Hypothesis 2-1) is the same as that for Study 1. The second and third research questions address two controversial issues on the use of mental ability tests for minority applicants: fairness and group preferences. We predict that the ASVAB scores will be fair under the Cleary (1968) model (Hypothesis 2-2). The practice of adjusting scores (also known as racial preference, within-group norming, and race-norming) remains legal, but controversial, in educational settings, despite being outlawed in U.S. employment settings (Civil Rights Act of 1991, Section 106 (1)). We could not locate any past research on the impact of these types of score adjustments on the criterion-related validity of mental ability tests. Given that there is more variance in *g* within groups than between groups (Brody, 1992; Jensen, 1980) and that applicant-faking adjustments have minimal effects on the criterion-related validity of personality measures (Rosse, Stecher, Levin, & Miller, 1998), Hypothesis 2-3 states that reducing between group variance will only slightly reduce the overall criterion-related validity of mental ability tests. We also hypothesized that HSPGA is *g*-loaded (Hypothesis 2-4) and reliable (Hypothesis 2-5), based on past research showing that HSPGA predicts college GPA, reliability is a prerequisite for validity, and both undergraduate GPA and job training grades are reliable.

3.1. Method

Archival data from the 1997 cohort of the National Longitudinal Survey of Youth (NLSY97) conducted by the U.S. Bureau of Labor Statistics (2005) were used. NLSY97 is a nationwide survey of a stratified-random sample of 8984 U.S. citizens who were between the ages of 12 and 16 at the end of 1996. Certain minority groups (i.e., Hispanics and Blacks) were oversampled in the study, allowing us to conduct fairness analyses. As part of the study, the Armed Services Vocational Aptitude Battery (ASVAB) was administered to 7127 (79%) of the participants, and the data was used to assist in norming the latest version of the ASVAB. Again, archival data were used and none of the authors were responsible for, or had control over, the data collection. Therefore, we did not employ our own data-collection stopping rules.

In terms of statistical power, Aguinis, Culpepper, and Pierce (2010) note that typical validation studies lack sufficient power to detect test bias. However, our study's sample sizes (minority group sizes ranging from 800 to 2400 cases) are so large that they do not appear on Aguinis et al.'s power analysis charts (which reached sufficient power in most cases).

Table 5
Demographic statistics.

Variable	Frequency	Percent
Gender		
Male	2385	49.5
Female	2431	5.5
Race/ethnicity		
White (non-Hispanic)	2626	54.5
Black (non-Hispanic)	1161	24.1
Hispanic (non-Black)	801	16.6
Native American	32	.7
Asian/Pacific islander	85	1.8
Missing/something else	111	2.3
Year of birth		
1980	994	2.6
1981	1087	22.6
1982	958	19.9
1983	875	18.2
1984	902	18.7
Total	4816	100

Table 6
Criterion-related validity of mental abilities tests for HSGPA-F.

Ability	r_{obs}	p	ρ_{OV}	ρ_{TS}	Incremental validity over g	p	β_{obs}	p	ΔR^2_{obs}	ΔR_{obs}	β_{TS}	ΔR^2_{TS}	ΔR_{TS}	SEM_{path}	p	$r_{SEM\text{-}implied}$	$r_{obs}-r_{SEM\text{-}implied}$
					$r_{obs \cdot g}$												
GS	.32	<.001	.34	.37	−.03	.019	−.05	.019	<.01	<.01	−.11	<.01	<.01	−.03	.026	.34	−.02
AR	.39	<.001	.41	.43	.08	<.001	.11	<.001	.01	.01	.12	.01	.01	.02	.171	.37	.02
WK	.31	<.001	.33	.34	−.04	.004	−.06	.004	<.01	<.01	−.09	<.01	<.01	−.05	<.001	.35	−.04
PC	.38	<.001	.40	.44	.08	<.001	.11	<.001	.01	.01	.15	.01	.01	.02	.135	.36	.02
NO	.18	<.001	.19	.23	.04	.002	.04	.005	<.01	<.01	.05	<.01	<.01	.04	.002	.15	.03
CS	.18	<.001	.19	.22	.01	.323	.01	.461	<.01	<.01	.01	<.01	<.01	.01	.354	.17	.01
AI	.13	<.001	.14	.15	−.09	<.001	−.09	<.001	.01	.01	−.12	.01	.01	−.06	<.001	.19	−.06
SI	.12	<.001	.12	.13	−.15	<.001	−.15	<.001	.02	.02	−.20	.03	.03	−.10	<.001	.22	−.10
MK	.43	<.001	.46	.48	.17	<.001	.23	<.001	.02	.03	.28	.03	.03	.07	<.001	.37	.06
MC	.29	<.001	.31	.34	−.05	.001	−.06	.001	<.01	<.01	−.12	.01	.01	−.03	.016	.33	−.04
EI	.23	<.001	.24	.26	−.10	<.001	−.13	<.001	.01	.01	−.18	.02	.02	−.07	<.001	.29	−.06
AO	.34	<.001	.36	.40	.08	<.001	.09	<.001	.01	.01	.12	.01	.01	.04	.002	.30	.04
g	.44	<.001	.47	.47										.49	<.001		
Math	N/A													.02	.222	.46	
Sci/tech.	N/A													−.03	.020	.44	
Verbal	N/A													.00	.90	.48	
Perc. speed	N/A													.02	.309	.31	

Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB was used as the mental abilities test and consists of 12 subtests: General Science (GS), Arithmetic Reasoning (AR), Word Knowledge/Vocabulary (WK), Paragraph Comprehension (PC), Numerical Operations (NO), Coding Speed (CS), Auto Information, (AI), Shop Information (SI), Mathematics Knowledge (MK), Mechanical Comprehension (MC), Electronics Information (EI), and Assembling Objects (AO).

3.1.1. HSGPA

High school transcripts were obtained for 6232 of the participants and a variety of HSGPAs were coded, recorded, and computed. The types of HSPGA used in the current study are listed in Table 4.

3.1.1.1. Armed Services Vocational Aptitude Battery (ASVAB). The ASVAB was used as the mental abilities test and consists of 12 subtests: General Science (GS), Arithmetic Reasoning (AR), Word Knowledge/Vocabulary (WK), Paragraph Comprehension (PC), Numerical Operations (NO), Coding Speed (CS), Auto Information (AI), Shop Information (SI), Mathematics Knowledge (MK), Mechanical Comprehension (MC), Electronics Information (EI), and Assembling Objects (AO). The ASVAB is one of the largest group-administered multiple-aptitude test battery programs in the world and was developed to match the curriculum and ability-levels of high school students for use in career counseling (as part of the no-cost Career Exploration Program) and military recruitment. The NLSY97 participants completed the ASVAB in a proctored setting using computer-adaptive testing. We used the theta (θ) estimates as the predictor variables in this study.

For the factor structure, we used the ASVAB's official factors, which consist of Math (AR, NO, MK) Science/Technical (EI, GS, AI, SI, MC), Spatial (AO), Verbal (PC, WK), and Perceptual Speed (NO, CS). A CFA using this structure gave acceptable fits: $\chi^2 = 738.5$, $df = 51$, GFI = .973, CFI = .966, NFI = .964. We considered alternative factor structures; however, we chose this factor structure as it produced models with slightly better fit statistics (see Table S9).

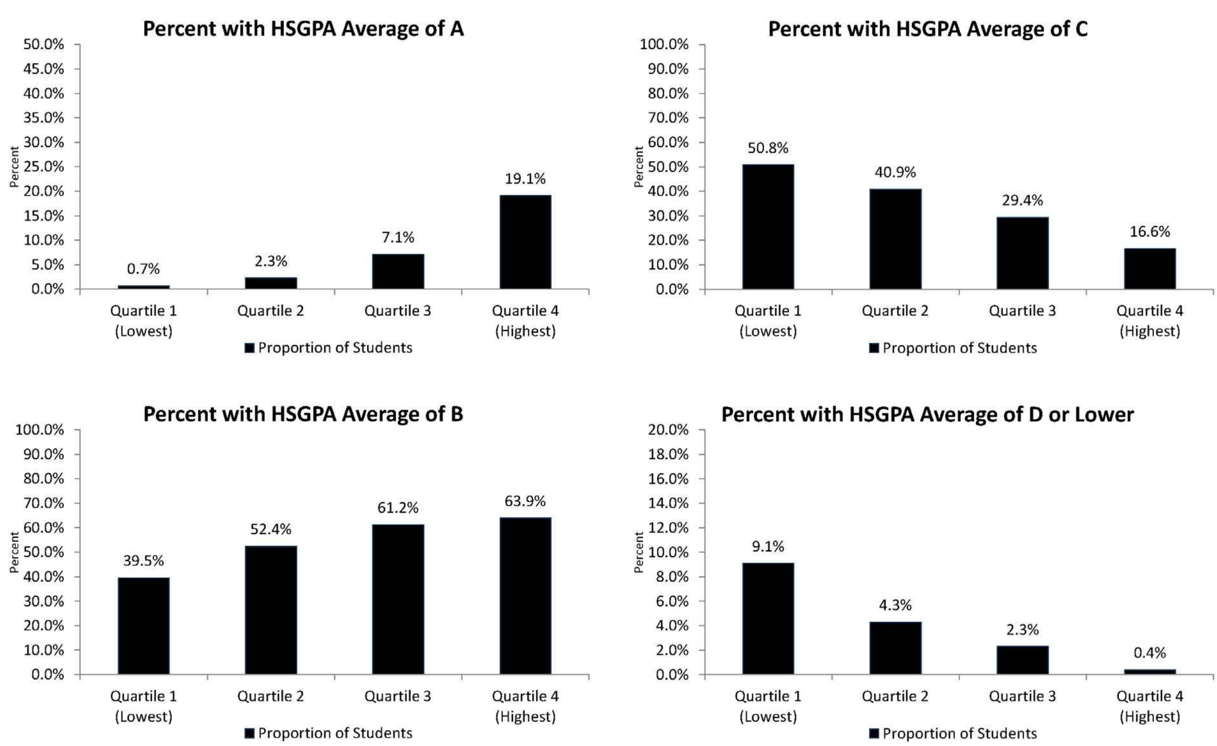

Fig. 2. Nester expectancy charts depicting percentage of students with average letter grade of A (≥3.70), B (2.70 to 3.69), C (1.70 to 2.69), D or lower (≤1.69) as a function of g-quartile.

Table 7
Cleary test-bias analyses.

Ability	Male vs fem.						B–W						H–W					
	$\Delta R^2_{int.}$	$\Delta R_{int.}$	p	$\Delta R^2_{slope.}$	$\Delta R_{slope.}$	p	$\Delta R^2_{int.}$	$\Delta R_{int.}$	p	$\Delta R^2_{slope.}$	$\Delta R_{slope.}$	p	$\Delta R^2_{int.}$	$\Delta R_{int.}$	p	$\Delta R^2_{slope.}$	$\Delta R_{slope.}$	p
GS	.04	.06	<.001	<.01	<.01	.001	.02	.02	<.001	.01	.01	<.001	<.01	<.01	.001	<.01	.01	<.001
AR	.04	.05	<.001	<.01	<.01	.009	.01	.01	<.001	.01	.01	<.001	<.01	<.01	.001	.01	.01	<.001
WK	.03	.05	<.001	<.01	<.01	.028	.02	.03	<.001	.01	.01	<.001	<.01	.01	<.001	<.01	.01	<.001
PC	.02	.02	<.001	<.01	<.01	.005	.02	.02	<.001	.01	.01	<.001	<.01	.01	<.001	<.01	.01	<.001
NO	.03	.07	<.001	<.01	<.01	.667	.05	.11	<.001	<.01	<.01	.854	.02	.05	<.001	<.01	<.01	.218
CS	.03	.06	<.001	<.01	<.01	.354	.05	.10	<.001	<.01	<.01	.826	.02	.05	<.001	<.01	<.01	.669
AI	.04	.11	<.001	<.01	<.01	.263	.05	.11	<.001	<.01	<.01	.032	.02	.08	<.001	<.01	<.01	.805
SI	.04	.12	<.001	<.01	<.01	.936	.05	.13	<.001	<.01	<.01	.221	.02	.08	<.001	<.01	<.01	.033
MK	.02	.03	<.001	<.01	<.01	.076	.01	.02	<.001	.01	.01	<.001	<.01	<.01	<.001	<.01	<.01	<.001
MC	.05	.07	<.001	<.01	<.01	.032	.02	.03	<.001	<.01	<.01	.001	.01	.01	<.001	<.01	<.01	.022
EI	.05	.08	<.001	<.01	<.01	.066	.03	.06	<.001	<.01	<.01	.067	.01	.02	<.001	<.01	<.01	.116
AO	.03	.04	<.001	<.01	<.01	.021	.02	.03	<.001	<.01	.01	<.001	.01	.02	<.001	<.01	<.01	.070
g	.04	.05	<.001	<.01	<.01	<.001	.00	.00	.009	<.01	<.01	<.001	<.01	<.01	.581	<.01	<.01	.001

3.2. Results

Demographics are displayed in Table 5 and the Appendix provides descriptive statistics and reliabilities for the research variables (Table S7) and their intercorrelations (Table S8). No corrections were needed for range restriction. To answer Research Question 1, we examined the criterion-related validity of the subtests. These analyses are summarized in Table 6 for the HSGPAF criterion. To begin, we computed the bivariate observed correlation (r_{obs}) between each of the 12 subtests and HSGPA. Just as in Study 1, we also computed operational validities (ρ_{ov}), the true score validities (ρ_{TS}), the *g* scores, the partial correlations ($r_{ST \cdot g}$), the incremental validities from hierarchical

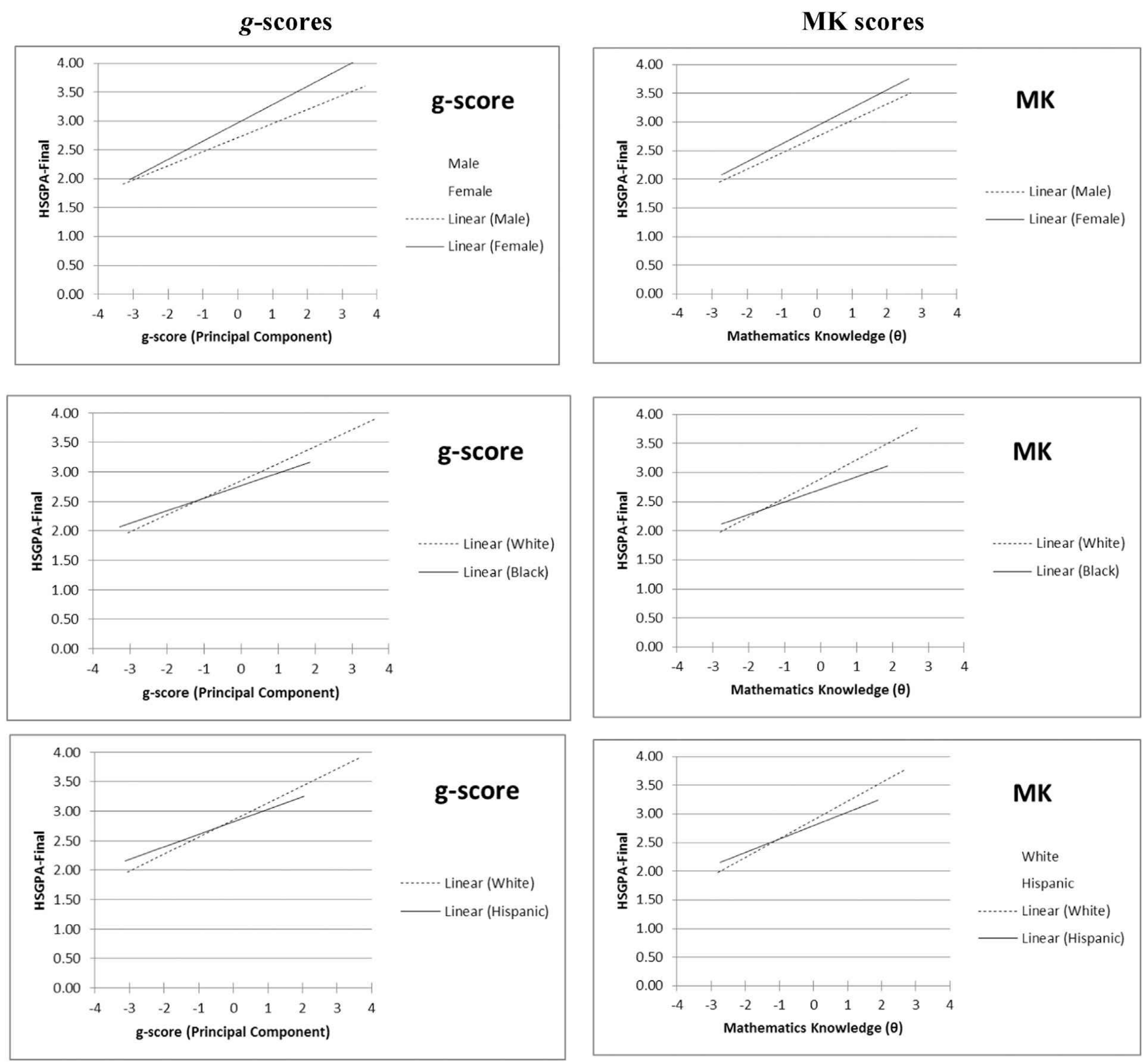

Fig. 3. Plots from Cleary analyses.

Table 8
Group differences with Cohen's d effect sizes.

Ability	Male		Female		White		Black		Hispanic		$d_{M\text{-}aF}$	p	d_{W-B}	p	d_{W-H}	p
	M	SD	M	SD	M	SD	M	SD	M	SD						
GS	−0.19	0.74	−0.32	0.74	−0.02	0.65	−0.61	0.71	−0.52	0.69	0.17	<.001	0.88	<.001	0.76	<.001
AR	−0.23	0.86	−0.30	0.86	−0.02	0.72	−0.69	0.87	−0.48	0.81	0.09	.001	0.88	<.001	0.62	<.001
WK	−0.36	0.82	−0.40	0.82	−0.13	0.70	−0.74	0.83	−0.65	0.83	0.05	.115	0.82	<.001	0.71	<.001
PC	−0.25	0.82	−0.09	0.82	0.02	0.73	−0.47	0.78	−0.38	0.77	−0.21	<.001	0.66	<.001	0.54	<.001
NO	11.19	8.84	11.87	8.84	12.18	9.20	10.69	8.61	10.64	7.96	−0.08	.008	0.16	<.001	0.17	<.001
CS	4.12	3.64	4.61	3.64	4.73	3.83	3.77	3.50	4.01	3.35	−0.13	<.001	0.26	<.001	0.20	<.001
AI	−0.69	0.67	−0.89	0.67	−0.66	0.60	−1.01	0.69	−0.91	0.66	0.31	<.001	0.56	<.001	0.42	<.001
SI	−0.57	0.67	−0.88	0.67	−0.52	0.59	−1.05	0.69	−0.91	0.71	0.48	<.001	0.85	<.001	0.63	<.001
MK	−0.05	0.88	0.04	0.88	0.20	0.81	−0.35	0.85	−0.23	0.85	−0.10	<.001	0.67	<.001	0.53	<.001
MC	−0.33	0.75	−0.52	0.75	−0.20	0.63	−0.80	0.75	−0.60	0.69	0.26	<.001	0.89	<.001	0.62	<.001
EI	−0.49	0.85	−0.72	0.85	−0.40	0.72	−0.91	0.77	−0.84	0.80	0.30	<.001	0.70	<.001	0.60	<.001
AO	−0.34	0.88	−0.29	0.88	−0.13	0.80	−0.71	0.81	−0.38	0.80	−0.06	.038	0.72	<.001	0.32	<.001
g	0.07	1.06	−0.06	1.06	0.38	0.88	−0.62	0.90	−0.37	0.92	0.13	<.001	1.12	<.001	0.84	<.001
HSGPA-F	2.73	0.60	2.95	0.60	2.96	0.60	2.64	0.59	2.74	0.59	−0.36	<.001	0.54	<.001	0.36	<.001
HSGPA-Jr.	2.84	0.55	3.04	0.55	3.04	0.55	2.72	0.54	2.86	0.49	−0.36	<.001	0.58	<.001	0.33	<.001

regression analyses (ΔR_{obs}), observed standardized regression weights (β_{obs}), true score incremental validities (ΔR_{TS}), and standardize regression weights (β_{TS}). Similarly, we conducted the same types of SEM analyses used in Study 1. That is, we ran three separate types of SEMs to estimate the standardized paths between (a) g and HSGPA, (b) each factor and HSGPA – controlling for g, and (c) each subtest and HSPGA – controlling for its associated factor and g. Table S9 presents the fit statistics for the various models and Table 6 presents the model implied correlations along with the difference between them and the observed validities.

In general, the ASVAB scores and the g scores were significantly related to HSGPA. The unique variance of the subtest scores had either negative validities or small validities that were less than .10 (i.e., "Irish validities" Kunin, 1965; Seberhagen, 2007). Furthermore, the observed validities were very close to those implied by the SEM model that only allowed g to predict HSGPA, suggesting that much of the correlation between the subtests and HSGPA is attributable to the relationship between g and HSGPA. The sole exception was MK, which rather consistently added incremental validity over g.

To illustrate the relationship between g and HSGPA, we divided the g scores into quartiles and computed the proportion of students with an "A," "B," "C," and "D or lower" HSGPA averages. These results are shown in Fig. 2 for the g scores. We also investigated the criterion-related validity of each subtest (and the g scores) against the HSGPAs for different areas of coursework. These validity coefficients are shown in Table S10.

To address Hypothesis 2-2, we used the Cleary (1968) model to examine the fairness of the g scores and each of the 12 subtests for Males–Females, Whites–Hispanics, and Whites–Blacks/African Americans. Power analyses were conducted to determine the power for detecting changes in the intercepts and slopes of .20 (which roughly represents a 50% decrease in validity) and using standardized predictors and criteria for the g scores (which exhibited the largest differences between the demographic groups). For the Males–Females, Whites–Hispanics, and Whites–Blacks/African Americans analyses, the power for intercepts and slopes exceeded .99. For the two remaining groups (Asians/Pacific Islanders and Native Americans) the power was too low, ranging from .23 to .50, therefore, we did not conduct Cleary analyses for these groups.

In Table 7, we present the ΔR^2 and ΔR for intercept and slope differences for each group. We present plots of the different groups for the MK subtest and g scores in Fig. 3. Given the large sample size, nearly all of the differences were statistically significant; however, most were practically small. None of the ΔR for the slope differences for Hispanics and Blacks exceeded .01 and none of the ΔR for the intercept differences exceeded .15. There was evidence of mild intercept bias (but not slope bias) for Females, with evidence of underprediction. The intercept differences (in ΔR units) for Females ranged from .02 to .12, with two largest differences for AI (.11) and SI (.12). Thus, we found partial support for Hypothesis 2-2.

Table 9
Criterion-related validity for within-group norming.

Ability	r_{obs}	p	Gender			White-Black			White-Hispanic		
			r_{wign}	p	q	r_{wign}	p	q	r_{wign}	p	q
GS	.32	<.001	.34	<.001	−.02	.27	<.001	.05	.29	<.001	.03
AR	.39	<.001	.39	<.001	<.01	.33	<.001	.07	.36	<.001	.03
WK	.31	<.001	.32	<.001	−.01	.26	<.001	.05	.27	<.001	.04
PC	.38	<.001	.36	<.001	.02	.34	<.001	.05	.36	<.001	.02
NO	.18	<.001	.17	<.001	.01	.16	<.001	.02	.17	<.001	.01
CS	.18	<.001	.17	<.001	.01	.16	<.001	.02	.16	<.001	.02
AI	.13	<.001	.16	<.001	−.03	.08	<.001	.05	.06	.001	.07
SI	.12	<.001	.16	<.001	−.04	.02	.143	.10	.03	.055	.09
MK	.43	<.001	.42	<.001	.01	.39	<.001	.05	.41	<.001	.02
MC	.29	<.001	.32	<.001	−.03	.23	<.001	.06	.25	<.001	.04
EI	.23	<.001	.26	<.001	−.03	.17	<.001	.06	.18	<.001	.05
AO	.34	<.001	.33	<.001	.01	.29	<.001	.06	.32	<.001	.02
g	.44	<.001	.45	<.001	−.01	.39	<.001	.06	.40	<.001	.05

Note. Cohen (1992) describes the use of q statistics as a measure of the effect size of the difference between two correlation coefficients. A q statistic is the difference between two Fisher-z transformed correlation coefficients and values of .10, .30, and .50 are considered small, medium, and large by Cohen.

Table 10
g-Loadings of ASVAB subtests and GPAs.

Variable	g-Loading	
	PCA	SEM
GS	.78	.71
AR	.78	.75
WK	.76	.74
PC	.76	.74
NO	.33	.30
CS	.39	.34
AI	.46	.40
SI	.52	.46
MK	.77	.75
MC	.73	.68
EI	.67	.61
AO	.66	.62
HSGPA-F	.44	.49
HSGPA-Jr. year	.43	.50
GPA-English	.37	.42
GPA-Foreign Lang.	.30	.37
GPA-Math	.34	.38
GPA-Social Sci.	.38	.43
GPA-Life/Phys. Sci.	.37	.41

Table 11
Exploratory analysis showing linearity of the relationship between mental abilities and HSGPA.

Ability	Linear			Quadratic			Cubic			Reshef et al.		
	ΔR^2	ΔR	p	ΔR^2	ΔR	p	ΔR^2	ΔR	p	MIC	$MIC - R^2_{linear}$	$MIC - R^2_{L,Q,C}$
GS	.11	.33	<.001	<.01	<.01	.009	<.01	<.01	.004	.13	.02	−.08
AR	.15	.39	<.001	.01	.01	<.001	<.01	<.01	.011	.16	.01	−.15
WK	.10	.31	<.001	.01	.01	<.001	<.01	<.01	.836	.12	.03	−.08
PC	.15	.38	<.001	<.01	.01	<.001	<.01	<.01	.086	.17	.02	−.13
NO	.03	.18	<.001	.03	.07	<.001	<.01	<.01	.028	.13	.10	.04
CS	.03	.18	<.001	.01	.03	<.001	<.01	.01	.001	.11	.08	.03
AI	.02	.13	<.001	.02	.06	<.001	<.01	<.01	.303	.08	.06	.03
SI	.01	.12	<.001	.01	.03	<.001	<.01	<.01	.033	.07	.06	.04
MK	.19	.43	<.001	<.01	.01	<.001	<.01	<.01	.005	.18	−.01	−.20
MC	.09	.29	<.001	<.01	<.01	.458	<.01	<.01	.046	.11	.03	−.06
EI	.05	.23	<.001	<.01	.01	.001	<.01	<.01	.061	.10	.05	−.01
AO	.12	.34	<.001	<.01	.01	<.001	<.01	<.01	.006	.13	.02	−.10
g	.19	.44	<.001	<.01	<.01	.003	<.01	<.01	.001	.18	−.01	−.21

Note. The Reshef et al. (2011) MIC (Maximal Information Coefficient) is a new method for detecting non-linear relationships between variables. According to the authors, the MIC statistic includes both linear and non-linear relationship between two variables and the equation MIC − R2 represents the non-linearity in the relationship that is not modeled by the Pearson r statistic. Two MIC − R2 statistics are included in the table, one (R^2_{linear}) represents MIC minus the R2 from the linear relationship between HSGPA and the ASVAB variables. The other R2 ($R^2_{L,Q,C}$) represents the R2 from the multiple regression equation including the linear, quadratic, and cubic terms. Thus, the MIC − $R^2_{L,Q,C}$ statistic represents the relationship between the ASVAB variables and HSPGA that is not accounted for by the linear, quadratic, and cubic terms.

Hypothesis 2-3 was addressed by adding points to the test scores for minority groups so that their means would now equal those of the majority group. The results for this analysis are in Tables 8 and 9. Within-group norming (WIGN) caused only slight reductions in the validity coefficients, as the *q* statistics mostly fell below Cohen's (1992) cutoff of .10 for a small effect. With respect to Hypothesis 2-4, HSGPA had high *g*-loadings (.44 and .49) indicating that it is substantially *g*-loaded (see Table 10). In fact, it had nominally higher *g*-loadings than 2–4 of the subtests. To answer Hypothesis 2-5 we computed coefficient alphas for HSGPAY1 through HSGPAY4, observing a value of .89 and a standard error of measurement of 0.20.

As an exploratory analysis, we examined whether or not the relationship between HSGPA and the test scores were linear. As shown in Table 11, the relationships were primarily linear (especially for the *g* scores) with evidence of a minor U-shaped relationship for NO and an inverted U-shaped relationship for AI. An anonymous reviewer inquired if ability tilt (i.e., math scores minus verbal scores; Coyle, Purcell, Snyder, & Richmond, 2014; Lubinski, Webb, Morelock, & Benbow, 2001; Park, Lubinski, & Benbow, 2007) added incremental validity over *g* — it did not (see Table S11). A reviewer also asked if the results would change if a bifactor model was used. The bifactor model was more likely to produce Heywood (1931) cases and it prevented the extraction of factors which only had two indicators. Therefore, we concentrated our results on the higher-order models. This being said, there were three subtests that had substantially larger standardized paths to HS grades and HSGPA when the bifactor model was used instead of the higher order regression models: T290ABST (the path increased from −.04 to .28), T107PHYS (the path increased from .02 to .49), and ASVAB-NO (the path increased from .03 to .24). One anonymous reviewer asked if the missing data in both studies biased the results in any way. A number of missing values analyses indicated that missing data had no effect on the overall results.[1]

4. Discussion

There are four main theoretical implications of our research. First, the classic multiplicative model between motivation (defined here as conscientiousness) and ability does not apply to academic performance; Sackett, Gruys, and Ellingson (1998) reported similar results for job performance. Second, specific aptitude theory received further disconfirmation, this time for academic performance. The main exception was that mathematics knowledge emerged as a unique predictor of *HS grades/GPA* in both studies. This finding is consistent with Coyle and Pillow (2008) who reported that the criterion-related validity of the ACT and SAT can be partially attributed to unique variance associated with mathematical knowledge (which the SAT and ACT math tests are closely linked to). It could be possible that mathematics knowledge is important for more high school courses (e.g., biology, chemistry, physics, business/accounting) than the other narrow abilities or that it invokes another ability which is related to performance in multiple types of courses. Nevertheless, the reason why mathematics knowledge adds incremental validity over *g* remains unknown and should be investigated by future researchers.

Third, there is some evidence of underprediction bias for females, meaning that the ASVAB subtests predict HSGPAs that, on average, were lower than observed when a common regression equation was used for males and females. This was especially true for the AI and SI subtests, and less of an issue with the *g*-scores. One possible explanation could be that the AI and SI are somewhat gender-biased due to their measurement of automotive and shop ability, which might not be well-reflected in HSGPA (e.g., not all schools might cover this material in courses that both males and females take). Operationally, this could be remedied by avoiding tests similar to AI and SI or by adding a constant value to females' scores on these two tests (although this would be illegal in U.S. employment settings). Fourth, HSGPA had reliability coefficients in the .80s, which are reasonably acceptable and consistent with those for undergraduate GPA and training performance. However, the reliability of HSGPA was less than that for a unit-weighted linear composite of standardized scores from the ASVAB (.97) as well as the SAT-Total (.90; College Board, 2011) and ACT-Total (.96; The ACT, 2007). That said, researchers examining the relationship between HSGPA and undergraduate GPA should consider correcting HSGPA for unreliability to estimate the true-score relationship between the HSGPA and undergraduate GPAs.

As for practical implications, colleges using HSGPA in admissions settings are indirectly measuring *g* and conscientiousness. Thus, colleges that drop the SAT or ACT in admissions may still be selecting applicants using a measure of *g* (i.e., HSGPA). Additionally, our findings shed light on which mental abilities (i.e., *g* and mathematics knowledge) predict HSGPA and could be measured in admissions tests for private, "charter," and "magnet" high schools for which students compete for admissions (often based on test scores).

[1] These analyses included a test of the relationships between dummy variables depicting whether a variable has a missing value and the numerical values for all other variables (see Tabachnick & Fidell, 2001), a SEM using full information maximum likelihood (FIML) analysis in AMOS, a comparison of the original correlation coefficients to those from the corFIML routine in the psyc package (Revelle, 2012) for R (Ihaka & Gentleman, 1996) and those from the expectation maximization (EM) described by Tabachnick and Fidell (2001) and implemented by Honaker, King, and Blackwell's (2011) Amelia II package in R (Ihaka & Gentleman, 1996).

In terms of societal implications, group differences continue to be a concern among users of standardized tests and our finding that these differences can be eliminated with only slight reductions in validity is an interesting one. However, we note that the use of WIGN presents legal and philosophical issues that require input and debate from non-psychologists.

One potential limitation of our estimate for the reliability of HSGPA is that we did not attempt to correct for the quality and rigor of the curriculum at each high school. However, most undergraduate institutions do not make such corrections when examining applications. Given the large number of high schools in the United States (~24,651 National Center for Education Statistics, 2011) it would be difficult to do so. Additionally, highly-qualified students attending schools of lesser quality would have their grades penalized, perhaps unfairly and with adverse impact against protected minority groups.

Future researchers should examine the role of *g* and specific abilities in predicting academic performance in primary, middle, and undergraduate schools. Additionally, Carroll's (1993) model included 69 specific abilities (far more than can be measured in any one study) and ones we did not study could be examined by future researchers.

Appendix A. Supplementary data

Supplementary data to this article, including Tables S1 through S11, can be found online at http://dx.doi.org/10.1016/j.intell.2015.11.007.

References

Aguinis, H., Culpepper, S.A., & Pierce, C.A. (2010). Revival of test bias research in preemployment testing. *Journal of Applied Psychology*, *95*(4), 648–668.

American Institutes for Research (1960). *Project TALENT, base year data, 1960: Codebook for combined classes.* Washington, DC: Author.

American Institutes for Research. Project Talent, Base Year Data, 1960. ICPSR33341-v2. Ann Arbor, MI: Inter-University Consortium for Political and Social Research [distributor], 2013-05-23. doi:10.3886/ICPSR33341.v2. Retrieved July 8, 2013 from http://dx.doi.org/10.3886/ICPSR33341.v2.

Anderson, R. (1971). A study of the predictability of high school grades and the Differential Aptitude Tests for success in vocational programs in health careers. Unpublished Masters Thesis, Wisconsin University. ERIC No. ED068697.

Armstrong, T. (1994). Multiple intelligences: Seven ways to approach curriculum. *Educational Leadership: Strategies for Success*, *52*(3), 26–28.

Arneson, J.J., Sackett, P.R., & Beatty, A.S. (2011). Ability-performance relationships in education and employment settings: Critical tests of the more-is-better and the good-enough hypotheses. *Psychological Science*, *22*(1), 1336–1342.

Barrick, M.R., & Mount, M.K. (1991). The Big Five personality dimensions and job performance. *Personnel Psychology*, *44*, 1–26.

Blythe, T., & Gardner, H. (1990). A School for all Intelligences. *Educational Leadership*, *47*(7), 33–37.

Brody, N. (1992). *Intelligence.* San Diego, CA: Academic Press.

Brown, K.G., Le, H., & Schmidt, F.L. (2006). Specific aptitude theory revisited: Is there incremental validity for training performance? *International Journal of Selection and Assessment*, *14*(2), 87–100.

Campbell, E.A. (1979). *Publications based on Project TALENT Data: An annotated bibliography.* Palo Alto, CA: American Institutes for Research.

Camara, W., & Michaelides, M. (2005). *AP® use in admissions: A response to Geiser and Santelices.* New York, NY: The College Board.

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York, NY: Cambridge University Press.

Cattell, R.B. (1957). *Personality and motivational structure and measurement.* World Book: Yonkers-on-Hudson.

Cattell, R.B. (1973). *Personality and mood by questionnaire.* San Francisco: Jossey-Bass.

Chamorro-Premuzic, T., Harlaar, N., Greven, C.U., & Plomin, R. (2010). More than just IQ: A longitudinal examination of self-perceived abilities as predictors of academic performance in a large sample of UK twins. *Intelligence*, *38*(4), 385–392.

Chamorro-Premuzic, T., Quiroga, M.A., & Colom, R. (2009). Intellectual competence and academic performance: A Spanish study. *Learning and Individual Differences*, *19*(4), 486–491.

Civil Rights Act of 1991, Pub. L. No. 102–166, 105 Stat 1071.

Cleary, T.A. (1968). Test bias: Prediction of grades of Negro and White students in integrated colleges. *Journal of Educational Measurement*, *5*, 115–124.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159.

Colberg, M. (1985). Logic-based measurement of verbal reasoning: A key to increased validity and economy. *Personnel Psychology*, *38*, 347–359.

Colberg, M., Nester, M.A., & Trattner, M.H. (1985). Convergence of the inductive and deductive models in the measurement of reasoning abilities. *Journal of Applied Psychology*, *70*, 681–694.

College Board (2011). Test characteristics of the SAT: Reliability, difficulty levels, completion rates. Retrieved from http://media.collegeboard.com/digitalServices/pdf/research/Test-Characteristics-of%20-SAT-2012.pdf

Conry, R., & Plant, W.T. (1965). WAIS and group test prediction of an academic success criterion: High school and college. *Educational and Psychological Measurement*, *25*(2), 493–500.

Coyle, T.R., & Pillow, D.R. (2008). SAT and ACT predict college GPA after removing g. *Intelligence*, *36*(6), 719–729.

Coyle, T.R., Purcell, J.M., Snyder, A.C., & Richmond, M.C. (2014). Ability tilt on the SAT and ACT predicts specific abilities and college majors. *Intelligence*, *46*, 18–24.

Cucina, J.M., Su, C., Busciglio, H.H., & Thompson Peyton, S. (2015). Something more than g: Meaningful memory uniquely predicts training performance. *Intelligence*, *49*, 192–206.

Cureton, E.E. (1968). *A factor analysis of Project TALENT tests and four other batteries.* Washington, DC and Pittsburgh, PA: American Institutes for Research and School of Education, University of Pittsburgh.

Dastgoshadeh, A., & Jalilzadeh, K. (2011). Multiple intelligences-based language curriculum for the third millennium. *International Conference on Education, Research and Innovation: IPEDR* (pp. 18).

de Raad, B., & Schouwenburg, H.C. (1996). Personality in learning and education: A review. *European Journal of Personality*, *10*, 303–336.

Digman, J.M. (1990). Personality structure: Emergence of the Five-Factor Model. *Annual Review of Psychology*, *41*, 417–440.

Digman, J.M., & Takemoto-Chock, N.K. (1981). Factors in the natural language of personality: Re-analysis, comparison, and interpretation of six major studies. *Multivariate Behavioral Research*, *16*, 149–170.

DiMaggio, P. (1982). Cultural capital and school success: The impact of status culture participation on the grades of U.S. high school students. *American Sociological Review*, *47*(2), 189–201.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160.

Faul, F., Erdfelder, E., Lang, A. -G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175–191.

Feinman, J. (2008). *High stakes, but low validity? A case study of standardized tests and admissions into New York city specialized high schools.* Boulder, CO and Tempe, AZ: Education and the Public Interest Center & Education Policy Research Unit Retrieved July 8, 2013 from http://epicpolicy.org/publication/high-stakes-but-low-validity.

Flanagan, J.C., Dailey, J.T., Shaycroft, M.F., Gorham, W.A., Orr, D.B., & Goldberg, I. (1962). *Design for a study of American youth.* Boston, MA: Houghton Mifflin.

Flanagan, J.C., Dailey, J.T., Shaycroft, M.F., Gorham, W.A., Orr, D.B., Goldberg, I., & Neyman, C.A., Jr. (1961). *Project TALENT.* Washington, DC: Project Talent Office, American Institutes for Research.

Flanagan, J.C., Davis, F.B., Dailey, J.T., Shaycroft, M.F., Orr, D.B., Goldberg, I., & Neyman, C.A., Jr. (1964). *The American high-school student: The identification, development, and utilization of human talents.* Pittsburgh, PA: Project TALENT Office, University of Pittsburgh.

Flynn, J.R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, *95*, 29–51.

Flynn, J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*, 171–191.

Flynn, J.R. (2007). *What is intelligence? Beyond the Flynn effect.* New York, NY: Cambridge University Press.

Flynn, J.R. (2012). *Are we getting smarter? Rising IQ in the twenty-first century.* New York, NY: Cambridge University Press.

Furnham, A., & Monsen, J. (2009). Personality traits and intelligence predict academic school grades. *Learning and Individual Differences*, *19*(1), 28–33.

Furnham, A., Rinaldelli-Tabaton, E., & Chamorro-Premuzic, T. (2011). Personality and intelligence predict arts and science School results in 16 year olds. *Psychologia*, *54*(1), 39–51.

Gardner, H. (1983). *Frames of mind.* New York: Basic Books Inc.

Gardner, H., & Hatch, T. (1989). Multiple intelligences go to school: Educational implications of the theory of multiple intelligences. *Educational Researcher*, *18*(8), 4–9.

Gesier, S., & Santelices, M.V. (2007). Validity of high-school grades in predicting student success beyond the Freshman year: High-school record vs. standardized tests as indicators of four-year college outcomes. *Research & Occasional Paper Series, 6.07* Berkley, CA: University of California, Berkley — Center for Studies in Higher Education.

Goldberg, L.R. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality psychology in Europe. Vol. 7*. Tilburg, Netherlands: Tilburg University Press.

Goldberg, L.R., Johnson, J.A., Eber, H.W., Hogan, R., Ashton, M.C., Cloninger, C.R., & Gough, H.G. (2006). The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, *40*, 84–96.

Gottfredson, L.S. (2002). Where and why g matters: Not a mystery. *Human Performance*, *15*(1/2), 25–46.

Grigorenko, E.L., Jarvin, L., Diffley, R., III, Goodyear, J., Sternberg, R.J., & Shanahan, E.J. (2009). Are SSATs and GPA enough? A theory-based approach to predicting academic success in secondary school. *Journal of Educational Psychology*, *101*(4), 964–981.

Heywood, H.B. (1931). On finite sequences of real number. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 134(824).* (pp. 486–501).

Hezlett, S.A., Kuncel, N.R., Vey, M.A., Ahart, A.M., Ones, D.S., Campbell, J.P., & Camara, W.J. (2001, April). The effectiveness of the SAT in predicting success early and late in

college: A comprehensive meta-analysis. *Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle, WA, April 2001*.
Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, *45*(7), 1–47.
Howard Gardner Multiple Intelligence School (2010). *Curriculum guide.* Vancouver WA: Author Retrieved July 8, 2013 from http://howardgardnerschool.com/wp-content/uploads/2012/07/Curriculum-Guide.pdf.
Hunter, J.E. (1983a). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In F. Landy, S. Zedick, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 257–266). Hillsdale, NJ: Erlbaum.
Hunter, J.E. (1983b). *The prediction of job performance in the military using ability composites: The dominance of general cognitive ability over specific aptitudes.* Report for Research Applications, Inc (in partial fulfillment of DOD Contract F41689-83-C-0025).
Hunter, J.E. (1984). *The validity of the Armed Forces Vocational Aptitude Battery (ASVAB) high school composites.* Report for Research Applications, Inc (in partial fulfillment of DOD Contract F41689-83-C-0025).
Hunter, J.E. (1985). *Differential validity across jobs in the military.* Report for Research Applications, Inc (in partial fulfillment of DOD Contract F41689-83-C-0025).
Hunter, J.E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*, 340–362.
Ihaka, R., & Gentleman, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, *5*, 299–314.
Jensen, A.R. (1980). *Bias in mental testing.* New York, NY: The Free Press.
Jensen, A.R. (1986). g: Artifact or reality? *Journal of Vocational Behavior*, *29*, 301–331.
Jensen, A.R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger Publishers.
Jensen, A.R., & Weng, L. (1994). What is a good g? *Intelligence*, *18*, 231–258.
Johnson, W., & Bouchard, T.J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence*, *33*(4), 393–416.
Keith, T.Z., & Reynolds, M.R. (2010). Cattell–Horn–Carroll abilities and cognitive tests: What we've learned from 20 years of research. *Psychology in the Schools*, *47*(7), 635–665.
Kell, H.J., Lubinski, D., Benbow, C.P., & Steiger, J.H. (2013). Creativity and technical innovation: Spatial ability's unique role. *Psychological Science*, *24*, 1831–1836.
Kobrin, J.L., Sinharay, S., Haberman, S.J., & Chajewski, M. (2011). An investigation of the fit of linear regression models to data from an SAT validity study. *College Board Research Report*, *3*, 1–21 (Retrieved July 16, 2013).
Kostal, J.W., Kuncel, N.R., & Sackett, P.R. (2014). Grade inflation marches on: Grade increases across 1990s–2000s. *Poster presented at the 29th meeting of the Society for Industrial and Organizational Psychology, Honolulu, HI*.
Kuncel, N.R., Crede, M., & Thomas, L.L. (2005). 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*, *75*(1), 63–82.
Kuncel, N.R., & Hezlett, S.A. (2007). Standardized tests predict graduate students' success. *Science*, *315*, 1080–1081.
Kuncel, N.R., Hezlett, S.A., & Ones, D.S. (2001). A comprehensive meta-analysis of the predictive validity of the Graduate Record Examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, *127*, 162–181.
Kuncel, N.R., Hezlett, S.A., & Ones, D.S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*, 148–161.
Kunin, T. (1965). Another look at validation. *The Industrial Psychologist*, *2*(2), 36–38.
Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences*, *49*, 344–351.
Lubinski, D., Webb, R.M., Morelock, M.J., & Benbow, C.P. (2001). Top 1 in 10,000: A 10-year follow-up of the profoundly gifted. *Journal of Applied Psychology*, *86*, 718–729.
Maier, N.R.F. (1955). *Psychology in industry: A psychological approach to industrial problems.* Boston: Houghton Mifflin.
Major, J.T., Johnson, W., & Deary, I.J. (2012). Comparing models of intelligence in Project TALENT: The VPR model fits better than the CHC and extended Gf–Gc models. *Intelligence*, *40*, 543–559.
Mattern, K.D., & Patterson, B.F. (2013). Test of slope and intercept bias in college admissions: A response to Aguinis, Culpepper, and Pierce (2010). *Journal of Applied Psychology*, *98*(1), 134–147.
McAbee, S.T., & Oswald, F.L. (2013). The criterion-related validity of personality measures for predicting GPA: A meta-analytic validity competition. *Psychological Assessment*, *25*(2), 532–544.
McGrew, K.S. (2005). The Cattell–Horn–Carroll theory of cognitive abilities: Past, present, and future. In D.P. Flanagan, & P.L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 136–181) (2nd ed.). New York, NY: Guilford.
McGrew, K.S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10.
McGrew, K.S., & Evans, J.J. (2004). Internal and external factorial extensions to the Cattell–Horn–Carroll (CHC) theory of cognitive abilities: A review of factor analytic research since Carroll's seminal 1993 treatise. *(Carroll Human Cognitive Abilities Project Research Report No. 2). Institute for Applied Psychometrics*.
New York City (2010). *Specialized high schools student handbook: 2010–2011.* New York, NY: Department of Education of the City of New York.
National Center for Education Statistics (2011). Digest of Education Statistics: 2011. U.S. Department of Education. Retrieved from http://nces.ed.gov/programs/digest/d11/index.asp. Retrieved July 16, 2013.
Norman, W.T. (1963). Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, *66*(6), 574–583.
Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill, Inc.
Olea, M.M., & Ree, M.J. (1994). Predicting pilot and navigator criteria: Not much more than g. *Journal of Applied Psychology*, *79*, 845–851.
Park, G., Lubinski, D., & Benbow, C.P. (2007). Contrasting intellectual patterns predict creativity in the arts and sciences: Tracking intellectually precocious youth over 25 years. *Psychological Science*, *18*(11), 948–952.
Ramist, L., Lewis, C., & McCamley, L. (1990). Implications of using freshman GPA as the criterion for the predictive validity of the SAT. In W.W. Willingham (Eds.), *Predicting college grades: An analysis of institutional trends over two decades* (pp. 253–288). Princeton, NJ: Educational Testing Service.
Raven, J., Raven, J.C., & Court, J.H. (2004). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* San Antonio, TX: Harcourt Assessment.
Ree, M.J., & Earles, J.A. (1991). Predicting training success: Not much more than g. *Personnel Psychology*, *44*, 321–332.
Ree, M.J., Earles, J.A., & Teachout, M. (1994). Predicting job performance: Not much for than g. *Journal of Applied Psychology*, *79*, 518–524.
Reeve, C.L. (2001). Using latent variable anlaysis to explore the contribution of general and specific abilities in academic achievement: Is there more than G? Unpublished Doctoral Dissertation. Bowling Green, OH: Bowling Green State University.
Reeve, C.L. (2004). Differential ability antecedents of general and specific dimensions of declarative knowledge: More than g. *Intelligence*, *32*(6), 621–652.
Reeve, C.L., Meyer, R.D., & Bonaccio, S. (2006). Intelligence-personality associations reconsidered: The importance of distinguishing between general and narrow dimensions of intelligence. *Intelligence*, *34*, 387–402.
Reshef, D.N., Reshef, Y.A., Finucane, H.K., Grossman, S.R., McVean, G., Turnbaugh, P.J., ... Sabeti, P.C. (2011). Detecting novel associations in large data sets. *Science*, *334*, 1518–1524.
Revelle, W. (2012). Procedures for psychological, psychometric, and personality research. *Computer software manual*. Evanston, IL: Northwestern University Retrieved from http://www.personality-project.net/r/psychmanual.pdf (R package version 1.3. 2)).
Rosse, J.G., Stecher, M.D., Levin, R.A., & Miller, J.L. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, *83*(4), 634–644.
Sackett, P.R., & Wilk, S.L. (1994). Within-group norming and other forms of score adjustment in pre-employment testing. *American Psychologist*, *49*, 929–954.
Sackett, P.R., Gruys, M.L., & Ellingson, J.E. (1998). Ability-personality interactions when predicting job performance. *Journal of Applied Psychology*, *83*(4), 545–556.
Schmidt, F.L. (1988). The problem of group differences in ability test scores in employment selection. *Journal of Vocational Behavior*, *33*(3), 272–292.
Schmidt, F.L., & Hunter, J.E. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods*, *1*(2), 199–223.
Schmidt, F.L., & Hunter, J.E. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*(1), 162–173.
Schmidt, F.L., Ones, D.S., & Hunter, J.E. (1992). Personnel selection. *Annual Review of Psychology*, *43*, 627–667.
Schmidt, F.H., Hunter, J.E., & Caplan, J.R. (1981). Validity generalization results for two job groups in the petroleum industry. *Journal of Applied Psychology*, *66*(5), 261–273.
Schmidt, F.L. (2011). An interview with Frank L. Schmidt. *The Industrial–Organizational Psychologist*, *48*(4), 21–29.
Schmidt, F.L., & Hunter, J.E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, *124*, 262–274.
Seberhagen, L. (2007). Technical affairs: Beauty may be in the eye of the beholder, but is the same true of a validity coefficient? In M. Aamodt (Ed.), *International Public Management Association for Human Resources: Assessment Council News. 5.* (pp. 2–10) Comment in.
Secondary Schools Admissions Test Board (2012). *All about the Middle and Upper Level SSAT for students.* Princeton, NJ: Author.
Seider, S., & Geiger, R. (2009). *Teachers are taking multiple-intelligences theory to heart.* San Rafael, CA: Edutopia, The George Lucas Educational Foundation Retrieved July 8, 2013 from http://www.edutopia.org/multiple-intelligences-schools-strategies.
Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, *15*, 201–293.
Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York, NY: Macmillan.
Strauss, V. (2013, February 26). SAT to be redesigned. The *Washington* Post. Retrieved from http://www.washingtonpost.com/blogs/answer-sheet/wp/2013/02/26/sat-exam-to-be-redesigned Retrieved July 16, 2013
Su, R. (2012). *The power of vocational interests and interest congruence in predicting career success. Unpublished doctoral dissertation.* Urbana, IL: University of Illinois at Urbana-Champaign.
Tabachnick, B.G., & Fidell, L.S. (2001). *Using multivariate statistics.* New York, NY: Pearson.
The ACT (2007). The ACT technical manual. Retrieved from http://www.act.org/aap/pdf/ACT_Technical_Manual.pdf Retrieved July 17, 2013
Thorndike, R.L. (1985). The central role of general ability in prediction. *Multivariate Behavioral Research*, *20*(3), 241–254.
Thorndike, R.L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior*, *29*, 332–339.
Tupes, E.C., & Christal, R.F. (1961). *Recurrent personality factors based on trait ratings (USAF WADC Tech. Note No. 58–61).* Lackland Air Force Base, TX: U.S. Air Force.
Tupes, E.C., & Christal, R.F. (1992). Recurrent personality factors based on trait ratings. *Journal of Personality*, *60*(2), 225–251 (Original work published 1961).
U.S. Bureau of Labor Statistics, U.S. Department of Labor. (2005). National Longitudinal Survey of Youth 1997 cohort, 1997–2003 (rounds 1–7) [computer file]. Produced

by the National Opinion Research Center, the University of Chicago and distributed by the Center for Human Resource Research, The Ohio State University. Columbus, OH: 2005.

Vroom, V.H. (1960). Some personality determinants of the effects of participation. *Englewood Cliffs*. N.J.: Prentice-Hall.

Vroom, V.H. (1964). *Work and motivation.* New York: Wiley.

Waters, S.D. (2007). General vs. specific ability tradeoffs in personnel selection: Consequences for validity and subgroup differences. Unpublished Doctoral Dissertation. Minneapolis, MN: University of Minnesota.

Wai, J., Lubinski, D., & Benbow, C.P. (2009). Spatial ability for STEM domains: Aligning over fifty years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*, 817–835.

Webb, E. (1915). Character and intelligence. *British Journal of Psychology Monographs*, *1*(3).

Webb, R.M., Lubinski, D., & Benbow, C.P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, *99*, 397–420.

Wise, L., McLaughlin, D.H., & Steel, L. (1979). *The Project TALENT data bank handbook.* Palo Alto, CA: American Institutes for Research.